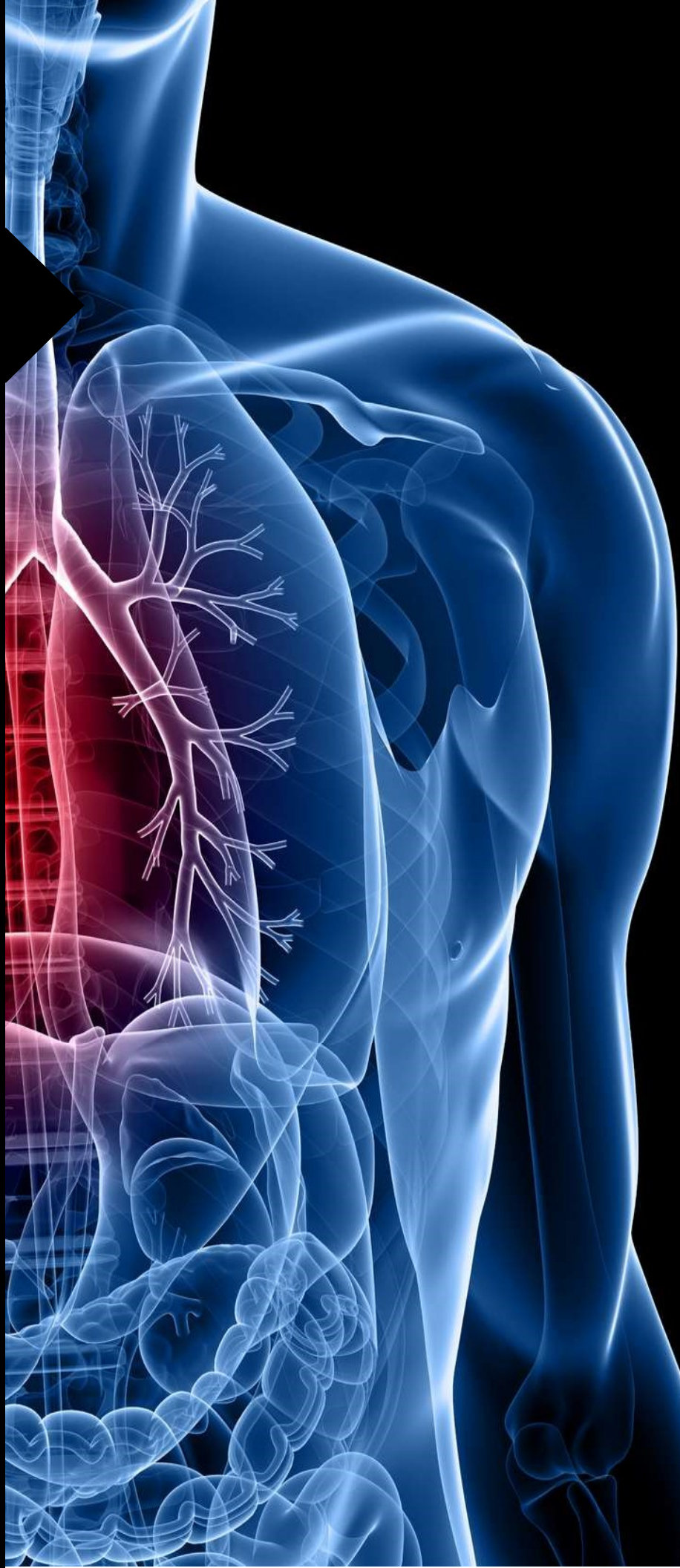
PATTERN
COMPUTER®

Computer-Assisted Detection and Diagnosis of Pediatric Pneumonia in Chest X-ray Images

Irshad Mohammed,[1] Nidhi Singh,[1,2] Meenakshi Venkatasubramanian[1,2]

[1]Pattern Computer Inc., 38 Yew Lane, Friday Harbor, WA 98250.
[2]Molecular Ecosystems Biology Department, Biosciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.



Pattern Computer Inc.

Pattern Computer Inc.
38 Yew Lane, Friday Harbor, WA 98250
USA

Computer-Assisted Detection and Diagnosis of Pediatric Pneumonia in Chest X-ray Images

In this paper, we demonstrate how PCI's deep learning-based algorithm allows detection and classification of pathological patterns in digital images for differential diagnosis of patients to strengthen the capacity of the health system.

Abbreviations:
ML: Machine Learning; DL: Deep Learning; AI: Artificial Intelligence; CNN: Convolutional Neural Networks; CAP: Community-Acquired Pneumonia; CXR: Chest X-ray; DCNN: Deep Convolutional Neural Network

Keywords:
Artificial Intelligence, Machine Learning, Computer Vision, Deep Convoluted Neural Network, Pediatric Pneumonia, Computer-Assisted Diagnosis, Localization

Abstract

Machine learning (ML) is a rapidly evolving research field attracting increasing attention in the medical imaging community. In radiology, ML aims at training computers to recognize patterns in medical images and to support diagnosis by linking these patterns to clinical parameters such as treatment or outcome. Conventional methods require extensive pre-processing, segmentation and manual extraction of specific visual features before classification. In contrast, ML approaches, such as deep learning, have exceeded human performance in visual tasks by utilizing automated hierarchical feature extraction and classification by multiple layers, to detect distinct patterns in high-dimensional datasets allowing quantification of the extent of the disease and the prediction of the course the disease will take. In this study, we present a deep learning-based algorithm that can accurately detect community-acquired pneumonia (CAP) in children and furthermore, distinguish its etiology in chest X-ray images. The results show that our model achieves an area under the receiver operating characteristic curve (AUROC) of 0.992 for normal vs. diseased classification and 0.976 for viral vs. bacterial classification, demonstrating the promising classification performance. We also report high recall (97.4% for normal vs. diseased; 98.3% for viral vs. bacterial) and the precision (97.2% for normal vs. diseased and 90.9% for viral vs. bacterial) for detecting CAP. Furthermore, our algorithm can localize areas that are most indicative of infection without using any localization data to train the system. Notably, our model outperformed state-of-the-art classifiers, in terms of all performance metrics across the classification tasks. Further research is necessary to determine the feasibility of applying this algorithm in a clinical setting and to determine whether the use of the algorithm could lead to improved care and outcomes compared with current assessment of pediatric pneumonia.

Introduction

Machine learning is a sub-discipline of artificial intelligence (AI) that is swiftly moving from an experimental phase to an implementation phase in many fields, including biomedical sciences. Recent progress in AI research, supported by enormous amounts of digital data and modern, powerful computational hardware has created major performance breakthroughs in the development of

ML applications. In the last 7 years, deep learning has delivered rapidly improving performance in image recognition, caption generation, and speech recognition. Early implementation of these techniques has a tremendous potential to influence radiology practice and research. In fact, integrating AI into diagnostic radiology has the potential to revolutionize disease diagnosis and management by performing many visual tasks – e.g., object detection, classification, and localization - by rapidly processing massive amounts of images which is difficult for humans [1]. Within medical imaging, deep learning has already shown immense potential, such as predicting the severity of diabetic retinopathy from retinal fundus images [2], classifying skin lesions [3], and analyzing histopathology [4,5]. Due to these successes, it is anticipated that the adoption of ML in radiology over the next decade will significantly improve the quality, value, and depth of radiology's contribution to patient care and population health. This will in turn revolutionize current clinical workflows and will help prevent diagnostic errors and reduce missed billing opportunities, thus enabling sustained productivity increases.

Community-acquired Pneumonia (CAP) is the most common infectious cause of morbidity and mortality among young children globally. Recent estimates suggest that approximately 120 million new cases of CAP occur each year with almost 1 million deaths among children under 5 years of age [6,7]. Establishing the causal agent of pneumonia (bacteria vs viruses) is essential to guarantee the most appropriate and effective therapy since each agent requires a very different form of management. While children suffering from viral pneumonia can be easily treated with supportive care, bacterial pneumonia requires urgent referral for immediate antibiotic treatment. However, the etiological diagnosis of pneumonia in children is difficult due to the low sensitivity of microbiological tests, and the low specificity of clinical signs and symptoms [8].

Chest radiograph or X-ray (CXR) is the preferred and most frequently used form of medical imaging to diagnose and differentiate between the different types of pneumonia [9]. Interpretation of CXR, however, remains a non-trivial task requiring both experience and expertise, as anatomic structures can overlap in a single two-dimensional image, and various physiological and pathological changes may appear similar or a single pathology may exhibit various features [10]. Thus, interpretation is prone to errors, with an earlier study reporting that 22% of all errors in diagnostic radiology were made in CXRs [11]. Compounding this difficulty is an increase in the number of examinations at a rate faster than the increase in the number of qualified radiologists, which has led to an increased workload for radiologists [12]. Therefore, there is an urgent need to develop efficacious, automated and cost-effective methods for accurate and timely determination of the causative agent especially in children. This is not only crucial to facilitating appropriate treatment and limiting the overuse of antibiotics but is also critical to preventing disease progression and complications including death.

In this study, we present a pattern-recognition algorithm based on a deep-learning framework that is capable of performing accurate detection and diagnosis of pediatric pneumonia in chest radiography which can provide clinicians with a reliable "second opinion". Furthermore, since visual evidence - such as spatial localization of abnormal regions - is a vital part of clinical diagnosis, this algorithm identifies regions of radiographs that are highly likely to contain pathology. We believe our AI algorithm when seamlessly integrated within the imaging workflow may provide pre-screened images and identified features for accurate and timely detection of pediatric pneumonia. This will be helpful to non-skilled technicians and trained radiologists alike, as this will allow them to work more efficiently against the rising pressure of increased workload, thus providing timely, efficient, and patient-centered delivery of quality health care.

Materials and Methods

Dataset and Image Pre-processing

For the purpose of model development, we used a set of pediatric CXRs that have been made publicly available by the authors of Ref [13]. The dataset includes anteroposterior CXRs of children from 1 to 5 years of age collected from Guangzhou Women and Children's Medical Center in Guangzhou, China. The characteristics of the data and its distribution are summarized in Table 1.

Before building deep learning models, the data is pre-processed to standardize the resolution. Higher resolution images dramatically increase computational time and cost, due to excessive number of weights to learn, increasing depth (number of network levels), and consequently, convolutions per layer. Accordingly, we resized the input images

to a 299 X 299 matrix while avoiding loss of visual features that distinguish one class from the other. We also used data augmentation techniques that permit the network to learn various alterations of each sample to provide true invariance. The prime data augmentation techniques include, but are not limited to, zooming, rotating/flipping, adjusting brightness, shearing, or offsetting.

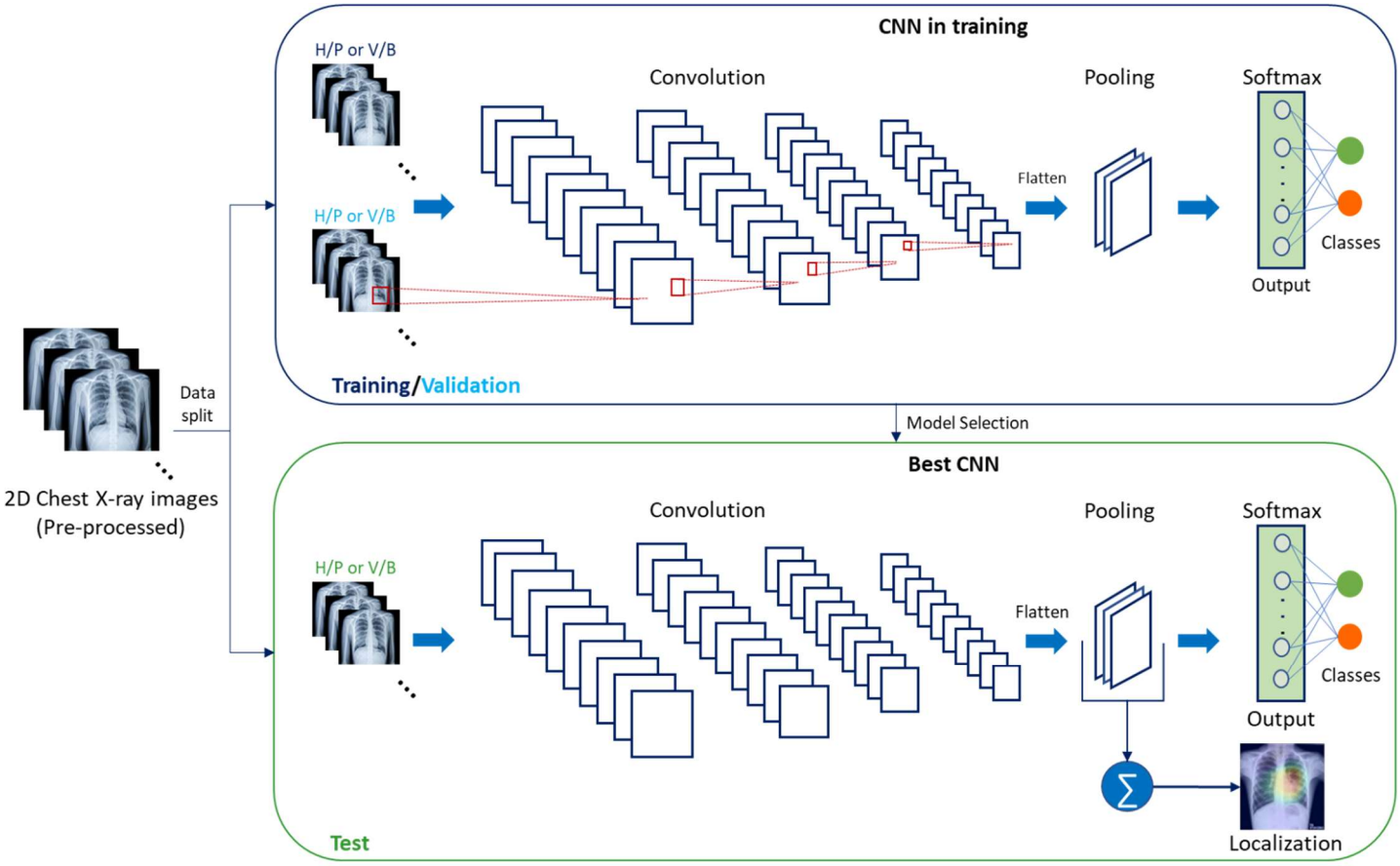
Table 1. The dataset and its characteristics.

Category	Training Samples	Test Samples
Normal vs. Pneumonia	5216	624
Bacterial vs. Viral	3875	390

Method

While various deep learning architectures have been explored to address different tasks, convolutional neural networks (CNNs) are the most prevalent deep learning architecture typologies for computer vision and image analysis [14]. A typical CNN comprises a series of layers that successively map image inputs to desired end points while learning increasingly higher-level imaging features, Figure 1. Starting from an input image, 'hidden layers' within CNNs usually include a series of convolution and pooling operations extracting feature maps and performing feature aggregation, respectively. These hidden layers are then followed by fully connected layers providing high-level reasoning before an output layer produces predictions. CNNs are often trained end-to-end with labeled data for supervised learning.

Figure 1. This illustration depicts the overall architecture of the CNN used in this work. Note that H and P stand for healthy and pneumonia, while V and B stand for viral and bacterial.



The algorithm developed for this work uses state-of-the-art CNN architectures (incuding InceptionV3, InceptionResnetV2, and Xception) with custom modifications that enable us to train and test the algorithms more efficiently and provide the insights of the decisions made by the network for localizing the signal. All these architectures are very deep and comprise hundreds of millions of free parameters/weights to be determined by training. We use only the convolutional parts of these architecture and map them onto the output layer using a SoftMax activation via one Global Average Pooling layer. This reduces the total number of parameters in the architecture, which otherwise come from the fully connected layers placed after the convolutional part. With this configuration, the network is extremely deep to see high-order features in the dataset, with much fewer parameters (~20 to 50 million).

The efficient training of such networks depends largely on the choice of hyper-parameters such as optimizing the algorithm, learning rate, momentum, loss function, and batch sizes. The wrong choice of these hyper-parameters can make training difficult and highly inefficient. As there are no rules for choosing the right values of these hyper-parameters, many groups in the industry use trial-and-error method. We employ sophisticated statistical algorithms, such as Markov chain Monte Carlo (MCMC) and genetic algorithms, to sample the hyper-parameter space. While we train several different architectures, as mentioned above, for each training set, the final model is a hybrid average decision made by all trained CNN. This model can make predictions on new images for detection or diagnosis, depending upon the task at hand.

Another important aspect is the localization of the signal that governs the decision of the model. The architecture defined above gives a natural way to localize this signal. We take the output of the last convolutional layer and weight it by the parameters used to map it onto the SoftMax layer. We then marginalize them to build a heat map that quantifies the strength of the signal coming from different parts of the original image. We recognize that this quantification might not be possible with conventional CNN architectures where fully connected neural network layers are placed at the end of the CNN.

Results and Discussion

Early and accurate diagnosis of Pediatric Pneumonia is an important and challenging task. CXR has proven to be a useful tool to detect pulmonary consolidation as evidence of pneumonia. However, diagnosis of pneumonia by x-rays has limitations: it is operator-dependent, and it needs to be carried out and interpreted by trained personnel. In fact, while expert radiologists can account for only a handful of qualitative features, automation through deep learning has the potential to assess a large number of quantitative features along with their degrees of relevance while executing the task of examining medical images to arrive at a repeatable conclusion, in a fraction of the time required for a human interpreter. As the most extensively used architecture of deep learning, CNN has attracted a lot of attention due to its great success in image classification and analysis. This strong ability of CNN motivated us to develop a CNN-based prediction method that can accurately diagnose community-acquired pediatric pneumonia and further, distinguish its etiology (i.e., whether the disease is bacterial or viral in nature) in chest X-ray images.

To evaluate the classification performance, four measures were computed: classification accuracy, recall, precision, and area under receiver operating characteristic curve (AUROC). The accuracy of a classification is its ability to differentiate diseased and healthy cases correctly. To estimate the accuracy of a model, we calculate the proportion of true positives and true negatives in all evaluated cases. The recall of a model is its ability to determine all patient cases correctly. For that, we calculate the proportion of true positives in diseased cases. The precision is computed as the proportion of correctly classified positive samples among the total number of positive samples. The AUROC is calculated by plotting the ROC curves (plotting the true positive rate against the false positive rate at various threshold settings) and calculating the area under the curve, a metric that falls between 0 and 1 with a higher number indicating better classification performance.

The performance statistics of the model on a test set is depicted in Table 2. Additionally, the model achieved an AUROC of 0.992 and 0.976 for normal vs. pneumonia and viral vs. bacterial prediction, respectively. Interestingly, the algorithm used in this study outperformed the state-of-the-art ML models described in Ref [13], see Table 2 that provides comparative performance statistics. Figure 2 presents the confusion matrices for the performance of the model with respect to classification task. Figure 3 shows a typical input radiograph (top row) and the corresponding heat map overlaid on the

radiograph (bottom row).

Table 2. Performance comparisons of PCI's model vs. state-of-the-art, Ref [13].

	Model	Normal vs Pneumonia	Viral vs Bacterial
Recall	PCI	97.4	98.3
	Ref [13]	93.2	88.6
Precision	PCI	97.2	90.9
	Ref [13]	-	-
Accuracy	PCI	96.6	92.9
	Ref [13]	92.8	90.7
AUROC	PCI	0.992	0.976
	Ref [13]	0.968	0.940

Figure 2. Confusion matrices for the performance of the model: a) Normal vs. Pneumonia; and b) Viral vs. Bacterial.

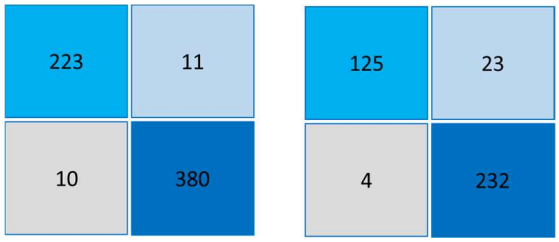
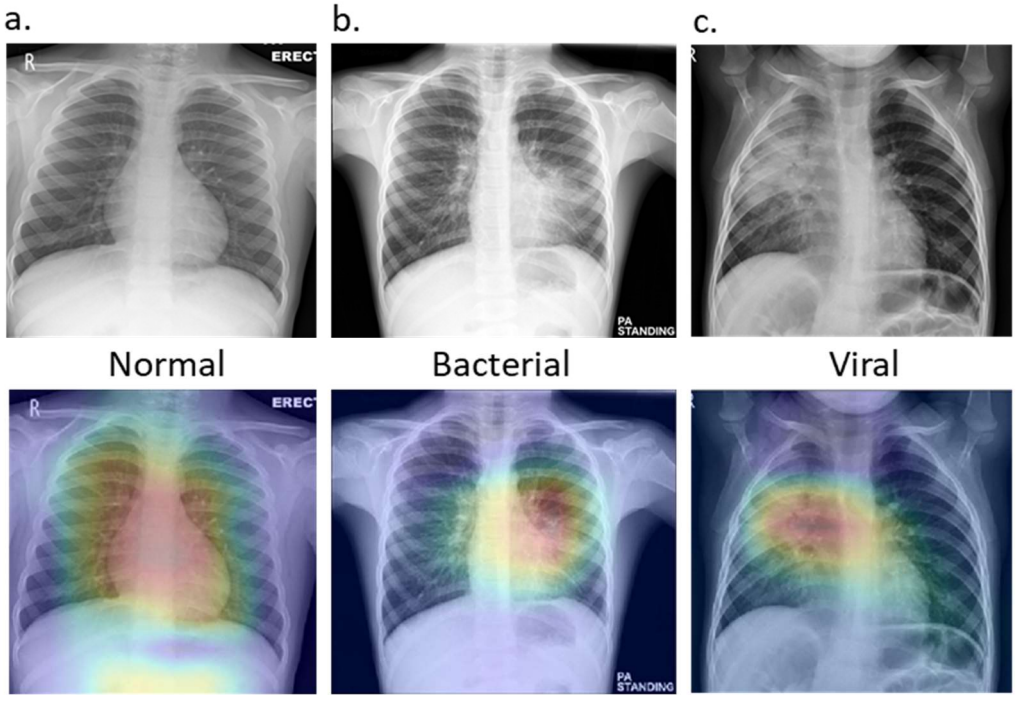


Figure 3. Chest X-ray images of pediatric patients: a) Normal CXR showing clear lungs with no abnormal opacification; b) Bacterial Pneumonia exhibiting focal lobar consolidation in the lower right lobe; c) Viral Pneumonia manifesting with diffuse interstitial patterns in both lungs. Processed images (bottom row) show localization of the decision-making signal as heat maps. Note that the heat maps are generated as a two-dimensional score grid, computed for each input pixel location. Pixels carrying high importance with respect to the expected class appear bright red with distinct color transitions for varying ranges. The generated heat maps are superimposed on the original input to localize image-specific region(s) of interest.



Our results demonstrate that DNNs can be trained, using large data sets and without having to specify localization data, to detect pneumonia and classify it into viral vs. bacterial in CXR images with high recall (97.4% and 98.3%, respectively) and high precision (97.2% and 90.9%, respectively). In general, a common criticism of computer-assisted diagnostic systems is that they increase recall at the expense of lowering precision, which results in unnecessary procedures and increased costs. It is noteworthy that the higher recall did not come at the expense of a lower precision which is probably due to the high diagnostic accuracy of the model.

To further mitigate the recall–precision tradeoff, and to decode the predictions made by the CNN-based model (which are otherwise perceived as black box), we attempted to visualize as heat maps the region important for classification within the input image that the model uses to identify a particular class in the CXR image. These heat maps not only assist in tuning and optimizing hyperparameters, but they also help in visualizing and debugging the decision process in classification network. They may further help in building the confidence of the end-user in model predictions prior to possible deployment by providing user-friendly explanations to how the model reached a specific conclusion. Thus, extracting the calculation process of CNN in a human-interpretable form, such as by visualization, shows the possibility to open the black box of deep learning.

There are several advantages to the diagnostic model developed in this study, including reproducibility and consistency of interpretation, high recall/precision, and near-instantaneous reporting of results. In addition, because the algorithm can have multiple operating points, its recall and precision can be optimized to match requirements for specific clinical settings - for example, high recall for a screening setting. Besides, there are several clinically important applications of the model. First, such a system may allow worklist prioritization. In this scenario, the cases detected as bacterial pneumonia could be moved ahead in the image interpretation workflow, allowing the sickest pediatric patients to receive quicker diagnoses and treatment. Second, the automated nature of the system may help combat radiologists' fatigue which occurs due to a shortage of trained personnel - particularly in resource-constrained areas - which leads to a dramatic increase in their workloads impacting their diagnostic accuracy. On average, a radiologist must interpret one image every 3–4 seconds in an 8-hour workday to meet workload demands [14]. With increasing numbers of medical images transitioning from traditional printed films to Picture Archiving and Communication Systems (PACS) and with the availability of multi-contrast, multi-phase and multi-planar imaging, there is an exponential growth of image data -both in terms of the size and overall complexity – due to which the interpretive task of radiologists has become further complicated. Then again, interpretation must often be delivered in the context of previous examinations. Deep learning systems may be well-suited for finding and presenting this complex task. Third, for healthcare workers and clinicians who lack subspecialized expertise, such an automated system may solicit a "second opinion" - especially when the etiology is difficult to assign or interpret.

There are several limitations in the present work. First, this is a proof-of-concept study in which the dataset was obtained from a single hospital and was limited to one pathological finding (pediatric pneumonia) and cannot evaluate other lung disorders. Real-world situations may be significantly more complex than the dataset used here, particularly regarding diversity of abnormalities on CXRs. Second, to lessen the number of parameters inherent to the neural networks, the images in our experiment were down-sampled to 299 X 299 pixels. While using higher resolution images may improve accuracy, particularly for subtle findings, it increases computational time and requires highly robust systems and graphics processing unit memory. Third, while the localization information provided by our algorithm can help visualize the logical background of classification output, there may be cases in which the resultant localization of abnormalities on CXR may not be clinically relevant. Finding these inconsistencies is central to establishing the reliability of the algorithms, particularly within the medical community. Finally, this work was performed using a dataset that was publicly available at the time of the study. Further evaluation of the effect of additional training cases on accuracy would be valuable as more datasets become accessible. It would be interesting to evaluate the use of DCNNs in a clinical practice for pediatric pneumonia, particularly in pneumonia-prevalent regions. Further work is needed to implement and evaluate other DNN architectures.

In the future, it will be interesting to develop applications that can combine the information extracted from medical images with genomic characteristics to facilitate the process of incorporating as much information as available to guide patient treatment.

Conclusion

To summarize, we developed a CNN-based decision support system and demonstrated the effectiveness of the algorithm to detect pneumonia in pediatric CXRs to expedite accurate diagnosis of the pathology. We applied state-of-the-art visualization strategies to explain model predictions which is important for better-informed clinical decision-making. Such a system can be used to support multifaceted diagnosis of childhood pneumonia particularly in resource-constrained settings, compensating for the shortage of expensive equipment and highly trained clinicians, leading to a broad clinical and public health impact. This system does not get fatigued, provides consistent read, and gains subspecialized expertise by being provided with labeled radiographs from human experts and accuracy improves over time with additional data provided to the system. Further research is necessary to determine the feasibility of applying this algorithm in the clinical setting and to determine whether use of the algorithm could lead to improved care and outcomes compared with current assessment of pediatric community-acquired pneumonia. Interestingly, the techniques we have described in this paper could potentially be employed in a wide range of medical images across multiple disciplines expediting the diagnosis and referral of similar treatable conditions, and thereby facilitating earlier treatment, resulting in improved clinical outcomes. In fact, this technology is easily transferrable across other domains such as the energy sector, finance, banking, transportation, and so forth, where data-driven decision-making is desirable.

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May 28;521(7553):436-44. doi: 10.1038/nature14539. (b) Hinton G. Deep learning-a technology with the potential to transform health care. JAMA. 2018 Sep 18;320(11):1101-1102. doi: 10.1001/jama.2018.11100.

[2] Gulshan V *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016 Dec 13;316(22):2402-2410. doi: 10.1001/jama.2016.17216.

[3] Esteva A *et al.* Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115-118. doi: 10.1038/nature21056. Epub 2017 Jan 25. Erratum in: Nature. 2017 Jun 28;546(7660):686.

[4] Sirinukunwattana K *et al.* Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging. 2016 May;35(5):1196-1206. doi: 10.1109/TMI.2016.2525803.

[5] Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. Med Image Comput Assist Interv. 2013;16(Pt 2):411-8.

[6] Liu L *et al.* Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet. 2016 Dec 17;388(10063):3027-3035. doi: 10.1016/S0140-6736(16)31593-8. Epub 2016 Nov 11. Erratum in: Lancet. 2017 May 13;389(10082):1884.

[7] (a) Katz SE, Williams DJ. Pediatric community-acquired pneumonia in the united states: changing epidemiology, diagnostic and therapeutic challenges, and areas for future research. Infect Dis Clin North Am. 2018 Mar;32(1):47-63. doi: 10.1016/j.idc.2017.11.002. (b) Jain S *et al.* Community acquired pneumonia requiring hospitalization among U.S. children. N Engl J Med. 2015 Feb 26;372(9):835-45. doi: 10.1056/NEJMoa1405870.

[8] Rodrigues CMC, Groves H. Community-acquired pneumonia in children: the challenges of microbiological diagnosis. J Clin Microbiol. 2018 Feb 22;56(3). pii: e01318-17. doi: 10.1128/JCM.01318-17.

[9] Cherian T *et al.* Standardized interpretation of pediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. Bull World Health Organ. 2005 May;83(5):353-9.

[10] Coche EE, Ghaye B, de Mey J, Duyck P, eds. Comparative interpretation of CT and standard radiography of the chest. New York, NY: Springer Science & Business Media; 2011. doi:10.1007/978-3-540-79942-9.

[11] Donald JJ, Barnard SA. Common patterns in 558 diagnostic radiology errors. J Med Imaging Radiat Oncol. 2012 Apr;56(2):173-8. doi: 10.1111/j.1754-9485.2012.02348.x.

[12] Nakajima Y *et al.* Radiologist supply and workload: international comparison - Working Group of Japanese College of Radiology. Radiat Med. 2008 Oct;26(8):455-65. doi: 10.1007/s11604-008-0259-2.

[13] Kermany DS *et al.* Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018 Feb 22;172(5):1122-1131.e9. doi: 10.1016/j.cell.2018.02.010.

[14] McDonald RJ *et al.* The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Acad Radiol. 2015 Sep;22(9):1191-8. doi: 10.1016/j.acra.2015.05.007.