Identifying Pneumonia with Neural Networks

By: Alexander Gambino

The Costs of Failing to Diagnose Pneumonia

- Increased Patient Mortality
- Prolonged Hospital Stays
- Higher Treatment Costs
- Legal Liabilities
- Damage to Reputation
- Reduced Patient Trust

Project Goals

- Train several neural networks to identify pneumonia
- Fine-tune them for better performance
- Identify the best-performing model
- Evaluate its ability to successfully identify pneumonia

The Chest X-Ray Image Dataset

- Downloaded from Kaggle.com
- Sourced from pediatric patients in Guangzhou, China
- Includes 5,863 images of chest x-rays
- Each image is classified as either pneumonia or normal
- Data is split into 3 smaller datasets:
  - Training
  - Validation
  - Testing

The Model-Building Process

- Data Preparation & Preprocessing
- Baseline Neural Network Training & Testing
- Adjusted Neural Networks Training & Testing
- Convolutional Neural Network (CNN) Training & Testing
- Adjusted CNNs Training & Testing
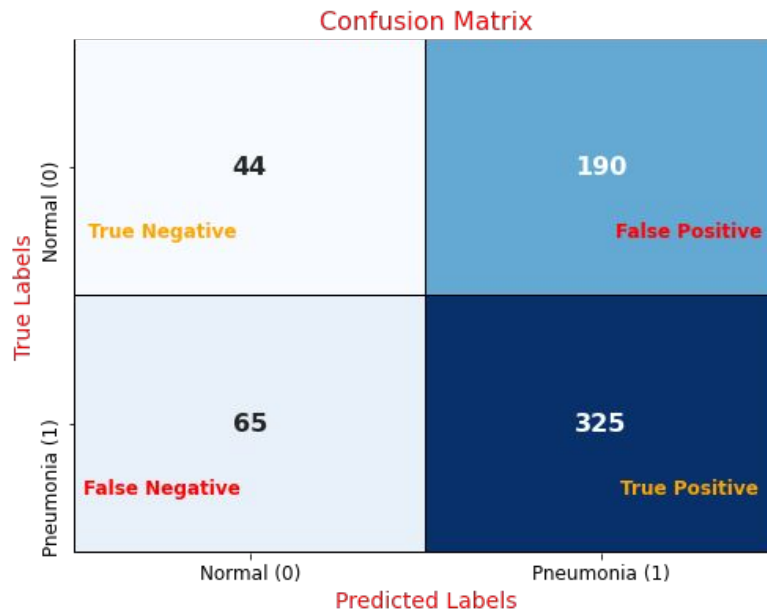- Final Model Evaluation

The Final Model

The Final Model:

- Class-balanced CNN with Dropout

Evaluation Performance:

- Predicted 83% of Pneumonia Cases
- 78% overall prediction accuracy

Confusion Matrix

	Normal (0)	Pneumonia (1)
Normal (0)	44 — True Negative	190 — False Positive
Pneumonia (1)	65 — False Negative	325 — True Positive

True Labels / Predicted Labels

How It can Benefit Your Healthcare Facilities

- Faster diagnosis times
- Reduced human error
- Lower treatment costs
- Improved patient outcomes
- Enhanced reputation and patient trust
- Increased productivity and profitability

Limitations

Data:

- Class Imbalance
- Limited Age Range
- Localized Demographics

Model:

- About 37% of pneumonia identifications are incorrect
- Roughly 17% of true pneumonia cases are missed

Summary

- Pneumonia diagnosis is critical
- Final model identified 83.33% of pneumonia cases with 78.37% overall accuracy
- Provides a tool to assist radiologists
- Enhances patient outcomes
- Saves time and money
- Limitations include the need for further development using diverse data

Thankyou! Any Questions?