Comparison of Deep Learning Approaches in Classification of the Chest X-Ray

DOI : 10.36909/jer.17121

Serap Karagol *, Kubra Nur Akpinar, Secil Genc
Ondokuz Mayis University
Electrical and Electronics Eng. Dept.,
Samsun, Turkey
*Corresponding Author : serap.karagol@omu.edu.tr

ABSTRACT

Chest X-Ray is a radiological examination that is commonly used in clinical practice and is easy to access. Deep convolutional neural networks (DCNN) are used to make the computer-aided diagnosis (CAD) of diseases on chest radiography. Deep convolutional neural networks help the radiologist to diagnose better. In this study, the ChestX-Ray14 data set was examined to assess performance modern deep learning networks in diagnosing chest diseases. X-Ray image quality was improved by applying a three-step process including crop, histogram equalization and contrast-limited adaptive histogram equalization to the data sets. For training and validation purposes, images in the dataset were applied to the model with and without preprocessing. It was determined that the processed datasets provided more accurate results than the original images. AlexNet, ResNet50, and GoogLeNet deep learning architectures were used to determine the presence of chest disease. The performances of these models, which generally classify normal and abnormal results from chest radiographs, were analyzed using preprocessed ChestX-Ray14 datasets and comparative evaluations were made. The most accurate was the ResNet architecture, where we used the preprocessed datasets to detect abnormalities, with 91.46% accuracy and 0.9584 area under curve (AUC) results.

Keywords: Deep Learning; Convolutional Neural Network; Medical Imaging; Chest X-ray; Chest Disease.

INTRODUCTION

Chest diseases are one of the most crucial and frequent health threats across the globe. Pneumonia affects an average of 450 million people each year, resulting in the death of around 4 million people (Wang et al., 2018). According to the World Health Organization (WHO) report, 1.6 million people die because of lung cancer per year (Kumar, 2018). Chest X-Ray (CXR) is one of the highly preferred methods for screening and diagnosing many lung diseases due to its accessibility and low cost. X-Ray imaging studies, including numerous radiological reports, are collected and kept in the Picture Archiving and Communication Systems (PACS) in many modern hospitals (He et al., 2016). Computer-aided detection tools help radiologists make quantitative and well-informed decisions. A significant problem in the medical field is the availability of reliable big data sets. The increasing population and life expectancy increases the demand for Chest X-Ray readings. Chest radiographs are analyzed partially by visual inspection. Interpretation of radiographs needs a high amount of skill and concentration. It is also prone to operator bias, expensive, time-consuming (Wang et al., 2018). As the volume of data grows, it will become challenging for radiologists to interpret and maintain the same efficiency level from X-Rays. Also, because of the complexity of chest radiographs, even radiologists have difficulty distinguishing between chest diseases, resulting in a lack of expert competent radiologists to read chest radiographs in numerous countries. It is known that this technology is not accessible in many regions of the world. Even if the materials required for radiography are available, the lack of specialists who can interpret the imaging results causes negative effects in patients with treatable illnesses (Rajpurkar et al., 2017).

Radiologists urgently need automation and augmentation to maintain diagnostic quality. It is advantageous to have a genuine computer tool to effectively and precisely determine the diagnosis and detect diseased CXR images (Liu et al., 2019). Automated classification of these CXR images can provide a great deal of help for the radiology interpretation process. Therefore, radiologists can concentrate on detecting diseases in abnormal images instead of checking abnormalities in every single image (Kumar, 2018).

Recently, deep learning has emerged as a promising technology that has achieved superhuman performance in classifying problems globally. This success in recognizing objects in chest images has led to an increased interest in deep learning practices in medical images (Islam et al., 2017). CAD techniques have been tried in many biomedical applications and successful results have been obtained. It has shown remarkable power in various tasks in medical image analysis, such as diagnosis of diabetic eye disease (Lakhani et al., 2017), detection and localization of cancer metastases (He et al., 2016, Behzadi-khormouji et al., 2019) detection of lung nodules (Ma et al., 2017), classification of diseases (Wang et al., 2018, He et al., 2016), and lesion segmentation (Bar et al., 2015). However, there are various challenges in deep learning practices for reading and understanding chest X-Ray images. For example, there are a wide variety of locations and sizes of diseases in images that are taken from different samples of chest X-Rays, a lack of additional explanation by radiologists, years of radiologist experience, and images of structures in the chest apart from the lungs and heart (medical tools, bone structures, etc.). For this reason, it was necessary for this study to preprocess the chest X-Ray images of patients and to understand how deep learning affects performance (Zech et al., 2018).

Our contributions to the literature with this study are summarized below:

- By applying three-stage preprocessing (crop, histogram equalization, CLAHE method) to the images in our data set, we reduced the confusing variables in the images and increase resolution.
- The presence of chest diseases was detected with 91.46 % accuracy using ResNet50, which is among the most popular deep learning network. AlexNet and GoogLeNet performance was also observed.
- The classification performance of the data processing applied, original and masked data sets were compared for all three networks.
- The effect of image processing on the classification of chest radiographs as normal or abnormal has not been examined in previous studies. The study also makes comparisons for all combinations of the segmentation dataset.

Much research has been done based on the automatic detection of different chest diseases. Islam et al. used the JSRT dataset, Indiana chest X-Ray dataset, and Shenzhen dataset and analyzed the performances of deep neural network (Alex + VGG + Google) architectures in 20 different abnormalities. They used heat maps acquired from the sensitivity of occlusion as a measure of localization in CXRs. DCN community models showed better performance than single models (Islam et al., 2017). Behzadi et al. proposed Chestnet, a problem-based architecture, which is proportional to the size of the data set examined to detect consolidation in chest X-Ray images. They reported that it could decrease the performance of radiological deep learning models in terms of changing lung images, strong edges of the rib cage, changing heart sizes, and density changes around the clavicle bones, and costophrenic sharp corners, difficulties in X-Ray imaging homogeneity, and suggested an effective preprocessing to improve images (Behzadi-khormouji et al., 2019). Kitamura et al. divided the open-source chest X-Ray8 (CXR8) data set into two as pneumothorax (PTX) and non-PTX cases, using the data set in their health systems between 2013 and 2017. They created a limited control machine learning model that includes both localized and non-localized pathology. The initial training of the model resulted in 0.90 AUC for pneumothorax detection using the CXR8 data set. They achieved 0.59 AUC when they used their validation datasets on the CXR8 trained model. After the model was retrained with its training datasets, the result was increased to 0.90 AUC (Kitamura et al., 2020). Guan et al. used Resnet50 and Densnet121 network architectures for the classification of chest diseases. Using heat map images, they masked and the cropped local areas through detection. They are classified by combining the last pooling layers of local and global branches (AG-CNN). Finally, they compared 14 disease classification performances of ResNet-50 and DenseNet-121 based AG-CNN architectures (Guan et al., 2020).

As a result of the literature review, it has been observed that the detection of abnormalities in chest radiographs becomes more difficult due to the poor quality of chest radiographs. Some of the data sets used to train the network were first preprocessed, such as segmentation operations, and some were only modified in size. This study aims to show how chest X-Ray image quality can be improved by applying three-step preprocessing, including crop, histogram equalization, and contrast-limited adaptive histogram equalization, to the dataset to improve the performance of the diagnosis of chest diseases. Normal and abnormal images were classified and their comparative evaluations were made using deep learning-based algorithm models.

CHEST X-RAY DATASET

Recently, two large chest X-ray datasets have been made available: the CXR dataset by Open-I (Wang et al., 2018) and the ChestX-Ray14 dataset from the National Institutes of Health (NIH) Clinical Center (Kumar, 2018). With the use of convolutional neural networks (CNN) in computer vision by Wang et al., various research groups have started the applications/practices for the classification of chest X-rays with CNN. The dataset does not contain the original DICOM images, however in Wang's work, the pixel depth was lowered to 8 bits and they applied a simple preprocessing to the images. In this study, ChestX-Ray14 was chosen as the data set to be used for different modern deep learning networks to detect chest abnormalities. This dataset contains the largest, highest-quality X-Ray images available to the public. Images are usually 3000×3000 pixels in size. The data set containing 14 different diseases were taken from more than 112,000 X-Ray images and more than 32,000 patients.

DEEP NEURAL NETWORKS ARCHITECTURES

The latest modern networks are frequently used in the detection of diseases present in major organs like the lungs, brain, and breast. In this study, chest X-Ray images were tested with three different deep learning networks: AlexNet, ResNet-50, and GoogLeNet after preprocessing was applied to the images. The networks were chosen for CNN by taking part in the top classification contests and performing at the highest level. When these models are compared, it shows that the convolutional layers differ from each other: as the number of layers increases, the classification accuracy increases, but the processing time is extended. AlexNet and ResNet are trained in the popular ImageNet database. Fully connected layers are selected for feature extraction. For example, the properties of ResNet-50 are determined using the res4f layer (Islam et al., 2017). All deep convolutional network models were applied in MATLAB and completed using Adam optimizer with learning rates of 0.0001.

Firstly, the performance of the AlexNet network, which was trained with 1.2 million images previously in the ImageNet dataset and whose weights are calculated, was examined. In this approach, deep learning models are pre-trained by using different numerical datasets to learn simple patterns such as edge, line, point, and spot. The model is then trained by using the main dataset to learn more abstract and complex patterns. Deep learning techniques also increase the interest in applications of Residual neural network (ResNet-50) architecture along with the increase in the accessibility of chest X-Ray images. The ResNet-50 deep convolutional neural network uses 3×3 and 1×1 size convolutional filters to reach local information between images. The convolution process of this dimension learns local and small patterns to distinguish images. Another popular network used in recent years is GoogLeNet architecture which is one of the first CNN architectures to avoid stacking convolution and pooling layers in a consecutive structure. Additionally, this new model has a crucial place with respect to power and memory usage. Since stacking multiple layers and adding a large number of filters burdens calculations and memory, GoogLeNet modules are used in parallel to overcome this issue.

TRAINING NETWORK

The parameters of the deep convolutional neural network structures used in this study are shown in Table 1. Seventy percent of the images chosen randomly from the data set are used are education and thirty percent are test data. The images have been reduced to 224×224×3 in size before being processed as network input data. As in the paper of Baltruschat et al., no dropout was applied in this study (Baltruschat, 2019). ADAM with default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ was used as the optimizer. The learning rate is set to $l_r = 0.00$l with the transfer-learning method. In the CNN, by looking at the loss of verification during education, the learning rate can be reduced in case the loss value does not decrease. According to the model architecture created, the mini-batch size is set to 15 due to different trials. Models are performed in the MATLAB simulation program and trained on GTX 960M GPUs.

Table 1. Parameters of Deep Evolution Neural Networks Used in the Study

Parameters	**AlexNet**	**ResNet50**	**GoogLeNet**
Image Size	224×224×3		
Learning Rate	$1e^{-4}$		
Batch Size	15		
Epochs	6, 10	6, 8, 10	6, 10
Optimizer	Adam		
Dilution	-	-	-
Number of Neural Networks in the First	1024		
Activation Function in the Final Classification Layer	Softmax		

RESULTS

Three datasets were tested with three different convolutional neural networks, both individually and together in a single cluster, to determine chest abnormalities by using chest X-Ray images and performance analysis. Image preprocessing is part of the proposed model and is performed as a preprocessing before pooling layer. It has been established through comparisons in the result tables (such as Table 2) that preprocessed images increase the network's ability for decision-making. The original image set, the preprocessed image set, the masked image set, and combinations of these image sets were applied as inputs to AlexNet, Resnet-50, and GoogLeNet neural networks. After adjusting the hyper-parameters of AlexNet, Resnet-50, and GoogLeNet networks, which are used to decide whether chest X-Rays are abnormal or normal, performance outputs for each of the three datasets are shown graphically. Also, the accuracy of the results, which are represented as a percentage, are measured using the probability that a randomly selected image belongs to the normal or abnormal class. Accuracy values, performances obtained according to the image set used, are shown in Table 2 comparatively. The correct estimation percentages of 4 randomly selected images from the set of images that are classified by AlexNet are also shown in Figure 1. ResNet-50, which is a strong network in classification, performed better than the other networks in many classification studies before. Training and validation accuracy and loss graphs of the preprocessed image set, which has the higher performance when compared the other networks. Four randomly selected images from the preprocessed image set with their accuracy percent are also shown in Figure 2. Classification examples and accuracy percentages of GoogleNet are shown in Figure 3.

Table 2. Accuracy of AlexNet, ResNet-50 and GoogLeNet

DCNs	**AlexNet**		**ResNet-50**		**GoogLeNet**		**# of Abnormal Images**	**# of Normal Images**
Image Type	Acc (%)	AUC	Acc (%)	AUC	Acc (%)	AUC		
Original	85.43	0.9179	89.95	0.9410	88.44	*0.9578	340	329
Pre-Processed	*89.95	*0.957	*91.46	*0.9584	*90.95	0.9364	335	325
Masked	72.94	0.8221	74.71	0.8092	78.24	0.8272	286	279
Original + Pre-Processed	86.11	0.9273	89.65	0.9517	88.64	0.9535	674	654
Original + Masked	74.63	0.8128	81.12	0.8783	71.39	0.8104	572	558

Original + Pre-Processed + Masked	81.63	0.8795	85.16	0.9278	82.16	0.9323	960	960
*is the best value								
Acc: Accuracy								

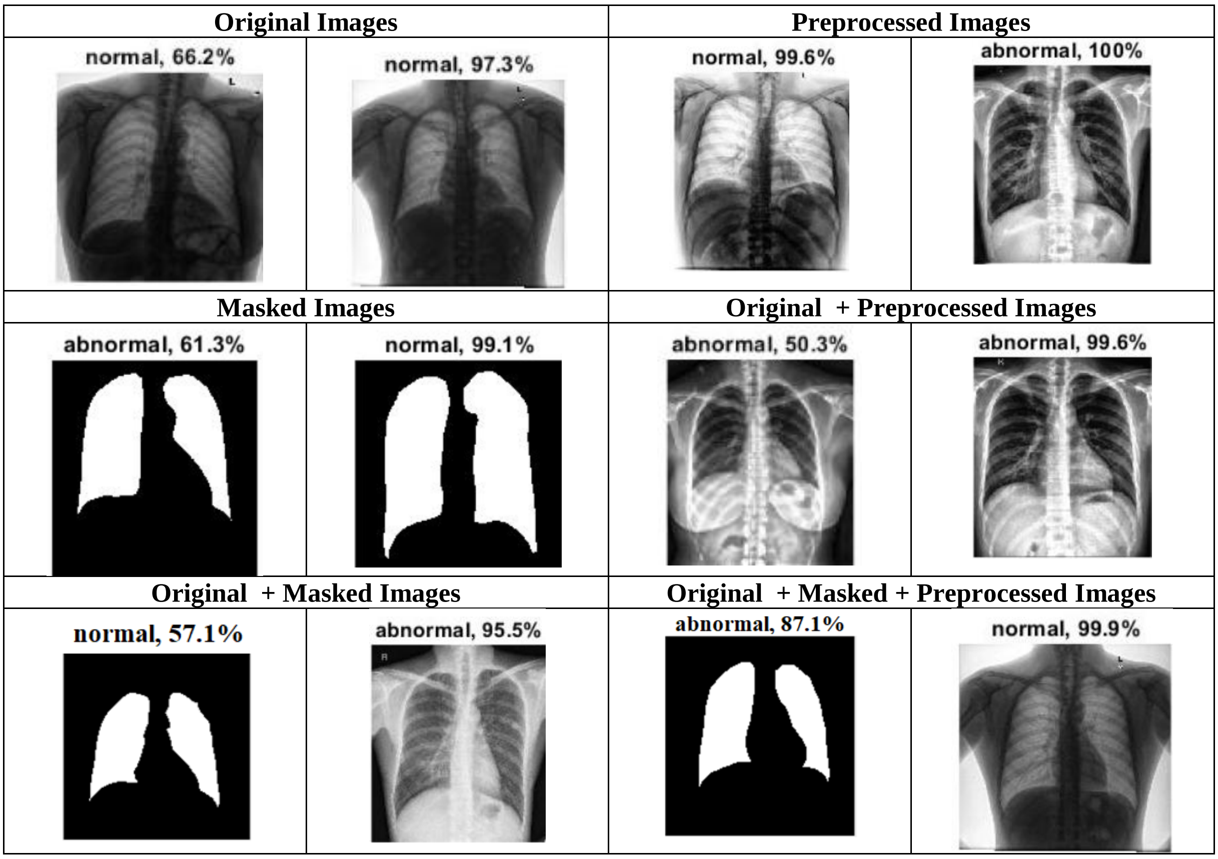
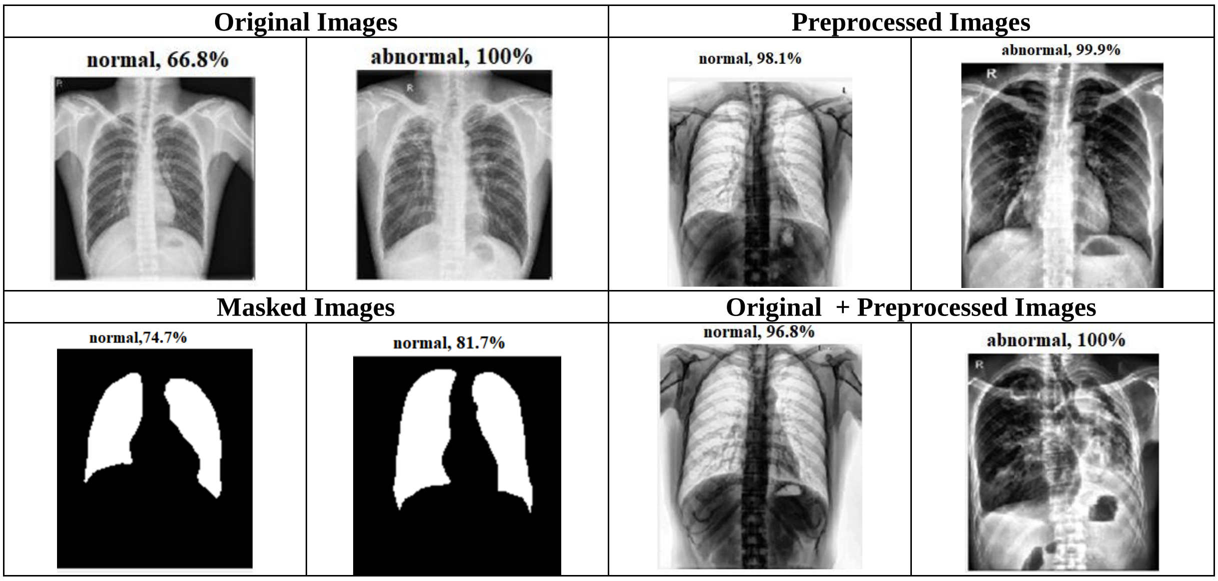


Figure 1. AlexNet Classification Result of 2 Randomly Choosen Images with Accuracy



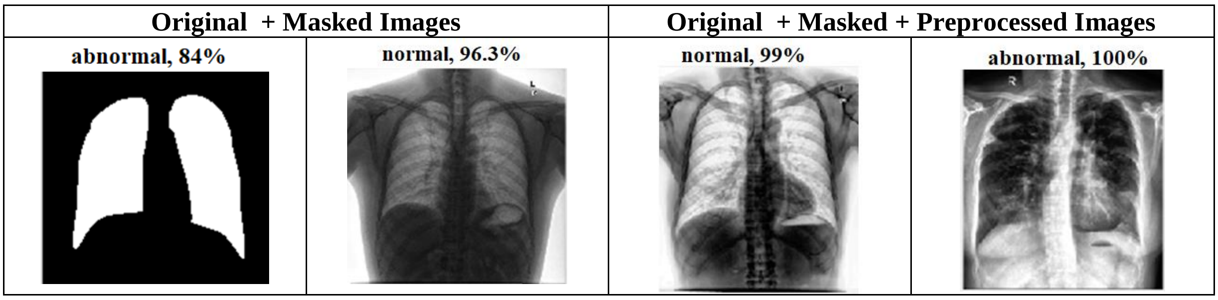


Figure 2. ResNet50 Classification Result of 2 Randomly Chosen Images with Accuracy

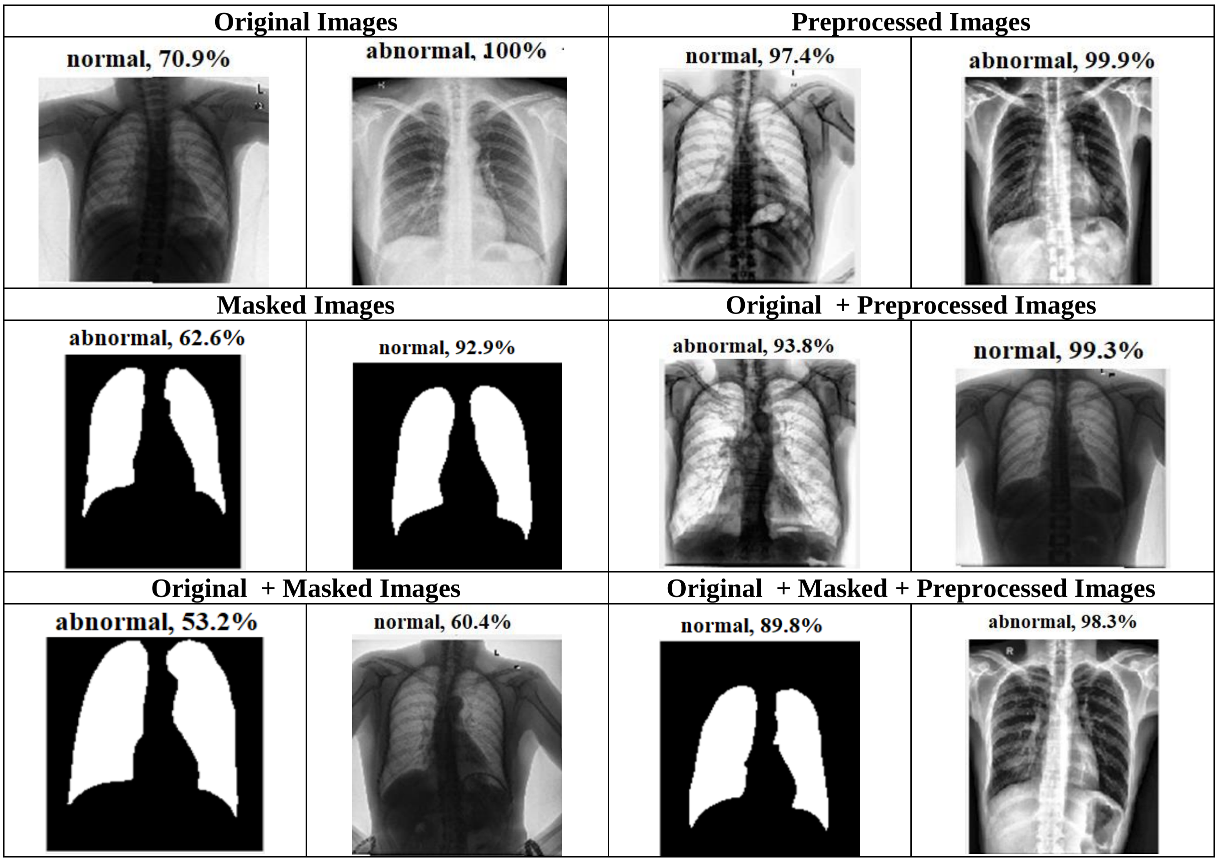


Figure 3. GoogleNet Classification Result of 2 Randomly Choosen Images with Accuracy

DISCUSSION

The classification quality of the study was measured in two units; receiver operating characteristics (ROC) - area under the curve (AUC) and accuracy. Accuracy is the correctly classified sample rate in the total sample. The ROC curve is a graphical plot that shows the true positive rate (TPR) and false positive rate (FPR) when the threshold value of the binary classifier changes from 0 to 1.

Specificity indicates the degree to which the classifier correctly classifies normal image samples as normal images. The aim of classification is obtaining high sensitivity and specificity, low error in diagnosis (Islam et al., 2017). The error matrix exists to measure the performance of machine learning classification problems where the output is two or more. It is a table with four different combinations with predicted and actual values. A more detailed discussion of the results is added as a table that can be found in Table 3.

Table 3. Overview of the other studies performances

Study	Year	DL technique	Acc (%)	Loss function	Evaluation metric
(Panwar et al.)	2020	VGG-16 (nCOVnet)	88.10	-	Acc, Sn, Sp, AUROC
(Mohammadi et al.)	2020	VGG-19	-	-	-
(Nigam et al.)	2021	Xception NASNET VGG-16	89.96 88.03 85.03 79.01	CCE, Weighted BCE	Acc, P, R, F1-score
(Azemin et al.)	2020	ResNet-101	71.9	cross-entropy	Acc, Sn, Sp, AUROC
(Jabber et al.)	2020	ResNet-101	82.5	-	-
(Hemdan et al.)	2020	VGG19 DenseNet201	90	cross-entropy	Acc, P, R, F1-score
(Das et al.)	2021	Ensemble model (ResNet-50, DenseNet201, InceptionV3)	91.62	Adam	5-fold cross validation, Acc, Sn, F1-score, AUROC
Current work	-	AlexNet ResNet-50 GoogLeNet	89.95 91.46 90.95	-	Acc

In this paper, ROC statistics were used to compare the abnormal classification performance of different approaches. In Figure 4, the results of the classification of 3 pre-trained AlexNet, GoogLeNet and ResNet-50 networks chest X-Ray images as normal and abnormal are shown with ROC curves. As a result of fine-tuning DCNs of all datasets created, Area-Under-Curve (AUC) and accuracy values, performances obtained according to the image set used are shown in Table 2 comparatively. Quantitative performance indicates that the ResNet-50 model achieves the best results with preprocessed images with AUC = 0.9584. Masked data appears to perform lower on all networks than other data sets.

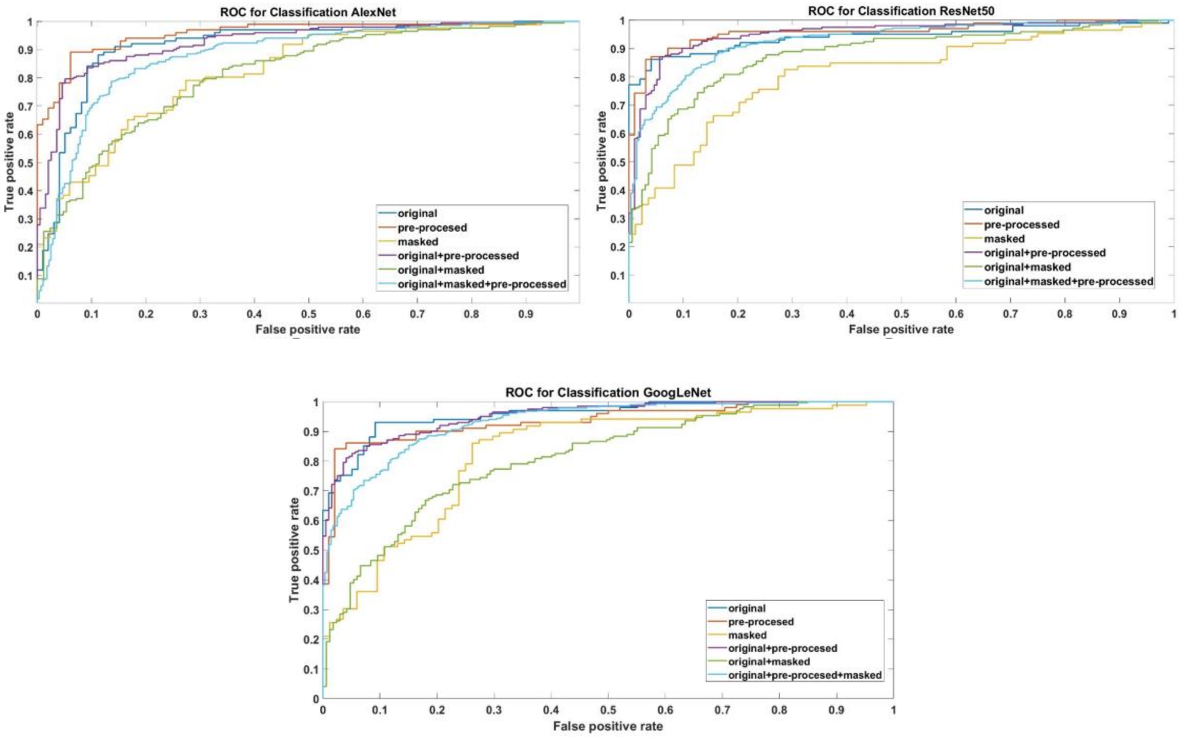


Figure 4. A comparison of AlexNet, ResNet-50 and GoogLeNet classification performance with different data sets

The results obtained in this study were compared with those of other authors working in the relevant field. In the classification of normal and abnormal images, Anuj (A. Kumar, 2018) used 1642 for training, 300 for verification and 200 (100 normal and 100 abnormal) images from the X-Ray dataset from the NIH website. The image size is 50×50 pixels. From the training parameters, the batch size is 64 and the number of epoch is 50. As a result, they classified the abnormal images with an accuracy of 61%. In addition, the ROC chart and AUC value were not included in determining the accuracy of the model used in their studies. Hisaichi et al. (H. Shibata et al., 2020) used the data set of the Radiological Society of North America (RSNA). They proposed the flow-based deep learning method to detect abnormalities in chest X-Rays (CXRs). The method using the logarithmic probability ratio metric detected abnormalities in general with AUC 0.783.

Successful deep learning models such as AlexNet, ResNet and GoogLeNet have been trained on over a million images that are hard to find in medicine. For this reason, these three architectures were preferred for the detection of chest abnormally in the study. Unlike other studies, preprocessing has been applied to increase the resolution of the images. Compared to the studies mentioned above, the preprocessed network using the GoogLeNet network reached 90.95% accuracy, while the presence of chest abnormally was more accurate by finding an AUC value of 0.9579 with the preprocessed image clusters of the AlexNet network. The challenge in this study is that the architectures used limit the volume of data for training and validation, as it requires a large amount of computation and additional memory load. Models are implemented in MATLAB simulation program and trained on GTX 960M GPUs. Despite having a sufficient number of X-ray datasets to apply with, the computer's GPU is not a high performance one. Therefore, processed dataset was applied to the deep neural network as 335 abnormal and 325 normal images input data and classified with 91.46% accuracy after 8 epochs. When the number of data used is increased, classification with higher accuracy will be possible. Actually, accuracy is high when compared to with the number of training samples. When the volume of data used is increased, classification with higher accuracy will be possible. As a result, useful information and concepts have been explained and presented to those who want to study the field of chest abnormality detection with deep learning.

CONCLUSION

In this paper, the data set used first was introduced, and then the preprocessing to improve this data set was explained step-by-step. In addition, the architectures and layers of the AlexNet, ResNet-50 and GoogLeNet networks, which are used in classifying images, were examined. The network performances were analyzed with the ROC curves and the results were discussed. Technical details are explained with explanations and visual supports. The performances of these models, which generally classify normal and abnormal results from chest radiographs, were analyzed using preprocessed ChestX-Ray14 datasets and comparative evaluations were made. The most accurate was the ResNet architecture, where we used the preprocessed datasets to detect abnormalities, with 91.46% accuracy and 0.9584 AUC results. Different image segmentation algorithms may be studied in future works. Also, the training phase can be tried with different sets of images (brain, etc.) in the field of medicine. Apart from the networks tested in this study, studies can be carried out on networks with higher classification performance. As seen at the end of the study, designs with different depths can be applied for the high performance ResNet-50 network.

REFERENCES

Wang, H., & Xia, Y. 2018. Chestnet: A deep neural network for classification of thoracic diseases on chest radiography. arXiv preprint arXiv:1807.03058.

Kumar, A. 2018. Automated Classification of chest X-ray images as normal or abnormal using Convolutional Neural Network. Asian Journal for Convergence in Technology (AJCT) ISSN -2350-1146 4(I).

K. He, X. Zhang, S. Ren, & J. Sun. 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., … & Ng, A.Y. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

Liu H., Wang L., Nan Y., Jin F., Wang Q., & Pu J. 2019. SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images, Computerized Medical Imaging and Graphics,75: 66-73. ISSN 0895-6111.

Panwar H., Gupta P., Siddiqui M.K.Morales & Menendez R., Singh. 2020. Application of deep learning for fast detection of COVID-19 in X-Rays using ncovnet. Chaos Solitons Fractals 138:109944.

Mohammadi R., Salehi M., Ghaffari H., Rohani A. & Reiazi R. 2020. Transfer learning-based automatic detection of coronavirus disease 2019 (COVID-19) from chest X-ray images, J. Biomed. Phys. Eng. 10(5):559.

Nigam B., Nigam A., Jain R., Dodia S., Arora N., & Annappa B. 2021. COVID- 19: Automatic detection from X-ray images by utilizing deep learning methods, Expert Syst. Appl 176:114883.

Azemin C., Zulfaezal M., Hassan R., Mohd Tamrin M.I. & Ali M.A. Md. 2020. COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest X-ray images as training data: preliminary findings. Int. J. Biomed. Imaging 2020.

Jabber B., Lingampalli J., Basha C.Z. & Krishna A. 2020. Detection of Covid- 19 patients using chest X-ray images with convolution neural network and mobile net. 3rd International Conference on Intelligent Sustainable Systems. Thoothukudi, India.

Hemdan E. E. D., Shouman M. A., & Karar M. E. 2020. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055

Das A.K., Ghosh S., Thunder S., Dutta R., Agarwal S. & Chakrabarti A. 2021. Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network, Pattern Anal. Appl. 1–14.

Islam M.T., Aowal M. A., Minhaz A.T., & Ashraf K. 2017. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850.

Lakhani P. & Sundaram B. 2017. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks, Radiology 284: 574-582.

Behzadi-khormouji, H., Rostami, H., Salehi, S., Derakhshande-Rishehri, T., Masoumi M., Salemi, S., Keshavarz, A., Gholamrezanezhad, A., Assadi,M., & Batouli, A. 2020. Deep learning, reusable and problem-based architectures for detection of consolidation on chest X-ray images. Computer Methods and Programs in Biomedicine, 185: 105162, ISSN 0169-2607.

Ma, J., Fan, X., Yang, S. X., Zhang, X. & Zhu, X. 2017. Contrast limited adaptive histogram equalization based fusion for underwater image enhancement, Preprints 2017, 2017030086. doi: 10.20944/preprints201703.0086.v1.

Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E. & Greenspan, H. 2015. Chest pathology detection using deep learning with non-medical training, Proc. - Int. Symp. Biomed. Imaging 294–297.

Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J. & Oermann, E.K. 2018. Confounding variables can degrade generalization performance of radiological deep learning models. arXiv preprint arXiv:1807.00431.

Baltruschat, I.M., Nickisch, H., Grass, M. et al. 2019. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. Sci Rep 9, 6381.

Kitamura, G. & Deible, C. 2020. Retraining an open-source pneumothorax detecting machine learning algorithm for improved performance to medical images, Clin. Imaging 61:15–19.

Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L. & Yang, Y. 2020. Thorax disease classification with attention guided convolutional neural network, Pattern Recognit. Lett. 131:38–45.

Karnewar, A. 2018 Back-to-basics part 1: Histogram equalization in image processing, Retrieved from: https://medium.com/@animeshsk3/back-to-basics-part-1-histogram-equalization-in-image-processing-f607f33c5d55. (Accessed: 29.12.2021)

Shibata, H., Hanaoka, S., Nomura, Y., Nakao, T., Sato, I., Sato, D., ... & Abe, O. 2020. A versatile anomaly detection method for medical images with a flow-based generative model in semi-supervision setting. arXiv preprint arXiv:2001.07847.