Pneumonia Detection from X-ray Images

ADVANCED MACHINE LEARNING (INT248)

Sujeet Maurya (RKM042B41)
Computer Science Department
Lovely Professional University
Jalandhar , Punjab.

Abstract:

Pneumonia causes the death of around 700,000 children every year and affects 7% of the global population. Chest X-rays are primarily used for the diagnosis of this disease. However, even for a trained radiologist, it is a challenging task to examine chest X-rays. There is a need to improve the diagnosis accuracy. In this work, an efficient model for the detection of pneumonia trained on digital chest X-ray images is proposed, which could aid the radiologists in their decision making process. A novel approach based on a weighted classifier is introduced, which combines the weighted predictions from the state-of-the-art deep learning models such as ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3 in an optimal way. This approach is a supervised learning approach in which the network predicts the result based on the quality of the dataset used. Transfer learning is used to fine-tune the deep learning models to obtain higher training and validation accuracy. Partial data augmentation techniques are employed to increase the training dataset in a balanced way. The proposed weighted classifier is able to outperform all the individual models. Finally, the model is evaluated, not only in terms of test accuracy, but also in the AUC score. The final proposed weighted classifier model is able to achieve a test accuracy of 98.43% and an AUC score of 99.76 on the unseen data from the Guangzhou Women and Children's Medical Center pneumonia dataset. Hence, the proposed model can be used for a quick diagnosis of pneumonia and can aid the radiologists in the diagnosis process.

Keywords: pneumonia, chest X-ray images, convolution neural network (CNN), deep learning, transfer learning, computer-aided diagnostics

Introduction:

Pneumonia is an acute respiratory infection that affects the lungs. It is a fatal illness in which the air sacs get filled with pus and other liquid solid There are mainly two types of pneumonia: bacterial and viral. Generally, it is observed that bacterial pneumonia causes more acute symptoms. The most significant difference between bacterial and viral pneumonia is the treatment. Treatment of bacterial pneumonia is done using antibiotic therapy, while viral pneumonia will usually get better on its own.

It is a prevalent disease all across the globe. Its principal cause includes a high level of pollution. Pneumonia is ranked eighth in the list of the top 10 causes of death in the United States. Due to pneumonia, every year, 3.7 lakh children die in India, which constitutes a total of fifty percent of the pneumonia deaths that occur in India. The disease frequently goes overlooked and untreated until it has reached a fatal point, especially in the case of old patients. It is the single largest cause of death in children (especially under the age of five) worldwide. According to the WHO, "Every year, it kills an estimated 1.4 million children under the age of five years, accounting for 18% of all deaths of children under five years old worldwide. Pneumonia affects children and families everywhere but is most prevalent in South Asia and sub-Saharan Africa. Children can be protected from pneumonia. It can be prevented with simple interventions and treated with low-cost, low-tech medication and care. Therefore, there is an urgent need to do research and development on computer-aided diagnosis so that the pneumonia-related mortality, especially in children, can be reduced.
One of the following tests can be done for pneumonia diagnosis: chest X-rays, CT of the lungs, ultrasound of the chest, needle biopsy of the lung, and MRI of the chest. Currently, chest X-rays are one of the best methods for the detection of pneumonia . X-ray imaging is preferred over CT imaging because CT imaging typically takes considerably more time than X-ray imaging, and sufficient high-quality CT scanners may not be available in many underdeveloped regions. In contrast, X-rays are the most common and widely available diagnostic imaging technique, playing a crucial role in clinical care and epidemiological studies . There are several regions across the globe where there is a

scarce availability of practiced healthcare workers and radiologists whose prediction on such diseases matter greatly. Computer-aided diagnosis using artificial intelligence based solutions is becoming increasingly popular these days

This facility can be made available to a large population at a minimal cost. Another issue with this disease is that sometimes, the features that describe the very existence of the disease often get mixed with other diseases, and hence, radiologists find it challenging to diagnose this disease. Deep learning techniques solve all these problems, and their accuracy in the prediction of the disease is the same and sometimes even greater than an average radiologist. Among the deep learning techniques, convolutional neural networks (CNNs) have shown great promise in image classification and segmentation and therefore are widely adopted by the research community. Biomedical image diagnosis that uses the techniques of deep learning and computer vision has proven to be very helpful to provide a quick and accurate diagnosis of the disease that matches the accuracy of a reliable radiologist. Currently, deep learning based methods cannot replace trained clinicians in medical diagnosis, and they aim to supplement clinical decision making. In this paper, a model is presented based on the applications of deep learning and convolutional neural networks that are capable of classifying automatically that the patient has pneumonia or not. The proposed methodology uses a deep transfer learning algorithm that extracts the features from the X-ray image that describes the presence of disease automatically and reports whether it is a case of pneumonia.

Related Work:

Deep learning based methods are already being used in various fields. Different authors have already proposed several biomedical image detection techniques discussed the challenges and the future of medical image processing. Much work has already been done for the detection of numerous diseases by using deep learning based techniques, as stated by Dinggang Shen Andre presented a deep learning model for dermatologist-level classification of skin cancer, and also proposed a methodology for the depiction of prostrate in MRI volumes using CNN. used the technique of deep learning for brain hemorrhage detection in CT scans, and proposed a method for detecting diabetic retinopathy in retinal fundus photographs. Y. Bar also discussed chest pathology detection by the techniques based on deep learning. Methods regarding the examination of the detection of disease by chest X-ray have also been worked on earlier by performing various examination techniques. The chest X-ray images are passed through the evaluation process of scan line optimization such that it eliminates all the other body parts to avoid any error in diagnosis. The algorithm used two deep three-dimensional (3D) customized mixed link network (architectures for lung nodule detection and classification. combined DenseNet and long-short term memory networks (LSTM) to exploit the dependencies between abnormalities. Several authors also have worked on pneumonia classification. proposed the use of EMD (earth mover's distance) to identify infected pneumonia lungs from normal non-infected lungs.A CNN model for pneumonia classification. Some researchers have shown assuring results such as Cohenn and Rajaraman . The performance of customized CNNs to detect pneumonia and further differentiate between bacterial and viral types in pediatric CXRs . A region based convolutional neural network for segmenting the pulmonary images along with image augmentation for pneumonia identification.neural networks with data augmentation and without any pre-training to obtain an area under the curve (AUC) of 0.94–0.95. used CheXNeXt, a very deep CNN with 121 layers, to detect 14 different pathologies, including pneumonia, in frontal-view chest X-rays. A localization approach based on pre-trained DenseNet-121, along with feature extraction, was used to identify proposed a novel multi-scale heterogeneous three-dimensional (3D) convolutional neural network (MSH-CNN) based on chest computed tomography (CT) images. used a hierarchical convolutional neural network (CNN) structure and a novel loss function, sin-loss, for pneumonia detection. used Mask-RCNN, utilizing both global and local features for pulmonary image segmentation, with dropout and L2 regularization, for pneumonia identification. Jung used a 3D deep CNN (3D DCNN), which had shortcut connections.combined the outputs of different neural networks and reached the final prediction using majority voting. None of the above-mentioned approaches except that of tried to combine predictions from different neural networks.

The main contribution is a weighted classifier that integrates five deep learning models. The weights for each model are based on each model's performance on the testing dataset.

Background of Deep Learning Methods:

3.1. Convolutional Neural Network:

LeCun first used CNN, in 1989, for handwritten zip code recognition. This is a type of feed-forward network. The main advantage of CNN compared to its predecessors is that it is capable of detecting the relevant features without any human supervision.

A series of convolution and pooling operations is performed on the input image, which is followed by a single or multiple fully connected layers, as shown in Figure 1. The output layer depends on the operations being performed. For multiclass classification, the output layer is a softmax layer. The main disadvantage with deeper CNNs is vanishing gradients.

3.2. Transfer Learning:

In transfer learning, a model that is trained for a particular task is employed as the starting point for solving another task. Therefore, in transfer learning, pre-trained models are used as the starting point for some specific tasks, instead of going through the long process of training with randomly initialized weights. Hence, it helps with saving the substantial computer resources needed to develop neural network models to solve these problems.

Domain, task, and marginal probabilities to propose a framework for better understanding the transfer learning. The domain D was defined as a two-element tuple consisting of the feature space, χ, with a marginal probability, P(X), where X is a sample data point. Hence, mathematically, domain D can be defined as,

D= {χ,P(X)}

1.Here, χ is the space of all term vectors, xi is the ith term vector corresponding to some documents, and X is a particular learning sample (X=x1,⋯,xn,∈χ).

For a given domain D, the given task T is defined as:

T={γ,P(Y|X)}={γ,η},Y={y1,⋯,yn},yi∈γ

2. where γ is the label space. η is a predictive function learned from the feature vector/label pairs (xi,yi), where xi∈χ and yi∈γ.

η(xi)=yi

3.Here, η predicts a label for each feature vector.

Due to the lack of a sufficient dataset, training a deep learning based model for medical diagnosis related problems is computationally expensive, and the results achieved are also not up to the mark. Hence, pre-trained deep learning models, which were previously trained on ImageNet dataset, were used in this paper. Further, all these pre-trained models were fine-tuned for pneumonia classification. All the layers of the architectures used were trainable. Further details, related to fine-tuning.

3.3. Performance Metrics for Classification:

All the models were tested on the test dataset after the completion of the training phase. Their performance was validated using the accuracy, recall, precision, F1, and area under the curve (AUC) score. All the performance metrics used in this paper are discussed below.

In the below-mentioned definitions and equations, while classifying healthy and pneumonia patients, true positive (TP) denotes the number of pneumonia images identified as pneumonia, true negative (TN) denotes the number of normal images identified as normal (healthy), false positive (FP) denotes the number of normal images incorrectly identified as pneumonia images, and false negative (FN) denotes the number of pneumonia images incorrectly identified as normal.

Accuracy: It tells us how close the measured value is to a known value.

Accuracy=(TP+FN)/(TP+TF+FP+FN)

Precision: It tells about how accurate the model is in terms of those which were predicted positive.

Precision=TP/(TP+FP)

Recall: It calculates the number of actual positives the model was able to capture after labeling it as positive (true positive).

Recall=TP/(TP+FN)

F1: It gives a balance between precision and recall.

F1=2×(Precision×Recall)/(Precision+Recall)

AUC Score and ROC Curve: ROC (receiver operating characteristics) is a probability curve, and AUC (area under curve) represents the degree of separability. The ROC curve is the plot of sensitivity (true positive rate) against specificity (false positive rate).

Object storage is the storage and retrieval of unstructured blobs of data and metadata using an HTTP API. Instead of breaking files down into blocks to store it on disk using a file system, block storage deal with whole objects stored over the network.These objects are in any form may be it could be an image file,logs,HTML files, or any self-contained blob of bytes. If you have studied database systems you are familiar about “Schema”. Object storage are unstructured because there is no specific schema or format they need to follow.

4. Proposed Methodology:

An optimum solution for the detection of pneumonia from chest X-rays is proposed in this paper. Data augmentation was used to address the problem of the limited dataset, and then, state-of-the-art deep learning models, as were fine-tuned for pneumonia classification. Then, predictions from these models were combined, using a weighted classifier to compute the final prediction.

4.1DataPreprocessingand Augmentation

Each image had to be preprocessed according to the deep neural network used. There were two important steps involved: resizing and normalization. Different neural networks require images of different sizes according to their defined architecture. ResNet18, DenseNet121, and MobileNetV2 expect images of size 224 × 224, while InceptionV3 and Xception require images of size 229 × 229. All the images were also normalized according to the respective architectures.

Adequate training of a neural net requires big data. With less data availability, parameters are undermined, and learned networks generalize poorly. Data augmentation solves this problem by utilizing existing data more efficiently. It aids in increasing the size of the existing training dataset and helps the model not to overfit this dataset. In this case, there were a total of 1283 images of the normal (healthy) case and 3873 images of the pneumonia case in the training dataset. Out of these, four-hundred images were reserved for optimizing the weighted classifier. This dataset was highly imbalanced. There were already enough images in the pneumonia case. Therefore, each image of only the normal (healthy) case was augmented twice. Finally, after augmentation, there were 3399 healthy chest X-ray images and 3623 pneumonia chest X-ray images.

4.2 Weighted Classifier

In this module of the proposed methodology, a weight (Wk) corresponding to each model was estimated. Wk can be defined as the belief in the kth model, with k being equal to 5 as 5 pre-trained models were used in this paper. Wk has values between 0 and 1, and the sum of all weights i. Each model, after it was fine-tuned, returned the probabilities for each class label, i.e., 2 classes in the form of a matrix (Pk). A weighted sum of all these predictions arrays was calculated .

P1W1+P2W2+P3W3+⋯+PkWk=Pf

W1+W2+W3+⋯+Wk=1

Loss=−1/N*(∑i=1ny×log(p)+(1−y)×log(1−p))

Pk is the prediction matrix, with shape: number of optimization images * class labels (400*2), corresponding to each architecture. In Equation (8), the contribution of each model is weighted by a coefficient (Wk), which indicates the trust in the model. First, we obtained the Pk for every model for an unseen image set (400 images). The maximum iterations for differential evolution algorithms were kept to be 1000. With the help of Pf, the prediction of a class label could be computed. Classification loss corresponding to this Pf was reduced N denotes the size of the image set (400) and p denotes the probability that the given image is pneumonia infected.

4.3 Class Activation Maps

Class activation maps (CAMs) can help in demystifying the results of deep learning models. Traditionally, deep learning based methods are considered to be a black-box approach. For clinical decision making, it is necessary that the results of the deep learning model can be interpreted. CAMs can help in identifying the parts of the image on which the model was focusing while making the final prediction and hence can provide insights into the working of the model. Such an analysis can further help in hyperparameter tuning and gain understanding of the reason behind the failure of the model. For obtaining the class activation map, the network needed to be trained with the global average pooling (GAP) layer. After the GAP layer, a fully connected network was maintained, which was followed by the softmax layer, providing the class, such as pneumonia, CAMS class activation maps were generated for both bacterial and viral pneumonia for all the fine-tuned model and are discussed in detail in the results section.

5. Experimental Results

In this section, the experiments and evaluation techniques used in the paper to test the efficiency of the proposed model are presented. The chest X-ray image dataset, . The Keras open-source deep learning framework with TensorFlow as the backend was used, first to load the pre-trained architectures on the ImageNet Dataset and then fine-tune them for the task at hand. All the computation work was done on a Standard PC with8 GB RAM, NVIDIA GeForce GTX 1060 6 GB GPU, and Intel i7, seventh-generation processor.

5.1. Result in Terms of Testing Accuracy and Testing Loss

To test and evaluate the performance of the proposed network, each experiment was conducted five times. Parameters and hyperparameters were tuned during the training. the training accuracy and training loss curves obtained while training the models for 25 epochs. The training accuracy for all the models exceeded 99%, and the training loss for all the models was below 0.03. Except for Xception, all the other models had similar training accuracy and training loss curves. summarizes the testing accuracy and testing loss for different networks and the final weighted classifier. DenseNet121 was able to attain the maximum testing accuracy and the minimum testing loss. Initially, all the weights of the weighted classifier were kept equal (W1=W2⋯W5=0.20). Hence, every model contributed equally towards the final prediction. A test accuracy of 97.45 and a loss of 0.087 was obtained. Then, the optimum weights were estimated for every model. With these weights, the final weighted classifier was able to achieve a testing accuracy of 98.43, and the testing loss was 0.062.

All the test images were pre-processed similarly as the training images and hence had the same size as required by the respective architecture. The test images were of size 224 × 224 for ResNet18, DenseNet121, and MobileNetV2, while for InceptionV3 and Xception, they were of size 229 × 229. The testing was also done on the same system on which training was done. The average inference time for all the models was 0.045 s (while the GPU was used), and for the weighted classifier, it was 0.203 s.

5.2. Performance Analysis

How's the ROC curves for different architectures and the proposed classifier. The maximum AUC score (99.76) was achieved by the proposed classifier. All the models had a similar AUC/ROC curve. After analyzing the results, it can be said the weighted classifier gave the best results with an AUC score of 99.75, F1 score of 98.63, and test accuracy of 98.43. Hence, the proposed weighted classifier was able to combine the predictions from all the individual architectures in an optimum manner. The differences in the performance of other models were not significant. This might be because all the models used in this paper were deep learning based and were fine-tuned on the same insufficient dataset.

The activation maps were plotted for every individual network. These activation maps helped in localizing areas in the image most indicative of pneumonia. The activation maps were obtained for the last convolutional layer of each network. In the case of bacterial pneumonia, all the networks detected the abnormal lung to predict the presence of pneumonia correctly. Viral pneumonia manifested with a more diffuse "interstitial" pattern in both lungs, which was detected by all the fine-tuned architectures

5.3. Comparative Analysis of Various Existing Methods

The accuracy of various existing methods and the proposed methodology were compared. All the results used CNN and achieved a validation accuracy of 92.4%. The test dataset used was smaller compared to this paper. achieved a validation accuracy of 93.73% with their own CNN model. No other metric was published in either of these works. used a model based on DenseNet-121. They reported an AUC score of 98.4%. Unfortunately, the other metrics were not reported in the paper. I used customized CNNs to detect pneumonia and reported a test accuracy of 96.2%. I combined features from different deep learning models for pneumonia classification and achieved an accuracy of 96.84%. I combined the outputs of different neural networks and reached the final prediction using majority voting. They achieved an AUC score of 99.34. I used deep learning based methods and achieved an accuracy of 94.4%, 84.5%, and 98.0%, respectively. In all of these papers, the dataset used was of a similar size. All the studies other than I used image augmentation techniques.

6. Discussion

The high test accuracy (98.43) and AUC score (99.76) showed that the proposed method could be used as a supplement in clinical decision making. It can only aid the radiologists in the decision making process; the final decision has to be made by an expert. The proposed weighted classifier, with optimum weights, showed an improvement of 0.98%, in terms of the testing accuracy, over the case in which equal weights were assigned to every model. The false positives were greater than the false negatives, and hence, the classification error of pneumonia suffering patients as healthy was comparatively lesser, which is ideally required in medical diagnosis. Further, the activation maps plotted in this paper showed that the deep learning based models used were able to identify pneumonia affected regions in the chest X-rays. When compared to DenseNet121, the proposed weighted classifier showed an improvement of 0.43% in terms of testing accuracy, which in the real world on a large test dataset would be a significant number.

One of the limitations of this approach was the scarcity of available data. Usually, deep learning models are trained over thousands of images. Training deep neural networks with limited data might lead to overfitting and restricts the models' generalization ability. Unlike large datasets like ImageNet, the variability in the chest X-ray data was several orders of magnitude smaller. The performance of the proposed methodology would only increase with the availability of more data.

Another limitation was that the results of the deep learning models could not be properly explained. A deep understanding of the radiological features visible in chest X-rays is required for the diagnosis of the disease from the X-rays. The proper explanation of the final prediction of the model is also required, and this is one of the drawbacks of the deep learning based models. To this end, the activation maps were plotted, but further work is required. In the future, with better annotated datasets available, deep learning based methods might be able to solve this problem.

Conclusion:

Pneumonia constitutes a significant cause of morbidity and mortality. It accounts for a considerable number of adult hospital admissions, and a significant number of those patients ultimately die (with a mortality rate of 24.8% for patients over 75 years). According to the WHO, pneumonia can be prevented with a simple intervention and early diagnosis and treatment. Nevertheless, the majority of the global population lacks access to radiology diagnostics . Even when there is the availability of imaging equipment, there is a shortage of experts who can examine X-rays. Through this paper, the automatic detection of pneumonia in chest X-ray images using deep transfer learning techniques was proposed. The deep networks, which were used in our methodology, had more complex structures, but fewer parameters and, hence, required less computation power, but achieved higher accuracy. Transfer learning and data augmentation were used to solve the problem of over fitting, which is seen when there is insufficient training data, as in the case of medical image processing. Further, to combine different architectures efficiently, a weighted classifier was proposed. The experiments were performed, and the different scores obtained, such as the accuracy, recall, precision, and AUC score, proved the robustness of the model. The proposed model was able to achieve an accuracy of 98.857%, and further, a high F1 score of 99.002 and AUC score of 99.809 affirmed the efficacy of the proposed model. Though many methods have been developed to work on this dataset, the proposed methodology achieved better results. In the future, it would be interesting to see approaches in which the weights corresponding to different models can be estimated more efficiently and a model that takes into account the patient's history while making predictions.

REFERENCES

[1] Horváth, G., et al.: A CAD system for screening X-ray Chest radiography. In: Dössel O., Schlegel, W.C. (eds.) World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings, Munich, Germany, vol. 25(5). Springer, Berlin, Heidelberg, 7–12 Sept 2009

Google Scholar

[2] Van Ginneken, B., Ter Haar Romeny, B.M., Viergever, M.A.: Computer-aided diagnosis in chest radiography: a survey. IEEE Trans. Med. Imaging 20(12), 1228–1241 (2001)

CrossRefGoogle Scholar

[3] Rubio, Y., Montiel, O., Sepúlveda, R.: Microcalcification detection in mammograms based on fuzzy logic and cellular automata. In: Melin, P., Castillo, O., Kacprzyk, J. (eds.) Nature-Inspired Design of Hybrid Intelligent Systems. Studies in Computational Intelligence, vol. 667. Springer, Cham (2017)

Google Scholar

[4] Pneumonia – National Center for Health Statistics – Center for Disease Control and Prevention, https://www.cdc.gov/nchs/fastats/pneumonia.htm, last accessed 2018/09/03

[5] González, C.I., Melin, P., Castro, J.R., Castillo, O., Mendoza, O.: Optimization of interval type-2 fuzzy systems for image edge detection. Appl. Soft Comput. 47, 631–643 (2016)

CrossRefGoogle Scholar

[6]

Melin, P., Sánchez, D.: Multi-objective optimization for modular granular neural networks applied to pattern recognition. Inf. Sci. 460–461, 594–610 (2018)

MathSciNetCrossRefGoogle Scholar

[7] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M., Ng, A.Y.: CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning (2017)

GoogleScholar

[8]. Johns Hopkins Medicine Pneumonia. [(accessed on 31 December 2019)]; Available online:

https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia.

[9]. Johnson S., Wells D., Healthline Viral Pneumonia: Symptoms, Risk Factors, and More. [(accessed on 31 December 2019)]; Available online: https://www.healthline.com/health/viral-pneumonia.

[10]. Healthcare, University of Utah. Pneumonia Makes List for Top 10 Causes of Death. [(accessed on 31 December 2019)];2016 Available online: https://healthcare.utah.edu/the-scope/shows.php?shows=0_riw4wti7.

[11]. WHO Pneumonia is the Leading Cause of Death in Children. [(accessed on 31 December 2019)];2011 Available online: https://www.who.int/maternal_child_adolescent/news_events/news/2011/pneumonia/en.

[12]. Rudan I., Tomaskovic L., Boschi-Pinto C., Campbell H. Global estimate of the incidence of clinical pneumonia among children under five years of age. Bull. World Health Organ. 2004;82:895–903. [PMC free article] [PubMed] [Google Schola