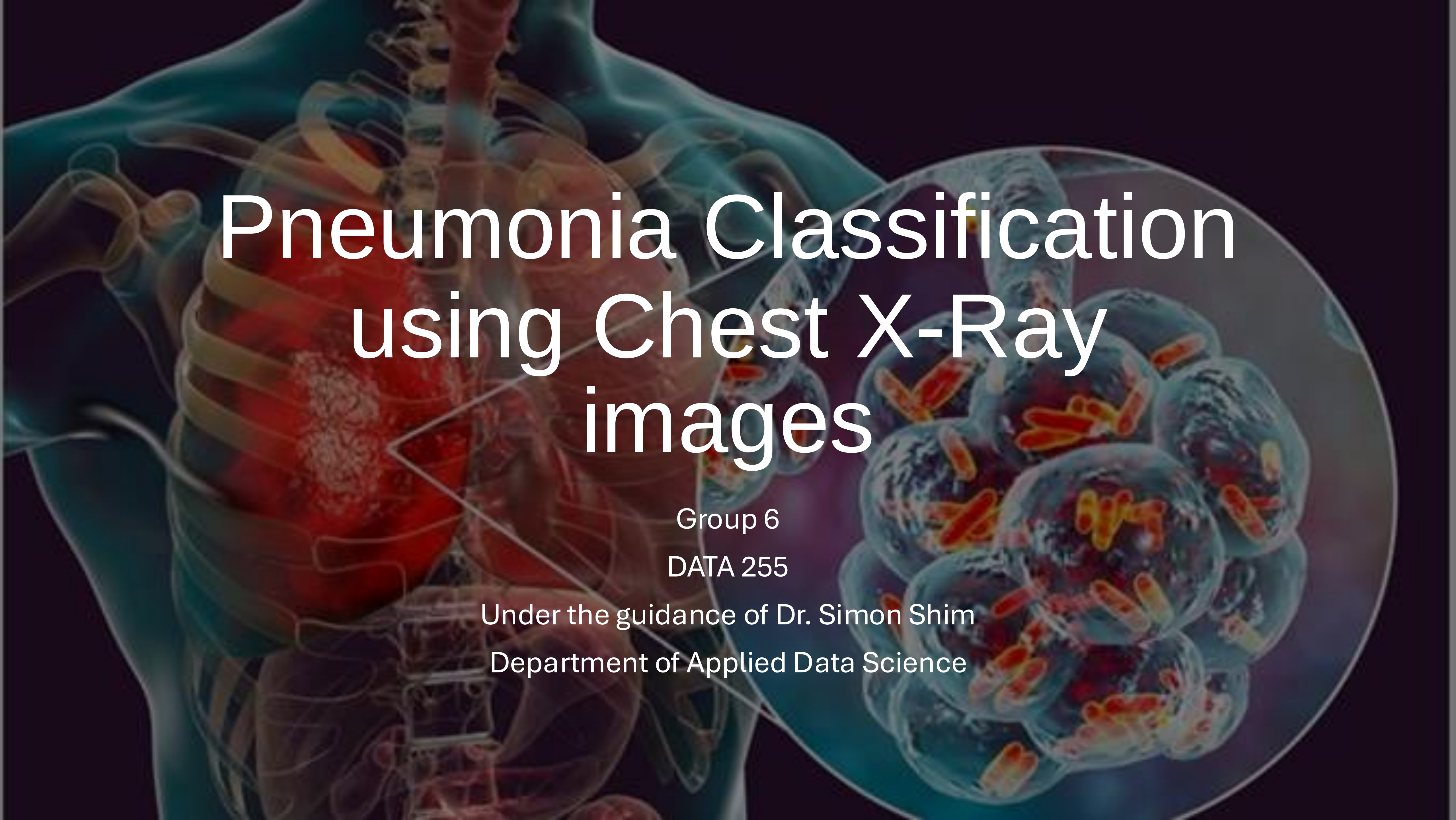


Pneumonia Classification using Chest X-Ray images

Group 6

DATA 255

Under the guidance of Dr. Simon Shim

Department of Applied Data Science

What is Pneumonia?

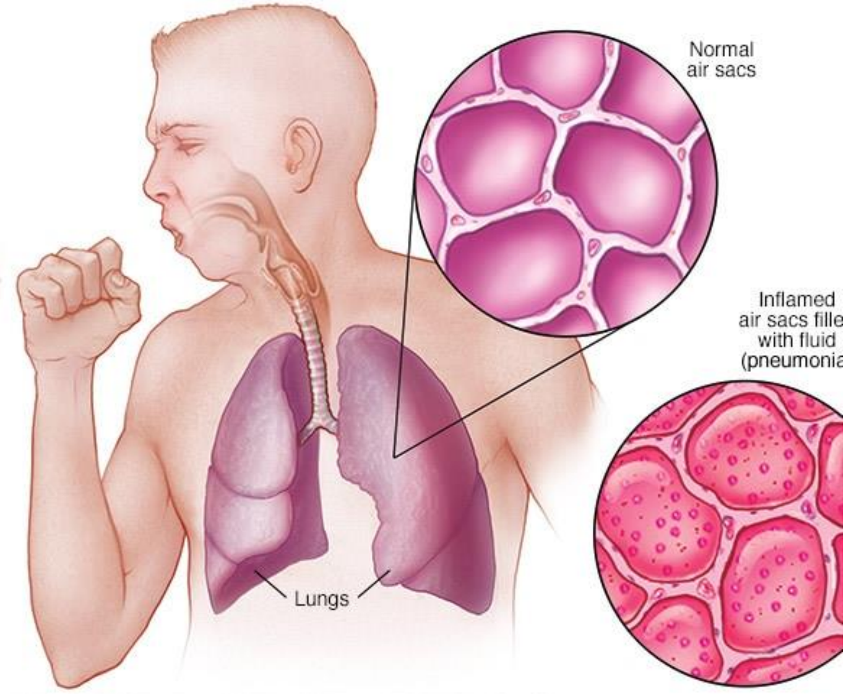
- Pneumonia is a lung infection causing inflamed air sacs filled with fluid or pus.

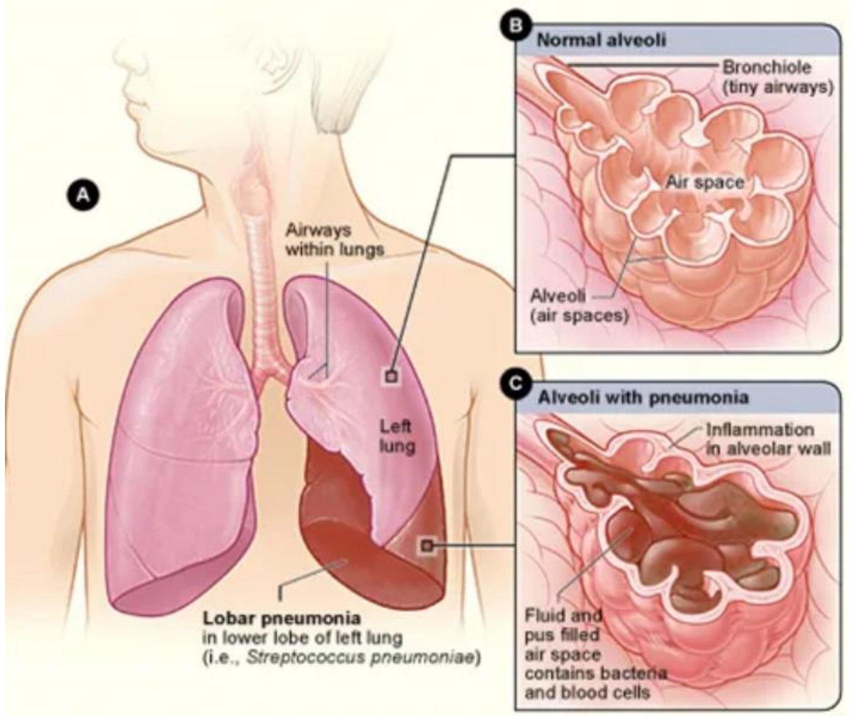
- Caused by germs like bacteria, viruses, or fungi, which enter through the air or spread from other infections.

- Germs multiply in the lungs, causing inflammation and reducing oxygen flow to the blood.

- Risk factors include weak immune systems, smoking, and pre-existing lung problems.

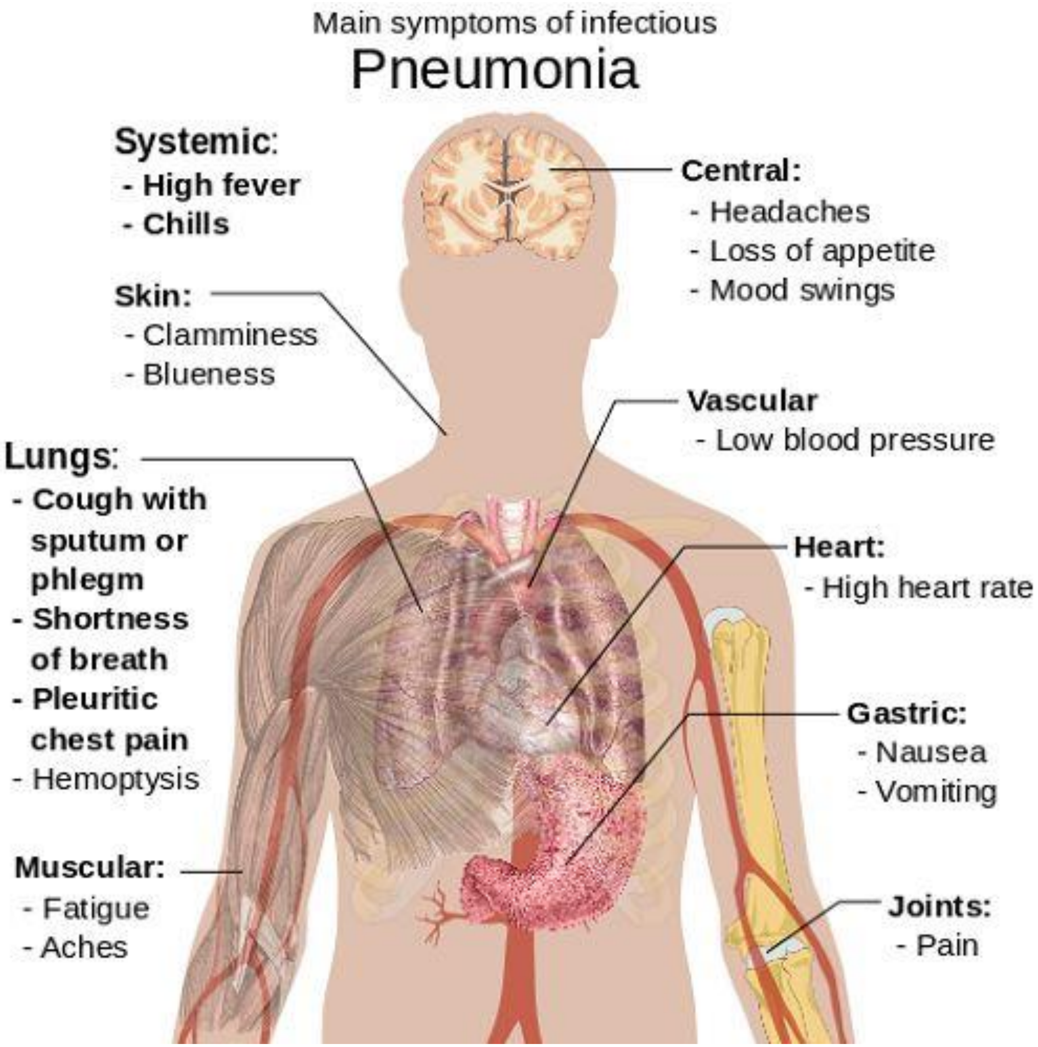
- Diagnosis is based on clinical exams: chest X-ray, blood count, and blood culture.





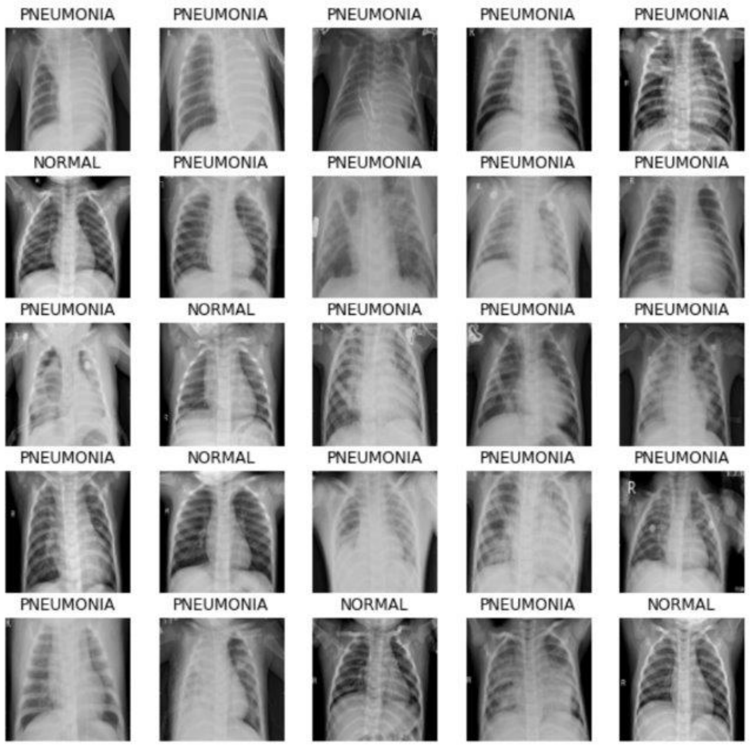
Introduction and Motivation

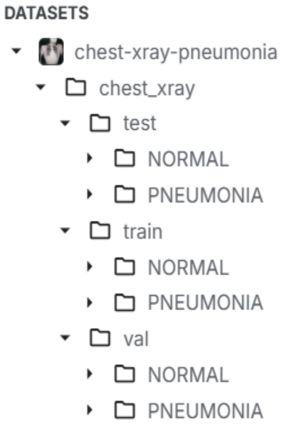
- Pneumonia caused over 15% of global deaths among children under 5 in 2015, totaling 920,000 deaths.
- In 2015, pneumonia led to 500,000 ER visits and over 50,000 deaths in the U.S., ranking among the top 10 causes of death.
- The U.S. sees over 1 million hospitalizations and 50,000 deaths annually due to pneumonia.
- The project aims to classify pediatric chest X-rays as "Pneumonia" or "Normal" using a deep learning model while leveraging explainable AI to identify contributing factors to the classifications.

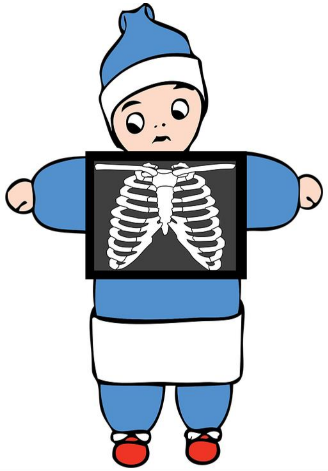


Data Source

- Kaggle Data set- [Chest X-Ray Images (Pneumonia)]() was leveraged for this project.
- 5,863 X-ray images (JPEG format) of pediatric patients (1-5 years), classified into **Pneumonia** and **Normal** categories.
- X-rays from pediatric patients at Guangzhou Women and Children's Medical Center, collected during routine clinical care.
- Images were screened for quality, ensuring only high-quality scans were included.
- For this project we have implemented the models using PyTorch

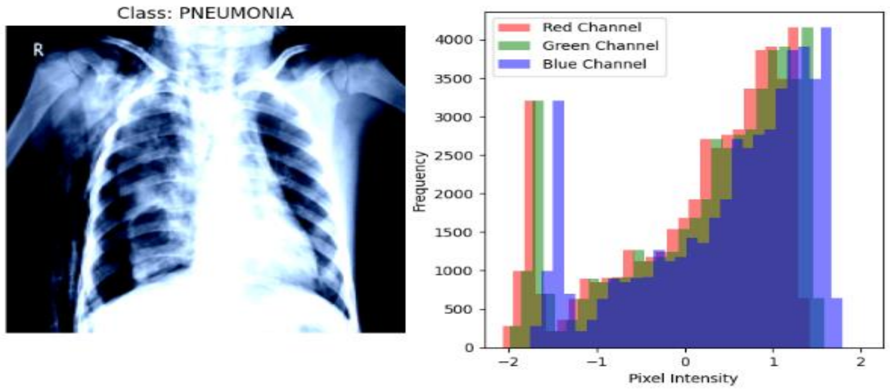


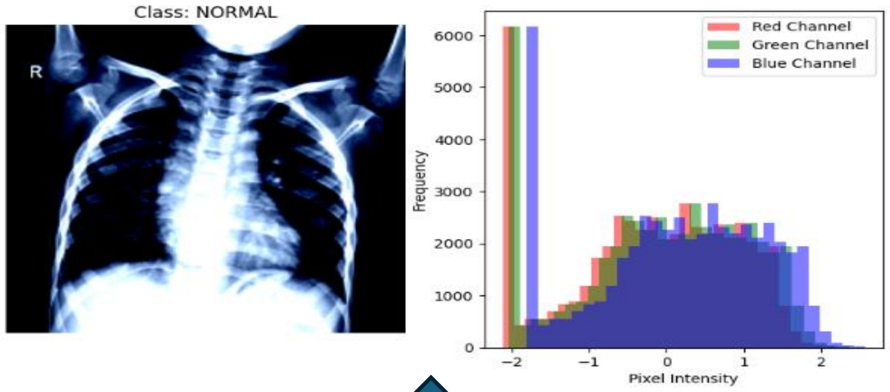




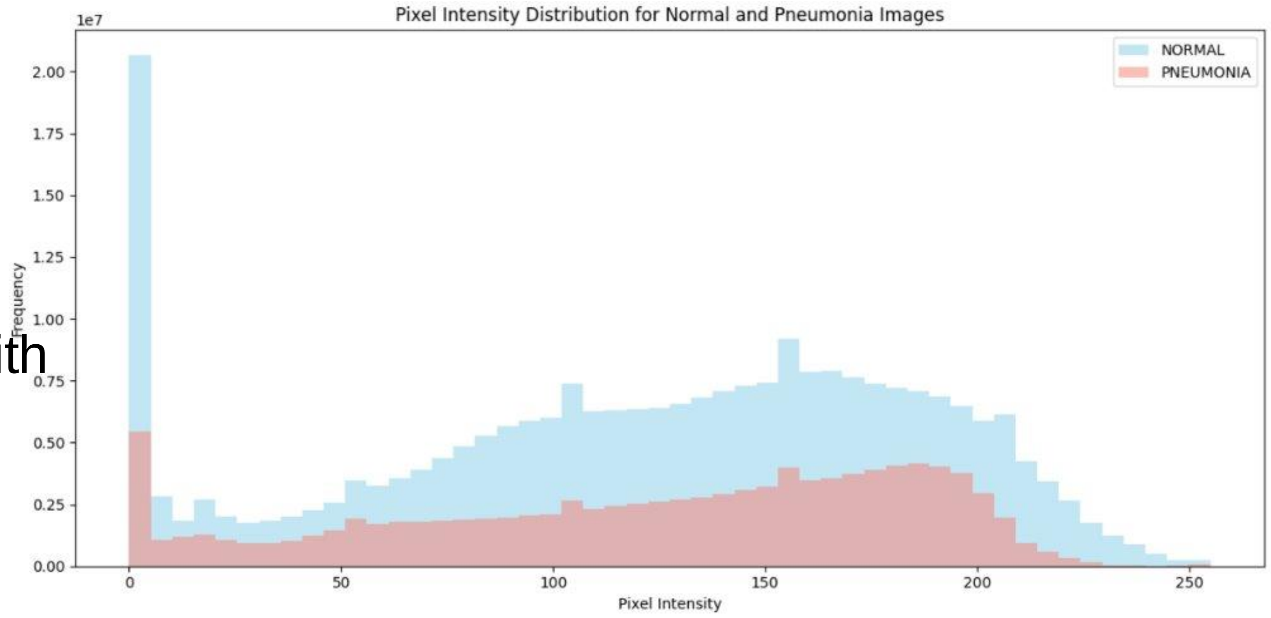
```
Class Names: ['NORMAL', 'PNEUMONIA']
```

Data Visualization





- NORMAL images show a sharp peak in low intensity (0–50) for dark lung areas and fewer bright pixels (100–150) for bones and tissues.
- PNEUMONIA images have a broader intensity range, reflecting variability from inflammation, with some overlap in patterns between classes.



•Normal X-rays show uniform, darker regions with less variation in pixel intensity.
•Pneumonia X-rays display brighter areas from fluid buildup, inflammation, and consolidation.

Data Cleansing

DATASETS

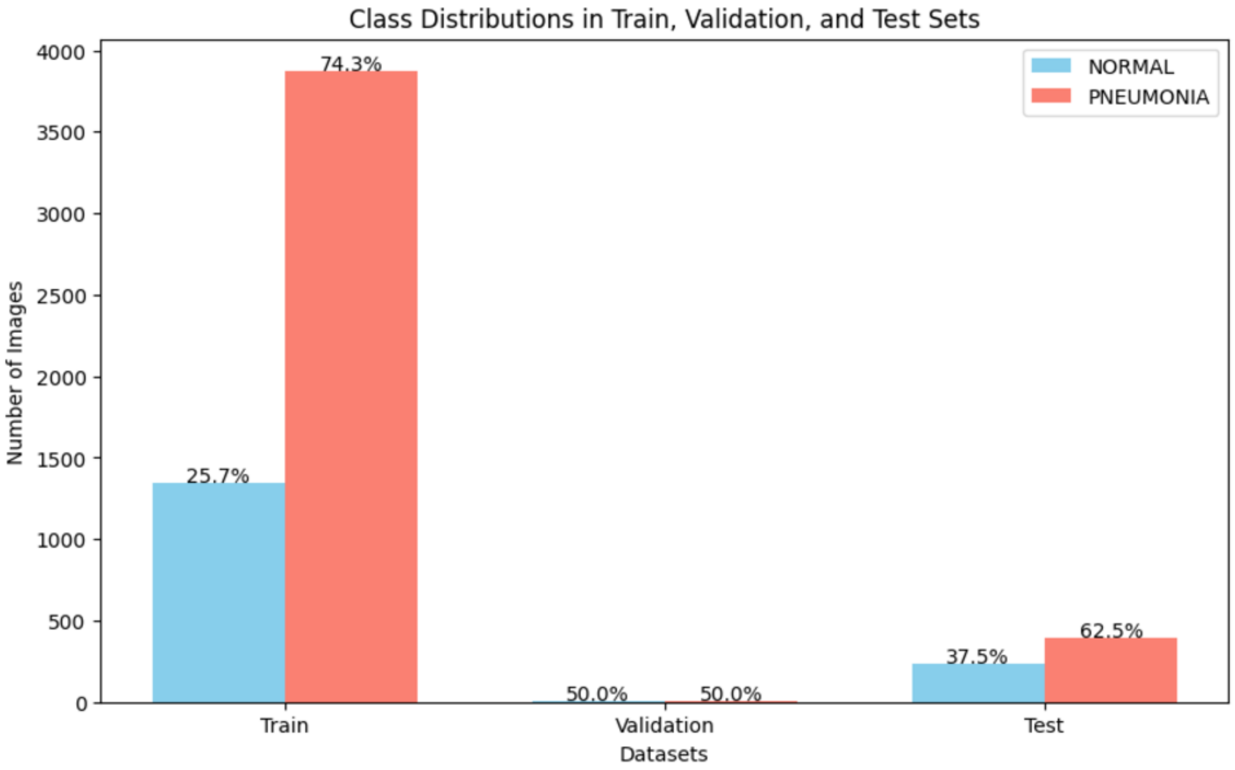
- chest-xray-pneumonia
  - chest_xray
    - test
      - NORMAL
      - PNEUMONIA
    - train
      - NORMAL
      - PNEUMONIA
    - val
      - NORMAL
      - PNEUMONIA

Uneven split

```
Total Images: 5856
Training Set: 5216 images (89.07% of total)
Validation Set: 16 images (0.27% of total)
Test Set: 624 images (10.66% of total)
```

Redistributed to correct the split ratios

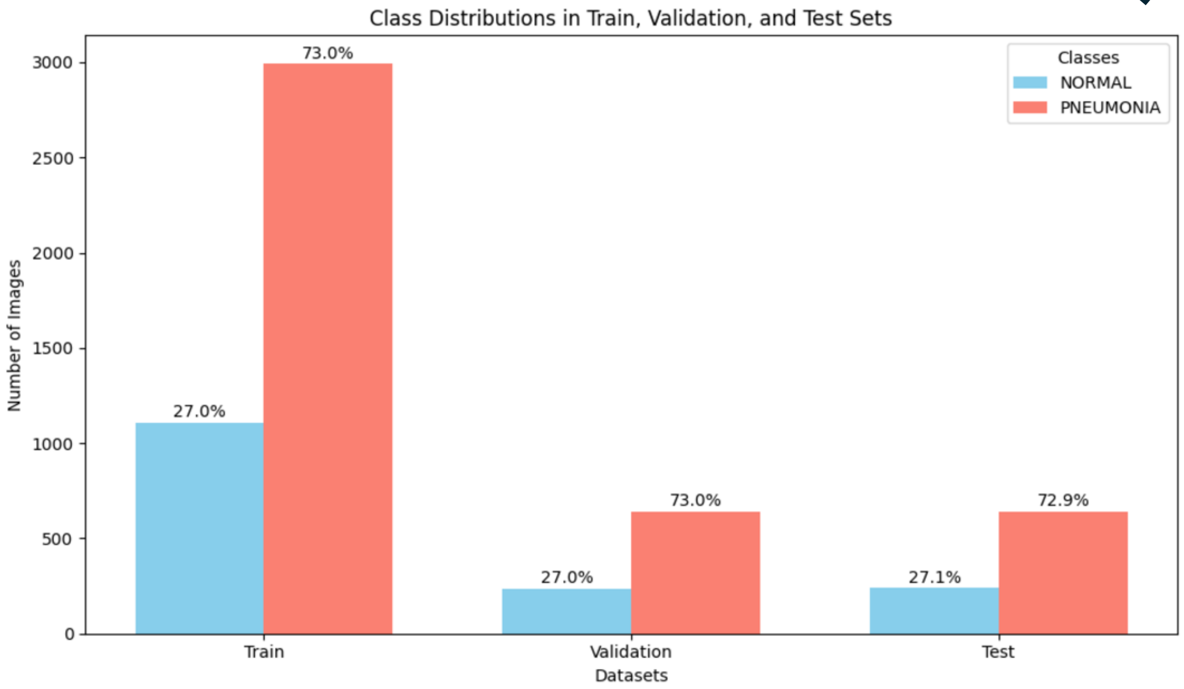
```
Redistributed Data Ratios:
Training Set: 4099 images (70.00% of total)
Validation Set: 878 images (14.99% of total)
Test Set: 879 images (15.01% of total)
```



Data Redistribution

- chest_xray_redistributed
  - test
    - NORMAL
    - PNEUMONIA
  - train
    - NORMAL
    - PNEUMONIA
  - val
    - NORMAL
    - PNEUMONIA

Class Imbalance



Still Classes are Imbalanced!

- To mitigate the Class Imbalance, we assign Class weights.
- Assign higher importance to the underrepresented class by adjusting the loss function to penalize incorrect predictions on the minority class more heavily.)
- class weights were calculated as:

$$w_c = \frac{\text{total number of samples}}{\text{number of classes} \times \text{number of samples in class } c}$$

For NORMAL:

$$w_{\text{NORMAL}} = \frac{4,099}{2 \times 1,108} = \frac{4,099}{2,216} \approx 1.8497$$

For PNEUMONIA:

$$w_{\text{PNEUMONIA}} = \frac{4,099}{2 \times 2,991} = \frac{4,099}{5,982} \approx 0.6852$$

```
Class weights: tensor([1.8497, 0.6852]
```

Feature Engineering/Data Transformation

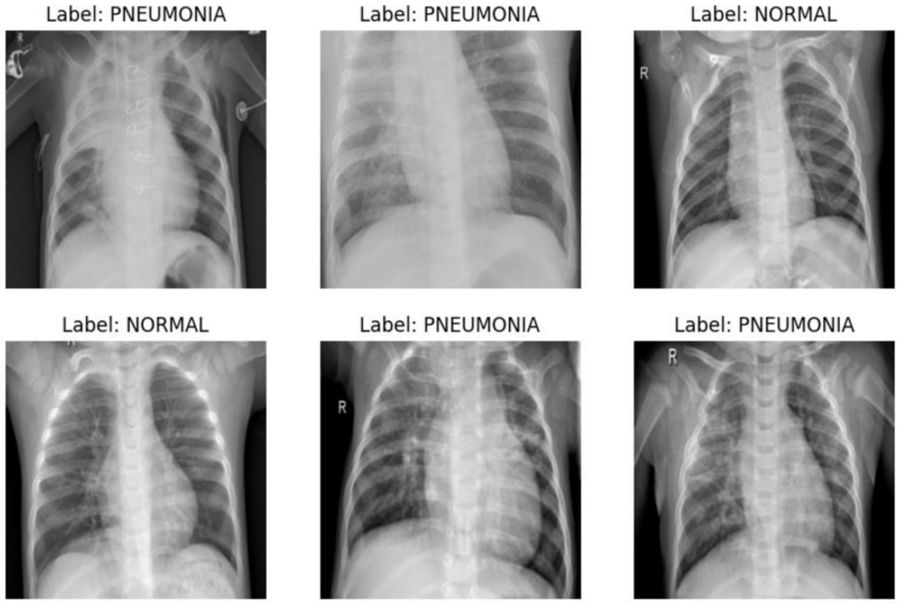
Data Transformations improve the model's generalization and help it learn robust features before training.

- **Resizing**: Standardizes image size (256x256) for consistent input to the model.

```
Average Image Sizes for Train, Validation, and Test Sets:
Training Set - NORMAL: (1692, 1382), PNEUMONIA: (1198, 822)
Validation Set - NORMAL: (1662, 1356), PNEUMONIA: (1195, 819)
Test Set - NORMAL: (1682, 1382), PNEUMONIA: (1179, 803)
```

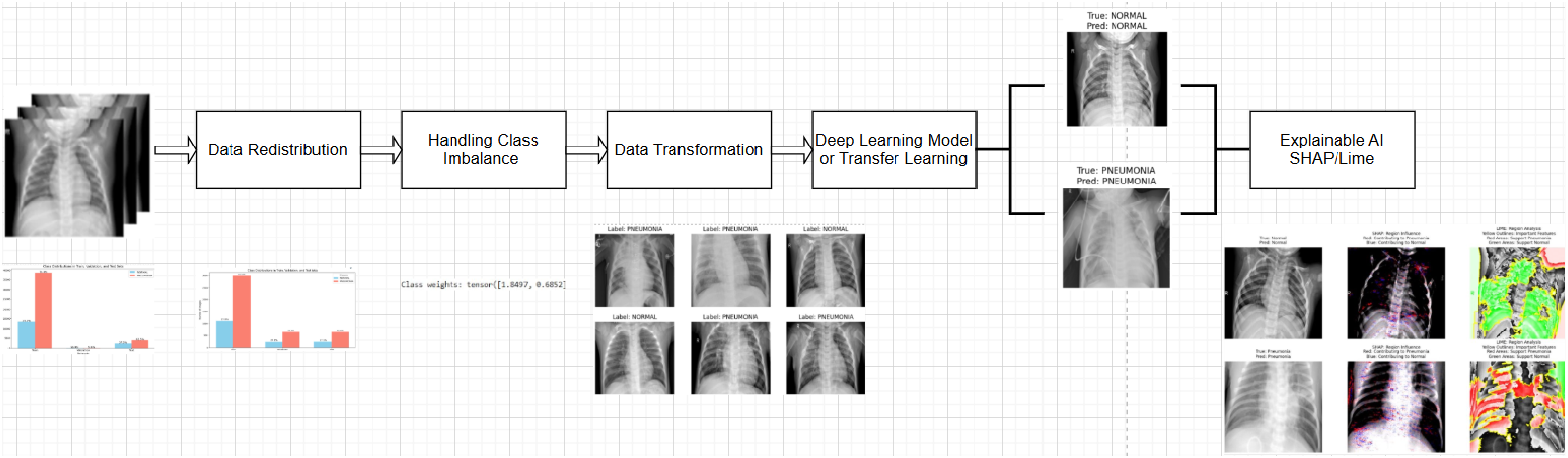
```
Shape of one batch of images: torch.Size([32, 3, 256, 256])
Number of images in the batch: 32
Image dimensions (C x H x W): 3 x 256 x 256
```

- **Resize to 256x256:** Ensures consistent input size, simplifies computation, and enables efficient batching by standardizing dimensions across all images.
- **Grayscale to 3 Channels**: Converts grayscale to 3 channels for compatibility with models expecting RGB input.
- **Normalization**: Scales pixel values to [-1, 1] for faster, stable training and improved model convergence.



Proposed Solutions

- By leveraging Deep Learning technologies like data augmentation, transfer learning, and hyperparameter tuning to develop an accurate and robust classifier that aids in early diagnosis of pneumonia, especially in pediatric patients.
- We implemented below models:
  - CNN
  - CNN Advanced
  - ResNet50
  - InceptionNet
  - CheXNet
  - Ensemble Model



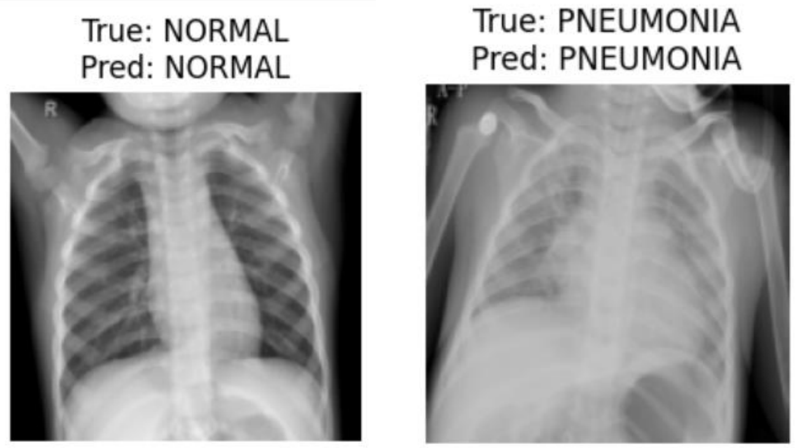
Models and Justifications

- **CNN**: Chosen for its ability to automatically learn **spatial hierarchies of features in images**, making it ideal for classifying medical images like X-rays with clear patterns for pneumonia.
- **CNN Advanced**: Utilized advanced CNN architectures to improve performance by **capturing more complex patterns and enhancing generalization**, especially for medical image classification
- **ResNet50:** Deep architecture with residual connections, **avoiding vanishing gradients** and enabling learning of fine-grained X-ray features.
- **CheXNet:** Pre-trained, optimized architecture specifically fine-tuned for pneumonia classification in X-rays. (Andrew NG)
- **InceptionNet:** Uses multiple kernel sizes per layer, capturing features at different scales to detect subtle differences in X-rays.
- **Ensemble Model**: Combines predictions from multiple models, improving classification robustness and reducing overfitting risk.

Convolutional Neural Networks (CNN)

- **Input Layer**: Accepts 256x256 RGB images for consistent input size and efficient training.
- **Convolutional Layers**: Conv1 (32 filters), Conv2 (64 filters), Conv3 (128 filters) with 3x3 kernels, stride 1, padding 1 for progressive feature extraction.
- **Activation**: ReLU activation for efficient gradient flow and non-linearity.
- **Pooling**: MaxPooling with 2x2 kernels to reduce spatial dimensions, prevent overfitting, and improve efficiency.
- **Fully Connected Layers**: FC1 (512 neurons) for feature aggregation and FC2 (2 neurons) for binary classification (Pneumonia vs. Normal).

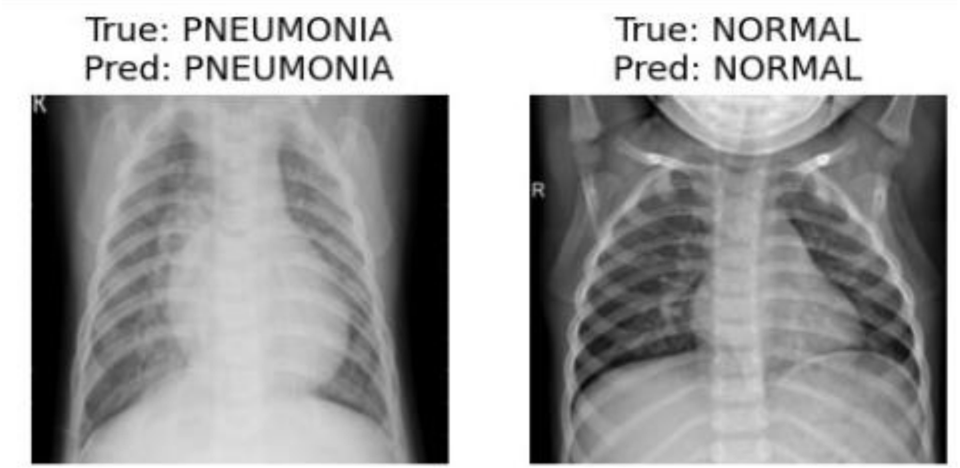
	precision	recall	f1-score	support
NORMAL	0.93	0.94	0.94	238
PNEUMONIA	0.98	0.98	0.98	641
accuracy			0.97	879
macro avg	0.96	0.96	0.96	879
weighted avg	0.97	0.97	0.97	879



CNN Enhanced

- **Activation Functions**: Experimented with different activations (e.g., ReLU, Leaky ReLU) for better feature extraction and non-linearity.
- **Dropout & Regularization**: Applied dropout (0.3) and **weight decay** to prevent overfitting while maintaining model capacity.
- **Batch Normalization & Optimizer**: Implement **batch normalization** to reduce training time and improve generalization; use **AdamW optimizer** for better weight updates and optimization.
- **High Resolution & Larger Kernels**: Retain higher image resolution for nuanced X-ray features, and use larger kernels (e.g., 7x7) with padding for better initial feature capture.
- **Gradient Clipping**: Used **gradient clipping** to handle high-contrast regions and subtle differences between normal and pneumonia X-rays.

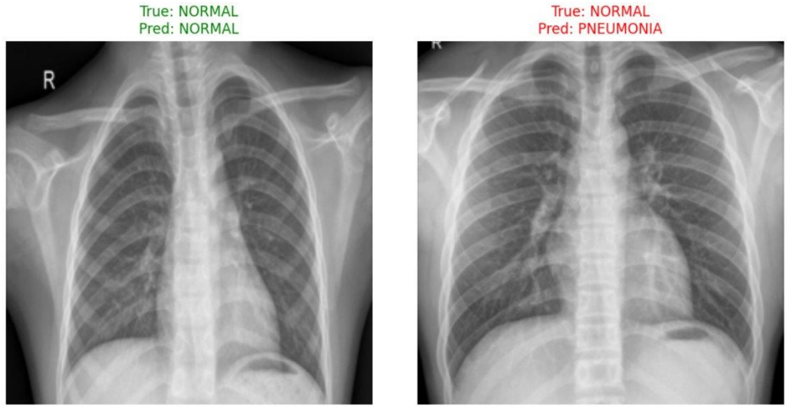
	precision	recall	f1-score	support
NORMAL	0.51	0.71	0.59	238
PNEUMONIA	0.87	0.75	0.80	641
accuracy			0.73	879
macro avg	0.69	0.73	0.70	879
weighted avg	0.77	0.73	0.75	879



ResNet50

- **Transfer Learning**: Utilizes a pre-trained ResNet50 model for effective knowledge transfer, enabling accurate pneumonia classification with a smaller dataset.
- **Model Architecture**: ResNet50, **with 50 layers and residual connections**, addresses the **vanishing gradient problem** and enhances training efficiency.
- **Fine-tuning**: Fine-tuned the last few layers of ResNet50 while freezing earlier layers to adapt the model to perform binary classification (Pneumonia vs. Normal) on chest X-ray features.
- **Optimization & Early Stopping**: Uses Adam optimizer, binary cross-entropy loss, and early stopping with dynamic learning rate adjustments for optimal training.

	precision	recall	f1-score	support
NORMAL	0.96	0.91	0.94	238
PNEUMONIA	0.97	0.99	0.98	641
accuracy			0.97	879
macro avg	0.97	0.95	0.96	879
weighted avg	0.97	0.97	0.97	879



CheXNet

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

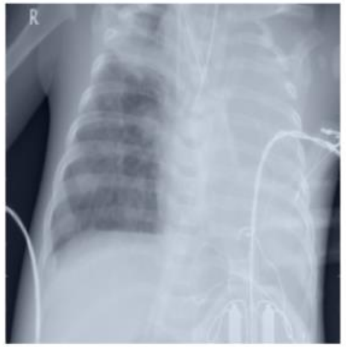
Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1]
Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Robyn L. Ball [2] Curtis Langlotz [3] Katie Shpanskaya [3]
Matthew P. Lungren [3] Andrew Y. Ng [1]

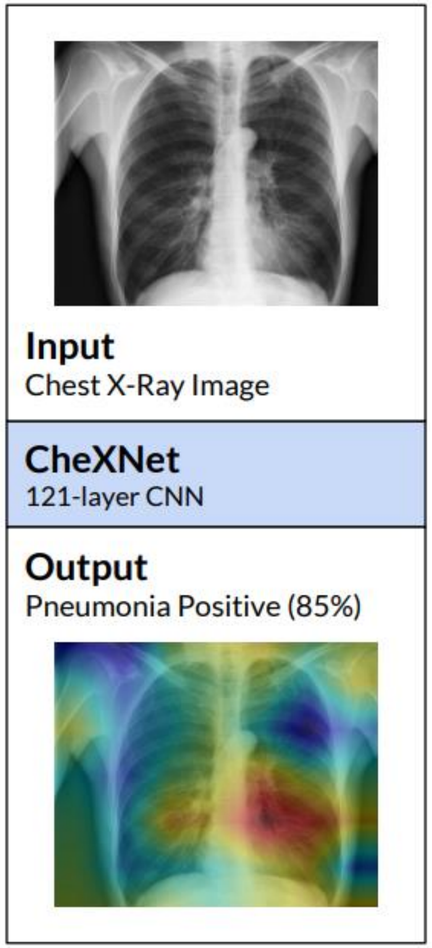
- CheXNet is a DenseNet-121 model fine-tuned specifically for chest X-ray classification. It was trained on the NIH ChestX-ray14 dataset and performs well on tasks like pneumonia detection
- Requires fewer parameters than models like ResNet, making it computationally efficient.
- **Final Layer**: Replaces the final layer with a fully connected network and sigmoid activation for binary classification (Pneumonia vs. Normal).
- **Loss Function & Optimizer**: Uses binary cross-entropy loss and Adam optimizer with a learning rate of 0.001, with decay when validation loss plateaus.

	precision	recall	f1-score	support
NORMAL	0.92	0.97	0.95	238
PNEUMONIA	0.99	0.97	0.98	641
accuracy			0.97	879
macro avg	0.96	0.97	0.96	879
weighted avg	0.97	0.97	0.97	879

True: Normal
Pred: Normal

True: Pneumonia
Pred: Pneumonia



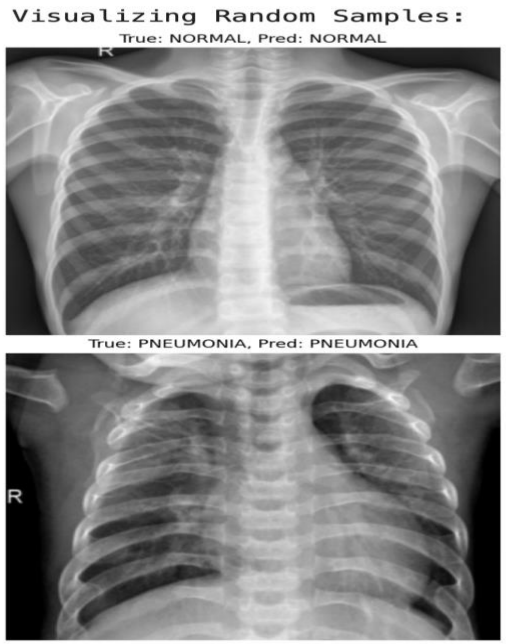


https://arxiv.org/abs/1711.05225

InceptionNet

- **Transfer Learning**: Utilizes a pre-trained **Inception V3** model for effective pneumonia classification using chest X-ray images by leveraging prior ImageNet knowledge.
- **Model Architecture**: Inception V3's modular structure enhances efficiency with inception modules. Custom top layers include a fully connected layer (2 units) and an auxiliary classifier for multi-output learning.
- **Data Augmentation**: Applies transformations like resizing (299x299), random flips, rotations, affine transforms and normalization to boost generalization.
- **Fine-tuning**: Fine-tunes later layers while freezing earlier layers to adapt to domain-specific features.

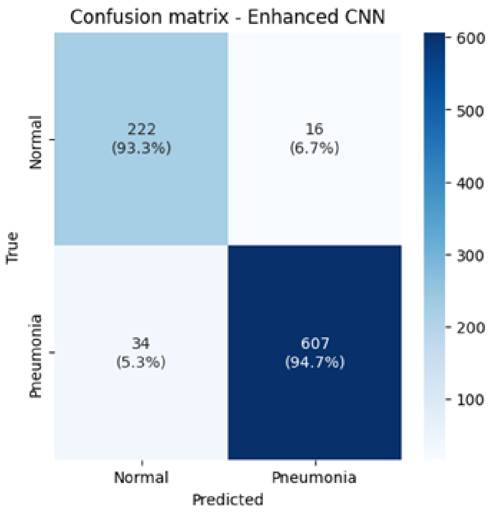
	precision	recall	f1-score	support
NORMAL	0.96	0.92	0.94	238
PNEUMONIA	0.97	0.99	0.98	641
accuracy			0.97	879
macro avg	0.97	0.96	0.96	879
weighted avg	0.97	0.97	0.97	879

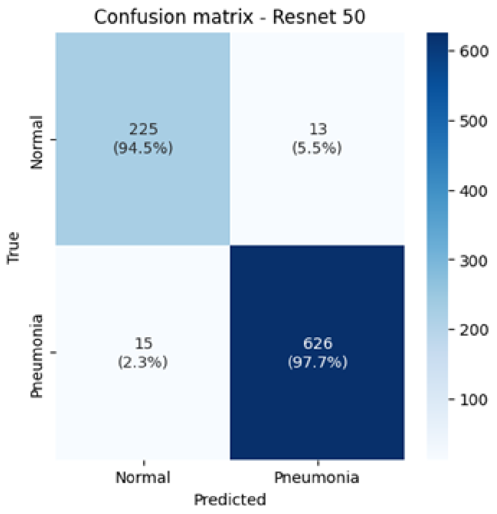


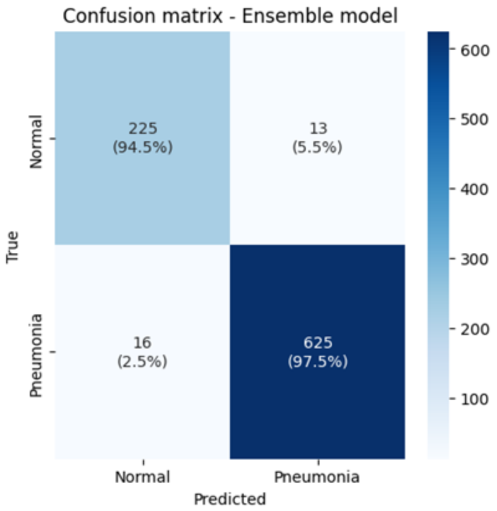
Ensemble

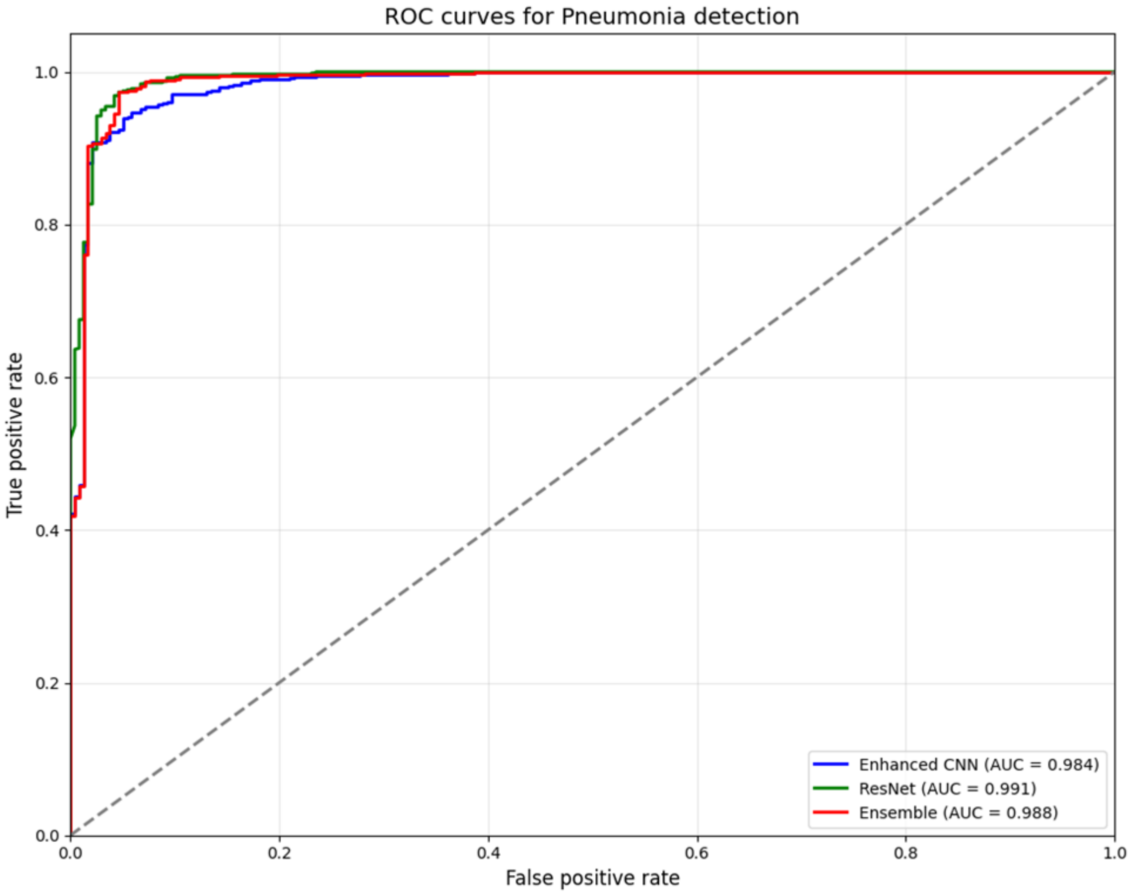
- **Ensembling** is a technique where we combine two or more models to make more accurate and robust predictions
- This approach leverages the strength of individual models while reducing overfitting and biases. There are multiple approaches for ensembling of deep learning models in increasing order of complexity:
  - Majority voting
  - Average probabilities
  - **Weighted voting** – we used this
  - Stacking
  - Bagging
- For our project, we considered two models for ensemble modeling – Enhanced CNN and ResNet50 (trained using transfer learning) using an accuracy-based weighting scheme
- We observed slightly higher weightage of 51% was given to the ResNet model while 49% weightage was with the CNN model since Resnet was the better performing model

Ensemble modeling - results









Model Evaluation & Result analysis

Model	Precision	Recall	F1-Score	Accuracy
CNN	N: 0.93, P: 0.98	N: 0.94, P: 0.98	N: 0.94, P: 0.98	0.97
CNN Advanced	N: 0.51, P: 0.87	N: 0.71, P: 0.75	N: 0.59, P: 0.80	0.73
ResNet50	N: 0.96, P: 0.97	N: 0.91, P: 0.99	N: 0.94, P: 0.98	0.97
CheXNet	N: 0.92, P: 0.99	N: 0.97, P: 0.97	N: 0.95, P: 0.98	0.97
InceptionNet	N: 0.96, P: 0.97	N: 0.92, P: 0.99	N: 0.94, P: 0.98	0.97
Ensemble Model	N: 0.93, P: 0.98	N: 0.95, P: 0.98	N: 0.94, P: 0.98	0.97

Overfitting!
Train Accuracy: 98.55

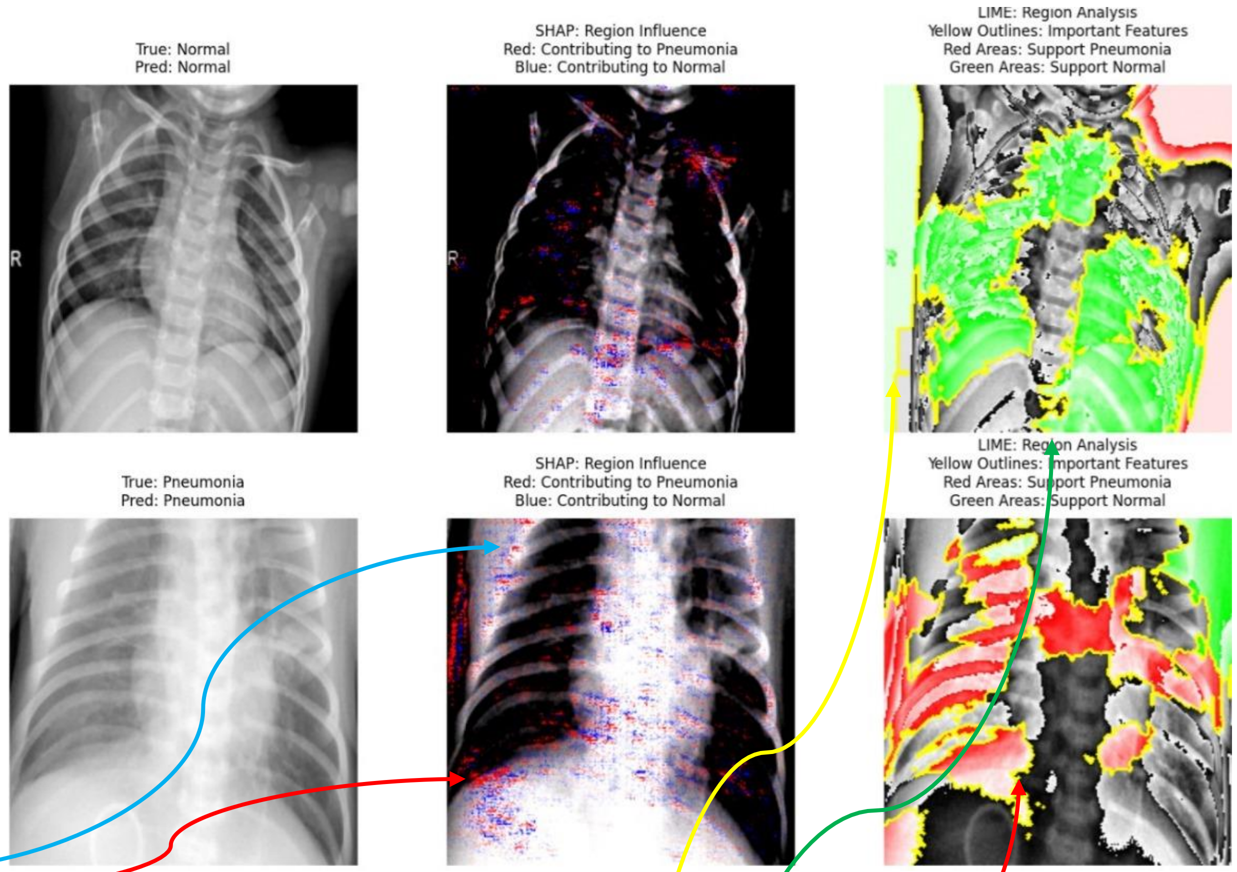
Best models

N-Normal , P - Pneumonia

- **Highest Accuracy**: CNN, ResNet50, CheXNet, and InceptionNet achieved 0.97 accuracy, with CheXNet and InceptionNet benefiting from transfer learning.
- **Highest AUC**: InceptionNet led with a ROC AUC of 0.9945, excelling in distinguishing pneumonia cases.
- **Ensemble Performance**: Slightly lower AUC (0.988) than ResNet50 (0.991), showing minor tradeoffs in combining models.

Explainable AI

- Most deep-learning models are black-box models – explainability and interpretability of these models is a major area of research
- There have been recent developments to bring explainability to Deep Learning models namely SHAP and LIME
- **SHAP** calculates each feature's contribution to understanding the input images' influence
  - SHAP values are visualized using a color-coded scheme wherein blue features predict Normal and red features predict Pneumonia



- **LIME** generates local explanations for a specific prediction by approximating the model's behavior around that instance
  - Important regions outlined in yellow
  - Green areas support Normal prediction
  - Red areas support Pneumonia prediction

SHAP: **SH**apley **A**dditive ex**P**lanations

LIME: **L**ocal **I**nterpretable **M**odel–agnostic **E**xplanations

Explainable AI – Applications

- SHAP and LIME overlays can act as a **decision support tool** to augment radiologists by highlighting suspicious areas, acting as an **AI powered "second opinion"** with explanations, **potentially reducing the chance of missed diagnoses or incorrect diagnoses**
- It can also play a crucial in **medical training and education - training new radiologists** by showing what areas of the X-ray to look for before diagnosing the patient
- In **emergencies like another pandemic, this approach could be used for emergency diagnosis** to provide a quick initial assessment with explanations
    - The doctor will provide human oversight while leveraging AI capabilities but this can help in quickly filtering out healthy cases and focusing on patients that require urgent care.
- Overall, this approach shows how explainable AI can make deep learning more practical and trustworthy in medical applications.

Conclusion

- Pneumonia detection achieved with deep learning models like CNN, ResNet50 and ensemble methods.
- ResNet50 and ensemble model outperformed individual models, providing robust and reliable predictions.
- Explainable AI tools (SHAP, LIME) ensured transparency in model decision-making.
- Addressed class imbalance and standardized data preprocessing for consistent results.
- **Future focus:** expanding dataset diversity, refining interpretability and clinical integration.

Related Work

- **Ensemble Learning**: Combined DenseNet169, MobileNetV2, and Vision Transformer to achieve 93.91% accuracy in pneumonia classification.
- **Deep Learning Architectures**: DenseNet169 achieved 95.72% accuracy for binary pneumonia classification.
- **Transfer Learning**: Pre-trained CNNs like VGG19 and DenseNet121 enhanced pneumonia detection accuracy.
- **Data Augmentation**: Techniques like cropping and histogram equalization improved image quality and model performance.

Deep Learning on Chest X-ray Images to Detect and Evaluate Pneumonia Cases at the Era of COVID-19

Karim Hammoudi [1,2,✉], Halim Benhabiles [3], Mahmoud Melkemi [1,2], Fadi Dornaika [4,5], Ignacio Arganda-Carreras [4,5,6], Dominique Collard [7,8], Arnaud Scherpereel [9]

Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network

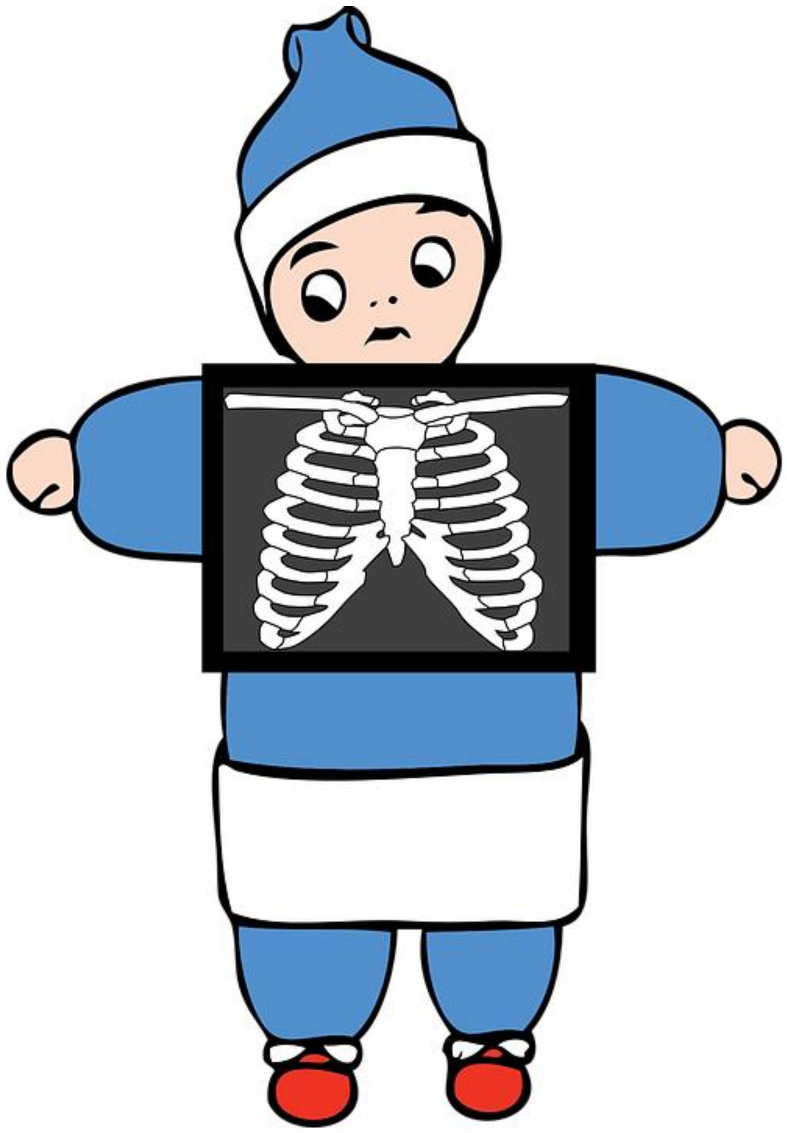
Muhammad Mujahid [1,†], Furqan Rustam [2,†], Roberto Álvarez [3,4], Juan Luis Vidal Mazón [3,5], Isabel de la Torre Díez [6,*], Imran Ashraf [7,*]

Pneumonia Detection on chest X-ray images Using Ensemble of Deep Convolutional Neural Networks

Alhassan Mabrouk, Rebeca P. Díaz Redondo, Abdelghani Dahou, Mohamed Abd Elaziz, Mohammed Kayed

References

- https://pmc.ncbi.nlm.nih.gov/articles/PMC9140837/
- https://arxiv.org/abs/2312.07965
- https://pmc.ncbi.nlm.nih.gov/articles/PMC8185498/



Thank you