Journal of Engineering Science and Technology
Vol. 16, No. 1 (2021) 861 - 876

PNEUMONIA DETECTION USING CNN THROUGH CHEST X-RAY

HARSHVARDHAN GM, MAHENDRA KUMAR GOURISARIA,
SIDDHARTH SWARUP RAUTARAY*, MANJUSHA PANDEY

School of Computer Engineering, KIIT Deemed to be University,
Bhubaneswar, Odisha, India
*Corresponding Author: siddharthfcs@kiit.ac.in

Abstract

In general, pneumonia affects children under 5 years and adults over 65 years of age which targets the lungs and fills the alveoli (air sacs) with liquid. In this paper, we employ convolutional neural networks (CNNs) of varying configurations on a machine learning based binary classification task with a given dataset of chest X-rays that depicts affected and unaffected cases of pneumonia. This paper primarily focuses on putting forth the performances of different simple CNN architectures and selecting the best architecture based on optimum corresponding minimum loss and maximum accuracy which can serve as a viable tool for physicians and the medicine community to correctly identify and diagnose viral, bacterial, fungal-caused and community acquired pneumonia given only the chest X-ray of the patient.

Keywords: Artificial neural network (ANN), Biomedical imaging, Deep learning, Machine learning, Neural networks.

1. Introduction

Pneumonia is a widely occurring severe form of acute respiratory disease that is caused due to infectious agents with high attack rates among individuals that belong to age groups of either of the extreme ends of the lifespan of an average human being [1]. These infectious agents may be viral, fungal or bacterial and the lungs react to the infiltration of these foreign microorganisms by an inflammatory response which causes the bronchioles and alveoli to get filled with liquid [2]. This in turn causes difficulty in respiration for the affected individual. In the case of children under 5 years of age, pneumonia is responsible for over 15% deaths recorded globally, with 920,000 deaths in 2015 alone. In the same year, the US witnessed over 500,000 cases of emergency admissions in hospitals due to pneumonia [3] with 50,000 casualties reported [4] which puts the disease among the top 10 causes of deaths in the US.

There are three main types of pneumonia: a) community acquired, b) viral, and c) bacterial. Community acquired pneumonia is mainly distinguished from nosocomial acquired pneumonia and roughly 1.5 million people in the US are hospitalized owing to community acquired pneumonia every year [5]. Globally, around 200 million people get affected by viral pneumonia annually [6], with a 50-50 infection rate between adults and children. Generally, bacteria are classified as typical or atypical. Typical bacteria can be seen on the Gram strain and cultured through standard media, which, on the other hand, is not seen in atypical class of bacteria [7]. Typical bacteria-caused pneumonia can be listed as staphylococcus aureus, streptococcus pneumoniae, haemophilus influenza, etc. whereas atypical bacteria-caused pneumonia is mostly caused by chlamydia pneumoniae, legionella, chlamydia psittaci, and mycoplasma pneumonia [8]. Other types of pneumonia are fungal, aspiration, and hospital-acquired.

The most commonly sought diagnostic technique for all forms of pneumonia involves studying the increased opacity in regions of the lungs as shown by the chest radiograph, or chest X-ray (CXR). The increased opacity is caused due to the inflammation of the lungs with the high amounts of liquid in the affected areas [9]. There can be complications to the diagnosis of pneumonia through CXR because of the possibility of existence of pulmonary edema [10] which is mostly caused by cardiac problems, or internal lung bleeding, lung cancer, or in some patients, atelectasis [11] which results in unilateral collapse or shutdown of a part of a lung or the whole lung itself. In this condition, alveoli are deflated to very low volumes, visible from the increased opacity of the affected part seen in the CXR. Due to these complications, it becomes vital for having trained physicians and specialists, equipped with the patients' clinical record, to study the CXRs at different time frames for comparison and proper diagnosis.

The medical field has witnessed lots of breakthroughs in better diagnosis of diseases through machine learning [12-16]. In this paper, we use the principles of deep learning, which is a branch of machine learning by employing convolutional neural networks (CNN) trained on normal and pneumonia positive CXR images to correctly identify whether or not a new CXR fed into the network is pneumonia positive when in the field for diagnostic purposes. The rest of the paper is organized as follows: 2. Related Work, 3. Methods and Data, 4. Experimentation and Results, 5. Conclusions and Future Work. We do a comprehensive evaluation of different

structures of the CNN to best identify the ideal architecture for the particular CXR dataset used in Section 4.

2. Related Work

Researchers in the past have emphasized on the use of deep learning frameworks, namely, CNNs to tackle the problem of diagnosis of pneumonia [17-20]. However, most of the work done in the literature focuses on detection of features through modern CNN architectures which are deeper than traditional, few layer architectures, for example, VGGNet proposed by Simonyan and Zisserman [21] and ResNet proposed by He et al. [22]. For instance, Kermany et al. [17] perform a comprehensive study of development of diagnostic tools to treat patients with treatable blinding retinal diseases and pneumonia with a state-of-the-art deep CNN. One of the problems with these deep, modern CNN architectures is the difficulty of training all the corresponding layers which proves to be quite time-taking and computationally extensive. To solve this, Kermany et al. [17] used a technique called "transfer learning" which essentially means the use of pre-trained weights in the neural network to kick-start the initialization of learning and expedite the whole training process through all its layers with the requirement of only a fraction of the training data.

Another example of the use of very deep convolutional architectures is demonstrated by ChexNet proposed by Rajpurkar et al. [18]. Their architecture correctly identifies pneumonia and goes further to localize the areas of maximum lung inflammation in a heat-map fashion. Zech et al. [19] studied the performance of training of ChexNet on an internal dataset of pneumonia and normal medical CXRs and tested it on an external dataset. Through the work done in [19] it became very clear that a generalized pneumonia detection model must be trained on pooled data from different sources (say, hospitals or different departments in a hospital) for better generalization of model behaviour.

Pankratz et al. [23] made use of machine learning algorithm namely logistic regression to detect usual interstitial pneumonia (UIP) distinguished from non-UIP cases with the area under the receiver-operator characteristic curve (AUC) to be as high as 0.92. Generally, there is noticed a trade-off between intelligibility of machine learning systems and the accuracy they attain in the field of medicine. The models that attain a high accuracy usually are not very intelligible. In other words, one cannot exactly understand every step of the process a less intelligible model undertakes and hence understanding, editing or validation of parameters of such models becomes difficult, even though they provide a high accuracy. We see such trade-offs when we are faced to choose between simple and intelligible machine learning algorithms like logistic regression or random forest put against more complex, less intelligible deep learning models like artificial neural networks which offer higher accuracy. In a field like medicine, a high accuracy may not always be the prime goal because these AI systems are augmented - with the supervision by a qualified person (say, the doctor) who has the final say.

To solve this problematic trade-off, Caruana et al., worked on development of an intelligible model using generalized additive models (GAMs) [24] to make generalized additive models with pairwise interactions (GA^2Ms) to achieve state-of-the-art accuracy on CXR data [25]. Wang et al., created a hospital-scale chest X-ray dataset and collected over 100,000 frontal view CXRs of over 30000 unique

patients for eight common thoracic diseases [26]. One of these was pneumonia which was detected through localization using a unified weakly-supervised multi-label image classification framework. For the CNN, they used transfer learning (ImageNet pretrained) to apply AlexNet, GoogLenet, VGGNet and ResNet except for the last fully connected layers for each of these models. For pneumonia, their method encountered a precision of 0.66, recall of 0.93 and an F1-score of 0.77. Sirazitdinov et al. [27] used a combination of two models, namely, Mask R-CNN [28] and RetinaNet [29] to form a deep ensemble model for the detection and localization of pneumonia. They achieved state-of-the-art mean Average Precision (mAP) for localization of the pathology and also comparable precision, recall and F1-scores of 0.758, 0.793 and 0.775. Following a similar approach as ours, Stephen et al. (2019) used a simple CNN architecture with a few layers to achieve high validation accuracy on a dataset of CXRs [30]. Table 1 summarizes and analyses the related work.

Table 1. Analyses of related work.

No.	Ref. no.	Technique	Advantage	Disadvantage
1.	[17]	Pre-trained (ImageNet) CNN for classification	Quick learning procedure even for very deep CNNs	Transfer learning from ImageNet may involve negative transfer; when the CNN is pre-trained on data not similar to CXRs
2.	[18]	121-layer CNN to output localization-based pathology detection	Density-based localization of pathology providing a very descriptive result	Only uses frontal CXRs which may not always be adequate for diagnosis (without lateral view)
3.	[23]	Genomic classifier distinguishing UIP from non-UIP through logistic regression	Achieves very high specificity and AUC (area under ROC curve) of 0.86 and 0.86, respectively	Takes into consideration a smaller cohort for experiments. Additionally, achieves a sub-par sensitivity of 0.63
4.	[25]	Use of a highly intelligible additive model for pneumonia detection	Work shows that intelligibility and achieving state-of-the-art accuracy is possible through their approach	Although the model is intelligible, the exact reason for the predictions being made are not known (various factors: overfitting, variable correlation, etc.)
5.	[26]	Creation of a thoracic diseases CXR-based dataset and applying pre-trained CNNs for heatmap localization	Creation and usage of a very diverse and vast dataset (108,948 frontal CXRs of 32,717 unique patients)	The ground-truth labels were attained through data mining which were an individual radiologist's judgement. Due to ambiguity of pathologies' appearance on CXRs, this ground truth may not always be accurate
6.	[27]	Creation of a deep ensemble	Deep ensemble learning achieves	Training both models is very intensive in terms of

		pathology detection and localization model using Mask R-CNN and RetinaNet	high mAP in terms of localization of pathology	CPU and may be tedious. Work only uses frontal CXRs without lateral ones which may be stated a disadvantage
7.	[28]	4 layered-CNN used for detection of pneumonia	Simple CNN architecture which achieves 93% accuracy on validation	Work does not mention important metrics such as precision, recall and F1 score

In this paper, we demonstrate that simpler CNN architectures can achieve state-of-the-art accuracy as compared to their bigger architecture counter parts with the added advantage of lower training times and lower computational expenses. Moreover, we do not employ any transfer learning techniques in our study. This is because, the sole purpose of our work is to explore the efficacy of smaller CNN architectures when put against very deep CNNs. There is no transfer learning required to train the architectures of the scale that we use in this paper since the weights are easily learnt due to the low number of layers. In fact, most transfer learning methods may have the issue of negative transfer where the model has pre-trained weights from images not related at all to the task that one tries to solve. Usually, models are pre-trained on vast number of images such as ImageNet and due to the amount of uncertainty involved in the extent of negative transfer, the transfer learning approach can be very tedious to optimize. It may be possible that transfer learning used for the architectures that take longer training time may help the model converge quickly, however, this may trade more important metrics such as accuracy, precision and recall for reduced training time due to negative transfer.

3. Methods and Data

In this section, we describe the methods used behind the classification process of CXRs determined to be pneumonia positive or negative. We also specify the dataset used for the task and the technology with which the experiments and computations were carried out. This section maps out as follows: 3.1. Dataset use, 3.2. Image augmentation, 3.3. Artificial Neural Network (ANN), 3.4. Convolutional Neural Network (CNN) and 3.5. Software and hardware.

3.1. Dataset use

The CXRs were collected from paediatric patients between one to five years of age from Gangzhou Women and Children's Medical Center, Gangzhou by Kermany et al. [17]. A sample distinction between a normal CXR and a pneumonia positive CXR is shown in Fig. 1. Table 2 describes the amount of CXRs taken for training and testing of the model.

Table 2. Dataset layout and proportions of images taken for training and testing.

	Normal	**Pneumonia**
Training	1341	3875
Testing	242	389
Total	1583	4264

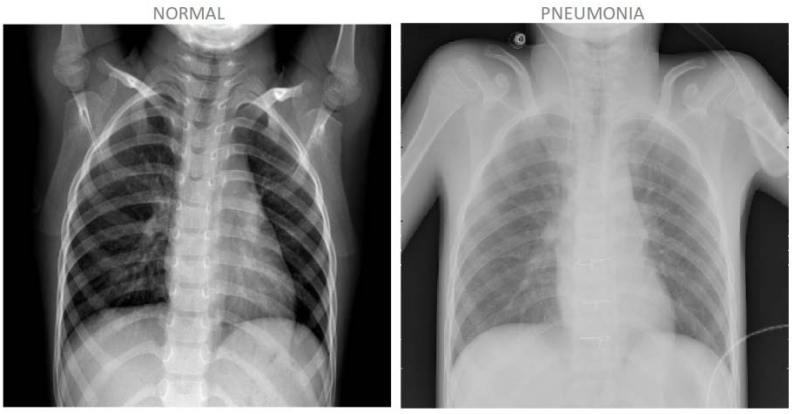

Fig. 1. (L) CXR of a normal individual, and (R) of a pneumonia positive patient. The increased opacity can clearly be seen in the pneumonia positive CXR which is a powerful indicator.

3.2. Image augmentation

Deep learning models usually require enormous amounts of data to train on for them to correctly function. In the case of CNNs, thousands upon thousands of images are required which, pragmatically speaking, is a difficult prospect. To solve this, a technique named image augmentation is used [31-34]. Image augmentation virtually increases the size of an existing dataset to a large extent with techniques like standardization of features (pixel values), whitening transforms [35], random rotations and shifts, flipping, rescaling, shearing, zooming, etc. In our approach, we rescale, shear, zoom and flip horizontally the training image for a wider variety and augmentation of training data.

3.3. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) are considered to be the backbone of deep neural networks and the deep learning framework. ANNs are powerful models which are used in a wide range of fields like regression and classification tasks, and some special applications like character recognition [36-40], speech recognition [41-44], face recognition [45-48], etc. They are also often used in combination with some specialized layers like convolutional and deconvolutional layers. Broadly, they consist of three layers: input, hidden and output. Figure 2 describes the structure of an ANN. Each layer's nodes are connected, in a directed fashion, to each of the nodes in the next layer through weights, which are adjusted during the training process. Let $L = \{L_1, L_2, \dots, L_Q\}$ be the set of layers of an ANN having a total Q layers with Q-2 hidden layers. Then we have a set of nodes $N = \{N_1, N_2, \dots, N_Q\}$ where $N_i = \{N_i^1, N_i^2, \dots, N_i^{P_i}\}, 1 \leq i \leq Q$ and P_i denotes the number of nodes in layer L_i. The links between the nodes are defined as weights by set $W = \{W_2, W_3, \dots, W_Q\}$ whose first element is W_2 because the first layer L_1 (input layer) does not have weights from a previous layer which is non-existent. Each element in W is defined by $W_i = \{W_i^{1,1}, W_i^{1,2}, \dots, W_i^{1,P_i}, W_i^{2,P_i}, \dots, W_i^{P_{i-1},P_i}\}$. For instance, the weight between 3rd node of L_4 and 2nd node of L_3 is given by $W_4^{2,3}$. Each node in an ANN is designed to mimic a *neuron* which has an option to either fire and

propagate a signal, or completely ignore it. Each node receives input from all other nodes in the previous layer multiplied by their connection weights which is summed as,

$$S = \sum_{j=1}^{P_{i-1}} W_i^{j,k} N_{i-1}^{j}, 1 \leq k \leq P_i \quad (1)$$

When an input is fed to the input layer, all the inputs get propagated through the network with some random initialized weights with each node having an activation function $\varphi(S)$ which decides whether or not to fire (propagate) the input forward. The activation function $\varphi(S)$ are of many types, such as sigmoid, rectified linear unit (reLU), leaky reLU, softmax, etc. The predicted values formed at the output layer are compared with test values and an error is backpropagated through the network to adjust the weights for better results in the next iteration or epoch.

Figure 2 shows an ANN having the following number of nodes, layer-wise (L-R): {12, 10, 10, 10, 2}. The input layer is the first layer with 12 nodes where the inputs are provided. Output layer has 2 nodes which have a softmax activation function for values to sum up to unity.

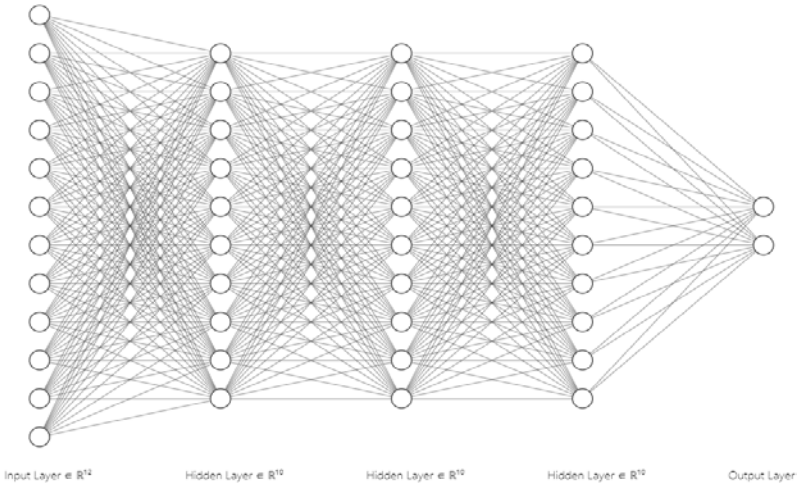

Fig. 2. The Artificial Neural Network (ANN).

3.4. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) [49] are one of the most popular deep learning framework models proposed by LeCun et al. which revolutionized deep learning and is also the model we use in our approach. CNNs are mainly used to operate on image data to help classify objects. CNNs have been used to detect faces [50, 51], scene labelling [52], action recognition [53, 54], etc.

CNNs are composed of two main types of layers - convolutional and dense (here dense refers to the ANN layers), Fig. 3 shows the procedure of usage of CNNs. The main objective of convolutional layers is to find features in an image using feature detectors which are 2D matrices called kernels or filters. This

obtains a feature map which stores the spatial relationships between pixels of the input image.

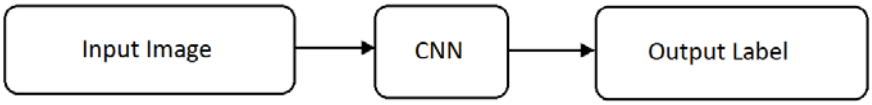

Fig. 3. CNN used for classification of images.

These spatial relationships are vital to finding differences between two objects in two images. A convolution operation of two functions $f(t)$ and $g(t)$ can be defined as,

$$f(t) * g(t) \triangleq \underbrace{\int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau}_{(f*g)(t)} \tag{2}$$

After the convolution layer, a pooling layer is used to further down sample the detected features which helps to detect the image whether it is tilted, stretched or rotated which helps provide spatial invariance to these models. Pooling reduces the parameters by 75% and also prevents overfitting. The next layer is to flatten all the pooled values and arrange them into a vertical array to be fed into an ANN (the dense layers). This step is called the full connection (FC). Finally, the values are propagated through the ANN to get an output at the output layer. Figure 4 displays a full convolutional neural network.

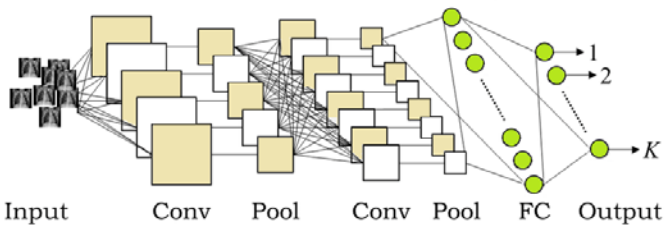

Fig. 4. Architecture of a 2 layered CNN.

3.5. Software and hardware

All the various architectures of CNNs that were trained were done through Python 3 with the Keras library (that uses TensorFlow backend) high level API for construction of neural networks on a workstation with Intel i5 8th generation processor with 8 GB RAM.

4. Experimentation and Results

In this section, we discuss the comprehensive experiments performed on the CXR dataset through different architectures as we select the most ideal one for

classification. We divide this section as follows: 4.1. Experiments and analysis, and 4.2. results of selected architecture.

4.1. Experiments and analysis

We carried out experiments with different CNN architectures pertaining to the number of convolutional layers, number of dense layers, inclusion, and exclusions of regularizations like L1, L2, BatchNorm (Batch Normalization) and Dropout, image input sizes, kernel sizes, pooling matrix sizes, and compared each architecture's performance based on the maximum accuracy achieved during training in addition to the least cross entropy loss encountered during training. Table 3 specifies the meanings of the abbreviations used in Table 4 which contains information of 15 different CNN architectures and their maximum accuracy and least loss along with the time taken in seconds to train them. One thing to note is that the number of nodes in each ANN layer is taken to be 128 except for the last output layer having 2 output nodes.

Figure 4 summarizes the statistics of each architecture based on LVCEL, MVA and TT. Figure 5(a) plots the training accuracy and Fig. 5(b) loss achieved per each epoch for each architecture. From Table 4 we notice that architectures 10 and 11 perform poorly with respect to MVA and LVCEL. This may be due to the L1 regularization added on these architectures (which is not applied on any other case) which restricts the model from correctly learning the relationships in the dataset. We also notice a directly proportional relationship between architecture size and TT. However, TT highly depends upon IS and as we see in architectures 5, 6 and 13, when IS (64, 64), TT is greatly reduced. It is helpful to note that TT is also directly proportional to the how computationally expensive training the model is. Training neural networks takes a lot of computational power, and the longer it runs, the more intensive the task is.

Finally, we can reduce from Table 4 that MVA and LVCEL do not improve drastically with increasing architecture size, whereas TT increases drastically (Fig. 6(c)). This brings us to the simple conclusion that for a given dataset, the simpler the architecture, the better. Adding extra layers to the neural network may cause it to underfit the data as it starts to attempt to detect features which are not really existent. Based on MVA, LVCEL and TT, for our purposes, we select architecture 5 for further consideration as the most ideal model and we use this simple architecture to evaluate model performance on the CXR dataset in the next sub-section.

Table 3. Abbreviations used in Table 4.

Abbreviation	Meaning
CL	Number of CNN Layers in specific architecture.
AL	Number of ANN Layers in specific architecture.
L1	Level 1 regularization
L2	Level 2 regularization
BN	Batch Normalization
DO	Dropout
IS	Input Size of image (64x64 or 128x128) fed to the CNN
FD	Features Detected in a convolutional layer, layer-wise. Varies from layer to layer, so for instance, {64, 32} refers to 64 features detected in first layer and 32 in the second layer.

KS	Kernel Sizes for each convolutional layer, layer-wise. Assumed to always be square matrices, for instance, {9, 3} refers to 9x9 in the first layer and 3x3 in the second layer.
PS	Pooling Sizes followed by every convolutional layer, always assumed to be a square matrix.
TT	Time Taken in time units to train the model.
LVCEL	Least Validation Cross Entropy Loss encountered during training. Here, validation refers to results on test set.
MVA	Maximum Validation Accuracy encountered during training.

Table 4. Performance of 15 different CNN architectures.

S. No.	CL	AL	Regularizations				IS	FD	KS	PS	LVCEL	MVA	TT
			L1	L2	BN	DO							
1	2	3	✗	✗	✗	✗	(128,128)	{64,32}	{9,3}	{4,2}	0.138	0.8984	2147
2	2	4	✗	✗	✗	✗	(128,128)	{64,32}	{9,3}	{4,2}	0.1963	0.8703	2440
3	2	4	✗	✗	✓	✓	(128,128)	{64,32}	{9,3}	{4,2}	0.3541	0.8703	2261
4	2	4	✗	✗	✓	✗	(128,128)	{64,32}	{9,3}	{4,2}	0.1733	0.9	2647
5	2	4	✗	✗	✗	✓	(64,64)	{64,32}	{9,3}	{4,2}	0.2017	0.9031	1563
6	2	2	✗	✗	✗	✗	(64,64)	{64,32}	{9,3}	{4,2}	0.2024	0.8938	1563
7	2	3	✗	✗	✓	✓	(128,128)	{64,32}	{9,3}	{4,2}	0.2849	0.8797	2237
8	3	4	✗	✗	✓	✗	(128,128)	{128,64,32}	{9,6,3}	{4,2,2}	0.2669	0.8	3549
9	3	5	✗	✓	✗	✗	(128,128)	{128,64,32}	{9,6,3}	{4,2,2}	0.2744	0.9	3591
10	4	4	✓	✗	✗	✓	(128,128)	{128,64,32,16}	{9,6,3,3}	{4,2,2,2}	0.5624	0.6996	3554
11	4	5	✓	✓	✗	✗	(128,128)	{128,64,32,16}	{9,6,3,3}	{4,2,2,2}	0.6898	0.6928	4125
12	4	5	✗	✗	✗	✗	(128,128)	{128,64,32,16}	{9,6,3,3}	{4,2,2,2}	0.2702	0.8953	4253
13	4	5	✗	✗	✗	✗	(64,64)	{64,32,32,16}	{9,6,3,3}	{2,2,2,2}	0.2492	0.8766	1410
14	5	5	✗	✗	✗	✗	(128,128)	{128,64,64,32,16}	{9,6,6,3,3}	{2,2,2,2,2}	0.1127	0.8859	8254
15	5	6	✗	✗	✗	✗	(128,128)	{128,64,64,32,16}	{9,6,6,3,3}	{2,2,2,2,2}	0.3088	0.8797	8200

4.2. Results of selected architecture

We evaluate the model performance of architecture 5 on the test set specified by Table 2. As the evaluation metrics, we use confusion matrix and calculate the *Sensitivity* and *Specificity* as follows,

$$Sensitivity = {TP}/{(TP + FN)} \quad (3)$$

$$Specificity = {TN}/{(TN + FP)} \quad (4)$$

where TP, TN, FN, FP refer to True Positive, True Negative, False Negative, and False Positive, respectively.

Sensitivity in our case is defined as the proportion of patients who test pneumonia positive. Specificity, on the other hand, is the proportion of patients not having pneumonia being tested negative. Figure 7 illustrates the confusion matrix of the selected model.

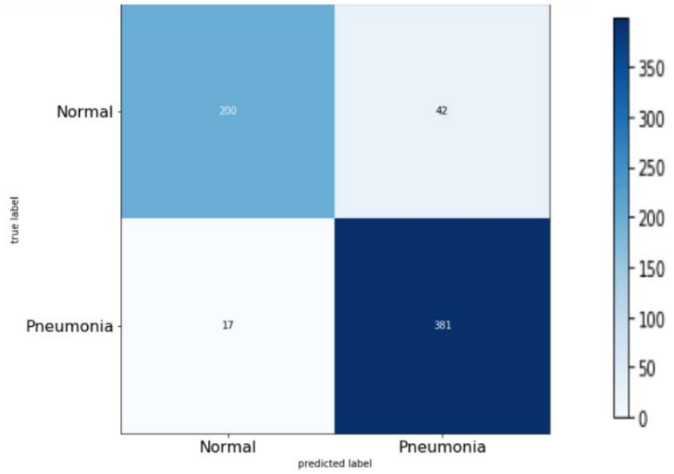

Fig 7. The confusion matrix for model of architecture 5 (see Table 4). From the figure, we see that TP=381, TN=200, FP=17, FN=42.

Hence, we calculate from Eqs. (3) and (4), $Sensitivity = 0.9007$ and $Specificity = 0.9216$. Both these metrics are above 90% which is a good indicator that the model performs excellently in both, having the ability to correctly identify most of the positive pneumonia cases, and also the ability of ruling out the negative cases. Moreover, the ROC curve illustrated in Fig. 8 whose area under the curve (AUC) is 0.9582 which is very high. These metrics are comparable to most of the state-of-the-art architectures applied on CXRs for pneumonia detection.

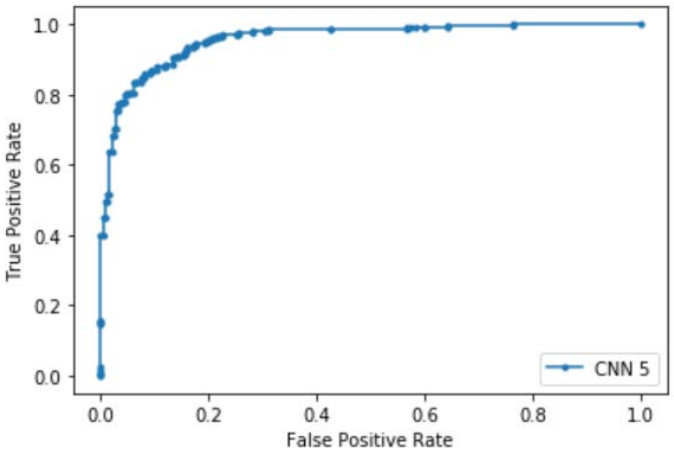

Fig 8. The receiver-operating characteristics (ROC) curve for 5th CNN architecture. Area under this curve is calculated to be 0.9582 unit.

5. Conclusions and Future Work

In this paper, we attempt to find a simpler approach for pneumonia detection based on CXRs by comparing the performances of 15 different CNN architectures trained on the same dataset. Based on our findings, we select the most ideal model which is easy to train (less computationally expensive and quicker), intelligible, and has one of the best performance metrics. The metrics of the selected architecture compare to some of the state-of-the-art architectures trained on CXRs that goes ahead to prove that striving for the simplification of CNN architectures is crucial for intelligibility without compromising accuracy and quality of performance. As learnt from the findings in this research work, we recommend that there be more research into fine-tuning simpler architectures to gain even higher levels of accuracy. We admit that the simple architecture 5 that was chosen after thorough experimentation can perform even better if fine-tuned and experimented upon more. Future work on detection of pneumonia may also be done through multimodal learning where symptoms of the patients as text could be taken by the model along with the CXR image for even better diagnosis.

Nomenclatures

L	Set of layers
N	Set of nodes
P_i	Number of nodes in layer i
Q	Total layers
S	Neural firing sum
W	Set of weights

Greek Symbols

φ	Activation function

Abbreviations

AL	Artificial neural network Layers
ANN	Artificial Neural Network
AUC	Area Under Curve
BN	Batch Normalization
CL	Convolutional Layers
CNN	Convolutional Neural Network
CXR	Chest X-ray
DO	Dropout
FC	Full Connection
FD	Features Detected
FN	False Negative
FP	False Positive
IS	Input Size
KS	Kernel Size
L1	L1 regularization
L2	L2 regularization
LVCEL	Least Validation Cross Entropy Loss
MVA	Maximum Validation Accuracy
PS	Pooling Size
ROC	Receiver-Operating Characteristics

TN	True Negative
TP	True Positive
TT	Training Time
UIP	Usual Interstitial Pneumonia

References

1. Marrie, T. (1994). Community-acquired pneumonia. *Clinical Infectious Diseases*. 18(4), 501-513.
2. Martin, E. (2015). *Concise Medical Dictionary*. Oxford University Press.
3. Rui, P.; and Kang K. (2015). National ambulatory medical care survey: 2015 emergency department summary tables. table 27. Retrieved April 27, 2020, from www.cdc.gov/nchs/data/nhamcs/webtables/2015edwebtables.pdf
4. Deaths: Final Data (2015). Supplemental tables. tables I-21, I-22. Retrieved April 27, 2020, from www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr6606tables.pdf
5. Ramirez, J.A.; Wiemken, T.L.; Peyrani, P.; Arnold, F.W.; Kelley, R.; Mattingly, W.A.; Nakamatsu, R.; Pena, S.; Guinn, B.E.; Furmanek, S.P.; Pesaud, A.K.; Raghuram, A.; Fernandez, F.; Beavin, L.; Bosson, R.; Fernandez-Botran, R.; Cavallazzi, R.; Bordon, J.; Valdivieso, C.; Schulte, J.; and Carrico, R.M. (2017). Adults hospitalized with pneumonia in the united states: incidence, epidemiology, and mortality. *Clinical Infectious Diseases*, 65(11), 1806-1812.
6. Ruuskanen, O.; Lahti, E.; Jennings, L.C.; and Murdoch, D.R. (2011). Viral pneumonia. *The Lancet*, 377(9773), 1264-1275.
7. Calik, S.; Ari, A.; Bilgir, O.; Cetintepe, T.; Yis, R.; Sonmez, U.; and Tosun, S. (2018). The relationship between mortality and microbiological parameters in febrile neutropenic patients with hematological malignancies. *Saudi Medical Journal*, 39(9), 878-885.
8. Sattar, S.B.A; and Sharma, S. (2020). *Bacterial pneumonia*. Treasure Island (FL): StatPearls Publishing.
9. Franquet, T. (2018). Imaging of Community-acquired Pneumonia. *Journal of Thoracic Imaging*, 33(5), 282-294.
10. Staub, N.C. (1974). Pulmonary edema. *Physiological Reviews*, 54(3), 678-811.
11. Woodring, J. H.; and Reed, J. C. (1996). Types and mechanisms of pulmonary atelectasis. *Journal of Thoracic Imaging*, 11(2), 92-108.
12. Nayak, S.; Gourisaria, M.K.; Pandey, M.; and Rautaray, S.S. (2019). Heart disease prediction using frequent item set mining and classification technique. *International Journal of Information Engineering and Electronic Business (IJEEB) 2019*, 11(6), 9-15.
13. Gourisaria, M.K.; Das, S.; Sharma, R.; Rautaray, S.S.; and Pandey, M.A (2020). Deep learning model for malaria disease detection and analysis using deep convolutional neural networks. *International Journal of Emerging Technologies (IJET)*. 11(2), 699-704.
14. Nayak, S.; Gourisaria M.K.; Pandey M.; and Rautaray S.S. (2019). Comparative analysis of heart disease classification algorithms using big data analytical tool. *International Conference on Computer Networks and Inventive Communication Technologies (ICCNCT 2019)*. Springer, Cham, 582-588.

15. Nayak, S.; Gourisaria, M.K.; Pandey, M.; and Rautaray, S.S. (2019). Prediction of heart disease by mining frequent items and classification techniques. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*. Madurai, India, 607-611.
16. Rautaray S.S.; Pandey, M.; Gourisaria, M.K.; Sharma, R.; and Das, S. (2020). Paddy crop disease prediction-A transfer learning technique. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(6), 1-6.
17. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; Dong, J.; Prasadha, M.K.; Pei, J.; Ting, M.Y.L.; Zhu, J.; Li, C.; Hewett, S.; Dong, J.; Shi, W.; Fu, X.; Duan, Y.; Huu, V.A.N.; Wen, C.; Zhang, E.D.; Zhang, C.L.; Li, O.; Wang, X.; Singer, M.A.; Sun, X.; Xu, J.; Tafreshi, A.; Lewis, M.A.; Xia, H.; and Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122-1131.
18. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; Lungren, M.P.; and Ng, A.Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.
19. Zech, J.R.; Badgeley, M.A.; Liu, M.; Costa, A.B.; Titano, J.J.; and Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, 15(11), e1002683.
20. Xu, X.; Jiang, X.; Ma, C.; Du, P.; Li, X.; Lv, S.; Yu, L.; Chen, Y.; Su, J.; Lang, G.; Li, Y.; Zhao, H.; Xu, K.; Ruan, L.; and Wu, W. (2020). Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv preprint arXiv:2002.09334.
21. Simonyan, K.; and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
22. He, K.; Zhang, X.; Ren, S.; and Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, 770-778.
23. Pankratz, D.G.; Choi, Y.; Imtiaz, U.; Fedorowicz, G.M.; Anderson, J.D.; Colby, T.V.; Myers, J.L.; Lynch, D.A.; Brown, K.K.; Flaherty, K.R.; Steele, M.P.; Groshong, S.D.; Raghu, G.; Barth, N.M.; Walsh, P.S.; Huang, J.; Kennedy, G.C.; and Matinez, F.J. (2017). Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Annals of the American Thoracic Society (ATS)*, 14(11), 1646-1654.
24. Hastie, T.; and Tibshirani, R. (1995). Generalized additive models for medical research. *Statistical Methods in Medical Research*, 4(3), 187-196.
25. Caruana, R.; Lou, Y.; Gehrke, J.; Koch, P.; Sturm, M.; and Elhadad, N. (2015). Intelligible models for healthCare. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. 1721-1730.
26. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; and Summers, R.M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, 3462-3471.

27. Sirazitdinov, I.; Kholiavchenko, M.; Mustafaev, T.; Yixuan, Y.; Kuleev, R.; and Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. *Computers amd Electrical Engineering*, 78, 388-399.
28. He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2980-2988.
29. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2999-3007.
30. Stephen, O.; Sain, M.; Maduh, U.J.; and Jeong, D.-U. (2019). An efficient deep learning approach to pneumonia classification in health care. *Journal of Healthcare Engineering*, 2019, 1-7.
31. Mikolajczyk, A.; and Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. *International Interdisciplinary PhD Workshop (IIPhDW)*. Swinoujście, Poland, 117-122.
32. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; and Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321, 321-331.
33. Wigington, C.; Stewart, S.; Davis, B.; Barrett, B.; Price, B.; and Cohen, S. (2017). Data augmentation for recognition of handwritten words and lines using a CNN-LSTM network. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Kyoto, Japan, 639-645.
34. Dellana, R.; and Roy, K. (2016). Data augmentation in CNN-based periocular authentication. *6th International Conference on Information Communication and Management (ICICM)*. Hatfield, England, 141-145.
35. Koivunen, A.C.; and Kostinski, A. (1999). The feasibility of data whitening to improve performance of weather radar. *Journal of Applied Meteorology and Climatology*, 38(6), 741-749.
36. Singh, R.; Yadav, C.S.; Verma, P.; and Yadav, V. (2010). Optical character recognition (OCR) for printed devnagari script using artificial neural network. *International Journal of Computer Science and Communication*, 1(1), 91-95.
37. Amin, A.; Al-Sadoun, H.; and Fischer, S. (1996). Hand-printed Arabic character recognition system using an artificial network. *Pattern Recognition*, 29(4), 663-675.
38. Mani, N.; and Srinivasan, B. (1997). Application of artificial neural network model for optical character recognition. *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 3, 2517-2520.
39. Patil, V.; and Shimpi, S. (2011). Handwritten English character recognition using neural network. *Elixir Computer Science and Engineering*, 41, 5587-5591.
40. Barve, S. (2012). Optical character recognition using artificial neural network. *International Journal of Advanced Research in Computer Engineering and Technology*, 1(4), 131-133.

41. Jaitly, N.; Nguyen, P.; Senior, A.; and Vanhoucke, V. (2012). Application of pretrained deep neural networks to large vocabulary speech recognition. *Proceedings of Interspeech* 2012, 2578-2581.
42. Dede, G.; and Sazli, M.H. (2010). Speech recognition with artificial neural networks. *Digital Signal Processing*, 20(3), 763-768.
43. Scanzio, S.; Cumani, S.; Gemello, R.; Mana, F.; and Laface, P. (2010). Parallel implementation of artificial neural network training for speech recognition. *Pattern Recognition Letters*, 31(11), 1302-1309.
44. Wijoyo, S.; and Wijoyo, S. (2011). Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot. *Proceedings of 2011 International Conference on Information and Electronics Engineering (ICIEE 2011)*. Singapore, 28-29.
45. Agarwal, M.; Jain, N.; Kumar, M.M.; and Agrawal, H. (2010). Face recognition using eigen faces and artificial neural network. *International Journal of Computer Theory and Engineering*, 2(4), 624.
46. Zhang, M.; and Fulcher, J. (1996). Face recognition using artificial neural network group-based adaptive tolerance (GAT) trees. *IEEE Transactions on Neural Networks*, 7(3), 555-567.
47. Réda, A.; Aoued, B. (2004). Artificial neural network-based face recognition. *First International Symposium on Control, Communications and Signal Processing*, 2004, 439-442.
48. Propp, M.; and Samal, A. K. (1992). Artificial neural network architectures for human face detection. *Proceedings of the 1992 Artificial Neural Networks in Engineering, ANNIE'92*. Fairfield, United State, 535-540.
49. LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
50. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; and Back, A.D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1), 98-113.
51. Ranjan, R.; Sankaranarayanan, S.; Castillo, C.D.; and Chellappa, R. (2017). An all-in-one convolutional neural network for face analysis. *12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*. Fairfield, United State, 17-24.
52. Pinheiro, P.H.; and Collobert, R. (2014). Recurrent convolutional neural networks for scene labelling. *31st International Conference on Machine Learning (ICML)*. Beijing, China, 82-90.
53. Ji, S.; Xu, W.; Yang, M.; and Yu, K. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221-231.
54. Du, Y.; Fu, Y.; and Wang, L. (2015). Skeleton based action recognition with convolutional neural network. *3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. Kuala Lumpur, Malaysia, 579-583.