Research Article

X-ray image based pneumonia classification using convolutional neural networks

Sarah Badr AlSumairi[*] and Mohamed Maher Ben Ismail

King Saud University, College of Computer and Information Sciences, Riyadh, Saudi Arabia

Abstract
Pneumonia is an infectious disease of the lungs. About one third to one half of pneumonia cases are caused by bacteria. Early diagnosis is a critical factor for a successful treatment process. Typically, the disease can be diagnosed by a radiologist using chest X-ray images. In fact, chest X-rays are currently the best available method for diagnosing pneumonia. However, the recognition of pneumonia symptoms is a challenging task that relies on the availability of expert radiologists. Such "human" diagnosis can be inaccurate and subjective due to lack of clarity and erroneous decision. Moreover, the error can increase more if the physician is requested to analyze tens of X-rays within a short period of time. Therefore, Computer-Aided Diagnosis (CAD) systems were introduced to support and assist physicians and make their efforts more productive. In this paper, we investigate, design, implement and assess customized Convolutional Neural Networks to overcome the image-based Pneumonia classification problem. Namely, ResNet-50 and DenseNet-161 models were inherited to design customized deep network architecture and improve the overall pneumonia classification accuracy. Moreover, data augmentation was deployed and associated with standard datasets to assess the proposed models. Besides, standard performance measures were used to validate and evaluate the proposed system.

Keywords
Pneumonia, Machine learning, Deep learning, Convolution neural networks, Data augmentation.

1.Introduction
Pneumonia is an infectious disease of the lungs, about one third to one half of pneumonia cases are caused by bacteria [1]. In fact, the risk of this disease is doubled, especially in a developing country where millions face poverty and depend on the polluting forms of energy. Regrettably, the World Health Organization (WHO) states that 4 million deaths of premature happen annually from pneumonia and air pollution-related disease. On an annual basis, over 150 million people are infected with pneumonia, especially children under five years old. Furthermore, imaging examination is important for diagnosing and detecting this disease. The main role of imaging examinations is to confirm the pneumonia diagnosis, so if the patient suffers from cough or fever, the diagnosis of pneumonia can be made. Using images in examination has a complementary role in evaluating treatment effects. Imaging examination is important as it helps doctors determine the initial treatment of pneumonia.

Again, imaging examination has its role in differentiating between infectious and non-infectious pneumonia. Additionally, the chest X-ray helps the doctor determine and diagnose the location and the extent of infection. It also helps the doctor know the kind of germ that caused pneumonia to the patient. Blood tests are also used to detect the organism that caused this infection to identify a precise diagnosis of the disease. Moreover, x-rays allow doctors to see the patient's lungs, blood vessels and the heart to determine if the patient has pneumonia or not. If the radiologist finds white spots in the lung, this means that the patient has an infection [2]. One should mention that physicians can also use sputum tests by taking a sample of fluid from the patient's lung after deep coughing to be analyzed to know the infection cause [3−5].

Early diagnosis is a critical factor for a successful treatment process. Typically, the disease can be diagnosed by an expert radiologist using chest X-ray images [6]. More than one million adults are hospitalized with pneumonia and yearly 50,000 of

---

*Author for correspondence

them die consequently in the US alone [6]. For diagnosing pneumonia, the current most reliable method is chest X-rays. In addition to what is mentioned above, Computer-Aided Diagnosis (CAD) is helpful to doctors in the interpretation of the medical images. CAD systems show images for typical appearance to clarify conspicuous sections. For this reason, the main goal of computer-aided diagnosis is to detect the disease accurately by reducing the negative false rates. Computer-aided diagnosis is increasingly used in the radiology mainstream and it has become a part and parcel of the clinical work for detecting breast cancer using mammograms. CAD is clinically related to detecting images such as lung nodule detection and mitosis detection [3].

The challenging task of recognizing pneumonia in chest X-rays requires the availability of expert radiologists [4]. However, such "human" diagnosis can be inaccurate and subjective. In fact, the disease symptoms can be unclear and yield an erroneous decision. Moreover, the error can increase more if the physician is requested to analyze tens of X-rays within a short period of time. Therefore, CAD systems were introduced to support and assist physicians and make their efforts more productive. Typically, CAD systems rely on image processing and machine learning techniques to automatically recognize Pneumonia cases using X-rays images. However, the accuracy of these CAD systems is constrained by the choice of the visual features used to represent the X-rays image as well as the performance of the machine learning method used to map the new cases to the positive or negative class.

The objectives of this research can be summarized as: (i) Survey existing Pneumonia detection and classification systems. (ii) Design an image-based deep learning system for automatic Pneumonia classification that is intended to learn the most relevant features to discriminate between "Pneumonia" and "Pneumonia-free", and (iii) Use a standard dataset and performance measures to validate and assess the proposed system.

The rest of this manuscript is organized as follows: A literature review relevant to Pneumonia detection and classification is conducted in section 2. Section 3 outlines the proposed system. The experiments are outlined in section 4. Finally, the report's conclusions and future works are presented in section 5.

2.Related works
2.1Deep learning-based solutions
The researchers in [4] developed a model to detect pneumonia as positive and negative if not, using a 121-layer convolutional neural network (DenseNet) [7]. Besides, they produced heat maps using Class Activation Mappings (CAMs) through looking at the output of the feature maps of the final convolutional layer to determine the most symptomatic image regions. The weights of the network were inherited from a pre-trained ImageNet [8] model. The authors in [5] addressed the pneumonia cases classification problem by developing a Convolutional Neural Network model. Their system starts by first augmenting and resizing the image collection. In [9], an ensemble of CNNs was introduced to overtake the pneumonia classification problem. Namely, RetinaNet [10] and Mask R-CNN [11] models were associated with the Feature Pyramid Network (FPN) algorithm. Specifically, RetinaNet was deployed as a primary model and Mask R-CNN as an auxiliary one that is intended to predict the pneumonia region. In addition, a non-maximum suppression (NMS) algorithm was used to aggregate the decisions of both models. The researchers in [12] introduced a CNN-based X-ray classification approach for Chest X-ray image classification. They first reduced the pixel depth to 8-bit, then fed the images into a ResNet-50 network. Additionally, non-image data such as patient gender, age, and acquisition type were used to build the deep learning model. In [13], an ensemble of two different deep convolutional neural networks (DCNNs) was outlined as a solution for tuberculosis detection in chest X-ray images. Namely, AlexNet [14] and GoogLeNet [15] models were associated to generate the final ensemble decision. Similarly, the authors in [16] developed an ensemble of Mask-RCNNs. Namely, they coupled a ResNet50 [17] and a ResNet101 [18] to address the pneumonia classification problem. The researchers in [19] introduced a CNN based feature extraction model followed by different shallow classification models intended to detect abnormal cases among chest X-Rays. Specifically, a DenseNet [7] model was adopted for feature extraction while the SVM algorithm was used for classification. In [6], Xception [20] and Vgg16 [21] models were used to classify pneumonia cases as positive or negative. The extreme version of the Inception model (Xception) consists of 36 convolutional layers for feature extraction followed by a logistic regression layer, the model is modified with depth wise separable convolutional layer and pointwise convolution.

Sarah Badr AlSumairi and Mohamed Maher Ben Ismail

On the other hand, VGG16 was pre-trained using ImageNet [8] dataset. In [22], a model to explore the benefits of deep residual learning along with separable convolution algorithm was introduced to detect pneumonia. The algorithm starts by removing noise through bilateral filtering to smooth the image while preserving edges. Then, a CNN is fed with the resulting X-ray images. This CNN model exhibits 4 distinct blocks intended to avoid the vanishing gradient problem. Note that each convolution layer in the model follows the depth-wise separable convolution algorithm. The authors in [23] outlined a model intended to automatically diagnose pediatric chest X-ray and know whether their pneumonia is viral or bacterial. They instigated various CNN models. Namely, they used a sequential CNN, a CNN with Inception modules (Inception CNN), a customized VGG16 [21] and a CNN with residual connections (Residual CNN). Particularly, lung region of interest (ROI) was pre-localized to avoid irrelevant feature representation. In [24], the researchers designed a cross-sectional system to assess model generalizability and by obtaining data from three separate hospital systems. Different CNN architectures such as ResNet-50 [17] and DenseNet-121 [7] were deployed along with an additional dense layer to facilitate extraction of bottleneck features in a reduced dimension. The authors in [25] investigated a DensNet based deep learning algorithm to detect the diseases in chest radiographs. Specifically, they aimed to solve 14-class classification problem based on X-ray images. In [26], chest pathologies classification problem using chest X-rays and deep learning was explored. Namely, convolutional neural network (CNN), Backpropagation neural network (BPNN) [27] and competitive neural network (CpNN) [28] were used to classify 12 common diseases. Besides, a histogram equalization is first performed as pre-processing phase prior to image segmentation. Then, feedforward neural network is deployed for the classification task. The researchers in [29] developed a model to detect pneumonia using pre-trained CNN models such as AlexNet [14], VGG-16 and VGG-19 [21]. They first performed data augmentation (brightness, rotation, cutting, width and height change, zooming and horizontal turning) to only normal images used to balance the distribution of samples over the classes. In [30], the authors proposed the consolidation of the state-of-the-art VGG16, DenseNet121 and pyramid convolutional structure [31] with an architecture design called ChestNet. Note that VGG16 and DenseNet121 were pre-trained using the ImageNet dataset that encloses 1.2 million color images [8].

56

2.2 Shallow model-based solutions

In [32], a logistic regression-based computer aided diagnosis system was introduced. The Logistic Regression baseline uses 32×32 image resolution, organizing the Google Cloud, and applying transfer learning. The authors in [33] used three classifiers, namely Naive Bayes, KNN and SVM, to detect pneumonia and improve their CAD system accuracy. The most accurate classifier was then used for childhood pneumonia detection. In particular, visual features such as coefficient of variation, contrast, correlation sum entropy and sum variance were extracted and fed into the different classification algorithms. The researchers in [34] proposed a model to classify pneumonia cases using DCNN. Additionally, they compared their system performance with ANN, SVM, Adaboost, logistic regression and Decision Tree based results.

3. Proposed approach

In this research, the proposed approach relies on transfer learning and investigates state-of-the-art pre-trained CNN models to overcome the image-based pneumonia classification problem. In addition, data augmentation is also deployed to improve the deep learning model generalization and address the overfitting problem.

In fact, the Deep Residual Learning for Image Recognition [17] was originally proposed to solve the gradient vanishing problem faced by CNN based architectures. In other words, the residual connections are intended to prevent information loss that may happen when training the deep networks. This yielded deeper layers for ResNet [21] compared to VggNet [17] and AlexNet [14]. The resent ResNet-50 architecture designed in this research in shown in *Figure 1*.

On the other hand, the feature reusability can be achieved using DenseNet. Particularly, the resulted condensed model is intended to be easily trainable and less complex. In fact, every layer in DenseNet takes additional inputs from all previous layers and passes on its feature-maps to all following layers [7]. *Figure 2* illustrates how we reduce the model size and complexity, by implementing BN-ReLU-1×1 Conv first then BN-ReLU-3×3 Conv.

In this research, we consider DenseNet169 which achieved promising performance on several applications. The proposed CNN architecture based on denseNet-161 is presented in *Figure 3*.
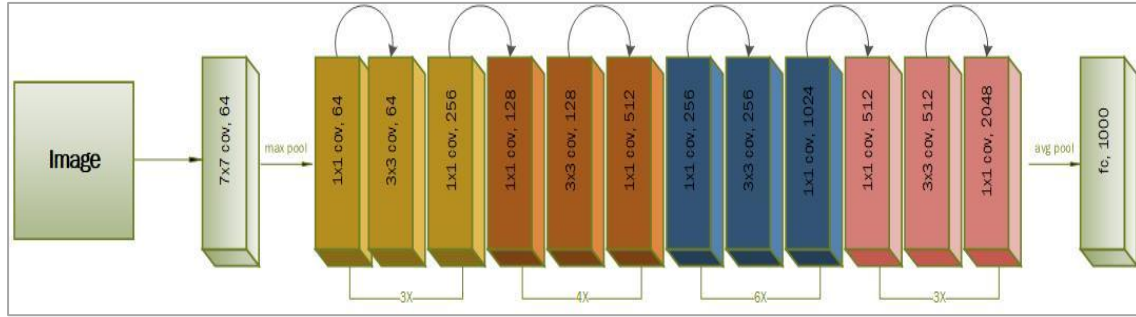
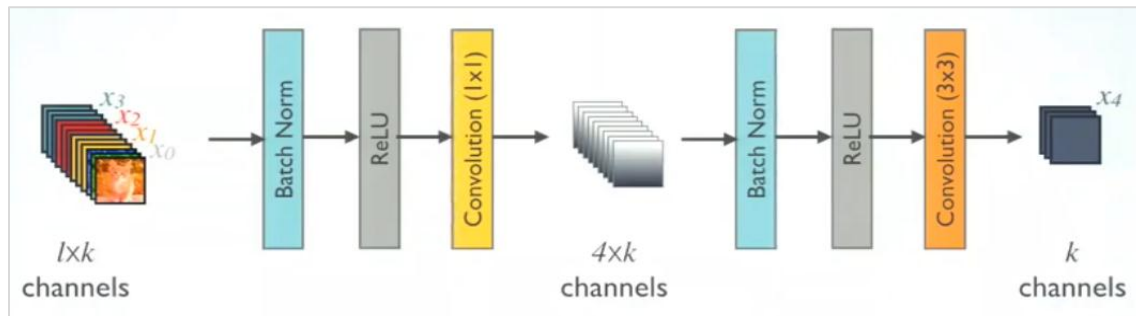**Figure 1** Resnet-50 architecture block diagram representation

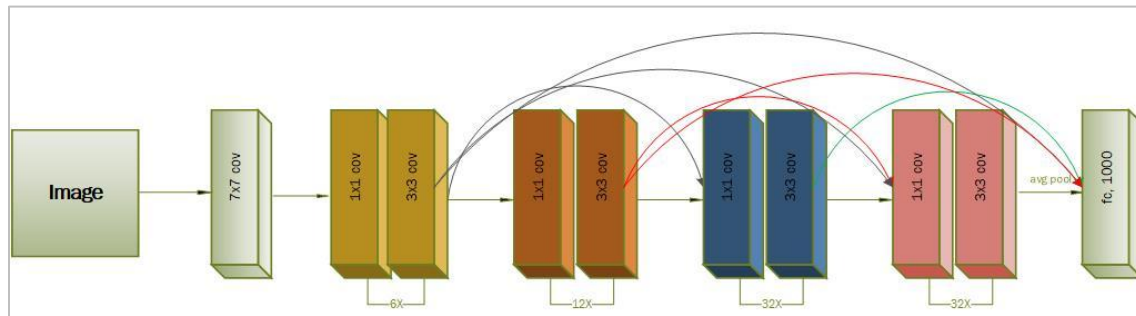**Figure 2** The bottleneck layers of the proposed DenseNet

Figure 3 Block diagram representation of the proposed DenseNet-161 architecture

As one can see, DenseNet can be perceived as a directed acyclic graph CNN (DAG-CNN) that overcomes the vanishing gradient problem and allows information transfer flow between the network layers. Literally, it contains direct connections from any layer to all following layers as illustrated in *Figure 3*. In particular, every two adjacent blocks are separated by a transition layer which changes the feature map size through convolution and pooling operations. The design of the proposed end-to-end model based using DensNet requires the specification of the number of convolution layers in each dense block to guarantee efficient information flow. Similarly, the number of dense blocks will be determined empirically. Furthermore, different optimization techniques will be investigated in this research.

Training the proposed deep networks which involve millions of parameters from scratch would take weeks. Moreover, this requires big data to avoid overfitting, and powerful Graphics Processing Units (GPUs) resources. Therefore, we rely on transfer learning to exploit the pre-trained networks to address this issue. Specifically, the original models are first pre-trained using a huge dataset, then the learned features and the model parameters are fed into our models to initialize them and resume their training using our application dataset. Besides, the last layer of the network, which is a fully connected (fc) layer is removed from the transfer learning task. On the other hand, the remaining layers of the network are used as a function extractor. The proposed CNN models pre-trained on large datasets

such as ImageNet learn simple visual patterns using the first few layers.

Then, more complex features are learned using the "deeper" layers. Finally, meaningful structures are recognized using the final layers by combining the patterns learned previously. One should note that the transfer learning allows our models to inherit knowledge (weight values) from the original pre-trained network.

In our case, to classify X-ray images, the basic structures which yielded the pre-trained models were meant to identify different objects already included in the ImageNet dataset. The graphical representation of the transfer learning approaches deployed in this research to classify X-ray images is depicted in *Figure 4*. Thus, for both models implemented in this

research, the fully connected layer was dropped and the pre-trained convolution layers were used as a base network in new architectures. As illustrated in *Figure 5*, Two blocks of the following have been attached to the base network respectively: *(i)* batch normalization, *(ii)* dropout, and *(iii)* fully connected layers. Particularly, the dropout layers were added to the base network to reduce the overfitting problem. Finally, to classify X-ray images into 2 classes the softmax activation function was used. Additionally, the proposed approach includes a data augmentation task because the dataset is relatively small. Specifically, this is intended to avoid overfitting and increase the model generalization. Moreover, the proposed data augmentation which includes various operations such as flipping and rotation, adds more variety to the considered X-ray image collection.
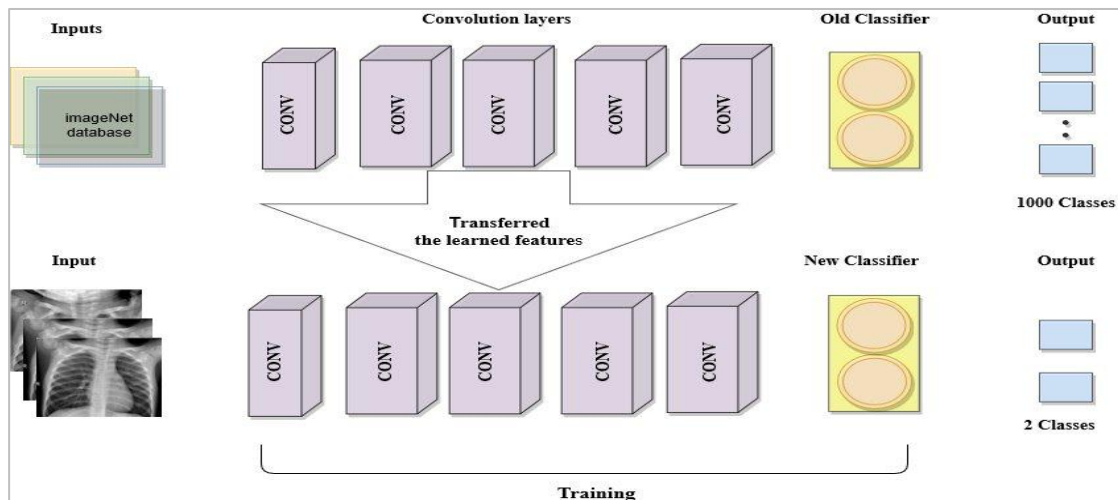

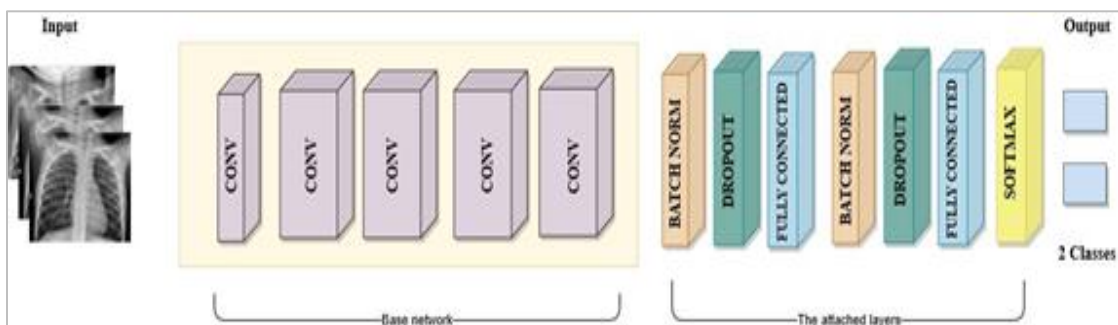**Figure 4** The transfer learning approach deployed to classify X-ray images

Figure 5 Customized CNN architecture

4.Experiments

In these experiments, we used a standard X-ray dataset [35] collected to address the detection of pneumonia infection. It consists of 5,863 JPEG Chest X-ray images (anterior and posterior) collected from paediatric patients of 1 to 5 years old. The X-ray images are categorized into two classes; Pneumonia and Normal cases. *Figure 6* shows sample images from both classes. As it can be seen, the main difference between pneumonia cases and negative cases is that the X-ray images of positive cases show white spots in the patient lungs.

In our experiments, 60% of the image collection was used for training, 20% for validation and the remaining 20% for testing. In particular, the validation set was used to perform an unbiased evaluation of the trained model fit while tuning its hyper parameters. In fact, the validation set was also used to check if the model is under fitting or overfitting. On the other hand, the test set was deployed once the model is completely trained and all parameters were fixed. This yields a better model generalization and higher overall system performance.

The purpose of the data augmentation was to increase the dataset size in order to learn a more robust deep learning model. Originally, the dataset was biased towards Pneumonia cases with 75% of positive cases versus 25% of negative (normal) cases. Therefore, the data augmentation was performed on the negative cases only in order to get a more balanced distribution of cases. Specifically, the size of the normal cases collection became equal to the size of the positive class. *Figure 7* shows sample images obtained using the data augmentation task. As it can be seen, various transformations were performed to increase the variance of the image's visual properties. *Table 1* and *Table 2* below detail the architectures of the implemented ResNet-50 and DensNet-161 networks. Note that all implemented models were trained for 10 epochs and a batch size of 32.

The Adam was adopted to optimize the parameters of the networks during the training phase. It is based on adaptive estimates of lower-order moments. Adam is space and computationally efficient. As learning rate, we empirically optimized it and set it to 0.001.

In this research, we evaluated the performance for each model using AUC, Precision, Recall and F1 measures. Typically, the confusion in *Table 3* matrix was generated for each classification task. As one can see, it reports the following cases:

- True Positive (TP): Cases correctly identified as positive pneumonia patient.
- False Positive (FP): Cases incorrectly identified as pneumonia patient (the positive class is predicted incorrectly).
- True Negative (TN): Cases correctly identified as healthy persons.
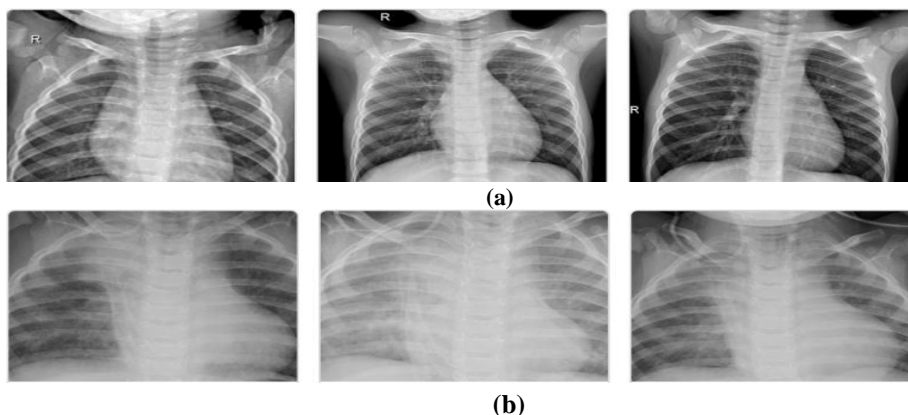- False Negative (FN): Cases incorrectly identified as healthy.

(a)

(b)

Figure 6 Sample X-ray images from standard dataset used in these experiments [35]: (a) negative cases and (b) Positive cases.

(a): Rotation

(b): Width shift

(c): Height shift

(d): Zoom
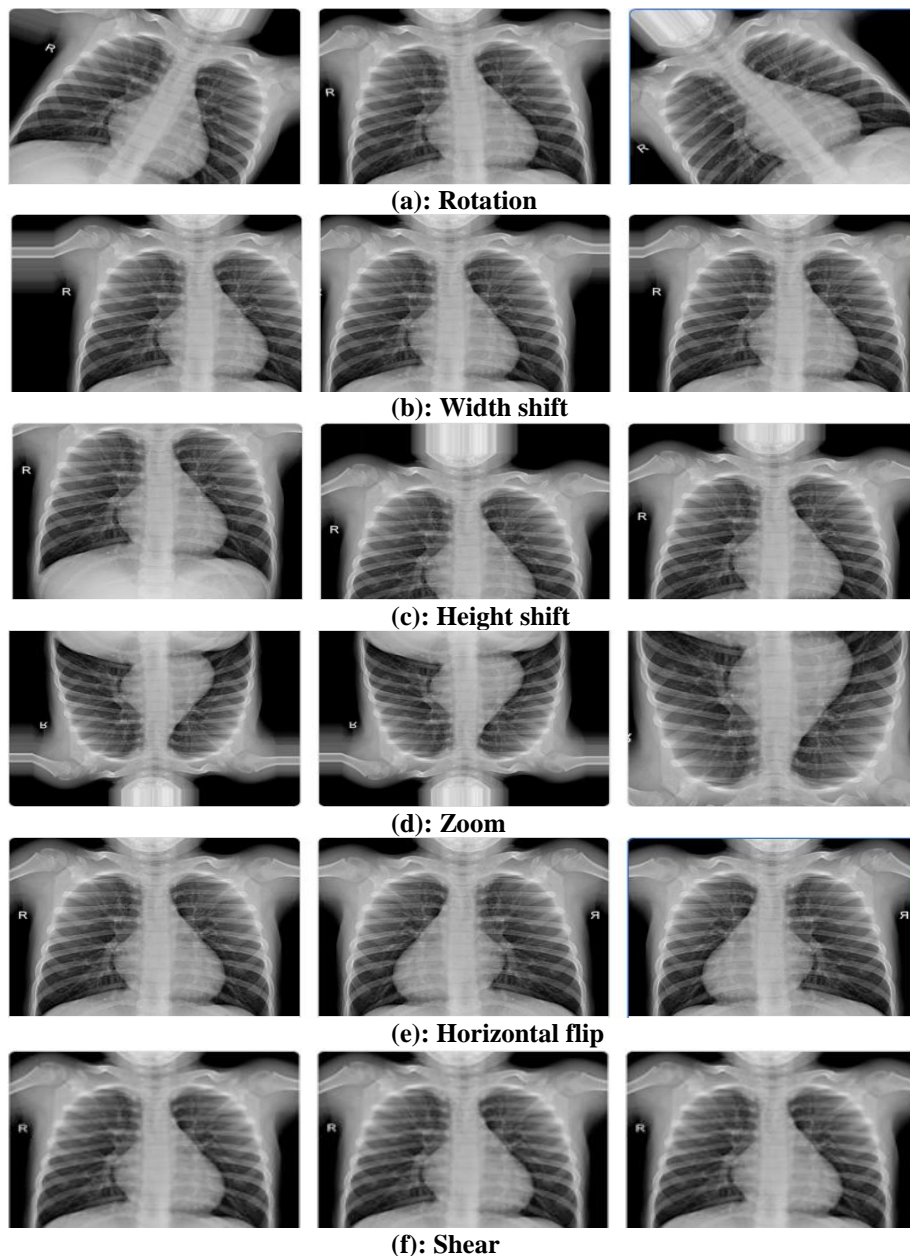

(e): Horizontal flip

(f): Shear

Figure 7 Samples X-ray images after the data augmentation

Table 1 Summary of the customized ResNet-50 implemented in this research

Layer name	Description
Data	Input Layer is normalized image of size 224×224 and with three channels for RGB colors
Conv0	A convolution layer of filter size 7×7 convolutions with 64 filters, 2×2 stride and 3 padding.
Bn0	BatchNorm2d size 64
Relu0	Rectified Linear Unit (ReLu) activation function layer.
Pool0	Max pooling of size 3×3 with stride 2×2 and 1 padding.
conv1.1	A convolution layer of filter size 1×1 convolutions with 64 filters and with 1×1 stride.
Bn1.1	BatchNorm2d size 64
Conv1.2	A convolution layer of filter size 3×3 convolutions with 64 filters, 1×1 stride and 1 padding.
Bn1.2	BatchNorm2d size 64
Conv1.3	A convolution layer of filter size 1×1 convolutions with 256 filters and 1×1 stride.

Layer name	Description
Bn1.3	BatchNorm2d size 64
Relu1	Rectified Linear Unit (ReLu) activation function layer.
Conv2.1	A convolution layer of filter size 1×1 convolutions with 128 filters and 1×1 stride.
Bn2.1	BatchNorm2d size 128
Conv2.2	A convolution layer of filter size 3×3 convolutions with 64 filters, 1×1 stride and 1 padding.
Bn2.2	BatchNorm2d size 128
Conv2.3	A convolution layer of filter size 1×1 convolutions with 512 filters and 1×1 stride.
Bn2.3	BatchNorm2d size 512
Relu2	Rectified Linear Unit (ReLu) activation function layer.
Conv3.1	A convolution layer of filter size 1×1 convolutions with 256 filters and 1×1 stride.
Bn3.1	BatchNorm2d size 256
Conv3.2	A convolution layer of filter size 3×3 convolutions with 256 filters, 1×1 stride and 1 padding.
Bn3.2	BatchNorm2d size 256
Conv3.3	A convolution layer of filter size 1×1 convolutions with 1024 filters and 1×1 stride.
Bn1.3	BatchNorm2d size 1024
Relu3	Rectified Linear Unit (ReLu) activation function layer.
Conv4.1	A convolution layer of filter size 1×1 convolutions with 512 filters and 1×1 stride.
Bn4.1	BatchNorm2d size 512
Conv4.2	A convolution layer of filter size 3×3 convolutions with 512 filters, 1×1 stride and 1 padding.
Bn4.2	BatchNorm2d size 256
Conv4.3	A convolution layer of filter size 1×1 convolutions with 2048 filters and 1×1 stride.
Bn4.3	BatchNorm2d size 2048
Relu4	Rectified Linear Unit (ReLu) activation function layer.
Pool	1×1 Avarage pooling with stride 2×2 and 1 padding.
Fc5	Fully connected layer in_features=2048, out_features=1024
Bn5	BatchNorm1d size1024
Drop5	50% Dropout
Relu5	Rectified Linear Unit (ReLu) layer activation function layer.
Fc6	Fully connected layer in_features=1024, out_features=512,
Bn6	BatchNorm1d size 512
Drop6	20% Dropout
Relu6	Rectified Linear Unit (ReLu) layer activation function layer.
Prob	Softmax function

Table 2 Summary the customized DenseNet-161 implemented in this research

Layer name	Description
Data	Input Layer is normalized image of size 224×224 and with three channels for RGB colors
Conv0	A convolution layer of filter size 7×7 convolutions with 96 filters, 2×2 stride and 3 padding.
Norm0	BatchNorm2d size 96
Relu0	Rectified Linear Unit (ReLu) activation function layer.
Pool0	Max pooling of size 3×3 with stride 2×2 and 1 padding.
Norm1.1	BatchNorm2d size 96
Relu1.1	Rectified Linear Unit (ReLu) activation function layer.
conv1.1	A convolution layer of filter size 1×1 convolutions with 192 filters and 1×1 stride.
Norm1.2	BatchNorm2d size 192
Relu1.2	Rectified Linear Unit (ReLu) activation function layer.
Conv1.2	A convolution layer of filter size 3×3 convolutions with 48 filters, 1×1 stride and 1 padding.
Norm	BatchNorm2d size 384
Relu	Rectified Linear Unit (ReLu) activation function layer.
Conv	A convolution layer of filter size 1×1 convolutions with 192 filters and 1×1 stride
Pool	2×2 Avg pooling with stride 2×2 and 0 padding.
Norm2.1	BatchNorm2d size 192
Relu2.1	Rectified Linear Unit (ReLu) activation function layer.
Conv2.1	A convolution layer of filter size 1×1 convolutions with 192 filters and 1×1 stride.
Norm2.2	BatchNorm2d size 192
Relu2.2	Rectified Linear Unit (ReLu) activation function layer.

Layer name	Description
Conv2.2	A convolution layer of filter size 3 × 3 convolutions with 48 filters, 1 × 1 stride and 1 padding.
Norm	BatchNorm2d size 768
Relu	Rectified Linear Unit (ReLu) activation function layer.
Conv	A convolution layer of filter size 1 × 1 convolutions with 384 filters and 1 × 1 stride
Pool	2 × 2 Avg pooling with stride 2 × 2 and 0 padding.
Norm3.1	BatchNorm2d size 384
Relu3.1	Rectified Linear Unit (ReLu) activation function layer.
Conv3.1	A convolution layer of filter size 1 × 1 convolutions with 192 filters and 1 × 1 stride.
Norm3.2	BatchNorm2d size 192
Relu3.2	Rectified Linear Unit (ReLu) activation function layer.
Conv3.2	A convolution layer of filter size 3 × 3 convolutions with 48 filters, 1 × 1 stride and 1 padding.
Norm	BatchNorm2d size 2112
Relu	Rectified Linear Unit (ReLu) activation function layer.
Conv	A convolution layer of filter size 1 × 1 convolutions with 2112 filters and 1 × 1 stride
Pool	2 × 2 Avg pooling with stride 2 × 2 and 0 padding.
Norm4.1	BatchNorm2d size 1056
Relu4.1	Rectified Linear Unit (ReLu) activation function layer.
Conv4.1	A convolution layer of filter size 1 × 1 convolutions with 192 filters and 1 × 1 stride.
Norm4.2	BatchNorm2d size 192
Relu4.2	Rectified Linear Unit (ReLu) activation function layer.
Conv14.	A convolution layer of filter size of 3 × 3 convolutions with 48 filters, 1 × 1 stride and 1 padding.
Fc5	Fully connected layer in_features=2208, out_features=1024
Bn5	BatchNorm1d size1024
Drop5	50% Dropout
Relu5	Rectified Linear Unit (ReLu) activation function layer.
Fc6	Fully connected layer in_features=1024, out_features=512,
Bn6	BatchNorm1d size 512
Drop6	20% Dropout
Relu6	Rectified Linear Unit (ReLu) activation function layer.
Prob	Softmax function

Table 3 Confusion matrix

		Actual class	
		Positive(P)	Negative(N)
Predicted class	Positive(P)	True Positive (TP)	False Positive (FP)
	Negative(N)	False Negative (FP)	True Negative (TN)

The Accuracy was calculated as the number of correct predictions over the total size of the dataset:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

The Specificity is the ability to determine the healthy cases correctly. It is obtained using:

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad (2)$$

The Precision measures the ability of the system to retrieve all relevant instances. It is computed as:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (3)$$

On the other hand, the Recall (Sensitivity) is defined as:

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (4)$$

Finally, the F-1 score represents a trade-off between Precision and Recall. It is calculated as:

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (5)$$

Besides, the Area Under the ROC Curve (AUC), where the ROC curve is a two-dimensional representation that illustrates the classifier performance. It is used as a standard measure of the classification performance. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0.

4.1Results

Table 4 reports the performance measures obtained using DenseNet-161 and ResNet-50 based scenarios. As it can be seen, DenseNet-161 based models outperform ResNet-50 based ones. For the accuracy, the highest achievement, 0.93, was attained using the proposed custom DenseNet-161 as confirmed by *Figure 8*. This performance can be attributed to the fact that DenseNet-161 architecture encloses more layers than ResNet50. Moreover, DenseNet-161 includes a feature layer (convolutional layer) that captures low-level features from the images. Additionally, several dense blocks and transition layers between adjacent dense blocks yield better,

higher-level features learning. This network depth is reflected by the high number of parameters of DenseNet-161 (29262914) compared to ResNet-50 (26135106). Note that DenseNet-161 was approximately 3 times faster than ResNet-50.

Regarding to the implementation with no data augmentation, the performance was drastically low for all models as reported in *Figure 9*. This confirms how data augmentation affected positively all model performances. Besides, Grayscale image-based scenarios yielded lower performance than RGB color images.

Table 4 The performance comparison of ResNet-50 for X-ray images

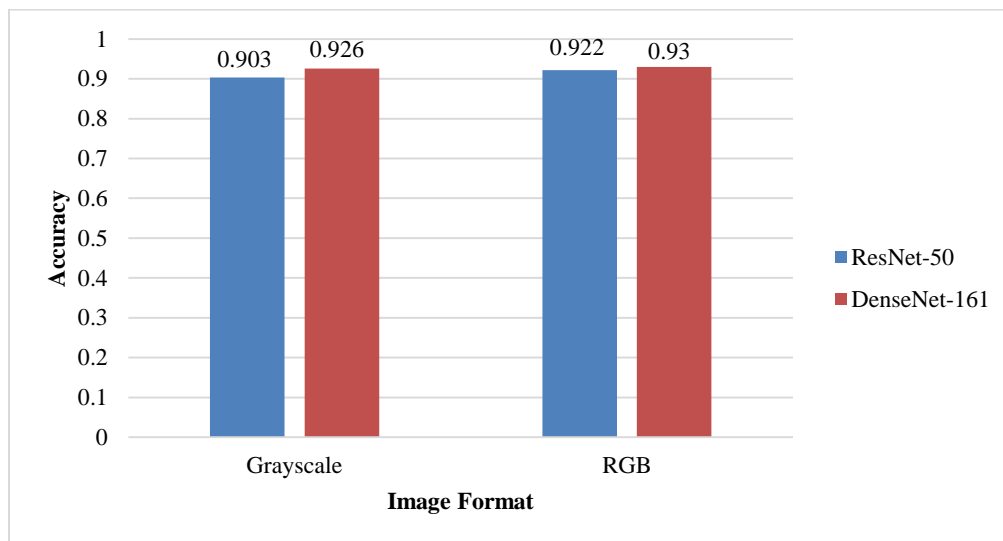| Model	Image format	Data augmentation	Accuracy	Recall	Precision	Specificity	F1	AUC	Misclassification
ResNet-50	RGB	Yes	**0.922**	**0.93**	**0.92**	**0.915**	**0.924**	**0.922**	**0.085**
Resnet-50	Grayscale	Yes	0.903	0.931	0.891	0.876	0.911	0.903	0.096
ResNet-50	RGB	No	0.838	0.979	0.80	0.62	0.869	0.799	0.168
DenseNet-161	RGB	Yes	**0.93**	**0.931**	**0.944**	**0.932**	**0.945**	**0.931**	**0.071**
DenseNet-161	Grayscale	Yes	0.926	0.928	0.94	0.941	0.927	0.925	0.078
DenseNet-161	RGB	No	0.845	0.993	0.84	0.79	0.881	0.88	0.157

Figure 8 Accuracy of ResNet-50 and DenseNet-161 models obtained using color and grayscale images
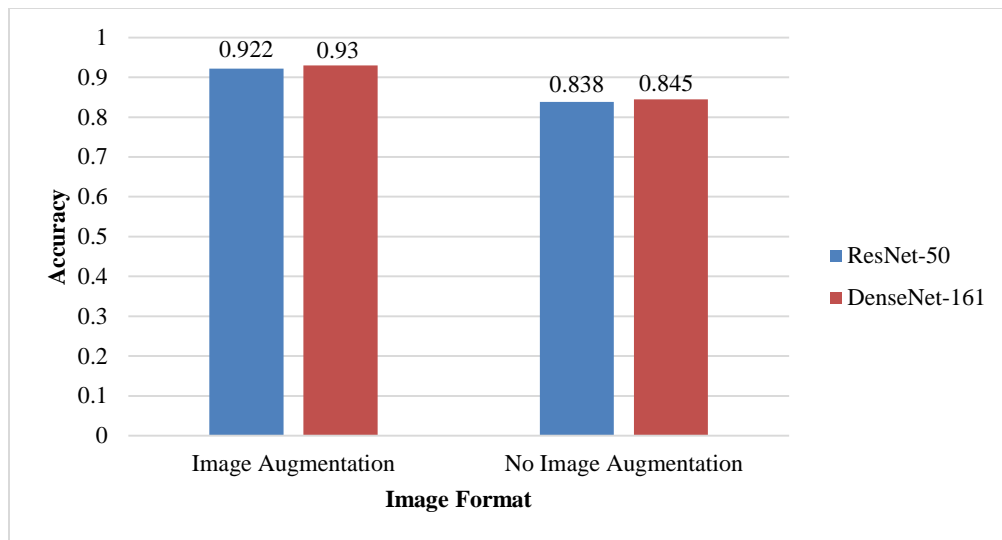
Figure 9 Accuracy of ResNet-50 and denseNet-161 models with and without data augmentation

5.Discussion

In this research, the main problem is to diagnose if a person is positive for pneumonia or not. Therefore, the main focus should be on false positive cases and true positive cases. In fact, for medical problems, the worst-case scenario would be to misdiagnose positive patients (FP) rather than conducting further investigation (even if he is actually healthy (FN)). So, the Recall value is critical and should be as high as the Precision.

As reported in *Figure 10*, Denesnet-161 overtakes ResNet-50 with respect to all considered performance measures. It achieved 0.93, 0.94, 0.94 F1 and 0.93 as recall, precision and F1-measure, respectively. Our

model overcomes one of the states of art results in [5] that has applied data augmentation and used the same Dataset with 0.9 accuracy as reported in *Figure 10*, Denesnet-161 overtakes ResNet-50 with respect to all considered performance measures. It achieved 0.93, 0.94, 0.94 F1 and 0.93 as recall, precision and F1-measure, respectively.

Sample images correctly classified using the customized ResNet-50 and DenseNet-161 are shown in *Figure 11*.

Similarly, sample cases misclassified using the customized ResNet-50 and DenseNet-161 are reported in *Figure 12*.
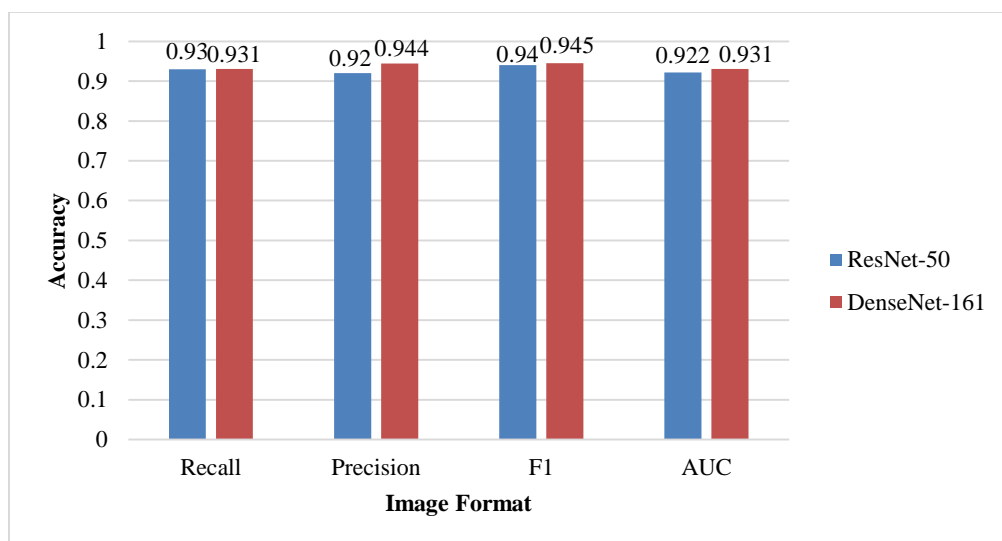

Figure 10 Recall, Precision, F1 measures and AUC obtained using the customized ResNet-50 and DenseNet-161
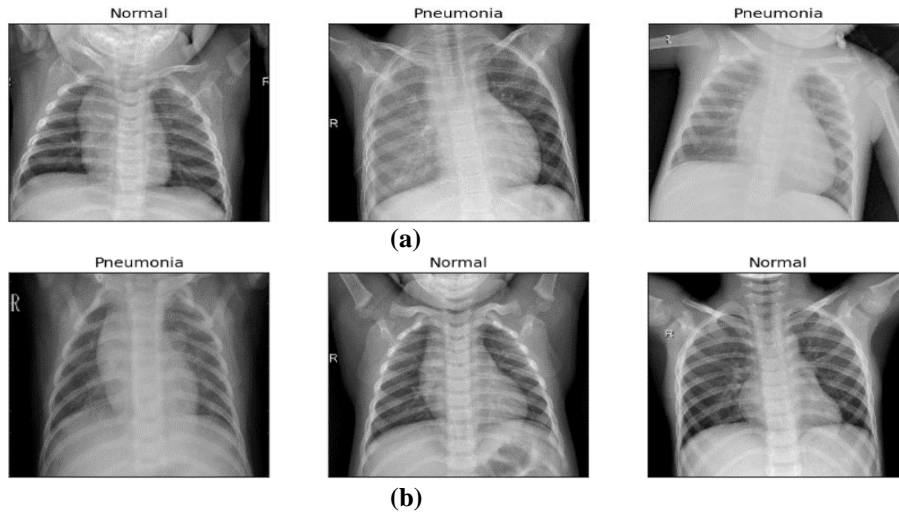
(a)

(b)

Figure 11 Sample correctly classified cases using: (a) ResNet-50 and (b) DenseNet-161

(a)ResNet-50 False Negative cases

(b)ResNet-50 False Positive cases

(c)DenseNet-161 False Negative cases
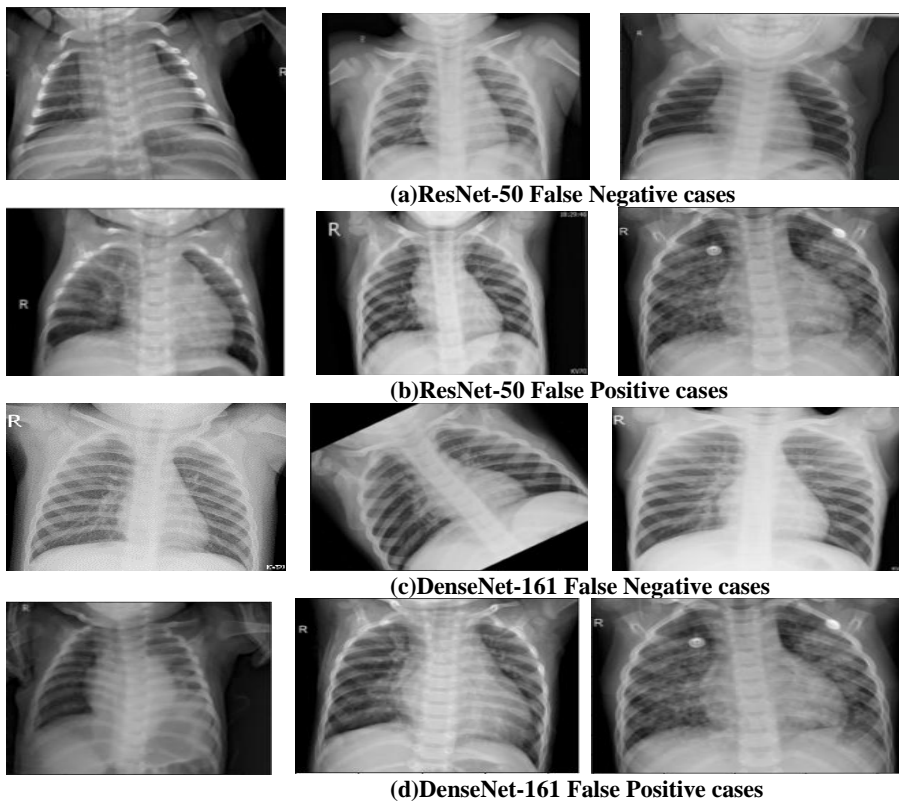

(d)DenseNet-161 False Positive cases

Figure 12 Sample cases misclassified using ResNet-50 and DenseNet-161

6.Conclusions and future works

In this research, different CNN based architectures were investigated and designed to address the Pneumonia classification problem. The proposed system was intended to recognize pneumonia cases based on X-ray images. Moreover, data augmentation was used to overcome the relatively low size of the data and avoid overfitting. ResNet-50 and DenseNet-65

161 models were adopted, customized and implemented to detect positive pneumonia cases. The standard datasets and performance measures used in our experiments proved that DenseNet-161 based models outperformed ResNet-50 based once.

One should note that the main limitation of the proposed system remains its inability to discriminate

Sarah Badr AlSumairi and Mohamed Maher Ben Ismail

between bacterial Pneumonia and viral Pneumonia. Particularly, this would support the medical community in diagnosing COVID-19. Thus, using relevant X-ray datasets including bacterial and viral Pneumonia would be considered as future works. Also, deep ensemble models can be designed using various pre-trained models in order to improve the pneumonia detection performance by exploiting the diversity of the base models.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References

[1] Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. Journal of the American Medical Informatics Association. 2000; 7(6):593-604.
[2] Förster A, Brandstetter S, Schulze-Briese C. Transforming X-ray detection with hybrid photon counting detectors. Philosophical Transactions of the Royal Society A. 2019.
[3] Sousa RT, Marques O, Soares FA, Sene Jr II, de Oliveira LL, Spoto ES. Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. Procedia Computer Science. 2013; 18:2579-82.
[4] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225. 2017.
[5] Stephen O, Sain M, Maduh UJ, Jeong DU. An efficient deep learning approach to pneumonia classification in healthcare. Journal of Healthcare Engineering. 2019.
[6] Ayan E, Ünver HM. Diagnosis of pneumonia from chest X-ray images using deep learning. In scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT) 2019 (pp. 1-5). IEEE.
[7] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In proceedings of the conference on computer vision and pattern recognition 2017 (pp. 4700-8).
[8] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In conference on computer vision and pattern recognition 2009 (pp. 248-55). IEEE.
[9] Sirazitdinov I, Kholiavchenko M, Mustafaev T, Yixuan Y, Kuleev R, Ibragimov B. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Computers & Electrical Engineering. 2019; 78:388-99.
[10] Hoang TM, Nguyen PH, Truong NQ, Lee YW, Park KR. Deep retinanet-based detection and classification of road markings by visible light camera sensors. Sensors. 2019.
[11] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In proceedings of the IEEE international conference on computer vision 2017 (pp. 2961-9).
[12] Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of deep learning approaches for multi-label chest X-ray classification. Scientific Reports. 2019; 9(1):1-10.
[13] Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 2017; 284(2):574-82.
[14] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In advances in neural information processing systems 2012 (pp. 1097-105).
[15] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
[16] Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ. Identifying pneumonia in chest X-rays: a deep learning approach. Measurement. 2019; 145:511-8.
[17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-8).
[18] https://www.kaggle.com/pytorch/resnet101. Accessed 26 June 2020.
[19] Varshni D, Thakral K, Agarwal L, Nijhawan R, Mittal A. Pneumonia detection using CNN based feature extraction. In international conference on electrical, computer and communication technologies 2019 (pp. 1-7). IEEE.
[20] Chollet F. Xception: Deep learning with depthwise separable convolutions. In proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1251-8).
[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
[22] Sarkar R, Hazra A, Sadhu K, Ghosh P. A novel method for pneumonia diagnosis from chest x-ray images using deep residual learning with separable convolutional networks. In computer vision and machine intelligence in medical image analysis 2020 (pp. 1-12). Springer, Singapore.
[23] Rajaraman S, Candemir S, Kim I, Thoma G, Antani S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Applied Sciences. 2018.
[24] Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest

radiographs: a cross-sectional study. PLoS Medicine. 2018.

[25] Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Medicine. 2018.

[26] Abiyev RH, Ma'aitah MK. Deep convolutional neural networks for chest diseases detection. Journal of Healthcare Engineering. 2018.

[27] Cilimkovic M. Neural networks and back propagation algorithm. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin. 2015.

[28] Abas AR. Adaptive competitive learning neural networks. Egyptian Informatics Journal. 2013; 14(3):183-94.

[29] Toğaçar M, Ergen B, Cömert Z, Özyurt F. A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. IRBM. 2020; 41(4):212-22.

[30] Behzadi-khormouji H, Rostami H, Salehi S, Derakhshande-Rishehri T, Masoumi M, Salemi S, et al. Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. Computer Methods and Programs in Biomedicine. 2020; 185:105162.

[31] Ullah I, Petrosino A. About pyramid structure in convolutional neural networks. In international joint conference on neural networks 2016 (pp. 1318-24). IEEE.

[32] Antin B, Kravitz J, Martayan E. Detecting pneumonia in chest x-rays with supervised learning. http://cs229 .stanford.edu/proj2017/final-reports/5231221.pdf. Accessed 26 June 2020.

[33] Karthick G, Harikumar R. Comparative performance analysis of naive bayes and SVM classifier for oral X-ray images. In international conference on electronics and communication systems 2017 (pp. 88-92). IEEE.

[34] Oliveira LL, E Silva SA, Ribeiro LH, de Oliveira RM, Coelho CJ, Andrade AL. Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. International Journal of Medical Informatics. 2008; 77(8):555-64.

[35] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge. Accessed 26 June 2020.

Sarah Badr AlSumairi is a Master's student in the Computer Science program at the Computer Science department, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh, Saudi Arabia.

Mohamed Maher Ben Ismail is an Associate Professor in the Computer Science Department of the College of Computer and Information Sciences at King Saud University. He received his PhD. degree in Computer Science from the University of Louisville in 2011. His research interests include Pattern Recognition, Machine Learning, Data Mining and Image Processing.
Email: mbenismaileksu.edu.sa