Convolution Neural Network X-ray Classification

Thinkful Data Science

Corey Coole

Presentation Contents:

- Background Information
- Research Goal
- Dataset Description
- The Model: Convolutional Neural Network
- Data Preprocessing, Augmentation
- Class Imbalance
- Model Training
- Model Testing Results
- Summary and Next Steps

Background Information:

The purpose of this study is to assist in the use and interpretation of chest X-ray images (anterior-posterior), as well as in the clinical referral and treatment of pediatric pneumonia.

Pneumonia kills approximately 2 million children under 5 years old every year. It is consistently estimated as the single leading cause of childhood mortality globally.

The World Health Organization reports that nearly all cases (95%) of new-onset childhood clinical pneumonia occur in developing countries, particularly in Southeast Asia and Africa.*

Bacterial and viral pathogens are the leading causes of pneumonia, but each require very different forms of management.**

*Epidemiology and Etiology of childhood pneumonia: Bulletin of the WHO; 2008

**Respiratory disease and its management: A. McLuckie, 2009

Background Information Continued:

Bacterial pneumonia requires urgent referral for immediate antibiotic treatment, while viral pneumonia is treated with supportive care.

Accurate and timely diagnosis is imperative.

A key element of diagnosis is radiographic data. Chest x-rays are routinely obtained as a standard of care and can help differentiate between unhealthy and healthy patients and further, the differents forms of pneumonia.

However, rapid radiological interpretation of images is not always available, particularly in low-resource settings where childhood pneumonia has the highest incidence and highest rates of mortality.

Research Goal:

Cases of delayed, missed, and incorrect diagnosis are common, with an incidence in the range of 10% to 20%. Some errors in diagnosis stem from mistakes in the interpretation of diagnostic tests. For example, pathology, radiology, and the clinical laboratory each have error rates of 2% to 5%.*

It is the goal of this study to create a robust and precise neural network that facilitates screening programs and supports a more efficient referral system in diagnosis of pneumonia, particularly in remote, or low-resource areas, leading to a broad clinical and public health impact.

*Bringing Diagnosis Into the Quality and Safety Equations: Graber M.D. et al., 2012

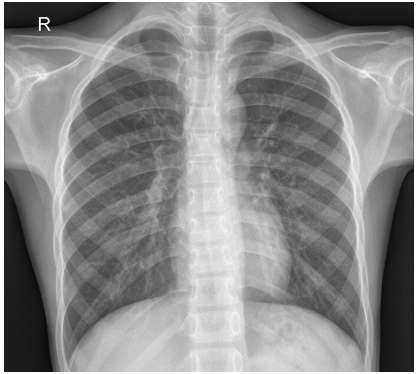
Dataset Description:

5232 pediatric chest x-rays selected from groups of pediatric patients of 1 to 5 years old from Guangzhou Women and Children's Medical Center, China. The x-rays were classified by medical experts and a label assigned to each image. Classifications are either Normal(healthy), or Pneumonia.

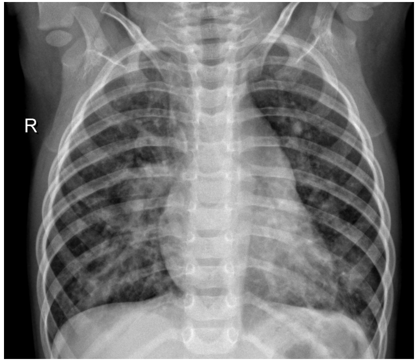
The following analysis asks with the image data provided, if can we classify with reasonable accuracy and precision, the presence of pneumonia in the lungs and what structure of neural network best encompasses the features and target data within the image dataset.

A successful model will be highly accurate with minimal required processing time, as well as minimal difference between the training and testing datasets by resulting type one and type two errors.

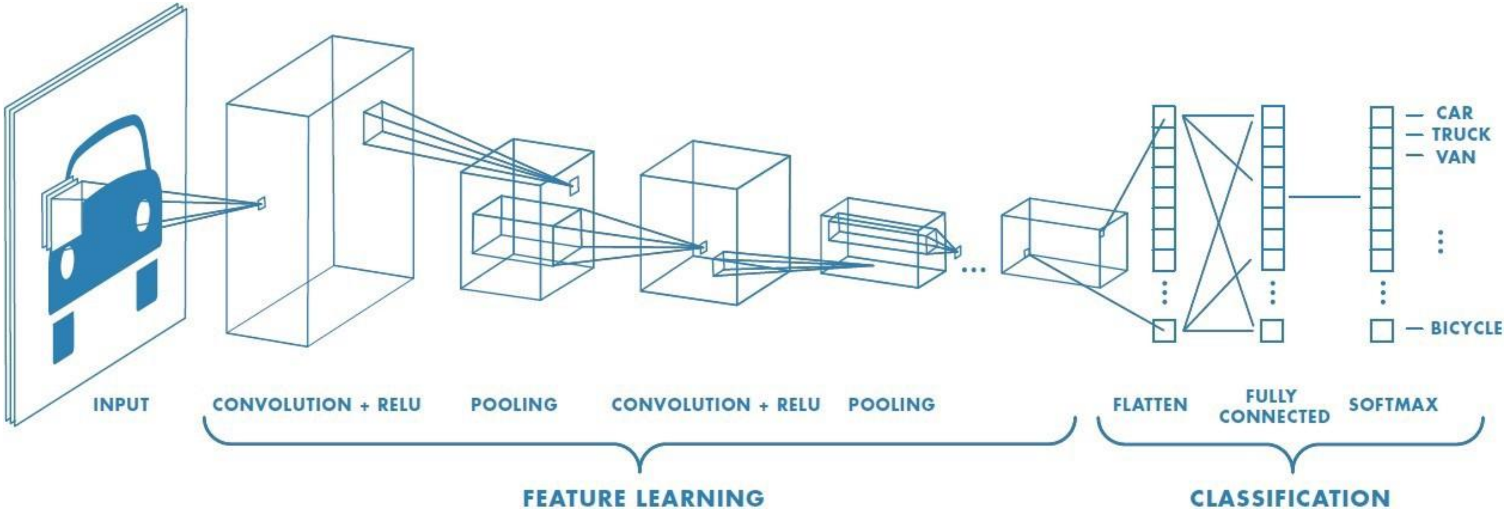
Normal X-ray



Pneumonia X-ray



The Model: Convolutional Neural Network



A Convolutional Neural Network is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.
The architecture of a ConvNet is analogous to that of the connectivity pattern of neurons in the human brain and was inspired by the organization of the visual cortex.

The Model Continued: CNN Structure

The role of the CNN is to reduce the images into a form which is easier to process, without losing features which are critical for returning a good prediction.

Each layer represent a different operation on the nodes of the neural network.

The weights of the neural network are adjusted with respect to gradient descent of error during backpropagation.

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 148, 148, 32)	320
activation_6 (Activation)	(None, 148, 148, 32)	0
max_pooling2d_4 (MaxPooling2	(None, 74, 74, 32)	0
conv2d_5 (Conv2D)	(None, 72, 72, 32)	9248
activation_7 (Activation)	(None, 72, 72, 32)	0
max_pooling2d_5 (MaxPooling2	(None, 36, 36, 32)	0
conv2d_6 (Conv2D)	(None, 34, 34, 64)	18496
activation_8 (Activation)	(None, 34, 34, 64)	0
max_pooling2d_6 (MaxPooling2	(None, 17, 17, 64)	0
flatten_2 (Flatten)	(None, 18496)	0
dense_3 (Dense)	(None, 64)	1183808
activation_9 (Activation)	(None, 64)	0
dropout_2 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 2)	130
activation_10 (Activation)	(None, 2)	0

Data Preprocessing, Augmentation:

Keras Image Data Generator: Generate batches of tensor image data with real-time data augmentation

- **Rescale:** rescaling factor, multiply the data by the value provided to Normalize values to 1 from 0 to 255 (256 values of pixels)
- **Shear Range:** Shear Intensity (Shear angle in counter-clockwise direction in degrees)
- **Zoom Range:** Range for random zoom
- **Horizontal Flip:** Randomly flip inputs horizontally

Images were also resized to a dimension of 150x150, as well as converted to single channel grayscale for less expensive processing.

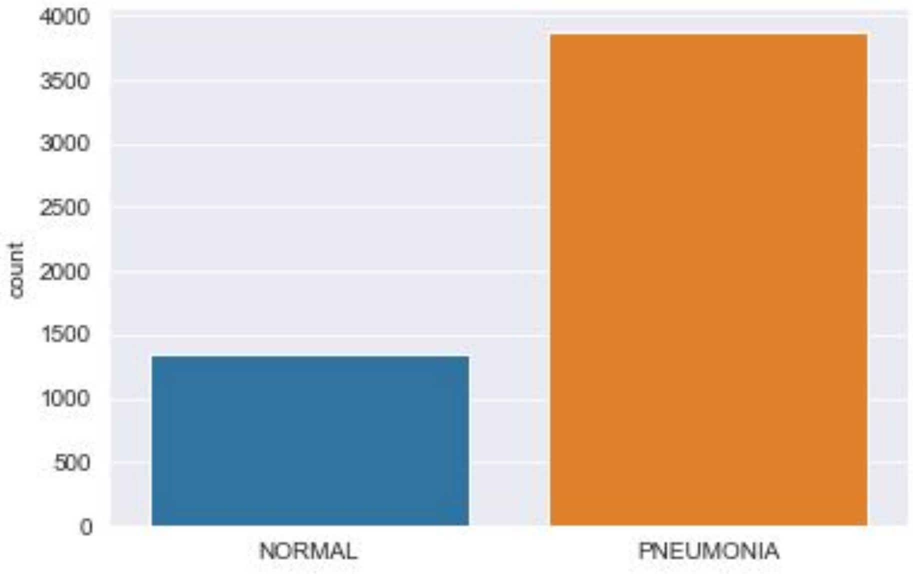
The full dataset was separated into 3 groups for model training(5216 instances), validation(16 instances), and testing(624 instances).

Class Imbalance:

The label data contained in the training group was found to be very unbalanced, 1341 instances of type Normal, 3875 of type Pneumonia.

To correct the imbalance, during model training, weights were applied to the label classifications.

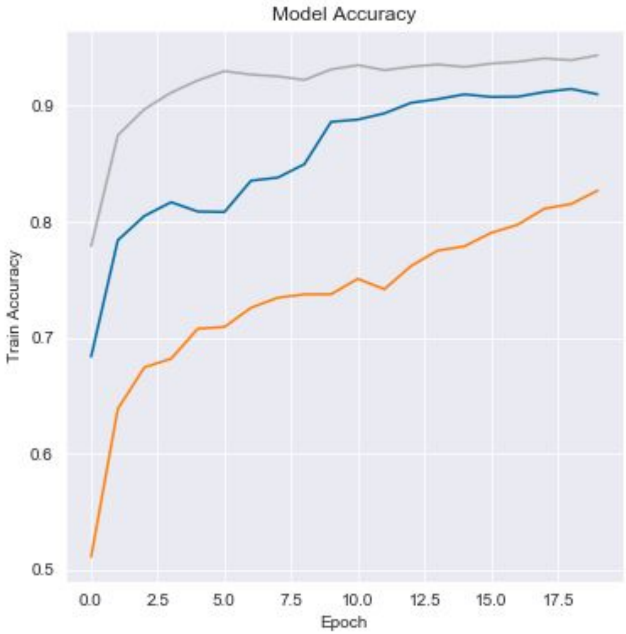
This allowed the model to pay more attention to samples from the under-represented class when it adjusts the model weights during backpropagation.

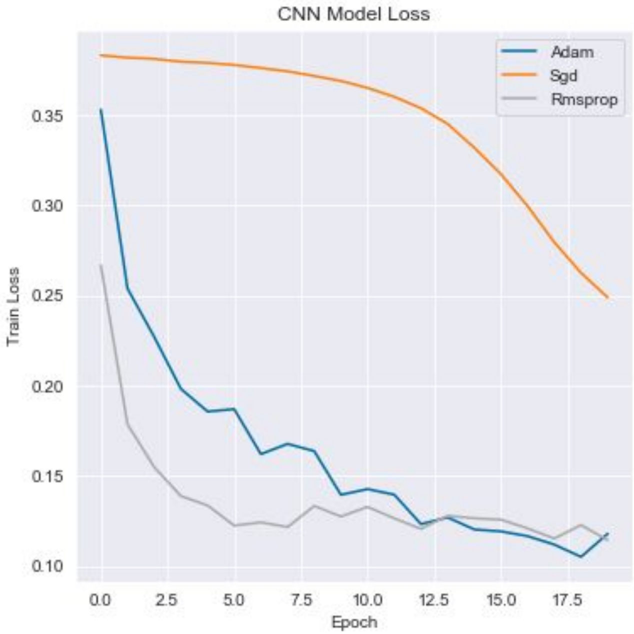


Model Training:

The CNN was trained in 3 separate sessions, each over 20 epochs and using a different optimizing function each session. The Adam, RMSprop, and Stochastic Gradient Descent functions were chosen.

The Model was trained over an average time period of 44.61 minutes per session.





Model Testing Results:

Similar to the training session results, the CNNs with the Adam and RMSprop optimizing functions performed best on the testing dataset.

Although all opt. functions saw an amount of overfitting issues.

Both Adam and RMSprop networks saw accuracy levels on the testing set over .90.

	Train Acc	Test Acc	Train Loss	Test Loss	Train/Test Difference
ADAM	94.56	90.06	0.12	0.40	4.50
SGD	87.61	85.26	0.57	0.59	2.35
RMSPROP	95.76	90.06	0.17	0.38	5.70

Model Testing Results Continued:

Due to the class imbalance, accuracy is not the optimal metric to assess the model's performance.

Looking at the precision, recall, and F1 scores of the optimizer function on the CNN, we see that the Adam and RMSprop iterations performed very similarly.

The clinical context of this study tells us to prioritize a minimal amount of false negatives in model predictive performance, in this way the recall holds the most significance.

	Recall	Precision	F1-score
ADAM	0.90	0.90	0.90
SGD	0.84	0.84	0.84
RMSPROP	0.91	0.90	0.90

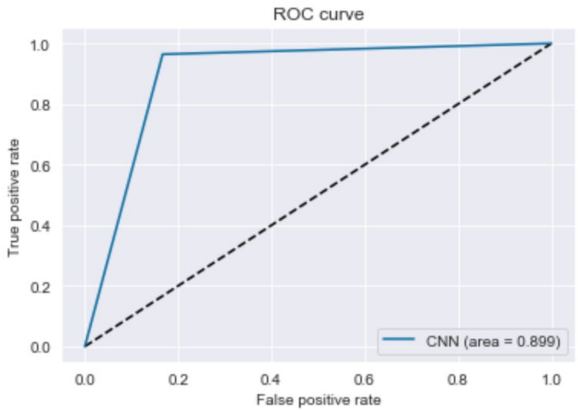
Model Testing Results Continued:

Here we see the depictions of the three network's Receiver Operating Characteristic graphs, as well as the Area Under the Curve scores, respectively.

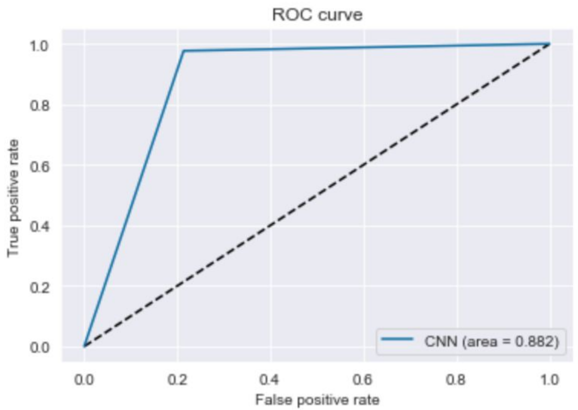
AUC scores reflect the probability that the model is able to distinguish between either target class.

In this way, we see the RMSprop network holds the best predictive magnitude over the other two models at .899.

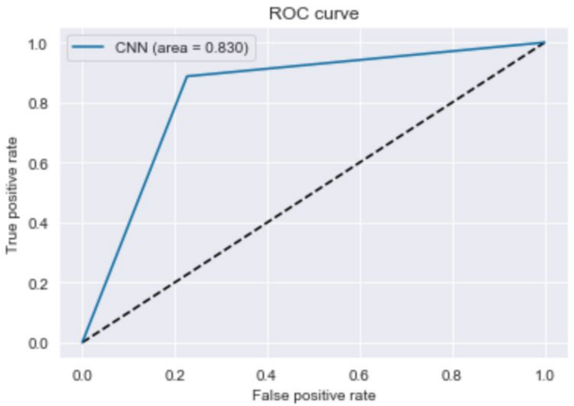
- RMSprop AUC: .899



- Adam AUC: .882



- SGD AUC: .83



Summary and Next Steps:

In final analysis, all three optimizers performed well on the CNN with respect to the recall score. However, using the RMSprop optimizing function returned the best results in terms of recall and AUC metrics.

I am confident that more investigation of the hyperparameters, increased dropout rate, and longer training cycles will result in a well-rounded CNN that will reach levels of classification accuracy equal to a medical expert.

A robust and precise neural network would facilitate screening programs and provide the opportunity to engage in a more efficient referral system in diagnosis of pneumonia, particularly in remote, or low-resource areas, leading to a broad clinical and public health impact.

Notebook link:

https://github.com/coreycoole/Final-Capstone-TensorFlow-Keras