Optimized CNN-based Diagnosis System to Detect the Pneumonia from Chest Radiographs

Mohammed Aledhari*, Shelby Joji†, Mohamed Hefeida‡, and Fahad Saeed§
*†College of Computing and Software Engineering, Kennesaw State University, Georgia, USA
‡Electrical and Computer Engineering, University of Idaho, Idaho, USA
§School of Computing and Information Sciences, Florida International University, Florida, USA
Email: *maledhar@kennesaw.edu, †sjoji@kennesaw.edu, ‡hefeida@uidaho.edu, §fsaeed@fiu.edu

***Abstract*—Pneumonia is a high mortality disease that kills 50, 000 people in the United States each year. Children under the age of 5 and older population over the age of 65 are susceptible to serious cases of pneumonia. The United States spend billions of dollars fighting pneumonia-related infections every year. Early detection and intervention are crucial in treating pneumonia related infections. Since chest x-ray is one of the simplest and cheapest methods to diagnose pneumonia, we propose a deep learning algorithm based on convolutional neural networks to identify and classify pneumonia cases from these images. For all three models implemented, we obtained varying classification results and accuracy. Based on the results, we obtained better prediction with average accuracy of (68%) and average specificity of (69%) in contrast to the current state-of-the-art accuracy that is (51%) using the Visual Geometry Group (VGG16 also called OxfordNet), which is a convolutional neural network architecture developed by the Visual Geometry Group of Oxford. By implementing more novel lung segmentation techniques, reducing over fitting, and adding more learning layers, the proposed model has the potential to predict at higher accuracy than human specialists and will help subsidies and reduce the cost of diagnosis across the globe.**

***Index Terms*—Machine Learning, Pneumonia, Convolutional Neural Network, Infection, Chest Radiograph**

I. Introduction

Pneumonia or pneumonitis is an inflammatory condition that primarily affects lungs. It can cause mild to severe illness in people of all age groups. Pneumonia related infection can be caused by bacteria, virus, and in some cases fungi. Every year approximately 1 million people are hospitalized in the United States due to pneumonia related conditions [1]. According to [1], nearly 50,000 people die each year from pneumonia-related infections alone. Pneumonia mostly affects children under the age of 5 and older adult population over the age of 65. It is one of the largest causes of infectious death in children under the age of 5. Nearly 1 million children died from pneumonia-related complications in 2015 [2]. Individuals who are suffering from Human Immune Virus (HIV), diabetes, malnutrition, renal failure, cancer, or underlying lung conditions have impaired immunity and are more susceptible to complications arising from Pneumonia [3]. According to [4], ninety-two serotypes have been documented as of 2011 and the 10 most common result in 62% of invasive disease worldwide as shown in figure 1. With discovery of newer antibiotics and improved medical care, mortality rates were significantly reduced but still ranks among the top ten causes of disease resulting in death worldwide. Chest X-ray (CXR)

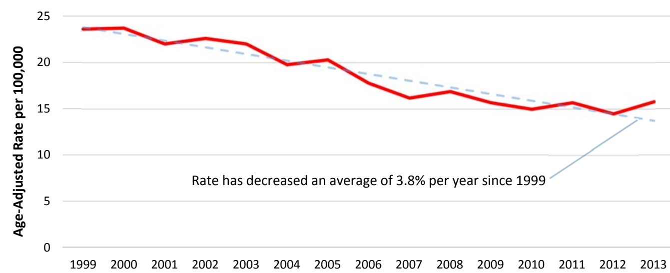


Fig. 1: Pneumonia and Influenza: Age Adjusted Death Rates by Year [5].

is one of the simplest techniques involved in diagnosing pneumonia. It is the first line of approach in diagnosing any lung disorders. However, detecting pneumonia on CXR is a challenging task. It heavily relies on expertise of the physician and is subject to misdiagnosis. Clinically, chest radiograph is one of two types - posteroanterior (PA) and lateral. When patients are severely ill, anteroposterior (AP) CXR is done. Chest radiograph depends on the ability to X rays to penetrate the matter. Depending on the type of pathogen, the appearance of CXR of pneumonia patients varies. In general, radiographic signs includes increased density of the infected area due to a mixture of consolidation and atelectasis [6] as shown in figure 2. As the infection becomes severe, air bronchogram appears due to the presence of air column surrounded by consolidation and atelectasis. As a result, dark bronchioles are surrounded by white area of infiltrates. This area appears as patchy and uniform throughout the area [6]. In some cases, pleural effusion due to the presence of fluid outside the lung and in the pleural space may also develop [6]. Pneumonia quite often presents with other disease process and appear with other radiographic findings which may include atelectasis, consolidation, and pleural effusion. Therefore, diagnosis of pneumonia from CXR is a difficult process. This is even complicated in the case of acute cases of pneumonia since infiltrates are beginning to develop on CXR. With the rapid advancements in computer vision and deep learning techniques, highly efficient techniques have been introduced in image classification, recognition, and segmentation. Deep learning

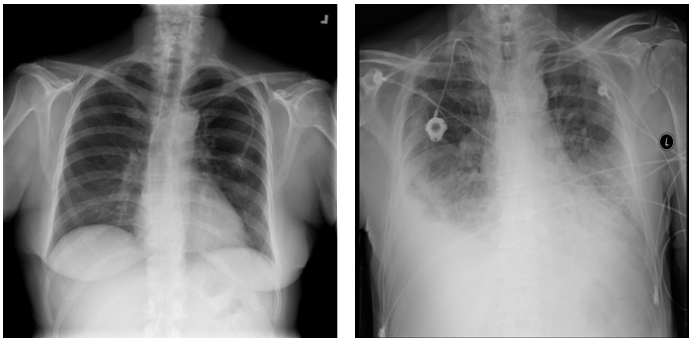


Fig. 2: Chest radiograph of normal (left) and pneumonia infected (right) lungs.

techniques can also be employed in medical data analysis such as in diagnosis of abnormalities from chest radiograph. Every day, hundreds of chest radiographs are being taken in hospitals across the nation. This creates a large number of unexplored medical images and datasets. In this paper, we proposed a deep-learning algorithm that can classify an x-ray image from a pneumonia patient.

A. Research Problem

Due to the increasing antibiotic resistance of newly emerging pathogens, and aging population, the economic burden of pneumonia is expected to continue increasing [7]. The effect of this on health care is inevitable and the financial burden is increasing by the day as shown in figures 3 and 4. Since CXR is simple and a first line approach in treating pneumonia, it is a great candidate for early detection of the disease. We

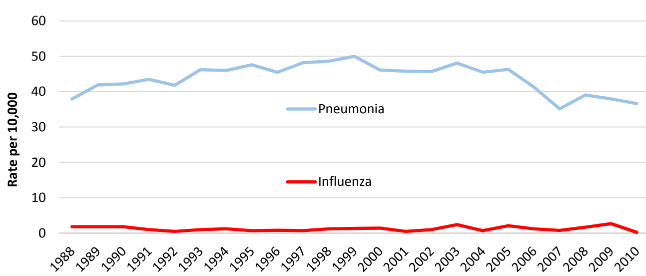
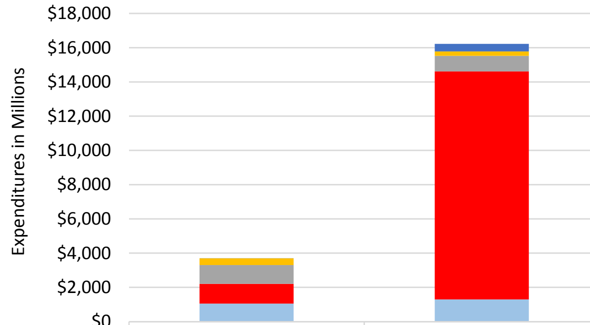


Fig. 3: Pneumonia and Influenza: Hospitalization Rate, 1988 - 2010 [5].

propose a model based on convolutional neural networks to help classify pneumonia cases from CXRs at an earlier stage. The input for this problem will be the chest radiograph and the output will be the presence or absence of pneumonia. For this reason, the research task was developed as a binary classification problem with an output y $\in$ {0, 1} for the coding of 0 that indicates to no pneumonia, while 1 refers to the presence of pneumonia. For this study, we optimized the binary cross entropy loss function as follows:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i)) \quad (1)$$



	Influenza	Pneumonia
Home	$3.7	$438.0
RX	$388.8	$259.6
ER	$1,099.9	$908.5
Hospitalization	$1,155.4	$13,319.0
Outpatient or Office	$1,051.7	$1,297.8

Fig. 4: Pneumonia and Influenza - Healthcare Expenditures by Disease and Type of Service, 2013 [5].

Where y is the outcome diagnosis of pneumonia, $p(y_i)$ is the predicted probability of having pneumonia and n is the number of radiographs analyzed in the training set.

B. Purpose of the Study

This research study focuses on providing a new paradigm for diagnosing the pneumonia from CXRs. We analyze open datasets available through the National Institute of Health consisting of 112K chest x-rays of more than 30,00 unique patients [8]. This dataset contains the CXR imagery of patients with advanced lung disease. Dataset consist of 14 different pathological findings which are pneumonia, atelectasis, cardiomegaly, consolidation, edema, effusion, emphysema, fibrosis, hernia, infiltration, mass, nodule, pleural thickening, and pneumothorax [8]. Additionally, the dataset contains CXR view type, patient gender, patient age, number of visits, original pixel spacing and original image height and width [8]. The purpose of the study is to perform analysis and implement several models using deep learning techniques to help correctly identify and classify the presence of pneumonia.

C. Contribution

This paper discusses the design and implementation of a deep learning model using convolutional neural network. Specifically, we aim to improve the accuracy of detecting pneumonia from CXRs. For machine learning to be incorporated in areas such as predicting pneumonia, high accuracy that exceeds that of human experts are critical. As this may not be possible in the initial stages, higher accuracy along with human input can together increase the throughput. In either case, accuracy is a metric that evaluates the performance of the model. For this paper, we optimized the pneumonia detection model using pertained model available from ImageNet [9] that is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories. Several notable work on the same National Institutes of Health (NIH) dataset yields promising results on predicting pneumonia. As seen in [8], Visual Geometry Group (VGG16 also called OxfordNet)-based model [10] yielded an accuracy of 51% [8] and ResNet-50 [11] yielded an accuracy of 63%. A 121-layer dense net model introduced by Rajpurkar et al [12] provides a better accuracy rate of 76%. Guan et al [13] improved on dense net 121 by adding attention guided layer to arrive at an accuracy rate of 78%. This work will be focusing primarily on fine tuning VGG16, ResNet-50, and Inception v3 [14] models. Additionally, all three models will be compared based on the results from a confusion matrix.

D. Motivation

Growing the health issues such as Pneumonia and Influenza, especially for children and seniors require better diagnosis methods such as computer-based solutions that aim to detect and treat those issues in early stages as possible. Also, the current advancement in the enabling technologies: hardware and software such as the computerized tomography (CT) scan [15], computer vision, machine learning, and deep learning algorithms in addition to the authors' knowledge and expertise in the field are encouraging to propose and implement a new Pneumonia diagnosis method that provides better accuracy and precision compares to the current ones. Moreover. the proposed method is aiming to participate in creating a healthy community, then enhancing the quality of life (QoF).

E. Paper Goals and Organization

The purpose of this paper is to design and implement a novel Pneumonia diagnosis computer vision-based algorithm to decrease related health issues and costs. In addition, we increased the accuracy of the Pneumonia detection that helps physicians and medical doctors to treat that issue in early stages using a computer. The contributions of this paper are outlined as follows:

- Summarizes the related works and current techniques to detect the Pneumonia with their features, advantages, and disadvantages.
- Provides an overview of the challenges of the tradition diagnosis methods that relies on specialist visions without utilizing the advancement of computer and deep learning techniques that provide better accuracy and precision.
- Presents the need for better Pneumonia diagnosis techniques to provide faster detection and treatment in lower costs.

The remainder of this paper is organized as follows: Section II presents the related works in terms of Pneumonia diagnosis methods. Section III discusses the model description and formulation (proposed method). Section IV presents the experiments and results of the proposed Pneumonia diagnosis algorithm. Section V presents our conclusions.

II. Related Works

Deep learning methodologies are becoming very common in the field of medical image classification. This can be attributed primarily due to huge success rate of these algorithms. The challenges of early machine learning models include low accuracy which was mostly depending on the ability of feature extraction layers. Traditional Machine learning techniques uses feature extraction using techniques such as Scale Invariant Feature Transform (SIFT) [16], Speeded Up Robust Features (SURF) [17], and other methods [18].

Many of the notable and most successful works in medical image classification involved implementing deep learning algorithms such as Convolutional Neural Networks (CNN). Roth et al [19] used Computerized Tomography (CT) scan [15] results to identify and classify Lymph Node (LN) detection using CNN. They were able to obtain high classification accuracy compared to earlier approaches using boosting-based feature selection which contains many false positives [19]. As seen in [20], CNN with U-net algorithm was used for cell image segmentation and tracking. U-net, a 23-layer CNN, was introduced in 2015 by Olaf Ronneberger, Philipp Fischer, and Thomas Brox [20]. Lan et al [21] also implemented deep CNN using U-Net algorithm with the addition of Residual network to correctly identify and classify lung nodules. Residual network was introduced in 2015 by Kaiming He and Xiangyu Zhang [21]. Residual network is a deep network that uses shortcut technique to input result to bottom of the layer. This reduces the number of parameters and efficiency of the network [21]. In this project, we will be using pre-trained residual network of 50 layers deep to classify pneumonia.

Organ segmentation is another important domain of medical image analysis. Some of the organs in our body exhibits high variability anatomically, especially pancreas, liver, kidneys and heart. This makes medical image analysis a much more complex problem. As mentioned in [22], deep CNN was used for pancreatic segmentation from Computerized Tomography (CT) results. Accurate segmentation of organs is a crucial step in identifying abnormal growth or cancer in organs that exhibit high anatomical variability. Bier et al. [23] proposes a new method in detecting anatomical landmarks from X-ray images irrespective of the viewing direction. Based on the CNN model, they were able to identify 23 anatomical landmarks of pelvis from single x-ray. [22] proposed a bottom up approach, which includes dense labeling of image patches to entire organ. This method has high accuracy compared to previous segmentation of organs using random forests [22] [24]. In order to better implement a model that accurately identify lung abnormalities, correct segmentation of field of interest must be implemented. This is often a challenging task as images of lung field consist of rib cages that exhibits various bone densities, presence of clavicles, and in some cases lung field can be altered by the presence of certain lung abnormalities. [25] proposed an algorithm to segment the lungs using dynamic programming approach. This method used to extract thoracic cage boundary from several manually handcrafted boundaries. However, this approach is only been tested on smaller dataset and I am unsure on how the model performs on a larger scale. Additionally, this approach tends to fail when certain pulmonary abnormalities are present like in cases on pneumonectomy. A much better approach was implemented by

[26] using U-net based convolutional neural network. They have used bone shadow exclusion techniques in extracting the lung borders. This approach tends to perform better with chest radiographs that has very low or no bone shadow. Therefore, this technique calls for additional preprocessing algorithm for bone shadow elimination or bone suppression making the solution more complex. There exists other shallow learning based on machine learning methods such as k- nearest neighbors (KNN), linear discriminant analysis (LDA) and so on but these approaches tend to rely on manually segmented lung fields.

Some of the recent studies used open datasets from openI and Indiana network of patient care [8]. For these datasets no quantitative disease results are reported creating a major hurdle in accurate image classification. For this project we use one of the largest publicly available datasets released by the National Institute of Health (NIH) [8]. The dependent labels are obtained for this dataset by performing text mining on medical reports. In [8], a model that predicts all 14 labels by fine-tuning CNN using stochastic gradient descent algorithm [27] was studied. They implemented CNN using VGGNet-16, AlexNet [28], GoogLeNet [29], and ResNet-50 pre-trained models. Using weights from cafe model zoo, [8] were able to achieve a highest AUC of 63% on detecting pneumonia using ResNet-50 model. Additionally, they were able to localize the infected region of lungs using heat map. A deep learning model using attention guided CNN and was able to achieve a pneumonia detection rate of 78% [13]. This approach identifies potential region of interest by using attention guided mask inference process. This process eliminates the needs of bounding boxes that was used in implementing similar models. In [30], a 2 stage end-to-end Neural Network combining densely connected image encoder and a recurrent neural network decoder to obtain AUC of 71% were implemented. In most clinical cases, there are other pathological abnormalities that show up along with pneumonia. In [30], this problem was rectified by exploiting statistical dependencies between labels. Another notable work on predicting pneumonia was implemented by the Stanford machine learning group. They were able to model the problem using state of the art 121-layer dense neural network and obtain an AUC of 76% [12]. Their work was compared to the reports of radiologists and results exceeded the accuracy of practitioners.

III. Proposed Method

Images provided by NIH were already preprocessed in a resolution of 1024 x1024 pixels. Dataset containing the labels were also clean with no missing values. It's worth noting that the dataset contains only 1431 labels that indicate pneumonia. Since the label of interest is only 1.2% of the dataset, we attempted resampling the dataset to obtain a fair amount of positive and negative labels before splitting into training, validation and testing. Additionally, we attempted to augment the data with random horizontal flipping. The model was built in a sequential format. Table 1 shows the architecture of the fine-tuned model.

Output from a pre-trained model was fed into the fine-tuning portion of the CNN. Feature output from pre-trained model was modified using batch normalization and was fed into three convolutional layers. Then average pooling was performed and passed through two additional convolutional layers. Output of this layer was multiplied by batch normalized result from previous result. Next, we performed global average pooling twice, one on the result of convolutional layer and another on multiplication layer, to get the average value of pixels from area of interest. These two results were used to rescale by dividing one by another to obtain the input for next dropout layer. Then, two more dense layers were added with a dropout layer in between. In the last stage, we used a sigmoid activation function. Additionally, we used binary cross entropy as loss function, Adam [31] as optimizer, and built in accuracy metric to get the result as we train. For all three pre-trained models, we employed the same approach of fine-tuning. Figure 5 displays the network architecture of proposed (fine-tuned VGG16) model. For the modeling process, we used a small training and validation batch size of 10, primarily due to the resource exhaustion of the GPU. Epochs were limited to 20 as the accuracy of the model was found to be decreasing on higher values. The result of the model was used in plotting the changes in accuracy and loss for each epoch on both training and validation values. Finally, the predicted outcome was calculated, confusion metric was obtained and Receiver Operating Characteristic (ROC) curve was plotted [32]. Figure 6 shows the pseudocode used in training the proposed model.

IV. Experiments and Results

TABLE I: Experimental setup

Specifications	Details
Cloud Platform	Floydhub
Processor	Intel(R) Xeon(R) CPU E5-2650 @ 2.00GHz
Memory	61GB
Graphics	Tesla K80 - 12 GB Memory
Operating System	Ubuntu
Programming Language	Python 3.6.5
Dataset Source	National Institute of Health (NIH)
Dataset Name	Data Entry 2017
Dataset Size	42.1 GB

A. Techniques Used

In the preprocessing phase, we resized the image using Python Image Library (PIL) to 224 by 224 pixel to fit the need of pre-trained models and reduce the computational time. For the Inception v3 model, we reshaped the image to 299 by 299 to fit the Keras input requirements. Inception v3 is a 48-layer CNN. For this project, we implemented three models that are: VGG16, ResNet-50, and Inception v3. All models were pre-trained and weights obtained from ImageNet were used. We used pre-trained features to fine tune the last layers to suit our objective. All modeling was done using the Keras library [33] with a TensorFlow [34] in the backend. Due to the increased needs of computational power, we used Jupiter notebook available through the cloud-based platform called

Floydhub [35]. All computational task ran on Tesla K80 GPU with TensorFlow v. 1.5.

For this study we had to resample the dataset primarily due the low positive labels of Pneumonia. First, we obtained a more balanced dataset by stratifying based on patient gender and pneumonia. We obtained 1431 label that were positive for pneumonia. Remaining 110,689 labels were under the negative class. After obtaining a more balanced dataset, we resample again with replacement to obtain 2000 positively and 2000 negatively labeled training sample. Figure 5 shows used CNNs in proposed work. After sampling, images were augmented and finally reshaped based on the need of appropriate models, and modeling was performed.

B. Results

Figure 6 shows the confusion matrix result of pneumonia classification done on all three models. Inception v3 model showed a test accuracy of 53% in correctly classifying the labels. Meanwhile ResNet-50 showed much better test accuracy of 58%. Among all three models, the proposed (fine-tuned VGG16) demonstrates the most test accuracy of 75%. Among all three model, test sensitivity of the proposed model was found to be highest at 76% and specificity was found to be highest for ResNet-50 model at 73%. The sensitivity refers to the probability that a test result will be positive when the disease is present (true positive rate, expressed as a percentage), while the specificity denotes to the probability that a test result will be negative when the disease is not present (true negative rate, expressed as a percentage). Sensitivity also called as true positive rate, is the proportion of correct labels that were accurately identified to have pneumonia. Similarly, specificity also called as true negative rate, is the proportion of correct labels that were accurately identified to have no pneumonia. It needs to be noted that we give more importance to specificity than sensitivity. This is because our model aims to increase the work flow of doctors. Therefore, we want high rate of identifying pneumonia while missing only a few cases of actual pneumonia. Precision for Inception v3 model was found to be lowest at 86% and ResNet-50 model showed the highest precision of 92%. Precision tells us how often the prediction of pneumonia is correct.

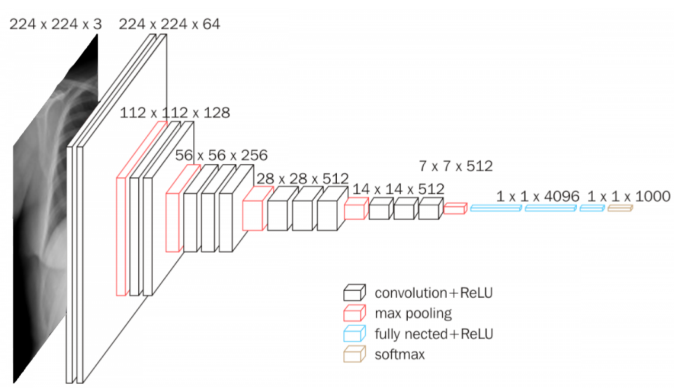


Fig. 5: The Proposed Architecture.

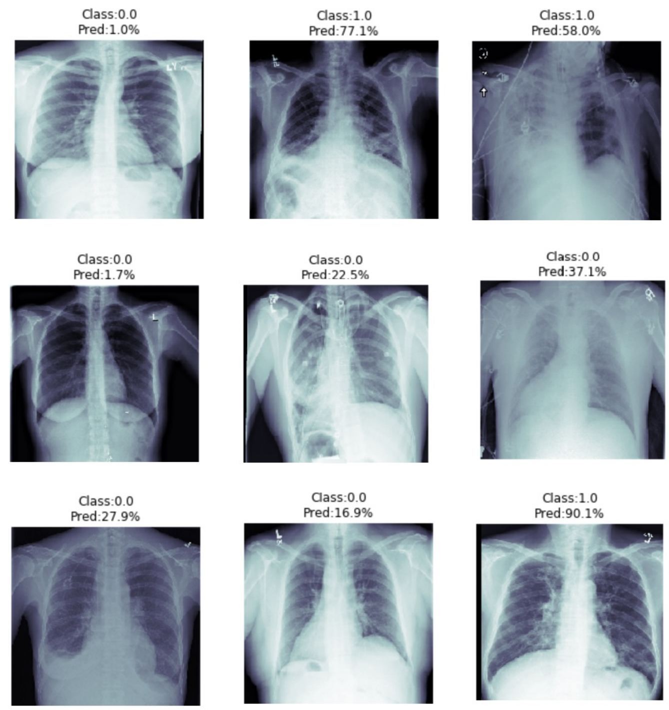


Fig. 6: Result of nine random chest radiograph from test dataset using Proposed model.

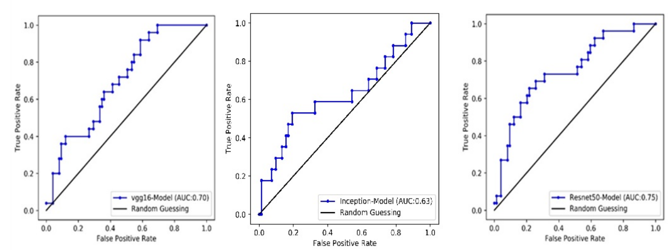


Fig. 7: Obtained ROC curve for different models.

F1- score was found to be comparably same for all three models and this was not included in the results figure. It measures the accuracy of the model. Figure 7 shows the ROC curve for all three models. It shows the trade-off between sensitivity and specificity. Area under the curve of ROC is a measure on the accuracy of the test. For Inception model displayed lowest area under the curve while ResNet-50 showed highest. Some of the models did exhibit some signs of overfitting during the modeling process. This is evident upon closely observing the validation and training results of various models in figure 6. Implementing same fine-tuning approach for all three model could be one reason behind overfitting. More hyper parameter optimization must be done in future to improve the accuracy of ResNet-50 and inception model. Figure 10 displays the output of proposed model on nine random chest radiographs from test dataset. Class indicates actual labeled class of the image and predicted value indicates predicted probability of pneumonia

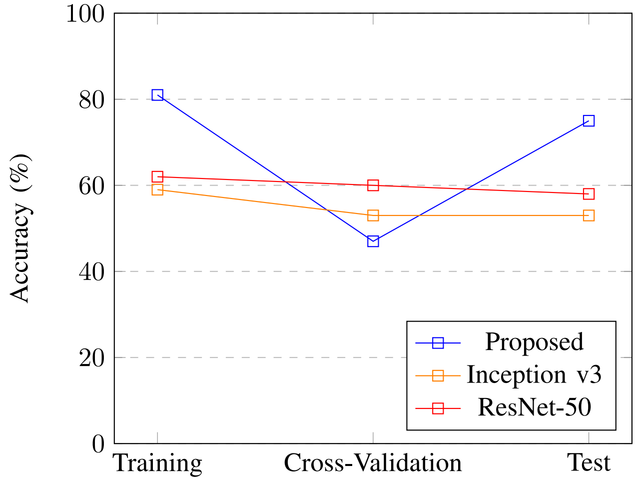


Fig. 8: Accuracy of Datasets using three models: Proposed, Inception v3, and ResNet-50

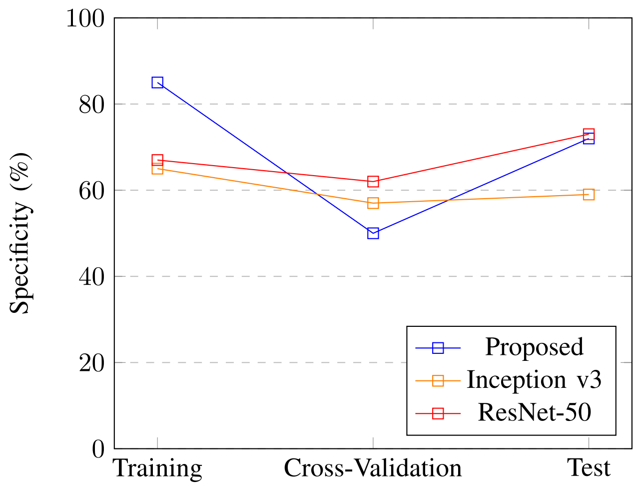


Fig. 9: Specificity of Datasets using three models: Proposed, Inception v3, and ResNet-50

by the model. We also compared the result with state-of-the-art results obtained on same dataset. Compared to results of similar implementation in recent literature VGG16 and ResNet-50 model showed better AUC as shown in the Figure 11. However, more novel approaches such as fine-tuning using dense 121-layer CNN model, and attention guided CNN model show more accurate results at the cost of complexity.

V. Conclusions

Pneumonia is an infectious disease that has tormented humanity for ages. Even with the significant advancement in technology, pneumonia is still listed among the top 10 causes of death in the world. Early and accurate interventions are critical for treating pneumonia. Chest x-ray is one of the cheapest and most widely used diagnostic tool in identifying pneumonia and other lung abnormalities. With the surprising advancements in new deep learning approaches, we have pro-

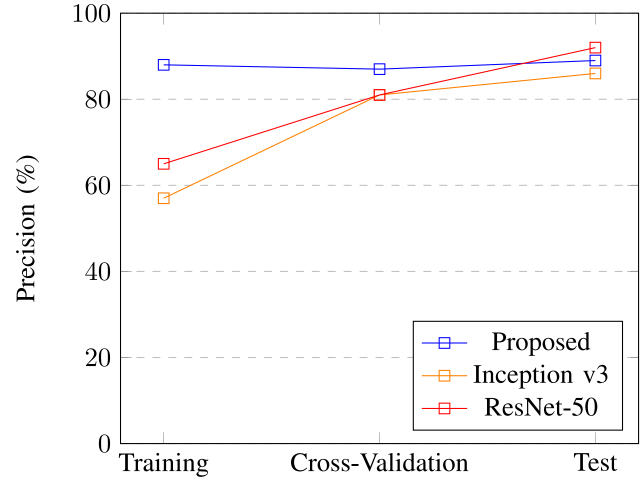


Fig. 10: Precision of Datasets using three models: Proposed, Inception v3, and ResNet-50

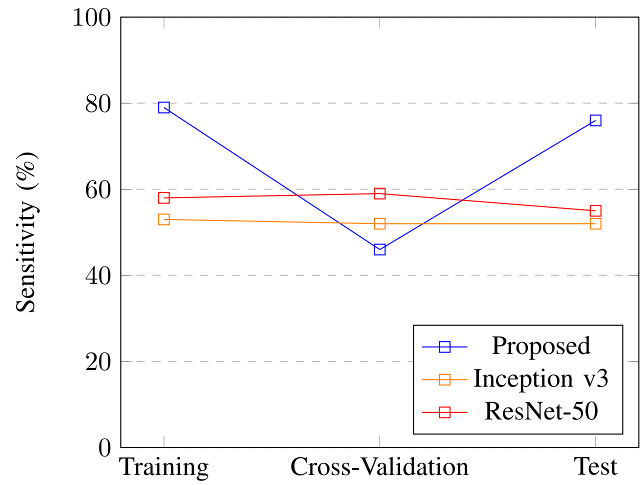


Fig. 11: Sensitivity of Datasets using three models: Proposed, Inception v3, and ResNet-50

posed a new tool that can help improve diagnostic accuracy of pulmonary abnormalities from chest radiograph. Throughout this project, we have implemented computer aided diagnostic of pneumonia using the most common deep learning approach known as convolutional neural network (CNN). In this model user will input the x-ray image and model will output the correct classification label for pneumonia with an average accuracy of (68%) and average specificity of (69%) in contrast to the current state-of-the-art accuracy that is (51%). All of the results were obtained for data sets provided by NIH [7]. Therefore, the proposed model is more efficient and accurate in contrast to the state-of-the-art models.

Acknowledgment

Fahad Saeed was partially supported by NSF CAREER OAC 1925960.

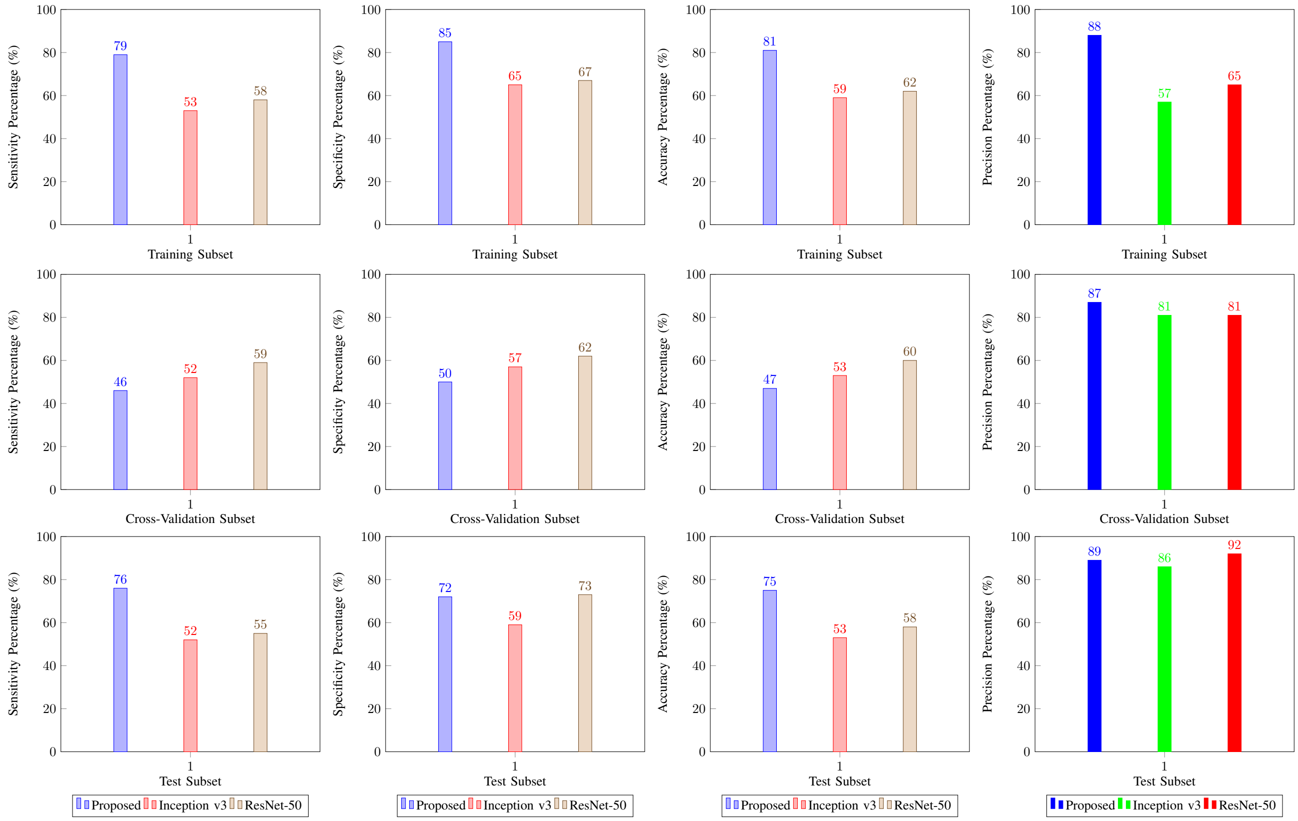


Fig. 12: Accuracy, Specificity, Precision, and Sensitivity of the three datasets: Training, Cross-Validation, and Test using Proposed, Inception v3, and ResNet-50 models

REFERENCES

[1] (2019, July) Preventing pneumonia. [Online]. Available: https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/pneumonia/preventing-pneumonia.html

[2] (2019, July) Pneumonia. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[3] D. E. Schraufnagel, *Breathing in America: diseases, progress, and hope*. American Thoracic Society, 2010.

[4] J. Hamborsky, A. Kroger, S. Wolfe, C. for Disease Control, Prevention *et al.*, *Epidemiology and prevention of vaccine-preventable diseases*. US Department of Health & Human Services, Centers for Disease Control and . . . , 2015.

[5] (2015, November) Trends in pneumonia and influenza morbidity and mortality. [Online]. Available: https://www.lung.org/assets/documents/research/pi-trend-report.pdf

[6] T. Jardins and G. Burton, "Clinical manifestations and assessment of respiratory disease," 2011.

[7] P. C. Wroe, J. A. Finkelstein, G. T. Ray, J. A. Linder, K. M. Johnson, S. Rifas-Shiman, M. R. Moore, and S. S. Huang, "Aging population and future burden of pneumococcal pneumonia in the united states," *The Journal of infectious diseases*, vol. 205, no. 10, pp. 1589–1592, 2012.

[8] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] 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.

[12] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[13] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *arXiv preprint arXiv:1801.09927*, 2018.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[15] S. L. Wellington, H. J. Vinegar *et al.*, "X-ray computerized tomography," *Journal of Petroleum Technology*, vol. 39, no. 08, pp. 885–898, 1987.

[16] D. G. Lowe *et al.*, "Object recognition from local scale-invariant features." in *iccv*, vol. 99, no. 2, 1999, pp. 1150–1157.

[17] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[18] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," *arXiv preprint arXiv:1803.01164*, 2018.

[19] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2014, pp. 520–527.

[20] 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. Springer, 2015, pp. 234–241.

[21] T. Lan, Y. Li, J. K. Murugi, Y. Ding, and Z. Qin, "Run: Residual u-net for computer-aided detection of pulmonary nodules without candidate selection," *arXiv preprint arXiv:1805.11856*, 2018.

[22] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564.

[23] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. Maier, "X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 55–63.

[24] A. Farag, L. Lu, E. Turkbey, J. Liu, and R. M. Summers, "A bottom-up approach for automatic pancreas segmentation in abdominal ct scans," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Springer, 2014, pp. 103–113.

[25] G. Horváth, G. Orbán, Á. Horváth, G. Simkó, B. Pataki, P. Máday, and S. Juhász, "A cad system for screening x-ray chest radiography," in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*. Springer, 2009, pp. 210–213.

[26] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*. Springer, 2018, pp. 638–647.

[27] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[28] 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.

[29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[30] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv preprint arXiv:1710.10501*, 2017.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] S. H. Park, J. M. Goo, and C.-H. Jo, "Receiver operating characteristic (roc) curve: practical review for radiologists," *Korean Journal of Radiology*, vol. 5, no. 1, pp. 11–18, 2004.

[33] (2019, July) Keras: The python deep learning library. [Online]. Available: https://keras.io

[34] (2019, July) Tensorflow. [Online]. Available: https://www.tensorflow.org

[35] (2019, July) Floydhub. [Online]. Available: https://www.floydhub.com