Pneumonia Detection in Chest X-ray Images using Convolutional Neural Networks

1st Esteban J. Palomo
Department of Computer Languages and Computer Science
University of Malaga
Malaga, Spain
ejpalomo@lcc.uma.es

2nd Miguel A. Zafra-Santisteban
Department of Computer Languages and Computer Science
University of Malaga
Malaga, Spain
miguelzafra98@gmail.com

3rd Rafael M. Luque-Baena
Department of Computer Languages and Computer Science
University of Malaga
Malaga, Spain
rmluque@lcc.uma.es

***Abstract*—Pneumonia is an infectious and deadly disease which strikes over millions of people. Usually, chest X-rays are used by radiotherapist to diagnose pneumonia. In this paper, a Computer-Aided Diagnosis (CAD) system for pneumonia detection in chest X-ray images is proposed. This system is based on Convolutional Neural Networks (CNNs) which are able to classify the image into two classes (pneumonia or normal). Experimental results show that the proposed system obtained an accuracy rate of 98.59%.**

***Index Terms*—pneumonia detection, chest X-ray images, convolutional neural networks, computer-aided diagnosis**

I. Introduction

Pneumonia is a respiratory infection affecting one or both lungs in humans commonly caused by bacteria or viruses [1]. Pneumonia is most common in under developed and developing countries, where unhygienic environmental conditions and the shortage of medical resources exacerbate the problem. Thus, early diagnosis of pneumonia is crucial to ensure curative treatment and increase survival rates. Chest X-Rays are a non-invasive and inexpensive method to examine the lungs, which need expert radiotherapists for evaluation. However, the examination of chest radio-graphs is not easy for a radiotherapist, since pneumonia can mimic with many other problems like congestive heart failure, lung scarring, etc. Therefore, Computer-Aided Diagnosis (CAD) systems that can detect pneumonia in chest X-ray images are more and more used by the medical personnel.

Convolutional Neural Networks (CNNs) have been extensively used for various image classification problems and, of course, for developing CAD systems for pneumonia detection. In [2], a diagnostic tool based on a deep-learning framework for the screening of patients with common treatable blinding retinal diseases was established. They also applied their AI system for the diagnosis of pediatric pneumonia using chest X-ray images. They used an InceptionV3 architecture pretrained on the ImageNet dataset. Another work employed pretrained CNN models along with supervised classifier algorithms to analyze chest X-ray images for pneumonia detection [3]. In [4], six convolutional neural networks were used for pneumonia detection, where two of them were proposed by the authors and the rest were pretrained models, namely, VGG-16, VGG-19, ResNet50, and InceptionV3. An ensemble of three convolutional neural networks (GoogLeNet, ResNet18, and DenseNet-121) was proposed as a CAD system for pneumonia detection [5]. In [6], they identified the presence of pneumonia using U-Net architecture based segmentation and classified the pneumonia as normal and abnormal (bacteria, viral) using pretrained models such as ResNet50, InceptionV3, Inception-ResNetV2. Their results were analyzed and compared with other CNN models such as DenseNet-169 + SVM, VGG-16, RetinaNet + Mask RCNN, VGG-16 and Xception, Fully connected RCNN, etc using various measures. Finally, in [7] the compound scaled ResNet50, which is the upscaled version of ResNet50, was used for pneumonia detection.

In this work, a CAD system for pneumonia detection in chest X-ray images based on CNNs is proposed. This system is capable of classifying chest X-rays images into two different categories: normal, and pneumonia. Also, an application web was developed to make easy the use of our proposed system by the medical personnel, in which the user can upload a chest X-ray image and the web shows whether pneumonia was detected

This work is partially supported by the Ministry of Science, Innovation and Universities of Spain under grant RTI2018-094645-B-I00, project name Automated detection with low-cost hardware of unusual activities in video sequences. It is also partially supported by the Autonomous Government of Andalusia (Spain) under project UMA18-FEDERJA-084, project name Detection of anomalous behavior agents by deep learning in low-cost video surveillance intelligent systems. It is also partially supported by the Autonomous Government of Andalusia (Spain) under project UMA20-FEDERJA-108, project name Detection, characterization and prognosis value of the non-obstructive coronary disease with deep learning. All of them include funds from the European Regional Development Fund (ERDF). It is also partially supported by the University of Malaga (Spain) under grants B1-2019_01, project name Anomaly detection on roads by moving cameras, and B1-2019_02, project name Self-Organizing Neural Systems for Non-Stationary Environments.

in the image or not together with the accuracy obtained by each considered model. Figure 1 shows a view of the developed web app where pneumonia is detected.

The remainder of this paper is organized as follows. The methodology used in this work is presented in Section II. Experimental results obtained by the different models are provided in Section III. Finally, Section IV is devoted to conclusions.

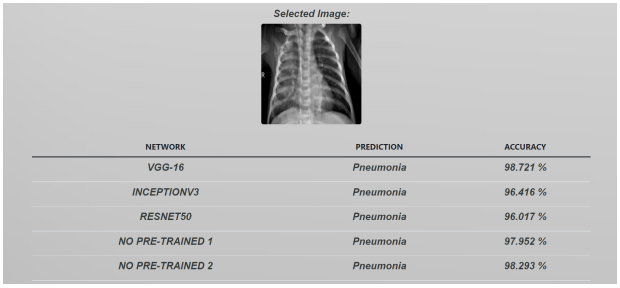


Fig. 1. A view of the developed web application.

II. Methodology

Our system consists of an application web developed with Flask, where we have five different trained CNN models. When an image is presented to the system, each of these models makes a prediction to classify the image into pneumonia or normal categories, providing the accuracy obtained during the prediction. A representation of this system can be seen in Figure 2.

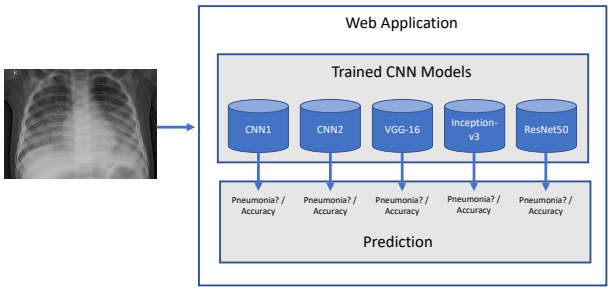


Fig. 2. A representation of the developed system.

Two different convolutional neural networks architectures were designed for pneumonia detection in chest X-ray images. The architectures were determined following a trial and error-based empiric optimization method. We also selected three pretrained models, namely VGG16, Inception-v3, and ResNet50. A detailed description of each model used in this paper is given below:

A. CNN 1

The first model consists of 3 convolutional layers of 3×3 kernel sizes, the first convolutional layer has 32 feature maps, the second convolutional layer has 64 feature maps, and the third convolutional layer has 128 features map, all of them employing the ReLU function. Max-pooling layers of 2×2 dimensions are used after each convolutional layer. A flattening layer is placed behind these layers. Three dense layers are used, the first dense layer has 120 output perceptrons employing ReLU, the second dense layer has 60 output perceptrons also employing ReLU, and the third dense layer has one output perceptron using the sigmoid function. Finally, a dropout layer is placed between the second and third dense layers. The summary of this architecture is shown in Fig. 3.

Layer (type)	Output Shape	Param #
conv2d_18 (Conv2D)	(None, 150, 150, 32)	320
max_pooling2d_18 (MaxPooling2D)	(None, 50, 50, 32)	0
conv2d_19 (Conv2D)	(None, 50, 50, 64)	18496
max_pooling2d_19 (MaxPooling2D)	(None, 25, 25, 64)	0
conv2d_20 (Conv2D)	(None, 25, 25, 128)	73856
max_pooling2d_20 (MaxPooling2D)	(None, 8, 8, 128)	0
flatten_6 (Flatten)	(None, 8192)	0
dense_18 (Dense)	(None, 120)	983160
dense_19 (Dense)	(None, 60)	7260
dropout_6 (Dropout)	(None, 60)	0
dense_20 (Dense)	(None, 1)	61

Fig. 3. Summary of the first proposed model (CNN 1).

B. CNN 2

The second model is similar to the first model but it is deeper. It consists of 4 convolutional layers of 3×3 kernel sizes, the first convolutional layer has 16 feature maps, the second convolutional layer has 32 feature maps, and the third convolutional layer has 64 feature maps, and the fourth convolutional layer has 128 features maps, all of them employing the ReLU function. Max-pooling layers of 2×2 dimensions are used after each convolutional layer. A flattening layer is placed behind these layers. Five dense layers are used, the first dense layer has 550 output perceptrons, the second dense layer has 400 output perceptrons, the third dense layer has 300, the fourth dense layer has 200, and the fifth dense layer has one output perceptron. All of these dense layers emply the ReLU function, except the last one which uses the sigmoid function. Finally, two dropout layers are placed after the first and the fourth dense layers. The summary of this architecture is shown in Fig. 4.

C. VGG16

VGG16 is a convolutional neural network that stood out in the localisation and classification tracks on the ImageNet Challenge 2014 [8]. This model has 16 layers in total and uses very small (3×3) convolution filters replacing larger convolution filters used in earlier models. Spatial pooling is carried out by five max-pooling layers, which follow some of the convolutional layers. The stack of convolutional layers is followed by three dense layers, and a sigmoid function. We have added an additional layer, so that the dense layers

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 16)	160
max_pooling2d (MaxPooling2D)	(None, 74, 74, 16)	0
conv2d_1 (Conv2D)	(None, 72, 72, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 32)	0
conv2d_2 (Conv2D)	(None, 34, 34, 64)	18496
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 64)	0
conv2d_3 (Conv2D)	(None, 15, 15, 128)	73856
max_pooling2d_3 (MaxPooling2D)	(None, 7, 7, 128)	0
flatten (Flatten)	(None, 6272)	0
dense (Dense)	(None, 550)	3450150
dropout (Dropout)	(None, 550)	0
dense_1 (Dense)	(None, 400)	220400
dense_2 (Dense)	(None, 300)	120300
dense_3 (Dense)	(None, 200)	60200
dropout_1 (Dropout)	(None, 200)	0
dense_4 (Dense)	(None, 1)	201

Fig. 4. Summary of the second proposed model (CNN 2).

have 1024, 512, 256, and 1 channel, respectively. Finally, a dropout layer between the second and third dense layers has been included.

D. Inception-v3

Inception-v3 is a convolutional neural network used for image classification which consists of 42 layers [9]. It has multiple variants such as inception-v1/google, inception-v2, inception-v3, and inception-v4. Inception-v3 is the variant used in this work. This model uses auxiliary classifiers to avoid or prevent the activation of each layer to converge to zero, batch normalization to address the problem of vanishing gradients and zero activations by reducing the internal covariate shift, and additional factorization to reduce the number of connections/parameters of the network without decreasing the network efficiency. Again, we used dense layers with 1024, 512, 256, and 1 channel followed by a sigmoid function, and a dropout layer between the second and third dense layers.

E. ResNet50

ResNet stands for residual network and is primarily used for image classification. This model won the 1st place on the ILSVRC 2015 classification task [10]. This model has multiple variants, such as ResNet50, ResNeXt, ResNet34, ResNetV2, etc. The variant used in this work is ResNet50, which has 50 layers. It also uses 3×3 convolution filters. The novelty of this model lies in the use of shortcut connections to address the problems of degrading accuracy and vanishing gradients present in deep neural networks. These shortcut connections allow the network to skip layers irrelevant for training. This reduces the training error and helps the network to converge faster in comparison to other networks. We have also used dense layers with 1024, 512, 256, and 1 channel followed by a sigmoid function. A dropout layer has been added between the second and third dense layers.

III. Experimental Results

A. Dataset

The dataset used is the "Chest X-Ray Images (Pneumonia)" available on Kaggle [11]. This dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). This dataset contains 5,863 images (JPEG) which are grayscale.

B. Performance measures

The accuracy, precision, recall, and F1-score were selected as performance measures. The accuracy provides an overall measure of the number of corrected predictions of the model, which is defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

where TP, FP, TN, and FN are the true positive, false positive, true negative, and false negative, respectively. The precision and recall values provide more insight into the performance of the model. Precision shows the accuracy of the model's positive label prediction. This provides the ratio of the correct predictions to the total predictions yielded by the model. Conversely, recall (or sensitivity) measures the percentage of ground truth positives that the model correctly predicted. Precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Finally, F1-Score provides a balance between precision and recall, considering both FPs and FNs, which is given as follows:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

C. Experimental setup

In order to perform a fair comparison among the five models used in this work, hyperparameters were set to the same values for all of them. Adam optimizer has been used with binary cross-entropy as the cost function. Learning rate was set to 0.001. As mentioned in Section II, for the three pretrained models (VGG16, ResNet50, and Inception-v3) we used four dense layers with 1024, 512, 256, and 1 channel followed by a sigmoid function. Also, a dropout layer has been added between the second and third dense layers. Each model was trained during a different number of epochs determined by the early stopping method with a patience of 5 and a maximum number of epochs of 50.

D. Comparative results

The five models were trained and tested on the selected dataset. We selected the VGG-16, InceptionV3, and ResNet50 as pretrained CNN models. We also designed two non-pretrained CNN architectures (CNN1 and CNN2) to compare the performance among pretrained and non-pretrained models.

The training was done during 10 runs. The model accuracy and loss graphs for the best run of each model are shown in Figure 5. The mean accuracy and loss obtained during training and validation are detailed in Table I. As we can see in this table, VGG-16 is the model with best accuracy and loss both, during training and validation. The second and third best models are CNN1 and CNN2, respectively. Moreover, confusion matrices for each model are given in Fig. 6, where we can see the errors made by the five models (false negatives and false positives).

TABLE I
ACCURACY, AND LOSS RESULTS OBTAINED DURING TRAINING AND VALIDATION BY EACH MODEL. BEST RESULTS ARE IN BOLD. STANDARD DEVIATIONS ARE SHOWN IN PARENTHESIS.

	Training		Validation	
Model	**Accuracy**	**Loss**	**Accuracy**	**Loss**
CNN1	0.9634 (0.03221)	0.0961 (0.0749)	0.9663 (0.0123)	0.1033 (0.0289)
CNN2	0.9599 (0.0397)	0.1036 (0.0878)	0.9594 (0.0128)	0.1330 (0.0410)
VGG-16	**0.9784 (0.0289)**	**0.0559 (0.0647)**	**0.9705 (0.0099)**	**0.0858 (0.0372)**
Inception-v3	0.9567 (0.0290)	0.1214 (0.0866)	0.9470 (0.0114)	0.1592 (0.0387)
ResNet50	0.8717 (0.0763)	0.2899 (0,1455)	0.8848 (0.0708)	0.2638 (0.1263)

After the training, the test dataset was fed to the different models. Table II shows the mean accuracy, precision, recall, and F1-score for the 10 runs, where the standard deviations are shown in parenthesis and the best results are in bold. Note that VGG-16 obtained the best results for all the performance measures. After the VGG-16 model, the two CNN models trained from scratch (CNN1, and CNN2) achieved the second best results outperforming the rest of the pretrained models.

TABLE II
ACCURACY, PRECISION, RECALL, AND F1-SCORE RESULTS OBTAINED DURING TEST BY EACH MODEL. BEST RESULTS ARE IN BOLD. STANDARD DEVIATIONS ARE SHOWN IN PARENTHESIS.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
CNN1	0.9742 (0.0039)	0.9727 (0.0089)	0.9629 (0.0277)	0.9676 (0.0157)
CNN2	0.9744 (0.0043)	0.9723 (0.0098)	0.9638 (0.0270)	0.9678 (0.0158)
VGG-16	**0.9859 (0.0017)**	**0.9830 (0.0105)**	**0.9833 (0.0087)**	**0.9832 (0.0072)**
InceptionV3	0.9604 (0.0029)	0.9594 (0.0133)	0.9415 (0.0451)	0.9497 (0.0236)
ResNet50	0.9474 (0.0137)	0.9339 (0.0382)	0.9381 (0.0298)	0.9359 (0.0327)

In addition, our proposal was compared with other state-of-the-art methods applied to pneumonia detection in chest X-ray images. This comparison is presented in Table III. By observing this table, we can see how the methodology based on an ensemble of three CNN models achieved the best accuracy result (98.81%). However, our best result achieved by the VGG-16 model yielded a similar result (98.59%) being a more simple model, since the three deep learning models used in the ensemble are GoogLeNet, ResNet18, and DenseNet-121. Furthermore, our two proposed CNN models also obtained very close results (97.42%, and 97.44%, respectively) with shallower architectures as commented before.

TABLE III
COMPARISON OF THE PROPOSED METHOD WITH THE EXISTING STATE-OF-THE-ART METHODS USING THE ACCURACY METRIC. BEST RESULTS ARE IN BOLD.

Methodology	**Topic**	**Accuracy**
InceptionV3	Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning [2]	92.8%
CNN Models along with DenseNet-169 and SVM	Pneumonia Detection Using CNN based Feature Extraction [3]	80.02%
VGG-19	Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning [4]	88.46%
Ensemble of three CNN models	**Pneumonia detection in chest X-ray images using an ensemble of deep learning models [5]**	**98.81%**
ResNet50	Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures [6]	93.06%
Compound Scaled ResNet50	Pneumonia detection in chest X-ray images using compound scaled deep learning model [7]	98.14%
Proposed VGG-16	—	98.59%

IV. CONCLUSIONS

In this paper, a Computer Aided Diagnosis (CAD) system for pneumonia detection is proposed. This systems is based on Convolutional Neural Networks (CNNs) which classifies chest X-rays images into pneumonia or normal classes. Two CNN architectures have been proposed for this purpose with 6 layers and 9 layers, respectively, which were trained from scratch. We also selected three pretrained CNN models, namely, VGG-16, Inception-v3, and ResNet50, where dense layers were changed. Experimental results on the "Chest X-Ray Images (Pneumonia)" dataset shows that our system achieves an accuracy rate of 98.59%, concretely this result is achieved by the VGG-16 model. However, our two proposed models obtained the best results after the VGG-16 and using shallower architectures, especially the CNN1 model which only uses 6 layers. The comparison performed with the existing state-of-the-art methods demonstrates the suitability and good performance in pneumonia detection of our proposal.

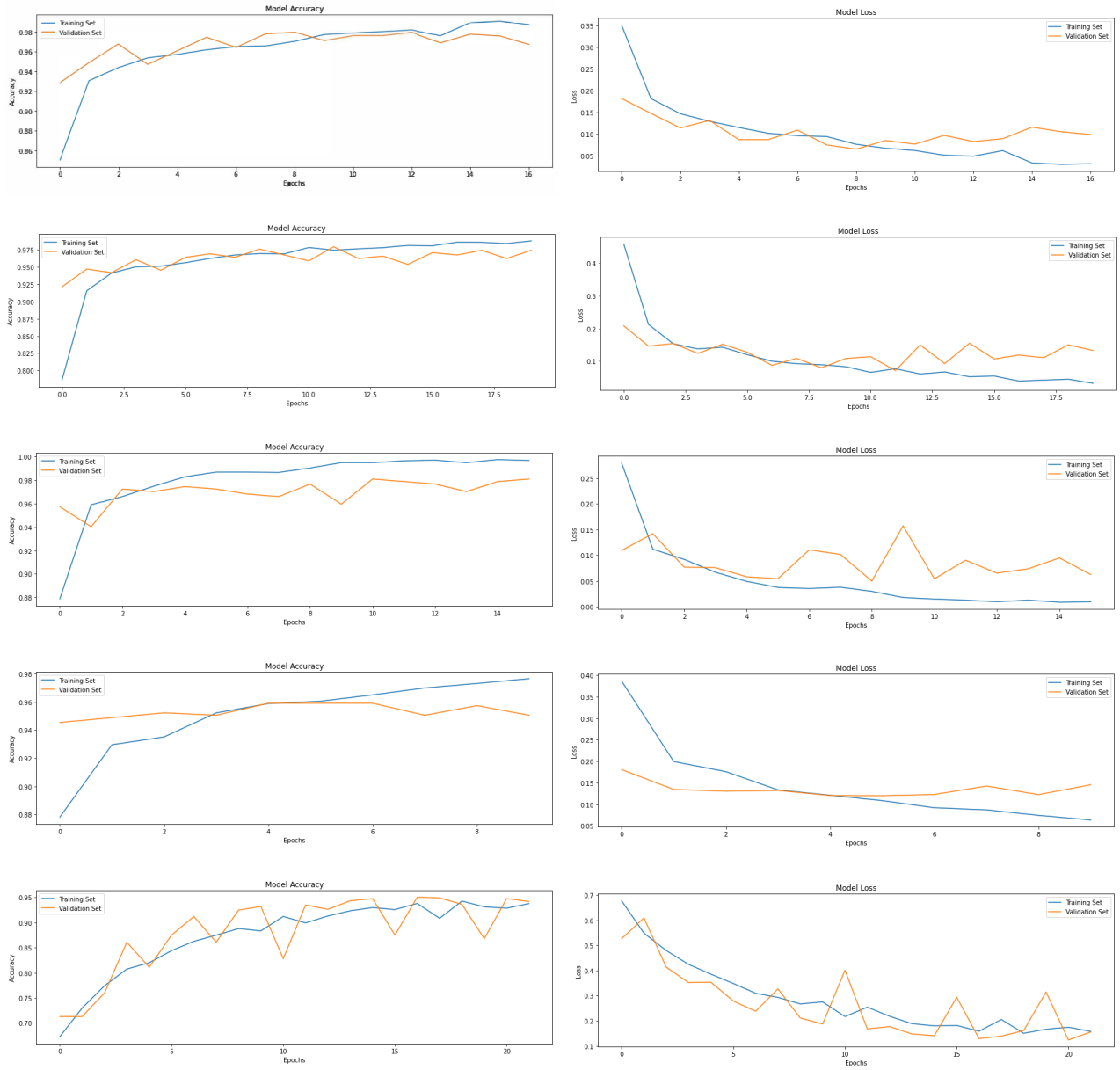
ACKNOWLEDGMENT

The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of two Titan X GPUs. The authors also thankfully acknowledge the grant of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga - IBIMA.

REFERENCES

[1] World Health Organization, "WHO Pneumonia," 2019. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying Medical



(a) (b)

Fig. 5. Training and validation (a) accuracy, and (b) loss graphs for CNN1, CNN2, VGG-16, Inception-v3, and ResNet50 models from the top row to the bottom row, respectively.

Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.

[3] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia Detection Using CNN based Feature Extraction," in *Proceedings of 2019 3rd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2019*. Institute of Electrical and Electronics Engineers Inc., 2019.

[4] R. Jain, P. Nagrath, G. Kataria, V. Sirish Kaushik, and D. Jude Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," *Measurement: Journal of the International Measurement Confederation*, vol. 165, 2020.

[5] R. Kundu, R. Das, Z. W. Geem, G. T. Han, and R. Sarkar, "Pneumonia detection in chest x-ray images using an ensemble of deep learning models," *PloS one*, vol. 16, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34492046/

[6] A. Manickam, J. Jiang, Y. Zhou, A. Sagar, R. Soundrapandiyan, and R. Dinesh Jackson Samuel, "Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures," *Measurement: Journal of the International Measurement Confederation*, vol. 184, p. 109953, 2021.

[7] M. F. Hashmi, S. Katiyar, A. W. Hashmi, and A. G. Keskar, "Pneumonia detection in chest x-ray images using compound scaled deep learning model," *Journal for Control, Measurement, Electronics, Computing and Communications*, vol. 62, pp. 397–406, 2021.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, 2015.

[9] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning,"

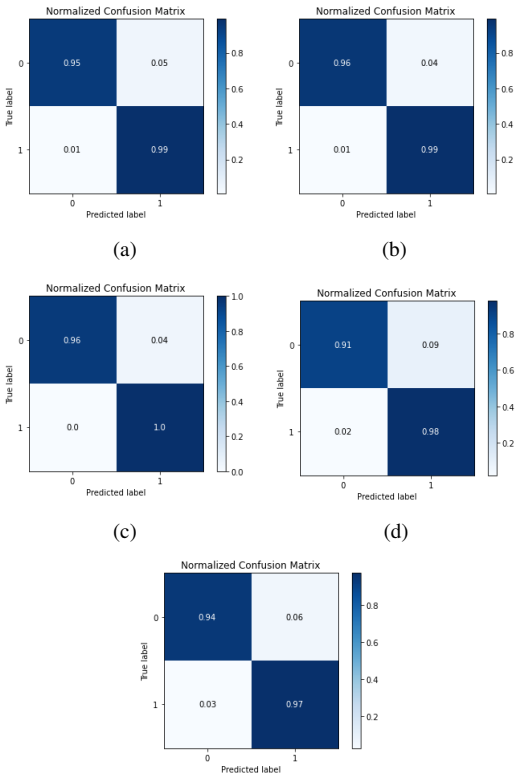


Fig. 6. Confusion matrices for each model: (a) CNN1, (b) CNN2, (c) VGG-16, (d) Inception-v3, and (d) ResNet50.

in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*. AAAI press, 2017, pp. 4278–4284.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem. IEEE Computer Society, dec 2016, pp. 770–778.

[11] C. W. et al. Kermany D, Goldbaum M, "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images," 2018.