An Advanced Convolutional Neural Network for Detecting Chest X-ray Abnormalities

Fady Tawfik and Yi Gu*

***Abstract*—In the field of medical images diagnoses, doctors need a valuable second opinion when diagnosing thoracic diseases in chest X-rays. Existing methods of interpreting chest X-ray images classify them into a list of findings without specifying their locations on the images, resulting in uninterpretable results. Convolutional Neural Network (CNN) is a popular model for thoracic diseases diagnoses, which is a deep learning technique that has shown high accuracy in image classification and feature detection. In this work, an advanced CNN model is proposed to identify 14 findings in chest X-rays. For each test image, the intended CNN model should predict a bounding box and class for all findings. The classes range from 0 to 13, with each number corresponding to a specific disease in the dataset. The results have demonstrated that the proposed model outperforms the CapsNet model with an accuracy of 94% in X-ray images classification and labeling.**

***Index Terms*—Convolutional Neural Network (CNN), deep learning, CapsNet model, X-ray image diagnoses**

I. Introduction

Thoracic diseases are serious health problems that plague a significant amount of people. Many peopledie each year from thoracic diseases, such as aortic enlargement, calcification, and pleural effusion. Thoracic diseases account for more than 500,000 deaths annually in the United States, which is a high death rate [1]. Detecting the thoracic diseases early and correctly can help clinicians to improve patient treatment effectively. Computerized tomography scans and chest X-ray imaging are currently the most widely available radiological examinations for screening and detecting thoracic diseases. However, the basic chest X-ray scans are preferred as they are cheap and fast and expose the patients to little radiation compared to other types of scans [2]. Chest X-ray imaging is the first diagnostic step for detecting any chest abnormalities.

Developing an automated system to understand medical images and to diagnose thoracic diseases in chest X-ray images is needed. Currently, a large number of chest X-rays are analyzed almost completely through visual inspection by radiologists [2]. The traditional analysis and diagnosis based on chest X-rays largely rely on the medical experience of the radiologists. With the ever- increasing amount of chest X-ray images, handling the workload to read these images can result in longtreatment periods for patients as well as mistakes even for the best practicing doctors [3]. Therefore, it is important to have a computer-aided system to solve these issues and automatically detect different findings from chest X-ray images.

Deep learning is a subset of machine learning in artificial intelligence that uses network learning capability [4]. Deep learning has gained wide attention because of its ability to obtain informative feature representation, which can be applied to achieve our goals. Deep machine learning techniques are becoming more popular due to the ease of applying them on X-ray images to detect specific features. The ability of some deep learning models to analyze and detect images has even reached human-level accuracy. It also has been explored in medical image analysis for segmentation and classification tasks [5], which has become an important part inthe medical industry.

This paper proposes an advanced convolutional neural network model to classify thoracic diseases in chest X-ray images. A total of 14 thoracic diseases are being classified and detected. The proposed solution has a high accuracy rate in detecting and localizing findings in each X-ray image and can achieve the following two specific aims:

- Localize and classify different types of thoracic abnormalities from chest X-rays.
- Build a valuable second opinion for doctors that could help accurately identify and localize findings on chest radiographs. This is done by comparing the performance of our models withthe CapsNet model.

The rest of the paper is organized as follows. Section II presents the related work. Section III describes the methodology of the proposed CNN model. Results and performance analyses are discussed in Section IV. Section V concludes the work.

II. Related Work

Convolutional neural network (CNN) is a special architecture of artificial neural networks proposed by LeCun in 1988, which has been widely applied to a variety of pattern recognition, image classification,image recognition, language translation, medical diagnostics, etc. [1]. The success of the CNN model lays in the fact that it is able to capture hidden features of the images through its large number of hidden layers. In the literature, researchers have proposed various CNN classifiers, such as LeNet-5 in 1998 [6], AlexNet in 2012 [7], VGG in 2014 [8], Inception V3 in 2015 [9], ResNet in 2015 [10], and Xception in 2016 [11]. To overcome the performance limitations of image classification, the latest CNN architecture, capsules (CapsNet), was proposed by Sabour *et al.* [12] in 2017. One of the aims of this project is to compare the accuracy of our model with the CapsNet model.

Manuscript received August 29, 2022; revised December 26, 2022; accepted January 30, 2023.

Fady Tawfik was with the Department of Computer Science, Middle Tennessee State University, Murfreesboro, TN, USA. E-mail: ft2p@mtmail.mtsu.edu (F.T.)

Yi Gu is with the Department of Computer Science, Middle Tennessee State University, Murfreesboro, TN, USA.

*Correspondence: Yi.Gu@mtsu.edu (Y.G.)

doi: 10.18178/ijml.2023.13.4.1141

There are several challenges in the image processing of chest X-rays, as many of them have irrelevantregions, such as neck and arm regions. The main focus should be on the lesion location instead of the global image. This challenge requires the model to be more flexible when localizing lesions for feature extractions [13]. Another challenge is the detection of thoracic diseases and their classifications. The complexity of thoracic diseases and the limited quality of chest X-ray images make the detection process a challenging task. Most publicly available chest image datasets have labels, but not the locations of the abnormalities that existed in each case.

The classification of abnormalities in chest X-ray images is also a challenge, since chest images may contain multiple types of thoracic diseases, and their positions and sizes are usually highly diverse [14]. This paper aims to overcome those challenges and specify the location of findings in each X-ray image,not just classify the images.

A lot of research and effects have been devotedto the field of medical image classification for thoracic disease identification in chest X-rays. In the world of medical image classification to abnormality identification, deep learning-based frameworkLeNet-5 [15] was used to classify MRI images of autism spectrum disorder from normal controls. The architecture uses two convolutional layers and three fully connected ones with some good resultsin terms of specificity and sensitivity. Some papers have focused on extraction of features from chest X-ray images. A new method for automatic determination and calculation of the number of visible vertebrae in pulmonary X-ray images was developedin [16], for the task of tuberculosis detection. The proposed method and system were then evaluated by using three X-ray lung datasets. It has shown that using the proposed system of out-of-distribution detection enhances tuberculosis classification resultsby up to 1.3% for the same classification model. The proposed system also allows for automatic training of a composite model that considers the X-ray radiation level of the image, which is more effective compared to the traditional one-part model. A fully convolutional network and a deep convolutional neural network model were used in [4] to identify bacterial and viral pneumonia, which has segmentation, classification, and assembly phases. The overall result was low because of the unbalanced dataset but it has a better performanceon other features. Mao *et al*. [2] introduced a deep generative classifier architecture to diagnose thoracic diseases with chest X-ray images. The deep generative classifiers contained an encoder and a classifier, where the encoder is used to encode the input chest X-ray images and the classifier is to classify the features of each image.

Many experiments have been done to compare the performance of various CNN models. Karnkawinpong *et al*. [17] studied the performance of three CNN architectures that could help in early diagnosis of Tuberculosis infection. The authors compared between AlexNet, VGG-16 and CapsNet modelsin terms of classification accuracy. CapsNet wasfound to outperform other models as it is more robust to affine transformations compared to those original CNNs that use pooling layers. Devnath *et al*. [18] compared the performance of seven models to detect black lung. The models include VGG-16 [8],VGG-19 [19], Inception V3 [9], Xception [11], ResNet50 [10], DenseNet121 [20], and CheXNet [21]. They have observed that the CheXNet model performs better overall with a high accuracy of 85%,in comparison to other models.

Our paper is different from other projects as it has a consistent dataset that helps to get accurate results. The accuracy achieved by all above-mentioned models was no higher than 91%. We have demonstrated that our model can achieve a higher accuracy and a better performance in classifying thoracic diseases in chest X-ray images.

III. Methodology

This section describes the methods used in this work, including description of the dataset and the system that consists of data preprocessing as well as the model implementation and training.

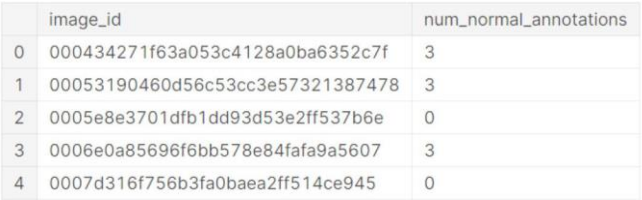
A. Dataset Description

In this work, VinDr-CXR [22] open dataset of chest X-rays with radiologists' annotation is used, which is collected from two major hospitals in Vietnam, Hospital 108 and the Hanoi Medical University Hospital. The dataset is divided into a training set of 15,000 and a testing set of 3,000. Each scan in the training set is labeled by 3 radiologists, while each scan in the testing set is labeled by 5 radiologists. At first, we check if the 3 radiologists' opinions differ on any image in the training dataset as shown in Figs. 1 and 2.

We confirm that 3 radiologists' opinions always match for normal-abnormal diagnosis. The results show either 0 (all radiologists think it is a normal case) or 1 (all radiologists think it is an abnormal case). Some public datasets of chest X-ray images, including ChestX-ray14 [14], Padchest [23] and CheXpert [24], have been released, and they depend on automated rule-based labelers to extract disease labels from images. The automated rule-based tool can produce labels on a large scale but introduces uncertainty and errors [25]. VinDr-CXR is found to be the best choice among these datasets as the images are manually annotated by a group of 17 experienced radiologists, making it more consistent and accurate compared to others.

```
// code to clean the images from unnecessary parts
train_df['area'] = (train_df['x_max'] - train_df['x_min']) *
    (train_df['y_max'] - train_df['y_min'])
valid_df['area'] = (valid_df['x_max'] - valid_df['x_min']) *
    (valid_df['y_max'] - valid_df['y_min'])
train_df = train_df[train_df['area'] > 1]
valid_df = valid_df[valid_df['area'] > 1]
train_df = train_df[(train_df['class_id'] > 1) & (train_df['class_id'] <
    15)]
valid_df = valid_df[(valid_df['class_id'] > 1) & (valid_df['class_id'] <
    15)]
train_df = train_df.drop(['area'], axis = 1)
train_df.shape
```

Fig. 1. Code to check radiologists' opinions on X-ray images if normal or not.



	image_id	num_normal_annotations
0	000434271f63a053c4128a0ba6352c7f	3
1	00053190460d56c53cc3e57321387478	3
2	0005e8e3701dfb1dd93d53e2ff537b6e	0
3	0006e0a85696f6bb578e84fafa9a5607	3
4	0007d316f756b3fa0baea2ff514ce945	0

Fig. 2. Radiologists' opinions in case of normal/abnormal scans.

B. Pre-processing of Images

In this phase, data is analyzed and prepared for training. First, we created a histogram in Fig. 3 that indicates the number of X-ray scans for each thoracic disease that exists in the dataset as wellas the number of clear/normal scans. This step is helpful to indicate the percentage of training and validation set split.

Secondly, the training-validation set is further randomly split into a training set and a validation set, where 2/3 is the training set and 1/3 is the validation set. The training set is used to train the model and the validation set is used to get a sense of how accurate the model is on the images that are not being used in the training. A split() method is used for the split process in Fig. 4.

Thirdly, data cleaning is applied to the original images. In our experiments, we use Pandas and NumPy [26] software libraries in python for data cleaning, which is an image processing technique that removes extra parts on both sides of the images. Unnecessary parts in the image are cropped according to the specified position and dimensions. The drop() function is called on our object, passing in the area parameter. The axis parameter is 1. This tells Pandas that we want the changes to be made directly in our object so it should look for the values to be dropped in the columns and rows of the object as shown in Fig. 5.

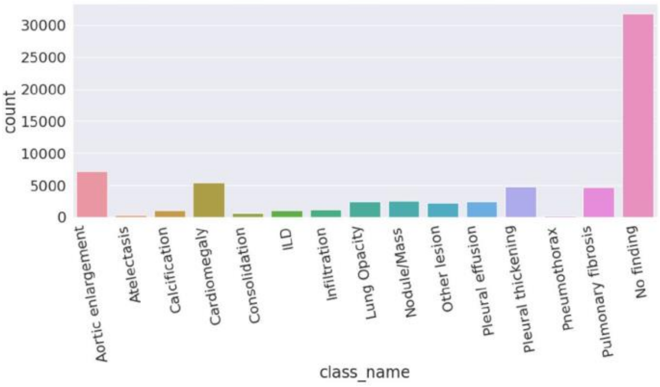
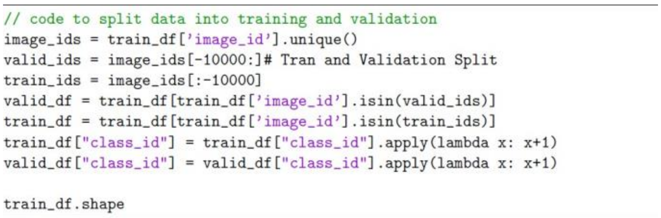


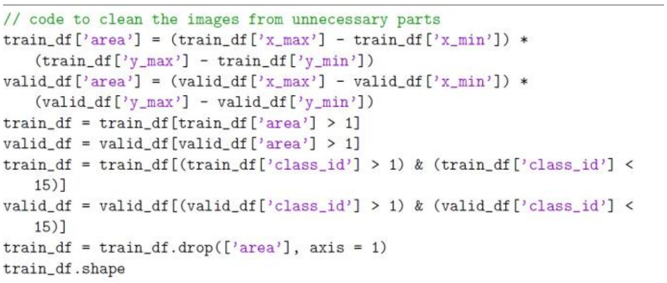
Fig. 3. Histogram shows the number of each finding in the dataset.



```
// code to split data into training and validation
image_ids = train_df[’image_id’].unique()
valid_ids = image_ids[-10000:]# Tran and Validation Split
train_ids = image_ids[:-10000]
valid_df = train_df[train_df[’image_id’].isin(valid_ids)]
train_df = train_df[train_df[’image_id’].isin(train_ids)]
train_df["class_id"] = train_df["class_id"].apply(lambda x: x+1)
valid_df["class_id"] = valid_df["class_id"].apply(lambda x: x+1)

train_df.shape
```

Fig. 4. Code to split data into training and validation sets.

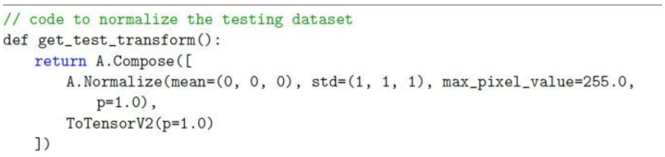


```
// code to clean the images from unnecessary parts
train_df[’area’] = (train_df[’x_max’] - train_df[’x_min’]) *
    (train_df[’y_max’] - train_df[’y_min’])
valid_df[’area’] = (valid_df[’x_max’] - valid_df[’x_min’]) *
    (valid_df[’y_max’] - valid_df[’y_min’])
train_df = train_df[train_df[’area’] > 1]
valid_df = valid_df[valid_df[’area’] > 1]
train_df = train_df[(train_df[’class_id’] > 1) & (train_df[’class_id’] <
    15)]
valid_df = valid_df[(valid_df[’class_id’] > 1) & (valid_df[’class_id’] <
    15)]
train_df = train_df.drop([’area’], axis = 1)
train_df.shape
```

Fig. 5. Code to clean the images from unnecessary parts.

C. Model Implementation and Training

All experiments for model implementation andtraining are conducted on the Kaggle Notebook kernels. They are essentially Jupyter notebooks in the browser which obtain their processing power from servers in the cloud instead of the local machine. The advanced CNN model is built using the Keras library and contains 3 types of layers, i.e., Convolutional, Max pooling, and SoftMax. Theconvolutional layer is combined with the ReLU function that transfers the summed weighted input from the node into the activation of the node or output for that input image. The output from the convolutional layer is the feature map. The max-pooling layer reduces the dimensions of the map produced by the previous layer. The SoftMax layer converts a real vector to a vector of categorical probabilities and is used for the last layer of a classification network because the result could be interpreted as a probability distribution. After layers are implemented, normalization is done on training, validation, and testing sets as shown in Fig. 6.



```
// code to normalize the testing dataset
def get_test_transform():
    return A.Compose([
        A.Normalize(mean=(0, 0, 0), std=(1, 1, 1), max_pixel_value=255.0,
            p=1.0),
        ToTensorV2(p=1.0)
    ])
```

Fig. 6. Code to normalize the testing dataset.

Batch normalization is used to reduce training time by normalizing the inputs to a layer for each mini-batch. A sigmoid function is the last step of themachine learning model that can convert the output into a probability score which is easier to work with and interpret. After combining all the layers andthe activation functions together, the final model is shown in Fig. 7.

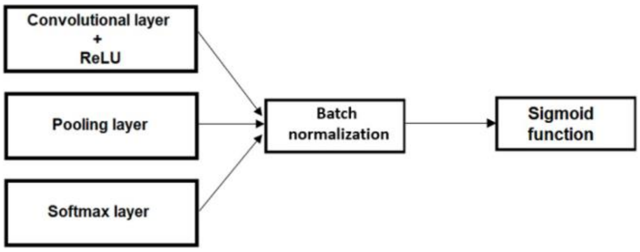
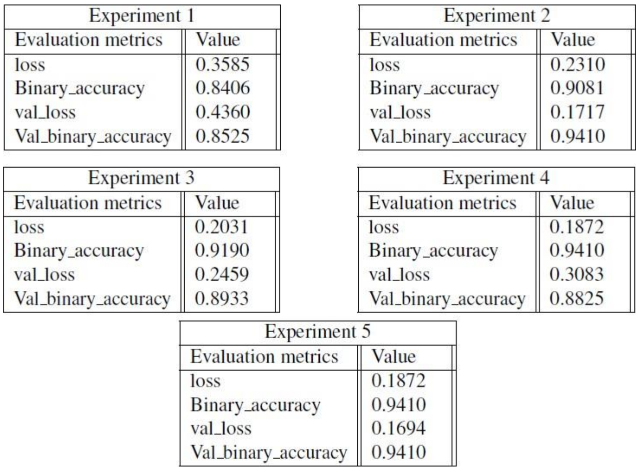


Fig. 7. The advanced CNN model architecture.



Experiment 1	
Evaluation metrics	Value
loss	0.3585
Binary_accuracy	0.8406
val_loss	0.4360
Val_binary_accuracy	0.8525

Experiment 2	
Evaluation metrics	Value
loss	0.2310
Binary_accuracy	0.9081
val_loss	0.1717
Val_binary_accuracy	0.9410

Experiment 3	
Evaluation metrics	Value
loss	0.2031
Binary_accuracy	0.9190
val_loss	0.2459
Val_binary_accuracy	0.8933

Experiment 4	
Evaluation metrics	Value
loss	0.1872
Binary_accuracy	0.9410
val_loss	0.3083
Val_binary_accuracy	0.8825

Experiment 5	
Evaluation metrics	Value
loss	0.1872
Binary_accuracy	0.9410
val_loss	0.1694
Val_binary_accuracy	0.9410

Fig. 8. Details of the five experiments.

For training and testing of the deep learning model on the dataset, k-fold cross validation is used to estimate the performance of the model on unseen data. Each experiment contains a different set of training and testing images based on the 5-fold cross validation. A comparison is carried out among all models' performance for automated thoracic diseases detection. In the experiments, the proposed model uses the Keras automated generator during the training. Details of the five experiments are tabulated in Fig. 8.

IV. Results and Discussion

As a result of the training phase, the proposed model has successfully classified and localized 6000 X-ray images with an accuracy of 94% success after7 cycles. The classification is done by mapping each X-ray image to a set of class scores, with each score corresponding to the probability that the input image belongs to a particular class. The class with the highest score is chosen as the final prediction. The localization is done by training the model to predict the coordinates of the bounding boxes, such as thex and y coordinates of the top left corner, the width and height of the box, and the class label of the object inside the box. Rectangular boxes and labels are created to be placed on each X-ray image in the dataset to locate the findings as plotted in Fig. 9.

The results have obtained satisfactory values for both accuracy and loss factor. During training, the accuracy value has been changing over time, and has reached 94% after the last cycle. In the accuracy graph shown in Fig. 10, the line with square markers shows the training accuracy performance of the network, while the line with circle markers shows the test accuracy performance of the network obtained as a result of each cycle. The accuracy is calculated using the BinaryAccuracy function during each training cycle by dividing the number of correct predictions (i.e., true positives and true negatives) by the total number of predictions.

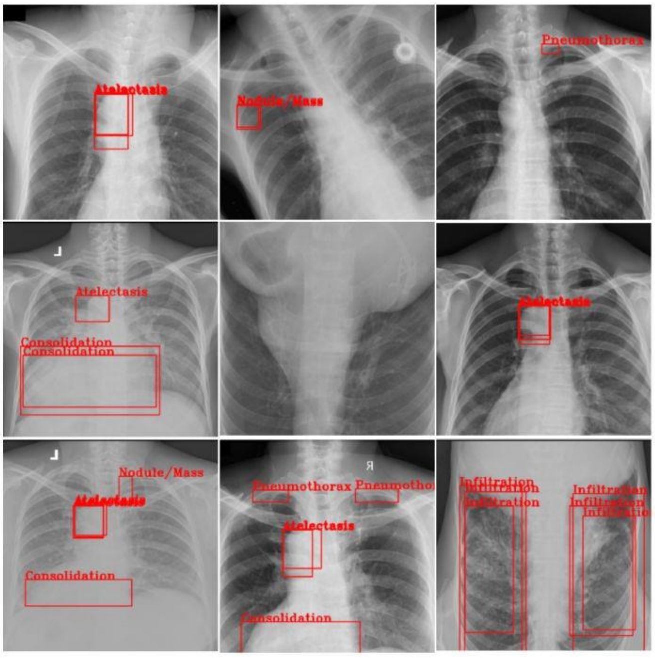


Fig. 9. Some examples of X-ray images classified and localized by the model during training, and the image in the middle indicates no findings found (normal).

The loss factor has also increased and decreased over the time. It started with 0.35 and ended with about 0.2 in the last cycle. In the loss factor graph shown in Fig. 11, the line with square markers shows the training loss value of the network, while the line with circle markers indicates the test loss value of the network obtained as a result of each cycle. The loss is calculated using the Binary Cross-entropy class that computes the cross-entropy loss between true labels and predicted labels. It is defined as:

$$loss = y * log(p) + (1 - y) * log(1 - p),$$

where y is the true label, p is the predicted probability of the positive class. The goal of the trainingis to minimize this loss function.

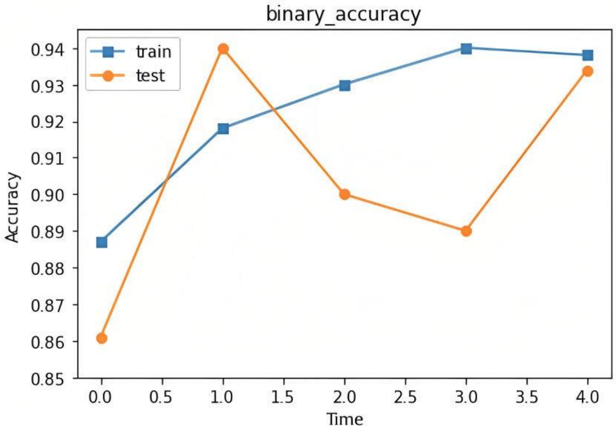


Fig. 10. The accuracy graph for the training and testing performance of the advanced CNN model.

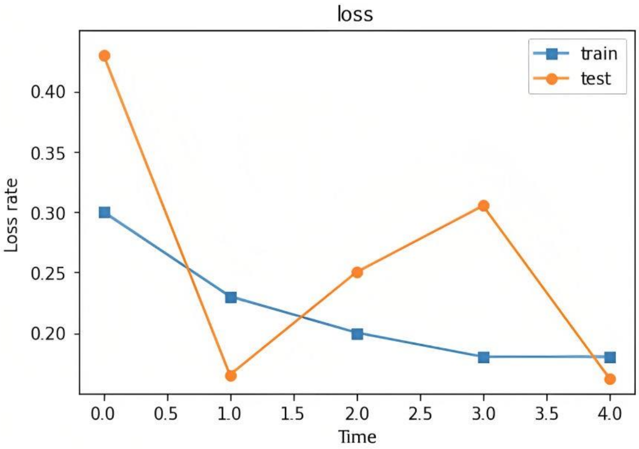
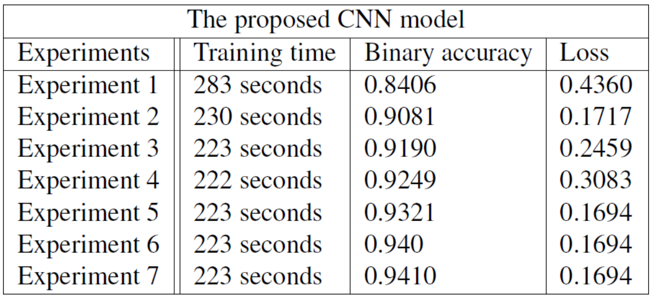


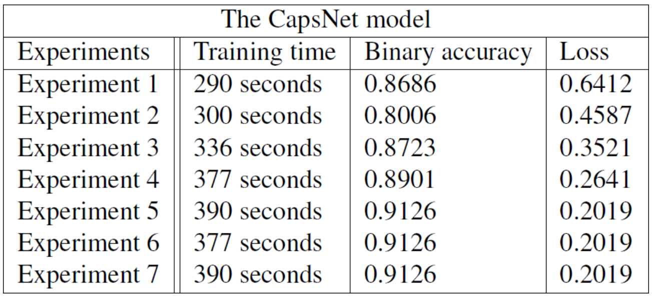
Fig. 11. The binary loss graph for the training and testing of the advanced CNN model.

The performance of the proposed model is compared with the popular CapsNet model in terms of accuracy and loss value. After 7 training cycles for both models, our model has reached a higher accuracy and a lower loss value. The comparisonis done using the VinDr-CXR [26] dataset on both models. The hardware specifications that were used to run both models, Processor: Core i5 4590, RAM: 16 GB, GPU: GTX1060 6 GB. Our model has achieved a higher accuracy of 94% while the CapsNet model is only 91.26%. Meanwhile, our model has achieved a lower loss value of 0.16 while the CapsNet model is 0.20. The detailed training time, accuracy, and loss value are plotted in Figs. 12 and 13, respectively, for a comparison purpose between the proposed CNN model and the CapsNet model.



The proposed CNN model			
Experiments	Training time	Binary accuracy	Loss
Experiment 1	283 seconds	0.8406	0.4360
Experiment 2	230 seconds	0.9081	0.1717
Experiment 3	223 seconds	0.9190	0.2459
Experiment 4	222 seconds	0.9249	0.3083
Experiment 5	223 seconds	0.9321	0.1694
Experiment 6	223 seconds	0.940	0.1694
Experiment 7	223 seconds	0.9410	0.1694

Fig. 12. The training time, accuracy, and loss value of the proposed CNN model.



The CapsNet model			
Experiments	Training time	Binary accuracy	Loss
Experiment 1	290 seconds	0.8686	0.6412
Experiment 2	300 seconds	0.8006	0.4587
Experiment 3	336 seconds	0.8723	0.3521
Experiment 4	377 seconds	0.8901	0.2641
Experiment 5	390 seconds	0.9126	0.2019
Experiment 6	377 seconds	0.9126	0.2019
Experiment 7	390 seconds	0.9126	0.2019

Fig. 13. The training time, accuracy, and loss value of the CapsNet model.

V. Conclusion and Future Work

Chest disease is an area where mortality rates are high all over the world. Studies of detection and classification of thoracic diseases are important. In this study, chest X-ray images taken from the VinDr-CXR dataset were performed as input data to the implemented model using data normalization and data cleaning techniques. In this research, an advanced Convolutional Neural Network wasproposed, which performs thoracic diseases classification and localization on chest X-ray images with a low loss rate of 0.2 after 7 training cycles. Our model outperforms the CapsNet model with an accuracy of 94% and 91.3% for the CapsNet model. This model can have a lasting impact on the fieldof medical images diagnoses as it produces a higheraccuracy than many other popular CNN models.The COVID-19 pandemic has highlighted the need to pull all available resources to fight against this virus. Because this virus affects an infected patient'slung, interpreting images can be an alternative for diagnosis. This work can help to interpret chest X-ray images and detect some early abnormalities, as well as be useful to the research community in the field of COVID-19 diagnosis.

As the future work, a bigger dataset will be used for more training of the CNN model to get a higher classification performance. With a high accuracy and a low loss rate of the current model, more training can be conducted to reach an even higher accuracy in the future. Additionally, 10-fold cross-validation can be employed to obtain a better approximation of optimum model accuracy since only 5-fold was done in this research.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Fady Tawfik conducted this research under Dr. Yi Gu's direct guidance and supervision. Fady Tawfik wrote the paper. Dr. Yi Gu edited and refined the paper. Both authors had approved the final version.

References

[1] P. Kieu, H. Tran, T. Le, and T. Nguyen, "Applying multi-cnns model for detecting abnormal problem on chest x-ray images," in *Proc. KSE: 10th International Conference on Knowledge and Systems Engineering*, pp. 300–305, 2018.

[2] C. Mao, L. Yao, Y. Pan, Y. Luo, and Z. Zeng, "Deep generative classifiers for thoracic disease diagnosis with chest x-ray images," in *Proc. IEEE International Conference on Bioinformatics and Biomedicine*, pp. 1209–1214, 2018.

[3] P. Kittiworapanya and K. Pasupa, "An image segment-based classification for chest x-ray image," in *Proc. CSBio the Eleventh International Conference on Computational Systems-Biology and Bioinformatics*, pp. 68–74, 2020.

[4] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," in *Proc. ICMP: International Conference on Multimedia and Image Processing*, pp. 88–93, 2018.

[5] L. Devnath, S. Luo, P. Summons, and D. Wang, "Performance comparison of deep learning models for black lung detection on chest x-ray radiographs," pp. 150–154, 2020.

[6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient- based learning applied to document recognition," in *Proc. the IEEE*, vol. 86, pp. 2278–2324, 1998.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Informa- tion Processing Systems*, vol. 25, 2012.

[8] K. Simonyan and A .Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2016.

[10] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, pp. 770–778, 2016.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," pp. 1800–1807, 2017.

[12] S. Sabour, N. Frosst, and G. Hinton, *Dynamic Routing between Capsules*, 2017.

[13] V. Teixeira, L. Braz, H. Pedrini, and Z. Dias, *Dualanet: Dual Lesion Attention Network for Thoracic Disease Classification in Chest X-rays*, pp. 69–74, 2020.

[14] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestxray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3462–3471, 2017.

[15] X. Yang, S. Sarraf, and N. Zhang, "Deep learning-based framework for autism functional mri image classification," *Journal of the Arkansas Academy of Science*, vol. 72, pp. 47–53, 2018.

[16] A. A. Dovganich *et al*., "Automatic out-of- distribution detection methods for improving the deep learning classification of pulmonary X-ray images," *Journal of Image and Graphics*, vol. 10, no. 2, pp. 56–63, June 2022.

[17] T. Karnkawinpong and Y. Limpiyakorn, "Chest x-ray analysis of tuberculosis by convolutional neural networks with affine transforms," in *Proc. CSAI: 2nd International Conference on Computer Science and Artificial Intelligence*, pp. 90–93, 2018.

[18] L. Devnath, S. Luo, P. Summons, and D. Wang, "Performance comparison of deep learning models for black lung detection on chest x-ray radiographs," pp. 150–154, 2020.

[19] B.-D. Liu, J. Meng, W.-Y. Xie, S. Shao, Y. Li, and Y. Wang, "Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification," *Remote Sensing*, vol. 11, p. 518, 2019.

[20] G. Huang, Z. Liu, L. Maaten, and K. Weinberger, *Densely Connected Convolutional Networks*, 2017.

[21] P. Rajpurkar *et al*., *Chexnet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning*, 2017.

[22] H. Q. Nguyen *et al*., "Vindr-cxr: An open dataset of chest x-rays with radiologists annotations," presented at VinDr: Vingroup Big Data Institute conference, 2020.

[23] A. Bustos, A. Pertusa, M. Salinas, and M. Vaya, "Padchest: A large chest x-ray image dataset with multi-label annotated reports," *Medical Image Analysis*, vol. 66, 101797, 2020.

[24] J. Irvin *et al*., “Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in *Proc. the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 590–597, 2019.

[25] L. Oakden-Rayner, “Exploring large-scale public medical image datasets,” *Academic Radiology*, vol. 27, 2019.

[26] Y. H. Li and Z. Yu, “Research on data cleaning technology based on instance level,” *Journal of Physics: Conference Series*, 1213, 2019.