BUDT751 Harnessing AI For Business

Final Project Report
AI-Driven Pneumonia Detection

Prasham Shah

1.0 Introduction

Pneumonia is a serious respiratory infection and a leading cause of death in children under the age of five, especially in low- and middle-income countries. Despite advancements in medicine, diagnosing pneumonia accurately remains a significant challenge, often relying on the interpretation of chest X-rays by trained radiologists. This task is inherently complex, time-sensitive, and heavily reliant on human expertise. In many regions around the world, there is a critical shortage of radiology professionals, leading to diagnostic delays and, in some cases, incorrect or missed diagnoses. These issues can result in poor treatment outcomes and increased mortality, especially among vulnerable populations. The role of AI in healthcare is rapidly evolving, with data science offering powerful tools to assist in medical decision-making. This project aims to develop an AI-powered assistant that can accurately detect pneumonia from chest X-rays. By integrating this solution into clinical workflows, we can reduce the burden on radiologists, increase diagnostic accuracy, and improve the speed at which patients receive treatment, ultimately saving lives and optimizing healthcare delivery.

2.0 General Overview of Role and Business

In the healthcare industry, early and accurate disease diagnosis is crucial for effective treatment, especially in critical cases such as pneumonia. Traditionally, diagnosing pneumonia requires a trained radiologist to examine chest X-rays, a time-consuming and complex task prone to human error due to visual fatigue, variability in interpretation, and resource constraints.

This project is the intersection of healthcare and data science, where artificial intelligence (AI) and machine learning (ML) are helping diagnostic support systems. Within this context, data scientists play a vital role in transforming raw medical imaging data into actionable insights through the development of predictive models.

By leveraging AI, specifically deep learning techniques using convolutional neural networks (CNNs), data scientists can automate the identification of pneumonia from chest X-ray images. This not only accelerates the diagnostic process but also enhances accuracy, consistency, and accessibility of care, especially in regions facing shortages of medical professionals.

The goal of this project is to build an AI-driven diagnostic assistant that can support medical professionals in the early detection of pneumonia. The system aims to function as a clinical decision support tool that complements radiologist expertise, reduces diagnostic delays, and helps deliver expert-level diagnostic capability to underserved areas.

3.0 Methods and Methodologies

3.1 Job Functions of Data Scientists in Healthcare AI

For AI-driven medical diagnostics, data scientists serve as the backbone of technological innovation. Their responsibilities are comprehensive and multidisciplinary, involving:

- Data Acquisition and Preprocessing: Data scientists gather large volumes of medical imaging data (e.g., chest X-rays) from open-source datasets or healthcare partners. They perform preprocessing tasks like image normalization, resizing, contrast enhancement, and data augmentation to improve model robustness.
- Model Development and Training: They select appropriate deep learning architectures and use transfer learning techniques to adapt pretrained models for pneumonia detection. They also manage model training, hyperparameter tuning, and optimization.
- Performance Evaluation: Data scientists use performance metrics such as accuracy, to validate the model's effectiveness. They also analyze false positives and false negatives to refine model behavior.
- Interpretability and Explainability: Since clinical decisions require transparency, data scientists need to provide visual explanations of model predictions, helping medical professionals trust and understand AI outputs.
- Collaboration with Domain Experts: Data scientists work closely with radiologists to validate model predictions, refine labeling processes, and ensure clinical relevance of the AI model.

3.2 AI’s Impact on the Data Scientist Role

The growing use of artificial intelligence in healthcare has changed the role of data scientists in important ways. Instead of focusing only on building accurate models, data scientists are now expected to develop solutions that are useful, safe, and trusted in real medical settings.

In healthcare, decisions made by AI can affect patient outcomes, so data scientists must ensure their models are not just accurate but also fair, explainable, and reliable. This includes being aware of biases, protecting patient privacy, and following regulations like HIPAA.

Data scientists now work more closely with doctors and healthcare staff to understand how their models will be used in real-life diagnosis and treatment. This teamwork helps make sure the AI adds real value to medical decision-making.

In addition, data scientists need to explain their work to people who may not have a technical background. This means they must be able to clearly communicate how the model works and what its predictions mean.

Overall, AI has made the data scientist’s role more important and more complex. They are not just building technology, they are helping shape the future of healthcare by creating tools that support better and faster medical decisions.

3.3 Future Technologies

Several emerging tools are shaping the future of AI in healthcare. Data scientists are beginning to use self-supervised learning to train models with less labeled data, and federated learning to work with sensitive data across institutions while maintaining privacy.

Explainable AI (XAI) methods like Grad-CAM are helping build trust by making model decisions more transparent. Synthetic data tools, such as GANs, generate new medical images to improve training without risking patient privacy.

Lastly, AutoML and edge AI are making it easier to deploy models quickly and in low-resource settings, expanding access to AI-driven healthcare worldwide.

4.0 Product Overview

This project delivers an AI-powered pneumonia detection system designed to assist healthcare professionals in rapidly and accurately diagnosing pneumonia from chest X-ray images. The core features include:

- AI-Driven Pneumonia Detection: Leveraging convolutional neural networks (CNNs), the system automatically identifies signs of pneumonia with high accuracy, reducing reliance on manual radiological interpretation.
- User-Friendly Visualization via Streamlit: The product features an interactive web application built using Streamlit, enabling users to upload patient x-rays and view AI-generated pneumonia detection results instantly.
- Fast and Reliable Diagnosis: The AI model is optimized for both speed and accuracy, supporting timely clinical decisions in various healthcare settings.
- Designed for Clinical Use and Accessibility: The tool integrates smoothly into clinical workflows and is optimized for deployment in resource-limited environments to improve diagnostic access globally.

5.0 Technical Implementation

The development process included data collection from public sources, thorough preprocessing, robust image augmentation, and model training using both a baseline CNN model and a transfer learning approach with VGG16 (Pretrained model with 138 million parameters for image classification). These steps were carefully chosen to ensure clinical reliability, scalability, and business relevance in healthcare environments.

5.1 Data Collection and Preprocessing

We worked with high-resolution frontal chest X-rays from several publicly available datasets

- Zhejiang University/Kaggle Pneumonia Dataset (~5,800 images)
- Mendeley Pneumonia Dataset (thousands of annotated samples)

Each image was labeled as either “Pneumonia” or “Normal,” based on clinical findings.

Preprocessing Steps:

- Converted all images to RGB format
- Resized to 512×512 pixels for consistency
- Pixel values normalized to [0, 1]
- Partitioned into training, validation, and test sets with an 80/20 validation split

5.2 Image Augmentation

To reduce overfitting and simulate real-world variations in chest X-rays, we applied data augmentation techniques using ImageDataGenerator. These transformations help improve the model's ability to generalize to new data.

The following augmentations were applied to training images:

- Rotation: +- 15 degrees
- Width and height shifts: up to 10%
- Shearing and zooming: +- 10%
- Horizontal flipping: enabled
- Pixel normalization: rescale=1./255

5.3 Implementation

We developed and evaluated two deep learning models to detect pneumonia from chest X-ray images. The implementation process involved building a custom CNN model as a baseline and then applying transfer learning using a pretrained VGG16 model to improve accuracy. We also developed a user-friendly interface using Streamlit to make the system accessible to non-technical healthcare professionals.

5.3.1 Model Development and Comparison

We first created a **baseline CNN model** consisting of three convolutional layers, each followed by batch normalization and max pooling. After the convolutional layers, we used global average pooling to reduce the spatial dimensions, followed by a dense layer with 128 neurons and a dropout layer (with a rate of 0.5) to prevent overfitting. The final output layer used a sigmoid activation function to perform binary classification—predicting whether pneumonia is present or not.
This baseline model was trained using binary cross-entropy as the loss function and the Adam optimizer with a learning rate of 0.0001. We tracked model performance using accuracy as the primary metric. To avoid overfitting and save the best-performing model, we implemented callbacks such as EarlyStopping and ModelCheckpoint. The model was trained for five epochs.

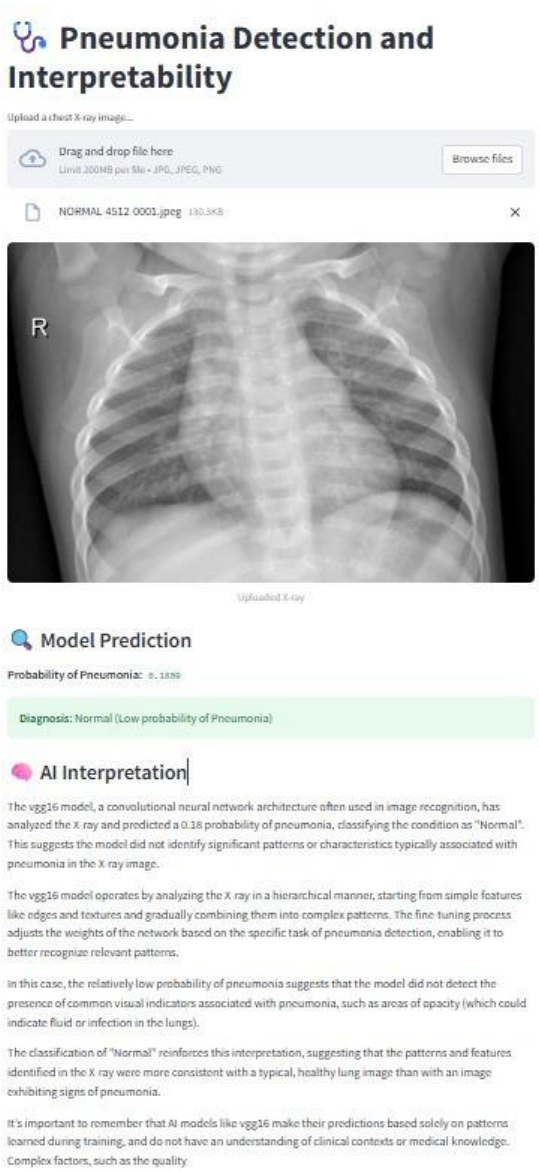
To improve performance further, we used **transfer learning with the VGG16 model**, which was pretrained on the large-scale ImageNet dataset. The convolutional base of VGG16 was frozen to retain its pre-learned weights. On top of it, we added a flatten layer, a dense layer with 256 units using ReLU activation, a dropout layer (0.5), and a final sigmoid output layer for prediction.
The VGG16-based model used the same training settings as the baseline model. This approach significantly improved the accuracy by leveraging rich visual features learned from a much larger dataset, making it more robust and reliable for medical image classification.

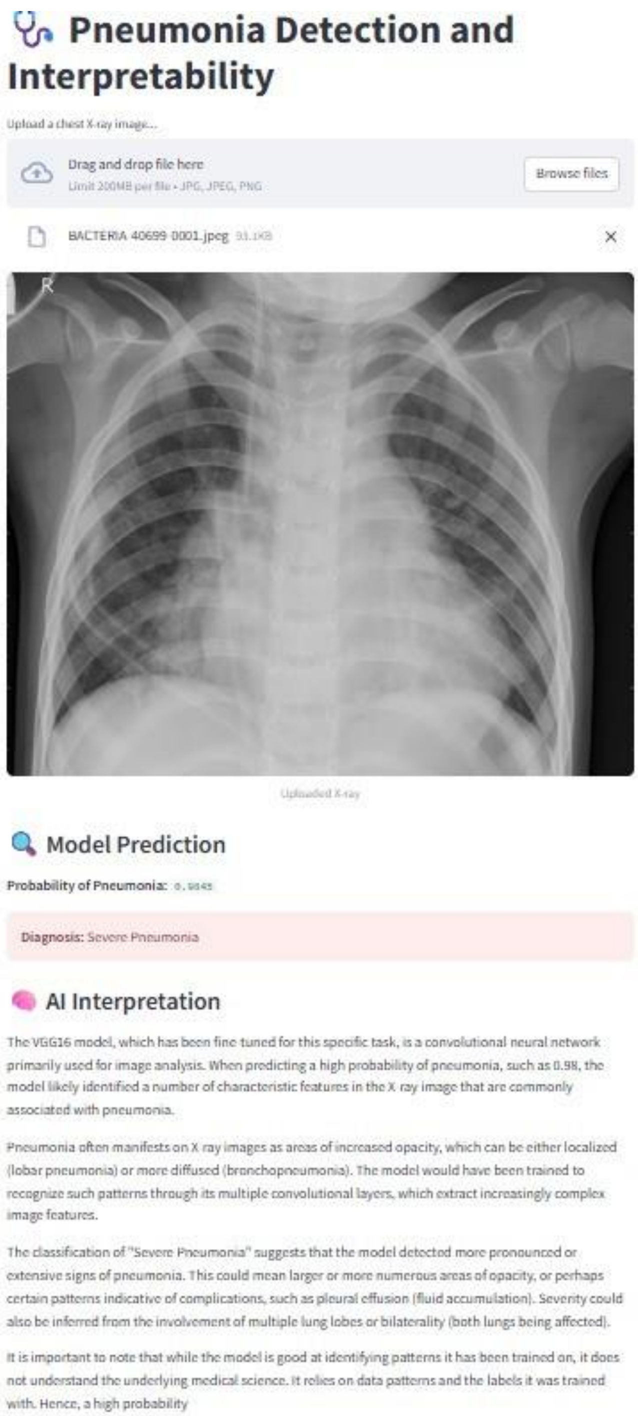
5.3.2 User Interface with Streamlit

To make the model usable in real-world healthcare settings, we built a simple and interactive **Streamlit UI**. This allows users, such as doctors, radiologists, or medical assistants, to upload chest X-ray images directly through the browser interface. Once an image is uploaded, the trained model automatically processes it and displays a prediction indicating whether the image suggests **pneumonia** or **normal** condition.

Additionally, we integrated ChatGPT to provide an **AI-driven summary** that interprets the model's results, offering insights into why it classified the image as healthy or indicative of pneumonia.

Sample model prediction on a held-out test case where the **subject was genuinely healthy.**





Sample model prediction on a held-out test case of a patient **confirmed to have pneumonia.**

5.3.3 Business Relevance

This end-to-end solution, from data preparation and model development to UI generation via Streamlit, provides an efficient, scalable, and cost-effective way to assist in pneumonia diagnosis. It addresses the shortage of radiologists, speeds up diagnosis, and reduces the chances of human error. Most importantly, it helps bring expert-level diagnostic support to under-resourced healthcare environments.

5.4 Model Performance and Optimization

Baseline CNN Model:

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 510, 510, 32)	896
batch_normalization (BatchNormalization)	(None, 510, 510, 32)	128
max_pooling2d (MaxPooling2D)	(None, 255, 255, 32)	0
conv2d_1 (Conv2D)	(None, 253, 253, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 253, 253, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 126, 126, 64)	0
conv2d_2 (Conv2D)	(None, 124, 124, 128)	73,856
batch_normalization_2 (BatchNormalization)	(None, 124, 124, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 62, 62, 128)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 128)	0
dense (Dense)	(None, 128)	16,512
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Total params: 110,785 (432.75 KB)
Trainable params: 110,337 (431.00 KB)
Non-trainable params: 448 (1.75 KB)

```
Classification Report:
              precision    recall  f1-score   support

      Normal       0.00      0.00      0.00       234
   Pneumonia       0.62      1.00      0.77       390

    accuracy                           0.62       624
   macro avg       0.31      0.50      0.38       624
weighted avg       0.39      0.62      0.48       624
```

Transfer learning with the VGG16 model:

Model: "sequential_1"

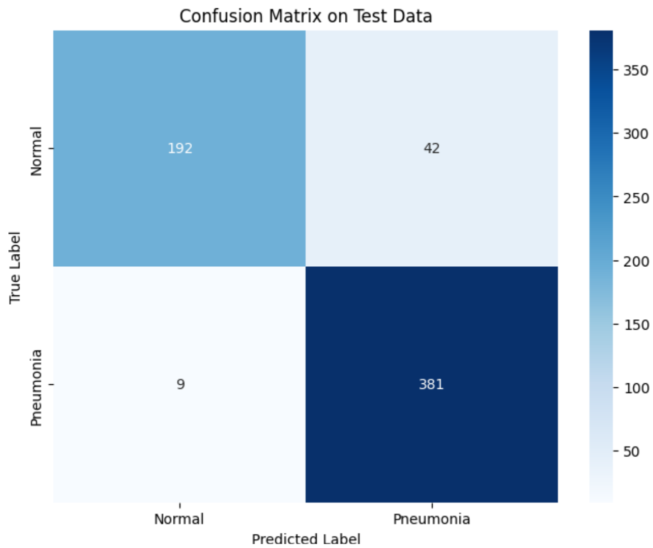
Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 16, 16, 512)	14,714,688
flatten (Flatten)	(None, 131072)	0
dense_2 (Dense)	(None, 256)	33,554,688
dropout_1 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 1)	257

Total params: 48,269,633 (184.13 MB)
Trainable params: 33,554,945 (128.00 MB)
Non-trainable params: 14,714,688 (56.13 MB)

```
Classification Report:
              precision    recall  f1-score   support

      Normal       0.96      0.82      0.88       234
   Pneumonia       0.90      0.98      0.94       390

    accuracy                           0.92       624
   macro avg       0.93      0.90      0.91       624
weighted avg       0.92      0.92      0.92       624
```



- The baseline model, with 110K parameters, is relatively small but underperforms, particularly on the Normal class, achieving only 62% accuracy.
- The VGG model, with over 48 million parameters, significantly outperforms the baseline, reaching 93% accuracy and balanced precision/recall.
- The substantial increase in model capacity and the use of a pre-trained architecture (VGG16) contribute to improved generalization and feature learning.
- Regularization techniques such as dropout were applied to reduce overfitting.
  Global average pooling and batch normalization in the baseline model helped with convergence but were insufficient alone.

Learning Curve of VGG16 model:

The model demonstrates strong performance based on the learning curves. Both training and validation accuracy are consistently high, reaching over 94%, which indicates that the model is making accurate predictions on both seen and unseen data. This level of accuracy reflects effective learning and a solid understanding of the patterns in the dataset.
Additionally, the loss curves show a steady decline in both training and validation loss over the epochs. This means the model's predictions are getting closer to the actual values, further confirming that the training process is working well. The trend suggests that the model is successfully minimizing error as it learns.
Overall, the model performs best around epoch 2 or 3, where validation accuracy is high and validation loss is low. It captures the underlying data patterns effectively and maintains high accuracy, making it a reliable choice for the given task.

6.0 Bias Detection and Evaluation

6.1 Bias Evaluation

The baseline model shows significant class imbalance in performance metrics, suggesting potential bias:

- The baseline model's recall for the Normal class is 0.00, indicating it completely fails to identify normal cases, while Pneumonia recall is 1.00.
- Precision for Normal is also 0.00, implying all Normal predictions are incorrect.
- This results in a weighted average F1-score of only 0.48, with macro average F1-score at 0.38, revealing poor balanced performance.
  Such a skew suggests the baseline model is biased towards predicting Pneumonia, likely influenced by class imbalance or insufficient learning capacity.

In contrast, the VGG-based model demonstrates balanced and high performance:

- Precision, recall, and F1-score for both classes are approximately 0.8 or higher.
- Accuracy improves drastically to 93%.
- Macro and weighted averages are aligned, indicating the model handles both classes fairly.

The baseline model exhibits significant bias favoring the Pneumonia class, while the VGG model mitigates this bias and offers robust performance across classes.

6.2 Mitigation Strategies

The baseline model’s bias towards Pneumonia was mitigated effectively by:

- Using a more powerful architecture (VGG16), which improves feature extraction and representation.
- Potentially applying better regularization and training strategies (e.g., dropout, dense layers).
- Incorporating more diverse and representative training data

7.0 Future Implementation

Future work will extend the model to perform multi-class classification, distinguishing between bacterial and viral pneumonia. This enhancement can improve treatment decisions by providing more specific diagnostic insights, further supporting clinical care.

8.0 Challenges

One major challenge was limited computational resources, which restricted the complexity of models we could train and the size of batches processed.
Another significant challenge was the lack of interpretability of deep learning models, making it difficult to provide clear explanations for predictions to clinicians, which is essential for clinical adoption.

9.0 Conclusion

In this project, we developed an AI-powered solution to detect pneumonia from chest X-ray images using both a custom Convolutional Neural Network (CNN) and a pretrained VGG16 model via transfer learning. The approach included robust data preprocessing, augmentation, and model training, followed using a Streamlit based interface. AI significantly reduced the manual workload for data scientists by automating key steps such as feature extraction, pattern recognition, and image classification, tasks that would otherwise require complex rule-based systems or intensive domain expertise. Transfer learning, in particular, eliminated the need to build models from scratch, saving time and computational resources. Overall, AI enabled faster development cycles, improved diagnostic accuracy, and provided scalable solutions that support healthcare professionals in making more informed decisions.

10.0 References

1. https://data.mendeley.com/datasets/rscbjbr9sj/3
2. https://www.geeksforgeeks.org/vgg-16-cnn-model/