Pneumonia Detection Using Convolutional Neural Networks (CNNs)

Conference Paper · April 2020

DOI: 10.1007/978-981-15-3369-3_36

4 authors, including:

Anand Nayyar
Duy Tan University

Rachna Jain
Jain Hospital & Research centre Pvt Ltd

Pneumonia Detection Using Convolutional Neural Networks (CNNs)

V. Sirish Kaushik, Anand Nayyar, Gaurav Kataria and Rachna Jain

Abstract Pneumonia, an interstitial lung disease, is the leading cause of death in children under the age of five. It accounted for approximately 16% of the deaths of children under the age of five, killing around 880,000 children in 2016 according to a study conducted by UNICEF. Affected children were mostly less than two years old. Timely detection of pneumonia in children can help to fast-track the process of recovery. This paper presents convolutional neural network models to accurately detect pneumonic lungs from chest X-rays, which can be utilized in the real world by medical practitioners to treat pneumonia. Experimentation was conducted on Chest X-Ray Images (Pneumonia) dataset available on Kaggle. The first, second, third and fourth model consists of one, two, three and four convolutional layers, respectively. The first model achieves an accuracy of 89.74%, the second one reaches an accuracy of 85.26%, the third model achieves an accuracy of 92.31%, and lastly, the fourth model achieves an accuracy of 91.67%. Dropout regularization is employed in the second, third and fourth models to minimize overfitting in the fully connected layers. Furthermore, recall and F1 scores are calculated from the confusion matrix of each model for better evaluation.

Keywords Convolutional neural networks (CNNs) · Pneumonia detection · ReLU · Max-pooling · Forward and backward propagation

V. Sirish Kaushik (✉) · G. Kataria · R. Jain
Bharati Vidyapeeth's College of Engineering, New Delhi, Delhi, India
e-mail: shirishkaushik@gmail.com

G. Kataria
e-mail: gaurav.kataria2291999@gmail.com

R. Jain
e-mail: rachna.jain@bharatividyapeeth.edu

A. Nayyar
Graduate School, Duy Tan University, Da Nang, Vietnam
e-mail: anandnayyar@duytan.edu.vn

P. K. Singh et al. (eds.), *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, Lecture Notes in Networks and Systems 121, https://doi.org/10.1007/978-981-15-3369-3_36

1 Introduction

One of the major factors associated with pneumonia in children is indoor air pollution. Apart from this, under-nutrition, lack of safe water, sanitation and basic health facilities are also major factors. Pneumonia is an interstitial lung disease caused by bacteria, fungi or viruses. It accounted for approximately 16% of the 5.6 million under-five deaths, killing around 880,000 children in 2016 [1]. Affected victims were mostly less than two years old. Timely detection of pneumonia can help to prevent the deaths of children. This paper presents convolutional neural network models to accurately detect pneumonic lungs from chest X-rays, which can be utilized in the real world by medical practitioners to treat pneumonia [2]. These models have been trained to classify chest X-ray images into normal and pneumonia in a few seconds, hence serving the purpose of early detection of pneumonia. Although transfer learning models based on convolutional neural networks like AlexNet, ResNet50, InceptionV3, VGG16 and VGG19 are some of the most successful ImageNet dataset models with pre-trained weights, they were not trained on this dataset as the size of dataset taken for our research is not as extensive compared to ones which generally employ transfer learning [3]. Four classification models were built using CNN to detect pneumonia from chest X-ray images to help control this deadly infection in children and other age groups. Accuracy of the model is directly correlated with the size of the dataset, that is, the use of large datasets helps improve the accuracy of the model, but there is no direct correlation between the number of convolutional layers and the accuracy of the model.

To obtain the best results, a certain number of combinations of convolution layers, dense layers, dropouts and learning rates have to be trained by evaluating the models after each execution. Initially, simple models with one convolution layer were trained on the dataset, and thereafter, the complexities were increased to get the model that not only achieved desired accuracies but also outperformed other models in terms of recall and F1 scores. The objective of the paper is to develop CNN models from scratch which can classify and thus detect pneumonic patients from their chest X-rays with high validation accuracy, recall and F1 scores. Recall is often favored in medical imaging cases over other performance evaluating parameters, as it gives a measure of false negatives in the results. The number of false negatives in the result is very crucial in determining the real-world performance of models [4]. If a model achieves high accuracy but low recall values, it is termed as underperforming, inefficacious and even unsafe as higher false-negative values imply higher number of instances where the model is predicting a patient as normal, but in reality, the person is diseased. Hence, it would risk the patient's life. To prevent this, the focus would be only models with great recall values, decent accuracies and F1 scores [5].

The paper is organized into 5 sections: Sect. 1 introduces the subject of this research paper, addresses its importance and relevance, the purpose and motive to undertake this research work and the objective of the paper. Section 2 explores the work related to this field that has been accomplished till now. Section 3 explains the methodology of the paper, explaining the architecture of the models, flowchart and the dataset used to train and test the four models. Section 4 presents the results achieved by the various CNN models and compares the performance of each model using accuracy and loss graphs and confusion matrices. Section 5 provides a brief conclusion to the paper and delivers the best-suited model. Furthermore, the future scope of this research work has also been discussed. All the references which are cited in the paper have been listed in the end.

2 Related Work

Many researchers have tackled the problem of classifying images with high accuracy. Here are some citations related to our paper:

Rubin et al. [6] developed a CNN model to detect common thorax disease from frontal and lateral chest X-ray images. MIMIC-CXR dataset was used to perform large-scale automated recognition of these images. The dataset was split into training, testing and validation sets as 70%, 20% and 10%, respectively. Data augmentation and pixel normalization were used to improve overall performance. Their DualNet CNN model achieved an average AUC of 0.72 and 0.688 for PA and AP, respectively. A deep convolutional neural network to classify pulmonary tuberculosis was developed by Lakhani et al. [7]. Transfer learning models such as AlexNet and GoogleNet were also used to classify chest X-ray images. The dataset was split into training, testing and validation sets as 68%, 14.9% and 17.1%, respectively. Data augmentation and pre-processing techniques were employed to get the best performing model achieving an AUC of 0.99. Precision and recall of the model were 100 and 97.3%. An AG-CNN model was developed by Guan et al. [8] to recognize thorax disease. ChestX-ray14 dataset was used to detect thorax disease from chest X-ray images. Global and local branch attention-guided CNN was used for classification purposes. Their model was better than other models mentioned in their research paper, achieving an AUC of 0.868. A deep convolutional neural network model was developed by Rajpurkar et al. [9] to classify chest X-ray images into pneumonia and other 14 diseases. ChestX-ray14 dataset was used for training the model. They compared their ChXNet model (121 layered model) with practicing academic radiologists. Their ChXNet model achieved an F1 score (95% CI) of 0.435 outperforming radiologists which achieved an F1 score (95% CI) of 0.387.

A deep convolutional neural network model having five convolutional layers some followed by max-pooling layers, having three fully connected layers was trained by Krizhevsky et al. [10]. This network contained 60 million different parameters. By employing dropout, this model achieved a top-five error percent of 17%. Simonyan et al. [11] developed a highly accurate model employing multiple small kernel-sized filters to achieve top-five test accuracy 92.7%. This model was trained on the ImageNet dataset and submitted to the ILSVRC 2014 competition. A convolution neural network for classification and segmentation of brain tumor MRIs was developed by Xu et al. [12]. Multiple techniques such as data augmentation, feature selection and pooling techniques were employed in this model. The validation accuracy for classification achieved by this model is 97.5%, and validation accuracy of segmentation is 84%, 256 × 256 pixels sized frontal chest radiographs which were fed to a deep convolution neural network to detect abnormalities. A convolutional neural network with five convolution layers employing leaky ReLU, average pooling and three fully connected layers was developed by Anthimopoulos et al. [13] to detect interstitial lung disease patterns in a dataset containing 14,696 images belonging to seven different classes. This model achieved a classification accuracy of 85.5%. He et al. [14] developed a residual neural network (RNN) to classify images present in the ImageNet dataset. RNN introduced the concept of shortcut connections to tackle the problem of vanishing gradients. This model which was submitted to ILSVRC 2015 attained state-of-the-art classification accuracy. A transfer learning model, extension of AlexNet using data augmentation techniques, was developed by Glozman et al. [15]. This model was trained on ADNI database. Two neural network models were presented by Hemanth et al. [16] which are MCPN and MKNN. These models classified MRIs with high accuracies and tackled high convergence time period for Artificial Neural Networks.

3 Methodology

CNN models have been created from scratch and trained on Chest X-Ray Images (Pneumonia) dataset on Kaggle. Keras neural network library with TensorFlow backend has been used to implement the models. Dataset consists of 5216 training images, 624 testing images and 16 validation images. Data augmentation has been applied to achieve better results from the dataset. The four models have been trained on the training dataset, each with different number of convolutional layers. Each model was trained for 20 epochs, with training and testing batch sizes of 32 and 1, respectively. The following sub-headings further explain the above stages in depth.

3.1 CNN Architecture

CNN models are feed-forward networks with convolutional layers, pooling layers, flattening layers and fully connected layers employing suitable activation functions.

Convolutional layer. It is the building block of the CNNs. Convolution operation is done in mathematics to merge two functions [17]. In the CNN models, the input image is first converted into matrix form. Convolution filter is applied to the input matrix which slides over it, performing element-wise multiplication and storing the sum. This creates a feature map. 3 × 3 filter is generally employed to create 2D feature maps when images are black and white. Convolutions are performed in 3D when the input image is represented as a 3D matrix where the RGB color represents the third dimension. Several feature detectors are operated with the input matrix to generate a layer of feature maps which thus forms the convolutional layer.

Activation functions. All four models presented in this paper use two different activation functions, namely ReLU activation function and softmax activation function. The ReLU activation function stands for rectified linear function [18]. It is a nonlinear function that outputs zero when the input is negative and outputs one when the input is positive. The ReLU function is given by the following formula:

This type of activation function is broadly used in CNNs as it deals with the problem of vanishing gradients and is useful for increasing the nonlinearity of layers. ReLU activation function has many variants such as Noisy ReLUs, Leaky ReLUs and Parametric ReLUs. Advantages of ReLU over other activation functions are computational simplicity and representational sparsity. Softmax activation function is used in all four models presented in this paper. This broadly used activation function is employed in the last dense layer of all the four models [19]. This activation function normalizes inputs into a probability distribution. Categorical cross-entropy cost function is mostly used with this type of activation function.

Pooling layer. Convolutional layers are followed by pooling layers. The type of pooling layer used in all four models is max-pooling layers. The max-pooling layer having a dimension of 2×2 selects the maximum pixel intensity values from the window of the image currently covered by the kernel. Max-pooling is used to down sample images, hence reducing the dimensionality and complexity of the image [20]. Two other types of pooling layers can also be used which are general pooling and overlapping pooling. The models presented in this paper use max-pooling technique as it helps recognize salient features in the image.

Flattening layer and fully connected layers. After the input image passes through the convolutional layer and the pooling layer, it is fed into the flattening layer. This layer flattens out the input image into a column, further reducing its computational complexity. This is then fed into the fully connected layer/dense layer. The fully connected layer [21] has multiple layers, and every node in the first layer is connected to every node in the second layer. Each layer in the fully connected layer extracts features, and on this basis, the network makes a prediction [22, 23]. This process is known as forward propagation. After forward propagation, a cost function is calculated. It is a measure of performance of a neural network model. The cost function used in all four models is categorical cross-entropy. After the cost function is calculated, back propagation takes place. This process is repeated until the network achieves optimum performance. Adam optimization algorithm has been used in all four models.

Reducing overfitting. The first model exhibits substantial overfitting; hence, dropout technique was employed in the later models [24]. Dropout technique helps to reduce overfitting and tackles the problem of vanishing gradients. Dropout technique encourages each neuron to form its own individual representation of the input data. This technique on a random basis cuts connections between neurons in successive layers during the training process [25]. Learning rate of models was also modified, to reduce overfitting. Data augmentation technique can also be employed to reduce overfitting.

Algorithm of CNN classifiers. The algorithms used in the convolutional neural network classifiers have been explained in Figs. 1 and 2. Figure 3 shows the flowchart of the overall schema of research. The number of epochs for all the classifier models presented in this paper was fixed at 20 after training and testing several CNN models over the course of research. Classifier models trained for more number of epochs have showed overfitting. Several optimizer functions were also trained and studied. Adam optimizer function was finalized to be used for all classifiers after it gave the best results. Initially, a simple classifier model with convolutional layer of image size set to 64 * 64, 32 feature maps and employing ReLU activation function was trained. Fully connected dense layer with 128 perceptrons was utilized. To improve the result, the second classifier model was trained with one more convolutional layer of 64 feature maps for better feature extraction. The number of perceptrons in dense layer was also doubled to 256, so that better learning could be achieved. The third model was trained for three convolutional layers with 128 feature maps in third convolutional layer for more detailed feature extraction. Dense layer was kept unchanged. Dropout layer was introduced at 0.3, and learning rate of optimizer was

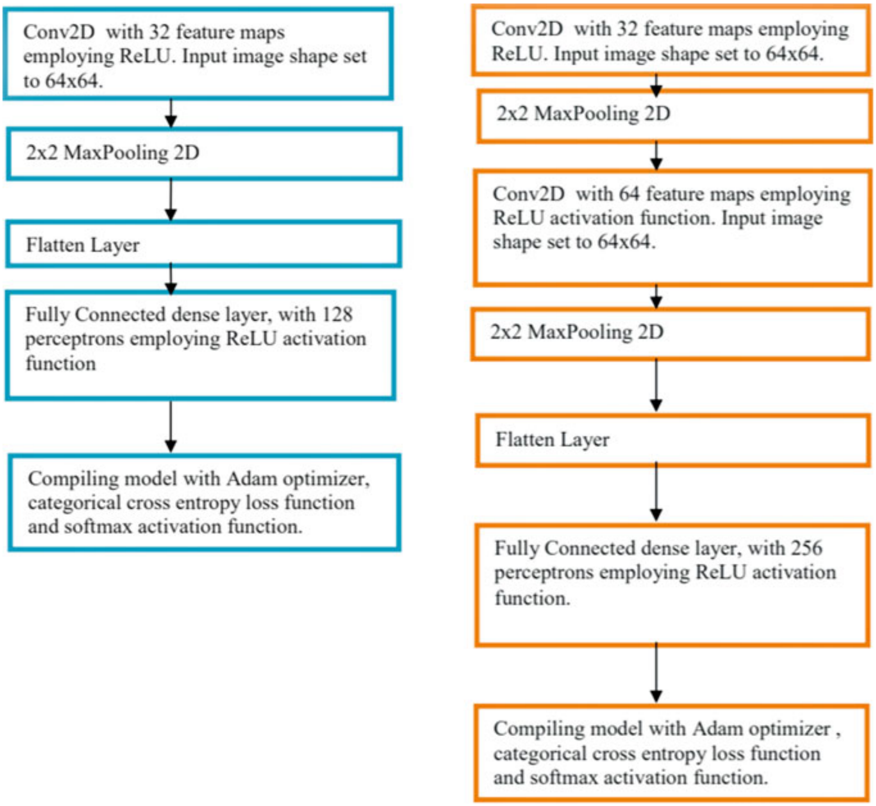

Fig. 1 Algorithms of CNN classifier model 1 (left) and model 2 (right)

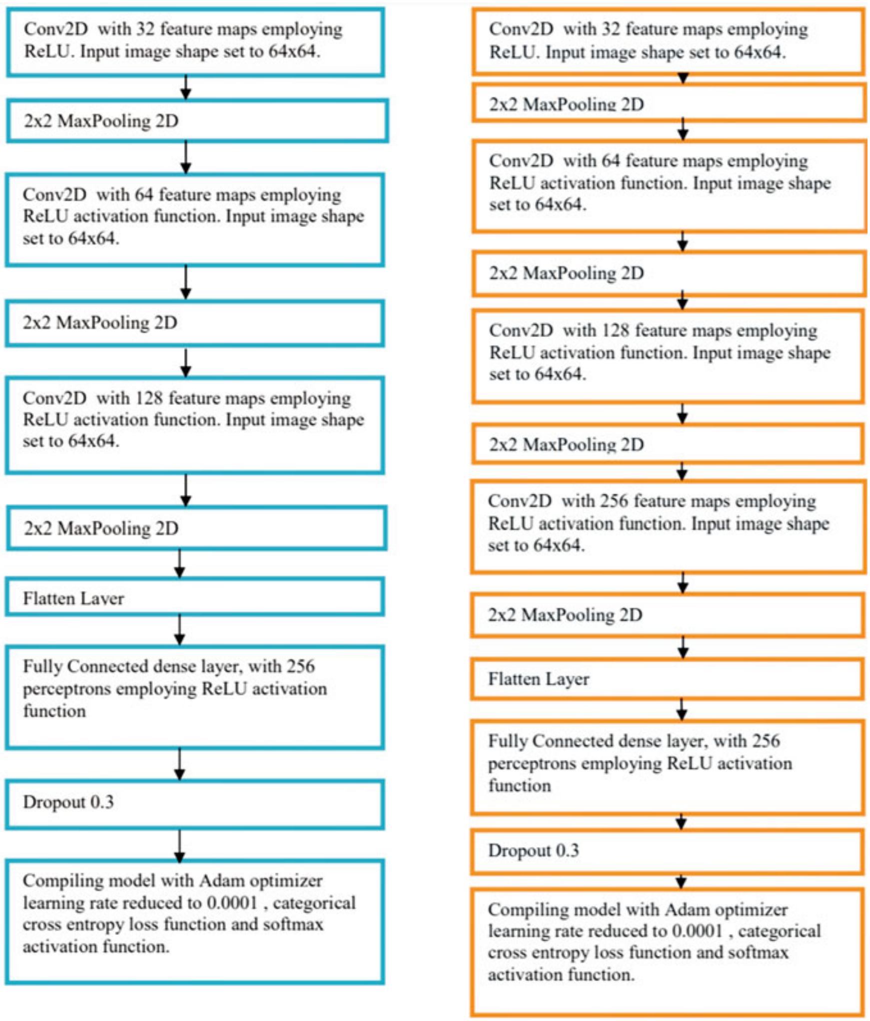

Fig. 2 Algorithms of CNN classifier model 3 (left) and model 4 (right)

lowered to 0.0001 to reduce the overfitting. The final fourth classifier model was trained for four convolutional layers with 256 feature maps in fourth convolutional layer. Dense layer, dropout layer and learning rate were kept same as third classifier model. The results have been summarized in the subsequent section of this paper.

Dataset. Chest X-Ray Images (Pneumonia) dataset of 1.16 GB size has been imported from Kaggle [26], with total of 5856 jpeg images split into Train, Test and Val folders each divided into category Pneumonia and Normal. Chest X-ray images (front and back) were selected from pediatric patients of one- to five-year olds from Guangzhou Women and Children's Medical Center, Guangzhou. Figure 4 provides

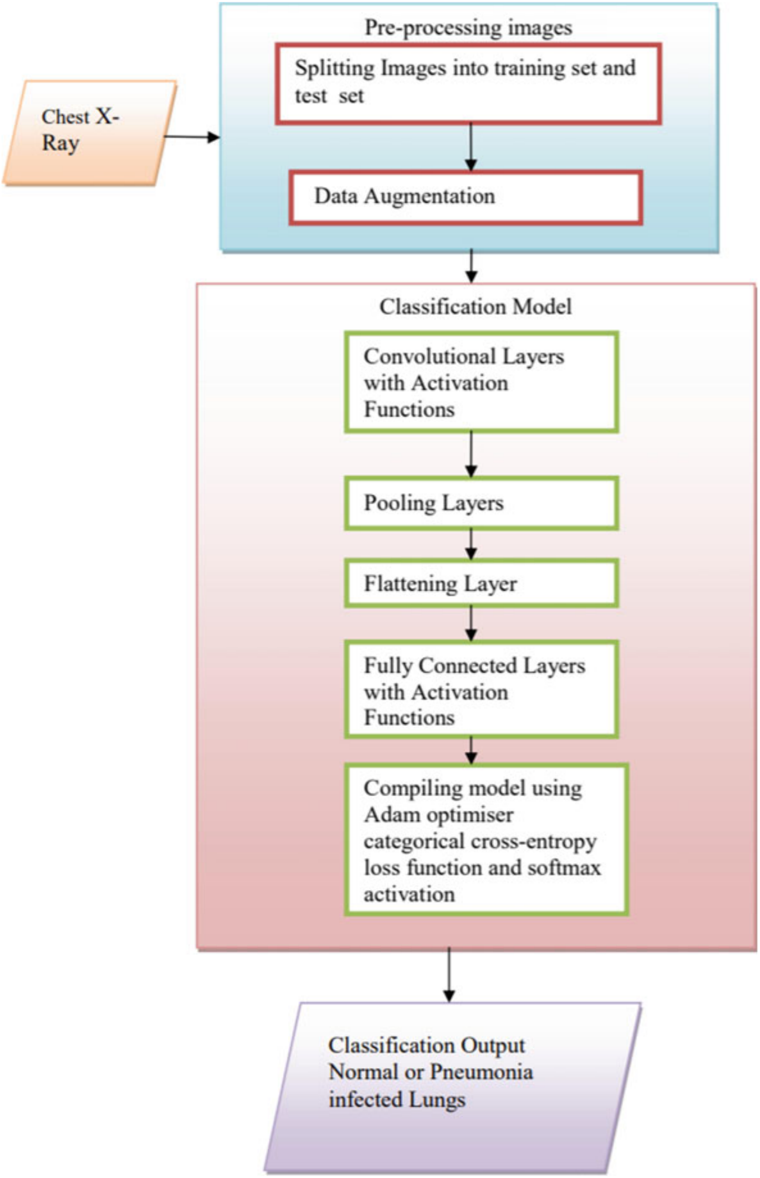

Fig. 3 Detailed schema of the experiment conducted

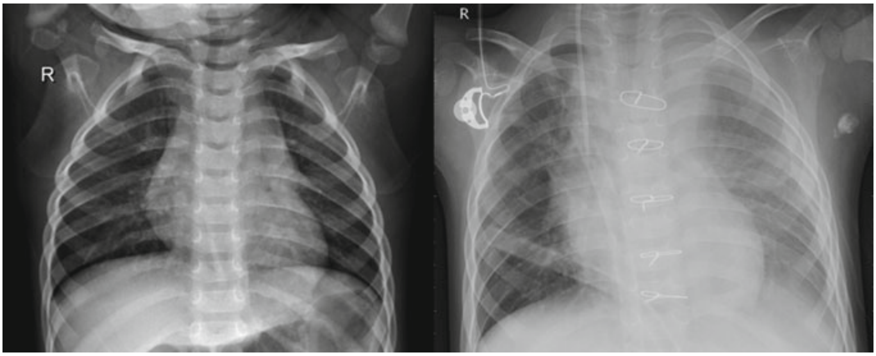

Fig. 4 Left image depicts normal lungs and right image depicts pneumonic lungs

the sample images from the dataset used during the research.

4 Experimental Results

To study the performance of each CNN classifier model, validation accuracy, recall and F1 score were evaluated as the performance measures [27, 28]. Accuracy and loss graphs were also studied. The confusion matrix was also computed for each model.

4.1 Comparison of Performance of Models

Figures 5 and 6 show the confusion matrices, accuracy graphs and loss graphs of all CNN classifier models. Table 1 and Figs. 5 and 6 show that classifier models 1 and 2 significantly underperformed compared to models 3 and 4. The accuracy graphs and loss graphs show overfitting. Accuracy, recall and F1 scores are also low. In addition to extra convolution layer, employing dropout and lowering the learning rate of optimizer in model 3 improved the performance considerably. It achieved the least overfitting along with highest accuracy and recall. Several attempts were made to better the performance by adding more convolutional layers and changing the parameters. Classifier model 4 with four convolutional layers showed good recall value and F1 score albeit with lower accuracy and higher overfitting compared to model 3. Thus, classifier model 3 performed the best among all CNN classifier models. In the following equations, tp = true positive, tn = true negative, fp = false positive and fn = false negative.

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \tag{1}$$

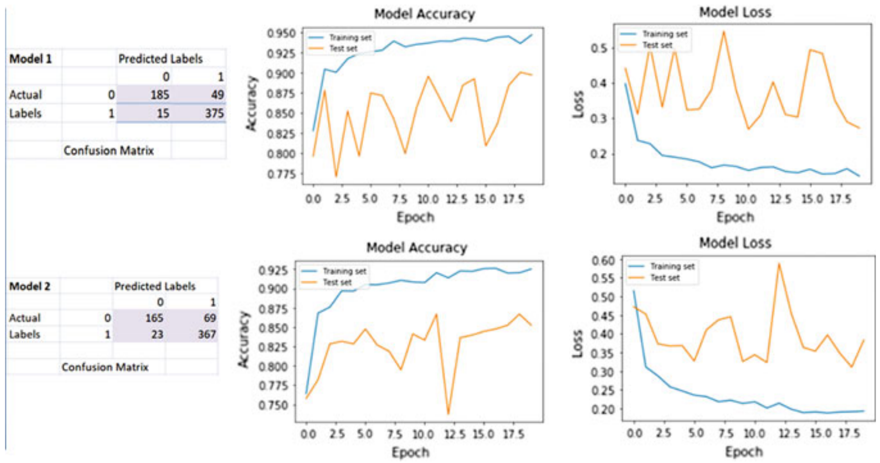

Fig. 5 Performance of classifier model 1 and model 2

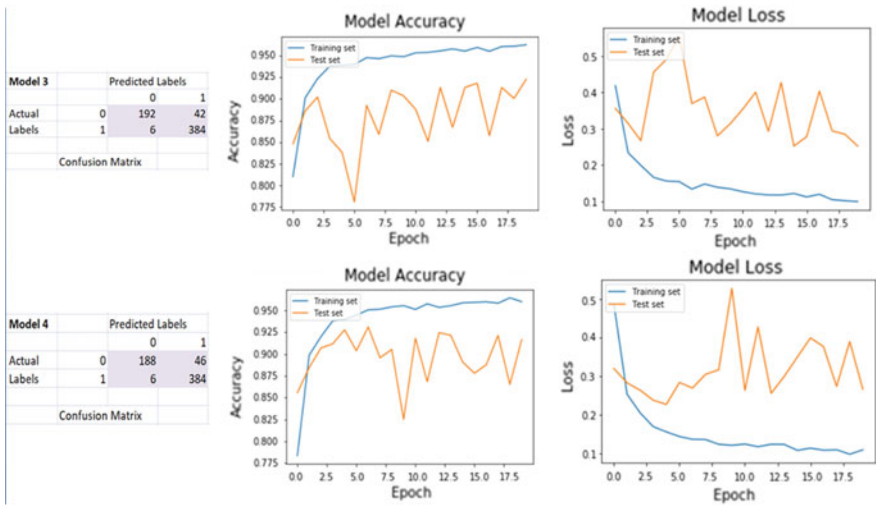

Fig. 6 Performance of classifier model 3 and model 4

$$\text{Precision} = \text{tp}/(\text{tp} + \text{fp}) \tag{2}$$

Table 1 Performance comparison of different CNN models

Classifier model	Validation accuracy (%)	Validation loss (%)	Recall (%)	F1 score (%)
Model 1 (one conv.layer)	89.74	27.31	96	92
Model 2 (two conv.layers)	85.26	38.36	94	89
Model 3 (three conv.layers)	92.31	25.23	98	94
Model 4 (4 conv.layers)	91.67	26.61	98	94

$$\text{Recall} = \text{tp}/(\text{tp} + \text{fn}) \tag{3}$$

$$\text{F1 Score} = 2(\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \tag{4}$$

5 Conclusion

The validation accuracy, recall and F1 score of CNN classifier model 3 with three convolutional layers are 92.31%, 98% and 94%, respectively, which are quite high compared to other models that were trained. CNN classifier model 4 with four convolutional layers also comes very close in performance with 91.67% validation accuracy, 98% recall and 94% F1 score. Both of these models have the same recall and F1 scores. The paper by Chakraborty [29] achieved the overall accuracy of 95.62% and recall of 95% trained on the same dataset. The paper by Liang [30] achieved recall of 96.7% on the same dataset. The models presented by us at best could achieve 92.31% accuracy which is lower, but 98% recall has been achieved. High recall values will ensure that the number of false-negative instances is lower, hence lowers the risk to the patient's life. Thus, it is concluded that CNN classifier model 3 and model 4 can, therefore, be effectively used by medical officers for diagnostic purposes for early detection of pneumonia in children as well as adults. A large number of X-ray images can be processed very quickly to provide highly precise diagnostic results, thus helping healthcare systems provide efficient patient care services and reduce mortality rates. These convolutional neural networks' models were successfully achieved by employing various methods of parameter tuning like adding dropout, changing learning rates, changing the batch size, number of epochs, adding more complex fully connected layers and changing various stochastic gradient optimizers [31].

In the future, it is hoped that transfer learning models would be trained on this dataset that would outperform these CNN models. It is intended that larger datasets will also be trained using the models presented in the paper. It is also expected that neural network models based on GAN [32], generative adversarial networks, would also be trained and compared with the existing models.

References

1. https://data.unicef.org/topic/child-health/pneumonia/. Accessed on 15 July 2019
2. Jaiswal, A.K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., Rodrigues, J.J.: Identifying pneumonia in chest x-rays: a deep learning approach. Measurement **145**, 511–518 (2019)
3. Kim, D.H., MacKinnon, T.: Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin. Radiol. **73**(5), 439–445 (2018)
4. Bernal, J., Kushibar, K., Asfaw, D.S., Valverde, S., Oliver, A., Martí, R., Lladó, X.: Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. Artif. Intell. Med. **95**, 64–81 (2019)
5. Arthur, F., Hossein, K.R.: Deep learning in medical image analysis: a third eye for doctors. J. Stomatology Oral Maxillofac. Surg.
6. Rubin, J., Sanghavi, D., Zhao, C., Lee, K., Qadir, A., Xu-Wilson, M.: Large Scale Automated Reading of Frontal and Lateral Chest X-Rays Using Dual Convolutional Neural Networks (2018). arXiv preprint arXiv:1804.07839
7. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology **284**(2), 574–582 (2017)
8. Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., Yang, Y.: Diagnose Like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification (2018). *arXiv preprint* arXiv:1801.09927
9. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P.: Chexnet: Radiologist-Level Pneumonia Detection on Chest X-rays with Deep Learning (2017). arXiv preprint arXiv:1711.05225
10. Krizhevsky, A., Sutskever, I., Hinton, G.E:. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
11. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2014). arXiv preprint arXiv:1409.1556
12. Xu, Y., Jia, Z., Ai, Y., Zhang, F., Lai, M., Eric, I., Chang, C.: Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation. In: 2015 international conference on acoustics, speech and signal processing (ICASSP), pp. 947–951 (2015)
13. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans. Med. Imaging **35**(5), 1207–1216 (2016)
14. 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, pp. 770–778 (2016)
15. Glozman, T., Liba, O.: Hidden Cues: Deep Learning for Alzheimer's Disease Classification CS331B project final report (2016)
16. Hemanth, D.J., Vijila, C.K.S., Selvakumar, A.I., Anitha, J.: Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification. Neurocomputing **130**, 98–107 (2014)

17. Bi, X., Li, S., Xiao, B., Li, Y., Wang, G., Ma, X.: Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. Neurocomputing (2019)
18. Eckle, K., Schmidt-Hieber, J.: A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw. **110**, 232–242 (2019)
19. Ren, S., Jain, D.K., Guo, K., Xu, T., Chi, T.: Towards efficient medical lesion image super-resolution based on deep residual networks. Sig. Process. Image Commun. **75**, 1–10 (2019)
20. Zheng, Y., Iwana, B.K., Uchida, S.: Mining the displacement of max-pooling for text recognition. Pattern Recogn. **93**, 558–569 (2019)
21. Bhumika, P.S.S.S., Nayyar, P.A.: A review paper on algorithms used for text classification. Int. J. Appl. Innov. Eng. Manage. **3**(2), 90–99 (2013)
22. Kumar, A., Sangwan, S.R., Arora, A., Nayyar, A., Abdel-Basset, M.: Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. IEEE Access **7**, 23319–23328 (2019)
23. Saeed, F., Paul, A., Karthigaikumar, P., Nayyar, A.: Convolutional neural network based early fire detection. In: Multimedia Tools and Applications, pp. 1–17 (2019)
24. Kukkar, A., Mohana, R., Nayyar, A., Kim, J., Kang, B.G., Chilamkurti, N.: A novel deep-learning-based bug severity classification technique using convolutional neural networks and random forest with boosting. Sensors **19**(13), 2964 (2019)
25. Khan, S.H., Hayat, M., Porikli, F.: Regularization of deep neural networks with spectral dropout. Neural Netw. **110**, 82–90 (2019)
26. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. Accessed on 15 July 2019
27. ALzubi, J.A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A., Thaventhiran, C.: Boosted neural network ensemble classification for lung cancer disease diagnosis. Appl. Soft Comput. **80**, 579–591 (2019)
28. Vora, J., Tanwar, S., Polkowski, Z., Tyagi, S., Singh, P.K., Singh, Y.: Machine learning-based software effort estimation: an analysis. In: 11th International Conference on Electronics, computers and Artificial Intelligence (ECAI 2019), pp. 1–6, University of Pitesti, Pitesti, Romania, 27–29 June 2019
29. Chakraborty, S., Aich, S., Sim, J.S., Kim, H.C.: Detection of pneumonia from chest x-rays using a convolutional neural network architecture. In: International Conference on Future Information & Communication Engineering, vol. 11, no. 1, pp. 98–102 (2019)
30. Liang, G., Zheng, L.: A transfer learning method with deep residual network for pediatric pneumonia diagnosis. In: Computer Methods and Programs in Biomedicine (2019)
31. Du, S. S., Zhai, X., Poczos, B., Singh, A.: Gradient Descent Provably Optimizes Over-Parameterized Neural Networks (2018). arXiv preprint arXiv:1810.02054
32. Radford, A., Metz, L., Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks (2015). arXiv preprint arXiv:1511.06434