CECS 456 Project Report - Chest X-Ray Pneumonia Classification Using Deep Learning

Author: Huy Nguyen

Course: CECS456 - Deep Learning

1. Introduction:

Pneumonia, a severe lung disease, is commonly diagnosed with chest X-rays. Manual interpretation is expert-dependent and slow. Deep learning, particularly convolutional neural networks (CNNs), has recently achieved strong performance in medical image classification.

This project develops and tests two deep learning models for classifying chest X-ray images as NORMAL or PNEUMONIA: a custom CNN and a ResNet18-based transfer learning model. Both are trained and evaluated on the Kaggle Chest X-Ray Pneumonia dataset.

This report covers the dataset, related research, methods, experiment setup, metrics, results, analysis, and my own contributions to the project.

2. Dataset and Related Work:

Dataset Description:

The Chest X-Ray Pneumonia Dataset from Kaggle is used in this project. It has X-ray images in two categories:

- NORMAL: healthy chest X-ray images
- PNEUMONIA: images showing signs of pneumonia infection

- The dataset structure provided includes:

- Training set: Contains the majority of images and is used to fit model parameters
- Validation set: Used for model selection and tuning
- Test set: Used for final evaluation and comparison

- The dataset is imbalanced, with many more pneumonia images than normal ones. Weighted sampling is used to help prevent the model from favoring the majority class.

Related Work:

Deep learning, especially CNN-based architectures, has shown strong performance in medical imaging tasks. Previous work includes:

- Transfer learning approaches using ImageNet-pretrained models, such as ResNet, VGG, or DenseNet, have demonstrated high accuracy for pneumonia detection.
- Custom CNNs have been used in smaller-scale research to explore simplified architectures tailored for binary classification tasks.

This project builds on previous work by explicitly comparing the performance, strengths, and limitations of a custom CNN and a well-known transfer learning model under the same experimental conditions.

3. Methodology:

This project implements and directly compares two model architectures by evaluating their performance using accuracy and other relevant metrics on the same dataset and experimental setup.

3.1 SmallCnn: Custom Model from Scratch

The custom model is a **4-layer CNN with progressive channels (32, 64, 128, 256), followed by global average pooling, fully connected layers with dropout, and a softmax output for binary classification.**

- Convolution → ReLU → BatchNorm → MaxPooling blocks
- Increasing channels: 32 → 64 → 128 → 256
- Global Average Pooling
- Fully connected layers with dropout
- Final softmax classification into 2 classes

This model is lightweight, easy to train, and acts as a baseline for comparison.

3.2 ResNet18Transfer: Transfer Learning Model

The second model is ResNet18 pretrained on ImageNet, with only the last fully connected layer modified for two outputs. Two configurations were explored: (1) frozen backbone, training only the final layer, and (2) unfrozen backbone for full fine-tuning.

- **Frozen backbone**: Only the final layer is trainable
- (Optional) Unfrozen backbone to fine-tune the full network

Due to computational limits, experiments mainly used a frozen backbone. Transfer learning leverages pretrained visual features, improving performance on limited chest X-ray data.

3.3 Data Preprocessing and Augmentation

All images were resized to **224×224**. Augmentation included:

- Random horizontal flips
- Random rotations
- Color jittering
- Normalization using ImageNet means and standard deviations

These steps help reduce overfitting and improve how well the model works on new data.

3.4 Loss Function & Optimization

- **Loss:** Cross-entropy
- **Optimizer:** AdamW (learning rate: 1e-3)
- **Scheduler:** Cosine Annealing
- **Imbalance Handling:** WeightedRandomSampler based on class frequencies

4. Experimental Setup

- **Environment:** Google Colab with GPU acceleration
- **Dataset Source:** Automatically downloaded using Kaggle API
- **Epochs:**
  - SmallCnn: 5–10 epochs
  - ResNet18Transfer: 5–8 epochs
- **Batch size:** 32
- **Image size:** 224×224
- **Evaluation:** Validation during training and final test evaluation

5. Measurement

The following performance metrics were used:

- **Accuracy** — overall proportion of correct predictions

- **Precision** — percentage of predicted pneumonia cases that were correct

- **Recall (Sensitivity)** — ability to detect pneumonia cases correctly
- **F1-score** — harmonic mean of precision and recall
- **Confusion matrix** — visual summary of correct/incorrect predictions

These metrics provide a well-rounded evaluation for medical classification tasks where false negatives can be critical.

5.1 SmallCnn Confusion Matrix & Classification Report

```
                                                     SmallCnn Test loss:
Confusion matrix:
 [[151  83]
 [  6 384]]

Classification report:

              precision    recall  f1-score   support

      NORMAL     0.9618    0.6453    0.7724       234
   PNEUMONIA     0.8223    0.9846    0.8961       390

    accuracy                         0.8574       624
   macro avg     0.8920    0.8150    0.8343       624
weighted avg     0.8746    0.8574    0.8497       624
```

SmallCnn Test Accuracy: 0.8574

Summary:

- NORMAL precision: 0.96
- PNEUMONIA precision: 0.82
- The model struggled with false positives (NORMAL misclassified as PNEUMONIA).

5.2 ResNet18Transfer Confusion Matrix & Classification Report

```
                                                     ResNet18Transfer Test
Confusion matrix:
 [[193  41]
 [ 17 373]]

Classification report:

              precision    recall  f1-score   support

      NORMAL     0.9190    0.8248    0.8694       234
   PNEUMONIA     0.9010    0.9564    0.9279       390

    accuracy                         0.9071       624
   macro avg     0.9100    0.8906    0.8986       624
weighted avg     0.9077    0.9071    0.9059       624
```

ResNet18 Test Accuracy: 0.9071

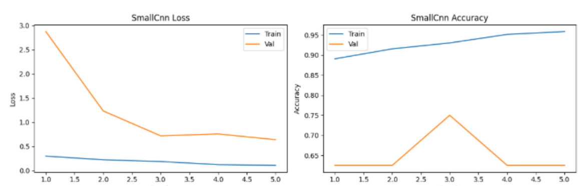
Summary:

- Precision for both classes is significantly higher
- False positives and false negatives reduced
- Stronger overall generalization

6. Results, Analysis, Intuitions, and Comparison .

This section contains all training curves and performance analysis.

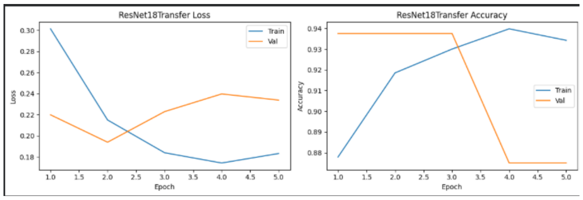
6.1 SmallCnn Training Curves



Interpretation:

- Training loss steadily decreases
- Validation loss fluctuates and does not improve consistently
- Validation accuracy peaks around epoch 3 then decreases
- Indicates overfitting due to limited model capacity

6.2 ResNet18Transfer Training Curves:



Interpretation:

- Training and validation curves are smoother
- Much lower validation loss compared to SmallCnn
- Higher and more stable validation accuracy

6.3 Model Performance Comparison

Model	Test Accuracy	Notes
SmallCnn	0.8574	Overfits; weaker generalization
ResNet18Transfer	0.9071	Best: strong accuracy

Key Takeaways:

- Transfer learning clearly outperforms the custom CNN
- ResNet18 has stronger feature extraction capabilities
- SmallCnn suffers from inconsistent validation accuracy, showing instability

7. Conclusion

This project used two deep learning models to detect pneumonia in chest X-ray images. The results show that transfer learning with **ResNet18** works better than a custom **CNN** for both accuracy and stability. The pretrained model fits medical imaging tasks well and reaches over 90% accuracy with only a small amount of fine-tuning.

Some possible ways to improve the project in the future are:

- Fully fine-tuning the ResNet18 model,
- trying deeper architectures like ResNet50 or DenseNet121,
- and using Grad-CAM to make the model's decisions easier to understand for medical professionals.

8. Contribution

All tasks were completed by Huy Nguyen including:

- Implementing dataset by using the kaggle API to import
- Write both model architectures
- Analyzing training curves and performance metrics
- Write and formatting the full project report