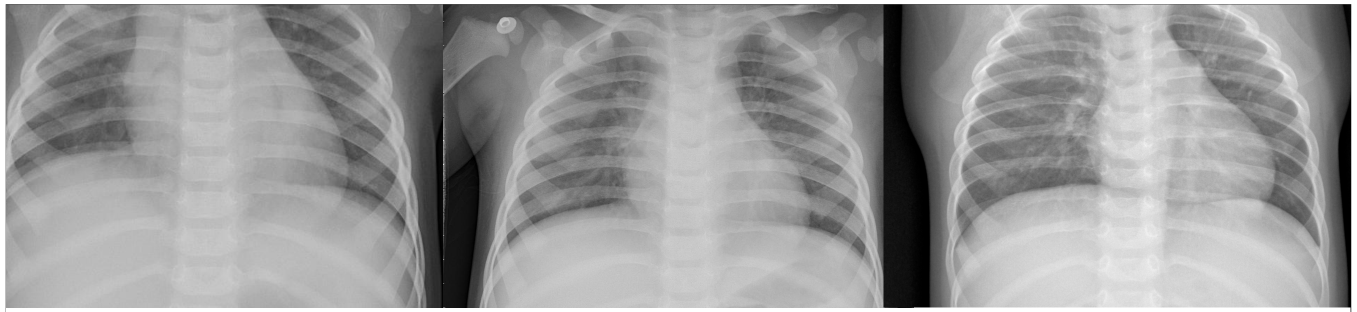
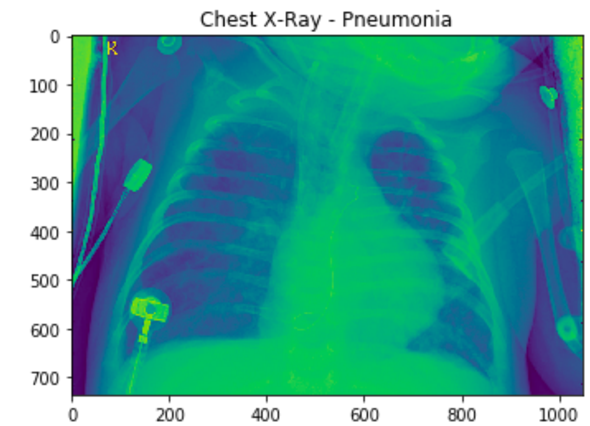


DEEP LEARNING IMAGE CLASSIFICATION

PNEUMONIA DETECTION

PROBLEM STATEMENT

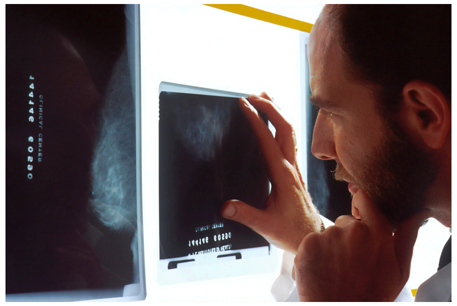
Can machine learning be used to detect pneumonia based on an X-ray image?



Hi, I'm Nadine from Flatiron Data Science team, here to present the results of our proof of concept project to see whether machine learning techniques can be used to classify X-ray images.

PROJECT VALUE

- Lower the risk of human errors
- Improve patient safety
- Minimise diagnosis time
- Allow for a better allocation of resources



Before we dive into the data science process and results, I just wanted to highlighted the benefits of applying machine learning to assist in x-ray image classification.

Photo: https://unsplash.com/photos/v_2FRXEba94

METHODOLOGY

1. Preprocess data
2. Explore and visualise data
3. Train and evaluate models
4. Recommendations

To solve our problem we performed the following steps. In the next slides, we will go through each step one by one.

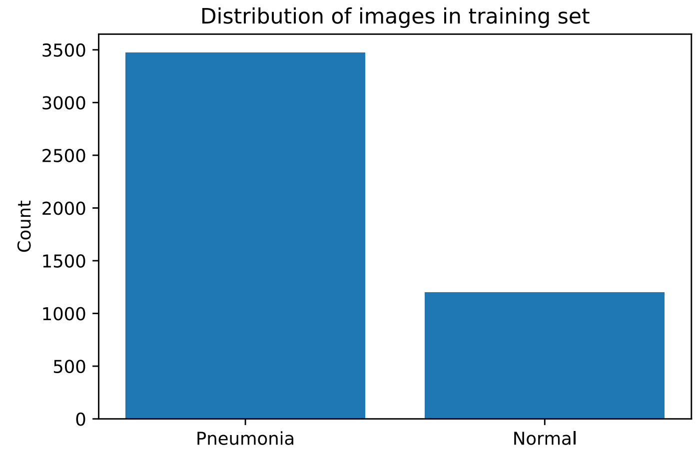
Icons source: from OSEMN framework originally by Hilary Mason and Chris Wiggins

1. PREPROCESSING

- Approx. 5,800 images, pneumonia and normal
- Train, test and validation sets
- Scaling 100 x 100 pixels
- Grayscale

The data gathered contained approximately 5000 chest x-ray images, divided into sick patients diagnosed with pneumonia and healthy normal patient x-rays.
Our first step was to divide this set into a training set, used to build the model, a validation set to evaluate each model and a final test set which shows how our model would perform on unseen data.
We then scaled the images to be 100x100 pixels, lost some aspect ratio and grayscale.
Image source: https://lab.getapp.com/importance-of-data-cleaning-and-governance/

2. EXPLORE AND VISUALISE DATA



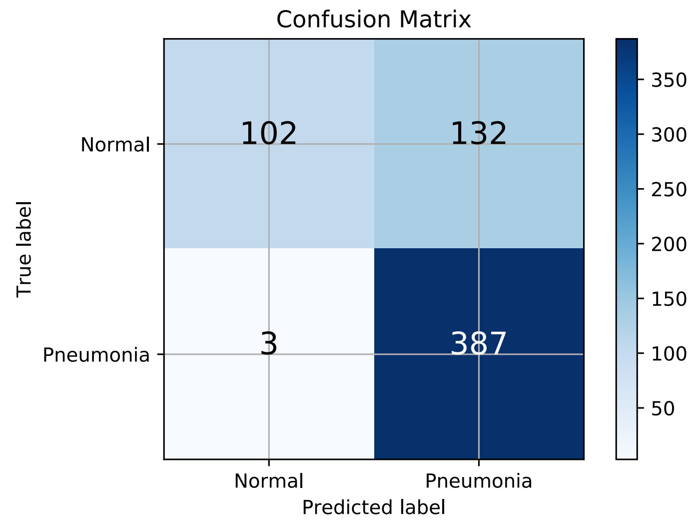
We looked at the distribution of pneumonia and normal x-ray images in our training set. As you can see there is a class imbalance by which I mean there are considerably more images of patients with pneumonia than normal. Whilst this is not initially a problem, it is something to be aware of as our model could achieve an inflated accuracy score by always predicting one response.

3. TRAIN AND EVALUATE MODELS

	Accuracy	F1	Precision	Recall
Baseline Model	0.73	0.85	0.73	1
Basic CNN	0.97	0.98	0.98	0.99
CNN with dropout	0.98	0.99	0.99	0.98
VGG19	0.93	0.96	0.92	0.99

We present in this slide 4 different models and the scores they achieved using various metrics. I'm not going to go into detail on the different scoring metrics, but know that it's expressed as a value from 0 to 1 with 1 being best. In the appendix I have some slides explaining them in more detail which I would happy to go over.
Our chosen model is a CNN which stands for convolution neural network.

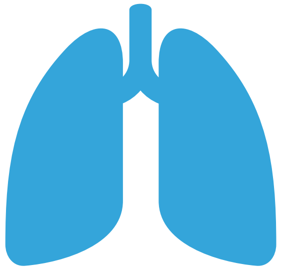
3. TRAIN AND EVALUATE MODELS



Here is a summary of our model's performance on the test set, that is unseen images.
Along the diagonal, 102 is the number of correctly identified normal cases and 387 is the number of correctly identified cases with pneumonia.
We then turn our attention to the miss-classifications. In the top right corner 132 is the number of cases incorrectly predicted as having pneumonia when in fact normal. It may seem quite high but for patient safety it is better for it to be cautious. On the other hand in the bottom left corner, we see that only 3 cases were predicted to be healthy when the patient was sick. We wanted to ensure these were low so no patients are sent home without treatment.

4. RECOMMENDATIONS

- Gather more data
- Use it as a complementary tool
- Consider other use cases

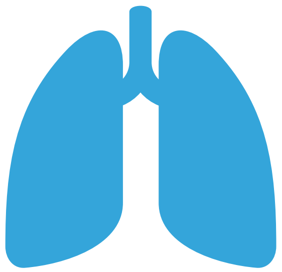


Gather more data - This dataset had only 5k images, which was sufficient for a proof of concept. Additional data will allow for a more robust model.

Apply the model to other x-ray images - Seek to identify COVID-19 or other illnesses using similar means.

FUTURE WORK

- Adjust for class imbalance
- Data augmentation
- More complex models
- Detect COVID-19



Adjust for class imbalance - As we saw in the data exploration section, the number of pneumonia vs. normal x-ray images is weighted in favour of those with pneumonia. We could use resampling techniques to achieve a more even split.
Data augmentation - Once we have a clearer idea of how x-ray images can vary, we can use data augmentation tools to generate new data.
Apply the model to other x-ray images - Seek to identify COVID-19 or other illnesses using similar means.

THAN YOU

NADINE AMERSI-BELTON

nzamersi@gmail.com
datascimum
nadinezab

I hope this presentation has shown you the value gained from applying deep learning to X-ray image classification.

APPENDIX 1 - CONFUSION MATRIX

	Predict No	Predict Yes
Actual No	TN	FP
Actual Yes	FN	TP

We begin with a confusion matrix, which shows all possible outcomes.
In our scenario, a **false negative** occurs when the model predicts that the patient is healthy when in fact the patient has pneumonia. This is hugely detrimental, as we would be mis-diagnosing a sick person and not giving them the treatment they need.
On the other hand, a **false positive** occurs when the model predicts that a patient has pneumonia when they are healthy. This also isn't ideal as additional resources and time would be potentially wasted, however it is much safer.
As such, **false negatives are worse than false positives**.

APPENDIX 2 - METRICS

	Predict No	Predict Yes
Actual No	TN	FP
Actual Yes	FN	TP

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

Recall = TP/(TP + FN)

Precision = TP/(TP + FP)

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

Recall is calculated as the number of correct positive predictions divided by the total number of positives.
Precision (PREC) is calculated as the number of correct positive predictions divided by the total number of positive predictions.