Improving Disease Classification on Rare Class distribution X-ray Images using Supervised and Few Shot Hybrid Learning

by

Soumodeep Biswas
21101291
Jannatul Nayeem
19301090
Md Ifty Rahman
24241357
Tanim Ahmed Saad
24241358
Sabiha Salam Ayesha
20201060

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2025

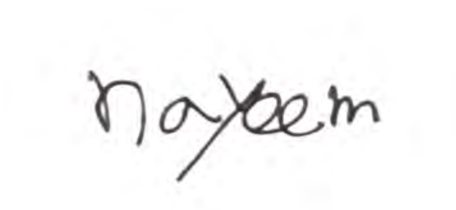
Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Soumodeep Biswas
21101291



Jannatul Nayeem
19301090

Md Ifty Rahman
24241357

Tanim Ahmed Saad
24241358

20201060
Sabiha Salam Ayesha

Approval

The thesis titled "Improving Disease Classification on Rare Class distribution X-ray Images using Supervised and Few Shot Hybrid Learning" submitted by

1. Soumodeep Biswas(21101291)
2. Jannatul Nayeem(19301090)
3. Md Ifty Rahman(24241357)
4. Tanim Ahmed Saad(24241358)
5. Sabiha Salam Ayesha(20201060)

Of Summer, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on October, 2025.

Examining Committee:

Supervisor:
(Member)

Md Tanzim Reza
Senior Lecturer
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Disease diagnosis through medical image analysis using various transfer learning models and neural networks have made significant progress in recent years. However, Medical Image datasets are highly imbalanced due to the minimal number of cases of rare diseases. As a result of this imbalance, pre-trained CNN models perform poorly in detecting rare diseases in supervised classification tasks. Classes with a high number of data samples dominate in conventional supervised learning setup. Therefore, our research focused on trying to minimize the effect of this class imbalance. We proposed a hybrid architecture which put together supervised learning and few-shot learning. For common class detection, we used a pre-trained MobileNet-V2 as the base model of the classical supervised learning. For Rare classes, a few shot learning model, Relation Network was responsible for detecting rare disease classes. Our proposed hybrid architecture achieved an average of 90% F1 score on the rare classes. In contrast, we experimented with 3 pre-trained CNN models for traditional supervised learning and observed that all of them had scored poor recall or precision value with an average of 45% F1 score on the rare classes. Therefore, our findings highlighted that our proposed hybrid architectural approach is more impactful and can achieve good results. For that reason, we believe our work opens the door for researchers for future study that a hybrid approach of combining supervised and few-shot learning can be effective instead of relying on only one.

Keywords: Supervised Learning; Few-shot Learning; Relation Network; Feature Embeddings; Support set; Query set;

Acknowledgement

First of all, we want to thank the almighty Allah, with his blessing and will we have been able to complete our thesis
We also want to say thank you to our supervisor Md. Tanzim Reza for his advice, support and guidance.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Deep neural networks (DNNs) have significantly advanced various domains by leveraging computational models that mimic the human brain's structure. These networks have revolutionized fields such as business, sports, biology, and especially healthcare, where DNNs have demonstrated remarkable efficacy in automating disease detection and classification. Particularly in the medical field, deep learning techniques have enabled rapid disease detection, assisting healthcare providers in diagnosing conditions efficiently from medical images, including X-rays LeCun, Bengio, and Hinton, 2015 [3]. Medical image classification using deep learning has been pivotal in diagnosing diseases such as pneumonia, tuberculosis, and lung cancer from chest X-ray CXR images Rajpurkar et al. 2017 [7].

However, despite the powerful capabilities of DNNs in medical image analysis, several limitations arise when it comes to rare diseases. A critical challenge in these applications is the scarcity of labeled data for rare conditions, which limits the generalizability of supervised learning models Chen et al. 2019 [13]. Factors such as patient privacy concerns, the rarity of specific diseases, and high costs of collecting medical images contribute to the data limitation in medical imaging Kermany et al. 2018 [11] . These challenges lead to models overfitting to the few available training samples, thus performing poorly when exposed to unseen data Kang et al. 2020 [13]. This issue is particularly concerning in the diagnosis of rare diseases, where the availability of images is often insufficient to build effective supervised models.

To address these challenges, several strategies have been proposed to adapt deep learning models for disease classification in domains with limited data. Transfer learning, where models pre-trained on large, diverse datasets are fine-tuned with smaller, task-specific datasets, has shown promise Pan and Yang 2010. [1] This technique allows for the transfer of learned features from large datasets to smaller ones, significantly improving performance in scenarios with limited data Shin et al. 2016 [4]. Additionally, few-shot learning (FSL) techniques have been introduced to enable models to classify diseases based on minimal labeled examples Lake et al. 2015. [2]. FSL approaches like metric learning, which learns to measure the similarity between examples, and meta-learning, which trains models to adapt quickly to new tasks, have shown success in overcoming the scarcity of labeled data Vinyals et al.2016 [5].

Hybrid learning frameworks that combine supervised learning with few-shot learning techniques have demonstrated great potential in improving the classification of rare disease classes in medical imaging. These hybrid models combine the strength of supervised learning's ability to learn from labeled data with the flexibility of few-shot learning to generalize from a small number of samples Ren et al. 2020 [20]. By utilizing both large pre-trained models and few-shot learning strategies, these hybrid frameworks can efficiently tackle the challenge of limited data while improving performance on rare diseases, making them highly effective in medical image classification tasks Chen et al. 2019 [13].

Furthermore, privacy concerns in medical data are another critical issue, particularly when working with sensitive patient information. Hybrid models can be adapted to be privacy-aware by focusing on techniques such as federated learning, which allows for decentralized model training across different institutions without the need to share raw medical data Yang et al. 2019 [17]. This approach ensures that patient privacy is maintained while still benefiting from large-scale model training.

In conclusion, the integration of few-shot learning with supervised approaches provides a robust framework for handling class imbalance and data scarcity in the medical domain. These hybrid architecture not only improve generalization and classification performance on rare disease classes but also contribute to developing efficient, privacy-aware, and data-efficient diagnostic tools.

By combining the strengths of these approaches, deep learning models can be adapted to overcome the inherent challenges in rare disease classification, making significant strides in the medical field Xie et al. 2020 [21]. Therefore, we propose to use a hybrid model that combines classical supervised learning and few-shot learning for rare disease classification of X-ray images on such dataset, which has high class imbalance, selected in our study.

1.2 Research Problem

Although deep neural network models can perform very well in classification tasks, it has been found that when the amount of data obtained in some classes is very insignificant or relatively less than other classes, which is often the case for rare diseases in medical X-ray images, in such situations, deep neural networks in classical supervised learning algorithms with conventional class balancing techniques often perform very poorly in diagnosing rare diseases. Therefore, it is very important to control the bias of the major classes of deep neural networks in a classical supervised learning setup to diagnose rare diseases efficiently and accurately. In our research, we focused on how to diagnose rare diseases in medical X-ray image datasets by freeing them from the dominant behavior of major common classes and at the same level as them comprising few-shot learning and supervised learning algorithms.

1.3 Research Objectives

Our research aims to develop a hybrid pipeline consisting of a deep neural network and a few shot learning to alleviate the negative effect of major common classes in medical x-ray image dataset on rare disease classes due to the severe class distribution imbalance. The followings are the goal of our study:

- To understand the limitations of Deep Neural Networks (DNNs) in imbalanced class distribution tasks.
- To understand the major role of DNNs in medical X-ray image classification to detect diseases at early stages.
- To explore the capability of few shot learning to detect rare disease X-ray images classification task. .
- To develop a combined hybrid pipeline capable of minimizing the limitation of DNNs majority class biasness.

Chapter 2

Related Work

2.1 Literature Review

In the paper titled as **"CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning"** [8] by Pranav Rajpurkar and his associates, they develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists by using an algorithm developed by them called CheXNet. It is a 121-layer convolutional neural network trained on ChestX-ray14 dataset. The standard way to diagnose pneumonia is through chest X-rays which requires very skilled radiologists who may not be available in every scenario and even if they are available large scale diagnosis would take a lot of time and effort because diagnosing pneumonia disease is quite difficult due to similar visual signs with other diseases in X-ray images. Therefore, CheXNet model was created with an aim to detect pneumonia at state-of-the-art level compared to regular radiologists. NIH ChestX-ray14 dataset was used by the authors which includes 112,120 frontal chest X-ray images of 14 disease classes. They used the pneumonia classes as positive and the rest as negatives as the dataset was split such as for training there were 98,637 images and for validation there were 6,351 images and for test there were 420 images which were manually labeled by four radiologists. The model takes chest X-ray images as inputs and then provides an output the probability of pneumonia detection along with a heatmap image of the area of the chest which contributed to the diagnosis for further analysis. CheXNet was first trained to perform a a binary classification task as to find out whether the patient has pneumonia or not. Weighted binary cross-entropy loss was used to deal with the class imbalance that appeared in the dataset. Downscaling the images to 224x224 pixels and normalizing them along with data augmentation helped in improving the model's generalization. The model was then further enhanced using a pretrained DenseNet on ImageNet and optimized with the Adam optimizer. The dataset was split into three parts: training, validation, and a test set which consisted of 420 images all labeled manually by the four radiologists. After the tests were done, the CheXNet model achieved an F1 score of 0.435 which outperformed the average score of radiologists 0.387. The result signified machine learning models capabilities of exceeding human expert performance in a controlled environment. Furthermore, the model was also evaluated on the other 13 diseases present in the CXR-14 dataset along with pneumonia and the model went on to beat the state-of-the-art models across every other disease class. Class Activation Maps (CAMs) was also used here to improve the interpret-ability of the predictions which included highlighting regions in the X-rays that influenced its decision making. The success of the model here demonstrates how deep learning can deliver expert-level diagnosis while being much faster and efficient especially since the availability radiologists can be scarce in many regions. The project offers a promising path forward in applying AI to real-world healthcare challenges

Additionally, the extended study of CheXNet by Xin Wu et al. aims to conduct multi-label chest X-ray (CXR) image classification and mitigate the challenges that come with it. In their paper [31], **"CheXNet-Combing Transformer and CNN for Thorax Disease Diagnosis from Chest X-ray Images"** The traditional CNN models have always struggled to capture the statistical dependencies between labels and there is the lack of direct interaction and information exchange between CNN and Transformer models. Furthermore, due to overlapping symptoms, which can appear in different sizes and locations, and can co-occur it is very difficult to diagnose Chest X-ray based diseases. These factors pose significant challenges for traditional deep learning models in accurately classifying these conditions. Traditional models that authors used in the past included Convolutional Neural Networks (CNNs), which performed well to extract detailed local features but under performed in realizing relationships between multiple disease labels and long-range dependencies. On the contrary, Vision Transformers (ViTs) excel at learning global image dependencies and comprehending label correlations. However, they often struggle to capture fine-grained visual cues, particularly when dealing with imbalanced data or small abnormalities. To address these issues, they proposed a hybrid deep learning network named CheXNet which consists of both CNN and Transformer networks in order to exploit their complementary strengths. The model consists of three main parts in the CNN and Transformer branches: Label Embedding and Multi-Scale Pooling module (MEMSP) for better feature extraction, Inner Branch module(IB), and Information Interaction module (IIM) to allow CNN-Transformer interaction. On the CXR11 dataset, the improved CheXNet demonstrated remarkable performance, achieving an impressive average AUC (Area Under Curve) of 82.56%. This outperformed both ResNet50 (80.96%) and standard Vision Transformers (75.7%). Notably, the model exhibited significant improvements in challenging cases, such as ETT-Abnormal, where its AUC was 9.52% higher than ResNet50. On the CXR14 dataset, the model achieved an average AUC of 76.80%, which was slightly higher than strong baselines like ConvNeXt and substantially outperformed Transformer-only models. Furthermore, removing any of the module such as- label embedding, multi-scale pooling, and feature interaction results in a performance drop signifying the importance of each module in the improvement of the results. These findings over previous models such as ResNet50 and Vision Transformers not only helped us learn about the model's ability to learn effectively even in the presence of class imbalance and overlapping features but also the effectiveness of hybrid CNN-Transformer approach especially in handling real-world diagnostic scenarios.

The paper **"Prototypical Networks for Few-shot Learning"** [9] proposes a model called prototypical networks for the problem of few-shot classification, where a classifier must generalize to new classes not seen in the training set, given only a small number of examples of each new class. Regularly used deep learning models require a large amount of labeled data in order to perform adequately, which is not always feasible in medical imaging scenario where such data might not be available. When trained on limited data the problem arises because traditional models tend to over-fit a lot. By representing each class with a prototype in an embedding space and classifying new instances based on their proximity to these prototypes the study proposes a simpler approach than previous meta-learning models. The model uses a neural network (usually a convolutional network for images) to implement the idea. The neural network is used to embed input samples into a lower-dimensional space where classes form compact clusters. For every training episode, a number of classes are sampled and a few examples from every class are sampled to form a miniature task. The model computes a prototype for every class by averaging the embedded vectors of its support examples, and query examples are then predicted based on the similarity of their embedding with each prototype. Episodic training mimics the few-shot setting at test time to enhance the generalization of the model. One of the key advantages of this approach is that it's test-time non-parametric—no new model update is needed when encountering a new class; the prototype can simply be computed from a few labeled samples of the class. The simple workflow of the model allows for the faster and lower computation as well as fast adaptation making the Prototypical Networks model a great choice for application where speed and simple workflow is needed. The authors tested their method on a series of benchmark datasets for few-shot learning, including Omniglot and miniImageNet. On the Omniglot dataset, which comprises handwritten characters from a variety of different alphabets, the model was near perfect for both the 1-shot and 5-shot learning scenarios, with improved performance compared to previous methods like Matching Networks. On the harder miniImageNet dataset, more indicative of real-world visual tasks, Prototypical Networks also had a good showing and beat out several meta-learning baselines. One finding from the experiments was that Euclidean distance worked better than cosine similarity in the embedding space, which supports the authors' theoretical motivation that an optimal classifier in a Euclidean space is linear under the class-conditional assumption they make. Lastly, we can see that the Prototypical Networks are faster more simpler way of solving the few shot learning problem. The model can do complex tasks without doing complex computation by representing each class with the average of its support embeddings and by contrasting the query points using a distance metric. The strong results compared to the traditional methods and practical advantages make it a viable approach in the field of medical imaging and has influenced many other followup works as a benchmark.

The paper [15] **"Survey on deep learning with class imbalance"** addresses one of the most common and challenging problems in machine learning: class imbalance, which occurs when some classes in a dataset have significantly more samples than others. It is especially problematic in applications like medical diagnosis, fraud detection, and anomaly detection, where the unusual cases (minority classes) are the most important to detect. Traditional deep learning models have a bias towards the majority classes and thus fail to perform well on underrepresented classes. The authors would like to give a comprehensive overview of how current deep learning techniques handle imbalanced data and offer directions for future research. They cover the significant problems class imbalance creates, such as biased loss optimization, inferior minority class generalization, and difficulty in performance measurement. The current methodologies are grouped into two categories by the survey: data-level techniques, which involve altering the training data, and algorithm-level strategies, which focus on modifying the model's learning process from the data. Techniques at the data level focus on re-balancing the class distribution of the data. The most widely used techniques in this aspect are oversampling and under-sampling. Oversampling involves synthesizing instances of minority classes (e.g., with SMOTE or GANs), whereas undersampling removes examples from majority classes to balance the data set. Although useful, oversampling may lead to over-fitting, whereas undersampling destroys useful information. These encompass more advanced techniques like adaptive sampling, data augmentation, and distribution alignment. Solutions at the algorithm level, on the other hand, change the training process but not the data. These involve cost-sensitive learning, in which the model is penalized harder for the misclassification of the minority class instances, as well as modified loss functions, for example, focal loss, that assign more importance to hard-to-classify or rare examples. Techniques such as meta-learning, transfer learning, and ensemble models which can be enhanced to handle the problem of class imbalance. Architectural innovations are also described by the authors, such as using attention mechanisms and multi-task learning paradigms, to force the model to focus on minority classes. In addition, they describe the evaluation metrics—citing how accuracy will be deceptive on imbalanced datasets. Instead, precision, recall, F1-score, AUC-ROC, and precision-recall curves provide a realistic estimate of performance, especially for minority classes. The survey evaluates how different deep learning applications, such as natural language processing, computer vision, and medical image analysis, apply these techniques in practical scenarios. For example, in medical imaging, where diagnosis of rare diseases is critical, practitioners are inclined to combine synthetic data generation, expert loss functions, and transfer learning to improve model sensitivity. Lastly, the paper determines that one-size-fits-all doesn't fit when class imbalance in deep learning is concerned. Instead, the best approach is to utilize a blend of methods—such as balancing data distribution, adjusting the loss function, and applying resilient metrics—to find a combination that performs well. The authors mention that future research must tackle more adaptable solutions that are capable of responding dynamically to imbalance at training time, reduce the requirement of significant tuning, and perform well even with sparse labeled sets. This survey is a valuable guide for researchers and practitioners to build reliable and fair deep learning models with class imbalance, especially in fields where outstanding but critical examples cannot be ignored.

The study by Paul et al. 2020 titled as, **"Fast Few-Shot Transfer Learning for Disease Identification from Chest X-Ray Images Using Autoencoder Ensemble"** [18] proposed a fast few-shot learning framework that uses transfer learning and an ensemble of auto encoders to identify different lung and chest diseases and conditions from chest x-rays. In this work, they divided different chest diseases into two disjoint categories: (i) base classes (with large training set) and (ii) novel classes (with a few training examples per class). First, they train a DenseNet-121 feature extractor solely on base classes from the NIH ChestX-ray14 dataset, ensuring that the model learns general chest X-ray features without exposure to novel classes. They introduce an autoencoder ensemble trained with just five examples per novel disease class for classification. The model here predicts the disease label using a majority vote and after the experiment was completed it was found that the model outperforms traditional methods by about 0.18 in F1-score for the novel disease classes along with very fast learning times showcasing the potential of the few shot learning framework for rapid and accurate diagnosis of rare diseases.

The authors in the paper titled as **"Assessing and Mitigating the Effects of Class Imbalance in Machine Learning with Application to X-ray Imaging"** [19] presents a thorough investigation of the effects of class imbalance and methods for mitigating class imbalance in ML algorithms applied to medical imaging. The authors here focused on class imbalance in training datasets, where pathological conditions are often underrepresented compared to normal cases. We first selected five classes from the Image Retrieval in Medical Applications (IRMA) dataset, performed multi-class classification using the random forest model (RFM), and then performed binary classification using convolutional neural network (CNN) on a chest X-ray dataset. They experimented with techniques such as oversampling, undersampling, and adjusting class weights to address class imbalance. The results showed that a close-to-balanced training set resulted in the best model performance, and a large imbalance with overrepresentation was more detrimental to model performance than underrepresentation. Oversampling and undersampling methods were also both effective in mitigating class imbalance, and efficacy of oversampling techniques was class specific. The study demonstrates that resampling methods, particularly combining oversampling with Gaussian noise, can significantly enhance model performance in imbalanced datasets. These findings can help further improve machine learning models, mitigate the problem of class imbalance, and generate more accurate disease classification results.

With a focus on the CXR14 dataset **Hellín et al.2024** [28] study in Applied Sciences investigates class imbalance in medical image classification using chest X-rays. It examines how deep learning, particularly CNNs like ResNet and DenseNet, can aid in diagnosis but struggle with unbalanced datasets. In order to address imbalance, earlier studies used methods like augmentation, oversampling, and cost-sensitive learning. The authors do note, though, that there are still unanswered questions regarding how these methods impact metrics like recall and precision in particular contexts. Although it is not balanced (the "No Finding" class is the most prevalent), the dataset CXR14 contains over 100,000 chest X-rays labelled for 14 thoracic disorders.To make the data more even, they employed CNNs (ResNet, DenseNet) with transfer learning, which means that they were trained on ImageNet and then refined on CXR14. They also employed augmentation and class separation.The minority class performance suffered due to the imbalance, which benefited the dominant class (low recall for pneumothorax).Improved recall and precision for the minority class (0.50–0.75 to 0.65–0.85).While balanced ratios improved metrics, an excess of them may exacerbate diversity.AUC-ROC (best models) ranges from 0.85 to 0.90, precision (minority classes) ranges from 0.65 to 0.85, and validation accuracy ranges from 0.70 to 0.85.The F1 score increased from 0.55 to 0.70–0.80 with balancing.Ten percent is used for testing, ten percent is used for validation, and eighty percent is used for training. The batch size ranges from 16 to 64, and the number of epochs is between 20 and 50. DenseNet achieved a 10–15% improvement in precision/recall and an AUC-ROC score of 0.85–0.90 with balancing.

Another paper titled as **"Survey on deep learning with Class Imbalance"**[14] looks at deep learning methods for dealing with class imbalance, a crucial problem in applications such as fraud detection and medical diagnosis. With an emphasis on convolutional neural networks (CNNs) for tasks like image recognition, it examines how conventional machine learning techniques (such as data sampling and cost-sensitive learning) have been adapted to deep neural networks (DNNs). The authors Johnson and Khoshgoftaar, 2019 point out that the majority of research focuses on computer vision and draw attention to the paucity of studies on deep learning with unbalanced data, especially big data. In their discussion of data-level (e.g., SMOTE, RUS), algorithm-level (e.g., cost-sensitive learning), and hybrid approaches, they highlight the necessity of robust performance metrics such as F1-score, G-Mean, and AUC-ROC because of the limitations of accuracy in unbalanced settings.The use of the CXR14 dataset in surveyed studies for the classification of thoracic diseases is mentioned in the paper, but no experiments are carried out with it directly.The purpose of CXR14 was to assess deep learning models under class imbalance using more than 112,000 chest X-ray images that were labelled for 14 thoracic diseases.In order to reduce imbalance in CXR14's multi-label classification tasks, surveyed studies used CNNs (e.g., ResNet, DenseNet) with transfer learning, data augmentation (rotations, flips), and sampling techniques (e.g., RUS, SMOTE).Minority classes like pneumothorax had fewer samples, which presented difficulties due to CXR14's imbalance ("No Finding" class).CNNs with augmentation and sampling enhanced minority class performance, according to the reviewed studies; however, CXR14's noisy labels and imbalance frequently skewed models in favour of majority classes.Although precise metrics differed between studies, one cited study (whose name is not specified) that used CXR14 improved recall for minority classes (such as pneumothorax) with SMOTE. Although they increased minority class detection, cost-sensitive approaches ran the risk of overfitting. Although not specifically mentioned in the survey for CXR14, the referenced studies generally had higher performance on majority classes and 70–85% validation accuracy.For minority classes, precision was between 0.60 and 0.80, and it increased by about 10% to 15% with methods like SMOTE or cost-sensitive learning. Training Specifics: Training included 20–50 epochs, Adam/SGD optimisers, batch sizes of 16–64, 70–80% train, 10% validation, and 10% test splits. For efficiency, GPUs were employed. Results were published in the Journal of Big Data (Vol. 6, Article 27) on March 19, 2019. According to the survey, no single approach consistently outperformed others across CXR14 and other datasets; some studies reported balanced models with AUC-ROCs between 0.80 and 0.90.

The problem of few-shot learning in medical image analysis, particularly for chest X-ray diagnosis, was discussed in the paper, **"Medical Image Analysis. Discriminative ensemble learning for few-shot chest x-ray diagnosis"** by Paul et al.2021, [22] which was published in Medical Image Analysis. The authors draw attention to the drawbacks of conventional deep learning techniques, which depend on sizable, annotated datasets which are frequently unavailable for rare diseases because of their low prevalence. To address this problem, they suggest a brand-new discriminative ensemble learning strategy that uses FSL to categorise thoracic diseases using a small number of training examples. Focussing on the particular difficulty of managing long-tailed, unbalanced datasets in medical imaging, the work builds on previous developments in deep convolutional neural networks (CNNs) for medical imaging tasks like as skin cancer classification, diabetic retinopathy and wrist fracture detection. In contrast to FSL, which seeks to generalise to unseen classes with few examples, earlier methods to address class imbalance such as data augmentation, modified sampling weights, or loss function adjustments require retraining for new classes. Citing studies like triplet network-based modality recognition and GAN-based brain MRI segmentation, the authors also discuss the lack of FSL applications in medical imaging and review FSL approaches in computer vision, such as meta-learning and metric-learning.The authors developed a two-step framework for diagnosing FSL chest X-rays: A CNN-based module (DenseNet-121 architecture, pre-trained on ImageNet) extracts general features from chest X-rays during meta-training.It uses a meta-train set with base classes to learn common characteristics of thoracic illnesses.During meta-testing, a novel discriminative autoencoder ensemble extracts salient features unique to a given disease from coarse-learner output. Bootstrap samples and feature subspaces chosen for diversity and class discrimination are used to train each autoencoder. Weighted voting is used for classification, and autoencoders are assigned intrinsic weights based on the quality of the training.The method handles high-dimensional data, noisy features, and chest X-ray inter-class similarity, and can handle novel classes with as few as five training instances.The system is modular and adaptable since only the coarse-learner is learnt during meta-training and the saliency-based classifier during meta-testing.The suggested approach outperformed baseline methods by 19% on the NIH ChestX-ray14 Open datasets. Fifteen novel classes (fibrosis, hernia, pneumonia, mass, nodule, pleural thickening, cardiomegaly, oedema, emphysema, consolidation, effusion, pneumothorax, atelectasis, infiltration, and no finding) were tested using five training examples per class. A single validation accuracy or precision measure for all classes is not specifically reported in the paper. Rather, it offers F1 scores, recall, and precision that are specific to a class.

The paper titled as, **"Few-shot learning for COVID-19 Chest X-Ray Classification with Imbalanced Data: An Inter vs. Intra Domain Study"** [27] review few-shot learning (FSL) for COVID-19 classification in chest X-ray images under severe data imbalance, focusing on intra- and inter-domain scenarios. It proposes a Siamese neural network-based methodology that combines techniques like data augmentation, transfer learning, weighted loss, and balanced sampling to address data imbalance and scarcity. The study evaluates different approaches and compares their outcomes to a common CNN baseline using four publicly available chest X-ray datasets. Few-shot learning is essential for medical imaging, particularly chest X-ray analysis, because there is a dearth of labelled medical data and expert labelling is expensive. The need for models that can generalise from small samples is highlighted by the rapid spread of diseases like COVID-19, especially in datasets that are unbalanced and have a notably low number of positive (pathological) cases. Increasing diagnostic speed, enabling computer-aided diagnosis in resource-constrained settings, and supporting clinical decision-making in life-threatening situations where timely identification can save lives all depend on this field. A small number of annotated images raise the risk of overfitting for data security. Minority class accuracy for data imbalance is decreased by bias towards the majority class. Poor generalisation across datasets was the domain variable.Diagnostic features may be obscured by standard transformations.Robustness in both intra- and inter-domain contexts was attained.Previous work includes FSL techniques such as Siamese networks, MAML, and prototypical networks, as well as deep learning for medical image tasks. By using FSL with Siamese networks and testing initialisation (random, ImageNet, transfer learning), augmentation (shifts, scaling, rotations), imbalance solutions, and classifiers on three combined datasets with different imbalance levels, this study builds on previous research.Performance in high/medium imbalance is enhanced by transfer learning and ImageNet pre-training.Convergence is a problem for random initialisation; bias may be introduced by transfer learning.In high imbalance, transfer learning fails, but it helps intra-domain performance.Broadens the variety of datasets and introduces non-realistic features that impair medical imaging performance.Thia requires domain-specific augmentation under the direction of experts.Enhances cases of low or no imbalance but runs the risk of overfitting in cases of high imbalance.It works well for severe imbalance, but too much repetition could cause overfit.Ideal for improving feature learning and low/no imbalance.Intra-domain low/no-imbalance performance is enhanced by negative pair emphasis, but high imbalance is not.The degree of imbalance determines the technique to use; for balanced data, combining techniques yields the best results.In high imbalance, Random Forest performs poorly.Classifier selection differs by imbalance and domain.Because Siamese networks share weights, they perform better in high/medium imbalance.CNNs perform marginally better in balanced situations.Siamese networks work well in settings that are few-shot and unbalanced.CNNs work well when there is a lot of data. This presents weighted contrastive loss for Siamese networks and tests pairing classifiers in FSL for medical imaging. Adapting techniques to other FSL architectures such as prototypical networks and using expert-guided augmentation could make performance and generalisation better. It's worth looking into self-supervised learning for initialisation.

In their paper, **Deep Learning for Understanding Multilabel Imbalanced Chest X-ray Datasets** [25] the researchers use deep learning techniques to mitigate class imbalance issues in multi label chest X-ray datasets. In the field of medical imaging, datasets are highly imbalanced which makes it harder for models to effectively learn underrepresented diseases. To solve this problem, the researchers use a combination of techniques like data preprocessing segmentation, cropping, and normalization, data augmentation on the multi labeled PadChest dataset which contains over 160,000 images and 174 labels with some labels like "normal" vastly outnumbering others. Several state of the art models like DenseNet-201, EfficientNetB0, InceptionV3, InceptionRestV2, Xception were used to train on the dataset. The researchers used ensemble learning where multiple models were combined for transfer learning and weighted cross-entropy loss in an effort to mitigate the effects of class imbalance. The key finding of their research was that an ensemble of models worked better than individual models for training to detect rare diseases. They also used a technique for explainable AI called Grad-CAM heatmaps to visualize and understand how the models made their prediction. The results show their methodology of assembling models with preprocessing achieves strong performance with AUC reaching up to 0.83 and F1-scores improving across the board. In conclusion, this paper clearly demonstrates that the performance of deep learning models on rare disease detection on class imbalanced multi labeled datsets by applying effective methods can be significantly improved, laying the groundwork for future research in this area.

The paper, **Impact of Class Imbalance on Chest X-ray Classifiers: Towards Better Evaluation Practices for Discrimination and Calibration Performance** [26] explores the impact of class imbalance in the performance evaluation of chest X-ray classifiers. Due to rare diseases being underrepresented in most medical datasets, standard evaluation metrics like AUC-ROC can provide misleading conclusions. To tackle this challenge of improper evaluation, the researchers evaluated the performance of several classifiers on two widely used chest X-ray datasets, ChestX-ray14 and ChestXpert under varying levels of class imbalance. The metrics used for evaluation included AOC-ROC, AUC-PR and a Balanced Brier Score which helps account for both discrimination and calibration.The study found that AUC-ROC alone is misleading in highly imbalanced datasets, whereas AUC-PR is found to be a more reliable metric for evaluating performance as it directly evaluates performance on minority class. The study also emphasized the importance of model calibration by showing that imbalanced datasets can lead to misleading conclusions using standard brier scores. so they used a balanced brier score which takes into account the performance of the model on both the minority and majority classes separately as a more reliable metric to assess both discrimination and calibrating performance. In conclusion the study recommends reporting AUC-PR alongside AUC-ROC and adopting the Balanced Brier score to ensure that performance is assessed fairly across both the majority and minority classes. The paper showcases further research into the impact of different class imbalance strategies, such as weighted loss functions and enrichment sampling, as well as a deeper dive into calibration techniques in chest X-ray classifiers that have much promise in improving performance evaluation.

In their paper, [29] the researchers try to address class imbalance on tuberculosis(TB) classification by introducing a novel approach where a few-shot learning (FSL) is combined with deep learning backbones that utilizes a prototypical network algorithm. Chest x-rays are a key diagnostic tool for TB detection in patients but class imbalance between TB positive and TB negative images in TB datasets makes it difficult for models to accurately classify the minority class that is TB-positive. When traditional deep learning methods perform poorly in classifying rare cases with limited labeled data, the few-shot learning approach offers a promising solution. The dataset used in this study, TBX11K is highly imbalanced with TB-negative images dominating. This imbalance leads to biased model performance in the cases of supervised learning. With few-shot learning, the authors try to address training situations where labeled data is scarce and generalize to new and unseen classes with only a few examples. To assess the effectiveness of FSL, the researchers examined the impact of different shot numbers (1-shot, 5-shot, 10-shot, and 20-shot) and backbone architecture on classification accuracy. Through meta-learning, The ResNet-18 backbone achieved 96.80% accuracy with 1-shot and 98.93% with 20-shots which is a significant improvement over supervised learning. Similar gains were also observed with the ResNet50 backbone indicating meta-learning greatly enhances the model's ability to generalize from a few examples. On the other hand, the prototypical network was found to be particularly effective at mitigating the issues caused by class imbalance to better classify the minority TB-positive class in the dataset. However, VGG16 did not benefit from meta-learning where the accuracy remained at 33.33% across all shots. Since VGG16 is an older model compared to the other ResNet models, suggesting that certain architectures may be well-suited for meta-learning. Therefore, selecting appropriate backbone models that have the ability to learn complex features and generalize from fewer examples, for few-shot learning tasks is underscored. According to the study, data augmentation, even though popular in medical image processing to handle class imbalance, its effectiveness can be limited due to overfitting and generation of unrealistic variations. When integrated with meta-learning, few-shot learning can provide a promising alternative for improving model performance with limited data. The study demonstrates how few-shot learning algorithms can be used to enhance model adaptability in medical imaging tasks, especially in the case of TB diagnosis from a highly imbalanced dataset. Therefore, further research focused on fine-tuning existing model architectures as well as exploring other few-shot learning algorithms could greatly advance disease diagnosis in imbalanced medical image datasets.

Tuberculosis, often called TB, remains one of the world's most lethal contagious diseases, taking lives every year. Quick, precise testing gives doctors a real shot at better outcomes and stops the disease from spreading or worsening. In their paper, titled as **"Few-Shot Learning Approach on Tuberculosis Classification Based on Chest X-Ray"** Pramana A. et al [30] stated that in under-resourced clinics, pairing AI with routine chest X-rays can lift diagnostic certainty when human eyes alone may falter. Yet because positive cases are scarce, standard classifiers struggle; the so-called class imbalance leaves them blind to TB on many datasets. That shortcoming pushes us to explore few-shot learning, especially Prototypical Networks, which are built to learn meaning from just a handful of samples.This is important for reliable TB diagnosis because there aren't many positive TB samples and it's hard to classify minority classes.TB CXR datasets, such as TBX11K, have too many negative (healthy or non-TB sick) cases compared to positive TB cases (17% positive in TBX11K). This makes models that don't work well on minority classes.Medical imaging often doesn't have enough labeled data, especially for uncommon conditions or certain types of TB (like active-latent TB), which makes traditional deep learning methods less useful.Techniques like zooming, flipping, and rotating can add variations that aren't realistic, which can lead to overfitting and unreliable model evaluations.Models need to be able to learn new or less common TB classes with only a few examples, which is hard for traditional supervised learning. (Yogi et al.2024) used techniques like as flipping and rotating to find TB with 96.6% accuracy and a 0.99 AUC. It makes the dataset more varied, stops overfitting, and is easy to use.This can make image variations that aren't real, which can lead to overly positive evaluations and possible model bias.While augmentation makes things more robust, it also risks making them less clinically relevant, so they need to be carefully validated. SMOTE stands for "Synthetic Minority Oversampling Technique". (Yogi et al.,2024) used SMOTE on TBX11K and got better results with XGBoost, going from 94.11% to 94.33% accuracy. Using SMOTE, got 99.74% accuracy.It makes synthetic samples of the minority class to fix class imbalance, which makes the model work better on classes that aren't well represented.It might add noise, data overlap, or small gaps, which could make the model worse. SMOTE works, but it needs to be carefully tuned to avoid adding artifacts that make the model less reliable.The literature review says that Random Forest and XGBoost are two methods that have been used to deal with data imbalance in TB detection.Strong for structured data, fast to compute, and useful in many situations.Not as good for high-dimensional image data and has trouble with severe class imbalance without preprocessing.These methods can't handle complex image features very well, which shows that we need more advanced deep learning methods. Data augmentation, SMOTE, and traditional methods are all easy ways to fix class imbalance and make TB detection more accurate. They are basic steps in analyzing medical images that show how to deal with datasets that aren't balanced.These methods often don't solve the real problem of having too little labeled data for minority classes. They also run the risk of adding artifacts like fake images or noisy synthetic samples and have trouble applying to new or rare TB cases.To get around the problems with each method, you need to use both preprocessing techniques and advanced learning paradigms. Traditional deep learning relies on large datasets, which shows how important new methods like FSL are, since they can learn well from only a few examples.The paper suggests a new way to use Prototypical Networks to fix the class imbalance in TB CXR datasets, specifically TBX11K. The idea is that Prototypical Networks, along with pre-trained CNN backbones (VGG16, ResNet-18, ResNet-50), can generalize well to TB classes that aren't well represented like active-latent TB with only a few samples. This would be better than traditional methods like data augmentation and SMOTE.This method uses metric-based learning, which means that new classes can be grouped together with only a few examples. This makes it less likely that the model would overfit.Earlier research that used oversampling or augmentation reached a different outcome than this one. By looking at VGG16, ResNet-18, and ResNet-50 in the context of FSL, we can learn more about the best way to get features from medical datasets that aren't balanced. Adding Meta-Learning by: Adding meta-training makes models more flexible, especially deep architectures like ResNet-50, which got 98.60% accuracy after 20 shots.This idea is worth looking into more since it helps to fill in important gaps in tuberculosis diagnosis and deals with the problem of not having enough data or having it in the wrong places. It also gives a method that could be used for other medical imaging tasks that have similar problems.

The **Bhimireddy et al. 2022** [23] work uses the CheXpert dataset, which is quite similar to the CXR14 dataset, to show how important Few-Shot Learning (FSL) is for diagnosing chest X-ray disease. It's highly important that FSL can make reliable models with less data because medical picture annotation is expensive and regulated by laws like HIPAA and GDPR (Bhimireddy et al., 2022). FSL's support for human-in-the-loop systems, which let radiologists swiftly retrain models and fix wrong inferences, fits with your CL aim of being able to react to changes in data domains without losing everything (Bhimireddy et al., 2022). FSL can apply what it learns to new or unique situations with only a few examples, which is better than static models. This is very important in medical contexts that are always changing

(Bhimireddy et al., 2022).Traditional deep learning is hampered by the cost and privacy law restrictions associated with curating sizable, well-annotated datasets. Despite their high performance, deep learning models generate false positives (FP) and false negatives (FN), necessitating the use of techniques to improve predictions.Model training and evaluation are made more difficult by inter-annotator variations among radiologists (Bhimireddy et al.,2022). Scalability in Different Pathologies for small datasets, it can be difficult for models to generalize across multiple pathologies (CheXpert) (Bhimireddy et al.,2022). Relying less on big datasets, FSL has been used in histopathology, endoscopy, COVID-19 CT diagnosis, and fundus image analysis (Bhimireddy et al.,2022). Large datasets are necessary for methods like CNNs, LSTMs, and f-AnoGAN, but they suffer from false inferences. The authors suggest Triplet Few-Shot Learning (TFSL), which uses 150 image triplets (FP/FN, TP, and TN) from CheXpert to refine a pre-trained DenseNet121 model. MarginRankingLoss is used to retrain the model over five epochs, increasing both the Positive Predictive Value (PPV) and the Negative Predictive Value (NPV). On an NVIDIA Quadro P6000 GPU, TFSL trains in 8–9 minutes, making it perfect for clinical settings that demand fast updates. TFSL not only lowers the amount of wrong conclusions, but it also raises the PPV and NPV for 14 diseases.One way that triplets based on feedback from radiologists make CL more flexible is by letting you quickly construct models. Using TFSL's modified DenseNet121 and Margin Ranking Loss is easier than using complex ensembles.The incremental FSL does better than the TFSL when it comes to NPV for pleural diffusion, which means that there is room for improvement.TFSL can't be used on datasets that don't have baseline models since it needs a robust baseline model to work. It is possible that 150 triplets will not be enough to get a clear picture of the noisy data from CXR14. The lower NPV improvement for pneumonia implies that there are problems that are unique to the disease , the way some diseases become worse is not always the same.Triplets with Margin Ranking Loss are a good way to cut down on the number of false positives. A greater NPV means that genuine positives are being handled better, but the fact that the training duration stays the same makes the calculation less relevant. The NPV is modest but the PPV is high which shows how important it is to make sure that fixes for FNs are as good as they can be. FSL works far better than regular fine-tuning when used on small datasets. However, the strength of the baseline model and the quality of the triplet are also very significant things to think about. To reach its goal of handling transitions in sequential data, CL is made more adaptable by using human-in-the-loop methods.The Margin Ranking Loss approach is better than the Triplet Margin Loss method when it comes to triplets, especially when it comes to using empirical loss functions.TFSL gives you a scalable way to use a human-in-the-loop technique.We can do this by fine-tuning a model for 14 diseases in CheXpert using image triplets.This method is different from previous FSL methods that worked well with fewer classes or larger datasets since it is simple and can be retrained quickly.

Significant study demonstrated by **Cores et al. 2022** [24] on few-shot learning (FSL) for COVID-19 screening using chest Xray (CXR) images. The ability of FSL to train with small datasets is crucial because new diseases like COVID-19 frequently lack large annotated datasets because of privacy regulations (like HIPAA/GDPR) and annotation costs (Cores et al.,2022). For managing pandemics in environments with limited resources, automated CXR-based triage systems can expedite diagnosis and lessen the workload of radiologists (Cores et al.,2022). CXR's benefits, like its portable devices and speedy scanning, make it perfect for quick deployment in a variety of clinical settings, enhancing your CXR14-based research (Cores et al., 2022). Traditional deep learning is limited by small datasets, necessitating the use of techniques. According to (Cores et al.,2022), COVID-19 increases false positives (FP) and false negatives (FN) by sharing visual characteristics with other lung diseases, such as ground-glass opacity. It can be difficult to distinguish between the four severity levels—normal, mild, moderate, and severe—particularly in cases that are asymptomatic and show no radiological symptoms (Cores et al., 2022). Model generalizability is impacted by heterogeneous CXR datasets, which differ in terms of acquisition methods, patient demographics, and severity (Cores et al.,2022). For COVID-19 detection, deep learning models such as VGG, Xception, CapsNets, and COVID-Net have been employed; however, they are ineffective with small, heterogeneous datasets and necessitate large datasets (Cores et al.,2022). The authors present a three-part FSL system: (1) a lung-aware region proposal network (RPN) that uses U-Net to target lung regions; (2) a cost function that is sensitive to misdiagnosis and incorporates expert knowledge to penalize errors according to their severity; and (3) an ensemble of multiple support sets (5-shot and 10-shot) to improve robustness. The technique is validated on the COVID-SC (1092 images, 653 COVID-positive) and COVIDGR-1.0 datasets, obtaining 87.40% and 79.10% accuracy, respectively, and compares prototypes using Earth Mover's Distance (EMD). outperforms ResNet-50 (71.62%) with a sensitivity of 91.42% on COVID-SC, which is essential for identifying COVID-positive cases during triage (Cores et al., 2022). By focusing on lung regions and eliminating noise from irrelevant areas, the RPN enhances feature extraction. In line with clinical requirements, the misdiagnosis-sensitive loss places a high priority on severe case detection. Stability (79.10% accuracy on COVIDGR-1.0) is improved by using ensemble support sets (5-shot and 10-shot), which lessen dependency on single support images. Due to their lack of radiological signs, asymptomatic COVID-positive cases (P-NORMAL) have an accuracy of only 38.26%, which limits their usefulness for triage (Cores et al., 2022). For non-COVID diseases with COVID-like characteristics, the accuracy was lower (25.41%), suggesting that it was difficult to differentiate between similar pathologies. The RPN and ensemble make the computation more expensive, which can be a problem in places with limited resources. The distribution of the data seems to affect performance, since different datasets do better or worse (87.40% on COVID-SC vs. 79.10% on COVIDGR-1.0). Focusing on lung areas makes features more relevant and makes it easier to find COVID-positive cases. Penalties directed by experts are better at finding serious cases, but they have trouble with cases that don't display any symptoms. Multiple support sets make things more resilient, but they don't fully take into account how different datasets are. FSL does well with tiny datasets, but the quality and variety of the datasets can affect how well it works. For medical imaging, lung-specific feature extraction and expert knowledge integration are very important, but they need to be correctly calibrated so that they don't match too closely to certain datasets. Cross-dataset variability shows how important it is to have models that can be used with different datasets. This helps CL reach its aim of being able to react to changes in data. Using a lung-aware RPN, a cost function that is sensitive to misdiagnosis, and a support set ensemble in an FSL framework is a new way to screen for COVID-19 CXR with only a few datasets.The COVID-SC dataset makes a big difference because it has a full severity and illness classification.

2.2 Background Study

2.2.1 Common Thorax Diseases

COPD: Chronic Obstructive Pulmonary Disease is generally caused by cigarette smoking, air pollution, workplace dust, genetic predisposition, Alpha 1 anti trypsin deficiency. It makes it hard to breathe, lowers our ability to exercise. It causes frequent flare ups. It can cause heart problems and risk of death. The economic burden is also on top because it raises healthcare expenditures, hospital staying and lost productivity. Some common symptoms like shortness of breath and prolonged cough people think that are just part of aging. The most effective diagnosis is not always available in many places, especially in developing countries. This means that

Figure 2.1: CXR image of Chronic Obstructive Pulmonary Disease

the disease is found so late when it has already gotten worse. Accordingly early diagnosis is essential since cause management via pharmacotherapy, pulmonary rehabilitation and lifestyle alterations can decelerate progression, enhance quality of life and reduce the effects. Deep learning has emerged to be highly promising in solving these diagnostic complexities over the past few years. Deep learning algorithms can find early indicators of emphysema and airway blockage that people often miss by looking at medical images like chest X-rays and CT scans. AI algorithms can also use spirometry data, electronic health records and even data from wearable technologies to identify how a condition will get worse in the future. Deep learning is valuable for doctors because of its accurate diagnosis. It helps them make decisions and lets them create individualized treatment plans for COPD patients.

Covid-19: The Severe Acute Respiratory Syndrome Coronavirus-2 causes Coronavirus Disease 2019 (COVID-19) which was invented in Wuhan, China at the end of 2019. From minor symptoms like fever, cough to severe pneumonia, acute respiratory distress syndrome, multi-organ failure and its effects can go to death. Additionally, its negative physical effects COVID-19 has had significant economic results across the globe. It also includes overburdened lockdowns and chronic difficulties. The fact that many people who are infected either show no symptoms at all or face mild symptoms which raises the possibility of silent transmission makes identification more difficult. RT-PCR are time-consuming conventional diagnostic techniques. For these techniques they need laboratory facilities. These aren't always available in low-resource environments. Early detection is important because timely treatment and isolation protects others against this virus. Deep learning has become a potent

Figure 2.2: CXR image of Covid-19

technique to help with identification and management in this area. Automatically and correctly find lung abnormalities connected to COVID-19 in medical images like X-rays and CT scans by applying AI type models. Diseases can be detected faster by deep learning techniques by analyzing electronic health records. By providing rapid, flexible and economical diagnostic support these approaches can help overburdened healthcare systems, facilitate early management and ultimately decrease the worldwide burden of COVID-19.
Emphysema: Emphysema is a chronic lung illness that is part of COPD. When the alveoli in the lungs are permanently destroyed there will be a progressive deterioration in breathlessness, reducing the surface area accessible for gas exchange.This condition leads to major effects, such as difficulty breathing that doesn't go away, a chronic cough, disability to exercise and more respiratory infections. In the worst cases, it can stop the body from breathing and put stress on the right heart. As a part of COPD, emphysema is a big health problem all over the world. COPD is the third most common cause of death in the world.

Figure 2.3: CXR image of Emphysema

It is strongly linked to smoking cigarettes being around air pollution for a long time and occupational contact with dust or chemicals. Even while emphysema is common, it is often not found in its early stages because symptoms like moderate shortness and cough are easy to miss or blame on getting older or smoking. To confirm the condition, you need spirometry and imaging studies like CT scans. These procedures aren't usually possible in places with few resources. Taking prescribed drugs, going to pulmonary rehabilitation, quitting smoking, and making other changes to your lifestyle can all help slow the development of the disease. It can reduce the risk of complications and improve physical and mental wellness.
Pneumonia: Pneumonia looks as a thick, white spot in the lung, and it often has air bronchograms. It can affect one lobe (lobar) or show up in patches (bronchopneumonia). Over time, the area may become less clear. Infections with bacteria, viruses, or fungi are to blame. It makes the oxygen to penetrate into the bloodstream difficult. Pneumonia is caused by bacterial infection in lungs. The very

common or known reasons for this disease are bacteria, fungi and viruses. Among the common problems some are sepsis, pleural effusion, lung abscess, severe breathing difficulty, and even death in some occurrences, specifically in infants, the older people, and people with weak immune systems. Because the initial complications of this condition, such as fever, coughing, fatigue, and soreness in the chest, are very likely to be the complications of other lung related infections, such as the flu or bronchitis, it can be problematic to find out. Finding out a problem during the

Figure 2.4: CXR image of Pneumonia

initial phase is very important because it permits the progressive treatment of bacterial health complications with antibiotics, virus related infections with antivirals, and supportive assistance, such as oxygen therapy or hospitalization, when needed. These treatments can essentially lower the possibility of mortality and challenges.A variety of significant paths to contribute to the global effort to reduce the number of cases of pneumonia, including vaccination, proper nutrition, and improved hygiene are available in the current world.

Tuberculosis: Tuberculosis (TB), a very widely known bacterial sickness that mostly impacts the lungs negatively but can often affect other parts of the body such as the kidneys, spine, or brain, is caused by Mycobacterium tuberculosis. The illness moves slowly and may not show any symptoms for years at a time. The time when people are sick due to tuberculosis, may spread to other people without intending to. Among the many worst cases that can happen if it's not treated in time, are frequent coughs with bloody sputum, pain in the chest, intense fatigue, damage of the organ, and even death. It can be hard to find out what's wrong because the early signs, including a persistent cough, a low fever, night sweats, and weight loss, are frequently not specific and might be confused with other respiratory diseases or general weakness. Tests that are run in the lab like culture, sputum microscopy

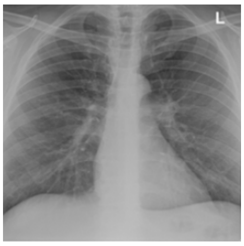


Figure 2.5: CXR image of Tuberculosis

and better molecular diagnostic methods are important to be sure, but many places with scarce resources still can't easily get these tests. The prompt initiation of antitubercular treatment after an early diagnosis is crucial for the patient's recovery and for reducing the likelihood of the disease's transmission to others. Drug-resistant TB is much harder and costs a lot more to treat, but it can be stopped from spreading quickly with treatment.

2.2.2 Rare Thorax Diseases

Edema: Edema is when fluid builds up in the body's tissues in an unusual way, which causes swelling and pain that can be seen. It can happen in particular areas such as the lungs, feet and ankles. It also can happen all over the body. The condition can happen for a few causes such as heart failure that increases venous pressure, kidney disease which can make it harder for the body to get rid of fluids, liver cirrhosis or malnutrition that lowers blood protein levels and lymphatic obstruction. It can stop fluids from leaving the body. Some drugs, steroids and antihypertensives also inflammation and allergic responses may also play a vital role. Doctors detect edema by visible swelling, shiny or stretched skin and pitting where it creates a temporary stain under pressure.In pulmonary edema fluid builds up in alveolar space of lungs causes cough chest pain dyspnea. To identify the issue, doctors perform a

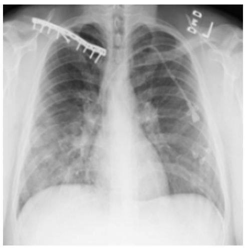


Figure 2.6: CXR image of Edema

physical exam, check the patient's medical history, do blood and urine tests to see how the kidneys, liver and heart are working. They may also utilize imaging tests like an ultrasound or chest X-ray. Management requires treating the underlying condition and taking steps to help, such as lowering salt intake, raising the affected limbs, wearing compression stockings, and taking diuretics to get rid of extra fluid. In critical situations pulmonary edema may need urgent treatment with oxygen help and hospitalized.

Fibrosis: Fibrosis is scarring of the lung tissues mainly caused by inflammation and long term injury. It causes idiopathic, autoimmune cause, environmental pollution, medications .This happens when healthy lung tissue is replaced with stiff scar tissue which makes the lungs inflexible and restricts oxygen intake and release. The consequences are critical include increasing shortness of breath, a dry cough, decreased stamina. In the worst cases, respiratory failure and death. Fibrosis is a major problem around the world. Idiopathic pulmonary fibrosis affects people and it can also cause death. In the beginning, it is commonly missed since the first signs, like a moderate cough or shortness of breath, are not distinct and might be confused for problems connected to aging, asthma, or smoking.

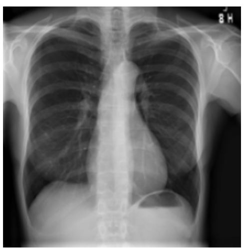


Figure 2.7: CXR image of Fibrosis

High-resolution CT scans, lung biopsies and pulmonary function tests are some of the specialized processes that can verify the condition. These tests aren't usually available in places with few resources. It is important to find out about fibrosis early because it cannot be reversed. Immediate treatments like antifibrotic medications, oxygen therapy, pulmonary rehabilitation and lifestyle modification can slow the development of the condition, ease symptoms and enhance well being.

Hernia: Hernia is abnormal protrusion of a viscus or part of it from its normal anatomical position due to weakness in the wall which contains it. In situations of lung hernia, organs in the abdomen, like the stomach or intestines, can be seen in the chest cavity.This looks like soft tissue lumps or loops filled with gas above the diaphragm. It might move the lung tissue out of the way. Often linked to birth abnormalities or injuries. This happens most often in the abdominal wall. A hernia can cause anything from sudden pain, discomfort and bulging to more problems including

Figure 2.8: CXR image of Hernia

blocking the intestine or strangling which cuts off blood flow to the trapped tissue, killing it and causing life-threatening situations.Small hernias can be hard to find because they don't often show many symptoms at first and might be mistaken for muscle strain or basic stomach pain, which makes it harder to get the right diagnosis. Early diagnosis really matters because timely surgery can avoid damaging consequences, pain and make life better.

Chapter 3

Methodology

3.1 Dataset

3.1.1 Dataset Acquisition

We have sourced our dataset from multiple open source dataset. Since we are working with a problem that requires an imbalanced dataset that is why we were looking for rare diseases among these open sourced available CXR image datasets and we have found that Edema Fibrosis and Hernia has very few data samples available compared to other CXR images.
In our dataset we have a total of 9 different chest X-ray diseases and a total of 17,615 frontal Chest X-ray radiology images.

Disease	**Count & Source**
COPD	3510 (PadChest)
Covid-19	3017 (Kaggle)
Edema	100 (NIH CXR-14)
Emphysema	2550 (NIH CXR-14)
Fibrosis	105 (NIH CXR-14)
Hernia	106 (NIH CXR-14)
Normal	3271 (Guangzhou Women and Children's Medical Center)
Pneumonia	4273 (Guangzhou Women and Children's Medical Center)
Tuberculosis	1683 (Kaggle)

Table 3.1: Disease classes with sample counts and dataset sources

We have collected Pneumonia and Normal chest X-ray images from Guangzhou Women and Children's Medical Center, Guangzhou. The dataset was uploaded in kaggle. All the CXR images were collected as part of patients' routine clinical care. The images were examined by the two expert doctors before the dataset received approval for training any AI system.
Another large CXR dataset we used was PadChest, a large scale high resolution chest x-ray image dataset introduced in the paper titled as, "PADCHEST: A LARGE CHEST X-RAY IMAGE DATASET WITH MULTI-LABEL ANNOTATED REPORTS". The author mentioned in the paper that the dataset has more than 160,000 CXR images from 67,000 different patients in between 2009-2017. Since we are working with multi-class dataset and PadChest has mostly multi-labeled classes therefore

we wanted to select images that are not multi-labeled and classes that have a good amount of single labeled data samples. This is why we have selected COPD (Chronic Obstructive Pulmonary Disease) from this particular dataset because among the single labeled dataset it has the highest amount of data samples.

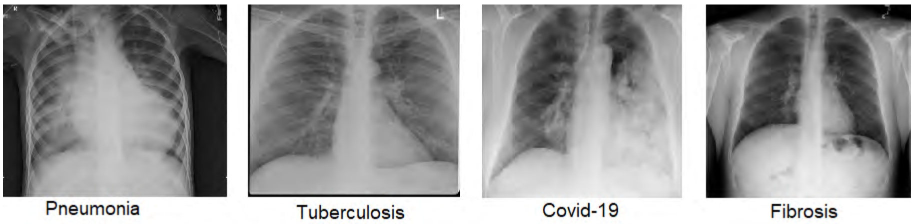


Figure 3.1: Some of the CXR Images

We have collected our 4 Chest X-ray disease Edema, Emphysema, Fibrosis, Hernia images from another large scale chest x-ray image dataset called NIH-CXR14. This dataset was published by . The dataset has a total 112,120 CXR images collected from 30,805 unique patients. Similar to PadChest this dataset is also mostly multi-labeled dataset hence first we had to single out single labeled images from the dataset and then we chose Edema, Emphysema, Fibrosis, Hernia from this particular dataset. Lastly we have sourced Covid-19 and Tuberculosis images from kaggle's open sourced dataset.

3.1.2 Data Analysis

The dataset we created by sourcing images from multiple datasets is highly imbalanced. Edema, Fibrosis, Hernia these three classes have very few data samples compared to the other 6 classes. Therefore, the dataset is perfectly suitable for our research problem because we needed a dataset where few classes will suffer from data scarcity and will represent as rare classes. From the frequency bar chart given below it is clearly visible that the class distribution is not uniform and the height of the aforementioned three classes Edema, Fibrosis, Hernia is much smaller than the other classes. All of the images of our dataset is in png format and since the dataset was sourced from multiple sources hence the size of the images varied. Therefore, we resized all the images of our dataset into 256*256 pixel size. Also it is worth mentioning that since CXR image datasets are mostly multi-labeled for that reason we first filtered out the single labeled images first and selected the desired images from the chosen class.For instance, NIH-CXR14 is mostly multi-labeled but it also has single labeled data samples. As a result, we filtered out the single labeled images first and then performed classwise separation of Edema, Emphysema, Fibrosis, Hernia. Some of the images were blurry or noisy hence we applied a standard procedure to enhance the quality of those images using a widely used technique called CLAHE. This technique is briefly described in the Data pre-processing section.
The following figure depicts the comparison of before and after applying CLAHE. Before applying CLAHE the original image looks blurry specially in the chest region and the edges inside the rib cage do not look much visible. After applying CLAHE the contrast of the image improved and the edges inside the rib cages looks visible better than the original image

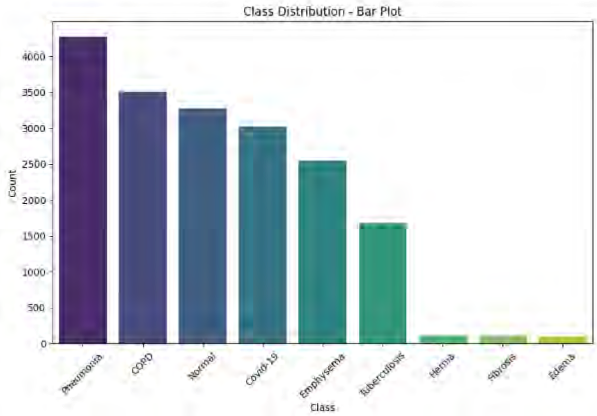


Figure 3.2: Class Distribution

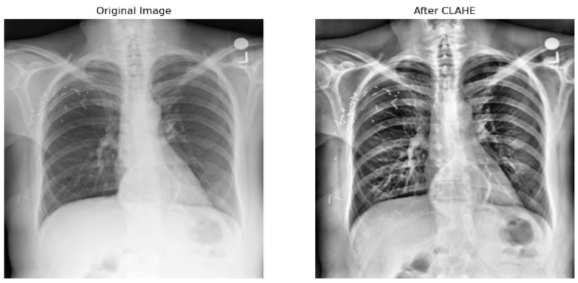


Figure 3.3: Before and After CLAHE

3.1.3 CLAHE Technique

Image enhancement plays an important role in modern image processing. Many images taken in real-world conditions have poor visibility or low contrast. Histogram equalization is a common way to make contrast better. On the other hand, this method uses the same change on the whole image which can make key local details disappear. A technique called Contrast Limited Adaptive Histogram Equalization, or CLAHE, was devised to fix this issue. CLAHE is used a lot in medical imaging, analyzing satellite images, taking pictures in low light, and security systems. It brings out small details without changing the picture.

CLAHE is better than the old Histogram Equalization (HE) and Adaptive Histogram Equalization (AHE) approaches. In simple histogram equalization, the brightness values of the complete image are spread out such that the contrast covers the whole range of intensities. This often makes things easier to see, but it also has certain problems. Global histogram equalization doesn't take into account local differences, therefore certain places may still look excessively dark or too bright. Adaptive histogram equalization was made to remedy this. AHE splits the image into distinct parts and does histogram equalization on each one separately. This makes it easier to see local details. AHE can also make noise louder, though, especially in areas where the pixel intensities are all the same. CLAHE was made to fix this problem by adding an extra step that limits contrast enhancement to make it better than AHE.

CLAHE operates through an orderly sequence of actions. Initially, the image is partitioned into small, square areas called tiles, with neighboring tiles kept apart so local yet distinct structures can be refined. Each tile is self-contained during processing, so individual regional patterns can be sharpened. A pixel intensity histogram is computed for the tile, and classical histogram equalization mechanism is used next, whose aim is to boost image contrast. To circumvent the aggressive equalization that large peaks can enqueue, the algorithm limits the total amount of pixel count that can pile into any single histogram bin. Any bin that would exceed the limit is adhered to the threshold, and any count that would overrun the threshold is redistributed evenly over the other bins. This reallocation retains detail without creating the over-enhanced stripes called contouring, especially on tile edges that lack texture. The neighboring tiles might still differ, however, continuing to save edges beside the tiles, CLAHE employs bilinear interpolation on the complete tile set. This averaging links interactions from the four tiles closest to an edge, creating smooth continuity across the tile boundaries and thus eliminating any modular appearance,constantly leaving the enhanced texture looking produced rather than sheeted.

3.1.4 Image Transformations

To introduce diversity and reduce any possible overfitting issue we have used several image transformations techniques in our training data pipeline. Image transformation helps the model to generalize better. The followings are the techniques we used in our experiments-

- Image Resize and Center Crop: We resized the CXR images by 256*256 pixel values from original image size at first then did a center cropping so that the images gets cropped from the center and the input image size becomes 224*224 finally. Since almost all of the pre-trained CNN model in computer vision requires an input size of 224*224 image

- Random Horizontal Flip: This transformation technique does a random horizontal flip of the input image.

- Random Rotation: This technique rotates any given image to a certain degree. We provided the degree to be 10.

- Random Affine: This transformation does multiple random operations on an image object including translation, rotation, shearing and scaling. It is a kind of geometric transformation that maintains the lines and parallelism

- Color Jitter: With the help of pytorch's color jitter transformation we can tweak an image's color properties such as brightness,contrast,hue and saturation.

3.1.5 Dataset Splitting

Our dataset has total 17,615 images. We have splitted our dataset into 80:20 ratio. 80% of the dataset was used for training and the rest of the 20% was used to create validation and test dataset. Class distribution table of Train/Test split is given below:

Disease Name	Train	Validation/Test	Total
COPD	2808	351 + 351	3510
Covid-19	2417	300 + 300	3017
Edema	80	10 + 10	100
Emphysema	2050	250 + 250	2550
Fibrosis	84	11 + 10	105
Hernia	84	13 + 9	106
Normal	2671	300 + 300	3271
Pneumonia	3418	428 + 427	4273
Tuberculosis	550	65 + 68	683

Table 3.2: Train, Validation/Test, and Total distribution of diseases

3.2 Model Specification

3.2.1 Supervised Learning

Supervised learning is a type of machine learning where the model is fed both the input and output called as labels so that the model or the algorithm can learn the function in order to map input features to its output. An efficient model learns this mapping function very well so that it can perform better to real world unseen data. In our proposed hybrid architecture, during phase 1 we have used the supervised learning approach to train a pre-trained CNN model to identify the common classes of our dataset.

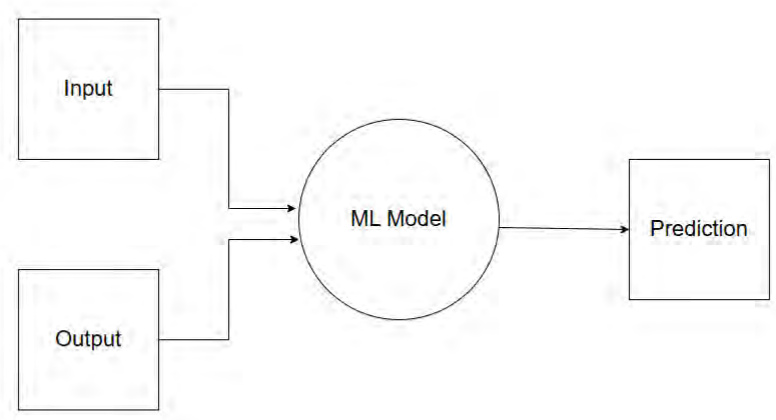


Figure 3.4: Supervised Learning

3.2.2 Transfer Learning

Training a deep neural network with relatively less amount of data is challenging. Deep neural networks require a sufficient amount of data to generalize the model for better performance but in the real world most of the time it is not practical to have a sufficient amount of high quality dataset. Therefore, to overcome this challenge transfer learning is introduced so that a model that generalizes itself to huge amounts of data can transfer its already learned knowledge so that they can be reused for different tasks. Pre-trained CNN models used in computer vision tasks are generally trained on a huge dataset containing millions of images with lots of classes such as ImageNet dataset that has 1.2 million images and 1000 classes. In our research we have leveraged the method of transfer learning using multiple Pre-trained CNN models like MobileNet-V2 , DenseNet-121 and EfficientNet-B1. In order to achieve an efficient and reliable performance, we fine tuned these models on our dataset in Phase 1 of our proposed hybrid architecture. We have basically initialized all of the models with their respective ImageNet weights for faster convergence. Following the strategy of continuing the update of pre-trained weights through backpropagation we have basically fine tuned all the layers of the Pre-trained CNN models and also retrained the classifier for our specific task. This strategy of fine tuning ensures that previous helpful knowledge these Pre-trained models had can transfer and provide a good start for the new learning task since X-ray images are very different from ImageNet dataset and also this helped to reduce the risk of overfitting problem. In Phase 2 of our proposed hybrid architecture where we have applied Few Shot learning models, we have used transfer learning as a means of collecting the vector embeddings of the input images so that a relation module can analyze the feature embeddings and can generate the relation score in order to classify the same class images. To achieve this we have removed the final classifier layer of the pre-trained CNN model and passed the output of the last convolution or pooling layer as an input to the relation module. In essence, during phase 2 we leveraged the transfer learning method as a fixed feature extractor. On the other hand, during phase 1 transfer learning helped us to generalize the model with faster convergence due to its pre-trained weight initialization.

3.2.3 Few-shot Learning

One of the key constraints of supervised learning is, it requires a humongous amount of data to generalize the given task therefore we can say that supervised learning is a data hungry learning technique. But there are some scenarios that exist in the real world where accumulating data in bulk is not possible or costly. For instance, in zoology there could be some rare species which images are difficult to collect, in healthcare there will always be some rare disease which will have scarce data since only few people are diagnosed with it. Therefore, to fight the obstacle of collecting more data, few-shot learning is introduced where models can learn without the need of seeing a bulk amount of data. There are several strategies available for Few Shot learning such as Prototypical network, Matching network, Relation network, MAML etc. In our hybrid architecture we have used the Relation Network approach. In Few shot learning training setup, there are support set images and query set images.

Support and Query Set: Support set has equal amount of labeled images across each class which means N number of classes will have exactly K amount of data points hence the size of a support set will become N*K. On the other hand, the query set will have an equal length of support set but it will contain unseen data points from N classes and each class will have the same K number of data points as the support set had.
Few shot learning operates on episode-based setup. In each of the episodes data points are collected through support and query sets and fed to the model. For instance, in a N way K shot setup each episode will collect N*K support and query set images and proceed for further training. The model uses the support set images to learn the features of the different classes and evaluates its learning through the new unseen images it has not seen in training.

3.2.4 Convolutional Neural Networks

Convolutional Neural Network is one of the most important class of deep neural networks which is used to solve tasks related to computer vision such as image classification , object detection , video recognition, speech recognition etc. CNN's are highly effective and excels in image classification because it automates the feature extraction part hence no manual feature engineering is required. Convolutional Neural Networks have three distinctive layers for different purposes.

- Input Layer
- Convolutional Layer
- Pooling/Subsampling Layer
- Fully Connected Layer

Input Layer: In this layer vector representation of the raw data samples/ input images are fed to the initial convolutional layer. For image classification task, typically the size of the input images vary based on the architecture of a CNN model but 224*224 is widely used.
Convolutional Layer: In a Convolutional Neural Network based model this layer is the most important and also it is the building block of a CNN model. The convolutional layer takes data as inputs then applies some filters over the inputs and extracts valuable features to generate a feature map. The filters or kernels move across the entire input images and find the best features; this operation is known as convolution. Filters are a 2D array of weights. The size of the filters vary but typically the size is 3x3 matrix. When the filter is applied to a part of the input image, it performs dot product with that particular part of the image and generates the output known as feature map. Stride is the amount of pixel filter will move for the dot production. If stride size is 1 it will move 1 pixel if the stride is 2 it will iterate the entire input image based on 2 pixels at a time. The filter continues the operation according to the stride size across the entire input image. After going through the entire image the output it generates is a complete feature map. A pre-trained CNN model has several numbers of Convolutional layers for generating feature maps. The early Convolutional layers generate lower level feature maps of an input image such as the horizontal, vertical edges. The top Convolutional layers

generate higher representations of an image. The below two equation determines the height and width of the output of a convolutional layer.

$$W_{out} = \frac{W_{in} - F + 2P}{S} + 1$$
$$H_{out} = \frac{H_{in} - F + 2P}{S} + 1$$
$$\text{Output feature map dimension:} \quad W_{out} \times H_{out} \times K$$

The output feature map is then passed as the next input layer of the next convolution layer.

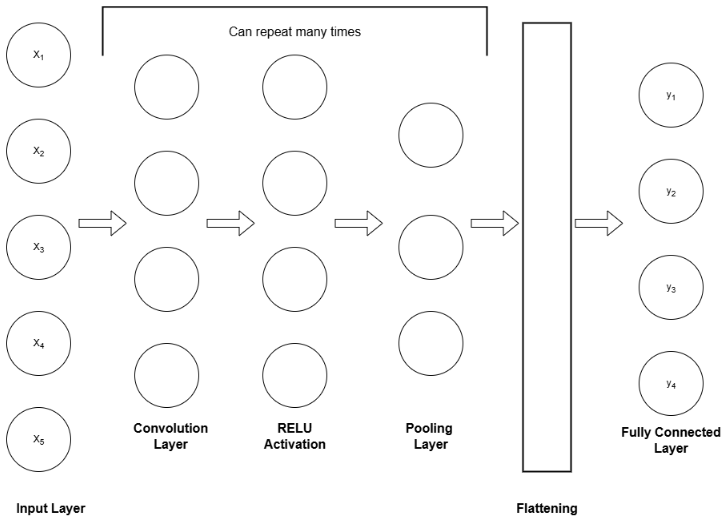


Figure 3.5: Standard CNN Layer

Pooling Layer: The purpose of pooling layer is to reduce the complexity, improve efficiency and minimize the possible overfitting problem. It reduces the dimension of feature maps by pooling the most important part of only. Similar to the convolution layer the pooling operation, a filter iterates over the input according to the given stride and extracts or pools useful information and leaves the redundant part. There are various types of pooling methods but Max pooling and average pooling are the most common. They reduce spatial dimensions by selecting the maximum or calculating the average value in pooling regions
Max pooling: The filter iterates over the entire input data with the given stride then it fetches the important pixel that has the maximum value. After the end of the iteration the output information has less dimension compared to its input with only the valuable and important features. This method is widely practiced compared to the average pooling method.
Average pooling: The operation is similar to the Max pooling but instead of pooling the maximum value, it calculates the average value within the input to generate the output array.
Fully Connected Layer: After the inputs are passed through the convolution and pooling layer the final output is flattened which means the higher dimensional feature maps or embeddings are converted into an one dimensional array and then this array of information is passed to the Fully Connected Layer with the purpose

of identifying the generalized input-output function that will help to predict the correct output after the completion of training the model over unseen data sample. In a fully connected layer, every neuron is connected to every neuron in the previous and subsequent layers as depicted in the following figure.

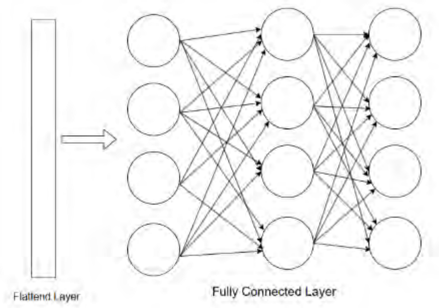


Figure 3.6: Fully Connected Layer

3.2.5 MobileNet-V2

MobileNet-V2 is an upgraded version of MobileNet-V1 architecture which was published by A. Howard et al. (2017)[12]. The purpose of MobileNet-V2 network is to deploy CNN architectures on mobile devices with more efficiency and less computational requirements. MobileNet-V2 was introduced by M. Sandler et al. (2018) based on the principle of MobileNet-V1. Similar to MobileNet-V1 it also has depthwise separable convolution operation which is a key operation that reduces the computation complexity. Novel ideas that are introduced in MobileNet-V2 are Inverted residuals, Linear Bottlenecks and depthwise separable convolutions.

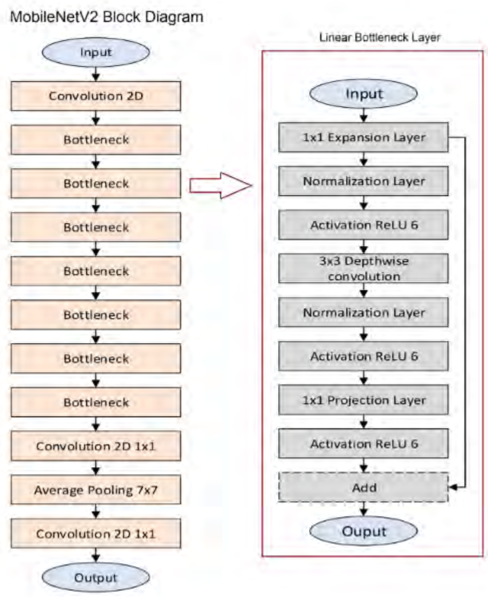


Figure 3.7: MobileNet-V2 Architecture block diagram

Depthwise separable convolutions: This particular operation is responsible for MobileNet architectures to reduce the computational requirements heavily. This convolutional operation is different from standard convolutional operations. Depthwise separable convolutional operation is divided into two parts:- depthwise and pointwise convolution. Standard convolution operations are done on the entire input channel which means if the input channel is 3 the convolution operation is performed into these 3 channels at a time. On the other hand, depthwise convolution operation is performed onto one single channel at a time which reduces the number of multiplications compared to a standard convolutional operation. Since

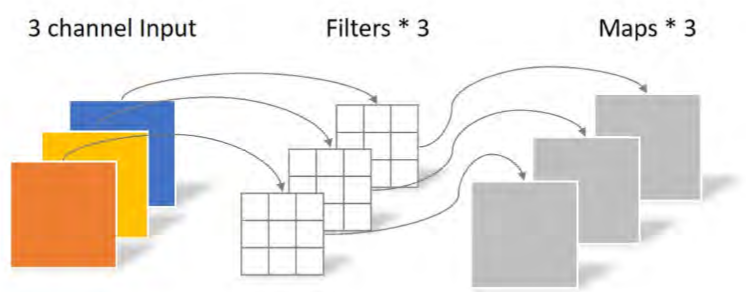


Figure 3.8: Depthwise Convolutional Operation

the depthwise convolution outputs independently for each of the input channels, there must be a way to combine them together. This is exactly what pointwise convolution does. By performing linear combination pointwise convolution binds all the outputs of depthwise convolution per channel and generates the final feature map. Pointwise convolution is a 1x1 kernel.

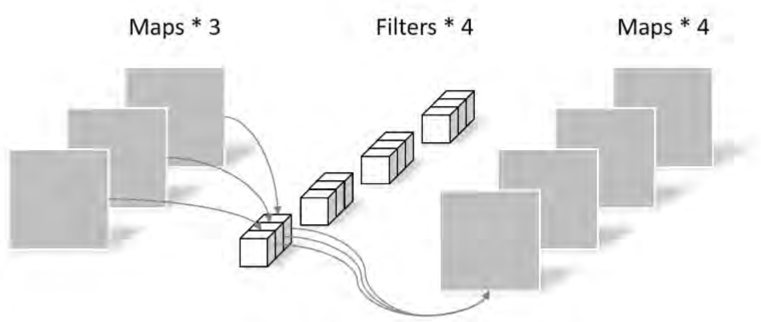


Figure 3.9: Pointwise Convolutional Operation

Linear Bottleneck: M. Sandler et al. (2018) claimed that expanding and compressing the feature maps using non-linear activation functions has its downsides. Exposure to non-linearity in low-dimensional spaces can lead to a good amount of information loss. Therefore, to minimize the information loss they introduced linear bottlenecks. Linear bottleneck layers reduces the feature map channels without using the non-linear activation functions like ReLU
Inverted residuals: Inverted residuals are different from Traditional residual blocks hence the name inverted. In traditional residual block expansion is done first and then compression is performed. On the other hand, inverted residuals do the opposite, it is first compressed and following the compressed operation, expansion is done.

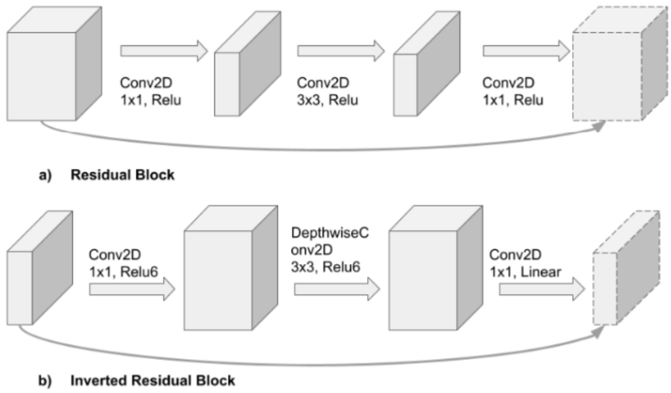


Figure 3.10: Difference between standard residual block and MobileNetV2 inverted residual block.

Due to the previous bottleneck layer, the input is already compressed. Therefore, following by a compressed input a 1x1 pointwise convolutional operation expands the compressed input to a high dimensional feature map then depthwise convolutional operation is performed and right after that 1x1 pointwise convolution is done so that the high dimensional feature map can be projected into a low dimension which is a compression operation. Between the low dimensional input and the last projected low dimensional output of a bottleneck layer there is a skip connection or residual connection shown in the figure with an arrow.The purpose of skip connection is during the model training gradients can flow smoothly.

3.2.6 DenseNet-121

DenseNet121 is a deep convolutional neural networks (CNNs) introduced in "Densely Connected Convolutional Networks" by Gao Huang et al. in 2017 [6]. At variant with traditional architectures such as VGG or ResNet, DenseNet121 fully connected across layers in a feed-forward way, thus making feature maps from the current layer combines features from all previous layers. This close connection allows feature propagation, nullifies redundancy and reduces the parameters significantly compared to other deep networks. The "121" in its name refers to the number of 121 layers of the model. DenseNet121 is an efficient architecture that reduces the vanishing gradient issue, improves accuracy and it also has been implemented in models as image classification, medical image processing and object detection. DenseNet121 is different from other networks mostly in layer connectivity such as that in the majority of standard CNNs such as VGG or even ResNet.
Every layer receives input from only the layer before. DenseNet121 puts forward dense connectivity where every layer receives input from all the preceding layers. Every layer delivers its feature maps to all succeeding layers. The structure encourages reuse of features, reduces the number of parameters, and enhances network efficiency in relation to depth when compared to similar-depth deep networks. It encourages relief from vanishing gradient problems, offers better performance and enhances information propagation with fewer parameters compared to most other deep CNNs. Also, DenseNet121 propagates more features since the earlier features are preserved and reused through the network, and it does a better job at extract-

ing fine-grained patterns. Although more memory-intensive since it keeps multiple feature maps, its better parameter efficiency and training stability make it a good candidate for deep learning tasks. As a result of these advantages, DenseNet121 has widespread use in applications such as medical image analysis, object detection and classification where accuracy alongside efficiency becomes the most important factor. With the help of its dense connectivity structure DenseNet121 minimizes

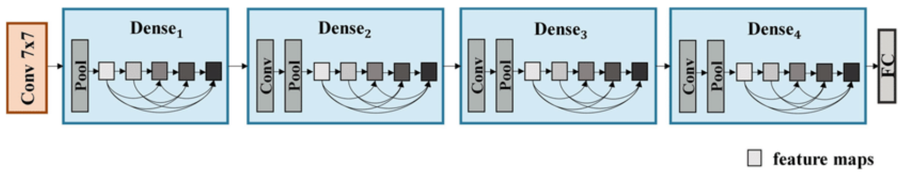


Figure 3.11: DenseNet-121

the vanishing gradient problem. In this framework, each layer is directly connected to all the earlier layers, and therefore backpropagation gradients have many short paths to follow. This means the error signals do not have to go through many stacked layers, which generally causes them to decline and eventually vanish in very deep networks. By ensuring that there is stronger gradient flow, DenseNet121 stops even the earlier layers from receiving negligible updates during training. Along with overcoming the vanishing gradient issue but also improves the reuse of features and allows the model to achieve high accuracy with less parameter usage.
The DenseNet121 architecture consists of dense blocks and 3 transition layers in between. Each dense block is composed of repeated layers, and every layer is composed of a set of modules: batch normalization, ReLU activation, a 1×1 convolution (bottleneck) to reduce feature dimensions, and a 3×3 convolution for extracting features.
The uniqueness of a dense block is that each layer is fed by all the earlier layers and its own output is passed to all the subsequent layers, thereby promoting feature reuse and efficient learning. Between dense block there is a transition layer which sole purpose is to reducing the feature map size. Transition layer incorporates batch normalization, a 1×1 convolution to scale down the feature maps and a 2×2 average pooling operation to scale down the spatial dimensions. Dense blocks increase the network by reusing and holding features, while transition layers scale down and compress the feature maps to keep the architecture computationally efficient and avoid overgrowth.

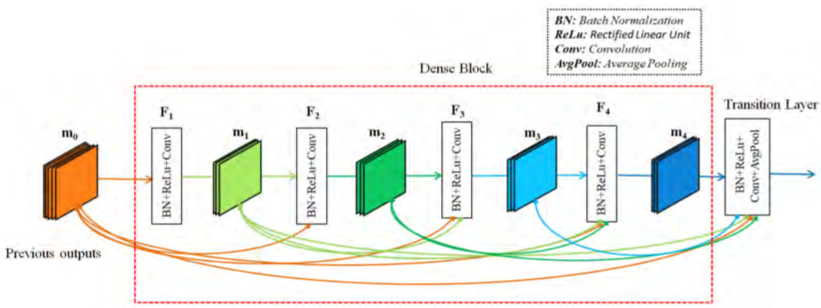


Figure 3.12: Dense Block

3.2.7 EfficientNet-B1

EfficientNet is a state of the art image classification architecture by Google AI Research [16] to scale the Convolutional Neural Networks up and achieve significantly better performance using fewer parameters. The goal here is to use a systematic and principle way to add compound scaling which means adding depth (adding more layers), width (adding more feature maps) and resolution (increasing the input image resolution) to the Convolutional Neural Network. Compound scaling is one of the features of EfficientNet that differentiates it from the other traditional CNN models such as ResNet. For example, in the ResNet model, if we want to scale the model up we add only layers, such as ResNet50 to ResNet200. On the contrary, EfficientNet can scale depth, width, and resolution all at the same time while being more efficient and performative. For example, on ImageNet, EfficientNet achieves state-of-the-art accuracy while EfficientNetB0 uses 5m parameters for the same task that ResNet50 uses 25m parameters. The body of the EfficientNet architecture also consists of MBConv blocks with different configurations whereas traditional CNN models use standard convolutional layers.
Architecture: EfficientNet through the use of compound scale method scales up the model efficiently and effectively. It uses Mobile Inverted Bottleneck (MBConv) layers which are made up of Depth-wise Separable Convolution layers and Inverted Residual Blocks. Through the use of Efficient Attention Mechanism the model further enhances its performance and efficiency. **Compound Scaling:** The main

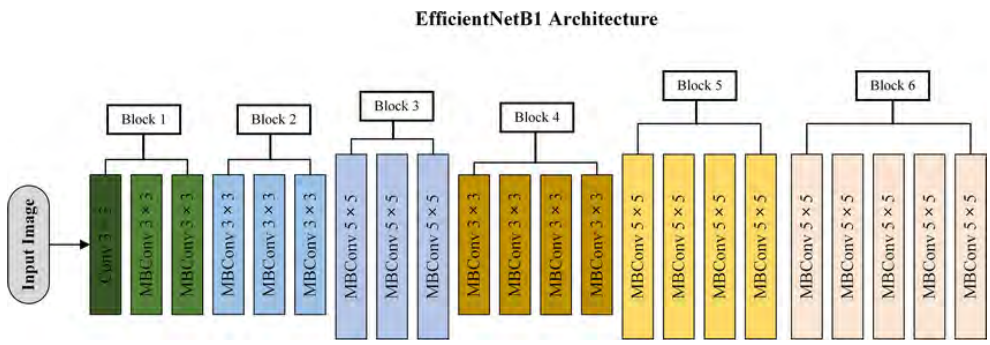


Figure 3.13: EfficientNetB1

component of the EfficientNet architecture is the compound scaling method which changes the depth, width and resolution of the model. But the adjustment is not done randomly, rather it is done uniformly scaling each dimension with a compound coefficient. Depending on the situation the user can choose the appropriate compound coefficient value for the model to make it lightweight or powerful. The constant coefficients are determined by a small grid search on the original model.
Depth-wise Separable Convolution: In order to lower computation expense EfficientNet uses Depth-wise Separable Convolution which replaces the standard convolution with two distinct steps: a depthwise convolution and a pointwise convolution. Firstly, Depthwise convolution is applied to a single channel at a time whereas for the standard convolution it will be applied for all channels of the input data. Secondly, Point-wise Convolution which is a convolution layer with 1×1 kernels is applied. It aggregates features from different channels.
Inverted Residual Blocks: EfficientNet uses inverted residual blocks which are inspired by MobileNetV2 to maximize on computational cost. The blocks work

in a 3 step workflow where at first comes the expansion phase where the number of feature maps are increased and then depth-wise convolution is done using a 3x3 convolutional bottleneck layer. Lastly, comes the projection phase where the number of feature maps is compressed back to the original input layer with 1x1 convolutional layer.

Efficient Attention Mechanism: In order to improve the feature representation, EfficientNet uses squeeze-and-excitation (SE) blocks which by learning channel-wise attention weights can selectively boost the enlightening features while remaining computationally cost efficient. There are many variants of EfficientNet from B0 to B7 where B0 is the base model with modest depth, width, and resolution and B1 to B7 are successively larger models which were scaled up with the help of compound scaling coefficient.

3.2.8 Relation Network

Relation Network is designed to build a classifier that will be able to classify only seeing a few images of each class. The Relation network was introduced by F Sung et al. in 2017 [10] the network learns distance metrics between support and query images through the relation module instead of only comparing them with the help of various distance or similarity functions like euclidean or cosine which many few-shot methods such as siamese network does. The relation network is divided into two parts: the embedding module and the relation module. Embedding module generates the feature embeddings of the input images and relation module helps to generate relation score of those input images through an artificial neural network which further uses softmax for classification. On the other hand, other network such as siamese that uses only distance metrics like euclidean distance metrics only depends on the plain euclidean score to differentiate among the classes but relation network tries to learn the distance function among the classes through its relation module hence it depicted a good performance in our study compare to the siamese network. The performance comparison was shown through a bar chart in the Result Analysis section

The workflow of Relation Network:

Dataset: In our study we have used 3 way 5 shot settings. In this setting support set images will have 5 images of each of the 3 classes and 15 query images. Since relation network is episode based training each episode will consist of 15 support and query set images for training.

Embedding Module: According to the author of the original paper the purpose of embedding module is to generate feature representation of the query images and the labeled input sample images. The generated feature embeddings will then be concatenated as shown in the above figure and will be used as input of the relation module. In our study we have used MobileNet-V2 as the architecture of the embedding module since it performed better than other pre-trained networks in our dataset. To reduce the output feature map of the MobileNet architecture we used a global average pooling. After that we flattened the pooled features and projected into a lower dimensional embedding of size 128.

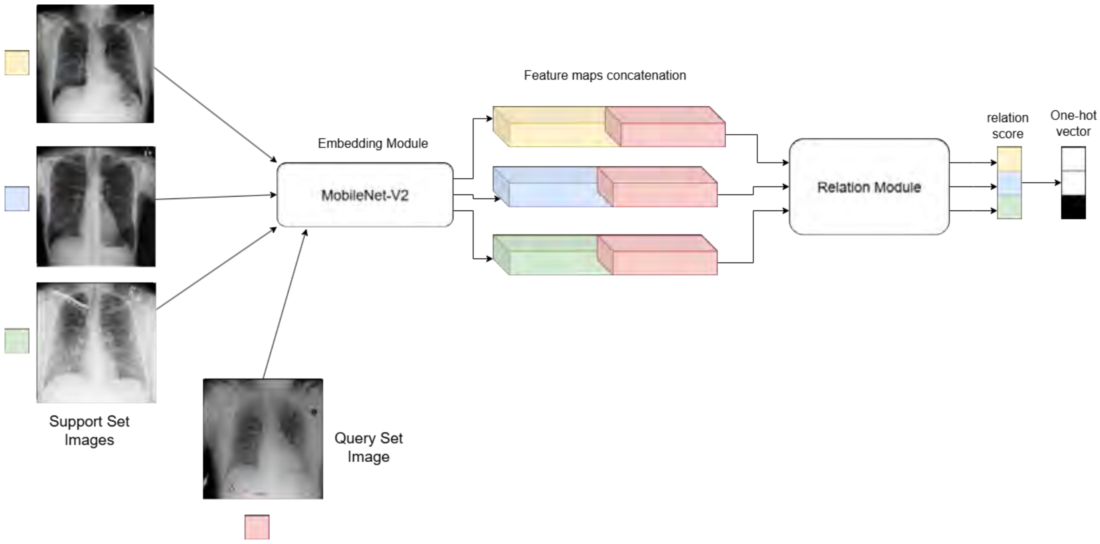


Figure 3.14: Relation Network

Relation Module: The purpose of this module is to learn the distance between support and query set images so that it can classify if the images belong to matching categories or not. The author of the paper stated that this module should use an artificial neural network hence we used a network of two fully connected layers of size 128 , to introduce non-linearity used ReLU activation and sigmoid function for classification.

3.3 Proposed Hybrid Pipeline

Because of the nature of our dataset which is imbalanced, a supervised model suffers to identify every class with the same good performance. The performance degradation of the single supervised model because of class imbalance was explored briefly in the result section. Therefore, to build a reliable architecture that will have good efficient performance across all the classes we propose a hybrid architecture by combining two different learning methods, Supervised and Few-shot learning. Since we can divide our dataset into two subset. The elements of subset one can be the classes that have vast amounts of data samples. We call this subset "Common class" and we named the other subset as a "Rare class" because the number of data samples of these classes are inadequate. The elements of the Common class subset are as follows,COPD,Covid-19,Emphysema,Normal,Pneumonia,Tuberculosis,Rare class[Edema, Fibrosis, Hernia]

Figure 3.15: Proposed Hybrid Architecture

We divided our entire problem into two sections. The first task is for identifying the Common class elements and we chose to use a supervised learning model for this task because this method is known for performing at a high level when there is enough data sample. On the other hand, since the other subset Rare class elements have less amount of data the second task was to find a learning method that is not dependent on the number of data samples. For that reason, we selected Few-shot learning since this learning method is designed with the purpose of performing at a

high level for scenarios when there is data scarcity.
To achieve this hybrid architecture we divided the entire architecture into phase 1 and phase 2.
Phase 1 has two goals-

- 1.Classifying the elements of the Common class subset with high precision and recall value.
- 2. Routing any data sample belonging to the Rare Class subset to the Phase 2 Few-shot learning method.

In Phase 2 we used a Few shot learning method so that the data sample rerouted from Phase 1 can be classified into the specific element of the Rare class subset.

3.3.1 Phase 1: Training of Supervised Learning

To achieve the goal of phase 1, first we introduced a new class called "Rare" into our dataset and this class has all the images belonging to any of the following elements of the Rare class subset which are Edema, Fibrosis and Hernia. All these three classes data samples are merged into one single class called "Rare". Therefore, with this modified dataset we trained several pre-trained CNN models to find the best performing model. The brief performance analysis of several pre-trained models are discussed in the result section. After conducting several experiments with different pre-trained models we observed that MobileNet-V2 architecture showed the best results compared to the other models by achieving an average value of 95.29% F1 score which indicates the high precision and recall value across all the classes. Hence, we selected the MobileNet-V2 architecture as the base model for classifying the common classes and rerouting the newly merged "Rare" class to the Few-shot model.

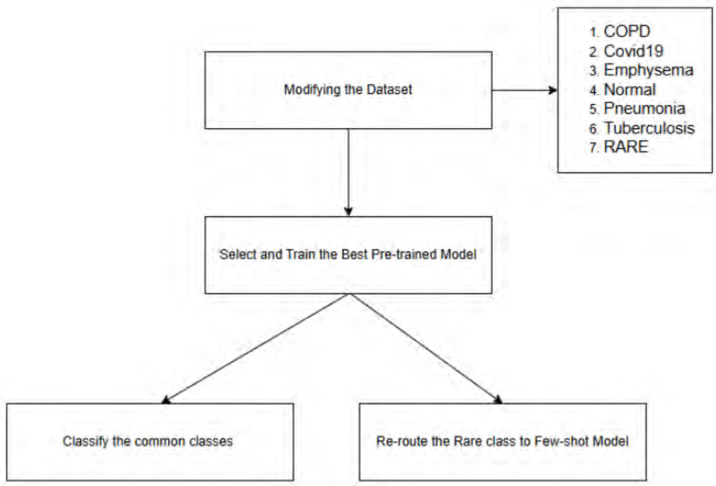


Figure 3.16: Steps in Phase 1

3.3.2 Phase 2: Few-Shot Learning model- Relation Network

To further classify the re-routed images from Phase 1 we have used a Few-shot learning model called "Relation Network". The sole purpose of Phase 2 is finding the best FSL model that performs much better than plain supervised models with limited data. We experimented with several FSL models and found that Relation Network achieved higher performance in our three Rare classes. The brief analysis of their performance was explored in the result section. The training procedure of Relation Network includes several hyper parameters such as number of classes known as N way and number of data examples per class known as K shot, then total number of episodes, base model for extracting features, input image size etc. We have used 3 way 5 shot training setup with 100 episodes and the size of input images were 224*224 since we have selected MobileNet V2 for extracting features. The relation network is divided into two parts. The first part is responsible for creating feature embeddings of support and query set images known as embedding module. The second part is known as the relation module which takes the support and query set images feature embeddings as inputs, analyzes them and generates the relation score between the support and query set images.Based on the relation score there is a softmax layer that does the classification of the images.

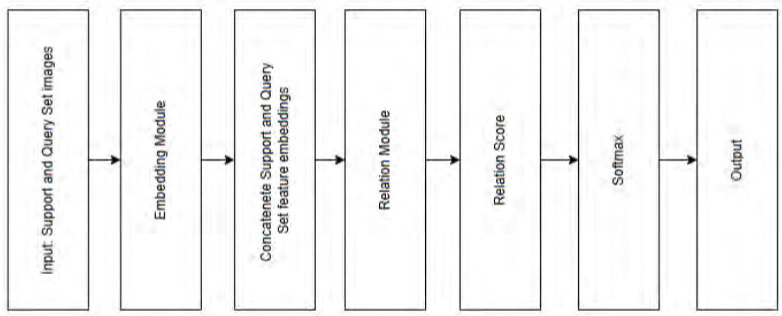


Figure 3.17: Relational Network Block Diagram

3.3.3 Hybrid Inference Workflow

After the completion of training both phase 1 and phase 2 we saved the best model of both phases. At inference time we used the saved best model to execute our hybrid architecture. As we discussed earlier, we treat Edema, Fibrosis and Hernia as a single Rare class in the phase 1 Supervised model. Therefore, if the input is any of the images belonging to these classes the model will predict them as the "Rare class" not Edema, Fibrosis or Hernia. For example, If we give an input image of Edema X-ray the supervised model should classify the image as RARE and re-route it to Phase 2 Relation Network to get the actual classification which is Edema in this case. After the Supervised model classified an image as "RARE" only then the Phase 2 Few-shot method will be applied to perform further classification of that image, which specific disease this image belongs to- Fibrosis, Hernia or Edema.

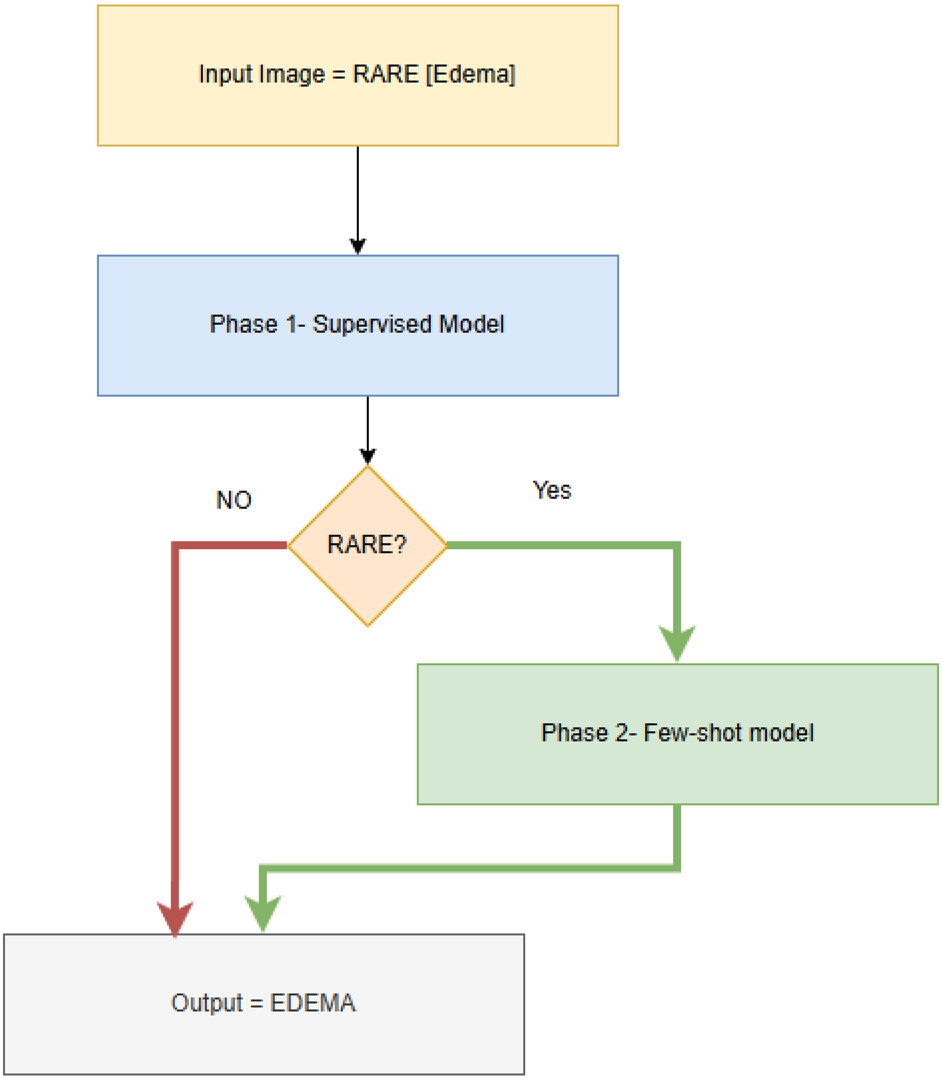


Figure 3.18: Hybrid Architecture work-flow when Input image is Edema

At the same time, the phase 1 supervised model will also treat the common classes as individual classes. If we input any of the images belonging to these common classes the model will predict which class it belongs to. For example, If we give an input image of Pneumonia X-ray the supervised model should classify the input image as Pneumonia.

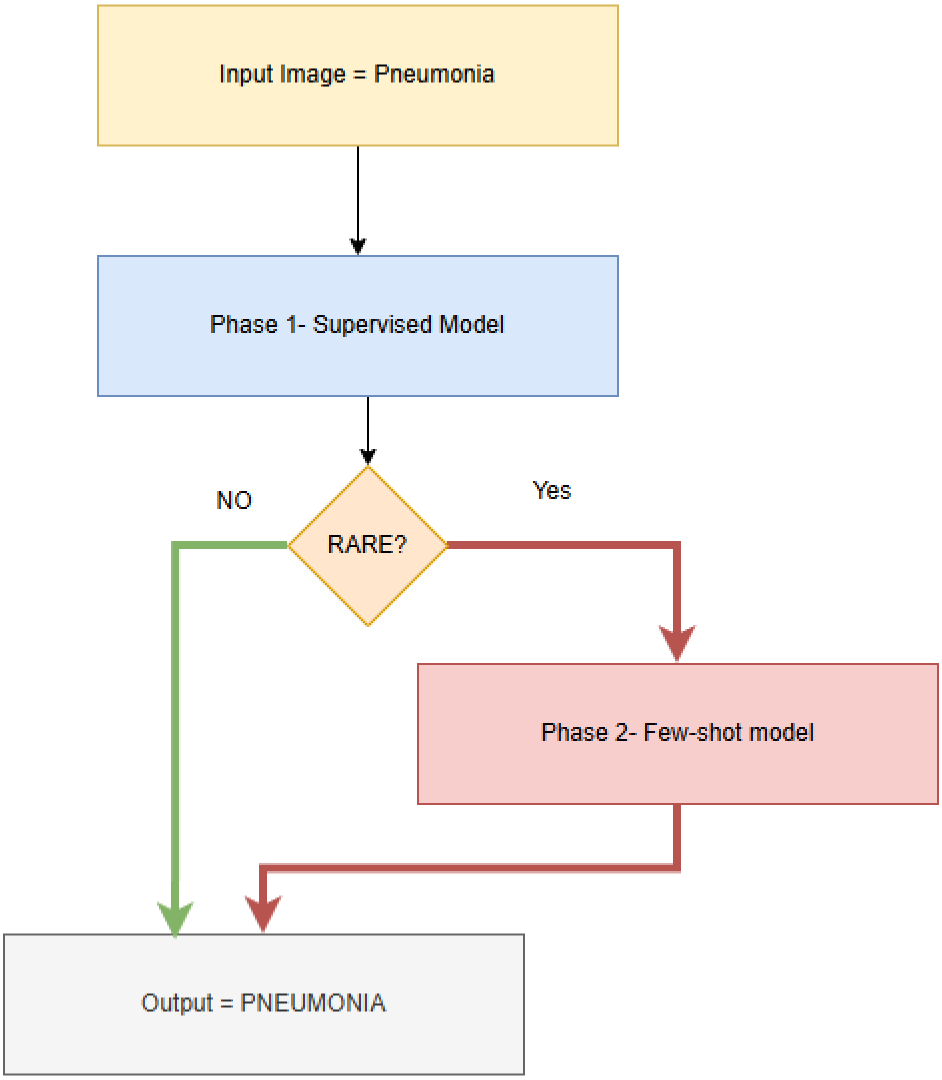


Figure 3.19: Hybrid Architecture work-flow when Input image is Pneumonia

To put it simply- Phase 1 Supervised model will answer the question if an image is a Rare disease or not? If the Supervised model's answer is, Yes, the few shot method will then return what specific rare disease the image has. Edema, Fibrosis or Hernia.

Chapter 4

Results

4.1 Result Analysis of Pre-trained CNN Models

4.1.1 Define Evaluation Metrics:

We have compared our hybrid model performance with classic supervised pretrained CNN models. To compare the models with the purpose of selecting the best scoring model, we have used Accuracy , Precision , Recall and F1 score as the performance metrics. Since we are dealing with medical X-ray image dataset it is extremely important to have a superior recall value and balanced precision value so that our model can have minimum false negative and positive cases while predicting the disease. To select the appropriate few-shot learning method we also used these metrics to select the most appropriate technique that suits our research purpose.
We know that accuracy is an elementary metric while evaluating a machine learning model's performance that basically indicates in general how many times a model was correct while performing the given task. The formula for accuracy is,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision is known for deducing what proportion of a model's prediction of positive cases were actually correct. The formula for precision is,

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: Recall is responsible for specifying the true positive rate of a model. This metric is known for showing how well a model captures the actual positive cases. Which means, it answers the question how much proportion of actual positive cases the model didn't miss. The higher the rate the better it is to never miss a positive case. The formula for recall is,

$$\text{Recall} = \frac{TP}{TP + FN}$$

Since our dataset is a medical dataset therefore the cost of missing a positive instance is high therefore it is highly important to minimize the False Negative cases as much as possible so that the model hardly misses a positive instance of a rare class disease.

- True Positive (TP): The model was right to predict the data sample as positive class.
- True Negative (TN): The model was right to predict the data sample as negative class.
- False Positive (FP): The model wrongly predicted the data sample of a negative class as a positive class. (false alarm).
- False Negative (FN): The model wrongly predicted the data sample of a positive class as a negative class (miss). For example, an image belongs to Hernia but the model predicted that image as not Hernia.

4.1.2 MobileNet-V2 Result and Analysis

We initialized the training of the lightweight CNN pretrained model MobileNet-V2 with its Imagenet weight. The input size was 224*224, learning rate was 0.001 and batch size was 32. The model was trained in 25 epochs and the best model was saved by evaluating the model on a held out validation set at epoch 21 with the best validation loss 0.21 during the whole training process. Model's performance on test set is shown in the below table-

Disease Name	**Precision**	**Recall**	**F1-Score**
COPD	100%	100%	100%
Covid-19	96%	88%	92%
Edema	50%	90%	64%
Emphysema	89%	96%	92%
Fibrosis	100%	18%	31%
Hernia	100%	8%	14%
Normal	99%	97%	98%
Pneumonia	98%	100%	99%
Tuberculosis	82%	100%	90%

Table 4.1: Precision, Recall, and F1-Score

From the table we can observe that the model performance was impeccable in all of the majority classes that have enough data samples but the performance on rare classes is totally unacceptable. Fibrosis and Hernia recall percentage is so poor that the model rarely captures these two diseases and misses almost every case from these two rare classes. Even if Edema recall is 90% the precision is 50%. And if we further analyze the confusion matrix we see that the model was confused between two rare classes, Hernia and Edema. Almost 70% Hernia positive cases were predicted as Edema.

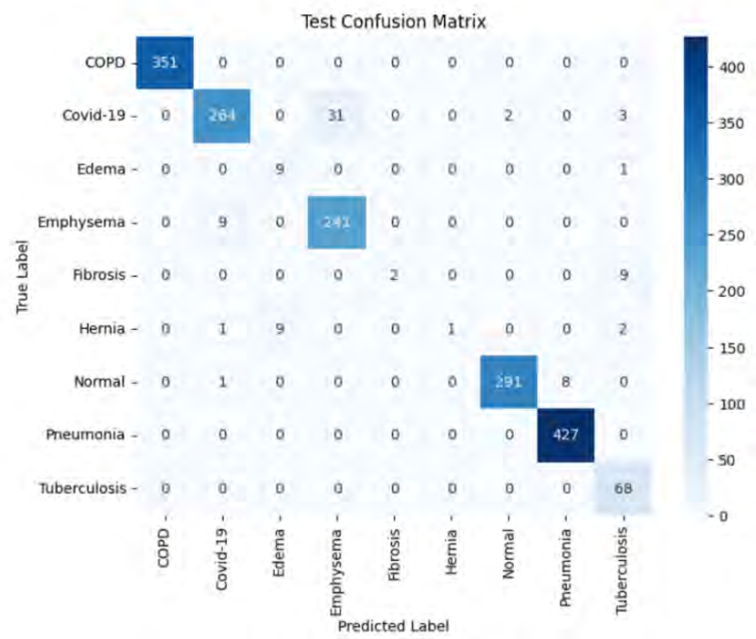


Figure 4.1: Confusion Matrix

Hence it is totally unacceptable even if the rare disease Edema scored 90% in recall value, models mixup with another rare disease Hernia discredits and makes the pretrained MobileNet-V2 model totally unreliable for all of the three rare diseases.

4.1.3 DenseNet-121 Result and Analysis

We initialized the training of the CNN pretrained model DenseNet121 with its Imagenet weight. The input size was 224*224, learning rate was 0.001 and batch size was 32. The model was trained in 25 epochs and the best model was saved by evaluating the model on a held out validation set at epoch 19 with the best validation loss 0.19 during the whole training process. Model's performance on test set is shown in the below table- From the table we can observe that just like the MobileNet-V2

Disease Name	**Precision**	**Recall**	**F1-Score**
COPD	100%	100%	100%
Covid-19	89%	90%	90%
Edema	100%	60%	75%
Emphysema	92%	84%	88%
Fibrosis	0.00%	0.00%	0.00%
Hernia	67%	15%	25%
Normal	97%	95%	96%
Pneumonia	97%	99%	98%
Tuberculosis	92%	97%	75%

Table 4.2: Precision, Recall, and F1-Score for different disease classes

model, the DenseNet121 model's behavior in Rare classes also is not up to the mark. Both of the models have performed in almost similar manner. They both showed very good results in all of the majority classes that have enough data samples but the performance on rare classes is poor due to the dominant factor of the majority classes. In this model it failed to recognize any images of Fibrosis and other two rare class Hernia and Edema recall percentage is so poor that the model rarely captured these two diseases and missed almost every case from these two rare classes.

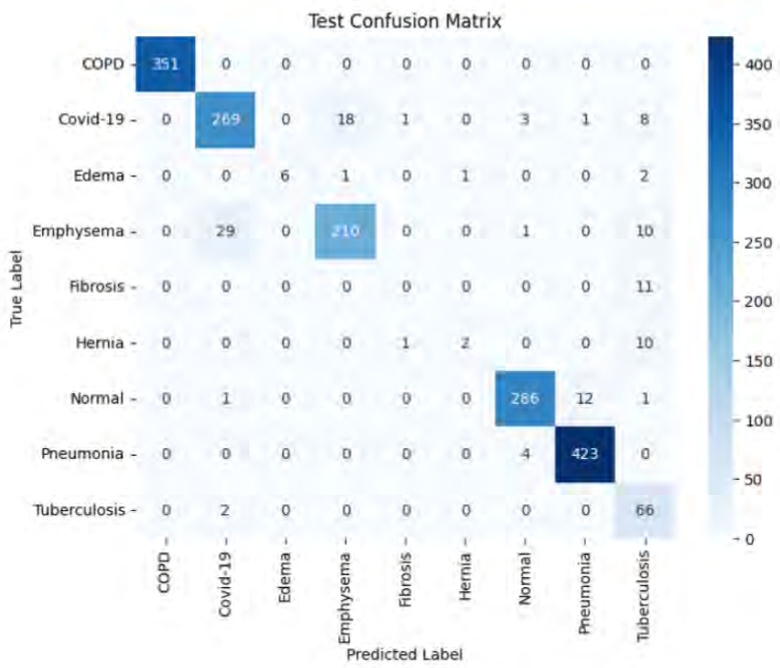


Figure 4.2: Accuracy and Loss Curve

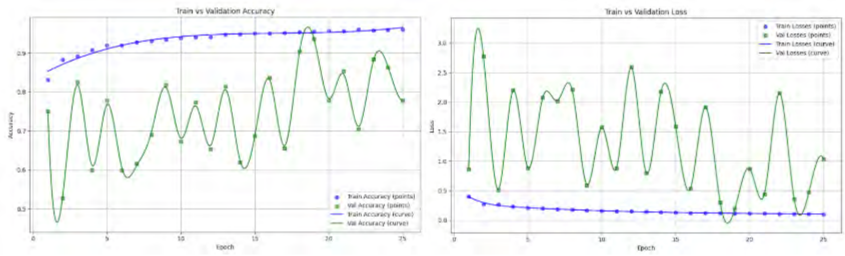


Figure 4.3: Confusion Matrix

4.1.4 EfficientNet-B1 Result and Analysis

We initialized the training of the CNN pretrained model EfficientNet-B1 with its Imagenet weight. The input size was 224*224, learning rate was 0.001 and batch size was 32. The model was trained in 25 epochs and the best model was saved by evaluating the model on a held out validation set at epoch 5 with the best validation loss 0.10 during the whole training process. Model's performance on test set is shown in the below table-
From the table we can observe that just like the above two pretrained CNN model's, EfficientNet-B1's behavior in rare classes is also not any different. Even though out of the three models EfficientNet-B1 has slight improvement in rare class metrics it is still not reliable. Therefore we can still claim that all three models have performed

Disease Name	Precision	Recall	F1-Score
COPD	100%	100%	100%
Covid-19	93%	86%	90%
Edema	100%	40%	57%
Emphysema	86%	92%	89%
Fibrosis	65%	100%	79%
Hernia	59%	77%	67%
Normal	99%	99%	99%
Pneumonia	100%	100%	100%
Tuberculosis	96%	99%	97%

Table 4.3: Precision, Recall, and F1-Score for different disease classes

in almost similar manner. EfficientNet-B1 has at least more than 50% F1 score in all of the three rare classes but in almost all of the common classes it has more than 90% F1-score which indicates that this model also shows biases like the other two models towards the classes that have vast amounts of data samples.

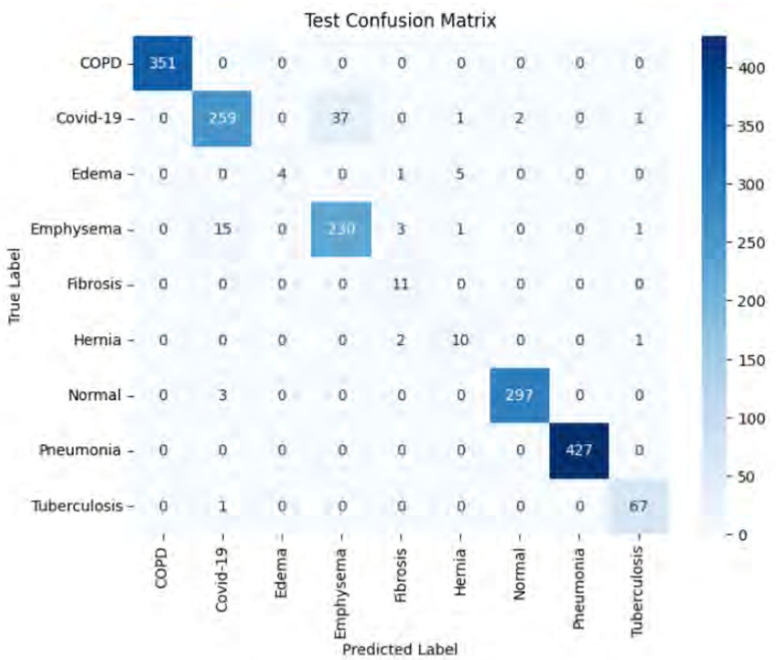


Figure 4.4: Confusion Matrix

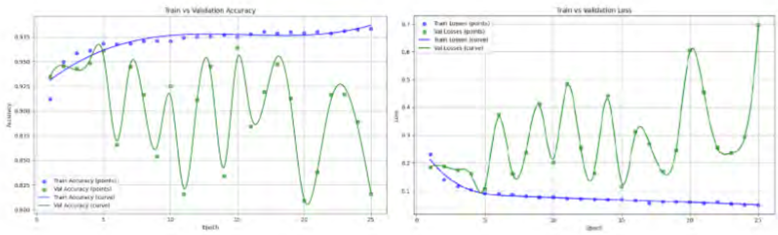


Figure 4.5: Accuracy and Loss Curve

Therefore based on the above experimentations and results on three different pretrained CNN model architectures we can conclude that the Classical Supervised learning approach fails miserably to identify the rare class diseases in an unbalanced data distribution. In the medical sector it is quite common to have unbalanced data distribution due to the fact that there will always be some rare disease with less data samples. Hence, it is important to find an optimal solution to tackle this problem with Artificial Intelligence so that the diagnosis can be reliable and trustworthy.

4.2 Phase 1 and Phase 2 Result and Analysis

4.2.1 Phase 1: Supervised Model

To find the best model suited for **Phase 1 supervised learning** we have experimented with three State of the Art pretrained CNN architecture- MobileNet-V2, EfficientNet-B1 and DenseNet121. All the three models achieved overall 91+ accuracy. Along with other class performance we mostly prioritized the **F1-score** of the class that was formulated combining the three rare disease classes from the actual dataset, named "RARE". In order to get the most benefit from the few-shot model in Phase 2, it is significantly important that Phase 1 model successfully captures almost all of the images belonging to the Rare class and hardly gets confused with other common classes. If this balance is not ensured then the purpose of building such hybrid architecture comes to nothing and F1 score correctly measures this balance since it is the harmonic mean of precision and recall value. The following table shows the F1-score performance metrics of all the three pretrained CNN models we have experimented on our dataset-

Disease Name	MobileNet-V2	EfficientNet-B1	DenseNet-121
COPD	100	100	100
Covid-19	89	91	85
Emphysema	89	90	83
Normal	99	97	92
Pneumonia	100	98	96
Tuberculosis	96	97	88
RARE	94	85	63

Table 4.4: F1 score comparison across different Models

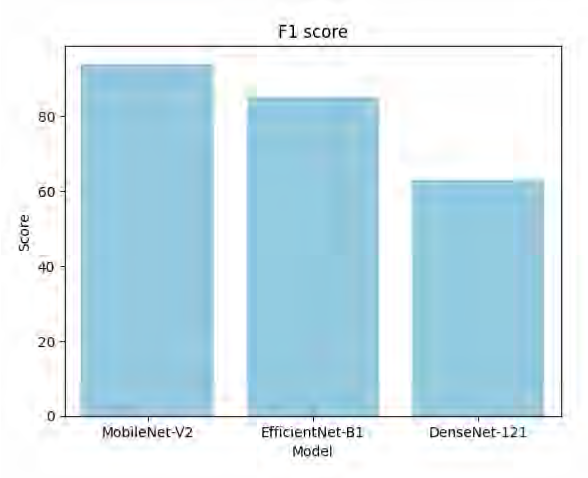


Figure 4.6: Different Model's performance in Phase 1

Therefore, from the table and barplot we can see that MobileNet-V2 clearly outperformed other two models by **achieving 94% F1 score on "RARE" class** hence it will be optimal to select the MobileNet-V2 model to complete Phase 1 supervised learning of the proposed hybrid architecture.

- **Phase 1:** In the phase 1 we trained three supervised learning models so that it captures all the unified rare class positive cases flawlessly and reroutes those cases to the Few-shot Relation Network to further classify them into their specific rare disease class. To achieve this goal, we used several pretrained models and among them MobileNet-V2 model performed better than the others. All of the three models were trained in the same setup. For all of the three models batch size was 32, total epoch was 25 and 0.001 learning rate. The best model was saved when the model scored lowest validation loss on a held out validation set. The selected pretrained MobileNet-V2 model was trained for 25 epochs. The best model was saved at epoch 19 with 8% validation loss. Training parameters like, batch size was 32 , learning rate was 0.001. For regularization purposes the dropout rate in the fully connected layer was 0.5 in order to prevent the overfitting. For all of the models, we used Adam optimizer and the loss function was cross entropy loss.

Disease Name	**Precision**	**Recall**	**F1-Score**
COPD	100%	100%	100%
Covid-19	94%	85%	89%
Emphysema	84%	96%	89%
Normal	100%	98%	99%
Pneumonia	100%	100%	100%
Tuberculosis	92%	100%	96%
RARE	100%	88%	94%

Table 4.5: Test Set Performance of Phase 1- Supervised Model

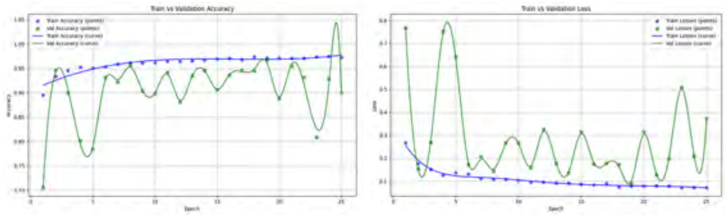


Figure 4.7: Train-Val Accuracy,Loss curve

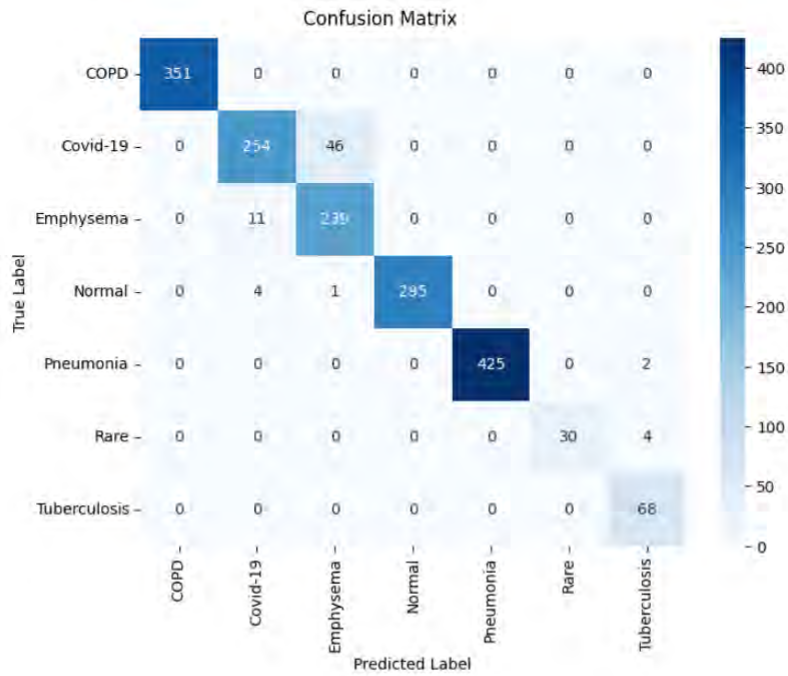


Figure 4.8: Phase 1 Confusion Matrix

4.2.2 Phase 2: Few-Shot Model

In order to select the best **Few-shot learning algorithms** we implemented two of the famous Few-shot learning methods Siamese Network and Relation Network. Between these two methods Relation Network outperformed Siamese Network in terms of overall performance. For that reason, we have selected Relation Network as the best candidate model for the Phase 2 of our proposed hybrid architecture. The below bar chart depicts the comparison between overall F1-score of Siamese Network and Relation Network. We have prioritized the F1-score so that the chosen network provides a good balance between Precision and Recall and after conducting several experiments we have seen that relation network excelled at our dataset compared to the siamese network.

- **Phase 2:** In phase 2 we trained a few shot learning model Relation Network introduced by F Sung et al. in 2017 . The training was done for a total of 100 epochs and the best model was saved at epoch 73 with 0.07 loss. Relation Network performed better to classify the rare disease images into their respective classes. Relation Network achieved overall 96% and 95% precision, recall value on Test set validation. The results indicate that the model superbly generalized itself and can identify the cases in unseen images.

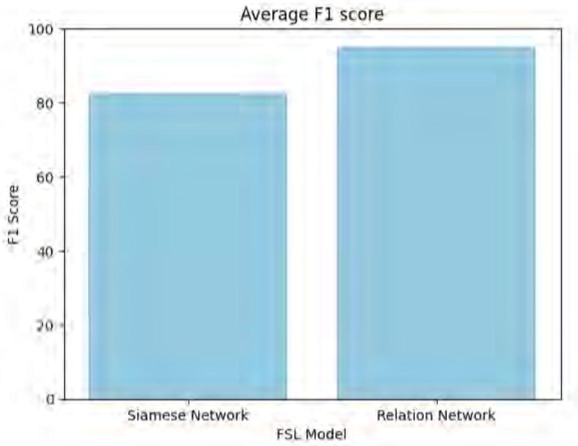


Figure 4.9: FSL Model's F1 score

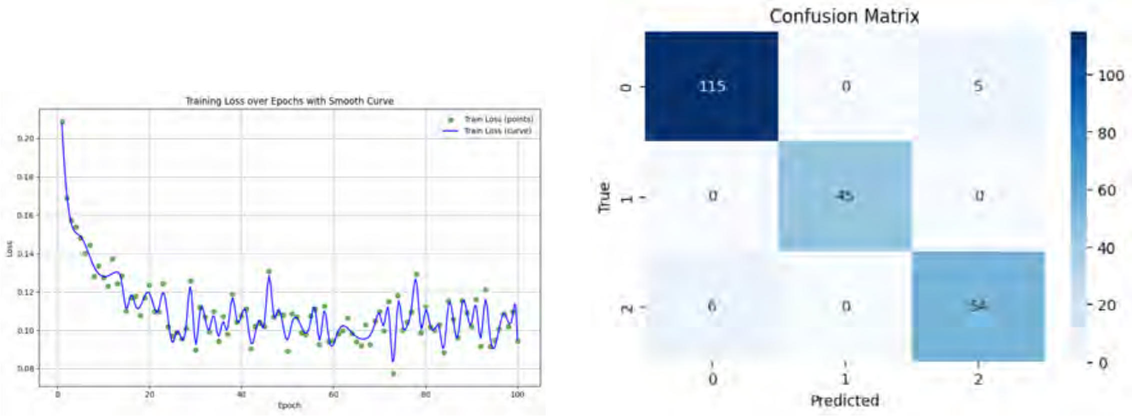
Disease Name	Precision	Recall	F1-Score
Edema	95%	96%	95%
Fibrosis	100%	100%	100%
Hernia	92%	90%	91%

Table 4.6: Test Set Performance of Phase 2- Relation Network

Disease Name	Precision	Recall	F1-Score
Edema	84%	78%	81%
Fibrosis	86%	83%	84%
Hernia	83%	82%	82%

Table 4.7: Test Set Performance of Phase 2- Siamese Network

The model was trained for a total 100 epochs and each epoch there were 50 episodes. Learning rate was 0.001 and Image size was 224*224. The training and testing of the model was done in a 3 way 5 shot manner. Which means, in each episode 3 classes were chosen with 5 randomly supported images. At the same time, from each 3 class 5 query images were randomly chosen as well. Hence, the length of the query set images was 15.



(a) Relation Network Loss Curve

(b) Relation Network Confusion Matrix

Figure 4.10: Relation Network training results: (a) loss curve and (b) confusion matrix.

4.3 Proposed Hybrid Pipeline Comparative Study and Result

4.3.1 Result

During hybrid inference any input belongs to one of the rare classes. Edema, Fibrosis, Hernia will be detected as RARE class by the phase 1 supervised model and then re-routed to Phase 2 Few-shot model Relation Network. Relation Network will then find out the specific class that input belongs to. For instance, if the input image is Hernia, phase 1 will predict the input as RARE and send it to the Relation Network to classify it as Hernia. The performance of our Hybrid architecture is shown in the table below.

Disease Name	**Precision**	**Recall**	**F1-Score**
COPD	100%	100%	100%
Covid-19	96%	87%	91%
Edema	90%	100%	94.7%
Emphysema	88%	96%	92%
Fibrosis	88.9%	88.9%	88.9%
Hernia	88.9%	88.9%	88.9%
Normal	99%	97%	98%
Pneumonia	98%	100%	99%
Tuberculosis	82%	100%	90%

Table 4.8: Precision,Recall and F1-Score of the proposed Hybrid Architecture

From the table above, it is clear that our hybrid architecture has significantly improved the performance of the three rare classes without degrading the performance of other common classes. If we observe the recall value of Edema we can see that it achieved 100% score with 90% precision which means the proposed model successfully captured all the positive cases of Edema with high precision. The same can be said with other two rare classes that achieved close to 90% recall and precision value.

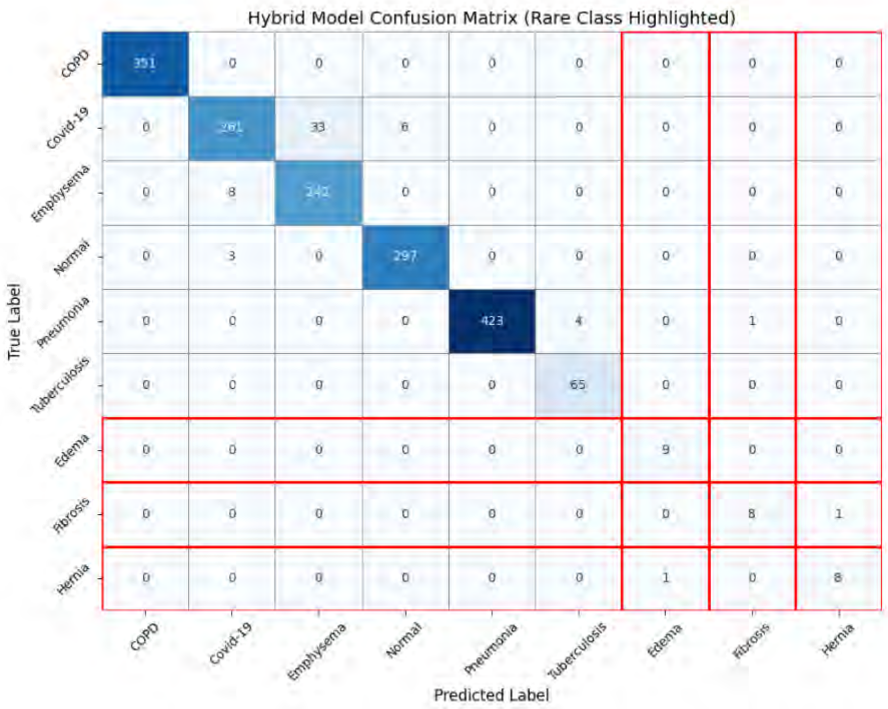


Figure 4.11: Confusion Matrix of the proposed Hybrid Architecture

4.3.2 Comparative Study with Plain Supervised Pre-trained CNN Model's

Previously we have seen that pre-trained models failed to classify the rare class images with high precision and recall. Let us perform a side by side comparison to see how much improvement the proposed hybrid architecture did classifying the rare class images. From the below side by side accuracy bar chart it can be shown that the three **rare classes Edema, Fibrosis, Hernia** performed poorly in pre-trained CNN model but there is a significant rise of accuracy in the proposed Hybrid model which indicates a good performance of the proposed model compared to the state of the art pre-trained CNN models.

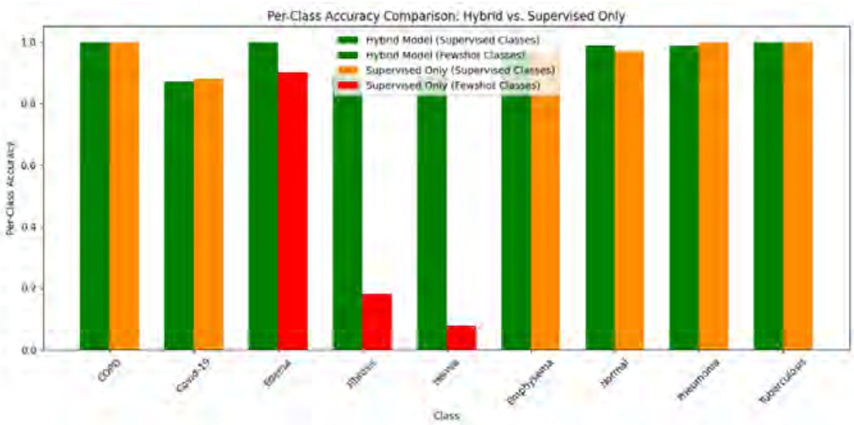


Figure 4.12: Per-Class Accuracy Comparison Hybrid vs Supervised Pre-trained CNN Models

Rare Disease Class	**MobileNet -V2**	**Efficient Net-B1**	**DenseNet -121**	**Our Model**
Edema	64%	57%	75%	94.7%
Fibrosis	31%	79%	0%	88.9%
Hernia	14%	67%	25%	88.9%
Average Score	**36.33%**	**68.33%**	**33.33%**	**90.83%**

Table 4.9: F1 Score across different Pre-trained CNN Models vs Proposed Hybrid Architecture

From the table 4.8 it can be inferred that our proposed hybrid architecture achieved an **average of 90% F1 score on the rare classes**. In contrast, all 3 SOTA pre-trained CNN models performed poorly with an **average of 45.98% F1 score on the rare classes**.

From the below two bar charts which represents the comparison of Recall and F1 Score of the Hybrid pipeline and plain supervised model across the all classes we can observe that significant improvement in both recall and F1 score value in the proposed hybrid pipeline.

We can see that among the three rare classes Fibrosis and Hernia performed awfully in terms of recall value whereas in the proposed hybrid architecture the recall value of Fibrosis and Hernia rose exceptionally high which indicates that compared to

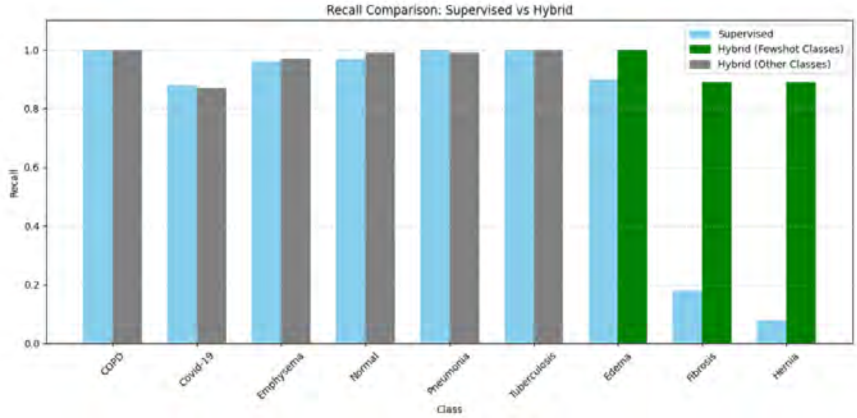


Figure 4.13: Recall score Comparison Supervised vs Proposed Hybrid Architecture

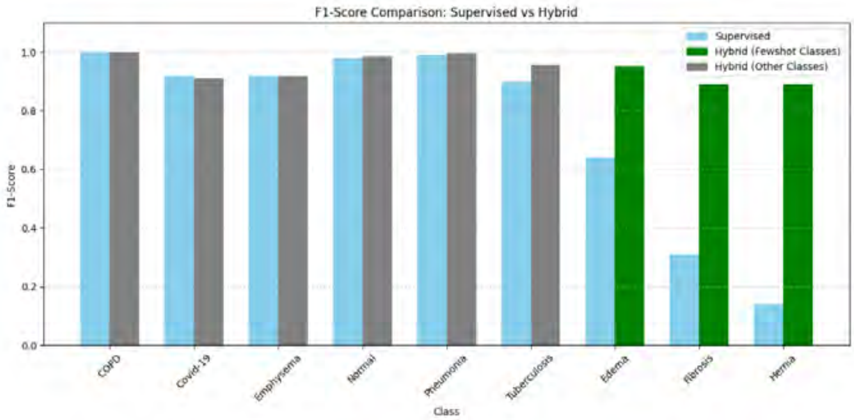


Figure 4.14: F1 score Comparison Supervised vs Proposed Hybrid Architecture

the plain supervised pretrained models proposed hybrid architecture successfully captured all the positive cases of Fibrosis and Hernia. On the other hand, even though in terms of supervised models, Edema has a good height in the recall bar chart but its height is significantly low in the F1-score bar chart slightly above 50% which indicates that the precision of Edema was low.Slightly above 50% F1 score means the performance of the supervised model in this class cannot be trusted because there must be so many negative cases of Edema hence it is important to have a good balance of Recall and Precision and high value of F1-score indicates such balance. Lastly, if we observe the bar chart height of other two rare class Fibrosis and Hernia in F1-score bar chart it is clearly understandable that not only the recall but also the precision value of the proposed hybrid model significantly improved compared to the pre-trained supervised model because the F1-score of these two classes in the supervised model was extremely low which clearly reflects in the bar chart as well. Therefore, we can conclude that a supervised model will suffer to identify the classes with less amount of data and will always excel at classifying categories with vast amounts of data. Hence, it is feasible to use a specialized architecture to categorize the classes with scarce data. And our experiment and results of the proposed hybrid pipeline supports the above mentioned approach.

Chapter 5

Conclusion

5.1 Conclusion

In the realm of medical diagnosis leveraging the most of Artificial Intelligence will be a game changing factor. But there are some situations where AI might deteriorate its effectiveness and its solution if not tackled properly and one of those problems is that most of the CXR image dataset are highly imbalanced and in image classification tasks AI becomes ineffective to determine class instances with less amount of data. Therefore, in the real world to overcome the problem of data scarcity one of the popular solutions is to collect more data but in some cases like for rare diseases collecting more data is not possible. For that reason, we have been motivated to study a solution that can be effective and robust without collecting more data instead it can leverage the most of the data already has. This was the motivation that led us to propose a hybrid architecture that unifies the two strong learning methods, supervised and Few shot learning. Since most of the CXR image dataset has both scenarios of having more images in common diseases and fewer examples in rare diseases, that is why using a pre-trained model only for supervised learning will not give expected results. For that reason, we combined the two learning methods that excel in those two scenarios. Supervised learning excels where vast amounts of data is available and Few-shot learning performs better than supervised when data is scarce. Since it is most common in the medical world that collecting more data is sometimes extremely challenging, this kind of architectural approach is more impactful and can achieve good results. For that reason, our work can be used as reference for future work because in our study we have shown that combining the two methods can produce a better result instead of relying on one.

5.2 Limitations

During the models training validation set loss and accuracy curve was not smooth even though models achieved a satisfactory loss and accuracy score. Possible reason for this unstable curve can be our dataset consisted of different distributions such as hospitals,devices etc. Another reason can be the high noise in the images which we tried to reduce by applying CLAHE. For deep neural network the size of the dataset is also important even though we are using transfer learning but we have allowed all the gradients of each layer to be recalculated during backpropagation and since the size of our dataset is not big enough the unstable curve indicates the limitation of the model failing to generalize smoothly.

5.3 Future Work

While detecting a disease with the help of an Artificial Intelligence assistant it is important to find the root cause of AI's decision to ensure reliability and transparency otherwise using AI for fast diagnosis will still not be normalized because of the sensitive nature of the medical diagnosis realm. Therefore, it is important for AI models to explain their decision to its users. This is why Explainable AI was introduced so that models can explain its output. Therefore in future we want to integrate the functionality of explainable AI in our proposed model which will shape our model to be much more reliable and trustworthy

Bibliography

[1] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191

[2] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015. DOI: 10.1126/science.aab3050

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539

[4] H. C. Shin, H. R. Roth, and M. Gao, "Deep convolutional neural networks for computer-aided detection: Caffe," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016. DOI: 10.1109/TMI.2016.2552205

[5] O. Vinyals, C. Blundell, and T. Lillicrap, "Matching networks for one-shot learning," in *Advances in Neural Information Processing Systems*, vol. 29, 2016, pp. 3630–3638.

[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[7] P. Rajpurkar, J. Irvin, and R. Ball, "Deep learning for chest x-ray analysis: A survey," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 12, pp. 2862–2873, 2017. DOI: 10.1109/TBME.2017.2702880

[8] P. Rajpurkar et al., "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[9] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017.

[10] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," *CoRR*, 2017.

[11] D. S. Kermany, K. Zhang, and M. Goldbaum, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, 1122–1131.e9, 2018. DOI: 10.1016/j.cell.2018.02.010

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *arXiv preprint arXiv:1801.04381*, 2018.

[13] J. Chen, S. Zhang, and Z. Liu, "Exploring the impact of data scarcity on deep learning models for rare disease classification," *Journal of Medical Imaging*, vol. 6, no. 2, p. 010 501, 2019. DOI: 10.1117/1.JMI.6.2.010501

[14] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, 2019. DOI: 10.1186/s40537-019-0192-5

[15] J. Johnson and T. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, 2019. DOI: 10.1186/s40537-019-0192-5

[16] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[17] Q. Yang and Y. Liu, "Federated learning: Challenges, methods, and future directions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 8, pp. 1337–1350, 2019. DOI: 10.1109/TKDE.2018.2822270

[18] A. Paul, Y.-X. Tang, and R. M. Summers, "Fast few-shot transfer learning for disease identification from chest x-ray images using autoencoder ensemble," in *Proceedings of the SPIE Medical Imaging*, vol. 11314, 2020, p. 1 131 407. DOI: 10.1117/12.2549060

[19] W. Qu, I. Balki, M. Mendez, et al., "Assessing and mitigating the effects of class imbalance in machine learning with application to x-ray imaging," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 2041–2048, 2020. DOI: 10.1007/s11548-020-02260-6

[20] S. Ren, K. He, and J. Sun, "Deep residual learning for image recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 22–31, 2020. DOI: 10.1109/TPAMI.2016.2528162

[21] Y. Xie, Y. Liu, and Y. Zhang, "Hybrid learning for rare disease classification in medical x-rays," *Journal of Artificial Intelligence in Medicine*, vol. 115, p. 101 025, 2020. DOI: 10.1016/j.artmed.2020.101025

[22] A. Paul, Y.-X. Tang, T. C. Shen, and R. M. Summers, "Discriminative ensemble learning for few-shot chest x-ray diagnosis," *Medical Image Analysis*, vol. 68, 2021. DOI: 10.1016/j.media.2020.101911

[23] A. R. Bhimireddy, J. L. Burns, S. Purkayastha, and J. W. Gichoya, *Few-shot transfer learning to improve chest x-ray pathology detection using limited triplets*, 2022. DOI: 10.48550/arXiv.2204.07824 arXiv: 2204.07824 [cs.CV].

[24] D. Cores, N. Vila-Blanco, M. Pérez-Alarcón, A. Martínez-de-Alegría, M. Mucientes, and M. J. Carreira, "A few-shot approach for covid-19 screening in standard and portable chest x-ray images," *Scientific Reports*, 2022. DOI: 10.1038/s41598-022-25754-6

[25] H. Liz, J. Huertas-Tato, M. Sánchez-Montañés, J. Del Ser, and D. Camacho, *Deep learning for understanding multilabel imbalanced chest x-ray datasets*, 2022. arXiv: 2207.14408 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2207.14408

[26] C. Mosquera, L. Ferrer, D. Milone, D. Luna, and E. Ferrante, *Impact of class imbalance on chest x-ray classifiers: Towards better evaluation practices for discrimination and calibration performance*, 2022. arXiv: 2202.12843 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2202.12843

[27] A. Galán-Cuenca, A. J. Gallego, M. Saval-Calvo, and A. Pertusa, *Few-shot learning for covid-19 chest x-ray classification with imbalanced data: An inter vs. intra domain study*, 2024. DOI: 10.48550/arXiv.2401.10129 arXiv: 2401.10129 [cs.CV].

[28] C. J. Hellín, A. A. Olmedo, A. Valledor, J. Gómez, M. López-Benítez, and A. Tayebi, "Unraveling the impact of class imbalance on deep-learning models for medical image classification," *Applied Sciences*, vol. 14, no. 8, 2024. DOI: 10.3390/app14083419

[29] A. A. G. Pramana, F. I. Permana, M. F. Maulana, and D. R. Fudholi, *Few-shot learning approach on tuberculosis classification based on chest x-ray images*, 2024. DOI: 10.48550/arXiv.2409.11644 arXiv: 2409.11644 `[eess.IV]`.

[30] A. A. G. Y. Pramana, F. I. Permana, M. F. Maulana, and D. R. Fudholi, *Few-shot learning approach on tuberculosis classification based on chest x-ray images*, 2024. DOI: 10.48550/arXiv.2409.11644 arXiv: 2409.11644 `[eess.IV]`.

[31] X. Wu et al., "Chexnet: Combining transformer and cnn for thorax disease diagnosis from chest x-ray images,"