Nat. Volatiles & Essent. Oils, 2021; 8(5): 2161 - 2169

Performance analysis of Pneumonia using Convolutional Neural Networks

Ranjith M S[1],Dr.Aruna Devi K[2], Dr. Lakshminarayana M[3], Dr. G. Somasekhar[4], APPASAMI G[5], Dr.R.SUNDAR RAJAN[6]

[1]Associate engineer from Ignitarium Technology solutions,

[2]Dr. Aruna Devi K, Assistant Professor in Computer Science (PG), Kristu Jayanti College (Autonomous), Bengaluru, Karnataka, India – 560077

[3]Assistant Professor at Dept of Electronics and Communication Engineering, SJBIT,

[4]Associate Professor of Computer Science and Engineering Department at CMR Technical Campus, (Autonomous), Hyderabad, Telangana, India – 501401

[5]Formerly worked as Assistant professor in the Dept of CS, Central University of Tamilnadu,

[6]Associate professor, IT at Kalasalingam Academy of Research and Education, Tamilnadu.

ranjithms523@gmail.com, k.arunadeviselvi@gmail.com, mlnphd101@gmail.com, giddalurisomasekhar@gmail.com, appa.govi@gmail.com, sundaradhira@gamil.com,

Abstract

Pneumonia is a bacterial or viral infection disease occurs in human lungs. Pneumonia is a critical issue and affecting the human lungs, most of the time infections caused by bacteria called Streptococcus pneumonia. Early stage of pneumonia diagnosis is very much essential for effective and efficient treatment process to avoid future problems. According to World Health Organization (WHO), in India death rate is one among three due to pneumonia infection. Along with primary check-up and test, chest X-ray image is also needed by a doctor for the disease evaluation. Chest X-ray is one of the best tools in examining pneumonia, because examining is quite cheaper, and this service will be available in most of the hospitals and clinic. However, several traditional and manual diagnosis methods have been used for the pneumonia diagnosis. The traditional and manual method of diagnosis requires human time and intervention. This work proposes the development of the automated pneumonia diagnostic method using different Convolutional Neural Networks (CNN), which helps doctors for treatment and planning diagnosis. CNN architecture is developed for automatic pneumonia detection, avoid manual pre-processing, classification and segmentation, instead of it use chest radiography pixel information. CNN architectures are a multi-layer architecture with and without data augmentation was carried out with casual and pneumonia plain chest radiography for network training. CNN network plays an important role in diagnosing the disease. The work includes four CNN networks, namely DenseNet 121, VGG 16, Inception V3 and ResNet 50 with and without data argumentation was carried out for pneumonia diagnose. By the end of 10^{th} epoch the accuracy reached 90% in all the cases but in DenseNet it reached 98.5% (training set) and in Inception it reached 98.6% on validation set indicating that the Inception network has low variance compared to others.

Keywords

Convolutional Neural Networks (CNN), Pneumonia, Inception, Vgg16, ResNet50, X-ray

1. Introduction

Pneumonia can be described as a bacterial or viral infection disease occurs in human lungs [1]. However, early diagnosis and detection of pneumonia may save the life with proper medication and treatment plan [2]. Along with primary check-up and test, chest X-ray image is also needed by a doctor for the disease evaluation. Chest X- ray is the best tool in examining pneumonia, because it's cheaper and available in most of the hospital and clinic [3]. Even though chest X- ray is available for pneumonia diagnosis, it requires expert doctor's intervention for the disease analysis and also disease identification is challenging in case of low resolution chest X-ray image. To reduce misclassification rate in low resolution chest X-ray image and also expert's intervention for analysis need computer based automatic system [4-6]. Diagnosis of pneumonia from chest X-ray makes it more accurate with the help of advancement in the field of machine learning. Machine learning based algorithms showed more improvement in the classification kind of problems, especially Convolutional Neural Networks (CNN) [7]. The traditional diagnosis process requires more human time and intervention. The manual methods for per-processing was used, instead of it use chest radiography pixel information [7, 8]. Different CNN architecture shows their vital role in the field pre-processing, handcrafted segmentation and classification, localization and object detection in computer vision application [9-11]. Apart from the computer vision problems, CNN pulled attention in the field of medical. Many medical classification and object detection problems using CNN as prime part [12], an automatic osteoarthritis severity predication [13], early detection of dementia disease before it gets worsen and Myocardial infarction detection [14, 15] etc. The complete deep learning application in analysing medical image is discussed here [16, 17]. DenseNet is a 121 layer CNN [18]. Only few researchers worked on the pneumonia detection using deep leaning. DenseNet 121 layer architecture with method of transfer leaning was used and obtains 0.60 % area under the curve (AUC) value [19] and other network called as CheXNet [20]. CheXNet network trained with ten thousand images of chest X-ray along with 14 different diseases. Other network used for pneumonia detection is based on the Xception model [21] and Vgg16 based model [22]. The work includes the performance analysis of different CNN networks like ResNet 50 [7], DenseNet 121 [8], Inception V3 and VGG 16, for the pneumonia classification purpose.

2. Related work

The traditional manual methods use handcrafted features for medical image classification. Advancements of machine learning algorithms and computer vision many of the medical image analysis applications using computer vision and machine learning for object detection as well as for the detection. Deployment of machine learning, in breast cancer field, arthritis, lungs nodules etc., plays more vital role [23,24,25]. CNN architectures having capable of extracting feature which are real needed for further processing [26]. Few CNN large scale online datasets like MNIST, CIFAR-10 and ImageNet are available. Keras produces few CNN's that been pre-trained with dataset of ImageNet are VGG19 [6] , VGG16, Inception V3 ResNet50, and Xception [7] and other available CNN networks are AlexNet [5], ResNet [8] and DenseNet [9]. Moreover CNN used for image classification application need high-rich extracted feature. CNN extract features stage by stage at convolution layers [10]. Along with primary check and test, chest X-ray is also needed by a doctor for the evaluation. Chest X-ray is the best tools in the examining septic region in the lungs. The work includes the evaluation of different performance matrices with different pre -trained CNN architecture which are helpful in the classification of the normal and pneumonia chest X-ray. The main contributions of this work include as follows: (i) accuracy based comparative study of different pre-trained CNN models (ii) variation of categorical cross entropy loss with different networks, (c) evaluation of different performance measures like precision, recall, specificity and F1 score.

The paper is structure as follows: In the section two, study about some related research were done in same field. In the section three, details about datasets used in the training and testing purpose. In section four about different CNN architecture used in the study and its study. In section five and six, explained about obtained results and its discussion.

Material and Methodology

A. Data

The work includes 5216 X-ray images of frontal chest which was provided by Kermany et al [23]. Complete 5216 dataset is split into 4695 images (90% of the total dataset) for training and 521 images (10% of the total dataset) for validation. Image size is restricted to 256 * 256. Fig 1 shows few x-ray samples of normal and pneumonia dataset.

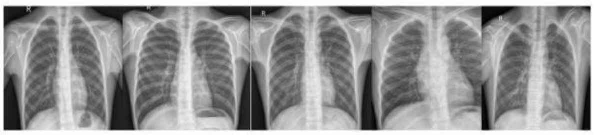

Fig 1: Normal chest X-ray images

Fig 2: Pneumonia chest X-ray images

The pixel values are normalized to have smooth loss surface. This model was trained using the Adam optimizer with 32 as batch size. Binary cross entropy is used for loss of function in training the model.

B. Data augmentation

Deep learning needs a large amount of dataset to produce correct production on testing phase. More number of datasets required expecting correct predication on test image but for some cases dataset may not be available for certain problems, especially for medical disease diagnosis. Along with increase in dataset can make the algorithms robust to different kind of inputs. Different methods are being used to increase available dataset; one among of them is data augmentation. Data augmentation method increase accuracy meanwhile it avoid overfitting of images. All algorithms have been trained with both raw dataset as well as augmented dataset. Initially images are resized to 256*256 sizes, normalized the pixel values and fed into the network. Whereas in augmented case we have performed few augmentations like height shift, width shift and along with the shear transformations (range of =0.01) along with normalization of pixels and resizing of images to 256 × 256.

Table 1: Describe about data augmentation operations

Data augmentation operations	Values
Width shift range	0.1
Height shift range	0.1
Shear range	0.01
Zoom range	0.15
Horizontal flip	True
Fill mode	Reflect
Rotation range	5

C. CNN architectures

The work include well known CNN networks DenseNet 121, ResNet 50, VGG 16 and Inception V3. Tensor Flow is used to create the different CNN networks. All algorithms have been trained with both raw dataset as well as augmented dataset. Initially images are resized to 256*256 sizes, normalized the pixel values and fed into the network. Whereas in augmented case we have performed few augmentations like width shift, height shift and shear transformations (range=0.01) along with normalization of pixels and resizing of images to 256 *256. The model is trained using Tesla-K80 GPU.

Binary cross entropy shall be given by,

$$CE = -\sum_{i=1}^{C'=2} t_i log\left(f(s_i)\right) = -t_1 log\left(f(s_1)\right) - (1 - t_1) log\left(1 - f(s_1)\right)$$

Where $f(s_i)$ represents sigmoid function given by

$$f(s_i) = \frac{1}{1 + e^{-s_i}}$$

3. Results and Discussions

Different CNN architectures like DenseNet 121, ResNet 50,VGG 16 and Inception V3 have been trained with normal and pneumonia chest x-ray images. Every model has trained for about 10 epochs. Training phase performed with both raw dataset as well as augmented dataset. The images are resized 256*256 sizes. The normalized the pixel values are fed into the network. The augmentations process include width shift, height shift and shear transformations. Table 2: describe about data augmentation operations to increase dataset and robust of the CNN networks.

Evaluation metrics

The four different CNN networks, namely DenseNet 121, VGG 16,ResNet 50 and Inception V3 with and without data argumentation evaluation is done by using different metrics like accuracy, loss, specificity, sensitivity, precision, recall and f1 score. Table 3 describes about different performance analysis formulas.

Table 3: describes about performance analysis formulas.

Performance parameters	Formulas
Accuracy	(TP+TN)/(TP+FN+TN+FP)
Loss	(FP+FN)/ (TP+FN+TN+FP)
Sensitivity	TP/(TP+FN)
Specificity	TN/(TN+FP)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F1 score	2 * (Precision * Recall) / (Precision + Recall)

Where TN,TP, FN, FP represents the total number of true negative, true positive, false negative, false positive accordingly.

Table 2: Describe about performance matrices values for different CNN architecture

Architecture	Inception V3	Inception V3	ResNet 50	ResNet 50	VGG 16	DenseNet 121	DenseNet 121
Augmentation	Yes	No	Yes	No	Yes	No	Yes
loss	0.0496	0.061	0.088	0.0284	0.1612	0.0406	0.0702
accuracy	0.9836	0.9796	0.9678	0.9883	0.9412	0.9851	0.9721
f1	0.9888	0.986	0.9782	0.9921	0.9594	0.9896	0.9809
precision	0.9895	0.9865	0.9781	0.9921	0.9657	0.9898	0.9819
recall	0.9887	0.9865	0.9795	0.9925	0.9562	0.9899	0.9812
specificity	0.9752	0.9704	0.9438	0.9828	0.9209	0.9798	0.9607
val_loss	0.0574	0.0515	0.1838	0.268	0.218	0.1565	0.1345
val_accuracy	0.9712	0.9866	0.9309	0.9386	0.8983	0.9482	0.952
val_f1	0.972	0.9913	0.952	0.9553	0.9105	0.9671	0.9638
val_precision	0.9926	0.9904	0.9304	0.9974	0.8731	0.946	0.9377
val_recall	0.9562	0.9926	0.9774	0.9179	0.9594	0.9908	0.993
val_specificity	0.9784	0.9719	0.7623	0.9935	0.6823	0.8321	0.817

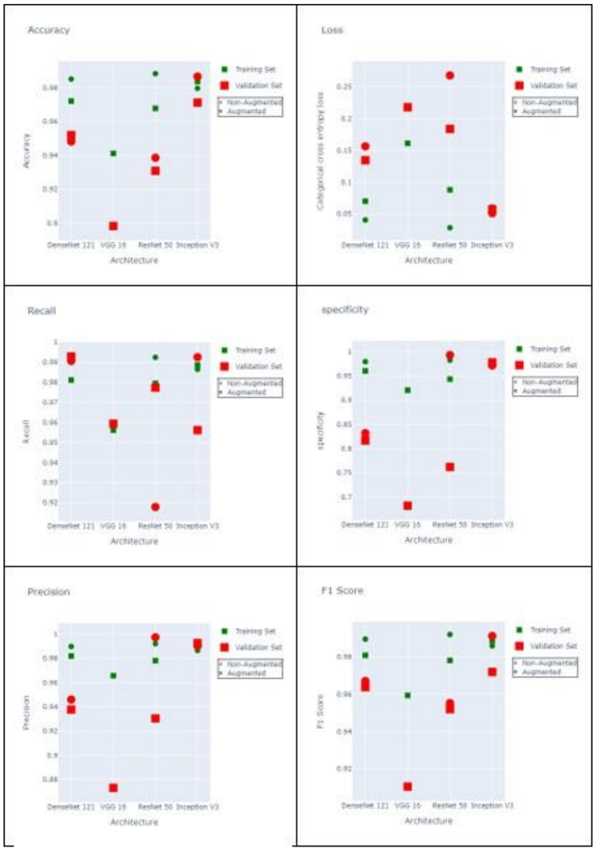

Fig 3: Performance matrices values variation with different CNN architecture.

Fig 3 shows performance matrices values variation with different CNN architecture. Accuracy plot specify that inception V3 performs well on training and validation sets of both augmented and non-augmented data compared with other models. Similar conclusion can be drawn that inception V3 performs better comparatively on other plot of F1 score, precision, recall, specificity. At the end of 10th epoch the accuracy reached 90% in all the cases but in DenseNet it reached 98.5% (training set) and in Inception it reached 98.6% on validation set indicating that the Inception network has low variance compared to others. Same inference can be drawn by looking at F1 scores, precision, recall and specificity. Also, it can be referred as there is wide distance between validation and training set in F1 score, accuracy, precision, specificity and recall of VGG 16 and ResNet 50 (VGG 16 being more) indicating that they are not as generalizable as Inception and DenseNet. From fig3 loss plot, shows that variation of categorical cross entropy loss with different networks. The average loss for Inception V3 over other network is comparatively low. Because of the low loss performance of Inception V3 making it the best among other networks for this task. From recall plot, DenseNet recall value is 0.993 on validation set. DenseNet recall value indicating slightly better robustness towards prediction of disease compared to other networks. From specificity plot that, specificity of DenseNet on validation set is around 0.83 whereas Inception has 0.9784 indicating that DenseNet is not as good as Inception when seen in terms of predicting non-disease cases. It is observed that the size of inception V3 which performs well on dataset is higher compared to DenseNet. In Resnet, Inception a layer only receives outputs from the previous second or third layer but it’s not true in case of DenseNet hence more memory is needed than the shown in case of DenseNet for prediction described in Fig 4.

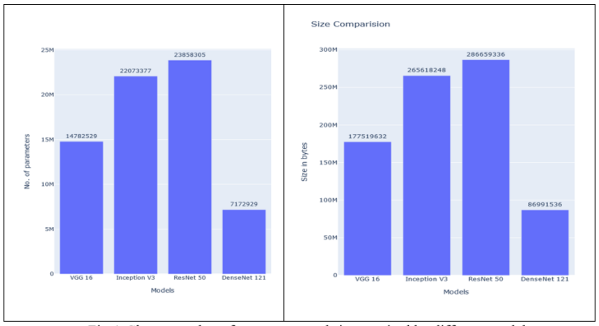

Fig 4: Shows number of parameters and size required by different models.

4. Conclusion

The work include the performance analysis of four different CNN network's for pneumonia disease diagnosis. The four different CNN networks, namely DenseNet121, VGG16, ResNet 50 and InceptionV3 with and without data argumentation carried out for pneumonia diagnose. At the end of 10^{th} epoch the accuracy reached 90% in all the cases but in DenseNet it reached 98.5% (training set) and in Inception it reached 98.6% on validation set indicating that the Inception network has low variance compared to others. Same inference can be drawn by looking at F1 scores, precision, recall and specificity. Loss plot shows that variation of categorical cross entropy loss with different networks. The average loss for Inception V3 over other network is comparatively low. Because of the low loss performance of Inception V3 making it the best among other networks for this task. From recall plot, DenseNet recall value is 0.993 on validation set. DenseNet recall value indicating slightly better robustness towards prediction of disease compared to other networks. From specificity plot that, specificity of DenseNet on validation set is around 0.83 whereas Inception has 0.9784 indicating that DenseNet is not as good as Inception when seen in terms of predicting non-disease cases. It is observed that the size of inception V3 which performs well on dataset is higher compared to DenseNet. Though it is observed that the models with augmentation have slightly lesser accuracy compared with the ones without augmentation, but they are robust to the variations in input images when deployed.

References

[1] CDC. (2019). Pneumonia Statistics. Available:
https://www.cdc.gov/pneumonia/prevention.html

[2] M. Aydogdu, E. Ozyilmaz, H. Aksoy, G. Gursel, and N. Ekim, "Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores," Tuberk Toraks, vol. 58, no. 1, pp. 25-34, 2010.

[3] W. H. Organization, "Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children," 2001.

[4] M. I. Neuman et al., "Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children," Journal of hospital medicine, vol. 7, no. 4, pp. 294-298, 2012.

[5] H. D. Davies, E. E.-l. Wang, D. Manson, P. Babyn, and B. Shuckett,"Reliability of the chest radiograph in the diagnosis of lower respiratory infections in young children," The Pediatric infectious disease journal,vol. 15, no. 7, pp. 600-604, 1996.

[6] R. Hopstaken, T. Witbraad, J. Van Engelshoven, and G. Dinant, "Interobserver variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections," Clinical radiology, vol. 59, no. 8, pp. 743-752, 2004.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.
[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, p. 436, 2015.
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015, pp. 234-241: Springer.
[11] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
[12] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," PeerJ, vol. 7, p. e6201, 2019.
[13] Antony, J., McGuinness, K., O'Connor, N.E. and Moran, K., 2016, December. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 1195-1200). IEEE.
[14] Arifoglu, D. and Bouchachia, A., 2019. Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. Artificial intelligence in medicine, 94, pp.88-95.
[15] Baloglu, U.B., Talo, M., Yildirim, O., San Tan, R. and Acharya, U.R., 2019. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recognition Letters, 122, pp.23-30.
[16] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," IEEE Access, vol. 6, pp. 9375-9389, 2018.
[17] G. Litjens et al., "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60-88, 2017.
[18] B. Antin, J. Kravitz, and E. Martayan, "Detecting Pneumonia in Chest X-Rays with Supervised Learning."
[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700- 4708.
[20] P. Rajpurkar et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," arXiv preprint, p. 1610.02357, 2017.
[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
[23] Dev Kumar Das, Madhumala Ghosh, Mallika Pal, Asok K Maiti, and Chandan Chakraborty. 2013. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron 45 (2013), 97106.
[24] MahdiehPoostchi,KamolratSilamut,RichardMaude,StefanJaeger,and George Thoma. 2018. Image

analysis and machine learning for detecting malaria. Translational Research (2018).
[25] Nicholas E Ross, Charles J Pritchard, David M Rubin, and Adriano G Duse. 2006. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Medical and Biological Engineering and Computing 44, 5 (2006), 427436.
[26] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 806813.
[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems. 10971105.
[28] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
[29] F Chollet. 2016. Xception: deep learning with separable convolutions. arXiv Prepr. arXiv1610 2357 (2016).
[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and