EARLY LUNG CANCER DIAGNOSIS USING ARTIFICIAL NEURAL NETWORKS

Name: Ibrahim Irfan
Supervisor: Mr. Waseem Ahmed
Course: Independent Research Project
Submission Date: 18th December, 2024

Acknowledgements

I would like to thank my supervisor, Mr. Waseem Ahmed, in assisting me and helping me fulfill the requirements of this project and for being cooperative with me during the course of this project. Mr. Waseem Ahmed was very helpful and was a good mentor to me throughout this independent research project.

Executive Summary

This report documents the procedure of the study conducted on the use of Artificial Neural Networks for Early Lung Cancer Diagnosis. The aim of this study was to develop a robust deep learning model that detects and classifies lung nodules to assist medical experts in early lung cancer diagnosis. What motivated the author to pursue this study is the latest advancements in Artificial Intelligence for healthcare and the ability to leverage artificial neural networks to improve the accuracy and efficiency of early lung cancer detection, ultimately reducing mortality rates through timely diagnosis. The first chapter introduces the topic of the study, explains the impact of lung cancer on mortality worldwide, discusses how AI addresses the challenges of lung cancer diagnosis and what makes AI and deep learning a crucial method for diagnosing lung cancer. The first chapter also addresses an overview of the use of ANNs in diagnosing lung cancer at an early stage, the research aims and objectives as well as the research questions. The second chapter discusses a thorough literature review of existing work related to early lung cancer diagnosis and discusses the strengths and limitations of the existing models. The third chapter discusses the research philosophy and methodologies used for the study. The research philosophy and methods used are pragmatism and quantitative approaches. The methods used for the practical implementation are transfer learning, convolutional neural networks and hybrid approaches. The third chapter also includes a step-by-step discussion of the project framework, implementation and computational approach, the computational resources required for the study and justifications for the decisions made. The third chapter concludes with the ethical and professional issues relevant to the study. The ethical and professional issues are data privacy and transparency, algorithm bias and dataset imbalance. The fourth chapter analyses the results of the practical implementation conducted throughout the study and the accuracies yielded by the training done on the ANN models. The fourth chapter also includes the testing done on the test sets of the datasets, the predictions that were made on the input CT scans of the test sets. The fourth chapter concludes with a comparison of the proposed models with the existing models discussed in the literature review and an analysis of how the proposed models fare with the existing models. The fifth and final chapter discusses the concluding remarks of the study, the results that were yielded throughout the practical implementation and model training. The fifth chapter also maps the work done in the study to the research questions in the first chapter. The fifth chapter concludes with a set of outcomes of the project, the limitations of the study and the future iterations of this project. The overall project report concludes with a glossary of terms and definitions that were used throughout the study as well as an appendix of the figures and tables that were presented throughout the report.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Lung Cancer has been the most harmful and dangerous form of cancer which has often proven difficult to diagnose. Among all cancer-related deaths, lung cancer is the most frequent cause. A total of 1-8 million are diagnosed every year and 1-6 million die every year **[1]**. As of 2020, the net five-year survival rate is 13.8%. 75% of lung cancer cases are only discovered in the advanced due to the absence of symptoms in the early stages **[2]**. In developing countries, where the death rate due to lung cancer is at its highest, CT Screening has become a method to cure the disease by routinely detecting potentially malignant lung nodules **[3]**. In 2020, WHO recorded 10 million deaths and 19 million new cases of lung cancer **[4]**. In 2018, lung cancer recorded the highest percentage of new cases and deaths worldwide at 11.6%. The type of cancer that affects 85% of lung cancer patients is non-small lung cancer **[5]**. Research conducted by the World Health Organization says that almost 7.6 million mortality happens worldwide annually due to lung cancer **[6]**. The American Cancer Society stated that lung cancer is the main cause of cause of death amongst all cancer disease in the USA **[7]**. Lung cancer comes fourth for the highest causes of death in the Mediterranean and third in the Middle East. The most recognized reason behind the diagnosis of lung cancer is heavy smoking. Other reasons include air pollution and unhealthy diet **[8].** In 2020, lung cancer became the first dangerous outline of cancer **[9].** From 1930 to 2019, lung cancer has risen as the main cause of death among males **[10].** The International Association of Cancer Society reported that in 2020, the estimated number of new lung cancer cases were 235,760 and the approximate deaths from lung cancer were 131, 880 **[11].** The two main types of lung cancers are small-Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC). The main reasons for lung cancer diagnosis include smoking, air pollution, and aging **[12].**

Artificial Intelligence is a field that is revolutionizing lung cancer diagnosis by addressing key challenges in early detection, accuracy, and efficiency. Deep Learning is a sub-field of AI that addresses these challenges by analysing CT scans, chest X-rays, and other imaging modalities to detect early-stage nodules with higher acurracy and precision. Artificial Neural Networks are crucial in detecting lung cancer because they often work well with processing images. Two common deep learning approaches that will be experimented with in this study are transfer learning and hybrid models.

1.1 Overview

The aim of the study is to develop a robust system for early lung cancer detection using artificial neural networks (ANNs). ANNs are often used in medical diagnosis because they work well with processing images and processing medical images such as CT scans is a very crucial to disease diagnosis. By analyzing medical imaging data, such as CT scans, and other relevant clinical information, the system will identify potential lung nodules and classify them as benign or malignant. This early detection could significantly improve patient outcomes by enabling timely medical intervention.

1.2 Objectives

- Gather a large and diverse dataset of medical images.
- Develop methods for the ANN to extract relevant features from the medical images.

- Use high-quality and annotated data combined with data augmentation, hybrid approaches and transfer learning to develop the model.
- Performance Evaluation: Ensure that the model achieves a minimum accuracy of 90%.
- Compare the performance of the ANN-based diagnostic system with existing models
- Document the research process, findings, and insights gained.

1.3 Research Questions

Q1. Which ANN architectures (e.g., CNNs, transfer learning, or hybrid models) are best suited for early lung cancer diagnosis, and how can they be optimized for performance?

Q2. How can transfer learning and pretrained ANN models be leveraged to improve diagnostic performance on lung cancer datasets with limited labeled samples?

Chapter 2: Literature Review

2.1 Foundational Benchmarks in Lung Imaging Challenges

In 2011, the **LOLA11** challenge took place. The goal was to evaluate and improve algorithms designed for medical image segmentation, particularly for dividing the lungs into lobes, which can be a difficult task due to abnormalities or the complexity of the lung structure **[13]**.

The LOLA11 challenge (2011) laid the groundwork for lung lobe segmentation by providing a standardized dataset for comparing segmentation algorithms. Despite its pioneering role, its limited dataset and reliance on semi-automatic techniques constrained its relevance for contemporary deep learning applications. This limits the capacity to train and test contemporary deep learning models, which require large-scale data for generalization. Overall, The LOLA11 challenge was among the early benchmarks for lung lesion segmentation, setting the stage for subsequent challenges such as LUNA16. The study could improve on expanding the datasets and revisiting the challenge with updated guidelines and incorporating deep learning-based approaches could yield significant performance improvements.

Building on the results of the LOLA11 challenge, **the LUNA16 challenge (2016)** significantly advanced lung nodule detection by incorporating modern deep learning techniques and larger datasets. It used a dataset from the LIDC-IDRI (Lung Image Database Consortium) containing 888 CT scans, where participants developed and evaluated Computer-Aided Detection (CAD) systems **[14]**.

By providing a standardized benchmark, LUNA16 has driven innovation in deep learning (DL) and traditional approaches for lung nodule detection. However, the absence of a gold standard diagnosis and limited clinical utility marked areas for improvement. The study could further be improved by expanding the dataset further and eliminating bias from the dataset. Incorporating clinically meaningful metrics, such as nodule malignancy risk scores or cost-effectiveness analysis, could bridge the gap between technical performance and clinical impact. Together, LOLA11 and LUNA16 challenges underscored the importance of standardized benchmarks, spurring innovation in lung imaging.

The **2017 Data Science Bowl** further bridged the gap between technical development and real-world clinical applications, not addressed in the LOLA11 and LUNA16 challenges, by focusing on lung cancer detection. **[15]**.

The use of 3D imaging catalyzed innovation, particularly in volumetric analysis with deep learning. However, ethical concerns regarding patient consent and dataset diversity highlighted the need for greater scrutiny in dataset design and competition frameworks. Overall, the focus on 3D imaging and early cancer diagnosis catalyzed innovation and inspired further research in medical imaging. However, limitations in dataset diversity, task scope, and clinical integration highlight areas for improvement. Future iterations that address these gaps could drive AI solutions toward broader adoption in clinical practice.

Collectively, the LOLA11 and LUNA16 Challenges as well as the 2017 Data Science Bowl not only benchmarked technical progress but also emphasized the importance of clinical relevance and ethical considerations in lung imaging research.

2.2 Machine Learning Approaches for Lung Cancer Detection

With the LOLA11 challenge, the LUNA16 challenge and the 2017 Data Science Bowl setting the foundation for Lung Cancer Diagnosis using AI, various studies explored Machine Learning approaches for lung cancer diagnosis.

Boban et al. [16] used similar approaches to analyze 400 lung cancer videos with the use of MLP, KNN and SVM algorithms with a focus on CT scan images with KNN achieving the highest accuracy of 99.2%. The study evaluates multiple ML models, including Decision Trees, Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and ensemble methods like Random Forests. This breadth allows a comprehensive comparison of algorithmic strengths and weaknesses for lung disease classification.

The paper does not discuss ethical issues related to handling sensitive patient data, which is critical for medical applications. The paper provides minimal details about the dataset used, including its source, size, and diversity. While it offers a broad overview of algorithmic performance, its reliance on classical techniques and lack of detail regarding datasets, preprocessing, and validation limit its impact.

Advanced ML techniques were utilized to detect lung cancer. **Banerjee et al. [17]** proposed an infrastructure for tumor detection using ANNs, SVMs and Random Forest algorithms. The study revealed that ANNs outperformed SVM and Random Forest in terms of Both area and texture-dependent characteristics were considered in the analysis with 92% accuracy. The achieved accuracy for texture-based features was 70% for Random Forest, 80% for SVM, and 96% for ANN.

In the study, the author evaluates multiple ML algorithms, providing a broader perspective on their applicability and effectiveness for lung cancer prediction. The author also acknowledges the importance of evaluating accuracy, sensitivity, and specificity, which are crucial for assessing the performance of ML models in medical diagnostics. The paper contains limitations like lacking a detailed explanation of preprocessing steps, such as feature extraction, normalization, and handling of missing data, which are critical in medical datasets and no discussion of how the models were tuned, including hyperparameter optimization, which significantly impacts ML model performance. The paper does not address the potential for biases in the dataset or models. Overall, while it highlights the potential of such methods, its lack of methodological depth, dataset transparency, and rigorous evaluation limits its impact.

While both research papers, **[16-17]**, utilized Machine Learning approaches to enhance lung cancer diagnosis, they lacked details about dataset diversity and ethical considerations.

2.3 Innovations in Deep Learning Architectures

Deep Learning has emerged as the frontrunner for disease diagnosis. **Yahya et al. and Chin et al. [18]** proposed hybrid models combining Convolutional Neural Networks (CNNs) and SVMs, leveraging CNNs for feature extraction and SVMs for classification. The study conducted achieved an accuracy of 96.71% on the training set and 97.91% on the testing set. The authors used CT images from the Chest CT-Scan Images Dataset. In the experiment, the dataset was divided into 20% for testing and the remaining 80% for training.

These hybrid approaches demonstrated superior performance but often lacked methodological transparency and robust validation techniques. The research has certain limitations such as not providing detailed information about the dataset used. The paper does not mention whether cross-validation or external validation was performed, which is essential to ensure the robustness and generalizability of the results. Future work should use well-documented, publicly available datasets (e.g., LIDC-IDRI) and provide detailed dataset descriptions. The paper proposes an intriguing hybrid CNN-SVM approach for lung cancer classification but lacks the methodological transparency, evaluation rigor, and real-world considerations needed for substantial impact.

Similarly, **Yu et al. [19]** proposed a Deep Learning system that predicts epidermal growth factor receptor mutations that are associated with lung cancer disease. The system was tested on two datasets and got an accuracy 76.16% and 67.55%.

The paper focuses on using CNNs, a proven technology for image analysis, to address molecular profiling in NSCLC, showcasing a novel application in the field of computational pathology. This aligns with the growing trend of using AI to integrate imaging and molecular data. The limitations of this study are that the methodological section appears to lack sufficient detail about data preprocessing, architecture design, and training process. For reproducibility, clear descriptions of these aspects are essential. The performance metrics, such as accuracy, sensitivity, or specificity, were not discussed in depth. Omitting comprehensive evaluation weakens the reliability of the reported results and the comparison with baseline models or existing methods is limited or absent. The paper could be improved by collaborating with hospitals or institutions to access larger datasets, which would improve model training and testing and providing architectural details, hyperparameters, training epochs, and augmentation techniques to enhance reproducibility.

Da Silva et al. [20] researched and tested the functionality of Deep Learning models for early lung cancer diagnosis through the LIDC database. The results were promising with the model achieving an accuracy of 79.40%. The researcher used Deep Belief Networks and Stacked Denoising Autoencoder using the same datasets and achieved even higher accuracies at 81.19% and 79.29%.

The paper has a focused objective and addresses an important clinical challenge: the differentiation of malignant and benign lung nodules. This application is crucial in early lung cancer detection and management. The paper however lacks methodological detail and the author doesn't specify whether they used any strategies to mitigate overfitting, such as dropout or regularization. There is no clear comparison with baseline or traditional machine learning methods, which would contextualize the performance of the proposed CNN. Areas for improvement include addressing the limitations of dataset size and diversity, improving methodological transparency, and exploring integration into clinical workflows.

Mohite et a. [21] used transfer learning to create an ANN-based model to detect lung cancer using CT scan images. The VGG19 architecture was used on the IQ-OTH/NCCD dataset and the model achieved a training accuracy of 90%, a validation accuracy of 93% and a test accuracy of 88%

The research sets out a clear objective: to detect and classify lung cancer using transfer learning. Such a focus ensures the research remains specific and actionable. However, the author doesn't does not compare the proposed model with existing approaches or state-of-the-art techniques. A comparative analysis would add credibility and context to the findings. The methodology lacks depth and preprocessing steps, such as resizing images, normalization, or augmentation, are not thoroughly discussed, which are critical in medical imaging tasks. Overall, the paper addresses a critical area in healthcare and demonstrates the potential of transfer learning for lung cancer detection however there is room for improvement in dataset transparency, methodology, and evaluation metrics and addressing these could make the study more beneficial.

The research conducted in **[18-21]** expanded the scope of AI applications beyond imaging to molecular profiling and utilized transfer learning as well as reinforcement learning for nodule-localization. However, challenges related to dataset size, overfitting mitigation, and clinical integration persisted.

2.4 Real-World Applications and Ethical Considerations

In order to address clinical relevance and ethical considerations which weren't discussed in the literature above, **Al-Yasriy et al. and Al-Husieny et al. [22]** developed a model using a Convolutional Neural Network technique with the AlexNet architecture to detect lung cancer. The researchers developed a CNN-based model to classify lung cancer cases as normal, benign or malignant with high accuracy through data collected by Iraqi hospitals**.** The trained model gave an accuracy of 93.548% on the 86th epoch out of a 100. The precision is 97.1015%, the sensitivity is up to 95.714% and lastly the specificity reaches 95%.

The paper addresses a critical challenge in oncology: the early and accurate diagnosis of lung cancer. This aligns with global efforts to reduce mortality through timely interventions, making it clinically relevant and potentially impactful in practical applications. The drawbacks of this study are that it lacks methodological transparency where the CNN architecture is not described in detail. There is no information on how hyperparameters (e.g., learning rate, batch size) were chosen or optimized. It lacks comparisons with existing state-of-the-art methods (e.g., classical machine learning models or other deep learning approaches) limits the ability to contextualize the proposed method's performance. There is room for improvement like describing the CNN architecture and training process comprehensively, using robust validation techniques, such as k-fold cross-validation, and test the model on external datasets to ensure generalizability and incorporate visualization techniques to explain how the CNN makes decisions. This will help build trust among clinicians and radiologists.

Similarly, **Ardila et a. and Kiraly et al. [23]** proposed a 3D CNN for Lung Cancer Screening. The. model achieved an accuracy of 94.4%. The researchers chose to learn new features using Deep Convolutional Neural Networks because they were superior to hand-engineered features. A Deep Learning model was developed using the NLST dataset, which was broken into 70% for training, 15% for tuning and another 15% for testing. During clinical evaluation, the model achieved better sensitivity and specificity than the average radiologist.

The researchers emphasized the limitations of initial lung cancer screening using AI, where there is relatively lack of cancer outcome information. The research had certain weaknesses such as over-reliance on retrospective data, where it uses retrospective data, which may not fully represent the complexities of prospective clinical implementation, such as variations in imaging quality and patient positioning. Despite its performance, the model still produces false positives and negatives, which could lead to unnecessary procedures or missed diagnoses. The clinical implications of these errors are not thoroughly explored. The study could be further improved by validate the model on datasets from diverse populations and imaging protocols to ensure broader applicability, conducting pilot studies in real-world clinical settings to evaluate the system’s impact on radiologists’ workflows, diagnostic accuracy, and patient outcomes and lastly incorporating visualization tools to make the model’s decision-making process more transparent for clinicians.

Masood et al. [24] proposed a 3D Deep CNN for detecting lung cancer using Chest CT scans. The process included training the data, using data augmentation, data pre-processing, and the training process. Like other proposed models, this model was created using the LIDC-IDRI and LUNA16 dataset. Experimentations showed the result is that the model can detect lung cancer at 98.5% accuracy.

The proposed system integrates key steps from image pre-processing to final diagnosis, offering a holistic solution rather than a single-stage model. The study has certain weaknesses where the author doesn’t address the data privacy and security challenges with regards to the cloud-based system. The paper doesn’t evaluate the computational costs or latency, which could affect usability in high-demand scenarios. The proposed model doesn’t provide any information into its decision-making process. Overall, the research is a forward-looking contribution to AI-driven healthcare, addressing the dual challenges of scalability and accessibility in lung cancer diagnosis however it can be improved by a deeper discussion of practical deployment challenges, such as data privacy, system integration, and interpretability.

The research in **[22-24]** demonstrated the potential of AI to match or exceed radiologists' performance. However, they also revealed critical gaps in addressing false positives and negatives, patient privacy, and clinical usability.

2.5 Summary

In summary, the above literature collectively illustrated the evolution of AI applications in lung cancer detection, from early segmentation challenges to sophisticated hybrid and deep learning models. While technical advancements have been significant, recurring limitations such as dataset transparency, methodological rigor, and ethical considerations underscore the need for more holistic research. Most of the studies achieved an accuracy of 90% and utilized the LIDC-IDR dataset. However, they didn't address the ethical considerations with regards to the use of AI in lung cancer diagnosis. The research gaps that were not explored in previous literature that this study intends to cover are:

- Integrating multiple data types (e.g., imaging, clinical data) into a unified neural network for diagnosis and prognosis.
- Addressing data scarcity through techniques like data augmentation, transfer learning, or synthetic data generation using generative models. High-quality, annotated datasets are limited in the field of lung cancer diagnosis, especially when it comes to rare subtypes of lung cancer or early-stage nodules.
- While some efforts are focused on early detection, there is still room for improving ANN models that can detect early-stage lung cancer.
- Address the ethical considerations of using AI for lung cancer diagnosis.

Chapter 3: Research Philosophy, Methodologies and Computational Approach

3.1 Research Philosophy and Methodology

The philosophical approach used for this study is Pragmatism. Pragmatism focuses on the practical application of ideas and the outcomes of actions **[25]**. The research methodology adopted for this project is the quantitative approach because it provides a structured framework for objectively analyzing and validating the performance of the system **[26]**. This alignment arises from the need to work with large-scale data, evaluate models using precise metrics, and ensure reliability in clinical applications.

Pragmatism ensures that the research remains focused on practical outcomes, such as improving diagnostic accuracy and integrating ANN models into clinical workflows. The pragmatic approach focuses on not just the technical performance of the models but also addressing challenges in data quality, model interpretability, and clinical acceptance. This aligns well with the quantitative methodology by emphasizing actionable insights, iterative improvement, and real-world impact—qualities that are essential for advancing early lung cancer diagnosis using ANNs.

3.2 Methodologies for Practical Implementation

3.2.1 Transfer Learning

Transfer Learning is a method which involves utilizing models that were trained on existing studies when conducting research on another related study. The method includes using pre-trained models for feature extraction to be integrated in completely new models **[27].** In this study, the Xception architecture will be one of the methods implemented. **Xception** (Extreme Inception) is a deep convolutional neural network architecture introduced by François Chollet in 2017.

3.2.2 Convolutional Neural Networks

Convolutional Neural Networks are multi-layered feedforward biologically influenced neural networks **[28-30].** CNNs are used to reduce images into a form that is relatively easier to process and maintains the crucial features that are necessary in order to get a good prediction **[31].** A traditional Convolutional Neural Network consists of three layers: convolutional layer (CONV), max-pooling layer (POOLING) and the fully connected layer (FC). In **[32],** the CNN is thoroughly discussed where the Convolutional Layer reads the input feature maps from the last layer and converts these inputs into a set of output feature maps. In any CNN, the CONV layer is followed by the POOLING layer, that subsamples the features of the CONV layer. Last but not least, the FC (Fully Connected layer) generates each class implied at the input data.

3.2.3 U-NET Architecture

The U-NET architecture is an encode-decoder deep neural networking model which is often used in medical imaging. It was first used in biomedical image segmentation. U-net contained three main blocks, down-sampling, up-sampling, and concatenation. Previous studies called it the encode-decoder architecture **[33].** Its main goal is to predict pixel-wise classification and identify regions of interest within images, such as tumors or other structures. The name "U-Net" comes from its distinctive **U-shaped structure**.

The **U-Net architecture** is particularly effective for lung cancer diagnosis, especially in tasks like the segmentation of lung nodules in CT scans. U-Net's ability to handle high-resolution, complex medical images**,** combined with its focus on precision segmentation, makes it a strong candidate for aiding in **lung cancer diagnosis**. Its architecture allows for detailed nodule localization**,** minimizes **false positives**, and integrates well into **automated diagnostic workflows**. This makes U-Net a valuable tool for **early detection**, **monitoring**, and **treatment planning** for lung cancer.

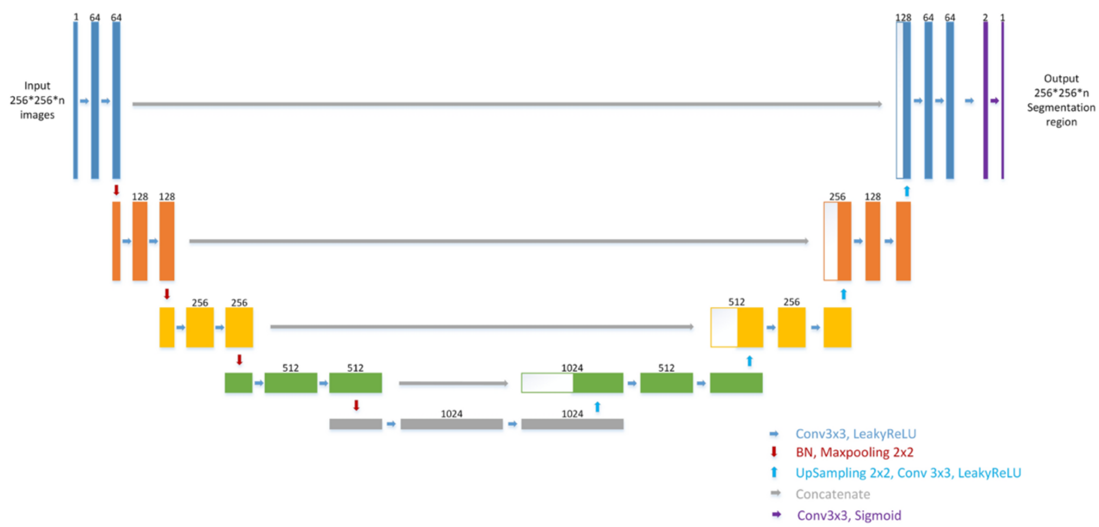

Figure 1: U-NET Architecture

3.2.4 DenseNet Architecture

DenseNet, short for *Dense Convolutional Network*, is a deep learning architecture designed to improve information flow and parameter efficiency by using dense connectivity among layers **[34]**. Unlike traditional CNNs, where each layer connects only to the next one, DenseNet connects each layer to all previous layers. DenseNet is frequently adapted in healthcare, especially in lung nodule detection. Studies like **DeepLung** combine DenseNet with region-based approaches to handle high-resolution CT scans for lung nodule detection and classification. The applications of DenseNet in Early Lung Cancer Diagnosis are:

- **Nodule Detection:** DenseNet is used to detect nodules in CT images by classifying regions of interest (ROIs) as either "nodule" or "non-nodule."

- **Nodule Classification:** Beyond detection, DenseNet architectures are also applied to classify nodules by type, which helps radiologists assess whether nodules are likely to be malignant or benign.

3.2.5 CNN-RF Architecture

The CNN-RF (Convolutional Neural Network - Random Forest) architecture is a hybrid model that combines the feature extraction power of CNNs with the classification capabilities of Random Forest (RF) algorithms **[35]**. This architecture is particularly useful in medical imaging and diagnostic tasks, such as lung nodule detection and classification, where it can leverage both deep learning's ability to extract complex features and machine learning's effectiveness in handling structured data for classification.

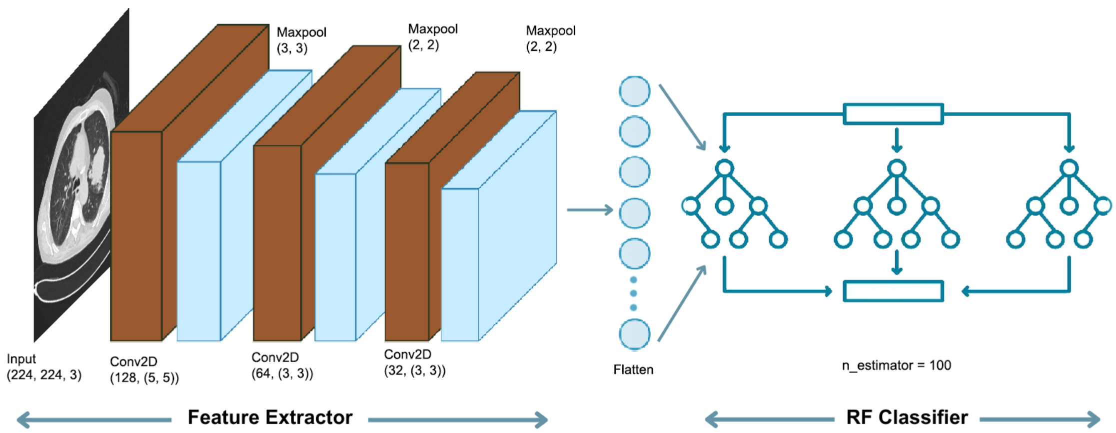

Figure 2: Hybrid CNN-RF Architecture

3.2.6 Lung ROI Nodule Mask Extraction

Lung ROI (Region of Interest) nodule mask extraction refers to the process of identifying and isolating the region of the lungs that contains nodules within a **CT scan**. It is a crucial method in medical image analysis, especially for tasks like detecting lung cancer, where the aim is to focus on areas of the lungs that might contain nodules.

3.2.7 Contrast Limited Adaptive Histogram Equalization

CLAHE (Contrast Limited Adaptive Histogram Equalization) is a method for improving image contrast, especially in medical imaging where finer details need to be enhanced without increasing noise excessively. CLAHE is widely used for preprocessing images in computer vision tasks, including those involving medical images like chest X-rays and CT scans. **[36]**

3.3 Project Framework and Implementation

The implementation and computational approach for this project follows an agile and iterative approach. The agile methodology suits this project because it offers flexibility, iterative development, and responsiveness to change—qualities that align well with the nature of experimental research and development.

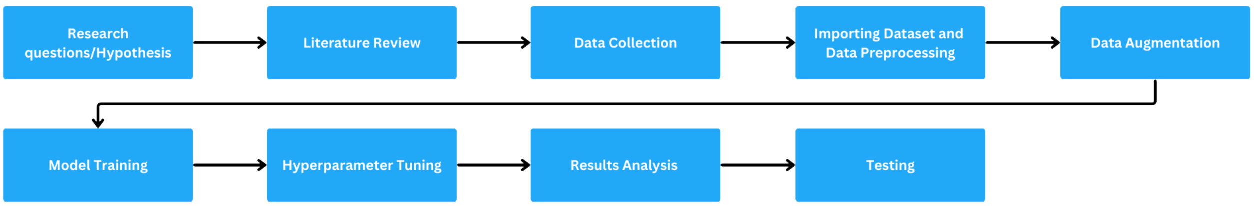

Figure 3: Project Framework

In figure 3, the project implementation framework presents each step that is to be carried out in the computational approach. Each step of the implementation process, starting with data collection will be explained in detail.

3.3.1 Data Collection

Three datasets were used in the implementation of this project, the Luna16 Dataset, the IQ-OTH/NCCD Lung Cancer Dataset, and the Chest CT-scan Dataset from Kaggle.

Luna16 Dataset

The **LUNA16 (**Lung Nodule Analysis 2016) dataset is part of a grand challenge in the field of **medical image analysis**, specifically aimed at improving algorithms for **automated detection of lung nodules** in computed tomography (CT) scans **[37].**

The **LUNA16 dataset** contains **888 CT scans**. These scans were derived from the **LIDC-IDRI** (Lung Image Database Consortium and Image Database Resource Initiative) database. The LUNA16 organizers selected a subset of scans from the LIDC-IDRI that met specific criteria, such as having consistent annotations of lung nodules by radiologists and being of sufficient image quality for analysis. The dataset is in RAW and MHD format.

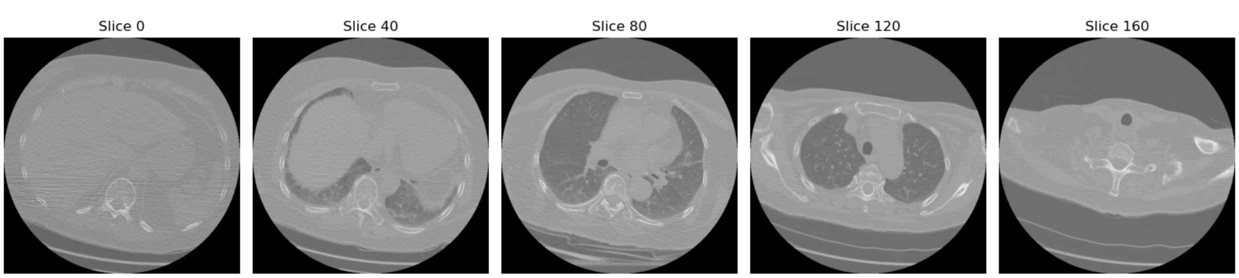

Figure 4: Slices of CT scans from Luna16 Dataset

Figure 5: Left to Right: Original CT scan, With Clahe CT scan, With Clahe and Weiner CT scan

The IQ-OTH/NCCD Dataset

The **IQ-OTH/NCCD (Iraqi Oncological Teaching Hospital / National Cancer Center Dataset)** is a medical image dataset specifically created for the classification of breast cancer tissue images. **[38].** It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. IQ-OTH/NCCD slide were annotated by oncologists and radiologists in the two centers. The dataset contains a total of 1190 images representing CT scans slices of 110 cases. These cases are grouped into three classes: normal, benign, and malignant.

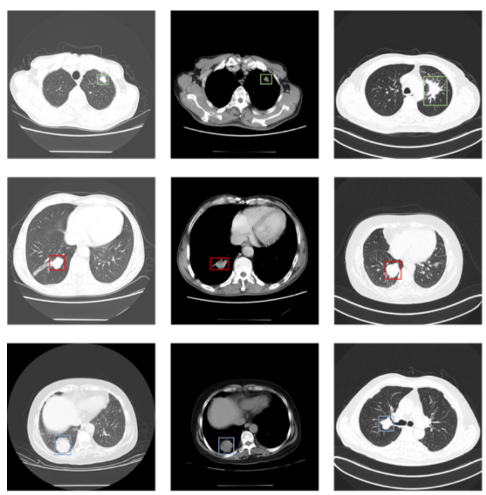

Figure 6: IQ-OTH/NCCD Dataset

Chest CT-Scans Dataset

The Chest CT scans dataset contains CT-Scan images with different types of chest cancer. The data contains 3 chest cancer types which are Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and 1 folder for the normal cell. Adenocarcinomas are found in several common cancers, including breast, prostate and colorectal. Large-cell undifferentiated carcinoma lung cancer grows and spreads quickly and can be found anywhere in the lung. Squamous cell lung cancer is responsible for about 30 percent of all non-small cell lung cancers, and is generally linked to smoking. The dataset contains 3 chest cancer types which are Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and 1 folder for the normal cell. The dataset is broken into 70% for training, 20% for testing and 10% for validation **[39]**.

3.3.2 Data Preprocessing

The data pre-processing stage of the experimental approach involved performing Lungs ROI nodule mask extraction on the LUNA16 Dataset, which is in RAW and MHD format. The RAW and MHD images first had to be converted to NumPy arrays. The process includes identifying the **region of interest (ROI)** in lung CT scans or X-ray images that corresponds to **nodules** (typically tumors or abnormal growths). This process involves creating a **mask** — a binary image (or labeled region) where the areas containing lung nodules are marked, typically with a value of `1`, and all other areas are marked as `0`. For accurate lung cancer detection, it is necessary to extract lung regions from the surrounding anatomical parts **[40].**

The CLAHE technique was applied for this purpose which involved identifying the threshold value for extracting the threshold mask with binary thresholding. The next step was to segment the lungs from the CT scan slices using binary thresholding. After extracting the region of interest from the lung nodules, the nodule masks and lungs ROI were saved as NumPy arrays.

The pre-processing techniques performed on **the IQ-OTH/NCCD** dataset involved defining the class labels of the dataset which are Benign, Malignant and Normal cases. Once the class labels were identified, the empty lists were initialized for each image and its corresponding label. Once the image and label lists were initialized, the image size was set to 224 and all the images were loaded using OpenCV and resized. The labels were converted to numeric values, Benign: 0, Malignant: 1, and Normal: 2. Once the labels were encoded to numeric labels, they were assigned to their corresponding images in the dataset and permutations were generation to perform random shuffling on the dataset. Normalization techniques were applied such as normalizing the pixel scale of the images.

3.3.3 Data Augmentation

The next step in the approach is data augmentation. Data Augmentation is a technique used to artificially increase the size and diversity of a dataset by creating modified versions of the existing data **[41-43].**

The data augmentation techniques used on the datasets in this study included applying random exposure on the datasets to adjust and enhance the brightness and exposure of the images. Horizontal Flipping was applied by about 50% in order to double the number of unique views the model sees. Random Rotation was another data augmentation technique applied on the dataset to rotate the images in random from an angle of -10 to 10.

The argument behind using these techniques for data augmentation is that Random Rotation allows the ANN model to learn the objects can appear in various orientations which is quite helpful in medical imaging, especially lung nodule detection and classification for Early Lung Cancer Diagnosis. Horizontal flipping is useful when the orientation of objects does not affect their classification. Randomly adjusting the exposure or brightness of an image helps the model become invariant to lighting conditions, which can vary significantly in real-world scenarios. Rescaling was another data augmentation technique used on the datasets.

3.3.4 Model Training

The four architectures to train the CNN models on the datasets are the U-NET architecture, DenseNet architecture, hybrid CNN-RF architecture and CNN with Pre-Trained Xception model. The U-NET model incorporates dropout and L2 regularization to enhance generalization and reduce overfitting. The architecture consists of an encoder-decoder structure with skip connections. The encoder down-samples input features using convolutional layers with ReLU activation, L2 regularization ($1\times10{-}4$), dropout (0.5), and max-pooling. The bottleneck captures high-level features with deeper convolutions. The decoder up-samples features via transposed convolutions, combining them with corresponding encoder outputs for spatial information recovery. Finally, a 1×1 convolution maps features to a single-channel output with a sigmoid activation for binary segmentation.

The DenseNet model employs dense blocks and transition layers for efficient feature propagation and compression. Each dense block consists of four convolutional layers, where every layer concatenates its output with previous layers to promote gradient flow and feature reuse. Growth rate controls the number of filters added per layer, set here to 48. Transition layers between blocks reduce spatial dimensions using 1×1 convolutions and average pooling, optimizing memory and computation. The network begins with a 7×7 convolution, followed by dense blocks and transitions, ending with global average pooling and a sigmoid-activated output for binary classification.

The CNN-RF model combines a convolutional neural network (CNN) for feature extraction and a Random Forest (RF) classifier for final predictions. The CNN includes three convolutional blocks with ReLU activations, pooling layers for down-sampling, L2 regularization, and HeNormal weight initialization. After feature extraction via flattening, the CNN outputs a 256-neuron dense layer with dropout to prevent overfitting, followed by a softmax output for classification. Intermediate features from the Flatten layer are used to train the RF classifier, which provides robust, ensemble-based predictions.

The CNN Model trained with Transfer Learning integrates a pretrained **Xception** architecture as the feature extraction backbone, optimized for large-scale image classification. Initially, the pretrained layers are frozen to retain their learned weights, preventing overfitting on smaller datasets. A custom classification head is added, comprising a **Global Average Pooling (GAP)** layer for dimensionality reduction and a Dense layer with **softmax activation** for multi-class output (4 classes). After initial training with the frozen base, the pretrained layers are unfrozen for **fine-tuning** with a lower learning rate.

Before the training process begins, the datasets are split into training, testing and validation sets.

The U-NET model is trained on a batch size of 16 and 100 epochs. The DenseNet model is trained on a batch size of 64, due to the large number of samples in the dataset and on 50 epochs. The CNN-RF model is trained on the IQ-OTHNCCD dataset with a batch size of 16 and on 50 epochs.

The Random Forest Classifier is trained using 100 n_estimators. The CNN model using Xception is trained with a batch size of 8 and on 50 epochs.

3.3.5 Hyperparameter Tuning

The hyperparameter tuning techniques used in the training process of the U-NET model involved using the ReduceLROnPlateu schedule whose role is to adjust the learning rate during training, specifically when the model's performance has "plateaued," meaning it's no longer improving. It is a useful technique in fine-tuning models and optimizing learning rates dynamically, often resulting in faster and more stable training. Early stopping is another hyperparameter tuning technique that was used to save training time by stopping the training process once the model's performance stops improving on a validation dataset. The Adam optimizer was used with a learning rate of ***2e-4*** and the evaluation metrics for the model were the model accuracy and model loss. L2 regularization is also implemented in order to reduce overfitting

On the DenseNet model, the hyperparameter tuning techniques involved using early call-backs. Early call-backs like **Early Stopping** monitor a model's performance on a validation set. When the model's validation metric (e.g., loss or accuracy) stops improving, training halts to prevent overfitting. This ensures the model doesn't continue learning details specific to the training data that could degrade performance on new data. The initial learning rate is set to 0.001, the Adam optimizer is used and the evaluation metric is model accuracy and model loss.

On the CNN-RF model, the learning rate is adjusted to 0.001 and the Adam optimizer is used. The evaluation metrics is model accuracy, precision, recall and f1 score.

On the CNN model trained using the pre-trained Xception architecture, the Early Stopping method is used as well as the ReduceLROnPlateau technique. The Adam optimizer is used and the learning rate is set to 1e-5.

Further hyperparameter tuning techniques involved adjusting the batch size and number of epochs to speed up the training process and produce better results.

3.3.6 Results Analysis

The results are analyzed using various evaluation metrics such as accuracy, precision, recall and f1 score. The primary evaluation metric used to compare the models is the model accuracy.

Accuracy
Model accuracy is described as how close a measured value to a known value. Three datasets were used using 4 model architectures. The accuracy of a deep learning model is crucial in deciding which algorithm is best suited for future experiments related to field of study. The accuracy of a model can be determined using the formula shown in **(Figure. 7).**

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Figure 7: Formula for determining Model Accuracy

Precision

Precision is an evaluation metric that measures the accuracy of positive predictions. It determines how many samples are correct from the total number of samples of a certain class that the model predicts. Precision is crucial when the cost of **false positives** is high. For instance, in medical diagnosis, mislabeling healthy patients as having a disease could lead to unnecessary treatments. Precision can be determined using the formula shown in **(Figure. 8).**

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Figure 8: Formula for determining Model Precision

F1 Score

The **F1 score** in of a machine learning model is a metric that provides a harmonic mean of **precision** and **recall**. It balances these two metrics, making it particularly useful when the dataset is imbalanced or when you need a single measure of model performance that considers both false positives and false negatives.

Precision is the fraction of correctly predicted positive samples out of all samples predicted as positive. **Recall** is the fraction of correctly predicted positive samples out of all actual positive samples. The f1 score of a model is calculated using the formula shown in **(Figure. 9).**

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 9: Formula for determining F1 Score

Recall

Recall, also known as **sensitivity** or **true positive rate (TPR)**, is a metric that measures how effectively a machine learning model identifies **all the positive instances** in the dataset. It is used to determine how many cases did a model correctly predict as positive from the total number of positive cases. While **recall** focuses on capturing all actual positive cases, **precision** measures how accurate the positive predictions are. The recall of a model can be determining using the formula in **(Figure. 10).**

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{80}{80 + 20} = 0.8\,(80\%)$$

Figure 10: Formula for determining Recall

3.3.7 Testing

Once the model is trained, it is tested on the test set of the Chest CT Scans and IQ-OTH/NCCD datasets. The Chest CT Scans and IQ-OTH/NCCD datasets are used for the model to make predictions.

3.4 Computational Resources

The computational resources used in this project include:

- **Hardware:** A 13” Apple MacBook Pro with 8GB RAM, an SSD of 512 GB and an M1 Chip was used for the overall project which includes practical implementation, report writing and presentation.
- **Software;** The software needed for the project includes:
  - Python, a high-level language used for Machine Learning coding.
  - Jupyter Notebook, a web-based open-sourced tool for practical implementation. The benefit of Jupyter Notebook is that it combines code with text in contrast to typical IDEs.
  - Google Colab and Kaggle Notebooks were preferred over Jupyter Notebook for model training due to their ability to run smoothly on GPU. GPU is preferred when training a model as it speeds up the training process.
  - Visual Paradigm, an editing tool for creating diagrams and charts.
  - Microsoft Office for report writing and presentation.

3.5 Justifications for Chosen Research Philosophy and Methodologies

The rational for the decisions made in choosing the philosophical approach and methodologies for early lung cancer diagnosis using ANNs are justified by their focus on practical applicability and technical rigor. Pragmatism emphasizes actionable insights, ensuring that the study addresses real-world challenges such as improving diagnostic accuracy, handling data quality, and achieving clinical acceptance. This philosophy aligns naturally with the quantitative methodology, which provides an objective, structured framework to evaluate large-scale data and validate models using precise metrics, crucial for reliability in clinical applications.

The methodologies further reinforce this alignment. Transfer learning reduces computational cost and improves feature extraction by leveraging pre-trained models like Xception. CNNs and U-Net architectures enhance image analysis, supporting accurate nodule detection and segmentation. DenseNet’s dense connectivity improves feature propagation, crucial for distinguishing nodules. The hybrid CNN-RF approach combines deep feature extraction with robust classification, while

CLAHE enhances image preprocessing for better model performance. Together, these methodologies ensure a comprehensive, impactful solution for early lung cancer diagnosis.

3.6 Justifications for Chosen Computational Approach

The argument for making the decisions behind the chosen implementation approach justify that each step of the implementation framework is carefully designed to address specific challenges in medical image analysis for early lung cancer detection.

1. **Data Collection**: Using multiple datasets (LUNA16, IQ-OTH/NCCD, and Chest CT-Scans) ensures diversity and robustness in the model, allowing it to generalize better across different data distributions.
2. **Data Preprocessing**: Preprocessing ensures data consistency, reduces noise, and focuses the model on relevant regions, like lung nodules, through techniques like ROI extraction, CLAHE, and normalization.
3. **Data Augmentation**: Augmentation increases data variability by applying random transformations like flipping, rotation, and brightness adjustment, enabling the model to learn invariant features and enhancing its robustness.
4. **Model Training**: Training diverse architectures (U-Net, DenseNet, CNN-RF, and Xception) ensures a comprehensive exploration of performance, balancing segmentation and classification challenges.
5. **Hyperparameter Tuning**: Techniques like early stopping and learning rate scheduling optimize training, ensuring efficiency and reducing overfitting.
6. **Testing**: Rigorous testing evaluates model performance on unseen data, ensuring reliability and real-world applicability.

3.7 Ethical and Professional Issues

The ethical and professional concerns that had to be adhered when applying the methodologies to the study are:

The methodologies applied in early lung cancer diagnosis using ANNs raise ethical and professional concerns. The pragmatic research philosophy prioritizes practical outcomes but may risk neglecting equity, interpretability, and inclusivity in healthcare applications. Similarly, the quantitative approach may overemphasize metrics like accuracy while underweighting broader implications such as patient impact and clinical trust.

Key ethical issues include biases in transfer learning and data quality, where pre-trained models and imbalanced datasets could propagate inaccuracies. CNNs, U-Net, and DenseNet models risk misdiagnoses due to segmentation errors, false positives/negatives, and challenges in interpretability. Hybrid models like CNN-RF complicate accountability, while preprocessing methods like CLAHE might introduce artifacts or variability, compromising diagnostic reliability. Professionally, these methodologies demand rigorous validation, reproducibility, and transparency. DenseNet’s complexity and resource requirements challenge scalability, while U-Net’s reliance on annotated datasets may limit its clinical generalizability. Hybrid architectures and preprocessing techniques add implementation challenges, increasing workflow complexity.

Chapter 4: Results Analysis and Testing

In this section, the results and findings of the experiment are discussed.

4.1 U-NET Model Performance

The U-NET model was trained with a batch size of 16 and number of epochs were 100. The Adam optimizer was used and the learning rate was 2e-4 with the model accuracy being the evaluation metric. The training stopped after 68 epochs and achieved a training accuracy of 99.95% and a validation accuracy of 99.97%. The model accuracy is shown in **(Figure. 11).**

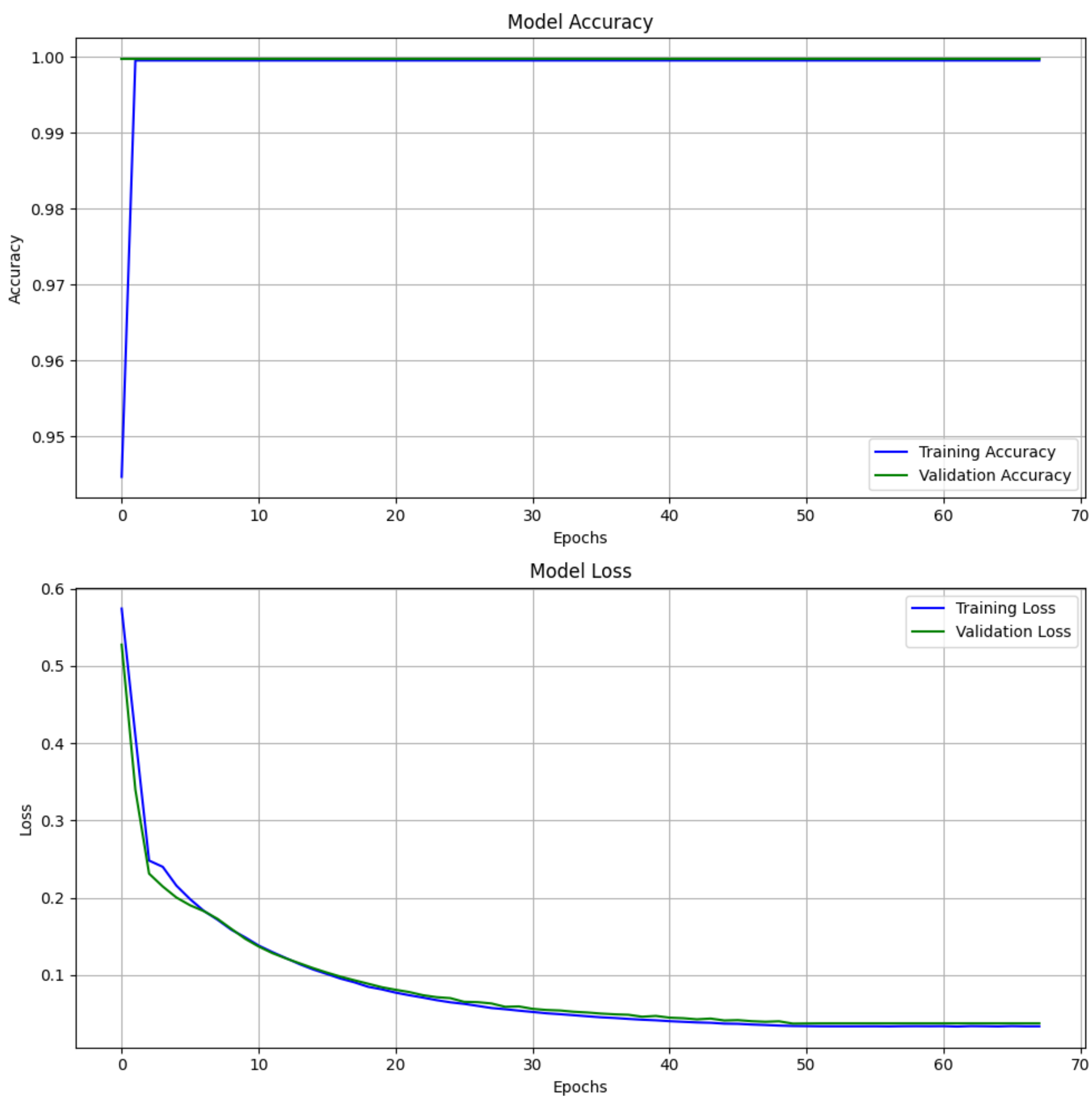

Figure 11: U-NET Model Performance

As you can see in **(Figure 11),** both training and validation accuracy are very close to 1 (almost 99.97%), indicating near-perfect classification. The model demonstrates excellent performance, achieving near-perfect training and validation accuracy with low loss. The consistency between training and validation curves indicates minimal overfitting, suggesting that the model generalizes well to unseen data.

4.2 DenseNet Model Performance

The DenseNet model was trained on the LUNA16 dataset with a batch size of 64. The batch size had to be increased due to the large number of samples in the dataset. The number of epochs were adjusted to 50 and the Adam optimizer was used with model accuracy and model loss being the evaluation metric. After 36 epochs, the model achieved a training accuracy of 99.05% and a validation accuracy of 97.42%. The performance for the DenseNet model is shown in **(Figure. 12).**

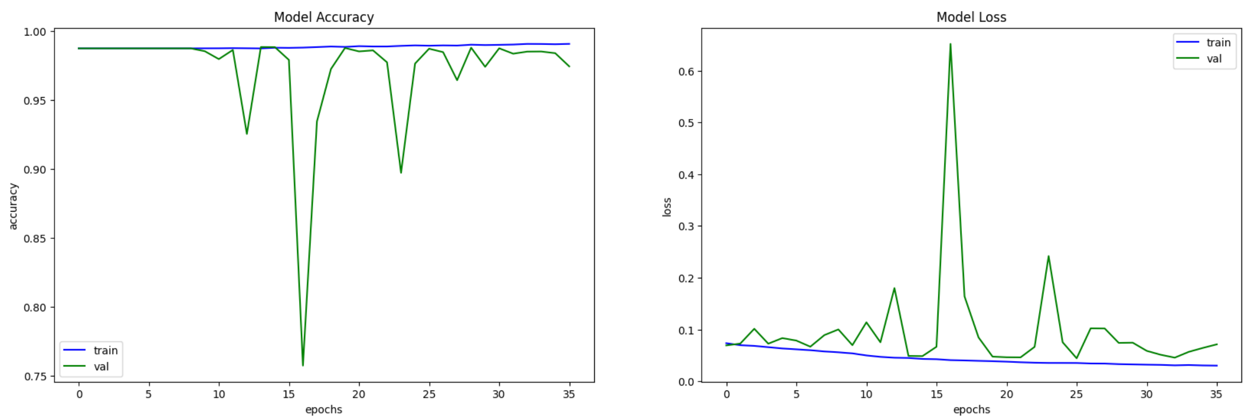

Figure 12: DenseNet Model Performance.

As seen in figure 12, the model shows high training accuracy, while the validation accuracy fluctuates significantly across epochs. The training accuracy is consistently high and improves steadily, indicating that the model is learning the training data well. The training loss is low and decreases gradually over time, which supports the indication that the model is fitting the training data effectively. Overall, while the model learns the training data effectively, the validation performance suggests it could benefit from strategies to improve generalization.

4.3 CNN-RF Model Performance

The hybrid CNN-RF model was trained on the IQ-OTH/NCCD dataset. The model was trained with a batch size of 16 and on 50 epochs. The Adam optimizer is used and the learning rate was set to 0.0001. After training, the model achieved an accuracy of 99.74 on the training set and an accuracy of 99.10% on the validation set. The model performance is shown in **(Figure. 13).**

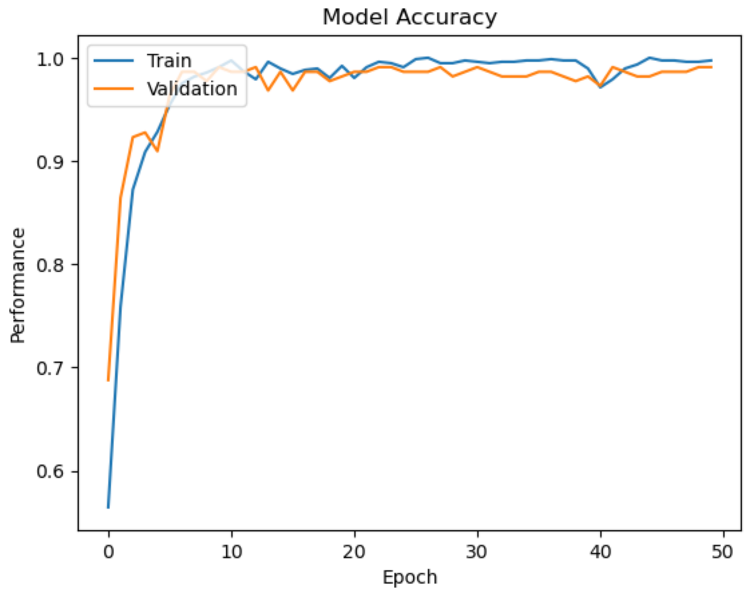

Figure 13a: CNN-RF Model Accuracy

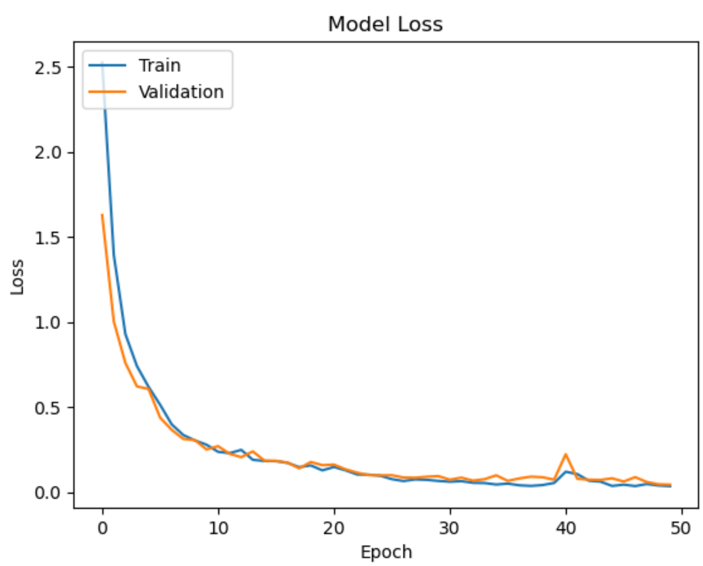

Figure 13b: CNN-RF Model Loss

In figures 13a and 13b, the training accuracy steadily increases and stabilizes around 1.0, indicating that the model learns the training data well. The validation accuracy also improves but shows fluctuations and does not reach the same level as the training accuracy. The training loss decreases over time and stabilizes at a low value, which is expected as the model learns from the training data. The validation loss decreases initially, then stabilizes with minor fluctuations, remaining slightly higher than the training loss.

Table. 1: CNN Model Classification Report

	Precision	recall	f1-score	support
0	1.00	0.92	0.96	12
1	0.97	1.00	0.98	56
2	0.97	0.95	0.96	41
accuracy			0.97	109
macro avg	0.98	0.96	0.97	109
weighted avg	0.97	0.97	0.97	109

In table 1, the classification report for the CNN of the hybrid CNN-RF model is presented. The classification report shows the CNN model's performance across three classes (0, 1, 2). The class labels are identified as 0 for Normal, 1 for Benign and 2 for Malignant. Precision, recall, and F1-score values are excellent, averaging **0.97** overall, reflecting high predictive accuracy. Class **1** achieves perfect recall (1.00), while slight performance dips occur for **class 0** in recall (0.92). Total support equals **109 samples**.

Table 2: Random Forest Model Classification Report

	Precision	recall	f1-score	support
0	1.00	0.83	0.91	12
1	1.00	1.00	1.00	56
2	0.95	1.00	0.98	41
accuracy			0.98	109
macro avg	0.98	0.94	0.96	109
weighted avg	0.98	0.98	0.98	109

In table 2, the classification report for the Random Forest algorithm of the hybrid CNN-RF model is presented. The model achieves high performance, with overall accuracy of **0.98**. Class **1** and **2** show perfect recall (1.00), while class **0** has a lower recall of **0.83**. Precision is excellent for all classes, and the macro average F1-score is **0.96**, reflecting strong overall predictive capabilities.

The confusion matrix for the CNN-RF Model is shown in **(Figure. 14).**

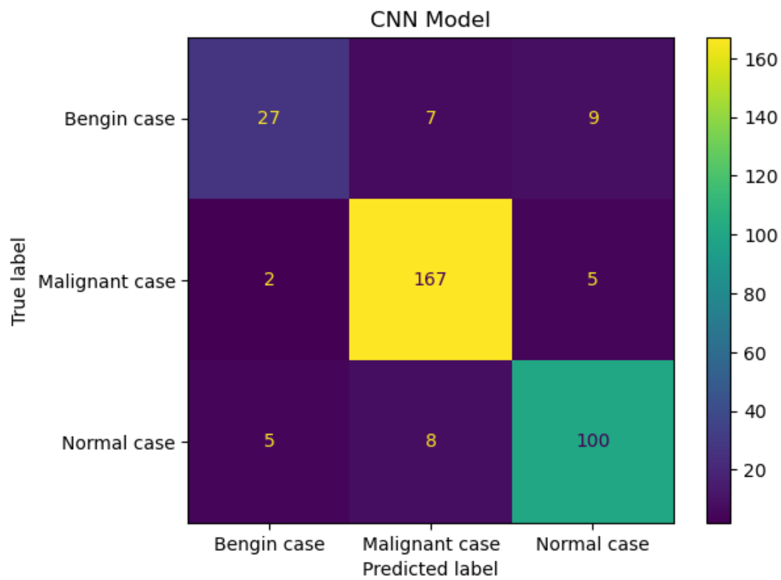

Figure 14: CNN-RF Model Confusion Matrix

4.4 Transfer Learning with Xception

In another method, the pre-trained model, Xception was used to train a model using the Chest CT-Scans dataset and IQ-OTH/NCCD dataset. The model was trained with a batch size of 8 and an output size of 4 on the Chest CT Scans dataset and an output size of 3 on the IQ-OTH/NCCD dataset. The number of epochs were 50 with the steps for every epoch being 25. The minimum learning rate was set to 1e-5 and the Adam optimizer was used. The model achieved a training accuracy of 99.93% and a validation accuracy of 90.28% on the Chest CT Scans dataset and on the IQ-OTJ/NCCD dataset, the model achieved a training accuracy of 99.34% and a validation accuracy of 97.27%. The model performance on the Chest CT Scans dataset is shown in **(Figure. 15)** and **(Figure. 16)** shows the model performance on the IQ-OTH/NCCD dataset.

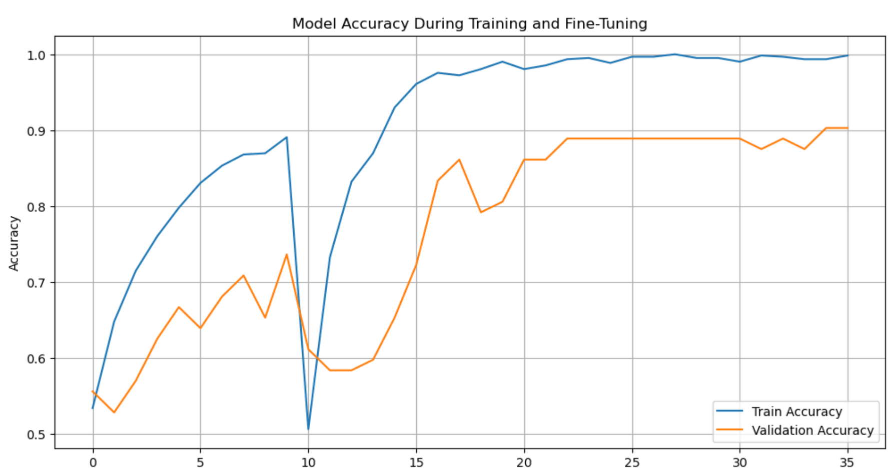

Figure 15a: Xception Model Accuracy on Chest CT Scans Dataset

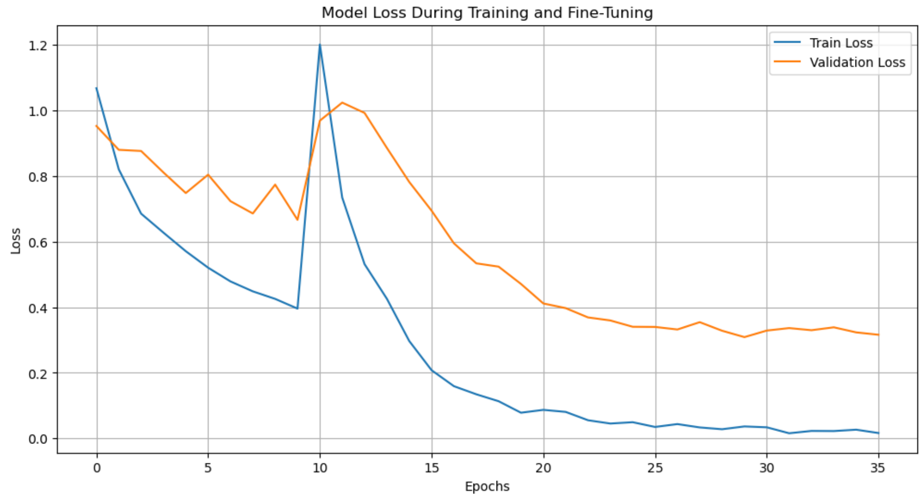

Figure 15b: Xception Model Loss on Chest CT Scans Dataset

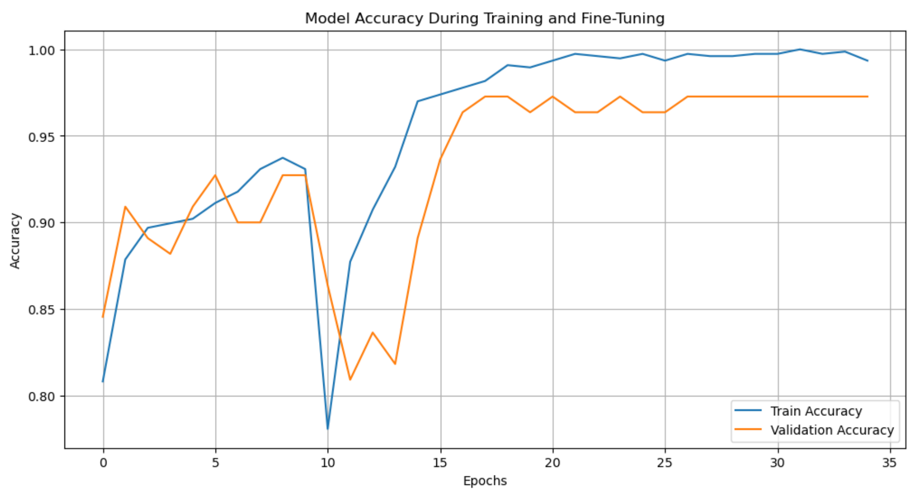

Figure 16a: Xception Model Accuracy on IQ-OTH/NCCD Dataset

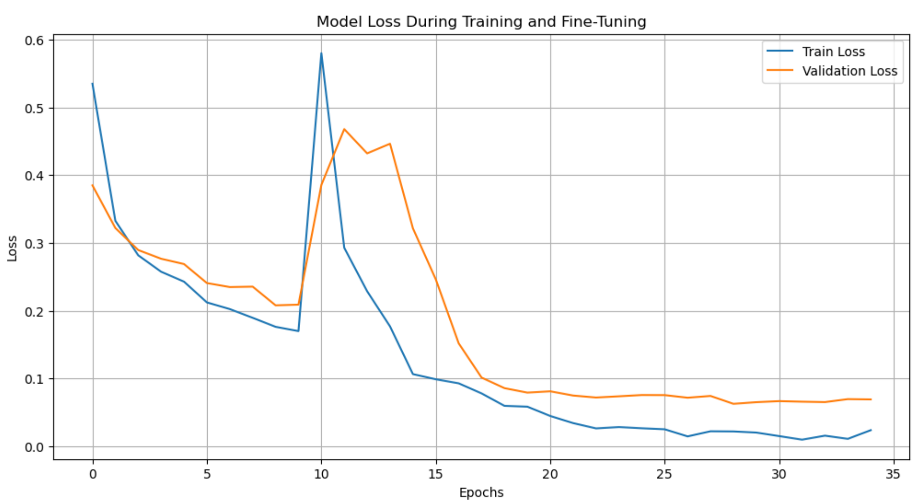

Figure 16b: Xception Model Loss on IQ-OTH/NCCD Dataset

Figures 15 and 16 show the model performance of the CNN trained using the Xception architecture. The model performed better on the IQ-OTH/NCCD dataset while the other version trained on the Chest CT scans dataset resulted in overfitting.

Table 3: Xception Model Classification Report

	Precision	recall	f1-score	support
adenocarcinoma	0.88	0.88	0.88	120
large.cell.carcinoma	0.85	0.88	0.87	51
normal	1.00	0.96	0.98	54
squamous.cell.carcinoma	0.91	0.91	0.91	90
accuracy			0.90	315
macro avg	0.91	0.91	0.91	315
weighted avg	0.91	0.90	0.91	315

Table 3 presents the classification report for the model trained using the Xception architecture. The model correctly classifies 90% of samples overall. It performs best on "normal" samples, with lower precision for "large cell carcinoma" and "squamous cell carcinoma".

4.5 Testing

Once the models were trained on the training data, they were tested on the test set. The UNET model was tested on the test set of the LUNA16 dataset and returned a test accuracy of 0.9997 and a test loss of 0.0373. The DenseNet model was tested on the test set of the Candidates data belonging to the LUNA16 dataset and returned a test accuracy of 0.9880 and a test loss of 0.0394. The hybrid CNN-RF model was tested on the test set of the IQ-OTH/NCCD dataset and returned a test accuracy of 0.9725 and a test loss of 0.0938. The CNN model trained using the pretrained Xception architecture was tested on the test set of the Chest CT Scans dataset and IQ-OTH/NCCD and returned a test accuracy of 0.9027 on the Chest CT Scans dataset and returned a testing accuracy of 0.9729 on the IQ-OTH/NCCD dataset.

Table 4: Model Comparison

Model	Training Accuracy	Testing Accuracy
U-NET	99.95%	99.97%
DenseNet	99.05%	98.80%
CNN-RF	99.74%	97.25%
Transfer Learning (Xception)	99.93%	97.29%

As shown in **Table 4,** the U-NET model returns the highest training accuracy at 99.95% and the highest testing accuracy at 99.97%. The DenseNet model comes second with a training accuracy of 99.05% and a testing accuracy of 98.80%. The CNN-RF model then returns a training accuracy of 99.74% and a testing accuracy of 97.25%. The CNN model training using transfer learning returns a training accuracy of 99.18% and a testing accuracy of 69.99%

4.5.1 Making Predictions

The model was used to make predictions on the Chest CT Scans dataset and the IQ-OTH/NCCD dataset. In **(Figure 17a to Figure 17d),** the model takes different samples belonging to each class of the Chest CT Scans dataset and predicts which class the Chest CT scan image belongs to.

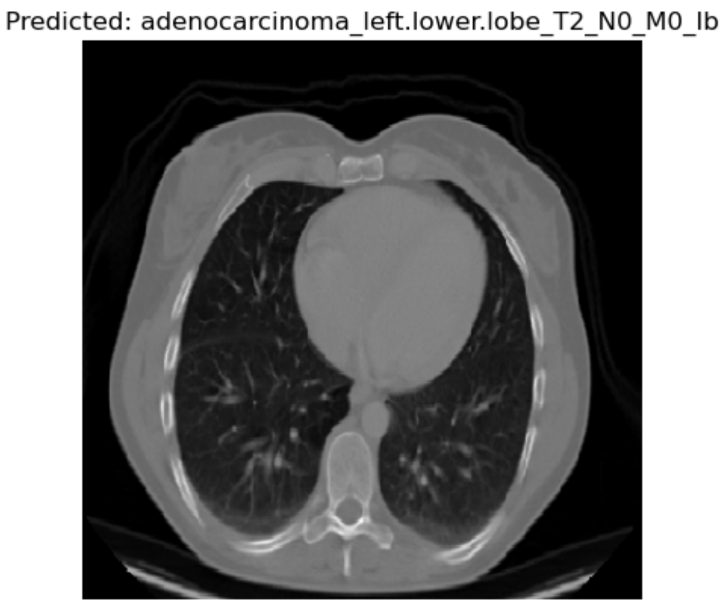

Figure 17a: Test Case 1

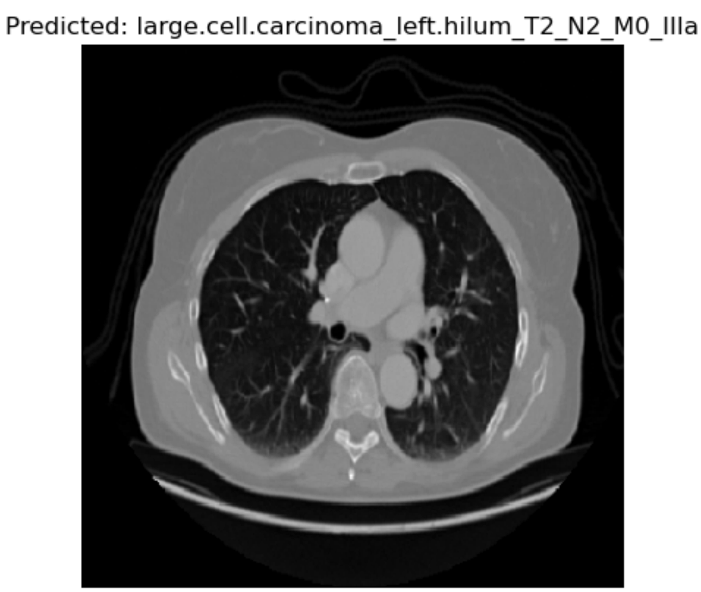

Figure 17b: Test Case 2

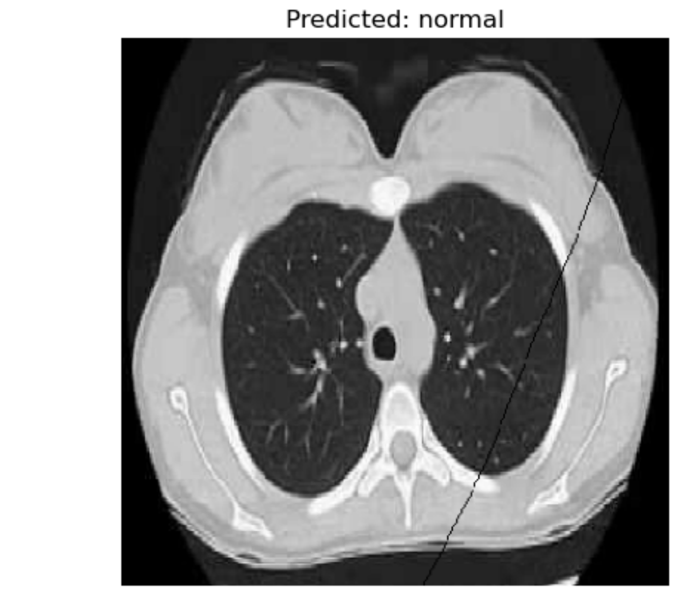

Figure 17c: Test Case 3

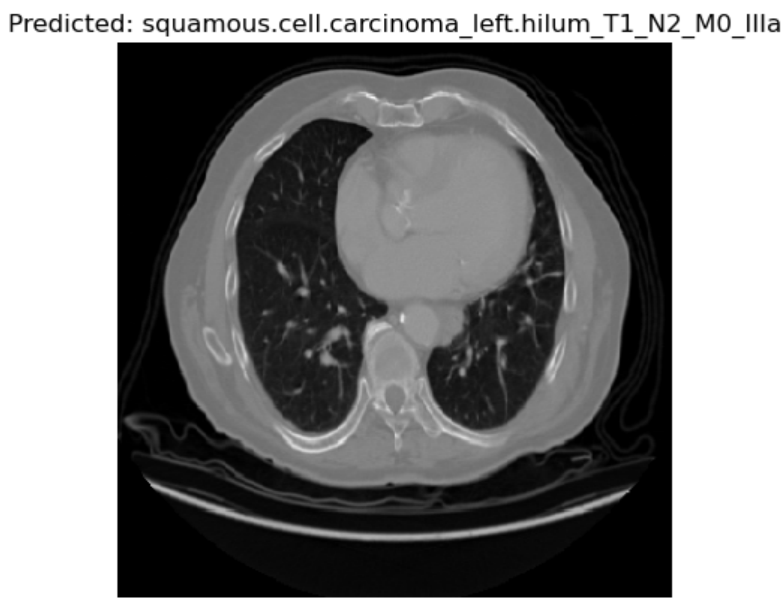

Figure 17d: Test Case 4

As shown in figures 17a to 17d, the model correctly predicted the class of each CT scan. Figure 16a shows that the CT scan has adenocarcinoma. Figure 16b shows that the CT scan has large cell carcinoma. Figure 16c shows that the CT scan is normal and Figure 16d shows that the CT scan has squamous cell carcinoma.

In **(Figures 18a to 18c),** the model takes an image as input from the IQ-OTH/NCCD dataset and predicts which class the image belongs to.

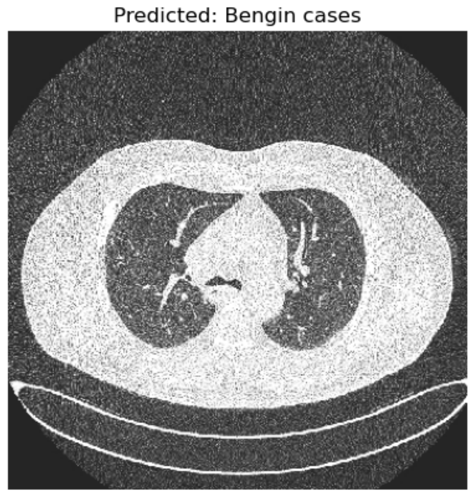

Figure 18a: Test Case 5

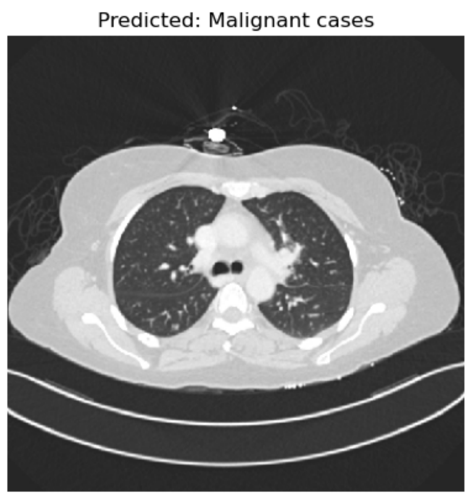

Figure 18b: Test Case 6

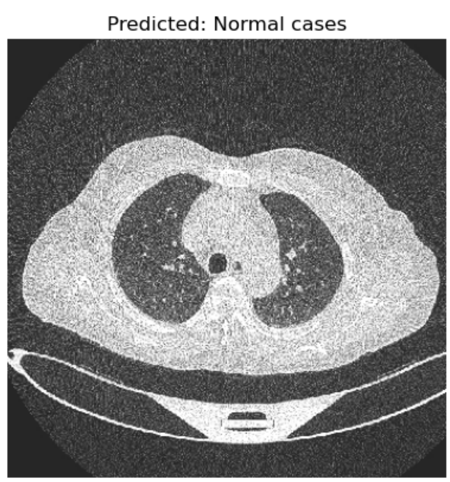

Figure 18c: Test Case 7

As shown in figures 18a to 18c, the model correctly predicted the class of each CT scan. Figure 18a shows that the CT scan is Benign. Figure 18b shows that the CT scan is Malignant. Figure 18c shows that the CT scan is Normal.

4.6 Comparing Results with Existing Literature

Table 5: Comparing Proposed Models to Existing Models in Literature

Published Work	Dataset	Accuracy	Precision	F1 Score	Recall
Multi-Layer Perceptron **[16]**	LIDC-IDRI	98.7%	-	-	-
KNN **[16]**	LIDC-IDRI	99.6%	-	-	-
SVM **[17]**	LIDC-IDRI	80%%	90%	69%	57%
Random Forest **[17]**	LIDC-IDRI	70%	89%	62%	47%
ANN **[17]**	LIDR-IDR	96%	92%	81%	69%
Hybrid CNN-SVM **[18]**	Chest CT Scans	96.71%	97.96%	-	-
CNN **[19]**	LIDC-IDRI	76%	-	-	-
Deep Belief Network **[20]**	LIDC-IDRI	79%	-	-	-
VGG19 with Transfer Learning **[21]**	IQ-OTH/NCCD	90%	43%	42%	42%
AlexNet CNN **[22]**	IQ-OTH/NCCD	93.548%	97.10%	96.40%	-
3D CNN **[23]**	NLST Dataset	94.4%	-	-	-
3D Deep CNN **[24]**	LUNA16	98.5	-	-	-
Proposed CNN Model using U-NET	**LUNA16**	**99.95%**	**-**	**-**	**-**
Proposed CNN using DenseNet	**LUNA16**	**99.05%**	**98.72%**	**98.06%**	**97.40%**
Proposed Hybrid CNN-RF Model	**IQ-OTH/NCCD**	**99.74%**	**98%**	**98%**	**98%**
Proposed CNN using Transfer Learning with Xception	**Chest CT Scans Dataset IQ-OTH/NCCD**	**99.34%**	**91%**	**91%**	**90%**

Table 5 presents a comparison of the proposed models with the models discussed in the literature. As shown in the table, the hybrid CNN-RF model in this study outperforms the hybrid CNN-SVM model in **[18]** with an accuracy of 99.74%. The CNN model implemented in this study using transfer learning on the pretrained Xception architecture outperforms the model trained using VGG19 in **[21].** The CNN models trained using U-NET and DenseNet outperform the model trained using AlexNet in **[22]** with accuracies of 99.95% and 99.05%. The models trained in **[17-17]** and **[19]** are outperformed by the hybrid CNN-RF model in this study.

Chapter 5: Conclusion and Future Work

5.1 Conclusion

Lung Cancer remains one of the leading causes of death worldwide. Therefore, in order to prescribe treatments, it is necessary to diagnose lung cancer within a patient at its early stage. Computer-Aided Design systems have been developed to assist radiologists in detecting lung nodules. In this study, an experiment was conducted using Deep Learning in which various ANN architectures were used on different datasets. The U-NET model was trained on the LUNA16 dataset and achieved a training accuracy of 99.95% and testing accuracy of 99.97. The DenseNet model was trained on the LUNA16 dataset and achieved a training accuracy of 99.05% and a testing accuracy of 98.80%. The hybrid CNN-RF model was trained on the IQ-OTH/NCCD dataset and achieved a training accuracy of 99.74% and a testing accuracy of 97.25%. The Xception model returns a training accuracy of 99.93% and a testing accuracy of 90.27% on the Chest CT Scans Dataset and a training and testing accuracy of 99.34% and 97.29% on the IQQ-OTH/NCCD Dataset.

From the results, it is evident that the proposed models in this study outperform the state-of-the-art models discussed in the existing literature. The proposed models take a CT scan as input from the Chest CT scans and IQ-OTH/NCCD datasets. The models correctly classify and predict the CT scans from the Chest CT scans dataset as either Adenocarcinoma, Large.Cell.Carcinoma, Squamous.Cell.Carcinoma and Normal. From the IQ-OTH/NCCD dataset, the proposed models take CT scan as input and correctly predict its class as either Benign, Malignant or Normal.

Looking back at the research questions in this study, we can conclude that:

Q1. Which ANN architectures (e.g., CNNs, transfer learning, or hybrid models) are best suited for early lung cancer diagnosis, and how can they be optimized for performance?

Based on the work in this study, we can conclude that CNN models developed using transfer learning on pretrained architectures such as Xception are best suited for early lung cancer diagnosis because they utilize pre-trained knowledge when trained on larger datasets and this makes them work well with both large datasets and small datasets. Hybrid models such as the proposed CNN-RF model are also well-suited because they combine machine learning algorithms like Random Forest with CNNs to achieve optimal performance.

Q2. How can transfer learning and pretrained ANN models be leveraged to improve diagnostic performance on lung cancer datasets with limited labeled samples?

Transfer learning and pre-trained models such as Xception leverage pre-trained knowledge when trained on a large dataset and fine-tune it on the target dataset. This makes them train well on smaller datasets such as IQ-OTH/NCCD and Chest CT scans dataset.

5.2 Project Outcomes

The set of outcomes of this project are:

- A processed dataset that can be used to train an ANN to classify lung nodules using CT scans
- A robust ANN model that can be utilized to classify lung nodules for lung cancer diagnosis.

5.3 Limitations of the Study

While this study serves as a good starting point for early lung cancer diagnosis, it has certain limitations. The U-NET and DenseNet models return a good accuracy however there may be some overfitting and further fine-tuning is necessary for them to work well with other datasets. The Xception model trained on the Chest CT scans dataset could be overfitting and needs further fine-tuning for improvement. The IQ-OTH/NCCD and Chest CT scans datasets were quite useful for training the hybrid CNN-RF model and the CNN developed using Xception, however the datasets lack clinical and patient data.

5.4 Future Work

Futuer iterations could include enhancing model performance by incorporating large and more complex datasets. The accuracy could be further improved by increasing the number of image samples. PET scans and MRI scans can be utilized in the future as alternatives to CT scans to train the model. Federated Learning techniques could be used to enable the proposed models to learn from more diverse and distributed datasets all the while ensuring data privacy and security **[44]**.

6. References

[1] Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Cur-ran WJ Jr, Wu YL, “Lung cancer: current therapies and new targeted treatments.” Lancet. 2017;389:299-311.

[2] Broggio J, John S. Cancer survival in England: Adult, stage at diagnosis and childhood - patients followed up to 2018. https://www.ons.gov.uk/peoplepopula-tionandcommunity/healthandsocialcare/condi-tionsanddiseases/bulletins/cancersurvivalineng-land/stageatdiagnosisandchildhoodpatientsfol-lowedupto 2018. 2019. Accessed 7 November 2020.

[3] C. J. Mathew, A. M. David, and C. M. J. Mathew, "Artificial Intelligence and its future potential in lung cancer screening," *EXCLI Journal*, published online Dec. 11, 2020, doi: 10.17179/excli2020-3095.

[4] Hussain, S. K., Ramay, S. A., & Arshad, R. (2023). “Machine Learning Approaches for Early Detection of Lung Cancer”. Available At: https://www.researchgate.net/publication/378964630_Machine_Learning_Approaches_for_Early_Detection_of_Lung_Cancer. *Journal of Medical Imaging and Health Informatics*, 13(1), 45-58.

[5] I. A. f. R. o. C. WHO, "New Global Cancer Data: GLOBOCAN 2018. World Health Organization," 2018. Available at: https://www.uicc.org/

[6] Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification. Multimedia Tools Applications. 2020:1-32.

[7] American Lung Association, URL www.lung.org.
[8] Republic of Iraq, Ministry of Health\Environment, Board. IC. Annual Report Iraqi Cancer Registry. 2016

[9] Ha R, Mutasa S, Karcich J, Gupta N, Sant EPV, Nemer J, Sun M, Chang P, Liu MZ, Jambawalikar S. Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm. J Digit Imaging. 2019;32:276–82.

[10] Gao Q, Wu Y, Qian Li. Fast training model for image classification based on spatial information of deep belief network. J Syst Simul. 2020;27(3):549–58.

[11] American Cancer Society, Key Statistics for Lung Cancer, 2020, Available from https://www.cancer.org/cancer/lung- cancer/about/key-statistics.html

[12] H. Jiang, He Ma, W. Qian, M. Gao and Y. Li, “An Automatic Detection System of Lung Nodule based on Multi-Group Patch-Based Deep Learning Network,”2018. [online]. Available: https://doi.org/10.1109/JBHI.2017.2725903.

[13] https://lola11.grand-challenge.org/

[14] A. Setio, A. Traverso, T. Bel, and S.N. Berend. (2016). “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge”. Published: 2016. Available: https://ar5iv.labs.arxiv.org/html/1612.08012

[15] K. Kuan, M. Ravaut, G. Manek, V. Chandrasekhar, and others, "Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge," ***arXiv preprint arXiv:1705.09435***, May 2017. DOI: 10.48550/arXiv.1705.09435

[16] Binila. M. Boban and R. K. Megalingam. “Lung diseases classification based on machine learning algorithms and performance evaluation”. In 2020 International Conference on Communication and Signal Processing (ICCSP), pages 0315– 0320. IEEE, 2020.

[17] Nikita Banerjee and Subhalaxmi Das. Prediction lung cancer–in machine learning perspective. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pages 1–5. IEEE, 2020.

[18] A. Y. Saleh, C. K. Chin, V. Penshie, and H. R. H. Al-Absi, "Lung cancer medical images classification using hybrid CNN-SVM," *Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak*, Sarawak, Malaysia; *Faculty of Engineering, Universiti Malaysia Sarawak*, Sarawak, Malaysia; *College of Science and Engineering, Hamad Bin Khalifa University*, Doha, Qatar.

[19] D. Yu *et al.*, "Convolutional neural networks for predicting molecular profiles of non-small cell lung cancer," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017: IEEE, pp. 569-572. doi: 10.1109/ISBI.2017.7950585

[20] G. da Silva, A. Silva, A. de Paiva, and M. Gattass, "Classification of malignancy of lung nodules in CT images using convolutional neural network," in *Anais do XVI Workshop de Informática Médica*, 2016: SBC, pp. 21-29. Available at: Google Scholar.

[21] A. Mohite, "Application of Transfer Learning Technique for Detection and Classification of Lung Cancer using CT Images," *International Journal of Scientific Research and Management (IJSRM)*, vol. 9, no. 11, pp. EC-2021-621-634, Nov. 2021. DOI: 10.18535/ijsrm/v9i11.ec02.

[22] H. F. Al-Yasriy, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of Lung Cancer Based on CT Scans Using CNN," in *IOP Conference Series: Materials Science and Engineering*, vol. 928, 2nd International Scientific Conference of Al-Ayen University (ISCAU-2020), Thi-Qar, Iraq, Jul. 2020, p. 022035, doi: 10.1088/1757-899X/928/2/022035.

[23] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019, doi: 10.1038/s41591-019-0447-x

[24] A. Masood, P. Yang, B. Sheng, H. Li, P. Li, J. Qin, V. Lanfranchi, J. Kim, D. D. Feng, "Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT," 2020. [online]. Available: https://doi.org/10.1109/JTEHM.2019.2955458

[25] N. Capes, H. Patel, I. Sarhan, and R. Shehata, "Artificial Intelligence: A Pragmatic Approach to Implementation in Medicine – A Review of the Literature and Survey of Local Practice in Midlands in UK," *International Journal of Intelligence Science*, vol. 13, no. 3, pp. 63–79, Jul. 2023, DOI: 10.4236/ijis.2023.133005.

[26] W. Shi *et al.*, "Quantitative analysis of imaging characteristics in lung adenocarcinoma in situ using artificial intelligence," *Thoracic Cancer*, vol. 15, no. 10, pp. 2345-2356, Oct. 2024, doi: 10.1111/1759-7714.15447.

[37] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC Medical Imaging*, vol. 22, no. 69, pp. 1–20, 2022.

[28] H.-y. Lee, "Deep learning tutorial," *Open Course, Online Available,* 2020. Available at: Google Scholar.

[29] H. Dao, "Image Classification Using Convolutional Neural Networks," 2020. Available at: Google Scholar.

[30] E. M. Hosseini, M. Hajabdollahi, N. Karimi, S. Samavi, and S. Shirani, "Splitting Convolutional Neural Network Structures for Efficient Inference," *arXiv preprint arXiv:2002.03302,* 2020. Available at: Google Scholar.

[31] Saha S. A comprehensive guide to convolutional neural networks–the ELI5 way. Medium; 2018.

[32] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into imaging. 2018;9(4):611-29.

[33] T.-W. Chiu, Y.-L. Tsai, and S.-F. Su, "Automatic detect lung node with deep learning in segmentation and imbalance data labeling," *Journal of Ambient Intelligence and Humanized Computing*, 2023

[34] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[35] Y. Lu, S. Aslani, A. Zhao, A. Shahin, D. Barber, M. Emberton, D. C. Alexander, and J. Jacob, "A hybrid CNN-RNN approach for survival analysis in a Lung Cancer Screening study," *Medical Image Analysis*, vol. 77, pp. 1–13, Jul. 2022, doi: 10.1016/j.media.2022.102368.

[36] Pizer, S. M. et al. "Adaptive histogram equalization and its variations. Compute. Vis. Gr. Image Process". 39, 355–368. (1987).

[37] https://luna16.grand-challenge.org/Data/

[38] K. Punithavathy, M. M. Ramya, and S. Poobal, "Analysis of statistical texture features for automatic lung cancer detection in PET/CT images," Available at: https://www.researchgate.net/publication/280620683_Analysis_of_statistical_texture_features_in_PETCT_images_for_lung_cancer_detection

[39] https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[40] S. Kaur, H. Aggarwal, and R. Rani, "Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation," *Multimedia Tools and Applications,* vol. 80, no. 7, pp. 10113-10139, 2021. doi: 10.1007/s11042-020-10114-1

[41] S. Dutta, P. Prakash, and C. G. Matthews, "Impact of data augmentation techniques on a deep learning based medical imaging task," in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, 2020, vol. 11318: International Society for Optics and Photonics, p. 113180M. doi: 10.1117/12.2549806

[42] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*, 2018: IEEE, pp. 117-122. DOI: 10.1109/IIPHDW.2018.8388338

[43] Z. Ren, Y. Zhang, and S. Wang, "LCDAE: Data Augmented Ensemble Framework for Lung Cancer Classification," *Technology in Cancer Research & Treatment*, vol. 21, no. 5, Sep. 2022. DOI: 10.1177/15330338221124372.

[44] G. Mostafa, Md. S. Hamidi, and D. Md. Farid**,** "Detecting Lung Cancer with Federated and Transfer Learning," *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2023, pp. 1–6. DOI: 10.1109/ICCIT60459.2023.10441256

7. Glossary

Agile Methodology	Agile methodology is an iterative approach to project management and software development that emphasizes flexibility, collaboration, and customer feedback to deliver small, incremental improvements.
Artificial Intelligence	Artificial intelligence (AI) is a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyze data, make recommendations
Artificial Neural Network	An artificial neural network (ANN) is a machine learning algorithm that uses a network of interconnected nodes, or neurons, to process data in a way that mimics the human brain
Accuracy	Model accuracy is the ratio of correctly predicted outputs to the total number of predictions, often expressed as a percentage.
Batch Size	Batch size is the number of samples fed to a neural network at each iteration of the training process
Benign	Benign refers to a condition, tumor, or growth that is non-cancerous and typically not harmful, as it does not spread to other parts of the body.
Classification Report	A classification report is a performance evaluation metric that provides detailed statistics such as precision, recall, F1-score, and support for each class in a classification model.
Confusion Matrix	A confusion matrix is a table used to evaluate the performance of a classification model by showing the true positives, true negatives, false positives, and false negatives.
Convolutional Neural Network	A convolutional neural network (CNN) is a deep learning architecture that uses data to learn and identify patterns in images, audio, time-series, and signal data
CT Scan	A CT scan is a diagnostic imaging procedure that uses X-rays and a computer to create detailed pictures of the inside of the body
Dataset	A dataset is a collection of structured or unstructured data used for analysis, training, or evaluation in machine learning and other applications.
Data Preprocessing	Data preprocessing is the process of cleaning, transforming, and organizing raw data into a suitable format for machine learning models.
Data Augmentation	Data augmentation is the process of artificially increasing the size and diversity of a dataset by applying transformations like rotation, flipping, or color changes to existing data.

Deep Learning	Deep learning is a type of machine learning that uses artificial neural networks to teach computers to process data in a way that mimics the human brain
Epoch	An epoch is one complete pass through the entire training dataset during the training of a machine learning model.
Federated Learning	Federated learning is a machine learning approach where multiple devices collaboratively train a model while keeping their data decentralized and private.
Framework	A framework is a structured platform or set of tools that provides a foundation for developing and deploying software or machine learning models efficiently.
Google Colab	Google Colab is a cloud-based platform that provides a Jupyter notebook environment, allowing users to write and execute Python code, with access to free GPU and TPU resources for machine learning and data analysis tasks.
Hybrid Model	A hybrid model is a machine learning or computational model that combines different algorithms, techniques, or approaches to improve performance and solve complex problems more effectively.
Hyperparameter Tuning	Hyperparameter tuning is the process of optimizing the configuration settings of a machine learning model to improve its performance.
Jupyter Notebook	Jupyter Notebook is an open-source interactive computing environment that allows users to create and share documents containing live code, equations, visualizations, and narrative text.
Kaggle Notebook	A Kaggle notebook is an interactive coding environment provided by Kaggle that allows users to write, run, and share Python code, particularly for data analysis, machine learning, and data science projects.
KNN	K-Nearest Neighbors (KNN) is a simple, non-parametric machine learning algorithm that classifies data points based on the majority class of their k closest neighbors in the feature space.
Lung Cancer	Lung cancer is a type of cancer that begins in the lungs, typically characterized by abnormal cell growth that can spread to other parts of the body.
Learning Rate	Learning rate is a hyperparameter that controls the size of the steps a model takes during optimization to adjust its weights in order to minimize the loss function.

MacBook	A MacBook is a line of laptop computers designed and manufactured by Apple, known for their sleek design, high performance, and macOS operating system.
Machine Learning	Machine learning is a branch of artificial intelligence that enables computers to learn from data and improve their performance on tasks without being explicitly programmed.
Malignant	Malignant refers to a harmful condition, often used to describe cancerous tumors that can invade surrounding tissues and spread to other parts of the body.
Model	A model is a mathematical or computational representation that is trained to recognize patterns in data and make predictions or decisions based on that data.
Methodology	Methodology is a system of methods, principles, and rules used in a particular discipline or research to approach and solve problems.
Multi-Layer Perceptron	A Multi-Layer Perceptron (MLP) is a type of neural network consisting of multiple layers of interconnected nodes (neurons), typically used for supervised learning tasks such as classification and regression.
Output	Model output is the result or prediction generated by a machine learning model after processing input data.
Python	Python is a high-level, interpreted programming language known for its simplicity, readability, and versatility, used in machine learning and data science.
Precision	Precision is the ratio of true positive predictions to the total number of positive predictions made by a model, indicating how accurate the model's positive predictions are.
Quantitative Methodology	Quantitative methodology is a research approach that focuses on collecting and analyzing numerical data to identify patterns, test hypotheses, and make predictions.
Random Forest	Random Forest is an ensemble machine learning algorithm that constructs multiple decision trees and merges their predictions to improve accuracy and reduce overfitting.
SVM	Support Vector Machine (SVM) is a supervised machine learning algorithm that finds the hyperplane that best separates data into different classes by maximizing the margin between them.

Testing	Model testing is the process of evaluating a trained machine learning model's performance on unseen data to assess its generalization ability and accuracy.
Training	Model training is the process of teaching a machine learning model to learn patterns from data by adjusting its parameters to minimize the error in its predictions.
Transfer Learning	Transfer learning is a machine learning technique where a model trained on one task is reused or fine-tuned for a different but related task, leveraging previously learned features to improve performance.
Validation	Model validation is the process of assessing a machine learning model's performance and generalization ability on a separate validation dataset to ensure it is not overfitting or underfitting the training data.
VGG19	VGG19 is a deep convolutional neural network architecture with 19 layers, known for its simplicity and effectiveness in image classification tasks, developed by the Visual Geometry Group at Oxford University.

8. Appendices

8.1 Appendix A

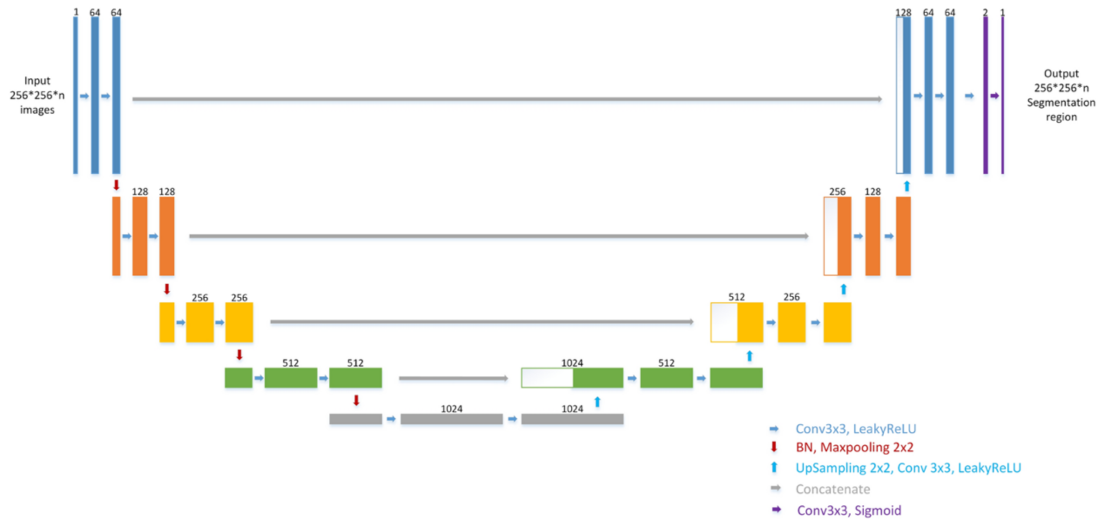

Figure 1: U-NET Architecture

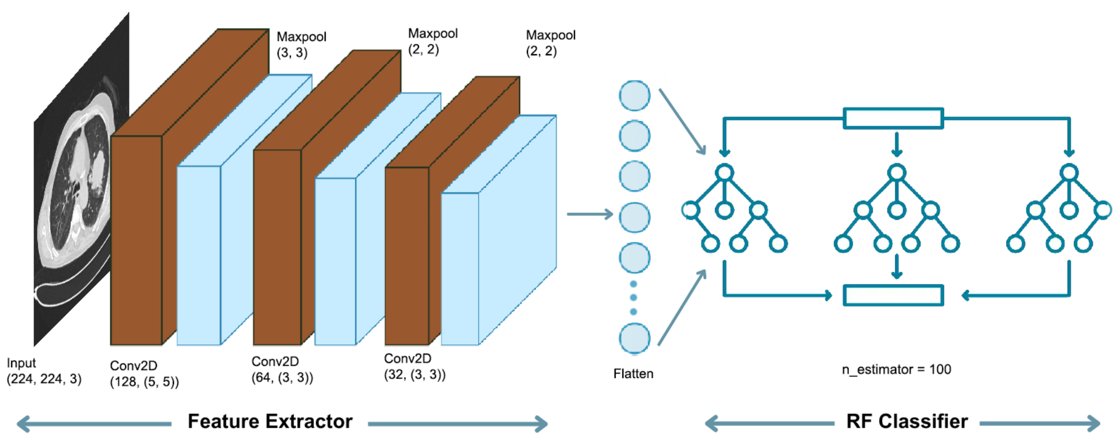

Figure 2: Hybrid CNN-RF Architecture

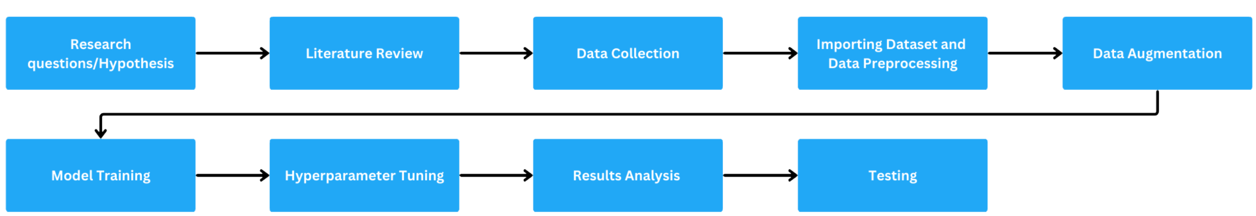

Figure 3: Project Framework

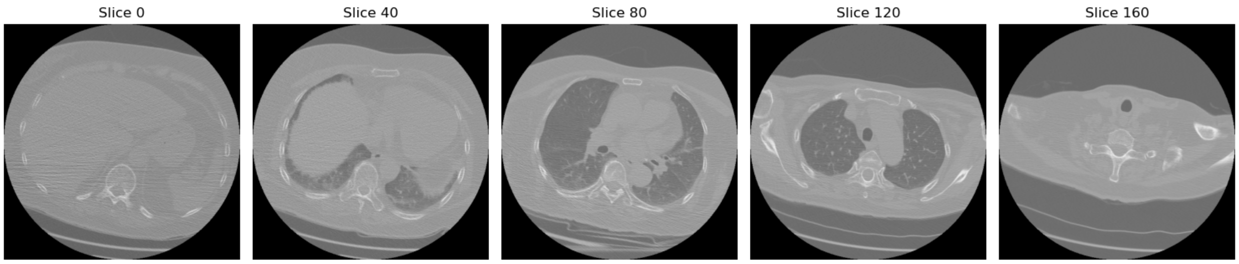

Figure 4: Slices of CT scans from LUNA16 Dataset

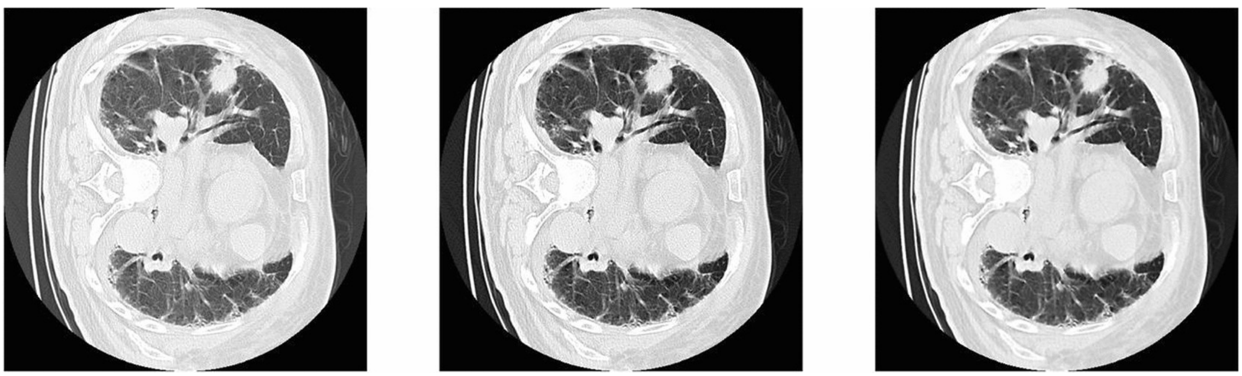

Figure 5: Left to Right: Original CT scan, With Clahe CT scan, With Clahe and Weiner CT scan

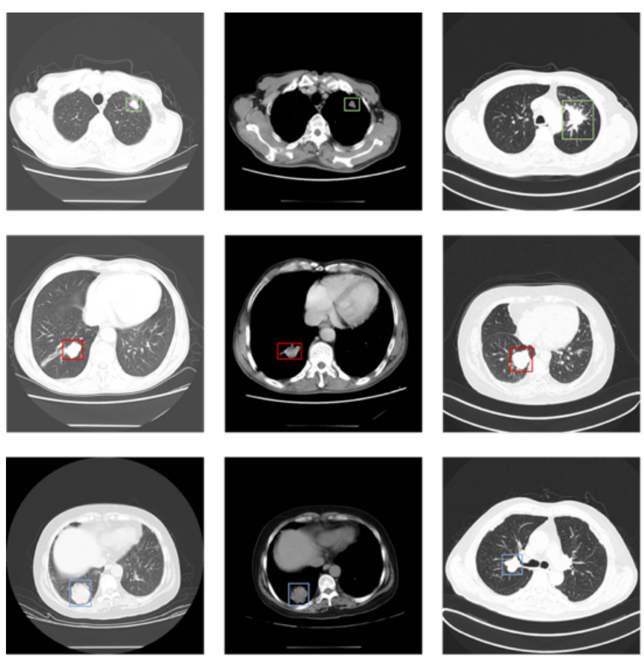

Figure 6: IQ-OTH/NCCD Dataset

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Figure 7: Formula for determining Model Accuracy

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Figure 8: Formula for determining Model Precision

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Figure 9: Formula for determining F1 Score

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{80}{80 + 20} = 0.8\,(80\%)$$

Figure 10: Formula for determining Recall

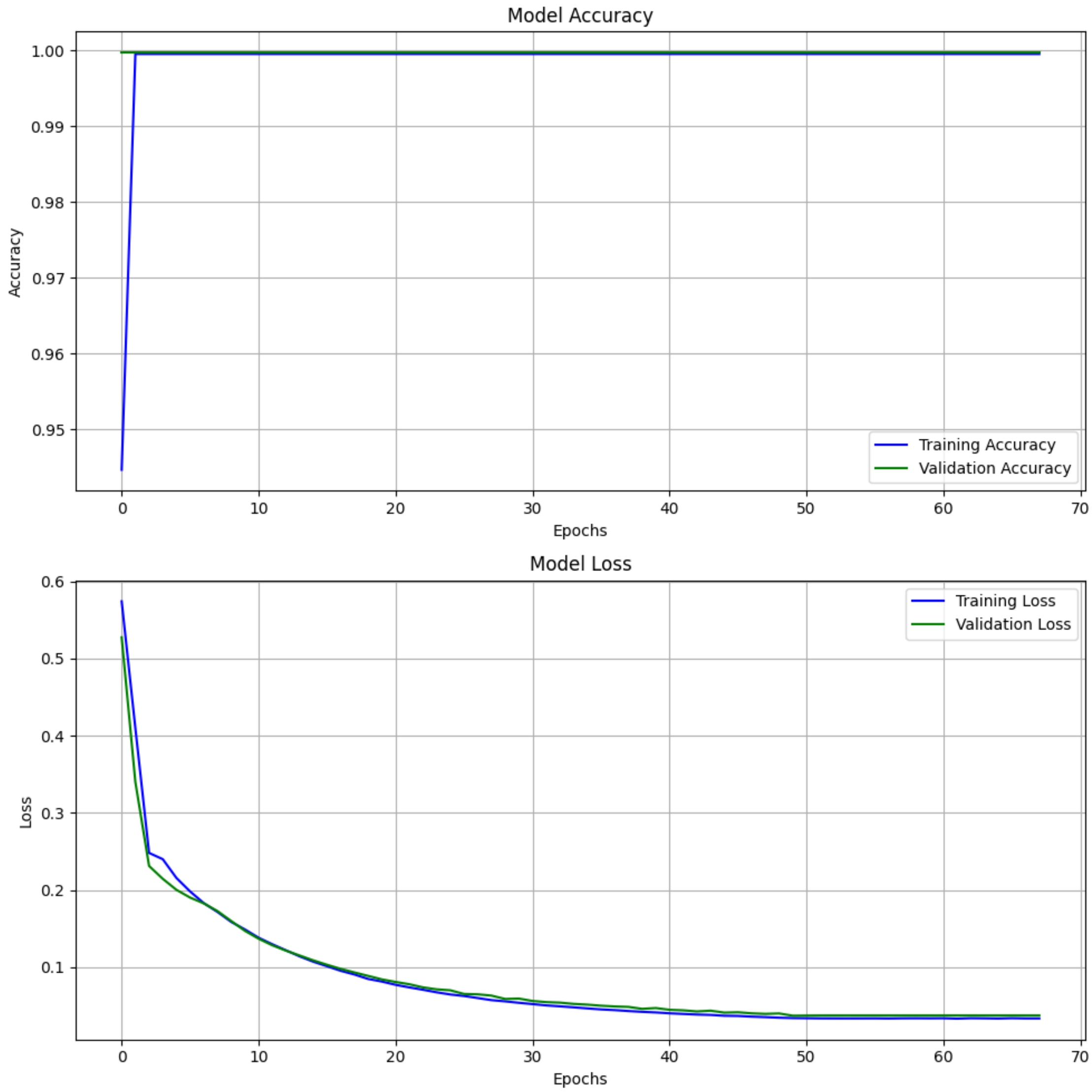

Figure 11: U-NET Model Performance

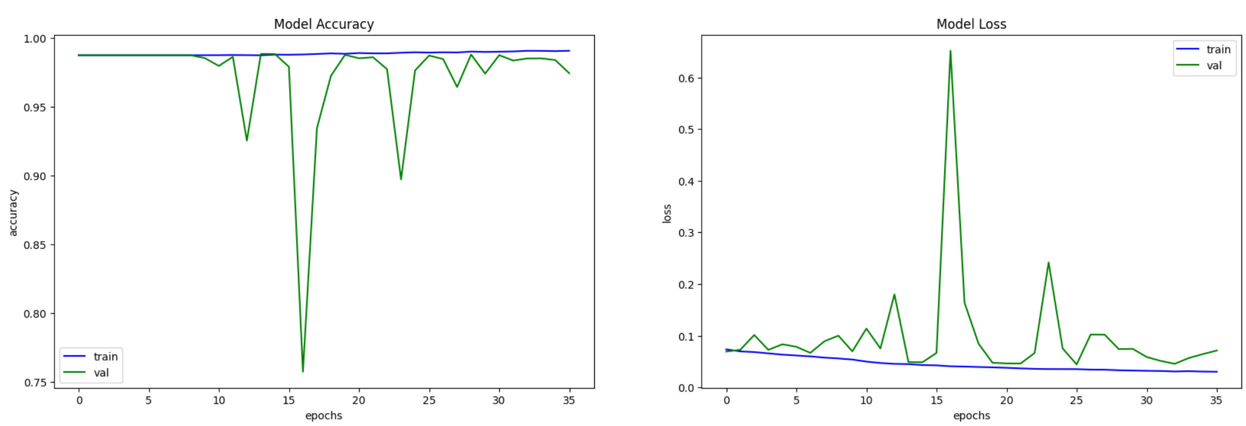

Figure 12: DenseNet Model Performance.

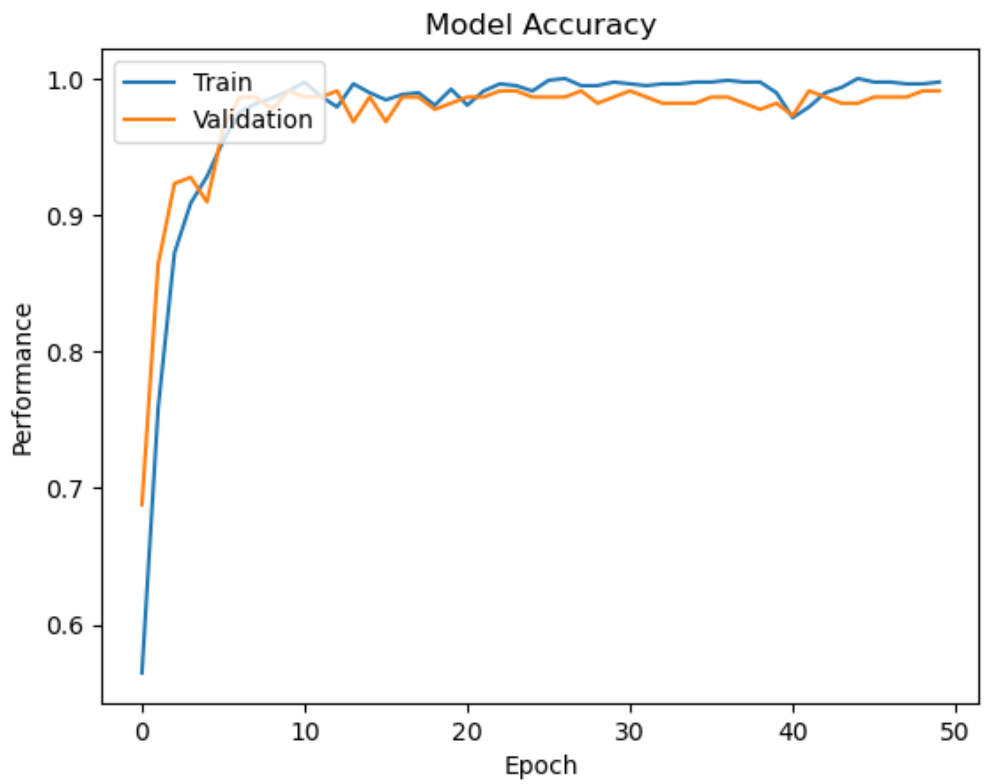

Figure 13a: CNN-RF Model Accuracy

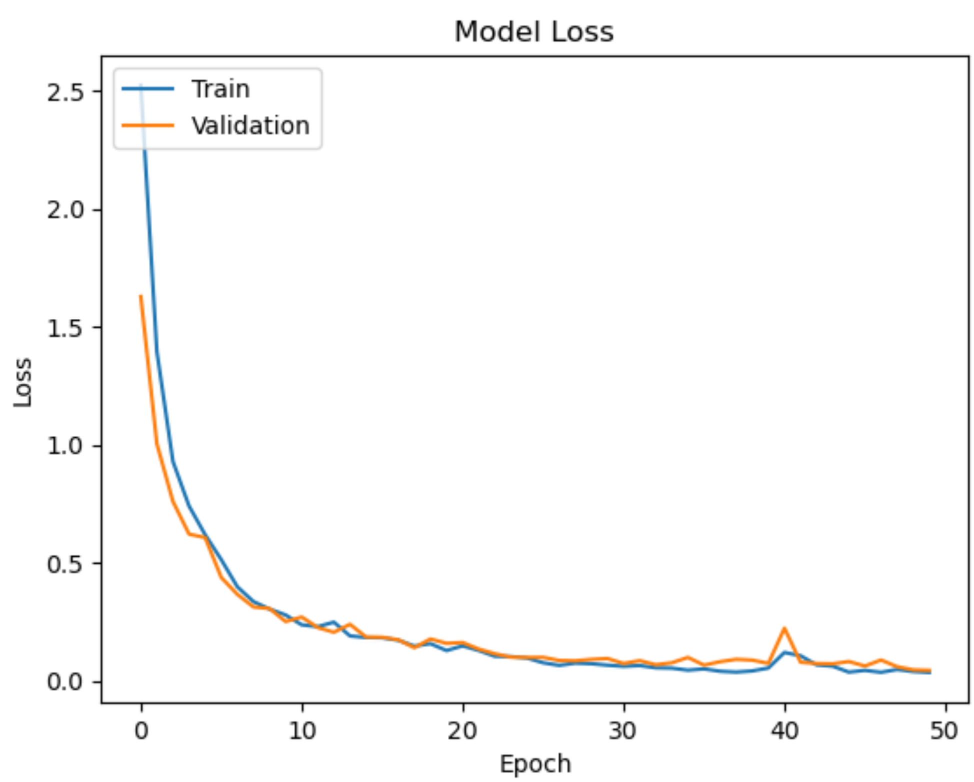

Figure 13b: CNN-RF Model Loss

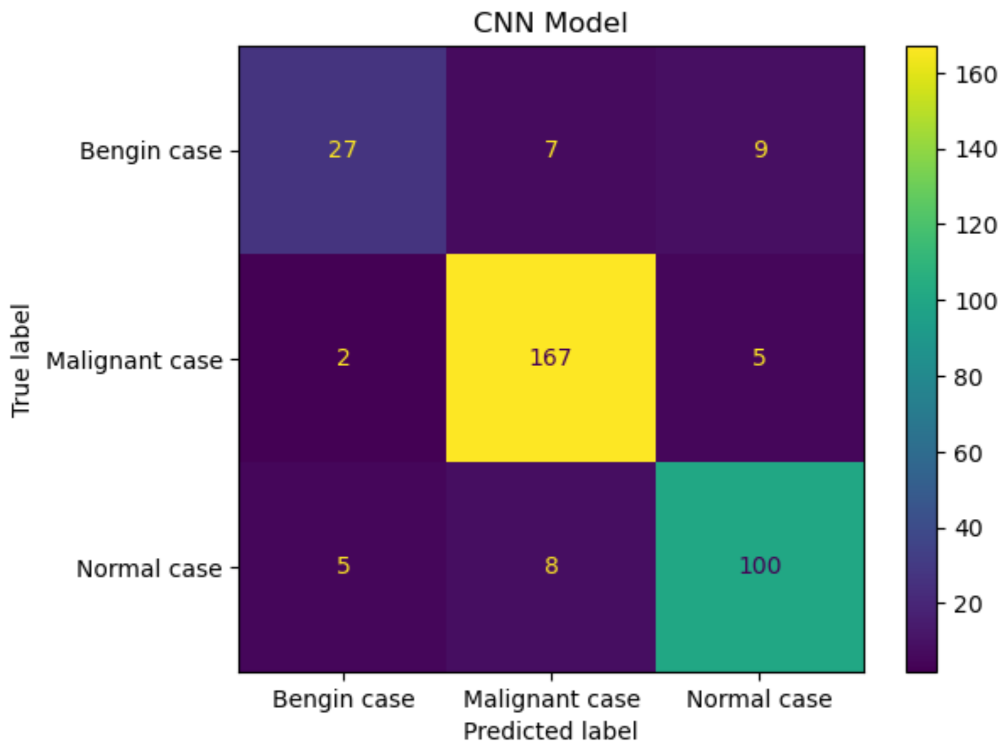

Figure 14: CNN-RF Model Confusion Matrix

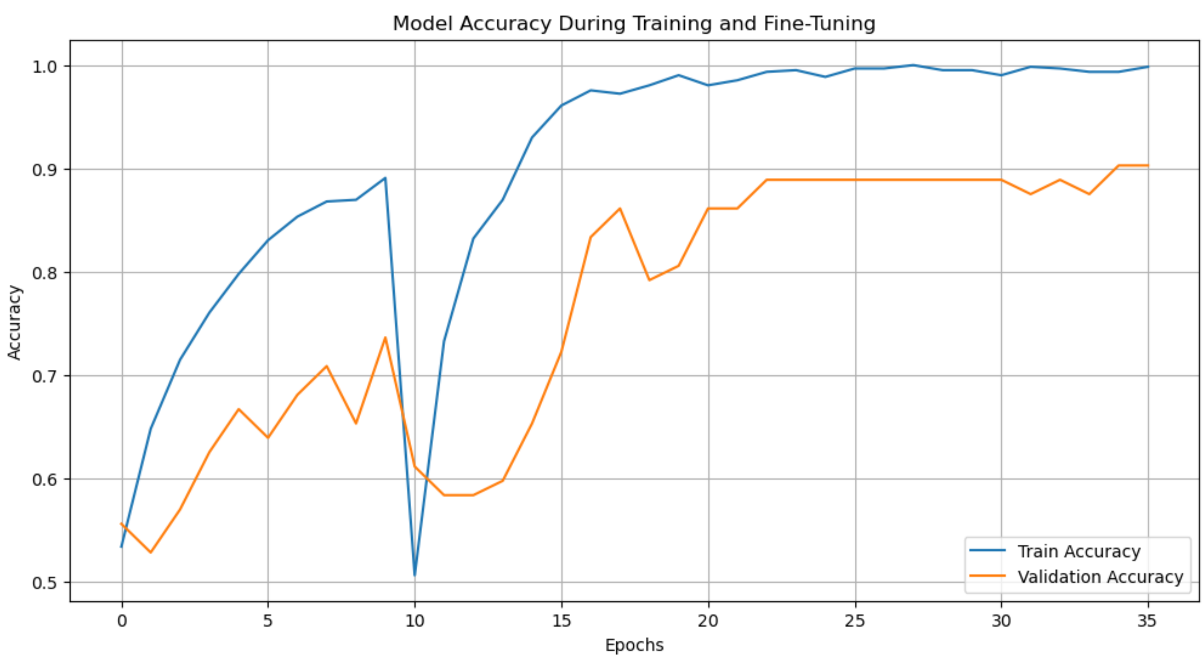

Figure 15a: Xception Model Accuracy on Chest CT Scans Dataset

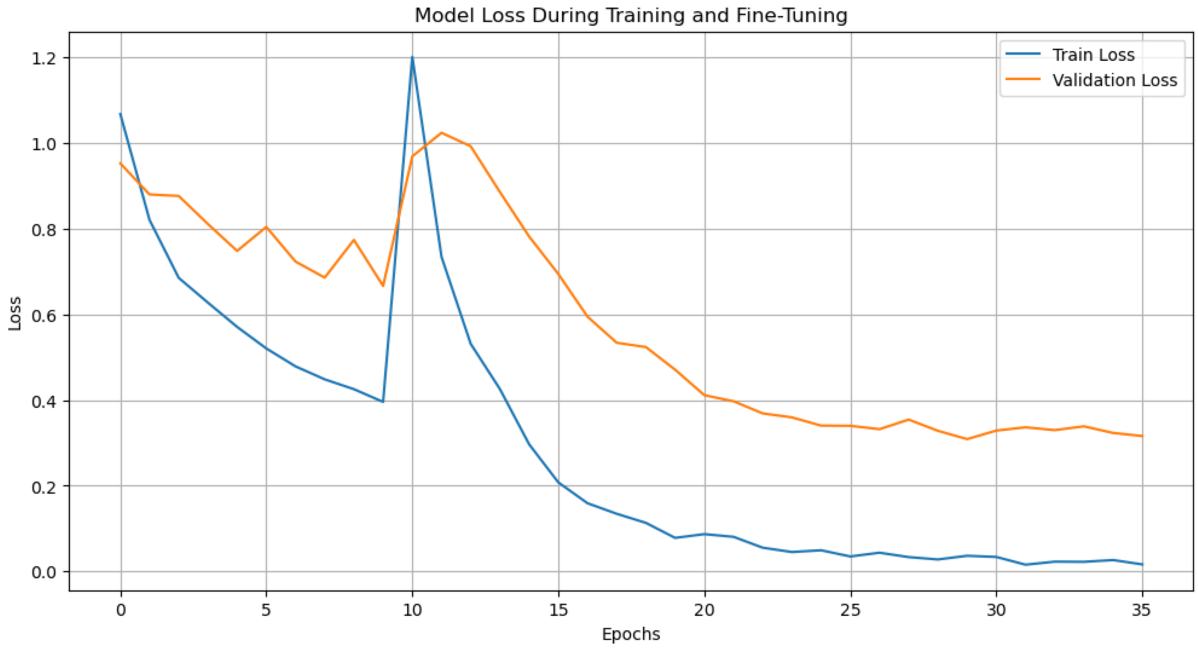

Figure 15b: Xception Model Loss on Chest CT Scans Dataset

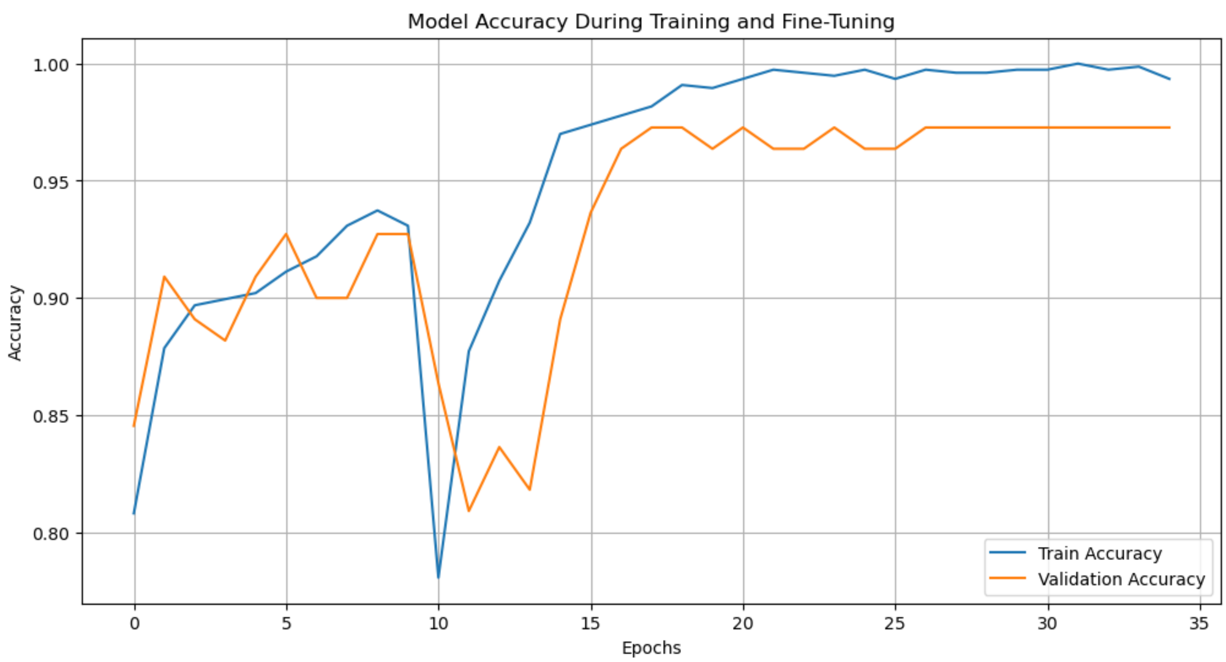

Figure 16a: Xception Model Accuracy on IQ-OTH/NCCD Dataset

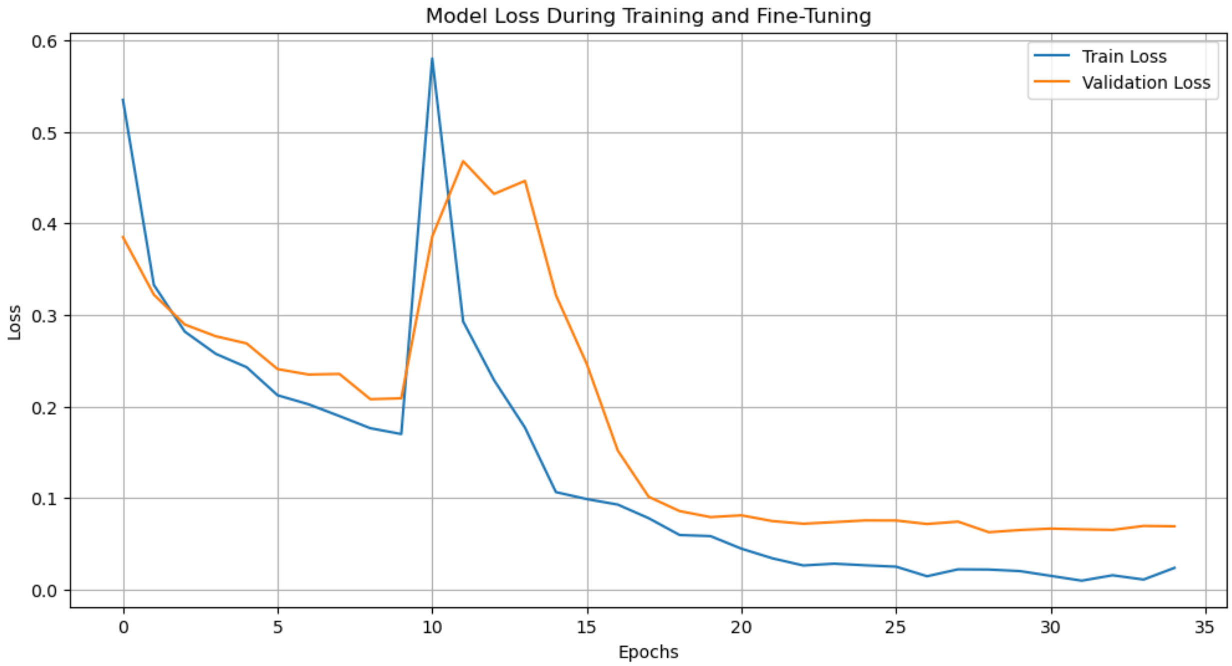

Figure 16b: Xception Model Loss on IQ-OTH/NCCD Dataset

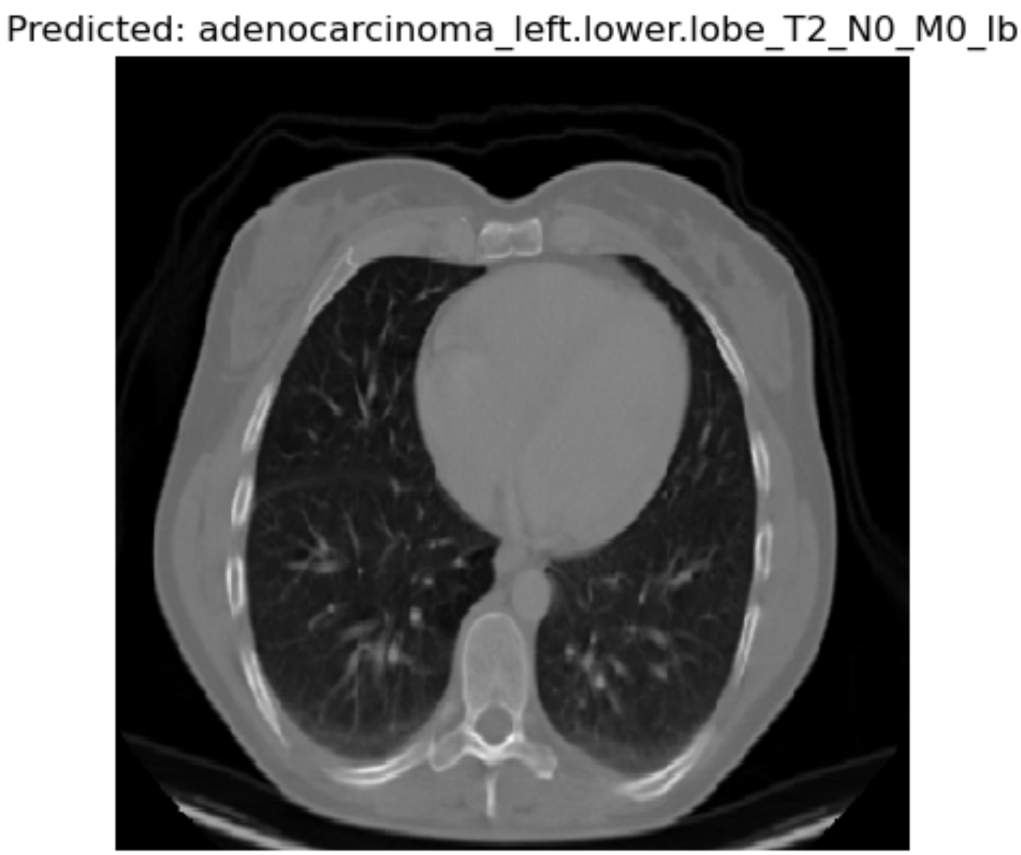

Figure 17a: Test Case 1

Figure 17b: Test Case 2

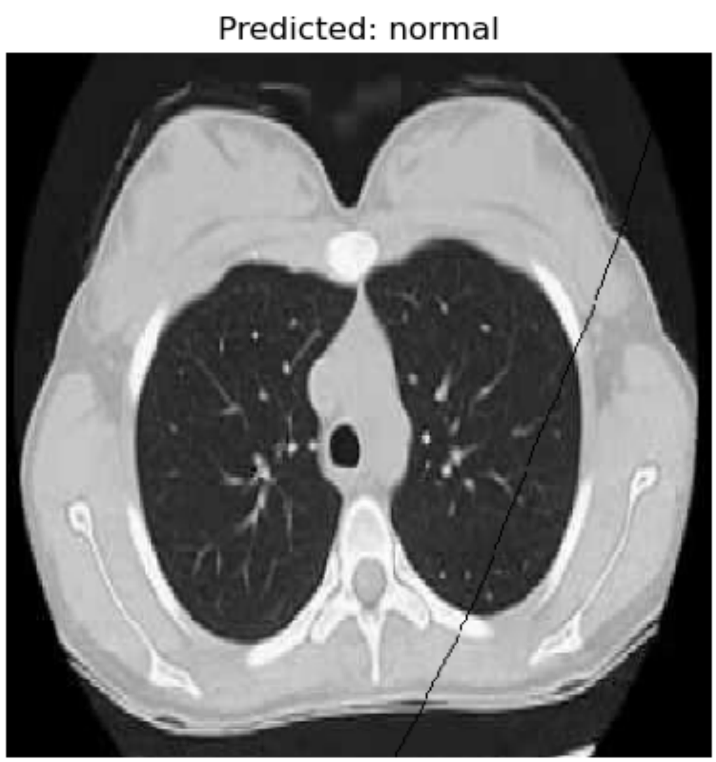

Figure 17c: Test Case 3

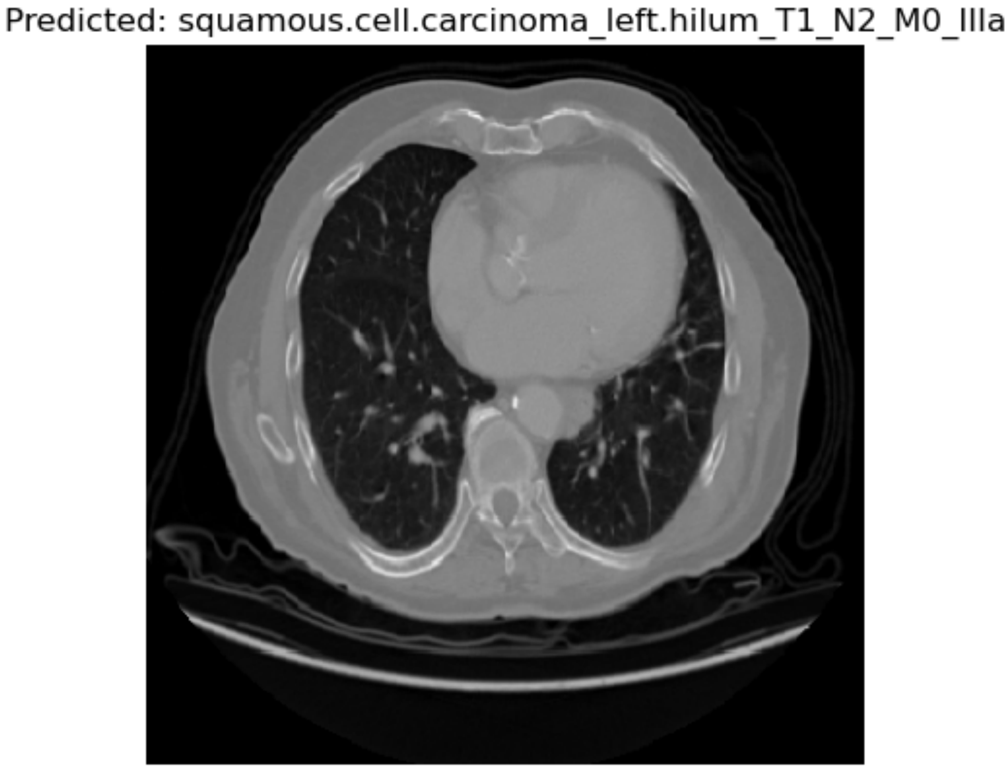

Figure 17d: Test Case 4

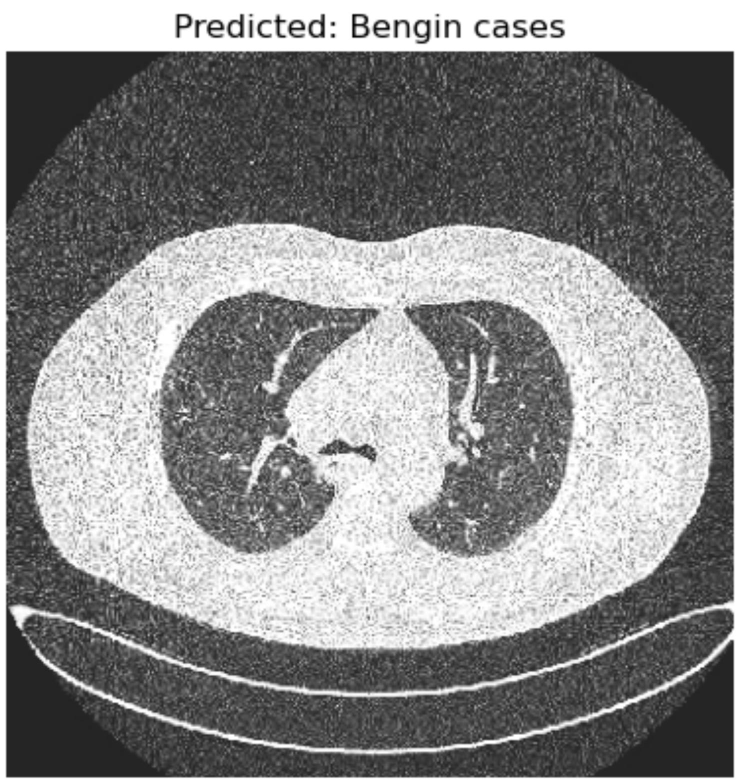

Figure 18a: Test Case 5

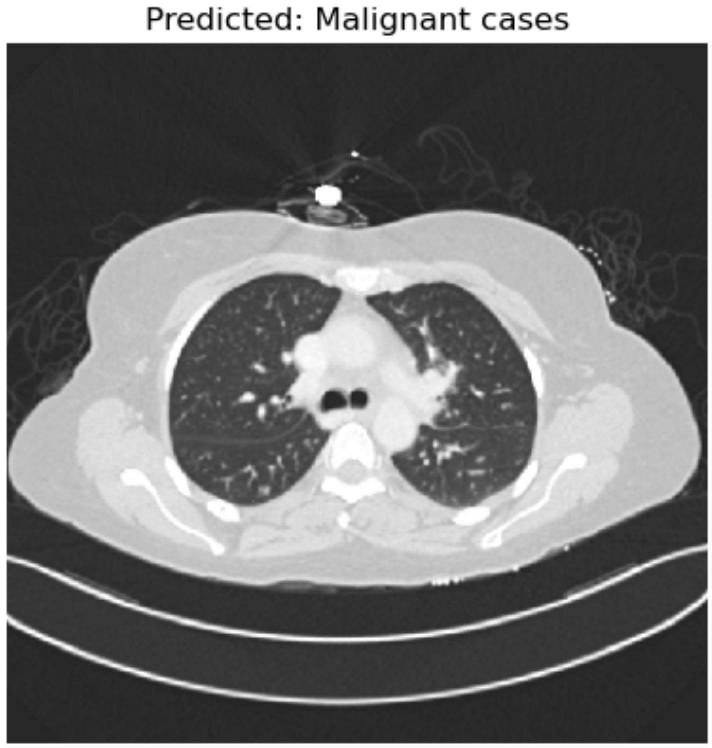

Figure 18b: Test Case 6

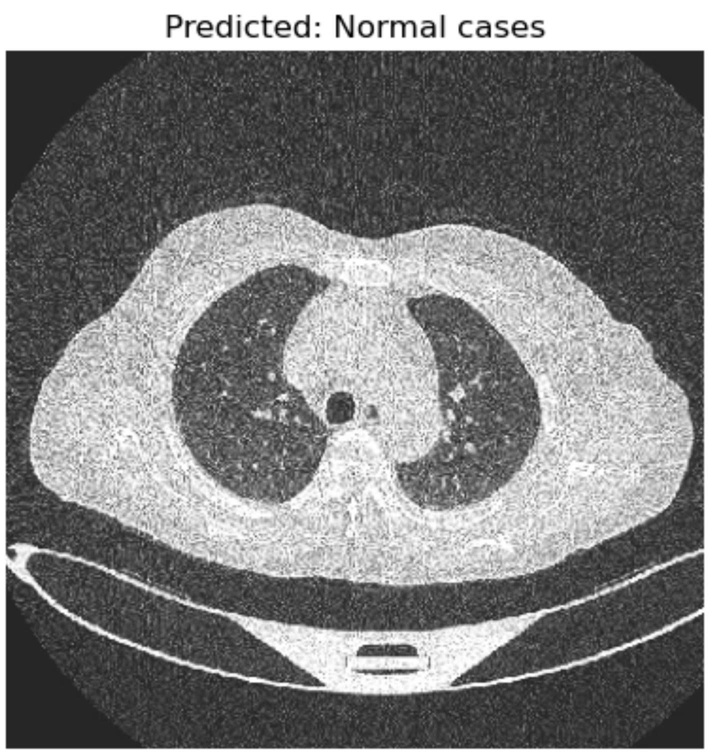

Figure 18c: Test Case 7

Table. 1: CNN Model Classification Report

	Precision	recall	f1-score	support
0	1.00	0.92	0.96	12
1	0.97	1.00	0.98	56
2	0.97	0.95	0.96	41
accuracy			0.97	109
macro avg	0.98	0.96	0.97	109
weighted avg	0.97	0.97	0.97	109

.

Table 2: Random Forest Model Classification Report

	Precision	recall	f1-score	support
0	1.00	0.83	0.91	12
1	1.00	1.00	1.00	56
2	0.95	1.00	0.98	41
accuracy			0.98	109
macro avg	0.98	0.94	0.96	109
weighted avg	0.98	0.98	0.98	109

In table 2, the classification report for the Random Forest algorithm of the hybrid CNN-RF model

Table 3: Xception Model Classification Report

	Precision	recall	f1-score	support
adenocarcinoma	0.88	0.88	0.88	120
large.cell.carcinoma	0.85	0.88	0.87	51
normal	1.00	0.96	0.98	54
squamous.cell.carcinoma	0.91	0.91	0.91	90
accuracy			0.90	315
macro avg	0.91	0.91	0.91	315
weighted avg	0.91	0.90	0.91	315

Table 4: Model Comparison

Model	Training Accuracy	Testing Accuracy
U-NET	99.95%	99.97%
DenseNet	99.05%	98.80%
CNN-RF	99.74%	97.25%
Transfer Learning (Xception)	99.93%	97.29%

Table 5: Comparing Proposed Models to Existing Models in Literature

Published Work	Dataset	Accuracy	Precision	F1 Score	Recall
Multi-Layer Perceptron **[16]**	LIDC-IDRI	98.7%	-	-	-
KNN **[16]**	LIDC-IDRI	99.6%	-	-	-
SVM **[17]**	LIDC-IDRI	80%%	90%	69%	57%
Random Forest **[17]**	LIDC-IDRI	70%	89%	62%	47%
ANN **[17]**	LIDR-IDR	96%	92%	81%	69%
Hybrid CNN-SVM **[18]**	Chest CT Scans	96.71%	97.96%	-	-
CNN **[19]**	LIDC-IDRI	76%	-	-	-
Deep Belief Network **[20]**	LIDC-IDRI	79%	-	-	-
VGG19 with Transfer Learning **[21]**	IQ-OTH/NCCD	90%	43%	42%	42%
AlexNet CNN **[22]**	IQ-OTH/NCCD	93.548%	97.10%	96.40%	-
3D CNN **[23]**	NLST Dataset	94.4%	-	-	-
3D Deep CNN **[24]**	LUNA16	98.5	-	-	-
Proposed CNN Model using U-NET	**LUNA16**	**99.95%**	**-**	**-**	**-**
Proposed CNN using DenseNet	**LUNA16**	**99.05%**	**98.72%**	**98.06%**	**97.40%**
Proposed Hybrid CNN-RF Model	**IQ-OTH/NCCD**	**99.74%**	**98%**	**98%**	**98%**
Proposed CNN using Transfer Learning with Xception	**Chest CT Scans Dataset IQ-OTH/NCCD**	**99.34%**	**91%**	**91%**	**90%**

Appendix B

The artifacts submitted, in agreement with the supervisor are listed below:

- Datasets
- Code
- Models
- Project Framework Diagram
- Signed Ethics Approval Form
- Link for RAW LUNA16 Dataset (Size of the dataset of 120 GB)