National College of Ireland

Few-Shot Thoracic Disease Classification Using Prototypical Networks

MSc Research Project
Data Analytics

Ramsha Amir
Student ID: X23218215

School of Computing
National College of Ireland

Supervisor: Aaloka Anant

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ramsha Amir
Student ID:	X23218215
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Aaloka Anant
Submission Due Date:	12/12/2024
Project Title:	Few-Shot Thoracic Disease Classification Using Prototypical Networks
Word Count:	9595
Page Count:	28

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ramsha Amir
Date:	28th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Few-Shot Thoracic Disease Classification Using Prototypical Networks

Ramsha Amir
X23218215

Abstract

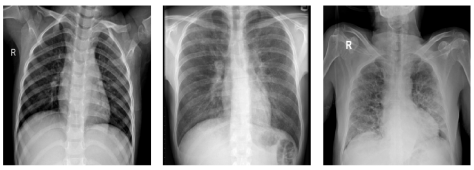
Thoracic diseases are among the most common health challenges worldwide, requiring accurate diagnostic tools to ensure timely intervention and better treatment outcomes. These diseases encompass a wide range of conditions affecting the lungs, heart, and other key organs within a thoracic cavity, which poses significant challenges for early diagnosis and effective treatment. For instance, diseases such as pneumonia, cardiomegaly, and COVID-19 lead to several complications ranging from respiratory failure and in some cases even death, hence the need for timely diagnosis to improve patient outcomes. However, the lack of labeled medical image datasets for these diseases makes it difficult to create reliable diagnostic models. This study investigates few-shot learning models, particularly Prototypical Networks into practice, for thoracic disease detection using minimal annotated chest X-ray images. Few-shot learning has proved significant for medical imaging, where large annotated datasets are not available. The approach proposed makes use of pre-trained models like VGG19, ResNet50, or DenseNet121 for feature extraction will subsequently classify diseases well with very few samples per category.

Key Words: X-ray Image, Few-Shot Learning, VGG19, ResNet50, DenseNet121, Prototypical Networks, Thoracic Disease Detection

1 Introduction

A chest radiograph commonly known as chest X-ray, is the most common imaging diagnostic used in the diagnosis of abnormalities in the airways, blood vessels, bones, heart, lungs, and other areas within the chest cavity @Çallı, Soganciogl u, van Ginneken, van Leeuwen and Murphy (2021).

Conditions such as pneumonia, COVID19, cardiomegaly if not detected promptly can lead to severe complications including respiratory failure and mortality. Early and accurate diagnosis is important, as the prevalence of these diseases are increasing across the world. The figure 1 below shows a normal x-ray and x-ray in which pneumonia and covid is detected.



(a) Normal (b) Pneumonia (c) COVD-19

Figure 1: Normal Vs Pneumonia and Covid 19 X-ray Images

The greatest challenge in the medical image domain is the diversity of the data set in terms of the number of samples. While the public data sets have recently tried to expand, these data sets are still limited and deal with the specific problem space. It is very expensive to enlarge such data sets, both in terms of increasing their number and diversifying them to best represent the problem. Moreover, the procedure to label these data sets is a process that should be worked under the supervision of radiologists and researchers together Soysal et al. (2023).

To overcome the challenges of limited data researchers have explored techniques such as artificial dataset augmentation and the application of machine learning algorithms to address data limitation. Few-shot learning algorithms are really highly effective. Unlike traditional deep convolutional neural network which require large amount of labelled data and cause overfitting if trained on small datasets, few-shot learning allows models to learn efficiently from a small number of labelled samples Cores et al. (2022).

By using few-shot learning techniques, with an emphasis on prototypical networks, this study seeks to close the diagnostic gap in thoracic diseases. The substantial difficulties in obtaining extensive, labelled medical datasets make this method necessary. Medical imaging files frequently need to be expertly annotated by radiologists, which is an expensive and time-consuming procedure. In order to overcome this constraint, few-shot learning allows models to learn efficiently from a small number of labelled samples.The goal of few-shot learning is to produce useful learning results using a small amount of labelled data in the training dataset, which consists of examples of inputs and their matching results.

1.1 Research Question

"How effectively can Prototypical Networks, combined with pre-trained feature extractors like VGG19, ResNet50, and DenseNet121, detect thoracic diseases from chest X-ray images in scenarios with minimal labeled data?"

1.2 Research Aim

This research primarily aims to create a few-shot learning framework that detects and classifies thoracic pathologies from NIH xrays. This research also aims to address the problems of limited labeled data using advanced pre-trained feature extraction and few-shot learning techniques to improve diagnosis and accuracy.

1.3 Research Objectives

The objective of this research is to implement few-shot learning using prototypical networks for thoracic disease classification and by utilising pretrained feature extractors such as VGG19, ResNet50 and DenseNet121. The list below highlights the steps carried out for developing the thoracic disease classification for this research:

- Conduct exploratory data analysis to understand the thoracic disease classes in the NIH Chest X-ray dataset.
- Preprocessing the images by applying augmentation and clahe and using feature extraction using pre-trained models such as VGG19, ResNet50, and DenseNet121.
- Developing few-shot learning framework using Prototypical Networks for thoracic disease classification with limited training samples.
- Applying 3-shot and 5-shot learning techniques are used in this framework to estimate its performance with different feature extractors.
- Asses the performance of the model based on the evaluation metrics

2 Related Work

Existing research on thoracic diseases classification reveals the methods, architectures, and approaches that have shaped the models in this field. By reviewing the previous system architectures, techniques, and datasets, this work will bring forward the advantages and disadvantages of each approach. Such aspects are essential needs in solving challenges while developing models with a particular emphasis on few-shot learning. This review is intended to bring forward the limitations of the existing methods and thus provide a basis for enhancement towards more efficient and accurate thoracic disease detection models.

2.1 Thoracic Disease Detection with CNNs

Deep learning's capacity to automate and improve diagnostic accuracy, especially with chest X-ray images, has led to its increasing use in the diagnosis of thoracic diseases. Most systems are like each other in that they start with the preprocessing of X-ray images before applying deep learning model, classifying the disease present in the images. Although most of the studies developed under this framework show similarities regarding the models and techniques considered, approaches, and data sets can be significantly different from one to another. It has worked on different thoracic diseases, though each method has its strengths and limitations discussed in the following studies.

Although big datasets are a plus for deep learning models, manual annotation poses a significant challenge in acquiring this kind of data, especially when it comes to multi-center or crowd-sourced data, owing to ethical and privacy issues. Some new innovations such as domain adaptation, generative models-like GANs, and semi-supervised learning seem very promising solutions for overcoming such limitations, but they are greedy for huge data and may become very tedious in training models. And active learning, which deals with uncertain predictions to be manually corrected, has proven beneficial in reducing burdens of annotation, though expert intervention is not completely offset.

To summarize, while CNNs have become a dominant technology for segmentation and classification in medical imaging, their generalization error, variability in data, and the lack of huge, annotated datasets remain stumbling blocks towards their deployment into clinical practice.

The development by Pant et al. (2021) is based on research work that focused on creating a deep learning system with the aim of early detection and diagnosis of thoracic ailments through chest X-ray images. The model, trained from the NIH Chest X-ray Dataset, was based on VGG16-based Convolutional Neural Networks (CNNs) that achieved high accuracies of 99% for training and 96.14% for validation. This was intended to be a great help to clinicians, especially in emergency settings, by automating the detection. It was tested by different classes of thoracic diseases, such as normal and pathological conditions, and had a web-oriented GUI for practical clinical end-use applications.

Although the system performed well, the authors pointed out a number of drawbacks, such as its reliance on a sizable, carefully selected dataset that might not translate well to smaller or more diverse datasets. Such limitation arises from applying only the VGG16 architecture within the progress of developing the envisioned deep learning architecture. the high stability of the model across the classes of diseases and different imaging scenarios. Some features such as image segmentation and spatial object detection weren't investigated while the system was designed for read posterior-anterior view of chest X-rays which brings a limitation as to its use in the environment where different imaging views or type are used. The study also extended the application to a mobile application to increase accessibility; however, real-time processing and constraint of existing devices still pose problems.

Abbas et al. (2021) n their research devised DeTraC which is a convolutional neural network aimed at classifying chest X-rays images of COVID-19 patients. DeTraC further segments classes to sub classes in order to induce refined predictions by addressing inconsistencies across the datasets through decomposing the classes into sub classes and utilised transfer learning with pre-trained models to achieve great accuracy in detecting COVID-19 from diverse global datasets. However, the working of its techniques may encounter hurdles when applied to a small, specialized dataset especially with a broader collection of thoracic diseases. Besides, the intricate architecture of DeTraC and that it requires significant computational resources might limit its adaptability to datasets with lower image count and multiple disease classes. Also, the other constraint in this research is that it only focuses on a single disease, COVID-19, and therefore, the model may not be directly applicable to the classification of multiple thoracic diseases with different characteristics.

The research conducted by Kesim et al. (2019) reviews the utility of small convolutional neural networks in real-time applications for chest X-ray images instead of using large pre-trained networks which are inefficient in terms of generalization for various imaging modalities. The proposed network attained 86% accuracy classification of images into twelve classes with real-time classification under one second on an embedded system. The present work evaluated images for data imbalance through augmentation and different CNN models using the ChestX-ray 14 dataset. The paper presents the necessity of small

network architecture for efficient functioning at embedded systems while proposing future work such as mobile application development and dataset enrichment. However, the study primarily depends on the partial dataset which poses an important limitation with the trade-offs of smaller networks against classification accuracy in very complex conditions.

2.1.1 Limitations of CNNs in Medical Imaging

CNNs, despite having achieved a lot in the field of medical imaging, have several limitations, one of which is the aspect on the clinical side. Generalization across datasets has proven to be one of the many hinderances to the feasibility of CNNs in clinical practice. It is especially problematic when the existence of different kinds of machines, either by vendor or model, bring training data or differences in acquisition parameters. These pose differences in aspects such as contrast, resolution, and signal-to-noise ratio, and consequently lead to a steep fall in model performance because these CNNs have the tendency to overfit to purely the statistical attributes of the data. The generalization gap then results because they cannot generalize well to novel or unseen distributionsPerone and Cohen-Adad (2019).

Although large datasets are advantageous for deep learning models, manual annotation for this kind of data acquisition has become quite a challenge, especially for multi-center or crowd sourced data due to ethical and privacy issues. Emerging domain adaptation, generative models like GANs, and semi-supervised learning appeared to be very promising solutions to overcome such limitations; however, they are hungry for huge data and might be complicated in training models. And active learning, which deals with uncertain predictions to be manually corrected, has proven to be useful in easing burdens of annotation, but does not remove the expert intervention totally Perone and Cohen-Adad (2019).

In brief, even though CNNs have significant utility in tasks like segmentation and classification within medical imaging, there are still challenges of generalization, data variability, and large annotated datasets preventing them from being deployed across the clinical spectrum. Perone and Cohen-Adad (2019)

2.2 Few-Shot Learning

Deep learning systems have made significant strides in many tasks within the medical domain, but they suffer from the same training data dependency as above. Few-shot learning algorithms typically try to deracinate this dependency through exploiting information available from very small amounts of data. Given that most diseases occur only very rarely, available data are minimal within the context of medical imaging, thus making the success of few-shot learning algorithms indeed a very potent advance.

2.2.1 Few-Shot Learning Algorithm

Few shot learning algorithms, as the name suggests, are primarily tasked with providing accurate results with a limited amount of data. The few shot learning algorithms are expected to represent small amounts of data in such a way, that it can generalize to represent a much broader range of data Kotia et al. (2021).

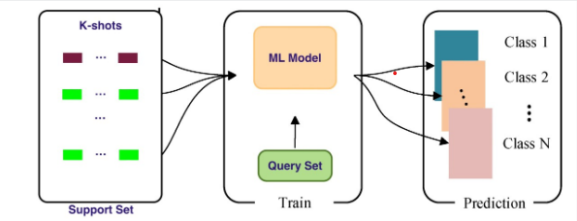


Figure 2: Few-shot Learning Process

Instead of having enormous training data like traditional supervised learning, few-shot learning models adapt to novel tasks quickly using prior knowledge and experiential learning with some extra data. Such models would require only a few cases to be able to make accurate predictions on something new by finding the important features and basic structure that make up a concept Shashi Thota (2023). A well known application of this design is in robotics, where this similarity function manages to allow rapid adaption for a new task using mappings between classes in the query and support sets with minimal additional information. The query set, in this case, refers to the examples in both old and fresh categories on which the model is generalized, while performance is evaluated based on the support set made up of a few labeled examples for every novel category of data.

Support Set (S): A small set of labeled example where each example consists of a data point and its corresponding label.DataCamp (2020)

Query Set (Q): A set of unlabeled data points for which the model needs to predict labels.DataCamp (2020)

2.2.2 Few-Shot Learning Techniques

Few-Shot Learning has emerged as a critical technique in machine learning. The main techniques in FSL include Metric Learning, Meta-Learning, and Transfer Learning, each of which approaches the problem of limited data from different perspectives.

Metric Learning: Metric Learning becomes an important central part in the framework of Few-Shot Learning. Basically, learning should be seen as the ability to estimate a distance metric or learn an embedment space such that the distance between instances defines their similarity. For example, now models in few-shot tasks can compare very few annotated examples with fresh data points based on similarity scores because they have already learned this. Two among the most frequently studied metric learning methods include Siamese Networks and Triplet Networks. These are generally policies of the model that act to minimize or maximize distances in the embedding space. A good example would be applications in medical imaging-facing verification or signature verification Shashi Thota (2023).

Meta Learning: Meta-learning helps for quick adaptation of the models with few data available for the new task. In meta-learning, training is conducted on several tasks that build strategies to generalize their learning. Such learned strategies can subsequently

be tuned to perform on a different task with very few examples.Model-Agnostic Meta-Learning and Prototypical Networks are very well known. For example, MAML can prepare the model for quick adaptation by training the model on how to effectively learn a good initial parameter set that will subsequently make it suitable for future data with minimal adaptivity. Prototypical Networks, in contrast, create a 'prototype' (aver- age representation) for each class and then determine to which prototype each unlabeled observation is closest. Such methods are particularly productive in the diagnostic task of medical image classifications where fewer labeled data are available Shashi Thota (2023).

Transfer Learning: Transfer learning is one of the most widely used methods in Few-Shot Learning. This technique takes an already trained model on a big dataset and fine-tunes it on a small dataset for a specific purpose. Such a method utilizes the knowledge gained by the model from the larger corpus of data and allows it to realize a new task on very few labeled data. It is one of the popular applications of such an idea as fine-tuning and domain adaptation. For instance, in medical image analysis, it might mean fine-tuning the original model trained on a larger dataset like ImageNet to diagnose rare diseases from very small datasets, boosting the likelihood of detection substantially Shashi Thota (2023).

2.2.3 Applications of Few-Shot Learning in Medical Imaging

Few-shot learning applications have shown great promise in dealing with the challenges posed by sparse medical imaging datasets, especially associated with rare disease Han et al. (2022), new methodologies have been suggested for screening ophthalmic diseases based on few-shot-learning and data augmentation techniques, for instance with detection for early onset of diabetic retinopathy, age related macular degeneration, and glaucoma. Integration of a combined dataset along with simulating different fundus cameras for data augmentation by style transfer will enhance the generalization ability of the model.

The approach advanced uses metric-based few-shot learning models such as the siamese network and it aims to accomplish high precision in disease classification with little data. With data augmentation, different metrics—accuracy, recall, and F1 scores—improved in handling the challenge of misclassification. The study presents how few-shot learning combined with innovative data augmentation techniques may make AI solutions more productive through easy access to medical imaging-with a focus on ophthalmology Han et al. (2022). The study also limits the generalization of the findings due to the use of synthetic data augmentation that does not capture the real world variability completely. Therefore, it would not generalize on unseen clinical data. Further, it confines its findings to a specific set of ophthalmic diseases, making them relevant only in this domain of medical imaging.

Another case study by Prabhu et al. (2019) is on diagnostic dermatology that adopts clinical image classification mainly under the shadows of long tailed data distribution and high intra class variability problems. Proposed by the authors of the present paper, the Prototypical Clustering Networks (PCN) learns a weighted combination of prototypes for each class, classifying this weighted combination in terms of the similarity between it and the input sample. The whole PCN idea is framed as a few-shot learning problem, the aim of which is to generalize a new classifier to understand new diseases given a very

small number of labeled examples. By using initial prototypes from clustering, which are refined as training is carried out, PCN improves generalization and performs particularly well for dermatological diseases.

Metric-learning approaches Hoffer and Ailon (2015) learn the optimal distance metric by comparing target examples and a small number of labeled samples in an embedding space. The goal is to create projection function mapping images into this space in such a way that images from the same class are brought closer together, while different classes are pushed farther apart. The principal assumption is that the feature representations learned for the base classes can be transferred to novel classes. Our solution borrows from the meta-learning paradigm, where a generic chest X-ray classifier is first trained on base classes and then adapted to novel classes with just a few labeled samples.

In a research study carried out by Anandhi et al. (2024), the author develops a new framework for diagnosing lung diseases from the image of chest X-rays by utilizing a combination of several techniques of deep learning and few-shot learning. The framework performs preprocessing using Contrast Limited Adaptive Histogram Equalization to enhance the quality of the images, followed by the segmentation of lung nodules using the UNET model. Features are extracted through a pre-trained CheXNet model using the Chest X-ray14 dataset.The framework was tested on lung diseases such as tuberculosis, pneumonia, and lung cancer, outperforming the state of the art. The study pointed out that such strategies for meta-learning could deliver output on the aspects of data scarcity and improved diagnostic accuracy in medical imaging. This framework pays attention to lung disease diagnosis, as well as to segmentation through UNET, which isolates lung nodules ahead of classification.

Soysal et al. (2023) presents research on improving disease detection in medical images through the integration of zero-shot learning with semantic ontology information, all under demanding conditions associated with limited data and labeling. The researchers addressed data scarcity and the inherent subjectivity found in the analysis of medical images by utilizing the ChestX-ray14 dataset, which is extremely rich in unlabeled chest X-ray images. With the adoption of a DBpedia ontology semantic class label, ZSL enabled the detection of unseen classes, and the authors state that such an approach when married to a ResNet50 neural network proved fruitful even in the gray areas of application where figures speak the hardest. The highest precision value scored was 29.59 but could not be far from the assertion of proponents that the paradigm of integrating ZSL and ontology improves medical image analysis on cases where labeled data could not survive, and further optimization of the idea would be expected for future work.

Another research conducted by Kshitiz et al. (2023), which evaluates and compares different few-shot learning models such as ProtoNet, MatchingNet, MAML, DSN, and a Proposed Model in the case of thoracic disease detection with the chest X-ray dataset with limited labeled data. The various diseases tested in the models included Fibrosis, Hernia, Pneumonia, Mass, Nodule, Pleural Thickening, Cardiomegaly, Edema, Emphysema, Consolidation, Effusion, Pneumothorax, Atelectasis, Infiltration, and No Finding. Results showed that DSN generally performs better than other models, achieving maximum accuracy for diseases like Edema and Pneumothorax where ProtoNet and MatchingNet

also performed well but had limitations in some diseases like Fibrosis and Pleural Thickening. MAML had a very low success rate for Hernia and Infiltration diseases compared to all others. The Proposed model was competitive, beating MAML for most diseases, but it has shown varied performance mainly on No Finding and Pleural Thickening. This study identifies the need for many strides to make few-shot learning applicable in medical imaging in order to improve robustness and consistency with the many diseases.

Authors	Methodology	Findings
Pant et al. (2021)	Used VGG16-based CNN trained on NIH Chest X-ray dataset for thoracic disease detection.	Achieved 99% training and 96.14% validation accuracy; limited by reliance on large datasets and restricted imaging views.
Abbas et al. (2021)	Developed DeTraC using transfer learning and decomposing classes for refined predictions.	High accuracy with global datasets; challenges with small datasets and high computational requirements
Kesim et al. (2019)	Proposed small CNN for real-time classification using ChestX-ray 14 dataset on embedded systems.	Achieved 86% accuracy; efficient for embedded systems but limited by dataset size and complex condition classification trade-offs.
Thota et al. (2023)	Explored Metric Learning techniques, including Siamese and Triplet Networks.	Demonstrated potential for classification tasks with small datasets; key for similarity-based classification like medical imaging and signature verification.
Han et al. (2022)	Proposed a few-shot learning model combined with data augmentation for ophthalmic disease detection.	Improved model accuracy and generalization; limited by reliance on synthetic data augmentation, which doesn't fully capture real-world variability.
Prabhu et al. (2019)	Introduced Prototypical Clustering Networks (PCN) for diagnostic dermatology under few-shot settings.	Effective for long-tailed data distributions and intra-class variability; achieved state-of-the-art performance for dermatological conditions.
Hoffer et al. (2015)	Developed a metric-learning approach for embedding space classification in medical imaging.	Enabled robust classification of new diseases with few labeled samples; challenges with generalization to unseen classes.
Kshitiz et al. (2023)	Compared ProtoNet, MatchingNet, MAML, DSN, and a Proposed Model for thoracic disease detection.	Found DSN and ProtoNet performed best for specific diseases like Edema and Pneumothorax; MAML showed weaker performance on diseases like Hernia and Infiltration.
Anandhi et al. (2024)	Applied CLAHE for preprocessing, UNET for segmentation, and Prototypical Networks for classification.	Outperformed state-of-the-art methods for lung diseases; highlighted the role of preprocessing and feature extraction in few-shot learning.
Soysal et al. (2023)	Improved disease detection in medical images by integrating zero-shot learning with semantic ontology information using the ChestX-ray14 dataset	ZSL allowed detection of unseen classes, and results showed precision up to 29.59\% (for matching at least one label)

Figure 3: Summary of Literature Review

To conclude, this paper on deep learning for thoracic disease detection states that substantial advancements have been made, particularly through CNN techniques, in defining conditions such as pneumonia, pneumothorax, and atelectasis.Few-shot learning is well positioned to remedy this situation; it allows accurate predictions with minimal labeled data. Metric learning, meta-learning, and transfer learning techniques are now considered by many as the forerunners in few-shot learning, offering frameworks that are ideally placed to handle scarcity in medical imaging data. The success despite these is still dragging the generalization challenges, variability in data, and computational complexity. Future investigations should therefore aim at improving the adaptability of models to diverse datasets and also on the performance of those few-shot learning systems in a clinical environment. More importantly, innovation in the existing methodologies by addressing limitations will be far reaching in transforming thoracic disease detection into applicable, accessible, and efficient technologies in the real world.

3 Methodology

This section will outline the methodology for developing few-shot learning using a prototypical network for thoracic disease classification in this research context as an answer to the research question. The research adopted a variation of the CRISP-DM process, as depicted in Figure 4 below.

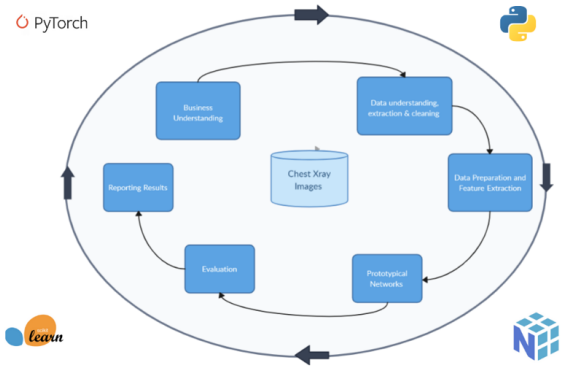


Figure 4: CRISP-DM Methodology

3.1 Business Understanding

Research for this project is mainly focused to study the health challenges of thoracic disease detection in the custodian of Prototypical Networks as transfer learning. Early and precise detection of diseases such as pneumonia, cardiomegaly, or atelectasis is crucial for positive patient outcomes and minimizing health costs. Acquiring large, labeled datasets for training models is a paramount challenges. The objective of this research is to implement transfer learning to pre-trained model feature extraction and few-shot learning by Prototypical Networks to realize a practical and scalable diagnostic solution for such low resource healthcare settings.

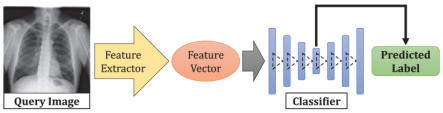


Figure 5: Key Components

3.2 Data Gathering

The NIH Chest X-ray Dataset of Health (2018) is the main dataset that was used in this study, along with an additional COVID dataset Fusicfenta (2020) that had cases of COVID-19, which were absent from the NIH dataset. Both of the datasets are available in Kaggle and the links for these datasets are available in the reference section. The datasets were explored and prepared in a systematic manner according to the specifications for few-shot learning.

3.3 Exploratory Data Analysis

During this initial exploring stage, the collected images were filtered and extracted based on their extensions such as .png, .jpg and .jpeg.It simplified the operation of identifying valid image files without involving unsupported formats that may have been introduced and may cause errors in training the model.

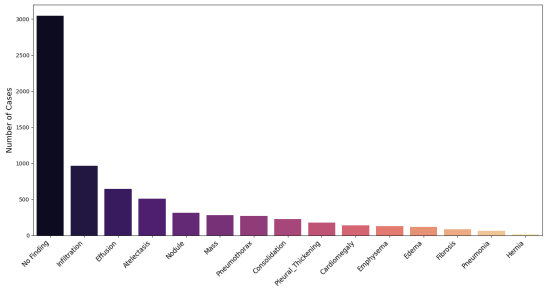


Figure 6: Class distribution of NIH Dataset

The class distribution analysis in figure 5 showed a considerable imbalance in the databases. For instance, common diseases such as No Finding have thousands of pictures, but quite the opposite is true regarding rare diseases, such as Hernia which is represented by very few. To address this, and guided by the few-shot learning framework, we selected 30 images per class for six diseases related to the thorax. This decision was made in reference to:

- **Consistency with Few-Shot Learning:** Few-shot learning techniques are designed to function effectively on small datasets. Selecting 30 images per class provides a balance between sufficient representation and maintaining a manageable dataset size. This number strikes a compromise between having enough data for learning and not overloading the system.
- **Class Uniformity:** Ensuring all classes have the same number of images minimizes the impact of class imbalance on the training process, thereby promoting fairness and consistency in model performance.

In this study, we adopted a systematic data balancing strategy using a class imbalance remedy.The first step of processing data was filtering the PA view X-rays to make them radiographically conform to the original sample set as uniformity in the X-ray images was important, as they were collected from different angles, creating different features and properties that would have potentially interfered with model learning.

Grouping as an important aspect in few-shot learning has a important role as it gives some organized way to examine the ability of the model to differentiate closely related classes in the presence of very limited data samples. The unique challenge lies in making all possible meaningful observations from a few labeled sample and grouping diseases together makes it possible to accomplish realistic learning tasks that simulate real work.

In this study, the diseases are grouped into three pairs such as Effusion and COVID-19 (Group 1), Pneumonia and Cardiomegaly (Group 2), and Atelectasis and Fibrosis (Group 3). These groups were built on clinical similarity and radiological overlap, a challenge to the model for diagnostically relevant tasks related to these two diseases. For instance, Effusion and COVID-19: both these entities typically display their presence in chest X-rays with fluid accumulation or diffuse opacities. The model would necessarily need to discriminate between subtle differences of patterns and distributions between these two diseases. Pneumonia and Cardiomegaly are two common diseases in clinical practice; they share an overlap of certain regions on chest X-rays, making it hard to separate pulmonary infection from cardiac enlargement. Likewise, Atelectasis and Fibrosis make a certain structural change in the tissues of the lungs, such as collapse or scarring; both of these would lead the model to pay more attention to texture, density, and certain radiological markers.

3.4 Modelling

The few-shot learning techniques in the Modeling section of this project are primarily adapted to the detection of thoracic diseases from X-ray images with a minimum labeled dataset. Prototypical Networks will be the main model. Feature extraction will be performed by ResNet50, VGG19, and DenseNet121 for small data by transfer learning. Cross-validation, data augmentation, and weighted loss functions must be applied during training to correct the class imbalance for the thoracic diseases and for prevention of overfitting. The performance was evaluated in terms of several measures like accuracy, precision, recall and accordingly comparisons were made to find the best-performing model in the context of limited labeled data.

3.5 Evaluation

To achieve these objectives, it is ineluctably necessary to identify the evaluation methods that have been used in prior studies. For instance, Kshitiz et al. (2023) proposed a software solution for evaluating performance scores of X-ray image classification systems by numerical and visual methods. Their evaluation framework is based on metrics, further enhanced by several additional visualization techniques to give more comprehensive insights into model performances.

4 Design Specification

A system diagram in Figure 7 above elaborately outlines the workflow describing classification of different thoracic diseases such as Covid-19, Effusion, Cardiomegaly, Pneumonia, Fibrosis, and Atelectasis.It begins with input x-ray images that are further enhanced by application of Contrast Limited Adaptive Histogram Equalization. Then, using deep learning models like VGG19, ResNet50, and DenseNet121, feature extraction processes applied to the segmented images create a feature vector which will be used to derive a prototype image depicting the condition and calculate a distance metric to compare the input image with the prototype for similarity. The process is aimed at classification and diagnosis of thoracic diseases using few-shot learning particularly prototypical network.

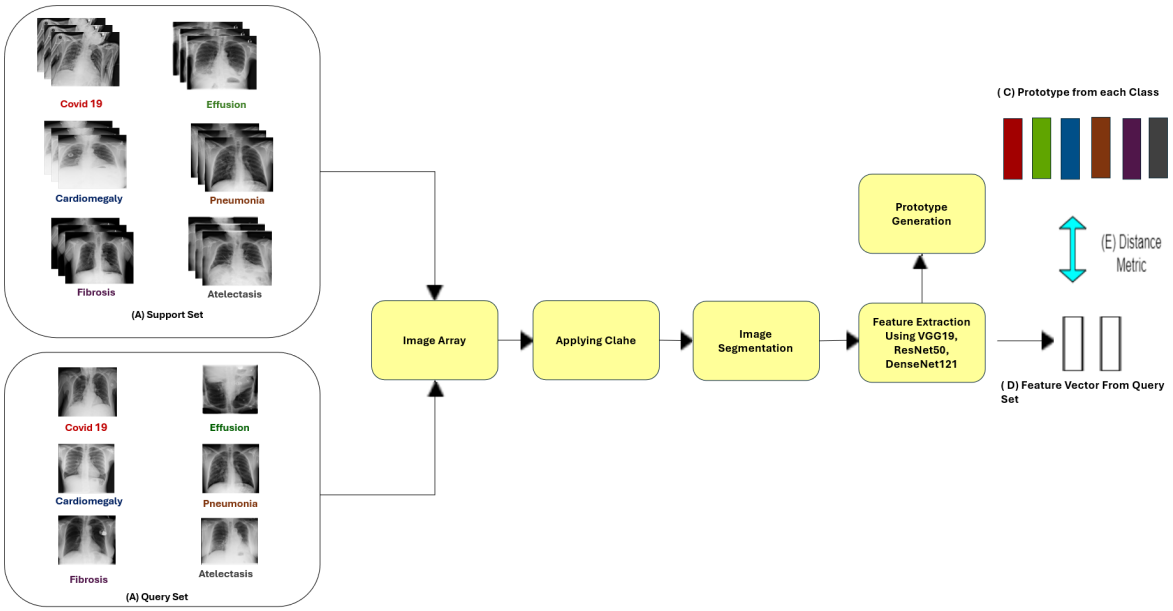


Figure 7: Design Specification Diagram

5 Implementation

The implementation of this thesis focuses few-shot learning techniques to the problem of thoracic disease detection from the NIH Chest X-ray dataset. It also involves pre-processing, feature extraction through ResNet50, and Prototypical Networks to effectively classify images with little labeled training. This entire process will be done using Google Colab with a T4 GPU for efficient and scalable computation while training and evaluating the model.

5.1 Data Preparation

Several pre-processing techniques were applied to ensure compatibility with the Prototypical Network model and to improve model performance:

- **Resizing:** All the images that are adjusted to dimensions of 224×224×3. This is because resizing the image is very important for meeting the input needs of the model such as that of a pre-trained deep learning model like ResNet50. Most models require an image of a specific size as input; for example, 224x224 pixels which is the size used to train most of them. Apart from that, it also ensures that all the images are standard and equalized in terms of resolution

- **CLAHE :** Contrast Limited Adaptive Histogram Equalization (CLAHE) has been used to render features in the images visible. An advanced image enhancement technique, CLAHE built an image by enhancing the distribution of pixel intensity values across every pixel of the whole image. Unlike traditional histogram equalization, CLAHE limited enhancement in regions where excessive contrast occurs, which preserves the quality of the image. CLAHE is very useful for medical images, such as those where the small differences in various regions are easier for the model to identify with diseases like pneumonia and cardiomegaly.

- **Data Augmentation :** Different data augmentation techniques have been used to make the model more robust and generalizable include:
  - **Horizontal Flipping**: Under this condition, images are flipped horizontally, thus allowing the model to view the different positions of the same object. This results in enlarging the dataset. The other aspect of the technique, which is important in the medical image, is that it represents possible angles in which the disease may be detectable in real-world use, therefore helping the model generalize better against memorization orientation.
  - **Zooming**: This involves random zoom transformations, giving rise to different scales and perspectives of the features within the sample images. If the image zooms in or zooms out, then the model will not be very sensitive to the size or location changes of the features, as it does occur with real-world images.

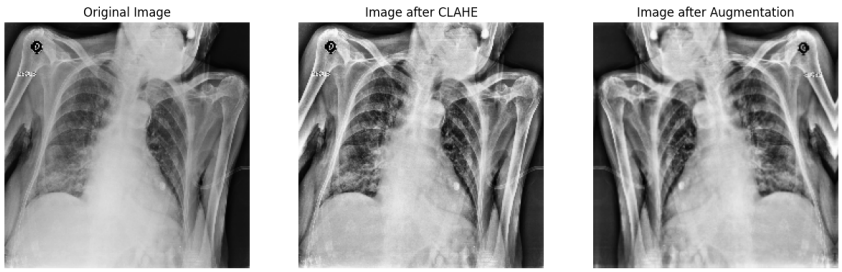


Figure 8: X-ray image with clahe and segmentation

5.2 Feature Extraction

The extraction of features in medical imaging is paramount for automated capturing of complicated features from images for incorporation in training representations. For this research we employed three pretrained deep learning models: VGG19 , ResNet50 , DenseNet121.The models selected for this work were VGG19, ResNet50, and DenseNet121, each selected for their unique architectural advantages. VGG19 was considered the solid base model for feature extraction because of its deep layers and simple yet effective convolutional structure, rendering it robust for an image classification task. ResNet50, with its residual connections through which the vanishing gradient is prevented, would allow for the training of deeper networks, while still holding good performance. DenseNet121 enables richer gradient flow because of dense connections between layers and has been proven to yield better performance with fewer parameters. Thus, they are selected based on the successful experimentation for image classification tasks but also for their predominantly common use in medical imaging tasks, thereby indicating their fitness for the complex patterns exhibited in chest X-ray data. These feature extractors make it possible to recover high-level image features critical for disease detection while minimizing the need for long, labeled data, which is one of the advantages in working with small datasets.

5.3 Prototypical Network

The Prototypical Network is used for overcoming the challenge of few-shot learning, particularly to address issues in medical image classification where labeled data are quite scarce. The prototypical Network model operates in a metric learning perspective where the prototype of a class is created based on the mean of the feature embeddings of the support set images. VGG19,DenseNet, and ResNet have been used as feature extractors in the network, and they are trained to map an input image to a very high-dimensional feature space. Euclidean distance,as suggested by Snell et al. (2017), between the query image embeddings and the class prototypes extracted from the support images once the feature vectors are obtained. The model assigns the query image to the class whose prototype is the closest one. In this way, the model learns the data structure so as to generalize well with a limited training set. The loss function of the network is based on the negative log of the probability of the true class label using softmax, which encourages the model to minimize classification errors. During testing, the Prototypical Network predicts the class of a query image by comparing its feature vector to the precomputed prototypes. This kind of modeling proves most of benefit for the task of thoracic disease detection, where one class has very few images versus the other classes, and there are generally few labeled samples for training. Using Prototypical Network, this work intends to enhance classification performance with few data, making it a very strong tool for applying few-shot learning in medical imaging, where collecting too many datasets most often becomes impractical.

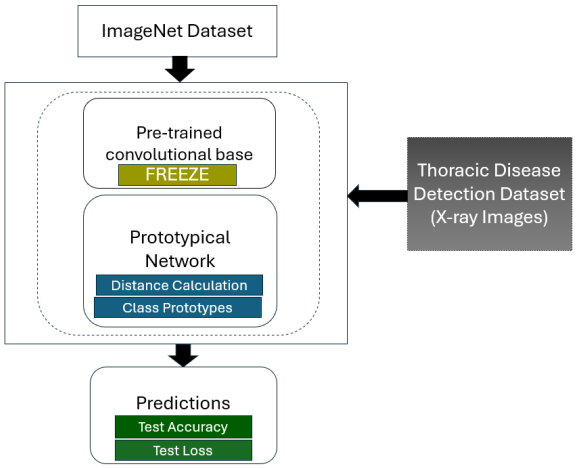


Figure 9: Thoracic Disease Detection Model Architecture

Having employed pre-trained weights of models for the transfer-learning process, certain layers were frozen to hold on to the learned feature representations from ImageNet without re-training them from scratch. This makes sure that the models can exploit both low- and mid-level learned features without overfitting on limited medical dataset. The final layers of these models were modified to adapt to the use case of thoracic disease classification. For this, the classifier layers were modified instead of the fully connected layers at the end to facilitate simple output in such a manner that the model emits a 1D vector of features for use by the subsequent classification task. The output layers were altered; besides that, the feature extraction layers were frozen to make the models more effective in the

identification of diseases in the thoracic cavity, as they will be able to provide a robust representation of the features needed for the research using this smaller data collection.

5.4 Episodic Learning

Episodic learning is a core concept in few-shot learning, where the model is trained and tested in a way that mimics the real-world scenarios of classifying unseen classes based on a limited number of labeled examples. This implementation leverages episodic learning using Prototypical Networks to test three popular backbone architectures: VGG19, ResNet50, and DenseNet121, in the context of medical imaging, specifically thoracic disease classification.

5.4.1 Classification Tasks in Episodic Learning

In this research, the classification assignments are framed within a few-shot learning paradigm to thoroughly assess the model's performance using limited amounts of labeled data. Each task pertains to a 2-way learning classification between two classes using support sets, comprising relatively small, labeled data and query sets, which are made up of unlabeled examples. The tasks are framed in such a way that they can adequately probe into the model's ability to generalize across different shot settings and diverse medical image data

- **2 Way Learning :** In this study we will be using 2-Way Learning, each task implies distinguishing between two distinct classes, chosen at each episode from the database for support and query sets.This two-way approach simplifies the learning problem so that evaluation of the learning and generalization capacity of the model is more targeted within a low-data environment. These experiments were then carried out across three divisions, with every division measuring performance under the varied architectures used in the experiment: VGG19, ResNet50, and DenseNet121 functionalities so as to make common comparisons.

Groups	**Classes**
Group 1	Effusion , Covid19
Group 2	Pneumonia , Cardiomegaly
Group 3	Atelectasis , Fibrosis

Table 1: Groups with their respective classes

- **Shot Settings** : Shot Settings are the numbers of labeled examples per class in the support set. This plays an important role in the model's ability to learn and generalize with few samples. In this research I will be doing 3-shot setting, the support set comprises only 3 labeled examples per class. Thus, a total of 6 examples for both classes combined. This setting forces the model to generalize from a very limited amount of labeled examples, which simulates the real-world medical imaging situation in which such labeled data is very scarce. Also we will implement a 5-shot setting, there are 5 labeled examples per class, making 10 examples from the two classes total. This increases the amount of class-specific knowledge learned in the model, allowing it to be understood with the additional labeled data and remake the outputs better. Furthermore, a 5-shot analysis would provide insight into how the accuracy of the model changes when learning has more label data to train on.

- **Query Settings** : Unlabeled examples from both classes in the support set compose the Query Set. These examples helps evaluating the model's proficiency in classifying fresh observations from properties learned from the support set. The support set helps the model discover its distinguishing traits for each class. Thus, the query set examples are classified. The query set is also defined in terms of the ability to generalize of the model because it brings data not in the training (support set) to bear on the judgment of model performance. This is like the common real-world diagnostic task problem, where one expects models to predict fresh unseen instances based on fresh knowledge learned from a limited set of labeled experiences.For 3 shot setting we have configured 10 query set x-ray images and for the 5 shot setting we have configured 15 query set x-ray images.

5.4.2 Testing setup

The underlying premise of the experiment is the application of episodic testing as a process. 5000 episodes are conducted for this experiment, and it is enough statistically reliable. Running 5000 episodes minimizes variance to most performance metrics and guarantees that the assessment of the generalization ability of this model toward heterogeneous tasks is quite accurate. This is because the class combination can vary depending on the tasks in real world, requiring an adaptive model. The classes have not been picked randomly, but pre-defined in terms of groups. In this way, diversity in the model testing is achieved with respect to sets of combinations of classes.

Metrics such as loss and accuracy are evaluated across these episodes to see how well the model performs. Overall, the loss is the difference between the output prediction and and the true labels assigned, while accuracy is the percentage of query set examples identified correctly in terms of classifications. Hence, these two fair metrics will be averaged over 5000 such episodes for a solid and trusted assessment of the model under consideration towards its performance evaluation.

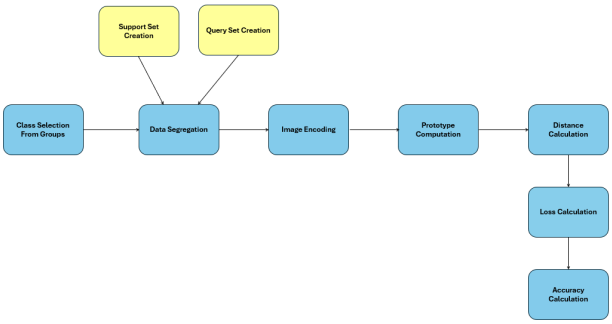


Figure 10: Workflow of Episodic Learning in ProtoNet

The experimental use of the models includes the Prototypical Network, which operates on generating prototypes for each support-set class. Simply put, a prototype is an average feature embedding for each class. To classify the query set, the model calculates the Euclidean distance between the feature of the query examples and the class prototypes

then it assigns to the examples the label of the nearest prototype. In other words, the distance is minimized in this process so that the model can accurately classify query images.The figure 10 above shows the workflow of episode learning in prototypical network

6 Evaluation

In this section different experiments are carried out for thoracic disease classification to get the best performing configurations. For the Evaluation we will use f1-score, precision and recall metrics. Tools like confusion matrices and ROC curves will also be used.

6.1 Experiment 1: Effusion and Covid19 Classification

In this experiment we will use the 3 selected feature extractors which are VGG19,ResNet50 and DenseNet121 with prototypical network for the diseases in group 1 which are COVID19 and Effusion.

6.1.1 VGG19 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 1: VGG19 was trained in a 2-Way 3 shot setting (figure 11) to classify between two class only with 3 examples per class. The confusion matrix shows that the model performed such that precision, recall and F1 score values are above 0.80 both for classes, indicating very high classification accuracies. Additional evidence for strong performance on the part of the model is through the ROC curve, which demonstrates that the area under curve (AUC) value approaches 1.

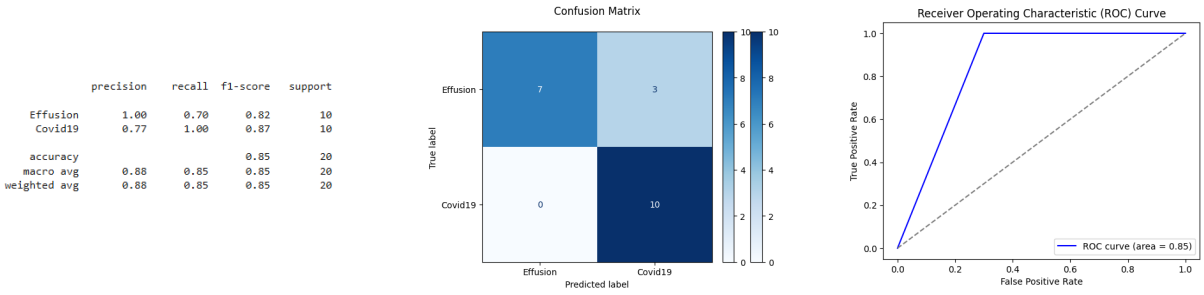


Figure 11: VGG19 Results 2 Way 3 Shot Group 1

2-Way 5-Shot Learning Results Group 1: VGG19 was trained in a 2-Way 5 shot setting in figure 12 below the confusion matrix shows that the model performed such that precision, recall and F1 score values are above 0.85 both for classes, indicating very high classification accuracies. Additional evidence for strong performance on the part of the model is through the ROC curve, which demonstrates that the area under curve (AUC) value approaches 1.

6.1.2 ResNet50 with Prototypical Network Analysis

In this experiment we will use ResNet50 feature extractor with prototypical network for the disease Covid19 and Effusion

	precision	recall	f1-score	support
Effusion	0.87	0.87	0.87	15
Covid19	0.87	0.87	0.87	15
accuracy			0.87	30
macro avg	0.87	0.87	0.87	30
weighted avg	0.87	0.87	0.87	30

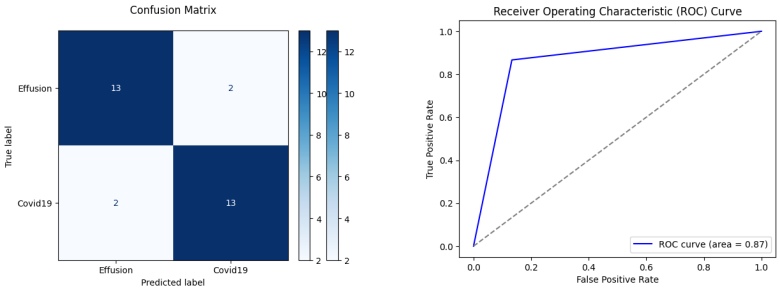


Figure 12: VGG19 Results 2 Way 5 Shot Group 1

2-Way 3-Shot Learning Results Group 1: ResNet50 was trained in a 2-Way 3 shot setting (figure 13) to classify between two class only with 3 examples per class. The confusion matrix shows that the model performed such that precision, recall and F1 score values are 0.85 both for classes, indicating very high classification accuracies. Additional evidence for strong performance on the part of the model is through the ROC curve, which demonstrates that the area under curve (AUC) value approaches 1.

	precision	recall	f1-score	support
Effusion	0.82	0.90	0.86	10
Covid19	0.89	0.80	0.84	10
accuracy			0.85	20
macro avg	0.85	0.85	0.85	20
weighted avg	0.85	0.85	0.85	20

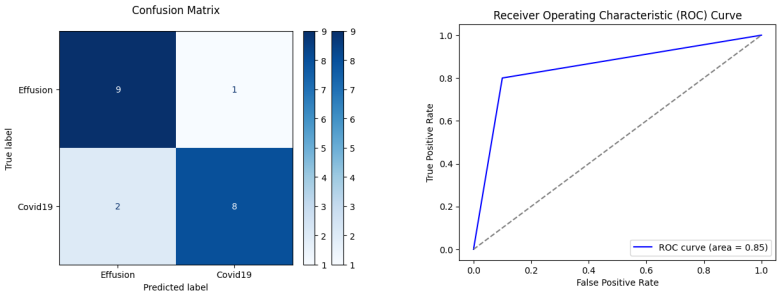


Figure 13: ResNet50 Results 2 Way 3 Shot Group 1

2-Way 5-Shot Learning Results Group 1: ResNet50 was trained in a 2-Way 5 shot setting (figure 14) to classify between two class only with 5 examples per class. The confusion matrix shows that the model performed such that precision, recall and F1 score values above 0.85 both for classes, indicating very high classification accuracies. Additional evidence for strong performance on the part of the model is through the ROC curve and it is performing better than 2 way 3 shot configurations.

	precision	recall	f1-score	support
Covid19	1.00	0.73	0.85	15
Effusion	0.79	1.00	0.88	15
accuracy			0.87	30
macro avg	0.89	0.87	0.86	30
weighted avg	0.89	0.87	0.86	30

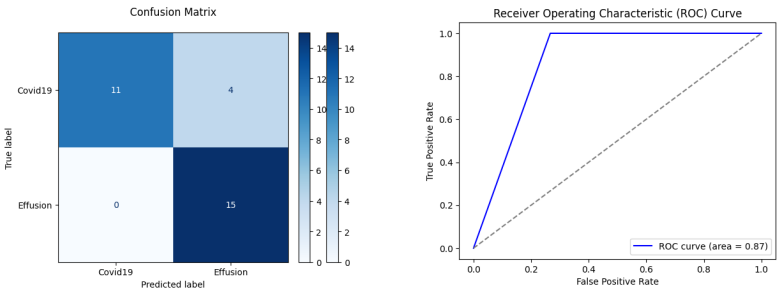
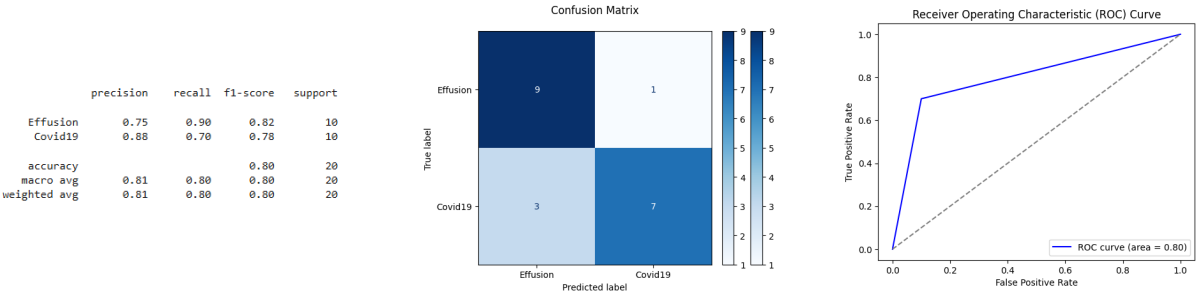


Figure 14: ResNet50 Results 2 Way 3 Shot Group 1

6.1.3 DenseNet121 with Prototypical Network Analysis

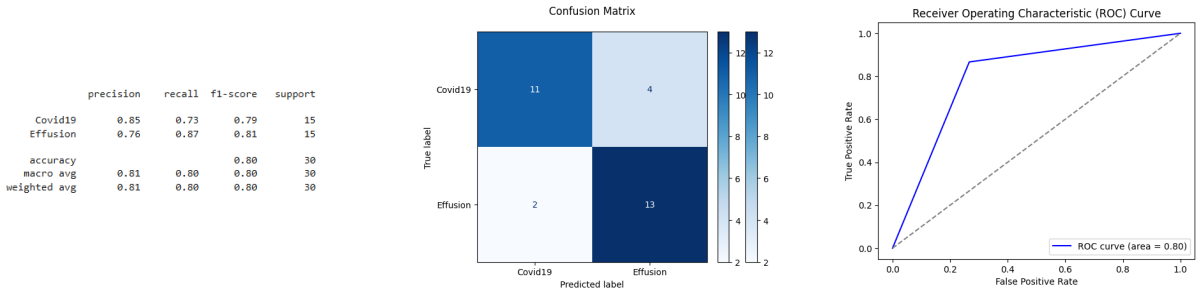
2-Way 3-Shot Learning Results Group 1: DenseNet121 was trained in a 2-Way 3 shot setting (figure 15) to classify between two class only with 3 examples per class. The confusion matrix shows that the model performed such that precision, recall and F1 score values is 80 both for classes, indicating good accuracies. Additional evidence for strong performance on the part of the model is through the ROC curve can be seen.



	precision	recall	f1-score	support
Effusion	0.75	0.90	0.82	10
Covid19	0.88	0.70	0.78	10
accuracy			0.80	20
macro avg	0.81	0.80	0.80	20
weighted avg	0.81	0.80	0.80	20

Figure 15: DenseNet121 Results 2 Way 3 Shot

2-Way 5-Shot Learning Results Group 1: DenseNet121 was trained in a 2-Way 5 shot setting (figure 16) to classify between two class only with 3 examples per class. The confusion matrix shows that the model performed such that precision, recall and F1 score values is 80 both for classes, indicating good accuracies and same as the previous densenet configurations.



	precision	recall	f1-score	support
Covid19	0.85	0.73	0.79	15
Effusion	0.76	0.87	0.81	15
accuracy			0.80	30
macro avg	0.81	0.80	0.80	30
weighted avg	0.81	0.80	0.80	30

Figure 16: DenseNet121 Results 2 Way 5 Shot

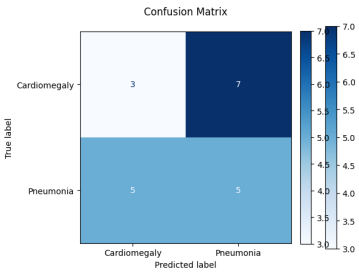
6.2 Experiment 2: Pneumonia and Cardiomegaly Classification

In this experiment we will use the 3 selected feature extractors which are VGG19,ResNet50 and DenseNet121 with prototypical network for the diseases in group 1 which are Pneumonia and Cardiomegaly.

6.2.1 VGG19 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 2: With respect to the 2-way 3-shot results in figure 17, they are significantly modest because of very low performance under the condition where Cardiomegaly achieved a 0.38 precision and a 0.30 recall, whereas Pneumonia's scores achieved 0.42 precision and 0.50 recall. The accuracy overall was 40%, and the F1-scores (0.33 for Cardiomegaly and 0.45 for Pneumonia) suggest struggles to balance precision and recall.

	precision	recall	f1-score	support
Cardiomegaly	0.38	0.30	0.33	10
Pneumonia	0.42	0.50	0.45	10
accuracy			0.40	20
macro avg	0.40	0.40	0.39	20
weighted avg	0.40	0.40	0.39	20



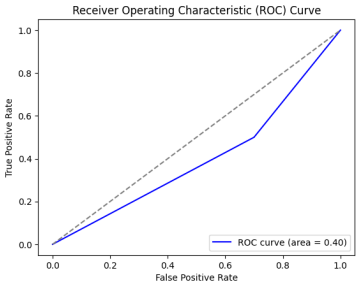
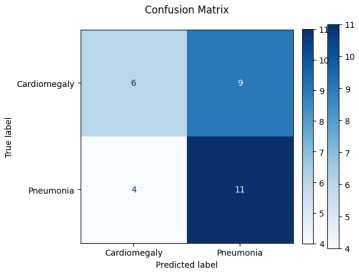


Figure 17: VGG19 Results 2 Way 3 Shot Group 2

2-Way 5-Shot Learning Results Group 2: 2-way 5-shot results in figure 18 are little better, it is because in the latter condition they used more support for the bigger support set. The precision for Cardiomegaly increased to 0.60, but the recall remained the same at 0.40. Pneumonia, on the other hand, had better metrics, such as 0.55 precision and 0.73 recall. The overall accuracy increased and reached 57%, with much higher F1 scores showing better generalization, especially for Pneumonia.

	precision	recall	f1-score	support
Cardiomegaly	0.60	0.40	0.48	15
Pneumonia	0.55	0.73	0.63	15
accuracy			0.57	30
macro avg	0.57	0.57	0.55	30
weighted avg	0.57	0.57	0.55	30



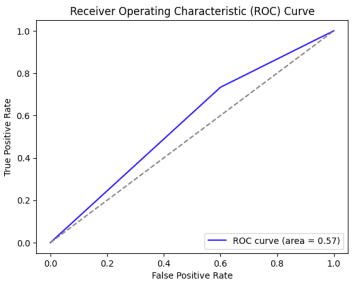
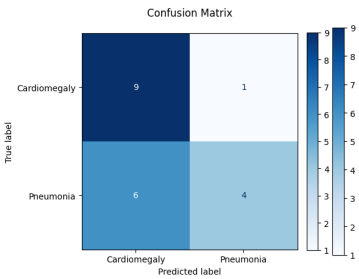


Figure 18: VGG19 Results 2 Way 5 Shot Group 2

6.2.2 ResNet50 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 2

	precision	recall	f1-score	support
Cardiomegaly	0.60	0.90	0.72	10
Pneumonia	0.80	0.40	0.53	10
accuracy			0.65	20
macro avg	0.70	0.65	0.63	20
weighted avg	0.70	0.65	0.63	20



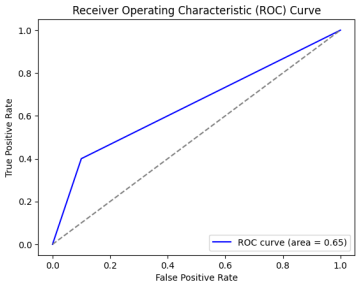


Figure 19: ResNet50 Results 2 Way 3 Shot Group 2

The 2-way 3-shot findings in figure 19 using ResNet50 tag strong performance for Cardiomegaly, citing a rather high recall of 0.90 and an associated F1-score of 0.72, which indicates that positive true detection is a good feature. Pneumonia managed to get precision at 0.80 but had much lower recall (0.40), leading to a middle-level F1-score of 0.53. Overall accuracy was 65% with a significant contribution from the Cardiomegaly result.

2-Way 5-Shot Learning Results Group 2: For the 2-way 5-shot results in figure 20, both classes gave much more balanced metrics because of the larger support set. Pneumonia precision was found to be 0.69 and its recall stood at 0.73, whereas the performance for Cardiomegaly was found to be slightly better at 0.71 for precision and 0.67 for recall. Both classes pushed close to F1 score of about 0.70, resulting in a final accuracy of 70%. These results indicate the ability of ResNet50 to improve this class balance and accuracy with greater support samples.

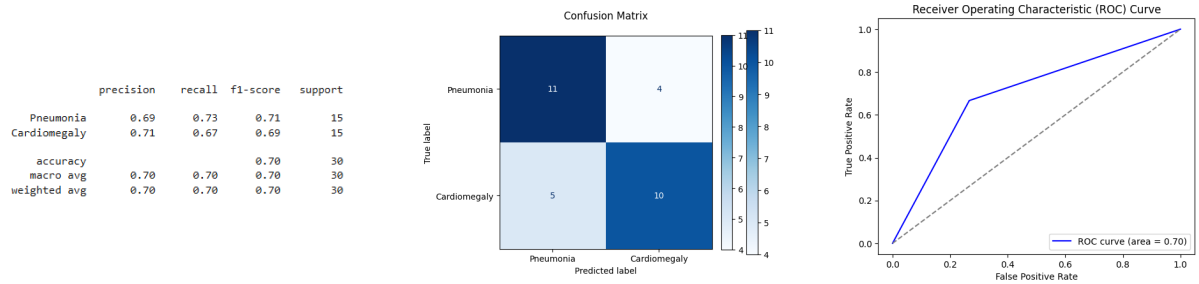


Figure 20: ResNet50 Results 2 Way 5 Shot Group 2

6.2.3 DenseNet121 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 2 In the case of the 2-way 3-shot experiment results in figure 21, DenseNet yielded just a moderate 55% accuracy performance. Thus, it was posted at 0.57 for precision in Cardiomegaly, but low recall (0.40) yielded an F1-score of only 0.47. In the condition of Pneumonia, although precision of 0.54 was determined, the high recall (0.70) allowed improvement of the F1-score to 0.61.

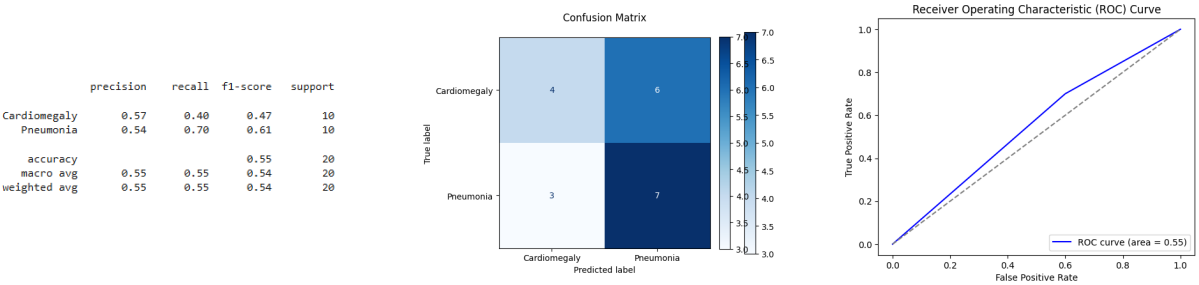


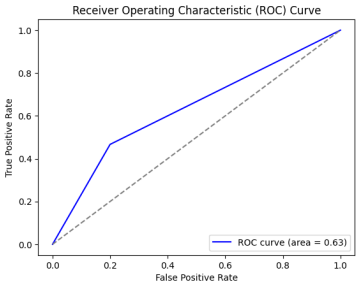
Figure 21: DenseNet121 Results 2 Way 3 Shot Group 2

2-Way 5-Shot Learning Results Group 2 In 2-way 5-shot experiment results in figure 22 improvements in performance were observed as much as 63 %. For Pneumonia, precision is 0.60, and recall increases to 0.80, yielding a strong F1-score of 0.69. Cardiomegaly, too, showed a precision level of 0.70, but the lower recall (0.47) gave an F1-score of 0.56.

6.3 Experiment 3: Atelectasis and Fibrosis Classification

In this experiment we will use the 3 selected feature extractors which are VGG19,ResNet50and DenseNet121 with prototypical network for the diseases in group 3 which are telectasis and Fibrosis

	precision	recall	f1-score	support
Pneumonia	0.60	0.80	0.69	15
Cardiomegaly	0.70	0.47	0.56	15
accuracy			0.63	30
macro avg	0.65	0.63	0.62	30
weighted avg	0.65	0.63	0.62	30



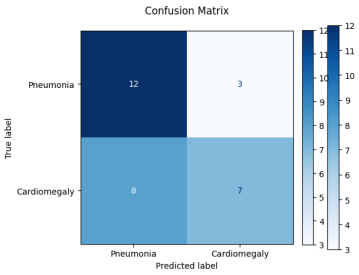
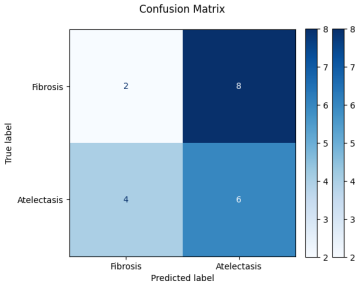


Figure 22: DenseNet121 Results 2 Way 3 Shot Group 2

6.3.1 VGG19 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 3: In two-way three-shot VGG 19 results in figure 23, model performance can be rated as 40%, and there exists a performance gap between different classes. For Fibrosis, there is a precision of 0.33 and a recall of 0.20, leading down to an F1 score of approximately 0.25, denoting a very poor detection ability

	precision	recall	f1-score	support
Fibrosis	0.33	0.20	0.25	10
Atelectasis	0.43	0.60	0.50	10
accuracy			0.40	20
macro avg	0.38	0.40	0.38	20
weighted avg	0.38	0.40	0.38	20



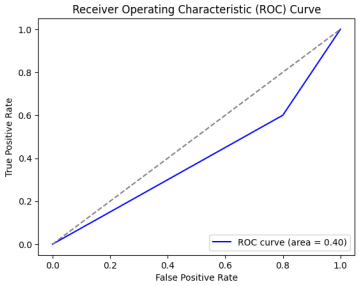
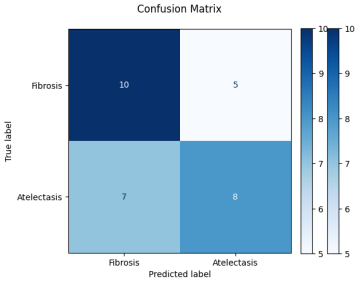


Figure 23: VGG19 Results 2 Way 3 Shot Group 3

2-Way 5-Shot Learning Results In two-way five-shot VGG 19 results figure 24, the accuracy of the model improves up to 60%, surpassing the previous score and showing a drastic enhancement in model performance. For Fibrosis, precision rises to 0.59 while recall improves to 0.67, summing with an F1 score of 0.62, indicating that performance had improved. The same pattern is seen with Atelectasis where 0.62 precision and 0.53 recall resulted in an F1 score of 0.57, showing detection better than the 3 shots. The macro average f1 scores increase to 0.60.The weighted average also rises to 0.60, showing a much better overall performance with the 5-shot setup.

	precision	recall	f1-score	support
Fibrosis	0.59	0.67	0.62	15
Atelectasis	0.62	0.53	0.57	15
accuracy			0.60	30
macro avg	0.60	0.60	0.60	30
weighted avg	0.60	0.60	0.60	30



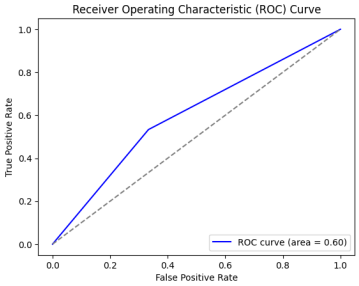


Figure 24: VGG19 Results 2 Way 5 Shot Group 3

6.3.2 ResNet50 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 3: The model achieves 55% accuracy from the results of 2-way 3-shot ResNet. Recall is good for Atelectasis, at 0.70, while Fibrosis lags with a recall of 0.40; hence the average F1 is at 0.54.

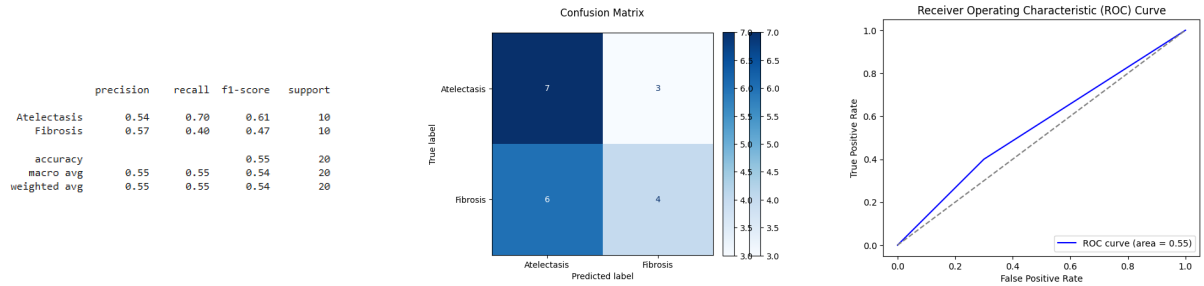


Figure 25: ResNet50 Results 2 Way 3 Shot Group 3

2-Way 5-Shot Learning Results Group 3: For the 2-way 5-shot ResNet results, accuracy is slightly down at 53%, with Atelectasis doing better than Fibrosis according to their respective F1-scores (0.59 and 0.46). Essentially, this does not change the overall observation: a macro average F1-score of 0.52.

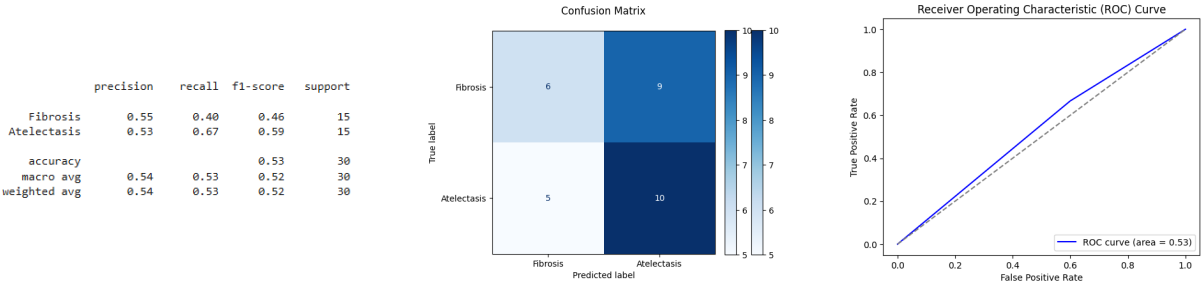


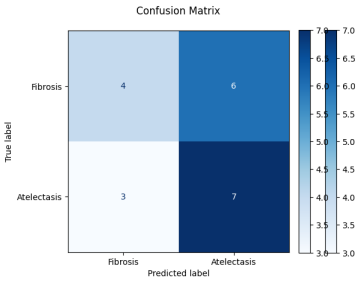
Figure 26: ResNet50 Results 2 way 5 shot Group 3 :

6.3.3 DenseNet121 with Prototypical Network Analysis

2-Way 3-Shot Learning Results Group 3: With the 2-way 3-shot results from DenseNet figure 27, a 55% accuracy is obtained. Atelectasis achieves a higher recall of 0.70, which results in a better F1-score of 0.61. However, Fibrosis has a low recall of 0.40, which gives an F1 score of 0.47. The overall performance is well balanced with both the macro and weighted averages being at 0.54.

2-Way 5-Shot Learning Results Group 3: For example, in the 2-way 5-shot DenseNet results in figure 28, they again fall flat at 50% accuracy, where both classes have very similar precisions and recalls. Fibrosis has a recall of 0.53 and an F1 score of 0.52, while Atelectasis appears a bit weaker with a recall of 0.47 and an F1 score of 0.48. But overall performance seems to show a consistent average of 0.50, both macro and weighted.

	precision	recall	f1-score	support
Fibrosis	0.57	0.40	0.47	10
Atelectasis	0.54	0.70	0.61	10
accuracy			0.55	20
macro avg	0.55	0.55	0.54	20
weighted avg	0.55	0.55	0.54	20



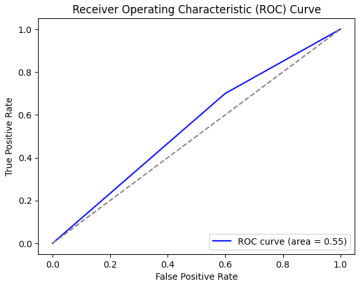
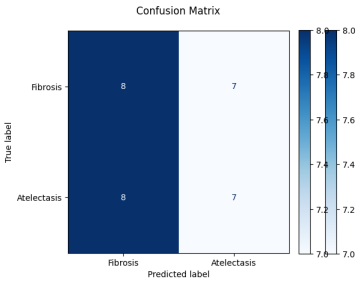


Figure 27: DenseNet121 Results 2 way 3 shot Group 3

	precision	recall	f1-score	support
Fibrosis	0.50	0.53	0.52	15
Atelectasis	0.50	0.47	0.48	15
accuracy			0.50	30
macro avg	0.50	0.50	0.50	30
weighted avg	0.50	0.50	0.50	30



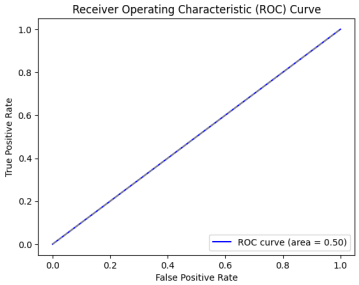


Figure 28: DenseNet121 Results 2 way 5 shot Group 3

6.4 Discussion

Various feature extractors, including VGG19, ResNet50, and DenseNet121, were put under the test with prototypical networks against each other in experiments detecting thoracic disease and results can be seen in the figure 29 below. Their purpose was to determine how well they performed in classifying diseases such as Effusion, COVID19, Pneumonia, Cardiomegaly, Atelectasis, and Fibrosis. The results indicated that VGG19 and ResNet50 performed well regardless of whether the materials used had greater or lower shot configurations such as with 5-shot input where even precision, recall, and F1-scores typically surpassed the threshold of 0.80. The analysis conducted showed that the curves for the area under the curve (AUC) were close to 1, demonstrating good accuracy. While DenseNet121 performed at a fair average precision and recall, it did not perform very well under the 3-shot cases, yet delivered respectable F1-scores, ranging between 0.50 to 0.69 across various conditions for the diseases under study. The performance for diseases such as Atelectasis and Fibrosis differed largely from the rest, with VGG19 considerably more accurate as the shot count increased, while DenseNet121 showed slight improvements with a 5-shot approach. Overall, experiments here call for larger support sets (5-shot) in few-shot learning tasks as they would necessarily be expected to show better class balance and improved performance, particularly on diseases that are often more difficult to detect such as Atelectasis and Fibrosis.

Based on the results, it can be deduced that ResNet50 serves as a more efficient feature extractor for all disease categories as it outperformed all others, including VGG19 and DenseNet121 in terms of metrics like recall and precision at varying shot configurations. The relatively balanced 3-shot and 5-shot performances thus proved robustness in disease classification among various other diseases such as Effusion, Pneumonia, and Cardiomegaly. Differing from VGG19, which varies on the basis of shot conditions in most instances, DenseNet121 disappoints in working with relatively smaller datasets. Hence, ResNet50

Disease Group	Feature Extractor	2-Way 3-Shot Precision	2-Way 3-Shot Recall	2-Way 3-Shot F1-Score	2-Way 5-Shot Precision	2-Way 5-Shot Recall	2-Way 5-Shot F1-Score
Group 1 Effusion & Covid	VGG19	0.88	0.85	0.85	0.87	0.87	0.87
	ResNet50	0.85	0.85	0.85	0.89	0.87	0.86
	DenseNet121	0.81	0.80	0.80	0.81	0.80	0.80
Group 2 Pneumonia & Cardiomegaly	VGG19	0.40	0.40	0.39	0.57	0.57	0.55
	ResNet50	0.70	0.65	0.63	0.70	0.70	0.70
	DenseNet121	0.55	0.55	0.54	0.65	0.63	0.62
Group 3 Atelectasis &Fibrosis	VGG19	0.38	0.40	0.40	0.60	0.60	0.60
	ResNet50	0.55	0.55	0.54	0.54	0.53	0.52
	DenseNet121	0.55	0.55	0.54	0.50	0.50	0.50

Figure 29: Results Summary

becomes more reliable and accurate in all the categories tested, making it most suitable for this analysis.

7 Conclusion and Future Work

Deep learning techniques through prototypical networks along with feature extractors such as VGG19, ResNet50 and DenseNet121 can be successfully applied for classification of thoracic diseases classification using few-shot learning. Among the three feature extractors, ResNet50 consistently outperformed the other two in all recall, precision, and f1 score across diseases which makes it the most suitable model for disease classification in this context. However, there was still much room for improvement in ResNet50 itself and such a robust feature extractor posed strong constraints in identifying more difficult diseases like Fibrosis. This means that further fine-tuning and even architectural or feature-extraction adjustments would be necessary in order to deliver more reliable results for such difficult to diagnose conditions. The results show that even with very limited training samples, deep learning models can compete, with ResNet50 striking a good balance between class precision and recall. Overall, the findings highlight how few- shot learning brings promise for medical image classification, even in some of the hardest situations with limited extractable data.

Future work could include fine-tuning the pre-trained models like ResNet50, VGG19 and DenseNet121, especially for difficult diseases like Fibrosis, using layer-wise fine-tuning or selective retraining methods. Besides, data augmentation and semi-supervised learning techniques can also be studied to improve performance at reduced labeling. Moreover, others of few-shot learning paradigms like Matching Networks or Model-Agnostic Meta-Learning may also lead to improvements. Ensemble methods may combine different models strengths; cross-disease generalization could be realized through multi-task learning or domain adaptation for enhancing detection in multiple thoracic diseases. Deeper insights into the clinical context, for example, patient history and demographics, can further improve model performance and real-world interpretability.

References

Abbas, A., Abdelsamea, M. M. and Gaber, M. M. (2021). Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network, *Applied Intelligence* **51**(2): 854–864.
URL: *https://doi.org/10.1007/s10489-020-01829-7*

Anandhi, S., Devakadacham, S. R., Manikandan, J. and Shanthalakshmi, M. (2024). Enhancing lung disease diagnosis through meta learning: A framework utilizing few-shot learning techniques, *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–6.

Cores, D., Vila-Blanco, N., Pérez-Alarcón, M. et al. (2022). A few-shot approach for covid-19 screening in standard and portable chest x-ray images, *Scientific Reports* **12**: 21511. Received 16 August 2022, Accepted 05 December 2022, Published 13 December 2022.
URL: *https://doi.org/10.1038/s41598-022-25754-6*

DataCamp (2020). What is few-shot learning? Accessed: 2024-12-06.
URL: *https://www.datacamp.com/blog/what-is-few-shot-learning*

Fusicfenta (2020). Chest x-ray for covid-19 detection. Accessed: 2024-12-11.
URL: *https://www.kaggle.com/datasets/fusicfenta/chest-xray-for-covid19-detection*

Han, Z., Xing, H., Yang, B. and Hong, C. (2022). A few-shot learning-based eye diseases screening method, *European Review for Medical and Pharmacological Sciences* **26**(23): 8660–8674.

Hoffer, E. and Ailon, N. (2015). Deep metric learning using triplet network, *International Workshop on Similarity-Based Pattern Recognition*, Springer, pp. 84–92.

Kesim, E., Dokur, Z. and Olmez, T. (2019). X-ray chest image classification by a small-sized convolutional neural network, *2019 Scientific Meeting on Electrical-Electronics Biomedical Engineering and Computer Science (EBBT)*, pp. 1–5.

Kotia, J., Kotwal, A., Bharti, R. and Mangrulkar, R. (2021). *Few Shot Learning for Medical Imaging*, Springer International Publishing, Cham, pp. 107–132.
URL: *https://doi.org/10.1007/978-3-030-50641-4*$_7$

Kshitiz, Garg, G. and Paul, A. (2023). Few-shot diagnosis of chest x-rays using an ensemble of random discriminative subspaces.

of Health, N. I. (2018). Chest x-ray images (sample). Accessed: 2024-12-11.
URL: *https://www.kaggle.com/datasets/nih-chest-xrays/sample*

Pant, H., Lohani, M. C., Bhatt, A. K., Pant, J. and Sharma, R. K. (2021). Thoracic disease detection using deep learning, *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1197–1203.

Perone, C. S. and Cohen-Adad, J. (2019). Promises and limitations of deep learning for medical image segmentation, *Journal of Medical Artificial Intelligence* **2**(0).
URL: *https://jmai.amegroups.org/article/view/4659*

Prabhu, V., Kannan, A., Ravuri, M., Chaplain, M., Sontag, D. and Amatriain, X. (2019). Few-shot learning for dermatological disease diagnosis, *in* F. Doshi-Velez, J. Fackler, K. Jung, D. Kale, R. Ranganath, B. Wallace and J. Wiens (eds), *Proceedings of the 4th Machine Learning for Healthcare Conference*, Vol. 106 of *Proceedings of Machine Learning Research*, PMLR, pp. 532–552.
URL: *https://proceedings.mlr.press/v106/prabhu19a.html*

Shashi Thota, Subrahmanyasarma Chitta, V. K. R. V. C. S. R. V. S. M. B. (2023). Few-shot learning in computer vision: Practical applications and techniques, *Human-Computer Interaction Perspectives* **3**(1): 29–59.
URL: *https://thesciencebrigade.com/hcip/article/view/332*

Snell, J., Swersky, K. and Zemel, R. S. (2017). Prototypical networks for few-shot learning, *CoRR* **abs/1703.05175**.
URL: *http://arxiv.org/abs/1703.05175*

Soysal, O. A., Guzel, M. S., Dikmen, M. and Bostanci, G. E. (2023). Common thorax diseases recognition using zero-shot learning with ontology in the multi-labeled chestx-ray14 data set, *IEEE Access* **11**: 27883–27892.
ΩÇallı et al.

Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K. G. and Murphy, K. (2021). Deep learning for chest x-ray analysis: A survey, *Medical Image Analysis* **72**: 102125.
URL: *https://doi.org/10.1016/j.media.2021.102125*