Radiographic Pneumonia Detection Using Convolutional Neural Network to Do Binary Classification

Aidan Daly
Boston College
dalybu@bc.edu

Luke Pisani
Boston College
pisanilu@bc.edu

For our GitHub Repository with our project see here. Or use this link https://github.com/aidandaly24/RSNA-Pneumonia-Detection.

1. Introduction

The purpose of our project was to apply what we've learned about Biomedical Image Analysis to a topic in medicine of which we were both interested. We though about doing something flashy like cancer detection or tumor identification, but we ultimately decided to focus on a lesser focused application of deep learning in the context of medicine, pneumonia. We believed that pneumonia detection would make a great project for a plethora of different reasons. Primarily, the data was easy to find and there was a lot of available data as x-ray scans of the lungs and upper respiratory system are often taken to examine a variety of conditions and are usually shared for research purposes as they are not easily identifiable. There being a large volume of quality data available was crucial to our process because data is the most important item needed to develop a successful model. The quantity of data out there also allowed us to test out quite a few different options, before settling on our approach. We also felt confident with our binary classification problem in determining if the scan either had pneumonia or didn't have pneumonia because this was something we had previous knowledge of and felt comfortable with from what we had learned in class. Finally, we were excited about our topic because pneumonia is the deadliest childhood disease, with one in six of all childhood deaths being due to the disease and is mainly concentrated in underdeveloped countries [6]. Boston College heavily promotes its motto of "men and women for others" and our project fits that perfectly as it has the potential to help people by diagnosing pneumonia before it gets to a point at which lasting harm is caused. In 2019, an estimated 2.5 million people died of pneumonia worldwide and we believe that deep learning technology will help to reduce this number and to save lives.

The high-level goal of our project was to develop and construct a model that when an x-ray image of the chest is input, detects whether pneumonia is present or not. The best approach we figured to achieve our goal was to use radiographic images of the chest because there are plenty of these images available to use for out data that are clear and contain the perfect amount of variation compared to other types of imaging like Positron emission tomography (PET) scans, Magnetic resonance imaging (MRI), and Ultrasonography.

Lately, as technology has been rapidly advancing, there have been major developments in using computers to solve important problems like medical imaging detection when it comes to things like cancers, fractures, and even diseases like pneumonia, leading to companies and organizations being formed with the express purpose of solving and refining these imaging processes. The problem we are seeking to solve is very important at this time because as people continue to innovate, they are making changes that will forever alter the process of medical diagnosis which will inevitably streamline and perfect the medical system. Technology has greater applications beyond entertainment or communication and with projects like ours this becomes evident as models like ours can be used to save lives and really make a difference, solidifying how and why this emerging field is not just important but critically life altering.

As for the source code production in our project, Aidan worked primarily on the downloading, visualizing, and processing of the data, building the model architecture, as well as helping with a lot of the testing and training. Luke worked substantially on background research, finding the dataset, training, and testing the model with particular emphasis. Throughout the project we both worked together and contributed equal amounts of work to the model. Further explanations of out individual contributions are elaborated upon in Section 6 Contribution.

2. Related works

Under the competition we based our model upon, "RSNA Pneumonia Detection Challenge," there were many models available for view [5]. From this challenge we got

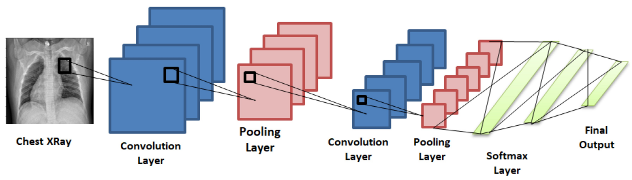

Figure 1. Flow of CNN used on X-Ray Image. The image is processed and input into the model. At this point, it is passed through a number of hidden convolutional layers. Finally, it is given to a fully connected layer and receives its prediction. [3]

our dataset and were able to see how other people solved this problem. This specific competition was in 2018 and has since closed in that same year with over 1,400 teams having participated and submitted models.

In the biotechnology field, there are many companies that use deep learning and artificial intelligence to aid the diagnosis and detection of diseases. Aidoc is an example of these kind of companies that uses artificial intelligence (AI) to analyze medical images and detect abnormalities that also uses X-ray imaging to examine chest scans and also screens for pneumonia in the same way our model does, albeit more accurately [1]. Additionally, they also focus on other sorts of diseases and different imaging techniques across different fields and can also predict the efficacy of new therapy types.

Similar to Aidoc, Enlitic uses AI in the pathology process to examine and interpret medical images from different types of scans with the goal of improving the speed and accuracy of deep learning models to further help patients by perfecting the diagnosis process, making it faster and more affordable [2].

In a study about deep learning and medicine, The National Cancer Institute found that deep learning (91% accuracy) was able to outperform doctors (69% - 71%) in cancer detection when the doctors were using traditional methods [5]. This illustrates that deep learning can be used in medicine effectively and has the potential to outperform doctors. Additionally, using models for pneumonia detection is useful because sometimes it is very challenging to see the disease even when it is present on scans, so models like ours would eliminate that visualization difficulty leading to better overall patient outcomes. It is also very difficult to see pneumonia when present early on in the disease presention, so creating a model that can detect it early on is crucial and could help save the lives of many children where the disease is sometimes not diagnosed early enough.

3. Method

Once we had decided on a project, our first task was to find a large data set full of quality training images which we would be essential to us in order to build a good, functioning model. We looked through lots of medical data each of different types including PET, MRI, ultrasound, and other varying pathologies. It became clear to us very early on in our search process that we should be focusing our efforts on looking at radiographic (x-ray) images. The images we settled on were simpler and easy to identify pneumonia on and were widely assessable compared to other imaging types. Furthermore, we identified the Kaggle data as the one we wanted to use after looking through lots of radiographic datasets. The dataset contains around 30,000 images split pretty much evenly between posteroanterior and anteroposterior frontal views of the chest and around 4,500 training images. These images were assigned a value: 0 for unknown, 1 for pneumonia present, and 2 for pneumonia absent. The images looked at were x-ray images of the chest, which is perfect for our goal of identifying pneumonia because the chest and lungs are where pneumonia is first visible in the body early on in the disease's progression.

The next step of our process was to consider the structure of our model. We ultimately decided to use a Convolutional Neural Network (CNN) because we felt that it would work best given that our task required binary classification. Throughout the duration of this semester we have used CNNs very frequently and we have learned to feel comfortable using them and have grown accustomed to their mechanics and we knew that we would be able to build a better model since we were familiar with the foundations. We specifically decided to use ResNet50. We also attempted to train a DenseNet and VGG19 model but Grad-Cam did not end up working as well as ResNet50, so we decided to stick with that. We also added Grad Cam which will help our model to make predictions and will allow us to visualize what are model is thinking about when given an image. Our final model ended up being a pretrained ResNet50 with GradCam so instead of training our model we are fine tuning it.

We partially mirrored problem sets 3 and 4 as well as some of the labs when training and testing our model. We also found similar projects to ours online and took inspiration from those and incorporated elements into our model which was extremely helpful for writing our training section because for that part it wouldn't make much sense to begin from scratch so having an example to go off of was incredibly useful. The time we saved using a guide allowed us more time to use on other parts of our model which we specifically spent on fine tuning. Once our training function was formulated, we constructed a means of testing our accuracy which would be used on the validation and data. After every epoch in our validation set, we checked the accuracy which was then incorporated into the training function which allowed us to see the effects of tuning hyperparameters nad to track the progression of the model's learning. Overall, we spent a lot of time validating and training which

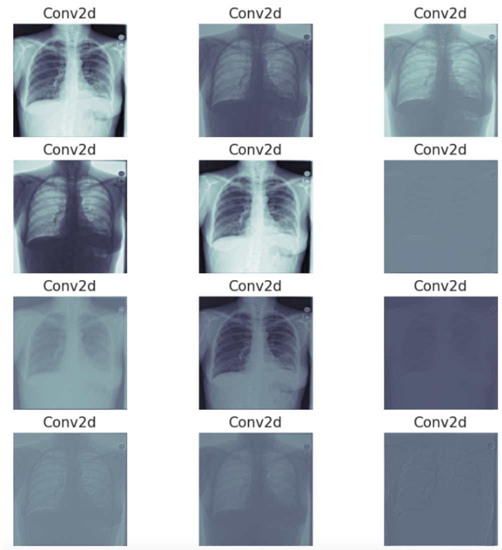

Figure 2. Convolutional layer filtering throughout model. This is an example image from our dataset that has gone through our model, which our model predicted to be Cancerous. This displays the filtering our model uses to make predictions after each convolution.

involved a lot of trial and error, all of which is detailed more in section 4.

After meeting with Professor Wei for the second time, he recommend that we look into a technique called Gradient-weighted Class Activation Mapping (Grad-CAM) which generates visual explanations for the predictions made by a CNN with the goal of identifying the regions of an input image that are most relevant for a specific prediction made by the CNN [7]. In the case of our model, we used Grad-CAM to better understand what areas of the image were used to determine whether or not pneumonia was present in the scan. At first it was a bit challenging to understand exactly how to use the technique in our model, but once we were able to utilize it, we were able to comprehend better how our model makes determinations. After implementing Grad-CAM, we were able to make changes to our model that lead to an improvement of accuracy, meaning that it added to our model's quality.

4. Experiments

The majority of our experimentation process was focalized around the training process. In order to create a model that would produce the best accuracy given a long training period time, we initially conducted training on smaller portions of the data for fewer epochs. This approach enabled us to train the model relatively more quickly which was helpful cause it allowed us to see the effects when we tweaked the hyperparameter settings. We would run the model, get an output, revise the parameters, and then repeat this cycle for hours until we were abel to obtain a version of the model we were satisfied with. This cycle also allowed us to be able to fine an optimal learning rate, number of epochs, and batch size in which we tried 32, 64, and 128 images per batch.

The dataset consists of over 30,000 images which meant that we needed to train the model on only a portion of the images to see the results of our changes. We began by running the model on one third of the dataset to finetune the learning rate, epoch number, and the size of the batch. We weren't happy with the results, so we went back to check the data. It turned out that the data was completely unbalanced and had lots of duplicates. Aidan went back and built out a remove duplicate and balance function which left our dataset perfectly balanced. We again finetuned our model on top of a pre-trained ResNet50 using "imagenet" weights and got much better results.

Additionally, we spent a lot of time configuring the format of the training function which we got to be able to check the accuracy of our validation set after every epoch which gave us an indication as to how the training was going. We also tested checking accuracy after every epoch with a smaller subset of the training data, but we quickly realized that his produced overly optimistic accuracy and did not accurately represent our model on unseen data which if used with real world data, would not generalize well. We ended up declining to check the accuracy of the validation set after each epoch in the way we just described for the reasons we just alluded to which is why we went with the other, alternate approach.

5. Conclusions

Our model had an accuracy of 88%, when we attempted to implement it as a legitimate pneumonia detector on data outside of our training and validation sets. We were happy with this as we trained it on top of a pre-trained ResNet50 and expected something of this type. One thing we weren't able to finish, but attempted to, was visualizing a Grad Cam heatmap on an image. Despite trying countless times prior to finishing up this report we could not get around some bugs that wouldn't allow us to visualize the Grad Cam heat map.

Deep learning models are being used more readily in the medical field as a whole with the Association of American Medical Colleagues recognizing some limitations but still models are beginning to be more accepted and used in medicine as time progresses. Specifically, the models usually are able to do great at performing one particular task, however they do not typically generalize well, but still they are being implemented with more frequency and will eventually be commonplace in medicine [4]. For now, doctors and computer scientists tend to agree that these models are best to be used alongside with a doctor's judgment which

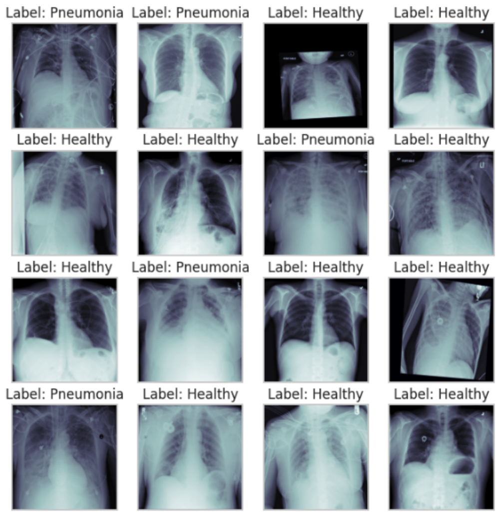

Figure 3. Examples of output on unlabeled test data. These images were run through our model, and given a label based on the model's prediction. There is no guarantee as to whether these labels are accurate. This is similar to if the model was given a fresh image from someone being tested for cancer.

will increase the accuracy of detection, reduce false positives, and prevent missed diagnoses overall. Deep learning models especially have the potential to make differences in underdeveloped or still developing countries that have limited healthcare access and doctor shortages because the models can help to make diagnoses that would otherwise go unnoticed. This is especially important in the context of pneumonia where the majority of cases occur in underdeveloped countries that have these aforementioned limitations in healthcare access.

We really enjoyed this project as a whole because while working we felt that being able to implement something like our model on a larger scale and not just a smaller problem set or lab was really important to our learning process and helped us to cement some of the ideas from the course. It was also very interesting to see how to combine our thinking and to work with another person as we had done a lot of the parts individually, but then had to combine our methods into one when putting the project together which was more challenging than we had anticipated originally. We both thoroughly enjoyed this process of creating something together that could be applied to a real world problem and have gain lots of experience not just in coding but in problem solving and in communication. The limitations of our model are known to the both of us, but we are totally satisfied with what we were able to build and accomplish and from all that we have learned this semester.

6. Contribution

In our work on this project we aimed to divide the work up as evenly as possible while both still putting a lot of effort into the project. We both believe we have completely accomplished this aim. A lot of working together on a project is collaboration and this team really put in the work. Despite specific contributions we both put the same amount of effort into this project and want to make sure that is clear.

Aidan worked on sampling and preprocessing our data, building out model architecture, training, and Grad Cam visualization. He also adapted his past work with filter visualization to this project which is found at the bottom of our source code. Aidan built functions specific to our data that would remove duplicates and balance the dataset. He also spent lots of time working on training alongside Luke. This specifically meant building out sections 2, 3, 4b, and 6 fully as well as split work on other portions.

Luke worked on training and testing mostly. This took lots of time and waiting for our model to train. Luke specifically spent time finding ways to increase speed of our model and learning. This meant switching out models and testing different pretrained models. Luke spent time looking into the differences between ResNet, DenseNet, and VGG, and was able to determine which would be a best fit. Luke also wrote and rewrote a lot of our training code until it worked perfectly. Luke's emphasis was on sections 4a and 5 but provided a lot of help with other sections as well.

As for this report, Aidan wrote the sections on Methods, Experiments, and Contribution. Luke wrote the sections on Introduction, Related Work, Conclusions, and Reference.

References

[1] Aidoc. Aidoc. https://www.aidoc.com/. 2

[2] Enlitic. Enlitic. https://enlitic.com/. 2

[3] Mohammad Farukh Hashmi. Efficient pneumonia detection in chest xray images using deep transfer learning. https://www.mdpi.com/2075-4418/10/6/417. 2

[4] American Association of Medical Colleges. Will artificial intelligence replace doctors? https://www.aamc.org/news/will-artificial-intelligence-replace-doctors. 3

[5] Radiological Society of North America. Rsna pneumonia detection challenge (2018). https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pneumonia-detection-challenge-2018. 1

[6] UNICEF. World pneumonia day 2015 infographic. https://data.unicef.org/resources/world-pneumonia-day-2015-infographic/. 1

[7] IEEE Xplore. Grad-cam: Visual explanations from deep networks via gradient based localization. https://ieeexplore.ieee.org/document/8237336. 3