Pneumonia prediction and detection through machine learning

A review of the use of transfer learning in pneumonia detection tools

Divyansh Verma
Faculty of Design and Creative Technologies
Auckland University of Technology
Auckland, New Zealand
Student ID: 23201983

Joshua Ladowsky
Faculty of Design and Creative Technologies
Auckland University of Technology
Auckland, New Zealand
Student ID: 18036421

***Abstract*—Pneumonia is a serious and often viral infection of the lungs that can be fatal in the worse scenarios and often leaves survivors with lasting side effects. One of the most common methods for detecting pneumonia is through chest X-rays, however as these X-rays must be examined by doctors this process is time consuming and if checked too early signs can be missed but if checked to late the condition can be lethal. As such many Deep Learning models have been developed to analyse chest X-rays and classify if pneumonia is present.**

One trend in this field of research is the use of transfer learning to develop generalized models that can be retrained to produce better results using smaller datasets. In this review we review eight different models from six different research papers and analyse how transfer learning is used comparing against models that did not utilize transfer learning. through analysis we find that a transfer learning model using the AlexNet architecture produces the best results of the eight however further research would be required as the model uses a large machine ready dataset rather than a dataset acquired from a clinical setting.

Additionally we conclude that it would serve many of these researchers well to not only use local clinical data in their training but develop a pipeline such that dynamic training can be done in a clinical setting for longer term review of a model that is produced.

***Index Terms*—Data mining, Deep learning, Transfer Learning, Pneumonia Detection, CNN, Literature Review, data mining, AlexNet**

I. Introduction

Pneumonia is a common and serious infection of the lungs that can be caused through many different difference viral outbreaks. Early detection of pneumonia is vital in improving the efficacy of treatment of this life treating infection. X-Rays of the chest are common methods of detecting pneumonia, however, the process of examining these images can be time-consuming and can lead to human error with both false positives and false negatives. As such much research has been done into the development of tools that can process these X-Rays and provide a prognosis quickly and efficiently.

The development of medical tools for the detection of ailments such as pneumonia are one of the more common points of research for machine learning tools. The development of these models are useful as often machine learning models, often leveraging deep learning tools can detect trends in X-Ray images that the human eyes cannot. While these tools are still in their infancy, it is vital that we examine the efficacy of the methods we are using to develop these models such that we can more effectively advance this area of research.

One such trend in the development of these tools is the use of transfer learning in the development of new models. Alternatively, many researchers opt to attempt the development of a model from scratch, either through the training of an untrained model that is available or creating their own from the ground up.

II. Background / Motivation

This research will focus on the methods that researchers have undergone to develop models that they have trained for use in pneumonia detection as well as the results of these models in their analysis. Our key metrics for successes will be the models' accuracy both with training data and testing data, but we will also factor the ratios of false negatives and false positives. Ideally, a model wouldn't produce either false positives or negatives however when they do, we should consider the fact that a false negative is much more dangerous compared to a false positive, this is because untreated pneumonia can cause death in the patient whereas a false positive can be filtered out with follow-up testing from the doctor.

We will examine three different scenarios where one model used transfer learning and the other did not. Firstly we will examine models that use clinical data rather than data that was procured from an online dataset, this will provide us with a better understand of the performance differences when the sample size of the pneumonia data is smaller than what is available online. Secondly, we wil examine two models using the VGG architecture to see how similar architectures can benefit from the use of transfer learning and how it can affect their results further along. Finally we will review two models that are trained on the same dataset procured from Kaggle to understand how these methods can change the outcome with the same dataset

A good deal of Pneumonia detection tools leverage Deep learning technology to better develop and build their systems, this is because Deep learning tools such as Convolutional Neural Networks (CNNs) are powerful tools for the purpose of image classification and have strong success in the field of developing medical tools powered by AI.

Many of these tools have been developed with their own models from scratch and or used an already developed but untrained model with new optimization algorithms and or methods of data preparation to create new systems for the goal of image classification, both broadly and for specific classifications. The use of these models are key in advancing our knowledge of how image classifiers are built and how they can best be improved upon.

Alternatively, tools can use Transfer Learning to teach a model that was trained for a related task how to achieve this secondary task. Transfer Learning can reduce the time it takes for a model to be developed and ready for its process and means that models that are already successful can be used in a wider range of uses. A model that is designed as a general object classifier can be instead retrained to detect and classify X-Rays that contain markers of Pneumonia for example.

Through the analysis of these trends we will be able to better understand what methods we should move forwards with developing and focusing on as we continue to build upon and develop tools that can work more effectively. as well as the direction we should take these tools further in deploying them so they can be used to better the direction we should take in their development.

III. Comparison of related work

A. *Models trained with clinical data*

Research done by Abdurahman and Yimer [1] procured a large dataset from two private medical institutes and a public machine learning ready dataset of Chest X-Rays for use in their model training. The researchers fed the data to their model and various pre-trained models. Results from the testing found that the researcher's proposed models created 8 false positives and 10 false negatives, of these 10 false negatives, it was concluded that 2 images contained were two noisy to even identify individual organs, these images came from the Merawi (MA) hospital while the other 8 images did not have this issue. Of the remaining 8 false negatives, 7 were procured from the hospital Felegehiwot (FHH) where only 15% of the dataset contained pneumonia X-Rays, and the final came from the online machine ready dataset.

Preforming Principal Component Analysis on the three data sets, they found that the FHH data set had pneumonia images were evenly spread along the graph whereas the online dataset and MA dataset both have higher separation of the data, this is likely due to the significantly lower ratios in the FHH data set with only 15% of the X-Rays containing pneumonia whereas the MA and online datasets contain 47% and 62% pneumonia to normal images.

The study had attempted to create a generalised deep learning model for pneumonia detection, and created a model that did preform better than previously trained models that they compared to, scoring 92% for accuracy, recall, and precision, however their scores on the generalised model were lowered by the images from both hospitals suggesting that their model isn't ready for use. Comparatively, the model only testing against the MH images achieved 89% for accuracy, precision, and recall

This compares with research by Sakib et al. [2] developed a model using the DenseNet-121 architecture utilized a transfer learning methodology to produce a generalised model designed to be tested against both a public dataset and a data set that was locally procured from and sanitized from a local diagnostic centre.

The model produced good results on the public data set, however testing on the local data set produced sub-par results, this is certainly due to the minuscule data sample acquired from the local clinic, a total sample size of 129 chest X-Rays clearly isn't adequate for the proper training of the model. The online dataset resulted in a scoring of 95%, 95%, and 95.5% for accuracy, recall, and precision respectively while the local dataset produced scores of 88%, 87%, and 84.5% respectively.

With the clear and drastic drop shows that while this method has promise the model will require a much larger data set to train the system for use in a clinical setting. The Online data set has nearly 100 times more samples than the local clinical data sets. Due to the drastic difference in datasets. However, further testing should be done to examine how much data would be required to produce a model that is similar in efficiency to the model trained with the online dataset.

Both of these models trained using clinical data from a local source and using some form of online dataset in the training of the model and in both cases the model preformed worse on the public datasets compared to their performance on the online dataset. Additionally, in both cases, there is a significant difference in the number of samples for both. However, Abdurahman and Yimer's [1] model preformed better trained on the MA data set than Sakib et al.'s [2] model preformed on the local data set. This is most likely attributed to the difference in the sample size of the datasets and as such it should be noted the fact that the results are comparable even with such a vast difference in sample sizes.

This is as a result of the pre-training the occurs prior to the transfer learning is undertaken for the model and shows the efficacy of the method and why it works well in a field such as this where sample sizes in clinical settings are smaller than are typically available online. Further research could be done into using the MA dataset with transfer learning to compare how this significantly larger sample size preforms when using transfer learning rather than directly training.

B. *Models using VGG architecture*

Research by Sharma et al. [3] preformed reviews on the effects of transfer learning on a total of 11 different CNN models. Initially, the models are pre-trained on images of X-Rays. After pre-training was concluded each model had additional layers add; drop-out layers were included to prevent over fitting and global pooling layers were added to assist with feature mapping and vectorizing.

Of the 11 models tested, the VGG-16 was deemed to be the best preforming after both the rounds of testing. However, before the transfer learning testing, the VGG-16 model preformed notably worse than other competing models. This suggests that the VGG-16 model adapts well to the process of transfer learning but may require higher levels of training data than was preformed in the first round.

before retraining, the VGG-16 model achieved 81.57%, 84.86%, and 98.97% for accuracy, precision, and recall respectively. This scoring makes the model the third worse model by F1 score. However, after retraining for the task of pneumonia detection the model became the highest scoring model by F1 scoring, achieving a scoring of 89.74%, 89.81%, 89.13% for accuracy, precision, and recall respectively.

Research preformed by Das et al. [4] attempted to develop a model to detect pneumonia through the use of the VGG-19 architecture. Using an online pneumonia data set after data pre-processing the study finds that their model preforms quite well, producing results of 95%, 92%, and 94% for Accuracy, Precision, and Recall respectively.
The study succeeded in producing a high-quality model that preforms well in detecting the presence of pneumonia in the testing set that was provided. However, it should be noted that the lack of clinical data, that being chest X-Rays from hospital settings does mean we cannot acquire a strong understanding of the results and how the model would preform in a clinical setting.

In both of these cases researchers used architecture from the Visual Geometry Group in their attempts to preform pneumonia detection. It should be noted that while Sharma et al. [3] did also test a VGG-19 model, its performance was lacking compared to the 16 model. This leads up to draw the conclusion that the VGG-19 model is well suited to the process of training as indeed the model pre-trained very well in the initial study, however this model is not suited for a transfer learning approach to Deep Learning and pneumonia detection. Additionally, the VGG-16 underperformed compared to the VGG-19 model produced by Das et al. [4].

C. *Models trained on the same dataset*

A study preformed by Athar et al. [5] set out to use transfer learning on the AlexNet deep learning model, pre-trained on a large ImageNet dataset to preform feature extraction on images of chest X-Rays and then fine-tuned the pre-trained model to detect pneumonia markers in chest X-Rays.

Once the model has been fine-tuned using the Kaggle pneumonia dataset, an adversarial training method was preformed to assist the training of the AlexNet model with its new task. Using this method the model was able to achieve an impressive 98%, 99 %, and 98% for the accuracy, recall, and precision of the model respectively. The study indicates that this method of transfer learning can provide high-quality results. However, the researchers do acknowledge that the lack of data from a clinical setting does limit the conclusions they can draw from their results.

Researchers Lin et al. [6] propose a system to combine a CNN with a State Vector Machines (SVM). The CNN will preform complex feature extraction, turning it into ”high-dimensional feature vectors” [6] that are passed into the SVM model for final decision-making.

The study produced two different versions of the ”CNVM” model, a linear model and a non-linear model that integrates a non-linear Radial Basis Function (RBF) into the SVM classifier. Of these two models the RBF model preforms marginally better than the linear model does, however both achieve high-quality results across the board. The RBF model achieving 85% 95% 70% for accuracy, recall, and precision respectively while the linear model produced 82% 97% 63% respectively.

The fine-tuning of the CNVM model can lead to drastic changes in the performance of the model due to the differences in decision-making between the CNN and SVM models. The study does recognize that a more efficient method of training of the system could be devised, and plans that in their future workings.

While both models preformed well, there is a clear showing that the AlexNet model preformed significantly better than the CNVM model did. This is most likely due to the approach of using transfer learning tied together with adversarial training. However, it is also possible that the CNVM model could be further fine-tuned to preform better results. As the researchers of the CNVM model discussed in their paper, the model requires a better optimization method as the slightest changes to the parameters of the model can result in drastic changes.

It could also be assumed that the adversarial training that was used on the AlexNet model has a significant effect on the model as it was retrained on the pneumonia dataset. while an better method of optimization is require for the CNVM model they should perhaps investigate modifying their methods of training also.

IV. Discussion

There are varying results from both the direct training approach and the transfer learning approaches but we can draw several conclusions. Firstly we can see that transfer learning has great promise for the development of generalized models that can be deployed into clinical settings. we can see this in the development of the DenseNet-121 model by Sakib et al. [2] who's model preformed quite well when trained of such a small dataset. this shows that the models adapability into new environments will be much more suitable than if you attempted to train the model solely from clinical data as we saw in Abdurahman and Yimer's CNN architecture [1]. More research needs to be done using larger clinical data sets as these are the use case for the models, this is a great weakness

Researcher(s)	Model	Acc	recall	precision
Abdurahman and Yimer [1]	CNN	92%	92%	92%
Sakib et al. [2]	DenseNet-121 (Online)	95%	95%	95.5%
Sakib et al. [2]	DenseNet-121 (Local)	88%	87%	84.5%
Sharma et al. [3]	VGG-16	89.74%	89.81%	89.13%
Das et al. [4]	VGG 19	95%	92%	94%
Athar et al. [5]	AlexNet	98%	99%	98%
Lin et al. [6]	CNVM (rbf)	85%	95%	70%
Lin et al. [6]	CNVM (ln)	82%	97%	63%

TABLE I
LIST OF MODELS AND THEIR SCORES

for many studies, as they are not as adaptable to a clinical environment.

As we can view in table I, the models that are built using transfer learning do not distinctly outpace those that don't. The different approaches are not able to outpace each other in either way this shows that neither method can directly outpace another. However, the top preforming model of those reviewed was the AlexNet model. This model also utilized adversarial learning as it was transfer learning to better further along the process, it is possible that we could improve all of these systems by implementing an adversarial approach. Further study should be done into weather this method would be beneficial to the research of pneumonia classification as a whole.

Finally the work of these models should be furthered into designing a system to pass new X-Rays through the model automatically such that a pipeline of diagnostic tools can be implemented into a clinical setting for quick and efficient analysis can be taken place for the use of medical professionals to ensure that they can provide the best care to their patients. It should be considered though that while these systems can be effective medical practitioners shouldn't take the model's work as perfect as any Deep Learning model can produce incorrect results.

V. CONCLUSION

The research into improving methods of detecting pneumonia is a vital as while there are many models that are being proposed if we neglect to review and analyse the methods and trends of other researchers we will slow the progress of development. We have reviewed six different attempts and analysed eight different proposed models that all achieve the process of detecting the signs of pneumonia in images of chest X-Rays.

Of the models examined, the AlexNet model using transfer learning and adversarial training to achieve the highest score. this has lead us to the conclusion that transfer learning does have great promise in the field of pneumonia detection as the models are able to be fitted locally to a smaller dataset. However, we cannot conclude that all attempts at pneumonia detection should use transfer learning as not all models using transfer learning beat out models that are directly trained.

We can however conclude that further research should endeavor to work alongside clinical practices to build better datasets from these practices as much of the current research uses the same dataset produced from Kaggle, while having the same dataset does make it easier for us to compare the results of different attempts because this model has been cleaned and pre-prepared it fails to reproduce the same conditions that a normal clinical environment would face. Alongside this, researchers should work with these clinics to design an interfaces with their systems such that a model can directly review new X-rays and review them.

With a pipeline developed to examine these X-rays as they are taken we can further test the efficacy in a clinical setting and build a greater understand how these models will preform in their intended purpose. This is currently a large gap in current research into pneumonia detection and classification and should thus be further looked into to better learn from work done.

REFERENCES

[1] H. S. Abdurahman and A. K. Yimer, "Application of Machine Learning Algorithms for Pneumonia Detection and Classification," in *2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*. Bahir Dar, Ethiopia: IEEE, Oct. 2023, pp. 67–72. [Online]. Available: https://ieeexplore.ieee.org/document/10302262/

[2] S. N. Sakib, S. Y. Rubaiat, R. Masud, M. Sarker, and C. Bepery, "Efficacy of Transfer Learning Models in Pneumonia Detection with Scarce Local Healthcare Data," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE, 2024, pp. 288–291. [Online]. Available: https://ieeexplore.ieee.org/document/10088848/

[3] G. K. Sharma, P. Harjule, T. Sadhwani, B. Agarwal, and R. Kumar, "Sequential Transfer Learning Models with Additional Layers for Pneumonia Diagnosis," in *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)*. Srinagar Garhwal, India: IEEE, Jun. 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/10262764/

[4] R. Das, D. S. K. Nayak, C. P. Rout, L. Jena, and T. Swarnkar, "Deep Learning Techniques for Identification of Pneumonia: A CNN Approach," in *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*. IEEE, 2024, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10507933/

[5] A. Athar, R. N. Asif, M. Saleem, S. Munir, M. R. Al Nasar, and A. M. Momani, "Improving Pneumonia Detection in chest X-rays using Transfer Learning Approach (AlexNet) and Adversarial Training," in *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*. Dubai, United Arab Emirates: IEEE, Mar. 2023, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/10111193/

[6] Y. Lin, T. Li, H. Xie, T. T. Toe, and L. Lei, "Optimizing Pneumonia Detection Model based on CNVM (CNN-SVM)," in *2023 2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)*. Zhuhai, China: IEEE, Oct. 2023, pp. 133–140. [Online]. Available: https://ieeexplore.ieee.org/document/10396157/