Assisted Pneumonia Diagnosis using Image Classification with Deep Learning

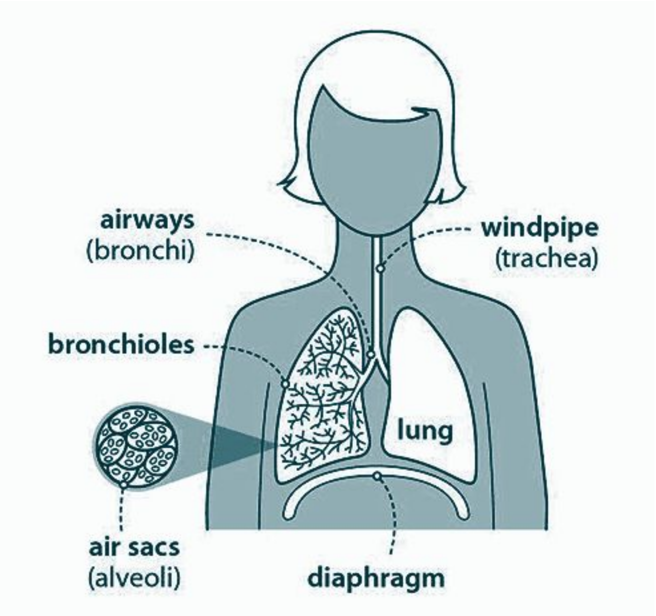
James Toop

Data Science, Module 4 – Final Project

Purpose –
Pneumonia is an acute respiratory infection

Air sacs (alveoli) in the lungs become inflamed, filling with fluid making it hard to breathe and limiting oxygen intake :

- Can be caused by viruses, bacteria, or fungi
- Can be prevented by immunisation, adequate nutrition, and by addressing environmental factors.
- According to the World Health Organisation, pneumonia is the single largest infectious cause of death in children worldwide.

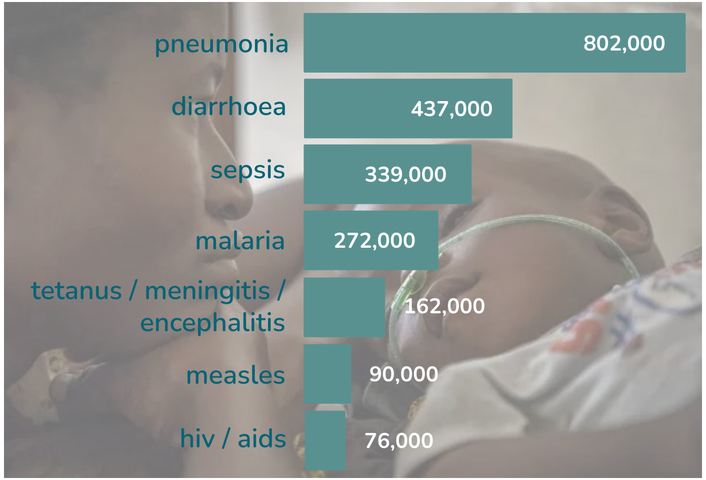


Source: British Lung Foundation, National Health Service (NHS) UK and World Health Organisation (WHO)

Purpose –
A child dies of pneumonia every 39 seconds

Pneumonia kills more children than any other infectious disease :

- 800,000 children under five die from pneumonia every year, that's around 2,200 every day.
- Greatest incidence occurs in South Asia (2,500 cases per 100,000 children) and West and Central Africa (1,620 cases per 100,000 children).
- Almost all of these deaths are preventable.



Deaths of children under 5 by leading infectious diseases, 2018

Sources: WHO Maternal Child Epidemiology Estimation (WHO-MCEE) 2018.

Purpose –
Too few children receive the care they need

An estimated 18 million more health workers are needed by 2030 to prevent, diagnose and treat pneumonia :

- Mortality due to childhood pneumonia is strongly linked to poverty-related factors such as undernutrition, lack of safe drinking water and sanitation, indoor and outdoor air pollution as well as inadequate access to health care.
- Globally, less than two-thirds (62 per cent) of children with symptoms of acute respiratory infection (ARI) are taken to a health care provider.

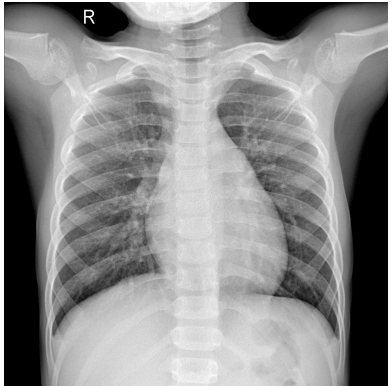
Source: unicef

Objective –

Use Deep Learning techniques to build a model that can classify whether a given paediatric patient has pneumonia, given a chest x-ray image.

Approach –
The raw dataset

Presenting pneumonia
chest x-ray image



"Normal"
chest x-ray image

Source: Kermany et al. on Mendeley / Kaggle

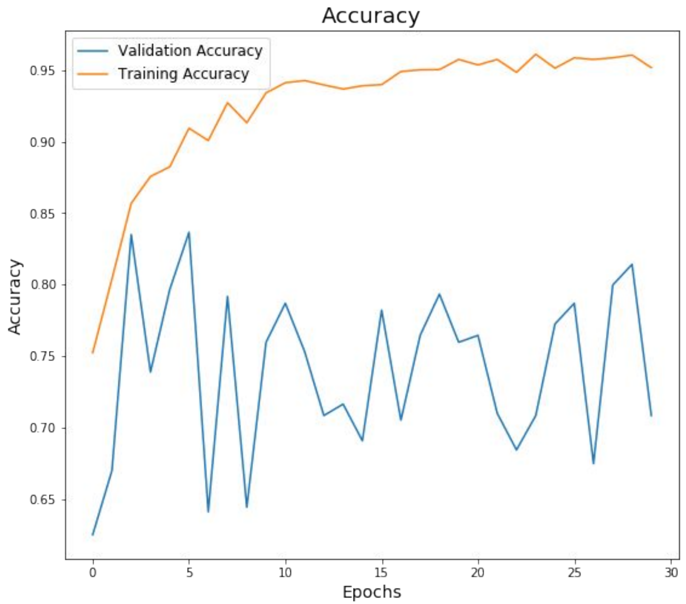
Almost 6,000 paediatric chest x-ray images :

- 5,856 images split into train, test and validation subsets and divided again into those labelled as "normal" and those presenting with pneumonia.
- 1,583 "normal" chest x-rays and 4,273 presenting with pneumonia.
- additional feature of bacterial or viral pneumonia available but unused.
- multiple images of the same patient within the dataset but removal of "duplicates" considered unnecessary for modelling purposes.

Approach –
Preparing the data and creating the models

Iterating through deep learning models, adjusting variables in response to previous model results :

- Rescale RGB values, resize images and create training and validation datasets.
- Create simple, baseline model and observe results.
- Further models created increasing the number of layers, learning rate, activation types etc.
- Final model uses a Convolutional Neural Network and data augmentation to artificially expand the training dataset.

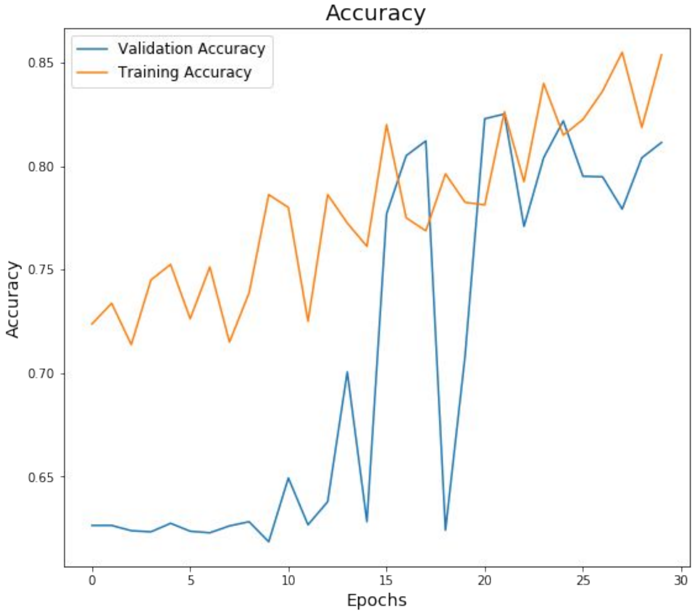


Example output of model accuracy during training

Conclusion –
Accuracy against unseen data an issue

Iterating a deep learning model that (reasonably) accurately identifies pneumonia in a paediatric patient based on a chest x-ray image :

- Initial models did not generalise well and tended to overfit the training data. Subsequent changes to parameters had little or no effect with significant fluctuations in accuracy against validation data observed.
- Augmentation of the training data by flipping the x-ray images vertically struck a better balance between training and validation accuracy.

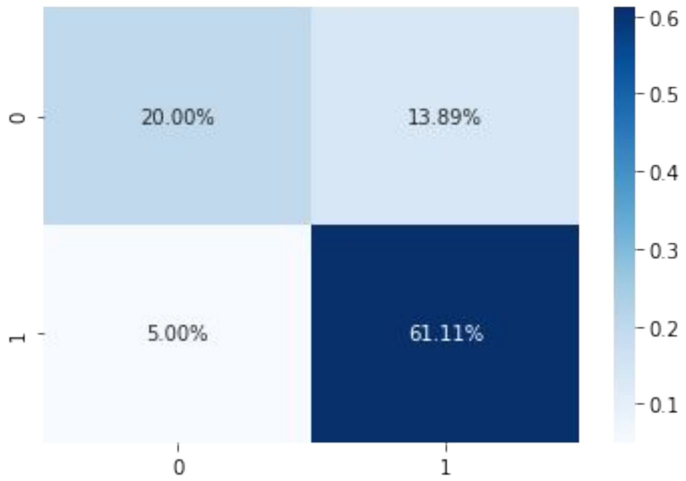


“Final” model accuracy during training

Conclusion –
Accuracy against unseen data an issue

Iterating a deep learning model that (reasonably) accurately identifies pneumonia in a paediatric patient based on a chest x-ray image :

- The final model using a Convolutional Neural Network and data augmentation produced the best results classifying 85% of the “unseen” chest x-ray images, but there is a significant resource overhead when using, taking over an hour to train the model.



Conclusion –
Potential next steps

Recommendations –

- Use weight balancing techniques to balance the dataset more evenly between “normal” chest x-ray images and those presenting with pneumonia.
- Source additional data in the form of further paediatric chest x-ray images and combine with data augmentation when training the model.
- Use callback methods and early stopping to optimise the model.
- Use pre-trained models to generate results with less computational overhead.

Thank you.

Any questions?

Appendix – Model Results

Model Iteration	Model Details	Training Accuracy	Validation Accuracy
1	Baseline model with single layer	96%	76%
2	Deeper model with increased neurons in each layer	96%	72%
3	Deeper model but with a different activation type and reduced number of neurons	96%	71%
4	Using regularization and reducing the learning rate	96%	72%
5	SGD learning rate = 0.001, output activation='softmax'	96%	76%
6	SGD learning rate = 0.001, output activation='sigmoid'	96%	72%
7	SGD learning rate = 0.001, output activation='softmax', batch_size=32	97%	75%
8	Adam learning_rate = 0.001, output activation='softmax', dropout layers added	96%	76%
9	CNN	98%	77%
10	CNN with data augmentation	97%	85%