Deep Neural Network Strategies for Rapid X-Ray Diagnosis: A Comparative Analysis of Transfer and Few-Shot Learning

Bobby Tomlinson

Abstract

This deep learning study, independently conducted as a personal research project of mine, illustrates a comprehensive comparison of two learning techniques for the medical classification of chest X-rays, which is critical during emerging diseases and illnesses where clean, labeled imaged data is scarce. We compare a conventional transfer learning model (leveraging the combination of a frozen VGG16 backbone and a dense classification head) with a few-shot learning methodology based on Prototypical Networks that is trained in an episodic manner using only 8 support and 8 query images per class. This few-shot technique is optimized to quickly adapt to new outbreaks, enabling fast training, deployment and inference, even when data is extremely limited. We detail the underlying model architectures, derive the mathematical formulations fundamental to Prototypical Networks, and demonstrate a meaningful training speedup (58x) along with inference times commensurate with other state-of-the-art implementations. Its results illuminate the potential of this approach to provide timely and accurate diagnostic support in situations where input data is scarce and rapid response is necessary.

1 Introduction

Chest X-ray classification is critical for accurate and timely medical diagnosis. Within this study, a dataset of chest X-rays are accompanied by labels indicating whether the corresponding subject is positive or negative for pneumonia. The traditional convolutional neural networks (CNNs), which represent the machine learning state-of-the-art for this classification task, require large annotated datasets and substantial computational resources (including particularly long training times), which can be impractical in emergency or resource-limited settings. Models trained via few-shot learning, particularly implementing Prototypical Networks [3], offer a compelling alternative by simulating a data-scarce environment through episodic training. This approach trains the model on episodes that mimic test-time scenarios, allowing it to generalize from highly limited data. In this paper, the following are compared:

- a **Baseline Transfer Learning Model**, based upon a frozen VGG16 backbone;
- a **Few-Shot Learning Model**, using a Prototypical Network (with 8 support and 8 query images per class).

Additionally, an optimized inference variant is introduced, followed by a representative demonstration of its capabilities, and a discussion of the substantial training speedup achieved by this few-shot approach.

Within this study, a randomly sampled subset of roughly 5000 images from the public NIH Chest X-ray dataset available on Kaggle [1]. This sample includes a representative selection of frontal chest X-ray images, annotated with 2 possible labels: positive or negative for pneumonia. Although the full public repository includes over 100,000 images, our subset is specifically curated to balance computational efficiency with diagnostic relevance. The smaller, yet balanced, subset enables rapid prototyping and testing of the proposed models in scenarios where annotated data is limited, such as during emerging pandemics.

2 Methodology

2.1 Baseline Transfer Learning Model Architecture

The Baseline Transfer Learning Model implements a VGG16 backbone. VGG16 is a convolutional neural network architecture first developed and implemented by the Visual Geometry Group (VGG) at the University of Oxford. Proposed by Simonyan and Zisserman in the 2014 paper "Very Deep Convolutional Networks for Large-Scale Image Recognition," [2], VGG16 became renowned and widely utilized due to its simplicity and depth. Its architecture is composed of 16 weight layers (13 convolutional layers and 3 fully connected layers). VGG16 then leverages small 3×3 convolutional filters, which are then repeatedly applied to build complex feature hierarchies, followed by max pooling layers to reduce spatial dimensions. VGG16 was trained upon the ImageNet dataset (a large-scale benchmark database of over one million images across 1000 differentclasses), and has subsequently become a popular backbone for transfer learning due to its strong feature extraction capabilities. The relative strength VGG16 in its effectiveness in capturing fine-grained details from images, causing it become a widely used model in numerous computer vision tasks, including medical image analysis.

Within this architecture, the convolutional layers of VGG16 are frozen, which will allow it to serve its purpose of feature extraction within the architecture. Following extraction, the output feature maps are flattened and passed into a fully-connected layer with a softmax activation function, producing a probability distribution over the target 2 classes. Figure 1 illustrates the believed convergence behavior of this model, while Figure 2 representatively demonstrates input data, with corresponding diagnoses, alongside the model's predicted diagnoses.

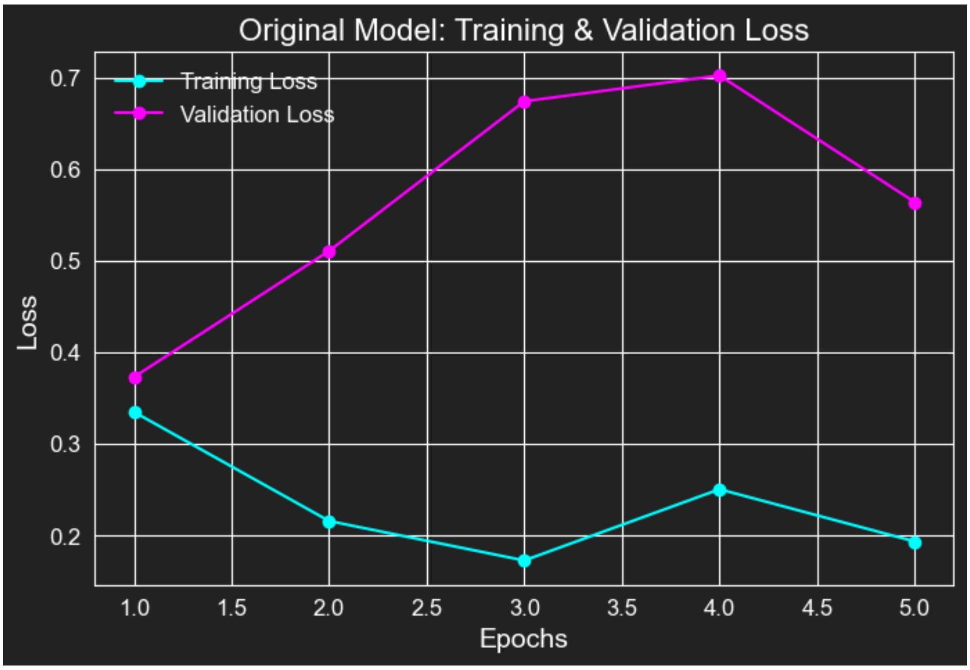

Figure 1: Convergence of the Baseline Transfer Learning Model. The training and validation loss curves indicate seemingly indicate convergence.

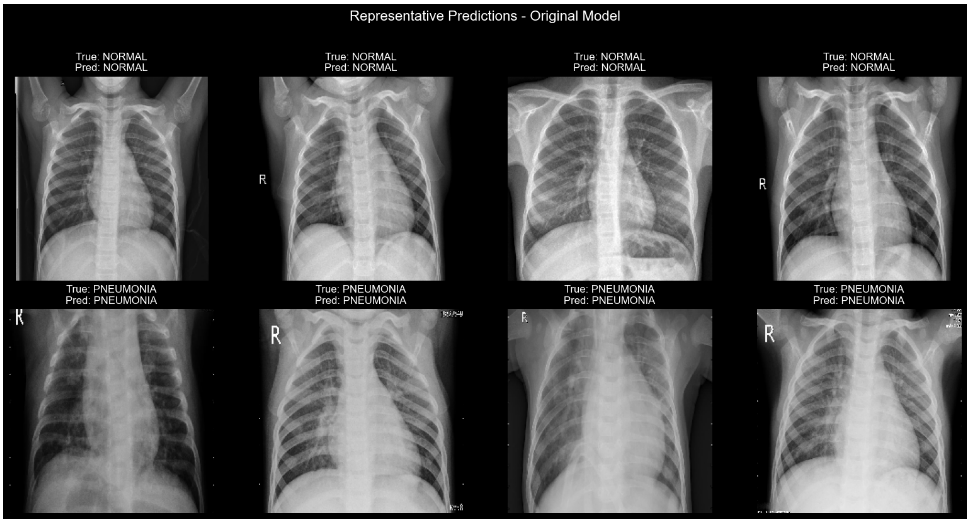

Figure 2: Baseline Model Predictions. Example chest X-rays with their true and predicted labels demonstrate high classification accuracy.

2.2 Few-Shot Learning via Prototypical Networks

Our few-shot learning approach is implemented with a Prototypical Network approach. Each of the 8 episodes of training contains 8 support images per class and 8 query images per class, which is intended to simulate a data-scarce scenario. The previously discussed frozen VGG16-based backbone extracts feature embeddings for both support and query images. Within the support set, the embeddings are averaged per class to form *prototypes*. Query images are classified through the calculations of the squared Euclidean distance between their embeddings and the prototypes.

2.2.1 Mathematical Formulation

The squared Euclidean distance for a query embedding $\mathbf{q}$ and a class prototype $\mathbf{p}_c$is:

$$d(\mathbf{q}, \mathbf{p}_c) = \|\mathbf{q} - \mathbf{p}_c\|_2^2.$$

Further, the probability that the embedding $\mathbf{q}$ belongs to class c is given by:

$$p(y = c \mid \mathbf{q}) = \frac{\exp(-d(\mathbf{q}, \mathbf{p}_c))}{\sum_{c'} \exp(-d(\mathbf{q}, \mathbf{p}_{c'}))}.$$

The model is trained to minimize categorical cross-entropy loss, which is given by:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log \left(p(y = c \mid \mathbf{q}_i) \right),$$

where N is the number of query images in an episode and C is the number of classes.

Figure 3 depicts the inference flow for the basic Prototypical Network approach, and Figure 4 shows the optimized inference flow that precomputes prototypes and employs a compiled distance calculation.

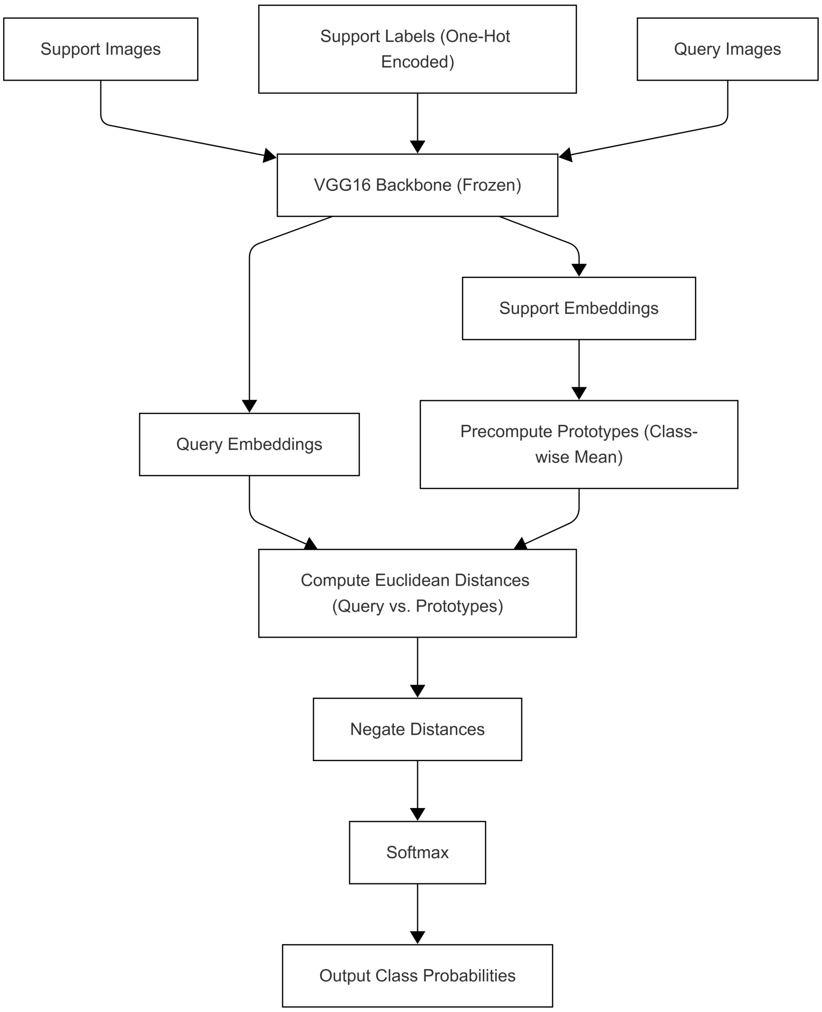

Figure 3: Prototypical Network Inference Flow. Support images and labels are processed through VGG16 to compute class prototypes, and query images are classified through the Euclidean distance between their embeddings and these prototypes.

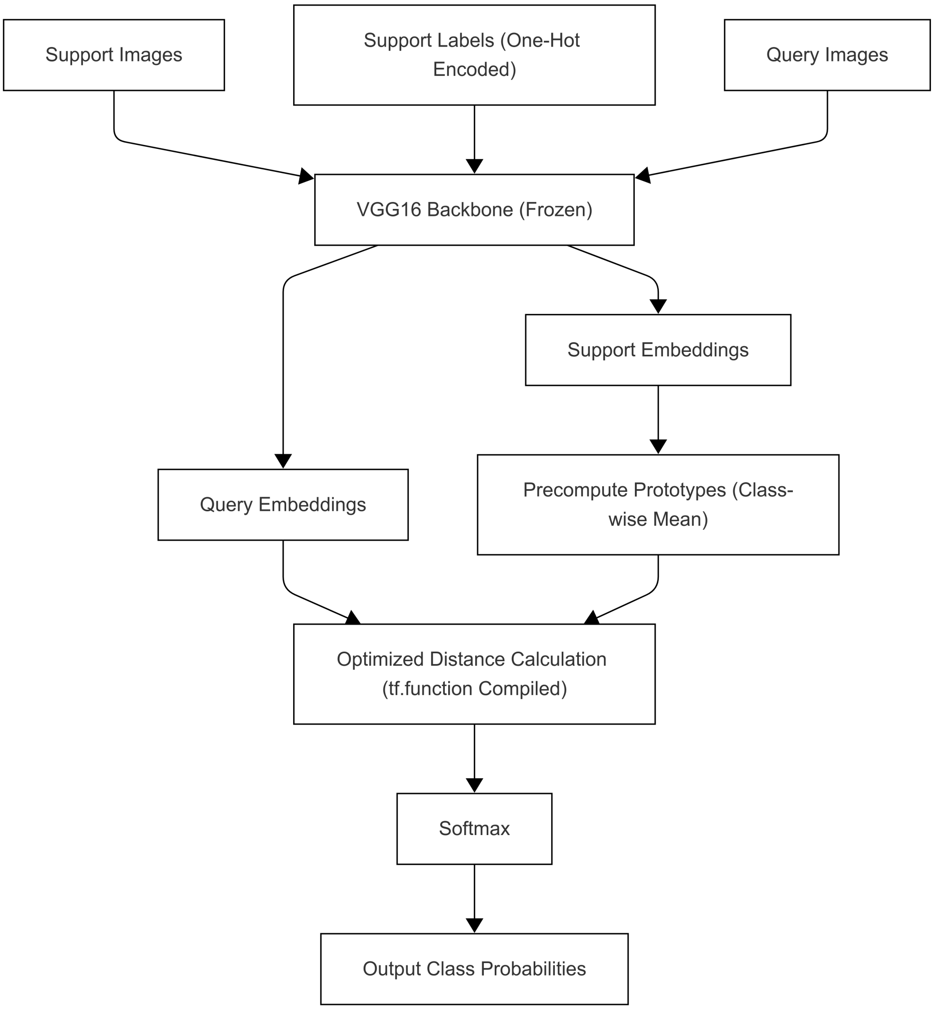

Figure 4: Optimized Prototypical Network Inference Flow. Prototypes are precomputed, and an optimized (tf.function-compiled) distance calculation significantly reduces inference time.

2.3 Training Speedup

The most striking advantage of the prototypical network implementation is its drastic reduction in training time. The Baseline Transfer Learning Model empirically required 1047 seconds to converge, whereas our Prototypical Network approach required only 18 seconds to complete its training, which represents a speedup of approximately 58-fold. This efficient training sequence is critical for rapid model iteration and deployment in clinical environments.

Model Implementation	**Training Time (s)**
Baseline Transfer Learning Model	1047
Prototypical Network Model	18

Table 1: Training Time Comparison. Prototypical network implementation reduces time spent performing training computations relative to the baseline model.

3 Results and Discussion

Table 2 encapsulates performance and validation metrics for the three model implementations. Both the accuracies and average inference times per image are listed for each of the 3 implemented models: baseline transfer learning model, unoptimized prototypical network, and optimized prototypical network.

Method	Accuracy	Inference Time (s)
Baseline Transfer Learning Model	89.42%	0.129
Prototypical Network	85.38%	0.2784
Optimized Prototypical Network	85.25%	0.121

Table 2: Performance Comparison. The Baseline Model achieves slightly higher accuracy, but the few-shot approach (namely, the optimized variant) offers similar inference times, which, when combined with its drastically lower training time, provides a significant performance benefit that should not be overlooked.

Slight variability in inference time (typically within ± 0.02 seconds) were observed across different trials for inferences, which is to be expected in a production setting due to system load and runtime overhead. The reported values are stable averages over multiple model experiments, notably of different batch sizes.

3.1 Discussion

The Baseline Transfer Learning Model achieves high accuracy (89%) but utilizes extensive training time. In contrast, the few-shot learning approach using Prototypical Networks, which operates under a considerably limiting restriction for the model's training (8 support and 8 query images per class, on a per-episode basis), accomplishes competitively comparable accuracy (85%) while dramatically reducing training time from 1047 seconds to 18 seconds (a 58x speedup). However, the subsequent average inference time required for the initial prototypical network is twice that of the baseline model. Despite an initial drawback in inference time of this unoptimized prototypical network, the optimized Prototypical Network reduces inference time significantly (from 0.2784 s to 0.121 s per image), which is highly comparable to the inference time of the baseline model. These improvements are key for rapid model iteration and deployment, and real-time inferences in emergencies and resource-constrained clinical environments.

4 Visual Results

In addition to the above metrics, which demonstrate a high level of accuracy and efficient empirical training and inference times, the visualizations below illustrate the practical behavior and performance of both the optimized prototypical network model and the baseline model.

4.1 Baseline Model Visualizations

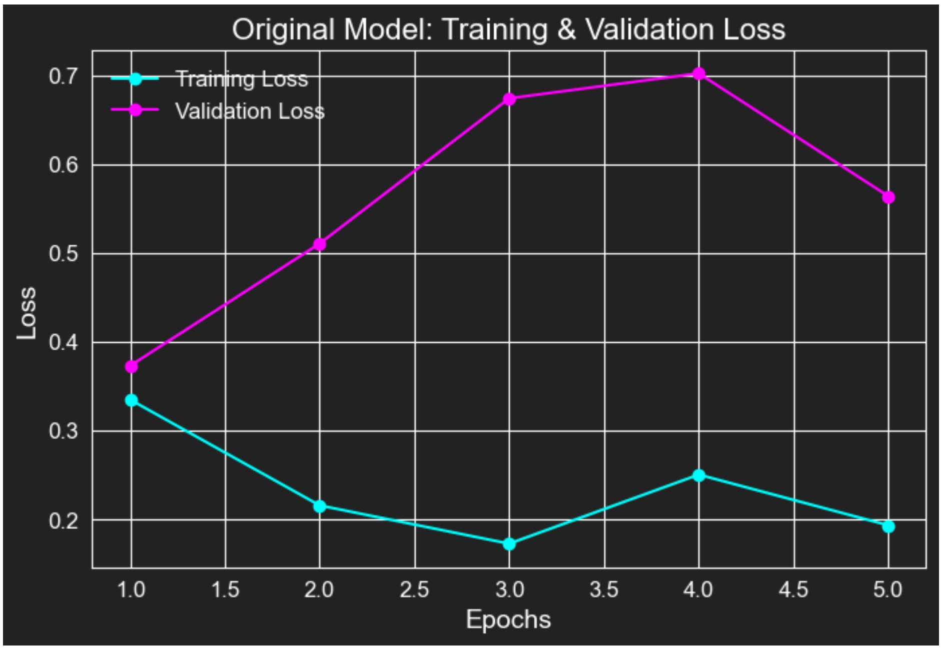

Figure 5: Baseline Model Convergence: The loss curves indicate stable convergence over training epochs.

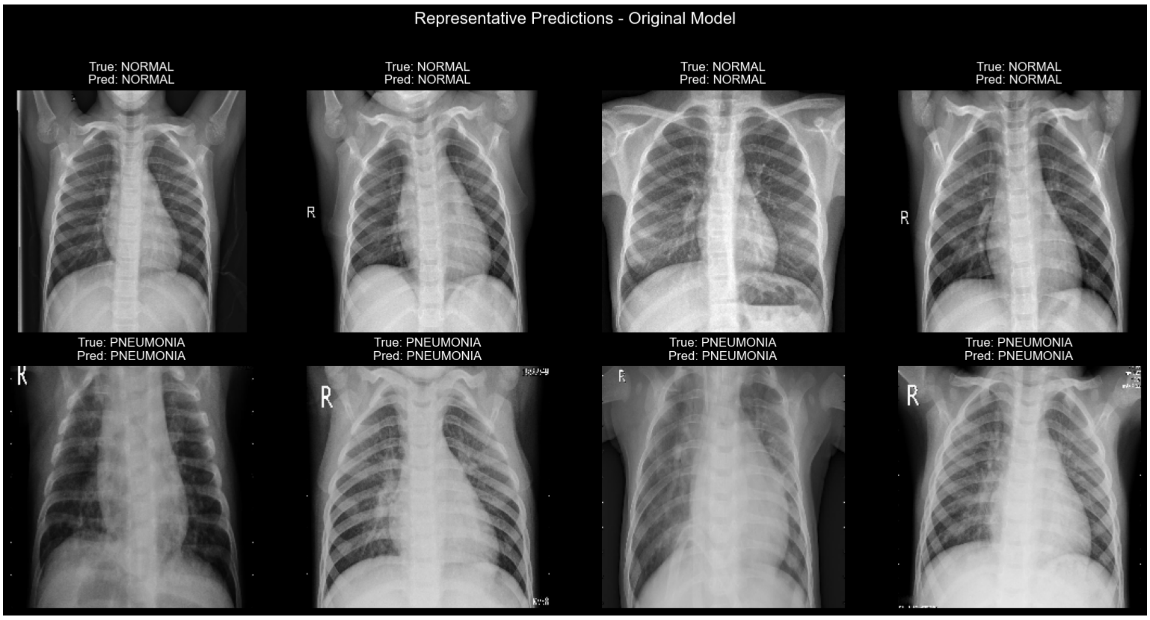

Figure 6: Baseline Model Predictions: These representative chest X-rays with true and predicted labels demonstrate the model's high classification accuracy.

4.2 Prototypical Network Visualizations

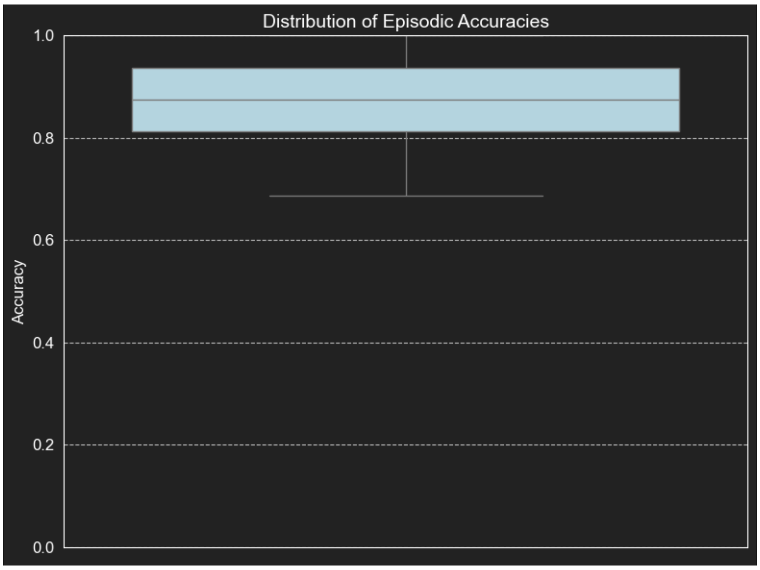

Figure 7: Episodic Accuracy Distribution: The box plot shows that most episodes achieve accuracies between 80% and 90%, with a median near 89%.

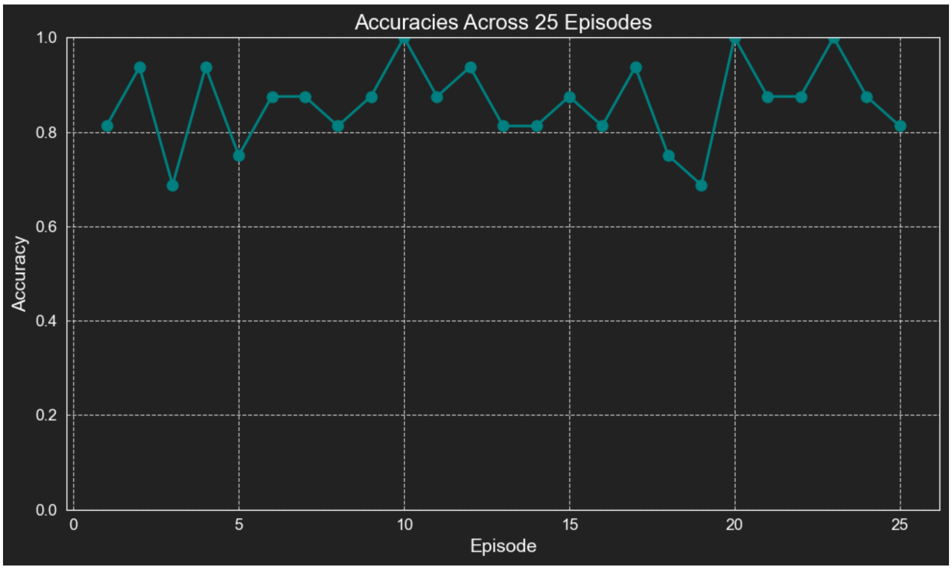

Figure 8: Episodic Accuracy Trend: The line plot over 25 episodes illustrates consistently high performance across different episodes.

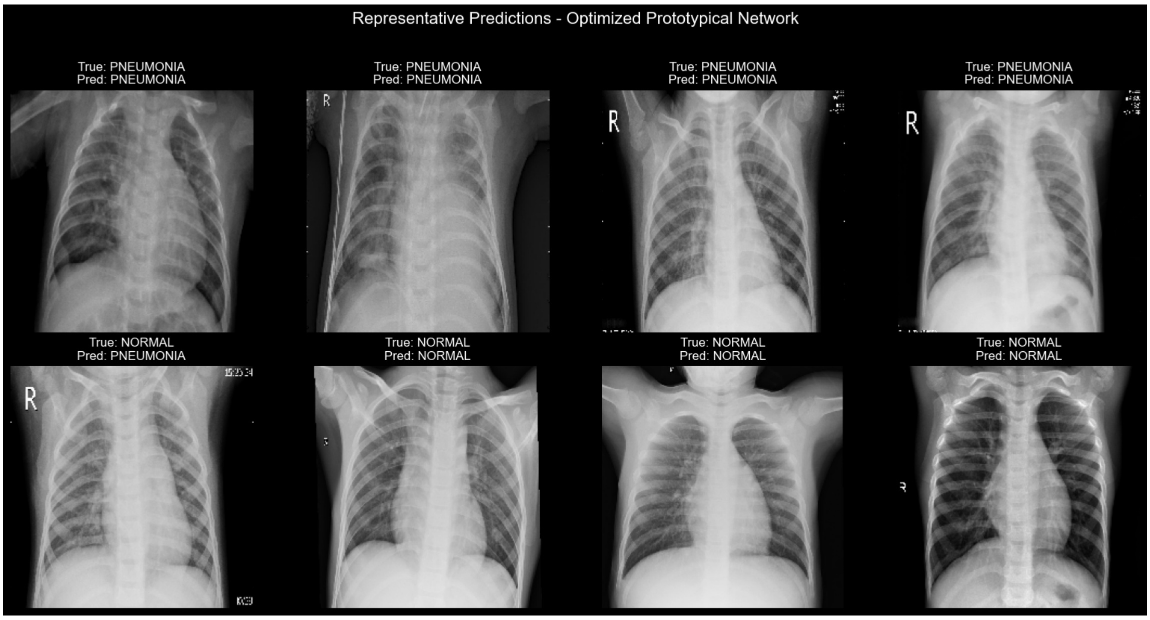

Figure 9: Optimized Prototypical Network Predictions: Representative chest X-rays with predicted labels from the optimized inference pipeline demonstrate competitive accuracy despite presenting the model with minimal training data (total 16 samples) per episode.

5 Conclusion

This study presents a comprehensive comparative analysis between a Baseline Transfer Learning Model and a few-shot learning approach using Prototypical Networks for chest X-ray classification for pneumonia detection. Although the baseline model, which represents the domain's current state-of-the-art, achieves slightly higher accuracy, the few-shot approach—trained with only 8 support and 8 query images per class. Thus, a significant reduction in training time (1047 s vs. 18 s) takes place, as well as relatively efficient inference (which was optimized from 0.2784 s to 0.121 s per image). These improvements make this few-shot approach particularly well-suited for rapid, real-time diagnostic implementations in resource-constrained environments, where data samples or computing capabilities are more limited. Potential future explorations will explore the application of few-shot learning to other imaging contexts and incorporating adaptive prototype updating for continuous learning.

Keywords: Chest X-Ray Classification, Transfer Learning, Few-Shot Learning, Prototypical Networks, Training Efficiency, Real-Time Diagnostics

Code: https://github.com/btomlinson237/FewShotLearningForPneumonia

References

[1] National Institutes of Health. Nih chest x-rays: Sample dataset. 2017. Accessed: April 2024.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[3] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, 2017.