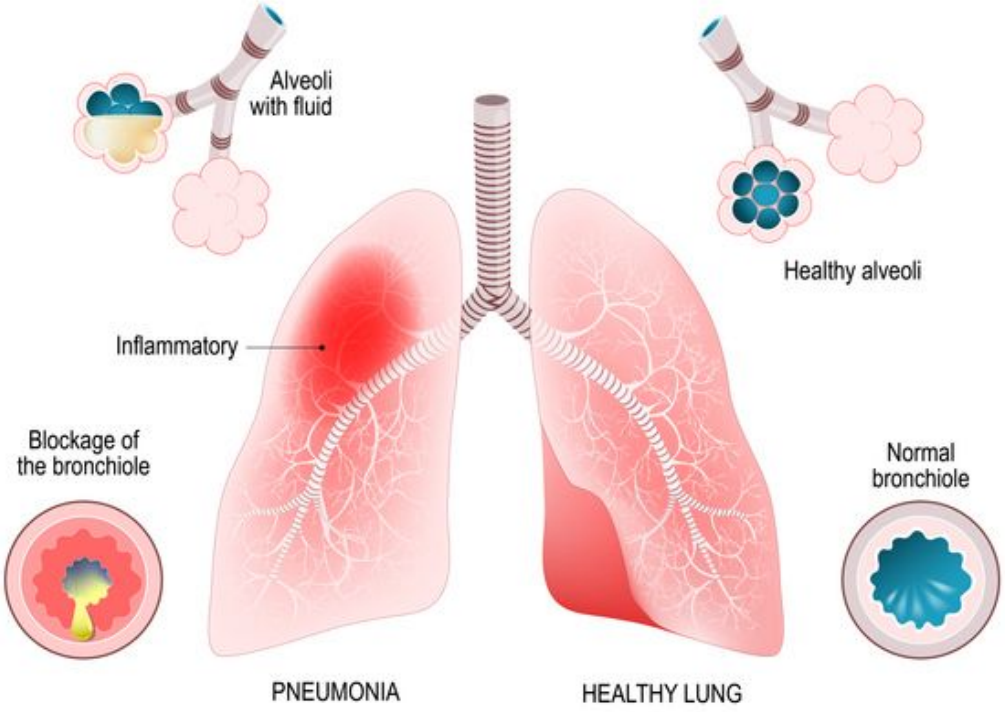
A Pneumonia Chest x-ray Detection Model

By John Mark Ang’awa

Overview

- Introduction
- Business understanding
- Project goal
- Business problem(s) solved
- Scope of the project
- Business problem
- Data understanding
- Data preparation
- Rationale
- Data analysis process
- Modeling
- Results
- Conclusion
- Visualization after analysis
- Recommendations
- Next steps

Pneumonia


Alveoli
with fluid
Healthy alveoli
Inflammatory
Blockage of
the bronchiole
Normal
bronchiole
PNEUMONIA
HEALTHY LUNG

Introduction

With change in time and technology, healthcare has drastically changed over the time. Doctors and Radiologists at JMarks Hospital have been affected by these changes hence compromising the level of efficiency of healthcare delivery in the hospital. Screening of thousands of patients for chest x-rays to determine the diagnosis for pneumonia has increasingly become tiresome and costly. Moreover, interpretation of chest x-ray images by doctors is limited to accuracy since its a judgement made by naked human eyes. The hospital administration is therefore seeking to mitigate this matter by building a system or tool that can assist doctors and radiologists in diagnosing a patient for pneumonia, but the lack knowledge on how to do that.

* N\B: One of the key questions that arises concerns the ability of a system to correctly work in order to ensure that the quality of healthcare being delivered meets the expectations of the doctors, radiologists and the patients receiving services.

What is Pneumonia?

It’s an inflammatory condition of the lung affecting primarily the small air sacs known as alveoli.Symptoms typically include some combination of productive or dry cough, chest pain, fever and difficulty breathing. The severity of the condition is variable. Pneumonia is usually caused by infection with viruses or bacteria and less commonly by other microorganisms, certain medications or conditions such as autoimmune diseases.Risk factors include cystic fibrosis, chronic obstructive pulmonary disease (COPD), asthma, diabetes, heart failure, a history of smoking, a poor ability to cough such as following a stroke and a weak immune system. Diagnosis is often based on symptoms and physical examination. Chest X-ray, blood tests, and culture of the sputum may help confirm the diagnosis.The disease may be classified by where it was acquired, such as community- or hospital-acquired or healthcare-associated pneumonia.

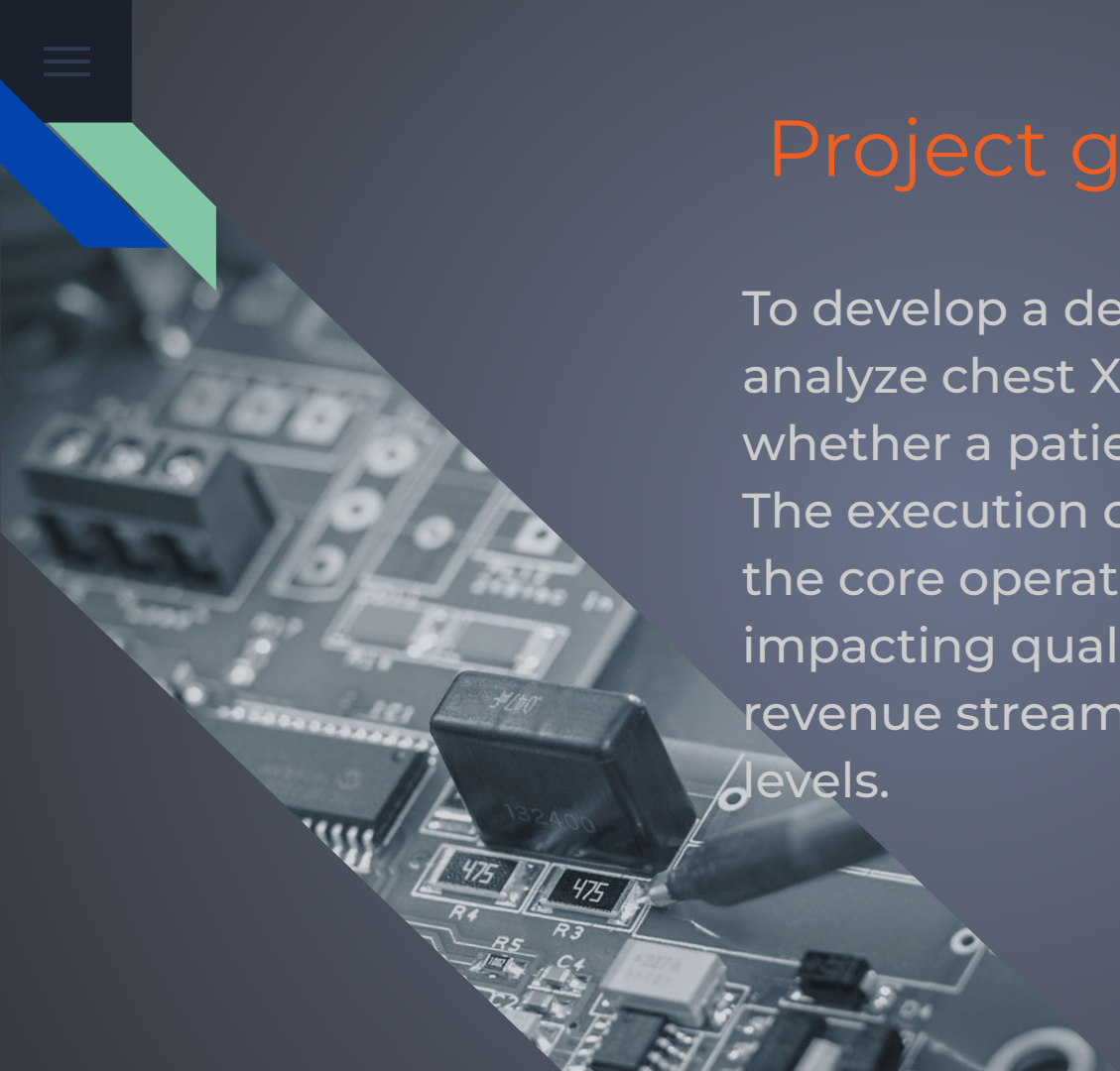
Business understanding

This project aims to develop a deep learning model that can analyze chest X-ray images and classify whether a patient has pneumonia. This classification can help doctors and healthcare providers at JMarkss Hospital identify cases of pneumonia more quickly and accurately, allowing for prompt treatment and improving patient outcomes.

The success of this project will be of great use to different JMarkss Hospital stakeholders:

* JMarkss Hospital Management: As the ultimate decision-makers, they're vested in the project's outcomes for improved healthcare delivery.
* Doctors, Radiologists and healthcare professionals: They are the primary users of this model in their daily routine.
* Data Science Team (Project Team): Responsible for executing the project, analyzing data, and creating classification models.

Project goal



To develop a deep learning model that can analyze chest X-ray images and classify whether a patient has pneumonia.
The execution of this project directly affects the core operations of JMarkss Hospital, impacting quality of healthcare delivery, revenue streams, and customer satisfaction levels.

Business Problem(s) Solved:

This Data Science endeavor addresses the critical business problem of pneumonia chest x-ray detection. It aims to identify patients with pneumonia and those without to improve efficiency and quality of healthcare system. In this sense, the research questions are:

* Is the model correcty working?

* What is the accuracy level of the model?

* How much time and resources is reduced by using this model?

Scope of the Project:

Inside the project's scope are the following components:

* Pneumonia Prediction: Developing an Image classification models to detect a sick patient and a healthy one.
* Correct Classification: Correctly classifying chest x-ray images.
* Recommendations: Offering actionable suggestions to curb pneumonia problem in the hospital and the community at large.

Business problem

The primary challenge lies in the correct classification of images since a patient's wellbeing is a life sensitive matter which is import to the stakeholders as well.

Benefits

By successfully addressing the challenge of correctly classifying chest x-ray images, this project stands to gain several benefits:

* Enhanced Patient Satisfaction: by the patients and doctors.
* Optimized Resource Allocation: this will enable optimizing operational efficiency.
* Business Sustainability: the hospital's quality of care is improved and its reputation built which a solid foundation for long-term growth.

Data understanding

The dataset is sourced from Kaggle, [document here](https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia) which provides a large collection of chest X-ray images labeled with pneumonia or non-pneumonia. This dataset is suitable for the project as it contains real-world medical images necessary for training a pneumonia classification model.

This project encompasses, machine learning model development, and the interpretation of model results. It involves understanding inner workings of the model to make correct predictions and make sure that the model w

Data preparation

Importing of relevant packages for use in the analysis process, loading and organizing the dataset, explore the data, preprocess it and data augmentation to ensure proper evaluation of deep learning models.

Rationale of the analysis

While the project tackles the formidable challenge of image classification, certain aspects lie beyond its immediate purview. Such as, the implementation of recommended strategies to running the model and the evaluation of the financial impact arising from the project's outcomes is a distinct consideration.

The data analysis process

Here, I performed data augmentation and preprocessing. The modeling phase involved an iterative approach, starting with a simple baseline model and then introducing more complex model, a Convolutional Neural Networks (CNNs). Visualizing intermediate activations followed, evaluation then prediction.

Modeling

Building baseline model- a Densely Connected Neural Network.

Iterating approach- a Convolution Neural Network.

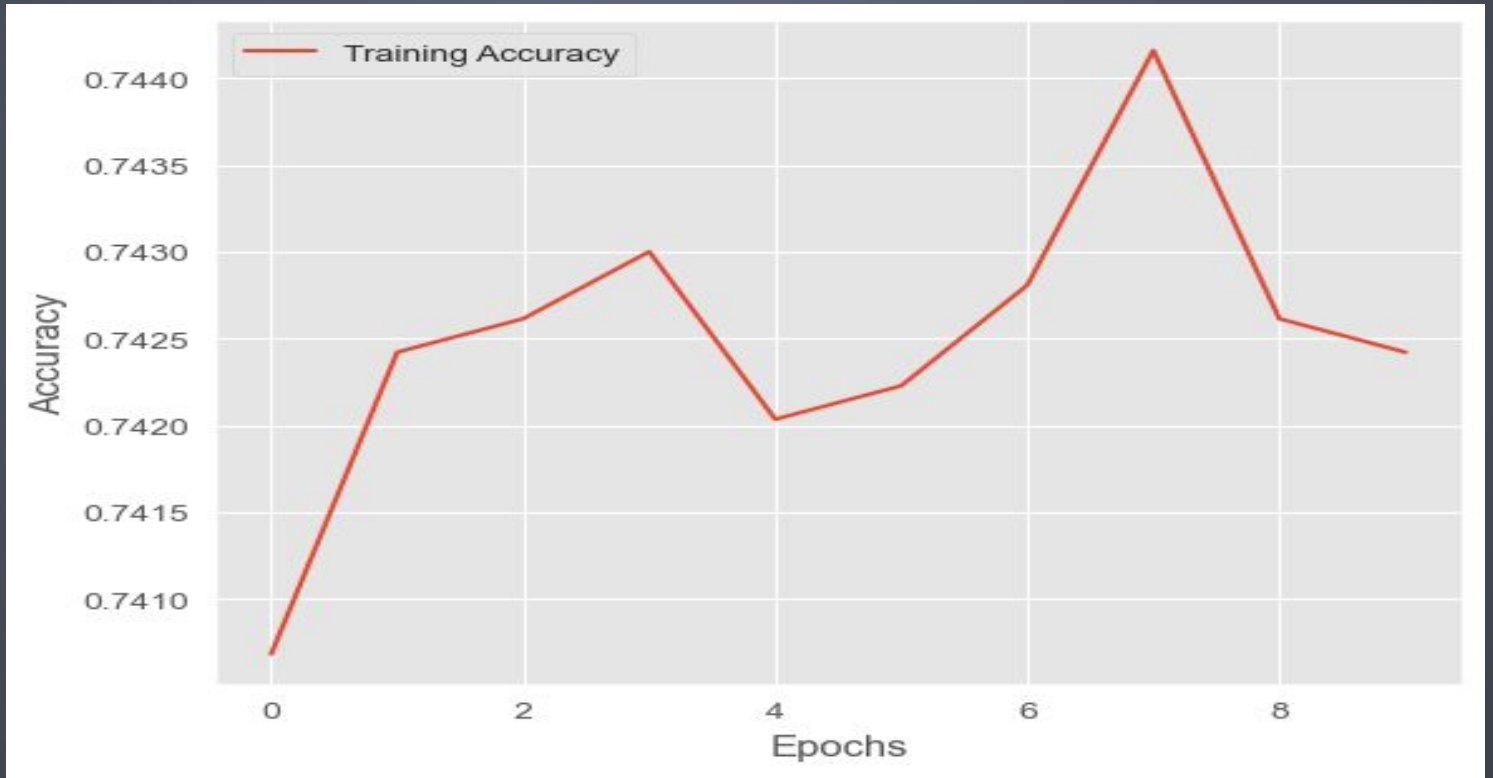
Visualizing intermediate activations.

Results:

1. Baseline model interpretation

- The fitted baseline model has an accuracy level of about 74 percent from ten epochs on the training data. This means that it can overally make correct predictons of about 74 percent on the x-ray images. The loss figures were relatively high with about 59 percent. We should also note that a learning rate was used here.
- The model explains an accuracy of about 62 percent on the testing data on the overal baseline model. This is a fair figure. We can introduce a second model and compare it to this model and see the performance on the data
- The model has a loss of about 67 percent on the training data which might be a bit of concern.

From the graph, the training accuracy is overly high as the number of epochs increase, even though it flops at certain points. This shows that the model is performing well overall.



2. Iterated model- CNN:

- The second model, convolution neural network, shows an improved result with about 89 percent accuracy o training data which is an improvement from the baseline model. This shows that the CNN model is better than the baseline model and can be deployed for prediction and classification of the problem at hand.
- The model has an average accuracy of about 75 percent on the evaluated testing data. This is a fair figure indicating that the model can overally works and can make predictions. The model also recorded a test loss of about 67 percent which is a bit of a high figure.

3. Final model results- from visualization

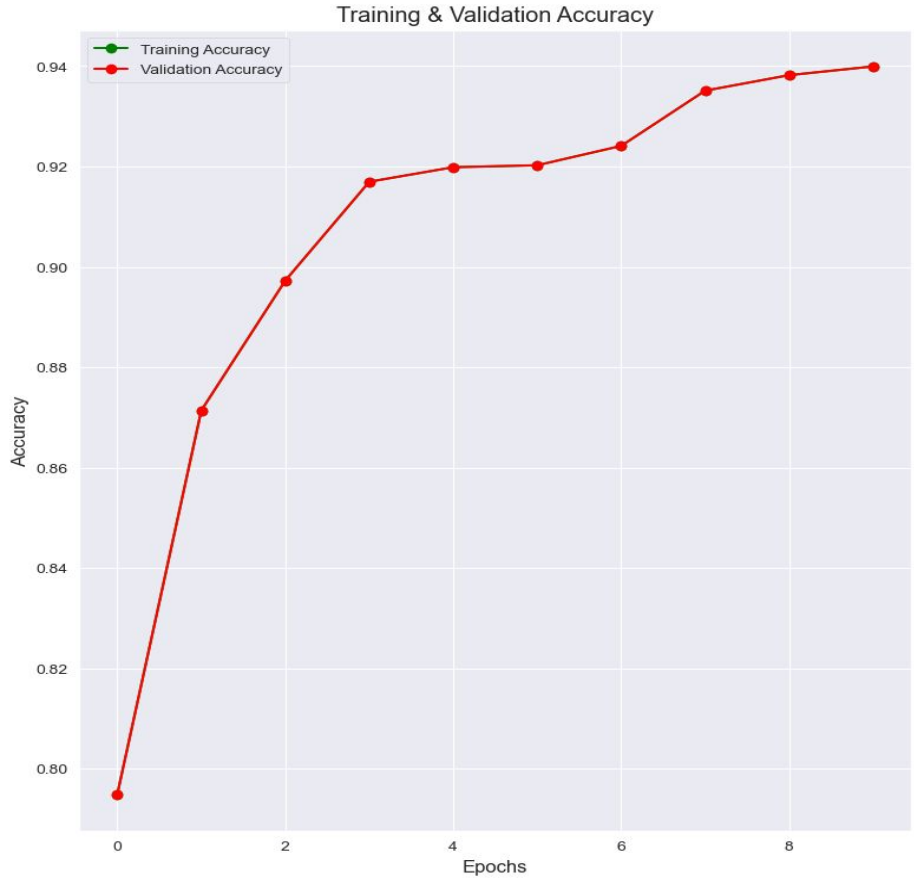
- The model is performing on average as from the classification report which shows an accuracy level of about 62 percent on the test data and a weighted average of about 54 percent in the test set.
- The model has an average accuracy of about 75 percent on the evaluated testing data. This is a fair figure indicating that the model can overally works and can make predictions. The model also recorded a test loss of about 25 percent which is a low figure indicating that the model generalized well to the data.
- The visualizations also show an upward trend where the accuracy was high and the loss rate was reduced indicating the CNN model generalized well to the data.

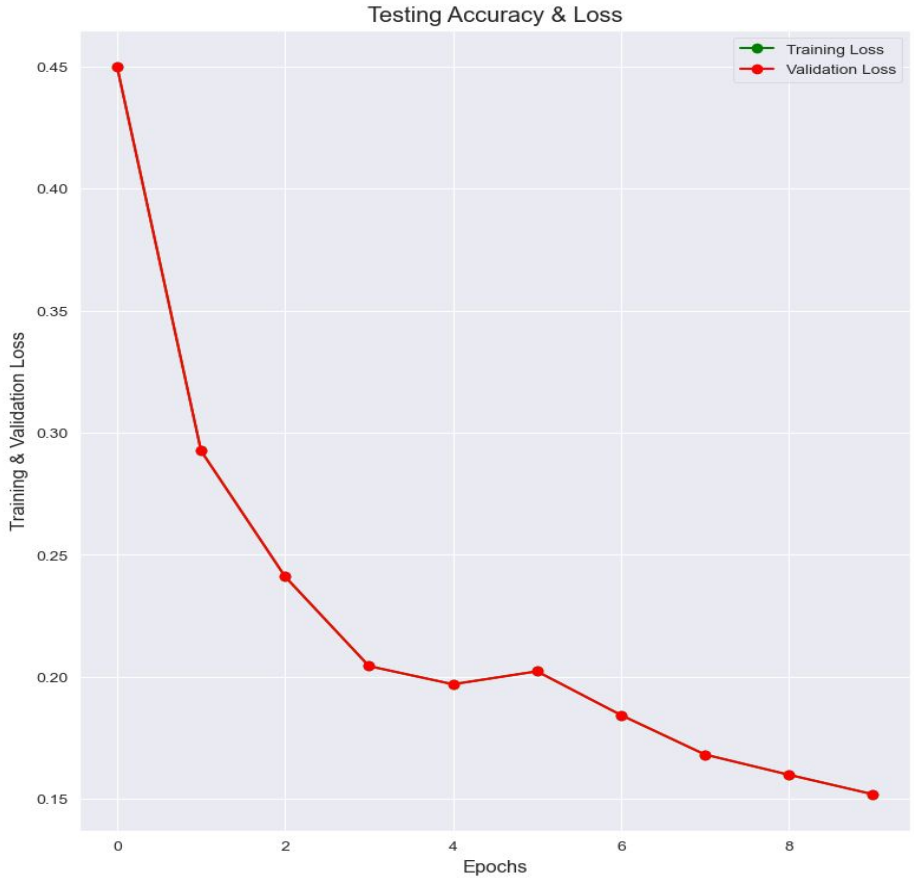
Conclusion

In conclusion, deploying the predictive models for chest x-ray image classification can significantly benefit JMarks Hospital by enabling them to correctly identify patients with pneumonia condition. The model should be regularly monitored, updated, and integrated into business processes for effective decision-making and improved healthcare delivery process.

- Model Performance: In general, the model is performing on average and is able to make predictions. The convolution neural network as the final model showed an improved performance and accuracy on the training and testing datasets.
- Usability of the model: The model is able to make correct predictions based on the evaluation report. The healthcare professionals using this model should continue keep track of its performance and provide continuous feedback

Visualisation after analysis





Recommendations

- Model Selection: Based on the model evaluations, we recommend deploying convolution neural network model which showed impressive results on the metrics.
- Data Collection: To further enhance the models, consider collecting additional data that might have predictive power. This could include more detailed x-ray images and other behavioral attributes.
- Regular Updates: The CNN prediction model should be updated regularly with new data. This ensures that the model remain relevant and effective in capturing changing behavior and preferences.
- Feedback Loop: Continuously monitor the performance of deployed models and gather feedback from business stakeholders. This feedback loop can help refine the models and identify areas for improvement.
- A/B Testing: Implement A/B testing to validate the effectiveness of different strategies for improving classification. Test new initiatives on a subset of images and compare the results to those predicted by the models.

Next steps

To improving the model, consider things like;

1. Collecting more diverse data.
2. Highlighting ethical considerations, privacy concerns, and regulatory compliance in implementing the model.
3. Exploring transfer learning from pre-trained models.

Thank you