Detection of Lung Diseases through Medical Imaging using Deep Learning

Author

Mudassar Ahmed

Registration Number 00000329828

Supervisor

Dr. Kashif Javed

DEPARTMENT OF ROBOTICS AND ARTIFICIAL INTELLIGENCE

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

OCTOBER, 2023

Detection of Lung Diseases through Medical Imaging using Deep Learning

Author

Mudassar Ahmed

Registration Number

00000329828

A thesis submitted in partial fulfillment of the requirements for the degree of MS

Robotics and Intelligent Machine Engineering

Thesis Supervisor:

Dr. Kashif Javed

Thesis Supervisor's Signature: 

DEPARTMENT OF ROBOTICS AND ARTIFICIAL INTELLIGENCE

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

OCTOBER, 2023

THESIS ACCEPTANCE CERTIFICATE

It is certified that final copy of MS/MPhil thesis written by Mudassar Ahmed (Registration No. 00000329828), of SMME (National University of Sciences and Technology) has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations/MS Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

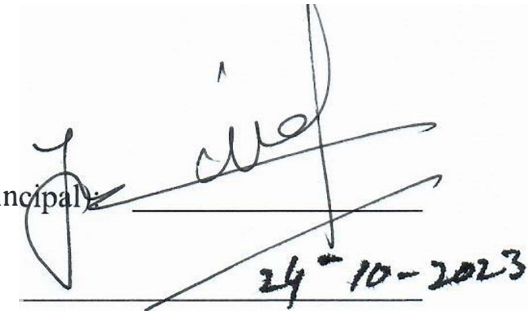
Signature of Supervisor: 

Name: Dr. Kashif Javed

Date: 23-10-2023

Signature of HoD: 

Date: 24-10-2023

Signature (Dean/Principal): 

Date: 24-10-2023

Certificate for Plagiarism

It is certified that PhD/M.Phil/MS Thesis Titled Detection of lungs diseases through medical imaging using Deep Learning by Mudassar Ahmed has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b. The work presented is original and own work of the author (i.e., there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c. There is no fabrication of data or results which have been compiled/analyzed.
d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor
Dr. Kashif Javed

Signature **:** _____

Declaration

I certify that this research work titled “Detection of Lung Diseases through Medical Imaging using Deep Learning” is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged/referred.

Signature of Student

Mudassar Ahmed

00000329828

Copyright Statement

- Copyright in the text of this thesis rests with the student author. Copies (by any process) either in full or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of NUST School of Mechanical & Manufacturing Engineering (SMME). Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights that may be described in this thesis is vested in the NUST School of Mechanical & Manufacturing Engineering, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the SMME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST School of Mechanical & Manufacturing Engineering, Islamabad.

Acknowledgments

I am thankful to my Creator Allah Subhana-Wattala for guiding me throughout this work at every step and for every new thought that Allah set up in my mind to improve it. Indeed, I could have done nothing without Allah's priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Allah will, so indeed none be worthy of praise but Allah.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout every department of my life.

I would also like to express special thanks to my supervisor Dr. Kashif Javed for his help throughout my thesis and for the Computer Vision course that he has taught me. I can safely say that I haven't learned any other engineering subject in such depth as the ones that he has taught. He helped me every time. whenever I got stuck in something; he came up with the solution. Without his help I wouldn’t be able to complete my degree.

I would also like to pay special thanks to my Co-Supervisor Dr. Sara Ali for her tremendous support and cooperation. Without her help, I wouldn’t have been able to complete my thesis. I appreciate her patience and guidance throughout the whole thesis.

I would also like to thank my evaluation committee members Dr. Khwaja Fahad Iqbal and Dr. Muhammad Sajid for being on my thesis guidance and for their support and cooperation throughout my course work and research work.

Finally, I am very thankful to my parents for supporting and loving me throughout my life journey. I would also like to express my gratitude to all the individuals, especially to MS scholar Engr. Mateen Ahmed who has rendered valuable assistance to my study.

Dedicated to my exceptional parents, my wife, and my adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Lung diseases are a leading cause of death worldwide, and early detection is crucial for effective treatment. Two commonly used imaging techniques for detecting lung diseases are Computed Tomography (CT) and X-rays. However, interpreting these scans can be time-consuming and subjective, requiring significant expertise and experience. Deep Learning techniques have shown potential in automating the detection and diagnosis of lung diseases through CT and X-ray scans. This approach involves training models on large datasets of scans and associated disease labels, allowing the models to learn and identify patterns and features indicative of various lung diseases. By using these models, clinicians can obtain more accurate and efficient diagnoses, which leads to improved patient outcomes. These techniques can potentially revolutionize lung disease detection and diagnosis, making them more efficient, accurate, and accessible. In this work, the automatic detection of lung disease named COVID-19 is performed using Deep Learning-based Vision Transformers. This research is conducted on the CT Scans dataset. The pre-trained model ViT-Base-Patch16-224-in21k on the ImageNet dataset is used. Data augmentation and transfer learning are applied to our CT-Scan dataset to attain higher accuracy. The proposed model has achieved the highest training and validation accuracies, which are 99.65% and 99.20% respectively.

Key Words: *Lung Disease, COVID-19, Deep Learning, Vision Transformer, Transfer Learning, ViT-Base-Patch16-224-in21k, Data Augmentation*

Table of Contents

List of Figures

List of Tables

List of Equations

List of Acronyms

Artificial Intelligence	AI
Artificial Neural Network	ANN
Batch Size	BS
Computer Vision	CV
Convolutional Neural Network	CNN
Data Split	DS
Deep Learning	DL
Gaussian Error Linear Unit	GELU
Layer Normalization	LN
Machine Learning	ML
Multilayer Perceptron	MLP
Multiheaded Self Attention	MSA
Vision Transformer	ViT
Vision Transformers	ViTs

CHAPTER 1: INTRODUCTION

1.1. Background and Motivation

Lung diseases are a leading cause of death worldwide [1]. According to the World Health Organization (WHO), there were more than 683 million confirmed COVID-19 cases worldwide and 6.8 million deaths as of now in 2023 [2]. The goal of this research is the early detection of lung diseases that make the treatment process easier and more effective. The method used for this technique is applied to CT scans. The word CT is Computerized Tomography (CT) which is an advanced and expensive imaging modality that consists of a sequence of X-ray radiology images captured at different angles around the body. The two commonly used imaging techniques for detecting lung diseases are Computed Tomography (CT) and X-rays. However, interpreting these scans can be time-consuming and subjective, requiring significant expertise and experience. Deep Learning techniques have shown potential in automating the detection and diagnosis of lung diseases through CT and X-ray scans. This approach involves training models on large datasets of scans and associated disease labels, allowing the models to learn and identify patterns and features indicative of various lung diseases. By using these models, clinicians can obtain more accurate and efficient diagnoses, which leads to improved patient outcomes. These techniques can potentially revolutionize lung disease detection and diagnosis, making them more efficient, accurate, and accessible.

1.2. Problem Statement

Lung Disease is the cause of 1.8 million deaths per year worldwide [1]. According to the World Health Organization (WHO), lung diseases account for a significant number of deaths worldwide.

Coronavirus Disease of 2019 (COVID-19): As of March 2023, there have been more than 761 million confirmed cases of COVID-19 worldwide, with over 6.8 million reported deaths [4]. That’s the reason, the importance of an automated system is required for fast lung disease detection. The automated system will be simpler to scan and automatically identify diseases in a CT scan [4]. Traditional machine learning architectures were effective at diagnosing a patient with disease, but their accuracy is limited to some extent with too much resource consumption. Lung disease detection from CT-scan requires the highly skilled radiologist to manually find abnormalities, the manual approach uses more manpower and economic support which is an

inefficient method prone to mistakes or undetectable diagnosis at the early stages of the disease. To avoid these problems the deep learning-based structural method of feature extraction is used, deep learning method will automatically diagnose and categorize the disease. In addition to this, the deep learning method has shown promising results in the medical and health sector, assisting more radiologists and doctors in the hospitals, dealing with CT-scan is necessary for lung disease detection without these scans diagnosis is impossible.

Many CNN-based architectures were used by different researchers for the recognition and classification of lung disease, e.g., DenseNet201 CNN used by Aayush Jaiswal et al. [5], Resnet50 CNN by Kelvin Leonardi Kohsasih et al, POA-LSTM CNN by R. Vinothini et al were used previously for classification of lung diseases.

The performance of the deep learning models is solely dependent on the dataset used for the training. As compared to the CNN model the Vision Transformers provide more generalizable results mid to large datasets. However, the lung disease used for COVID-19 malignancy is very small compared to other publicly available datasets. Because of these reasons, firstly it is resolved by enhancing the dataset using different image processing techniques including data augmentation. Secondly, the model used should be trained through a large dataset and fine-tuned on the required dataset using transfer learning techniques.

1.3. Objective

The very first objective of this research is to use deep learning model designing techniques to classify lung disease precisely and highly accurately. One of the profound Machine Learning (ML) methods that is currently used by Deep Learning (DL) researchers is Transfer Learning (TL), which transfers or translates the weights of pre-trained models to a new classification outcome by optimizing computer resources. As a beneficial result, training of model becomes simpler and effective. With the given data for training, the main goal of this research is to use the concept of transfer learning to resolve the problem of CT scan classification.

The significant contributions this work brings to the field of research are the following:

1. A pre-trained model is employed to train the proposed model using data augmentation.
2. Upon using the proposed technique, a more accurate, precise, and early detection of COVID-19 will become possible.

3. The effectiveness of this model is evaluated by comparing this technique with previous techniques used for COVID-19 classification.

1.4. Thesis Overview

The chronological breakdown of this article is in the following order:

1. Chapter 2 shows the relevant work done by other researchers.
2. Chapter 3 describes the methodology and implementation employed in this research.
3. Chapter 4 will explain the experimental results obtained by this research.
4. Chapter 5 contains the conclusion of this work.
5. At last, Chapter 6 describes the future work possibility in research performed.

CHAPTER 2: LITERATURE REVIEW

For decades, researchers have been trying to obtain diagnosis techniques for lung diseases. With the advent of technology in CT scans, X-rays, and Machine Learning (ML) models now it's possible to diagnose lung diseases beforehand. Numerous approaches already exist for identifying and classifying lung diseases. These approaches were proposed by researchers in the domain of computer vision and machine learning. The recent death toll due to COVID-19 has given insignificant importance to the diagnosis of lung diseases in advance. To protect the lungs from diseases, a study of numerous techniques for the identification and classification of lung diseases is described below.

EGFR mutation status detection for Lung Adenocarcinoma treatment was performed by Shijie Zhao et al. [3] by using many techniques but the Densformer-based Vision Transformers (ViTs) technique achieved a higher accuracy of 80.7% among other methods.

The work related to the severity level of COVID-19 infection was performed by Mercy Ranjit et al. [4] which gives three levels of severity information such as mild, moderate, and severe. The mechanism used in that study was a space-time transformer-based attention mechanism with an accuracy of 93.3% at the Sequence Level and 99% accuracy at the Patient Level.

Among other lung diseases, a severe disease is lung cancer. It is usually diagnosed at that stage when the treatment becomes very difficult and redeems chances of survival. A study carried out by Shweta Tyagi et al. [5] differentiates between the improvement of a patient after and before the treatment by using the DL technique a combination of Vision Transformers (ViTs) and CNN for segmentation with an accuracy of 75.61%. Another study on lung cancer was carried out by Khalil Barbouchi et al. [6] which used the DETR Transformer model to assist physicians with staging information of lung cancer by employing the TNM staging system and histologic subtype classification.

Pulmonary Tuberculosis (TB) is also a type of lung disease, Junlin Tian et al. [7] used a lightweight model that consist of fussed CNN & Transformers architecture for TB disease classification and had achieved classification accuracy of 97.23%.

Previous work on COVID-19 using deep convolutional neural networks that led to our work on this disease is discussed below.

Classification of COVID-19 performed by Aayush Jaiswal et al. [8] has achieved 97% accuracy by using the DenseNet201 transfer learning-based model.

Kelvin Leonardi Kohsasih et al. [9] also worked on COVID-19 classification by employing a Resnet-50-based CNN network with a reported accuracy of 95%.

Yet another COVID-19 classification performed by R. Vinothini et al. [10] had achieved one of the highest accuracies of 98.89% and proposed the pelican optimization algorithm-based long-short-term memory (POA-LSTM) model for COVID classification using CT-Scans.

Last but not least, Dilbag Singh et al. [11] used multi-objective differential evolution–based convolutional neural networks and reported an accuracy of 93.2% for COVID-19 infection classification.

CHAPTER 3: METHODOLOGY AND IMPLEMENTATION

Vision Transformers have gained increasing popularity in recent years for automatic disease detection so has for lung diseases as well. A brief overview related to the proposed framework of the deep learning (DL) model and the use of image processing techniques for the detection and classification of COVID-19 disease in lung diseases is given below. **Figure 3.*1*** shows the general schematic diagram of our proposed deep learning model, it includes the input CT-scan dataset stage, the preprocessing & data augmentation stage, the DL model stage, Transfer Learning (TL) stage, the disease classification stage, and model performance assessment stage.

The proposed methodology consists of four main modules:

- Dataset Collection
- Data Augmentation and Image pre-processing
- Data Splitting
- Proposed Architecture

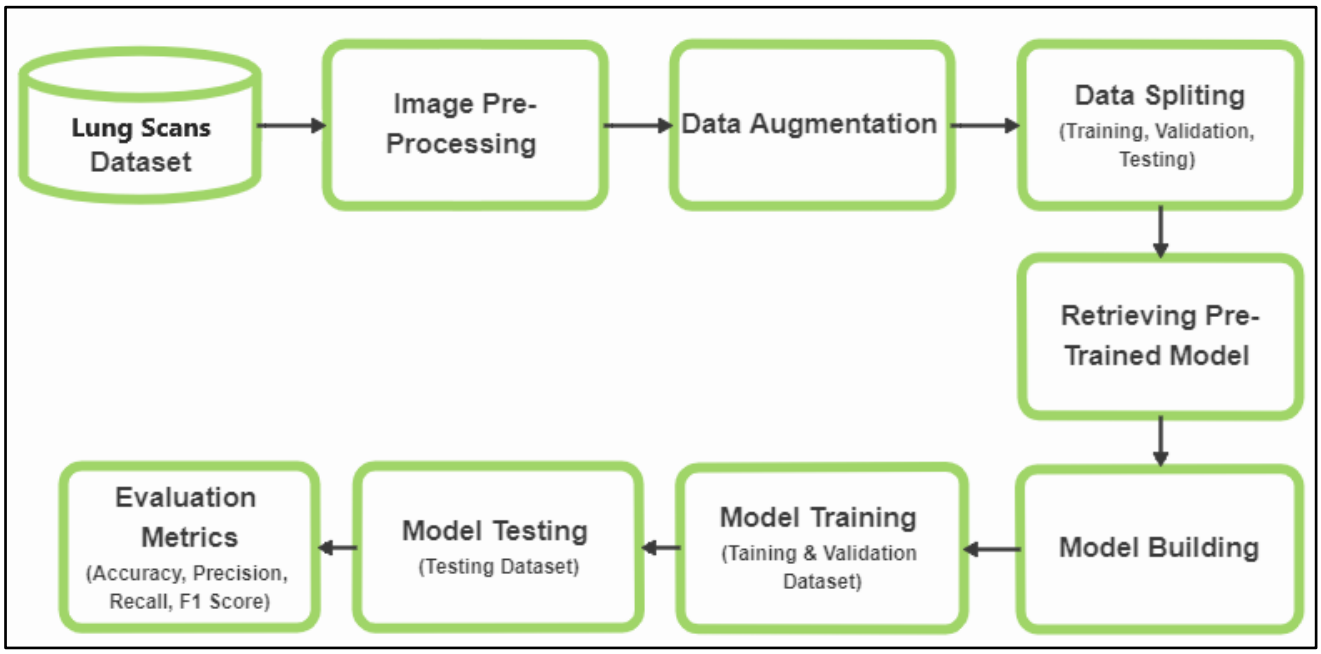


Figure 3.1: General Schematic of Proposed System

3.1. Dataset Collection

Deep learning algorithms require a dataset that contains tens of thousands of images to effectively learn features for specific classes. In this work, SARS-CoV-2 dataset of CT-Scans [12] for COVID-19 lung disease is employed. This dataset includes 2482 scans of lung disease, both COVID and non-COVID. All scans are 342 pixels wide and 254 pixels in height. The dataset contains separate folders for COVID-19 disease samples and non-COVID samples of CT scans. **Table 3.*1*** shows the details of the dataset for healthy and unhealthy scans. The samples of COVID disease and the samples of non-COVID are shown in **Figure 3.*2*** and **Figure 3.*3***.

Table 3.1: COVID-19 Dataset

Lung Disease	Images	Males	Females
COVID	1252	32	28
Non-COVID	1230	30	30

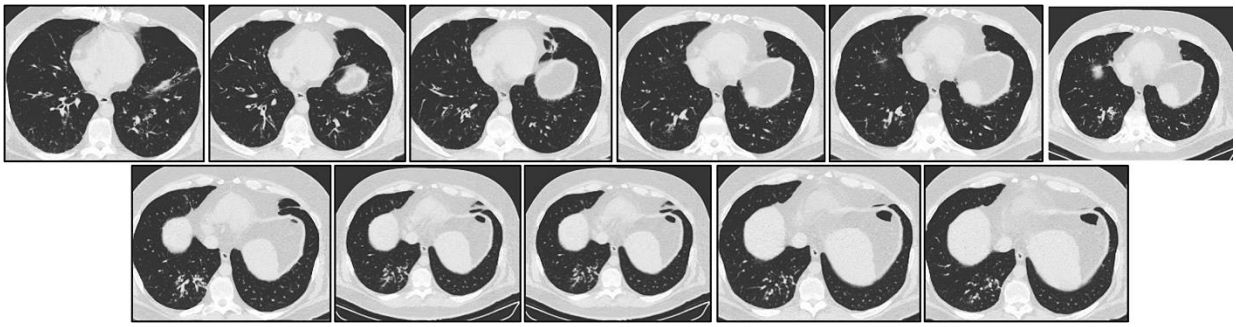


Figure 3.2: Lung Disease CT-Scans

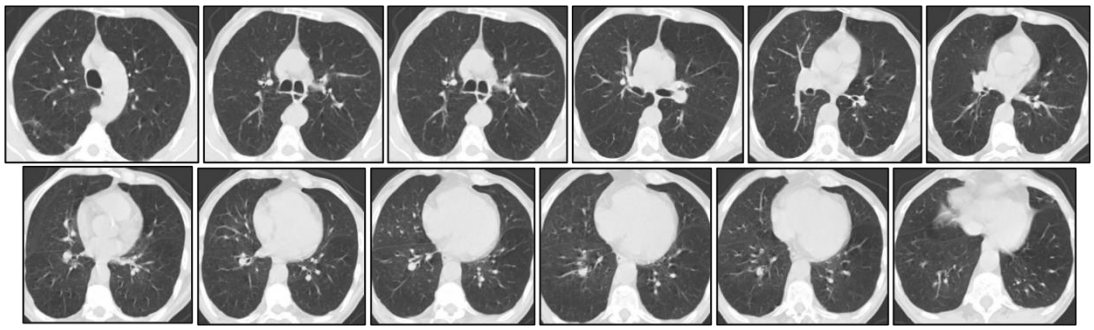


Figure 3.3: Non-COVID-19 Lung CT-Scans

3.2. Dataset Augmentation and Resizing

The dataset augmentation helps in the effective training of deep neural networks including Vision Transformers (ViTs), and a large amount of dataset of images required for training purposes. Unfortunately, the non-annotated ground truths, the amount of currently available lung disease scan collections, and other factors make automatic diagnosis difficult for lung diseases. The Vision Transformers-based deep neural network requires a huge quantity of training data during the training stage. These requirements are fulfilled by performing augmentation operations on the training dataset to increase the training scans. A set of data augmentation techniques are used for a variety of effects including Center Crop, Random Horizontal Flip, Resize, and Rotate. Data augmentation is also used for generating multiple copies of an image in various settings using the above-mentioned techniques as shown in **Figure 3.*4***. Initially, 2,482 original CT-Scan images were available. After data augmentation a total of 6,720 images of CT-Scans are produced as listed in **Table 3.*2***.

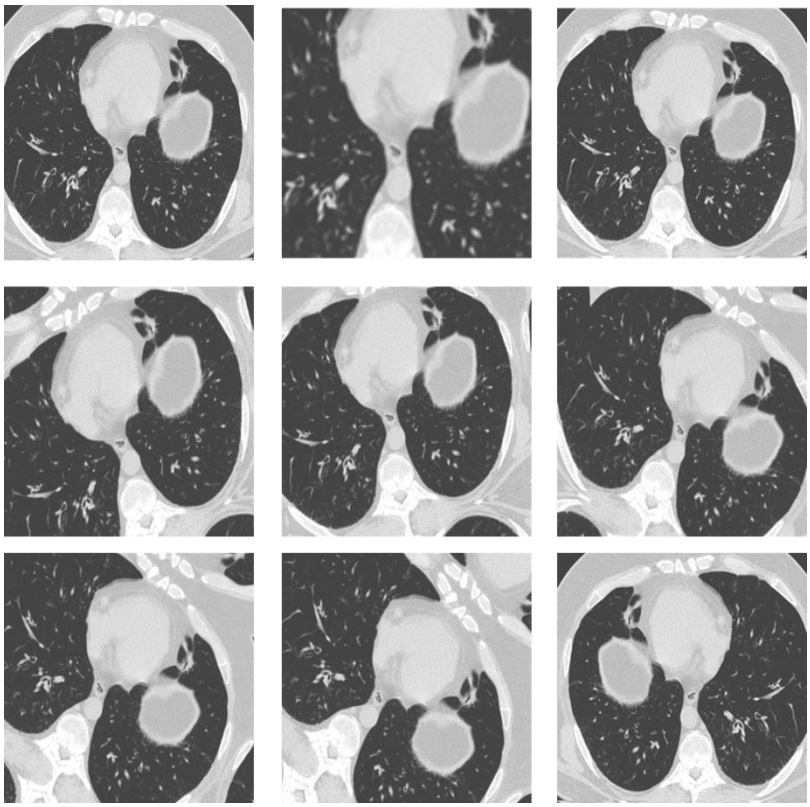


Figure 3.4: Data Augmentation

Table 3.2: Augmented COVID Dataset

Categories	Diseases	Original images	After Data Augmentation
Lung	COVID	1252	3390
	Non-COVID	1230	3330
Total Images		2482	6720

3.3. Dataset Distribution

The dataset is divided into three subsets as depicted in **Figure 3.*5***:

(i) Training Data,
(ii) Validation Data
(iii) Testing Data.

Training, Validation, and Testing are done in a batch size of 10 and data split (80, 10, 10) respectively.

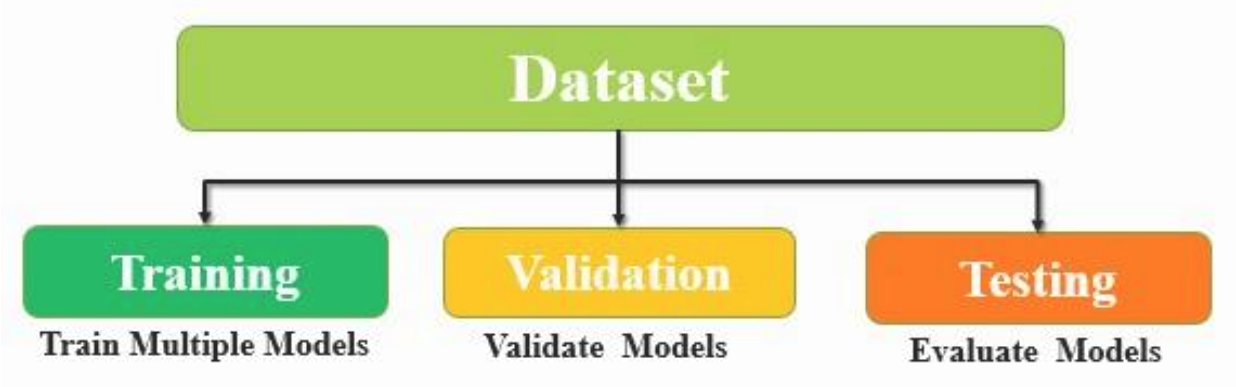


Figure 3.5: Dataset Distribution

3.3.1. Training Set (A Subset of Dataset)

A Vision Transformer-based deep neural network model is built that utilizes 80% of the original dataset specified for training, this ratio could be altered as per the requirements of the model. The training dataset contains labeled images that correspond to the desired output. The data augmentation is performed on this subset of the dataset.

3.3.2. Validation Set (A Subset of Dataset)

The validation subset of the dataset is 10% of the original dataset and it is used to validate the selected model performance during training. The result obtained from this validation process is used to modify the model hyperparameters and settings. To avoid the overfitting problem a split of dataset named validation dataset is used.

3.3.3 Test Set (A Subset of Dataset)

The proposed model is finally assessed on unseen data using a test set that contains a 10% split of the original dataset. Once the model is trained successfully, it is available for the evaluation process. Based on this test set the model is further processed using evaluation metric in terms of accuracy, precision, recall, and F1-Score that concludes the model performance. In simple words, it shows the response to the question “How effective is the working of a model with unseen data?"

3.4. Proposed Architecture

This study proposes a unique approach to identifying and classifying lung disease. The fundamental architecture consists of five main steps: (a) data acquisition (b) image Pre-Processing and Data Augmentation (c) Vision Transformer based feature extraction (d) final classification (f) Performance Evaluation. **Figure 3.*6*** shows a detailed flow of the suggested framework.

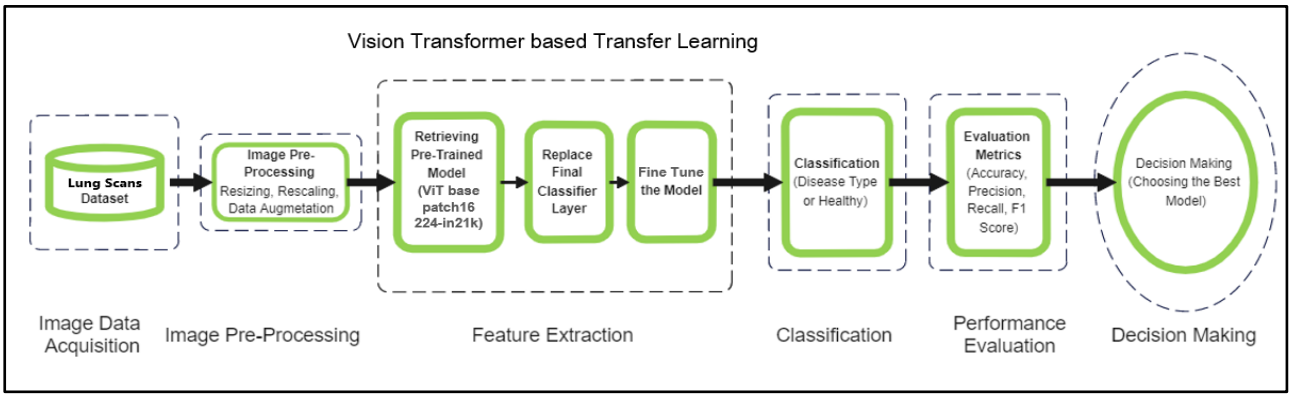


Figure 3.6: Detailed architecture of Transfer Learning based lung disease classification

3.4.1 Feature Extraction Using Vision Transformer-based Transfer Learning

The model used in this work is ViT-Base-Patch16-224-in21k pre-trained Vision Transformer-based architecture by applying Transfer Learning (TL), as shown in **Figure 3.7**.

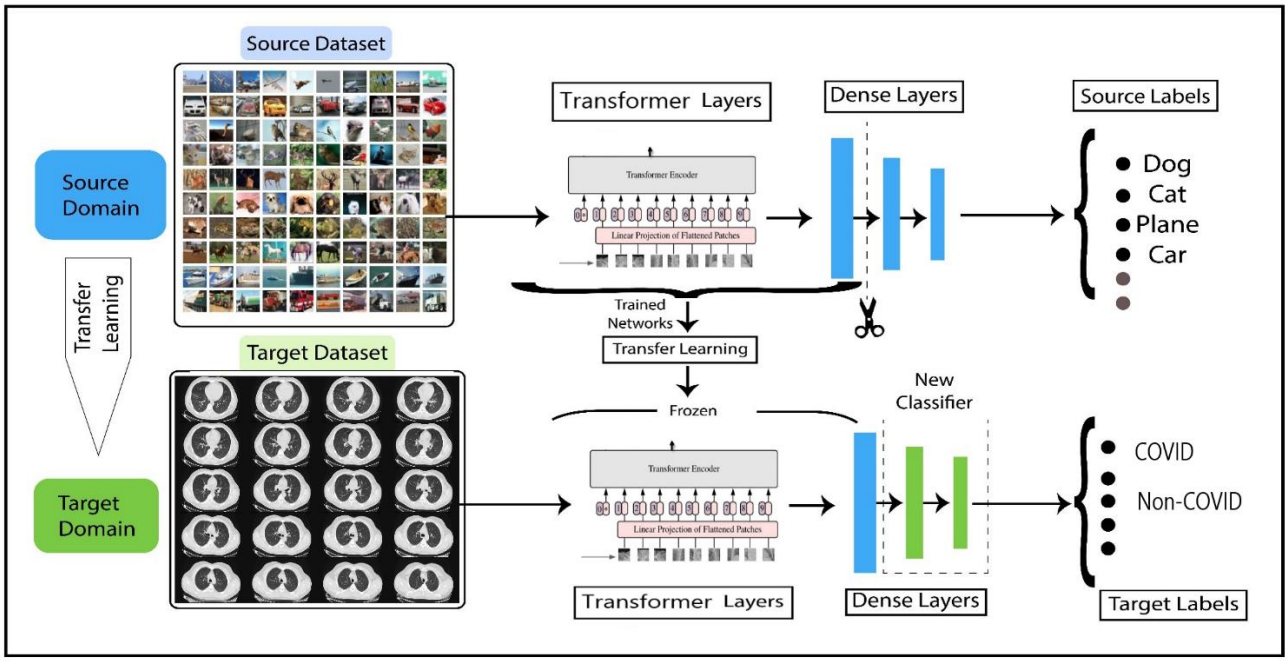


Figure 3.7: Representation of Vision Transformer-based feature extraction using Transfer Learning

3.4.2. Vision Transformer Architecture (ViT-Base-Patch16-224-in21k):

The Transformers model usage as a de-facto standard for NLP had given birth to the application of Transformers in Vision applications as well with the name of Vision Transformer (ViT) [13]. Through this method patches of images were provided to the transformer's encoder after position and patch embedding, further, this transformer encoder is fed to a multilayer perceptron (MLP) head that provides the final classification results. The transformer encoder consists of multiple layers of MLP blocks, MSAs, LNs, and residual connections [13]. The MLP blocks contain GELU as an activation function.

The ViT-Base main objective is to use fewer parameters (86 million) as compared to ViT-Large and ViT-Huge which contain 307 million and 632 million parameters respectively. ViT-Base has 12 layers and ViT-Large and ViT-Huge have 24 and 32 layers respectively [13]. ViT-Base-Patch16-224-in21k is pretrained on ImageNet-21k dataset [14].

The comparison of Vision Transformers versus CNN indicates that Densenet-201 has 201 layers and 20 million learnable parameters and Resnet-50 has over 23 million parameters with 50 layers, but the ViT-Base algorithm has 12 layers with 86 million total parameters and has greater accuracy as compared to CNN architectures. As accuracy is the main concern.

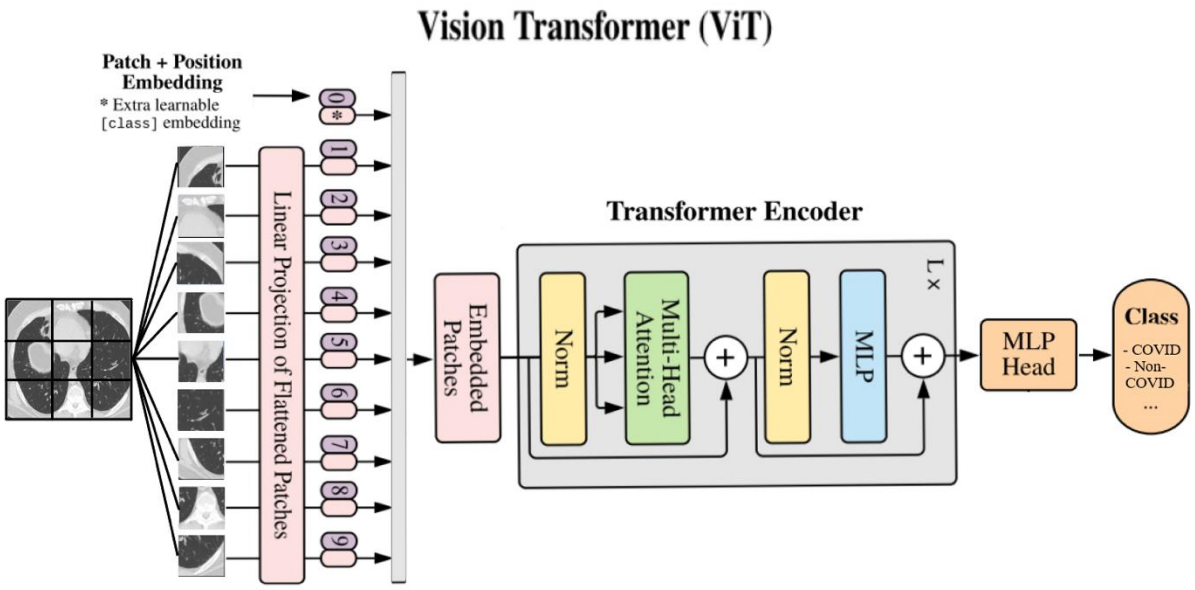


Figure 3.8: Vision Transformer Architecture (ViT-Base)

The Vision Transformer model shown in **Figure 3.*8*** represents the depiction of the actual model, the input patches shown here are 3x3 in size the same pattern is followed by the ViT-Base model using 16x16 patches having L as 12 layers.

The flowchart of the implementation of the ViT-Base transformer explains the flow of the algorithm and depicts the interconnection of different modules as shown in **Figure 3.*9***.

The flowchart describes that at the very beginning of the model, data acquisition is performed that leads toward image pre-processing and data augmentation further this acquired dataset is split into a train, validation, and test sets after this pretrained model parameters are retrieved for application of Transfer Learning (TL) the model is Fine Tuned on the desired dataset. After achieving the desired accuracy, the final classification results are obtained.

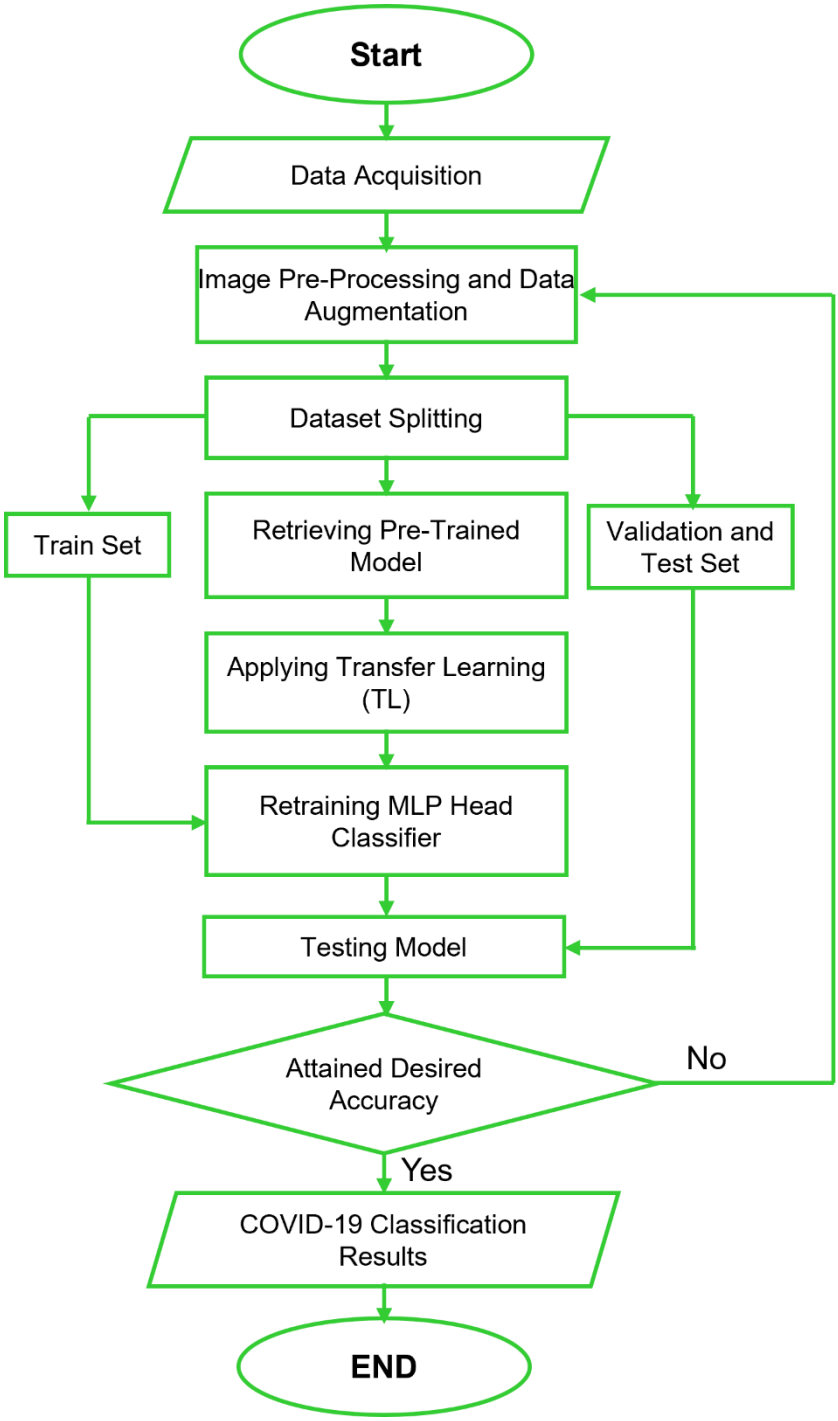


Figure 3.9: Flowchart of the proposed model

The algorithm for the model is enlisted in **Figure 3.*10***. It verbally describes the steps that are taken during the implementation of the model. The details of the model implementation are given as well. Such as batch size, image size, number of classes, and dataset splitting details.

1. ***START***
2. ***Input:*** *COVID-19 Dataset (CT-Scan Image Files, Class Name)*
3. ***Output:*** *Disease Classification with high prediction accuracy.*
4. *BATCH_SIZE = 10, IMAGE_SHAPE = 224, num_classes = 2;*
5. *dataset = load_dataset_from_directory("Path of directory ")*
6. *Define create_dataset_partitions(ds, train_split=0.8, val_split=0.1, test_split=0.1)*
7. *Apply Image Pre-Processing and Data Augmentation on train_split.*
8. *Load Pre-Trained Model*
9. *Freeze classification features learning layers and add a new MLP heads Classifier.*
10. ***FOR*** *All training examples* ***DO***
    Fine-tuned the model with a new MLP head classifier on training data
    Validate the model on val_split
    Test the model with Test Data
11. ***IF*** *the desired accuracy is achieved* ***THEN***
    Show COVID-19 disease classification results.
    Show Table containing Accuracy, Loss, Precision, Recall, F1-Score
12. ***ELSE***
    Go back to step number 7.

END
END

END

Figure 3.10: Algorithm for our Vision Transformer-based model

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION

Experimental results are discussed in this section. The Vision Transformers (ViTs) based pre-trained model ViT-Base-Patch16-224-in21k architecture is used in this experiment through training and validation datasets. These experimental results are used to obtain the accuracy, loss, and performance matrix.

4.1. Model Results

The Experimental results show that ViT-Base-Patch16-224-in21k outperforms all CNN-based models in terms of performance, with a training accuracy of 99.197%. The comparison of the Vision Transformers (ViTs) based model is given in section 4.5.

The accuracy and loss curves obtained during the model learning phase are shown in the graphs of **Figure 4.*1***. The number of epochs used for training and learning rate decay curves are shown in the graphs of **Figure 4.*2***.

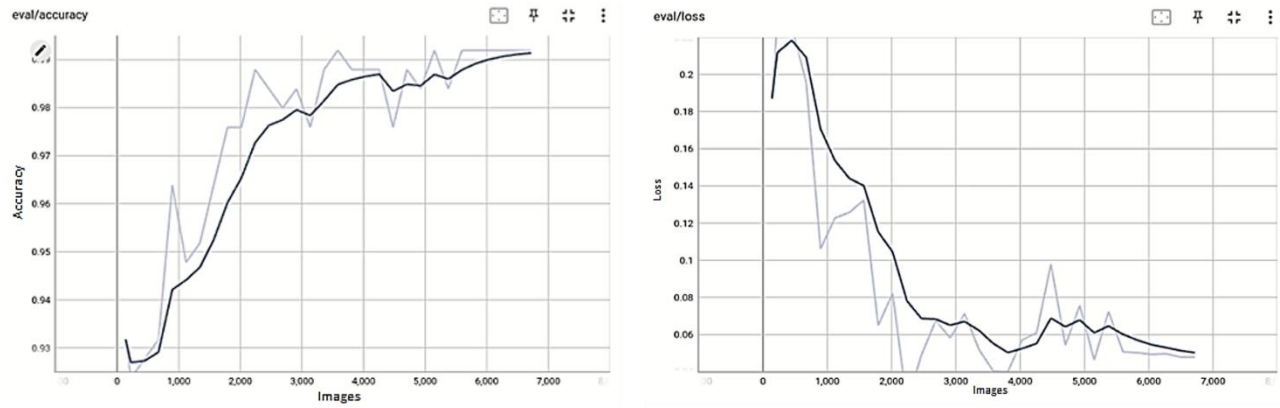


Figure 4.1: Training v/s Validation Accuracy and Loss graph for the ViT-Base-Patch16-224-in21k model

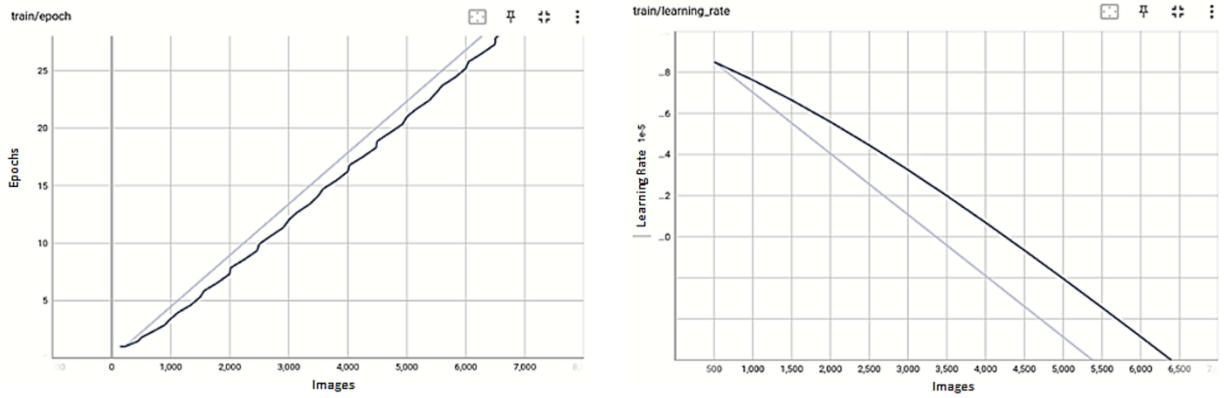


Figure 4.2: Number of Epoch and Learning Rate Decay graph for the ViT-Base-Patch16-224-in21k model

4.2. Model Performance Evaluation:

Table 4.*1* displays the performance evaluation matrix between the trained network and the test dataset, including each model's accuracy, precision, recall, and F1-Score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \textbf{Equation 1}$$

$$Precision = \frac{TP}{TP + FP} \quad \textbf{Equation 2}$$

$$Recall = \frac{TP}{TP + FN} \quad \textbf{Equation 3}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \textbf{Equation 4}$$

Here the TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

The model performance evaluation matrix obtained from the trained model using Equation 1, Equation 2, Equation 3, and Equation 4 is shown in **Figure 4.*3***.

The test results showed that ViT-Base-Patch16-224-in21k performed better than all CNN-based architecture models in regard to accuracy (99.197%), precision (99.22%), recall (99.19%), and F1-Score (99.19%).

Table 4.1: Performance Evaluation of trained model with test data

Pre-Trained Model	Data Augmentation	Precision	Recall	F1- Score	Accuracy (%)
ViT-Base-Patch16-224-in21k	COVID	0.9845	1.0000	0.9922	99.197
	Non-COVID	1.0000	0.9836	0.9917	99.197

	precision	recall	f1-score	support
COVID	0.98450	1.00000	0.99219	127
Non-COVID	1.00000	0.98361	0.99174	122
accuracy			0.99197	249
macro avg	0.99225	0.99180	0.99196	249
weighted avg	0.99209	0.99197	0.99197	249

Figure 4.3: Models Performance Evaluation

4.3. Confusion Matrix for the Model:

The confusion matrix of the ViT-Base-Patch16-224-in21k architectural model is shown in **Figure 4.*4***. The table shown in **Table 4.*1*** is genuinely obtained from the model results. **Figure 4.*3*** and **Table 4.*2*** illustrate the same accuracy, precision, recall, and F1 Score values obtained from this confusion matrix.

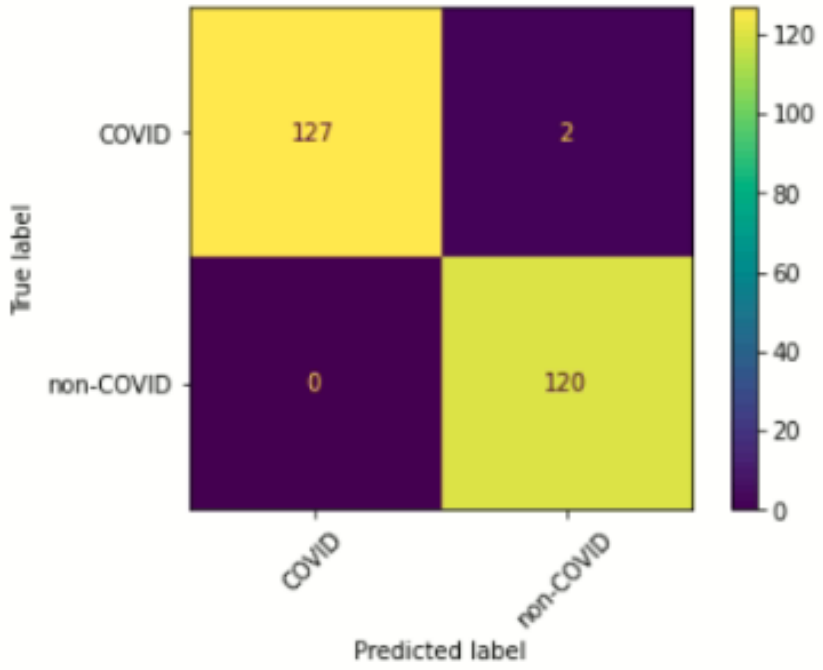


Figure 4.4: Confusion matrix of EfficientNetB3

The ViT-Base-Patch16-224-in21k model that contains Vision Transformers (ViTs) based architecture obtained an accuracy of 99.197%, **Figure 4.*4*** demonstrates the tangible results that the model is the most accurate for the identification and classification of COVID-19 lung disease.

4.5. Proposed Model Comparison

Additionally, the proposed deep learning model comparison with other methods that are used for the classification and diagnosis of COVID-19 lung disease is given below. According to **Table 4.*2*** and **Figure 4.*5***. Based on classification accuracy, the suggested technique outperforms the existing techniques.

Table 4.2: Proposed Model Comparison with previous studies

References	Year	Accuracy (%)
Aayush Jaiswal et al. [8]	2020	97
Kelvin Leonardi Kohsasih et al. [9]	2022	95
R. Vinothini et al. [10]	2023	98.89
Proposed Model	2023	99.19

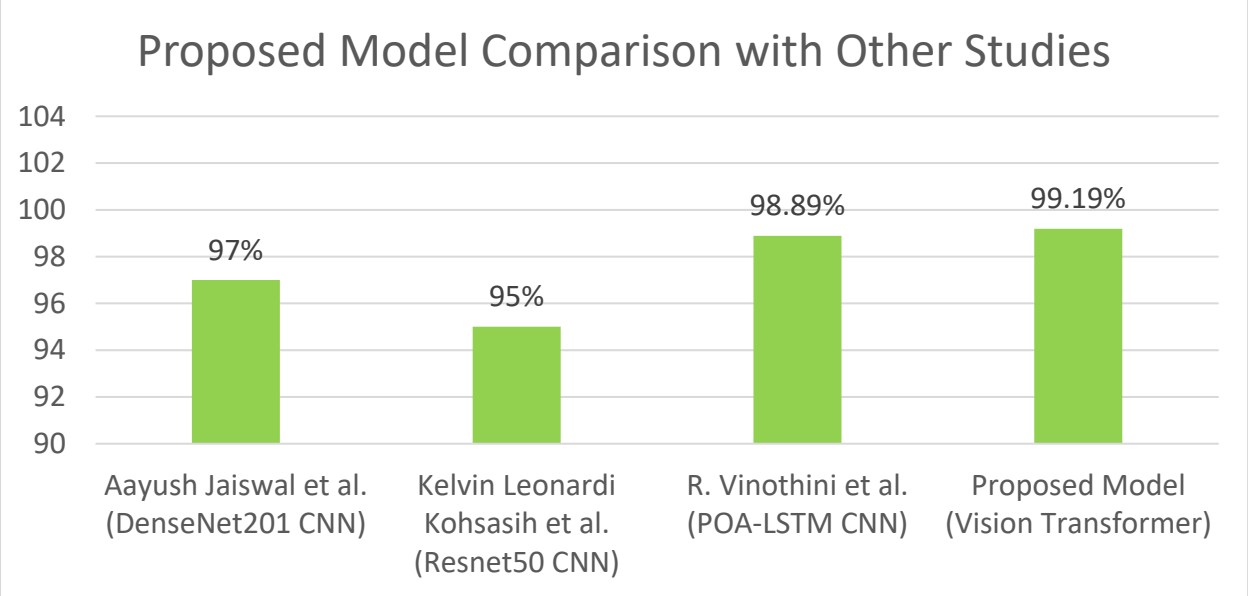


Figure 4.5: Proposed Model Comparison with previous studies

CHAPTER 5: CONCLUSION

In this work, the Implementation of the vision transformer (ViT) model is performed by employing the transfer learning (TL) technique. The improvement in lung disease detection is achieved by accurately identifying COVID-19 lung disease based on the SARS-COV-2 dataset. To achieve the highest accuracy image preprocessing and data augmentation techniques are used. Finally, the vision transformer-based pre-trained model (ViT-Base-Patch16-224-in21k) on ImageNet-21K is employed. The proposed method has achieved training and validation accuracy of 99.5980% and 99.197% respectively on CT scan images. The proposed technique has outperformed previous work in terms of accuracy, sensitivity, recall, and F1-score which are 99.197%, 99.225%, 99.180%, and 99.196% respectively.

CHAPTER 6: FUTURE WORK

The primary challenge in this work is the small dataset, which is reduced by employing the data-augmentation technique. Furthermore, this work is limited to COVID-19 lung disease detection. In future studies, the inclusion of other lung diseases and further data augmentation techniques are intended to be included and it will ultimately cater to the problem of small datasets as well. The proposed pre-trained model was limited to 12 layers. To achieve higher accuracy and performance measures other vision transformer models that consist of large layer models should be used.

REFERENCES

[1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71(1):7-33.

[2] Abenavoli, Ludovico, and Ivan Gentile. "COVID-19: Where We Are and Where We Are Going." Diseases 11.1 (2023): 40.

[3] Zhao, Shijie, et al. "End-to-end Prediction of EGFR Mutation Status with Denseformer." IEEE Journal of Biomedical and Health Informatics (2023).

[4] Ranjit, Mercy, et al. "AUTOMATED SEVERITY SCORING OF COVID-19 CT SEQUENCES USING SPACE-TIME TRANSFORMERS." ICTACT Journal on Image & Video Processing 13.4 (2023).

[5] Tyagi, Shweta, and Sanjay N. Talbar. "Predicting lung cancer treatment response from CT images using deep learning." International Journal of Imaging Systems and Technology (2023).

[6] Barbouchi, Khalil, et al. "A transformer-based deep neural network for detection and classification of lung cancer via PET/CT images." International Journal of Imaging Systems and Technology (2023).

[7] Tian, Junlin, et al. "Lightweight Classification Network for Pulmonary Tuberculosis Based on CT Images." Journal of Artificial Intelligence and Technology 3.1 (2023): 25-31.

[8] Aayush Jaiswal, Neha Gianchandani, Dilbag Singh, Vijay Kumar & Manjit Kaur (2021) Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning, Journal of Biomolecular Structure and Dynamics, 39:15, 5682-5689, DOI:10.1080/07391102.2020.1788642.

[9] Kohsasih, Kelvin Leonardi & Hayadi, B. (2022). Classification SARS-CoV-2 Disease based on CT-Scan Image Using Convolutional Neural Network. Scientific Journal of Informatics. 9. 197-204. 10.15294/sji.v9i2.36583.

[10] Vinothini, R., Niranjana, G. & Yakub, F. A Novel Classification Model Using Optimal Long Short-Term Memory for Classification of COVID-19 from CT Images. J Digit Imaging (2023). DOI:10.1007/s10278-023-00852-7.

[11] Singh, Dilbag, et al. "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks." European Journal of Clinical Microbiology & Infectious Diseases 39 (2020): 1379-1389.

[12] Soares, Eduardo, et al. "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification." MedRxiv (2020): 2020-04.

[13] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[14] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R'emi Louf, Morgan Funtowicz, and Jamie Brew. Huggingface's transformers: State of the art Vision Transformer. ArXivpreprint: 1910.03771,2019.