PULMONARY DISEASE CLASSIFICATION USING RESPIRATORY SOUNDS

A PROJECT REPORT

Submitted by

ALLEN MANOJ (190501015)
JANANI K (190501045)
RITUNJAY M (190501099)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

SRI VENKATESWARA COLLEGE OF ENGINEERING
(An Autonomous Institution; Affiliated to Anna University, Chennai-600025)
ANNA UNIVERSITY, CHENNAI 600 025

MAY 2023

SRI VENKATESWARA COLLEGE OF ENGINEERING
(An Autonomous Institution; Affiliated to Anna University, Chennai-600025)
ANNA UNIVERSITY, CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"PULMONARY DISEASE CLASSIFICATION USING RESPIRATORY SOUNDS"** is the bonafide work of **"ALLEN MANOJ (190501015), JANANI K (190501045) and RITUNJAY M (190501099)"** who carried out the project work under my supervision.

SIGNATURE

Ms. G. R. KHANAGHAVALLE
INTERNAL SUPERVISOR
ASSISTANT PROFESSOR
COMPUTER SCIENCE & ENGG

SIGNATURE

MR. DINESH KOKA
EXTERNAL SUPERVISOR
CEO, ONWARD ASSIST
1999 BATCH ECE ALUMUS, SVCE

SIGNATURE

Dr. R. ANITHA
HEAD OF THE DEPARTMENT
COMPUTER SCIENCE & ENGG

Submitted for the project viva-voce examination held on ……..

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ABSTRACT

Pulmonary diseases, such as asthma and COPD, affect millions worldwide due to factors like infections, smoking, and air pollution. Pulmonary auscultation with a stethoscope has long been a safe and cost-effective diagnostic method. To enhance accuracy and speed, researchers are now exploring digital stethoscopes and machine learning. Using the ICBHI 2017 Dataset, a notable study focuses on developing an 8-class classification system for pulmonary diseases. Data preprocessing involves audio slicing, and feature extraction is done using Short-Time Fourier Transform (STFT) - Wavelet Spectrogram. The audio files are then converted into spectrogram images, fed into pre-trained deep neural networks like ResNet50, VGG16, and EfficientNet-B0. These models learn to classify respiratory sounds and identify different pulmonary diseases. This approach shows promise in reducing the time and cost associated with pulmonary disease diagnosis. By employing digital stethoscopes and machine learning, pulmonologists can offer accurate and timely diagnoses, leading to more efficient treatments and improved patient outcomes while lessening the burden on healthcare systems globally.

ACKNOWLEDGEMENT

We thank our Principal **Dr. S. Ganesh Vaidyanathan,** Sri Venkateswara College of Engineering for being the source of inspiration throughout our study in this college.

We express our sincere thanks to **Dr. R. Anitha,** Professor and Head of the Department, Computer Science and Engineering for her encouragement accorded to carry this project.

With profound respect, we express our deep sense of gratitude and sincere thanks to our internal guide **Ms. G. R. Khanaghavalle, Assistant Professor** and external guide **Mr. Dinesh Koka, CEO Onward Assist, Hyderabad (1999 batch ECE alumnus, SVCE)** for their valuable guidance and suggestions throughout this project.

We are also thankful to our Project coordinators **Dr. R. Jayabhaduri, Professor, Dr. N. M. Balamurugan, Professor and Dr. V. Rajalakshmi, Associate Professor** for their continual support and assistance.

We thank our family and friends for their support and encouragement throughout the course of our graduate studies.

ALLEN MANOJ
JANANI K
RITUNJAY M

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

AKG	Automatic Keyphrase Generation
ALSC	Aspect Level Sentiment Classification
ANN	Artificial Neural Network
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Networks
COPD	Chronic Obstructive Pulmonary Disease
CQT	Constant Q Transform
DFE	Deep Feature Extraction
DL	Deep Learning
FF	Feed Forward
GB	Gigabyte
Hz	Hertz
LRTI	Lower Respiratory Tract Infection
LSTM	Long Short-Term Memory
MBConv	Mobile inverted Bottleneck Convolution
MFCC	Mel-Frequency Cepstral Coefficients
ML	Machine Learning
PCA	Principal Component Analysis
RBN	Radial Basis Networks
RDC	Random Dot Cancellation
ReLU	Rectified Linear Unit
ResNet	Residual Network

RGB	Red Green Blue
RNN	Recurrent Neural Networks
RSV	Respiratory Syncytial Virus
SGD	Stochastic Gradient Descent
SIANN	Shift Invariant or Space Invariant Artificial Neural Networks
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TL	Transfer Learning
URTI	Upper Respiratory Tract Infection
VGG	Visual Geometry Group
VTLP	Variable Time and Length Padding
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Pulmonary diseases have a significant impact on a global scale, affecting millions of people worldwide and posing significant challenges to public health systems. Addressing the world impact of pulmonary diseases requires a comprehensive approach involving public health interventions, healthcare system strengthening, education, environmental regulations, and socioeconomic improvements. By promoting prevention, early diagnosis, and effective management, the burden of pulmonary diseases can be mitigated, improving global health outcomes.

1.2 PROBLEM STATEMENT

India has 18% of the global population and an increasing burden of chronic respiratory diseases. A Pulmonary disease may be caused by infections, smoking tobacco, or other forms of air pollution. According to the world health organization report in 2017, more than 235 million people are suffering from asthma worldwide. In addition, chronic obstructive pulmonary disease (COPD) is expected to be the third leading cause of death by 2030. This project aims to diagnose pulmonary diseases with the provided respiratory sounds using the latest technologies such as deep learning and digital stethoscope to achieve high accuracy in a safe, cost-effective and non-invasive way.

1.3 RESPIRATORY DISEASE

Respiratory diseases encompass a broad range of conditions that affect the respiratory system, which includes the organs involved in breathing and oxygen exchange. These conditions can vary in severity, from mild respiratory infections to chronic, life-threatening diseases.

The respiratory system plays a vital role in the body's overall health and well-being. It is responsible for taking in oxygen from the environment and expelling carbon dioxide, a waste product of metabolism. Any disruption or impairment of the respiratory system can lead to a variety of symptoms and complications.

Respiratory diseases can significantly impact an individual's quality of life, leading to symptoms that range from mild discomfort to severe disability. Proper prevention, early detection, and effective management of these conditions are crucial for reducing their burden on individuals and society as a whole. This includes lifestyle modifications, such as smoking cessation, maintaining good air quality, and receiving appropriate vaccinations, as well as timely medical interventions and ongoing treatment plans.

1.4 RESPIRATORY DISEASE CLASSIFICATION

1.4.1 Pneumonia

Pneumonia is an infection that inflames the air sacs in one or both lungs. It can be caused by bacteria, viruses, or fungi and often leads to symptoms such as cough, fever, chest pain, and difficulty breathing. Pneumonia can range in severity from mild to life-threatening and requires prompt medical treatment.

1.4.2 Chronic Obstructive Pulmonary Disease

COPD is a chronic lung disease characterized by persistent airflow limitation. It typically results from long-term exposure to irritants, such as cigarette smoke, and is mainly caused by chronic bronchitis and emphysema. COPD leads to symptoms such as coughing, wheezing, shortness of breath, and recurrent respiratory infections. It is a progressive condition that requires ongoing management to relieve symptoms and slow disease progression.

1.4.3 Asthma

Asthma is a chronic respiratory condition characterized by inflammation and narrowing of the airways. It causes recurring episodes of wheezing, coughing, chest tightness, and shortness of breath. Asthma symptoms can be triggered by various factors, including allergens, exercise, cold air, and respiratory infections. Management of asthma involves avoiding triggers, using medications to control inflammation and bronchoconstriction, and having a personalized asthma action plan. The prevalence of asthma has been on the rise worldwide over the past few decades. According to the Global Asthma Report 2018, approximately 339 million people globally have asthma.

1.4.4 Upper Respiratory Tract Infection

URTI refers to infections that affect the upper respiratory tract, including the nose, throat, and sinuses. Common causes of URTI include viruses, such as the common cold or flu. Symptoms may include nasal congestion, sore throat, cough, headache, and mild fever. URTIs are typically self-limiting and can be managed with rest, fluids, and over-the-counter medications to relieve symptoms.

1.4.5 Lower Respiratory Tract Infection

LRTI refers to infections that affect the lower respiratory tract, which includes the lungs and bronchial tubes. Examples of LRTIs include pneumonia and bronchitis. LRTIs are often caused by viruses or bacteria and can lead to symptoms such as coughing, chest congestion, fever, and difficulty breathing. Treatment for LRTIs depends on the specific infection and may involve antibiotics, antiviral medications, or supportive care.

1.4.6 Bronchiectasis

Bronchiectasis is a chronic condition characterized by abnormal widening and thickening of the bronchial tubes, which results in a build-up of mucus and recurrent respiratory infections. It can be caused by underlying conditions such as cystic fibrosis or as a result of previous lung infections. Symptoms may include chronic cough with large amounts of sputum, shortness of breath, and recurrent chest infections. Management of bronchiectasis involves airway clearance techniques, antibiotics to treat infections, and addressing underlying causes.

1.4.7 Bronchiolitis

Bronchiolitis is a common respiratory infection that primarily affects infants and young children. It is usually caused by a viral infection, commonly respiratory syncytial virus (RSV). Bronchiolitis leads to inflammation and blockage of the small airways in the lungs, resulting in symptoms such as cough, wheezing, rapid breathing, and difficulty feeding. Treatment focuses on supportive care, including maintaining hydration, ensuring proper oxygen levels, and monitoring breathing patterns.

1.5 DEEP LEARNING

Deep Learning, a sub-field of machine learning, is inspired by the working of the human brain and involves the replication of neurons through a system of interconnected nodes. In the context of pulmonary disease classification, Deep Learning (DL) algorithms play a significant role. Artificial neural networks, the standard neural network architecture, consist of nodes and weights across different layers. These layers include input nodes where data is passed, output nodes for predictions, and hidden layers that identify patterns in the input data. Initially, each node is assigned a weight bias, and the neural network is trained using existing datasets. Back propagation is used to adjust the weights between nodes based on necessary feedback. This iterative process finds the most efficient weight combination by comparing the weights and the output cost function through gradient descent. By calculating the difference between desired and existing output, the algorithm identifies the efficient local minima with lower cost functions, allowing the adjustment of weights. In recent years, attention-based architectures have gained significant attention in the deep learning community. These architectures use mechanisms that allow the model to focus on relevant parts of the input data while disregarding irrelevant information. Transformer models, such as the widely known BERT (Bidirectional Encoder Representations from Transformers), have revolutionized natural language processing tasks by employing attention mechanisms to capture contextual relationships within text data.

Deep learning architectures have found applications in various fields, including pulmonary disease classification. Different neural network architectures are designed to work with specific data types or domains. Feed Forward (FF) networks consist of interconnected neurons and hidden layers, with data flowing only in the forward direction. Multi-layer Perceptron incorporate multiple hidden layers and activation functions, allowing supervised learning through forward and backward propagation. Radial Basis Networks (RBN) use radial functions to predict targets based on the

comparison of feature values with stored classes. Convolutional Neural Networks (CNN) are commonly used for image classification, extracting features from images through convolution layers. Recurrent Neural Networks (RNN) are suitable for sequential data, incorporating time-delayed inputs for predictions.

While RNNs face challenges with vanishing gradients, Long Short-Term Memory (LSTM) neural networks address this issue by incorporating a memory cell. LSTM uses gates to control the input, output, and forgetting of data, allowing for the storage of information for longer periods.

In the domain of pulmonary disease classification, Deep Learning techniques offer promising capabilities for accurately categorizing and diagnosing respiratory conditions. By leveraging the power of neural networks, these algorithms can analyze complex data patterns and improve the accuracy and efficiency of pulmonary disease diagnosis.

1.6 CONVOLUTIONAL NEURAL NETWORKS

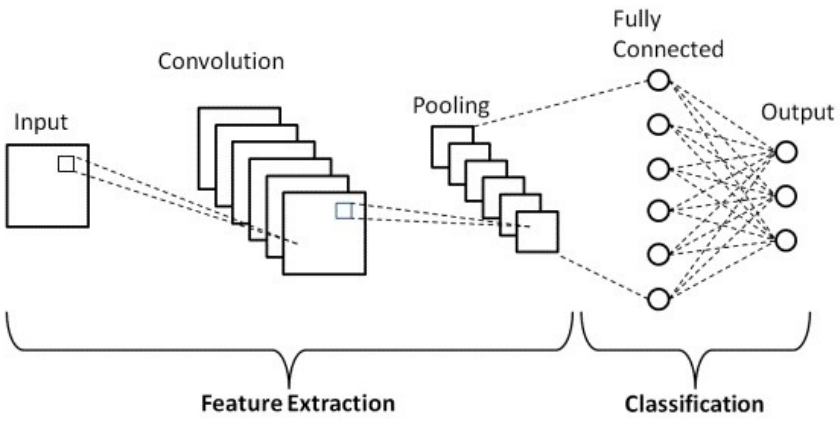

Figure 1.1 CNN Architecture

A Convolutional Neural Network (CNN) as shown in Figure 1.1 is a type of Artificial Neural Network (ANN) specifically designed for image processing tasks. It

operates by taking an input image and applying learnable weights and biases to different aspects or objects within the image, enabling it to distinguish between different entities. CNNs are also referred to as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN) due to their shared-weight architecture, where convolution kernels or filters slide across input features and produce translation-equivariant responses known as feature maps. In CNNs, convolution is a mathematical operation used to merge two sets of information. It involves filtering the input data to generate a feature map.

The filter, also known as a kernel or feature detector, typically has dimensions like 3x3. The convolution process involves the kernel traversing the input image and performing element-wise matrix multiplication. The results for each receptive field, where convolution occurs, are recorded in the feature map. Feature maps are valuable for several purposes, including reducing the size of the input image. Larger strides during convolution lead to smaller feature maps. While a one-pixel stride was used in this example, real-world images often require wider strides due to their larger and more complex nature.

The concept of CNNs was inspired by the functioning of neurons in the brain, where each neuron acts as a node transmitting electrical signals. CNNs leverage a multi-layer perceptron design that has been optimized for reduced processing requirements. The layers of a CNN include an input layer, an output layer, and a hidden layer that incorporates multiple convolutional layers, pooling layers, fully connected layers, and normalization layers. CNNs are predominantly employed in image recognition tasks, as images can be converted into matrices and passed as input. They find applications in image and video recognition, image classification, medical image analysis, natural language processing, and brain-computer interfaces.

The main advantage of CNNs over their predecessors is their ability to automatically detect important features without human supervision. CNNs have been successfully applied in various real-world applications, including cancer detection and biometric authentication. They are also employed in visual question answering and image captioning tasks, where CNNs take an input image and generate natural language descriptions or answers based on that image.

1.7 STATISTICS OF PULMONARY DISEASE

Pulmonary diseases pose a significant global health challenge, contributing to high mortality rates and causing substantial healthcare burden. Preventive measures, such as smoking cessation, avoidance of second-hand smoke, and minimizing exposure to air pollution, can play a crucial role in mitigating the risk of developing pulmonary diseases. Treatment options for these conditions include medication and surgical interventions. According to the World Health Organization (WHO), pulmonary diseases ranked as the fourth leading cause of death worldwide in 2019, accounting for 12.8% of all deaths. The most prevalent pulmonary diseases include chronic obstructive pulmonary disease (COPD), lung cancer, and pneumonia. Disturbingly, in 2020, these diseases resulted in 4.1 million deaths globally. Smoking remains the primary risk factor for pulmonary diseases, exerting a significant impact on disease development and progression. Additionally, air pollution is a major contributor to the incidence and severity of pulmonary diseases.

The consequences of pulmonary diseases extend beyond mortality rates. They also inflict substantial disability and impair the quality of life for affected individuals. Figure 1.2 provides statistical insights into the prevalence of pulmonary diseases among males and females. Efforts to address pulmonary diseases must encompass comprehensive strategies that emphasize prevention, early detection, and effective management. Implementing measures to reduce smoking rates, enhance air quality, and

promote awareness of the risk factors associated with pulmonary diseases can help reduce their impact on individuals, communities, and healthcare systems worldwide.

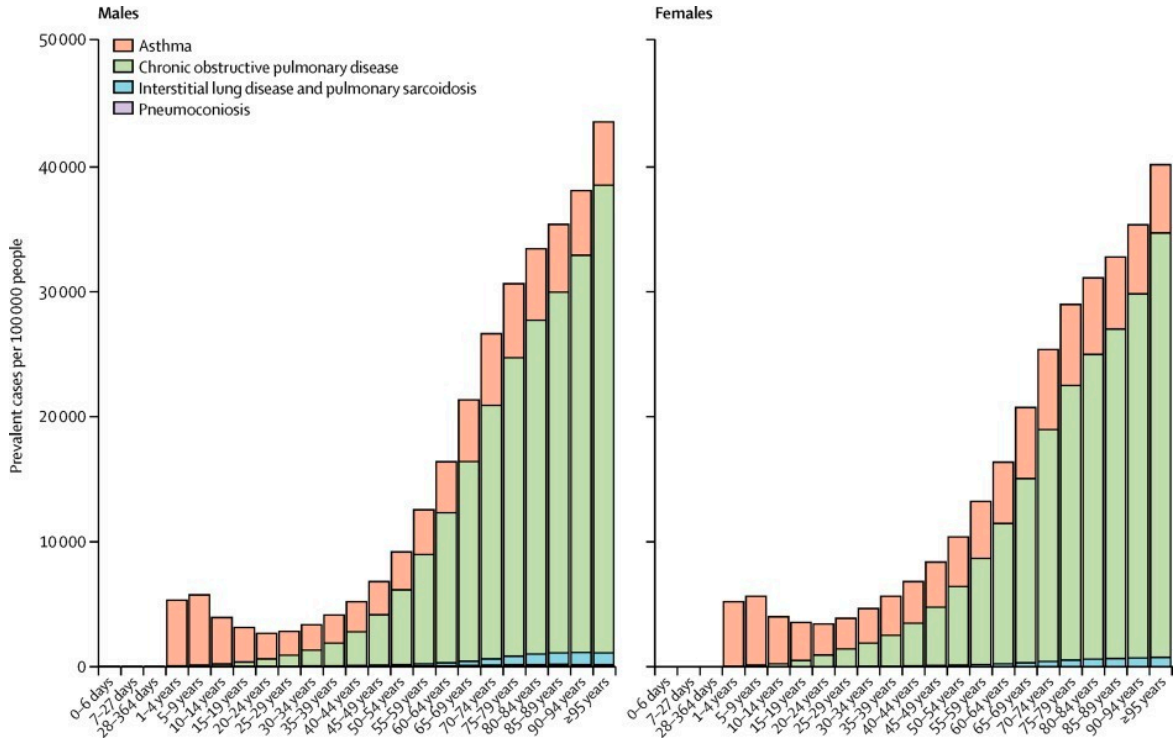

Figure 1.2 Statistics of people affected by Pulmonary Diseases

1.8 ISSUES AND CHALLENGES

In current medical practices, one of the commonly used methods for diagnosing respiratory diseases involves the utilization of Chest X-Ray Scans, which are then carefully examined by trained medical professionals. However, there are certain limitations associated with this approach. X-Rays can be costly, making it less cost-effective for widespread implementation. Moreover, the safe disposal of X-Ray materials in large quantities can also be a concern.

Another traditional tool used for diagnosis is the manual stethoscope. However, this method has its limitations as well. Many diseases exhibit characteristic respiratory sounds that may go undetected or be misdiagnosed due to the inability of the physician

to accurately hear or interpret these sounds. Factors such as the quality of the stethoscope and the experience of the physician can further impact the accuracy of the diagnosis.

To address these challenges, machine learning (ML) models have been developed to aid in respiratory disease diagnosis. These models can analyze large datasets, such as the ICBHI dataset, and extract valuable insights for improved detection and classification of respiratory conditions. However, it should be noted that current ML models have primarily focused on binary and ternary-class classification, meaning they can distinguish between a limited number of disease categories.

Efforts are being made to enhance the capabilities of ML models for respiratory disease diagnosis. Researchers are exploring ways to increase the complexity and accuracy of these models to accommodate a wider range of disease classifications. By training ML models on larger datasets and leveraging advanced algorithms, it is anticipated that they will become more adept at recognizing and differentiating various respiratory conditions.

CHAPTER 2

LITERATURE REVIEW

Truc Nguyen et al. (2022) explores the use of pre-trained ResNet models as backbone architectures for classifying adventitious lung sounds and respiratory diseases. The researchers leverage the transferred knowledge of pre-trained models from different ResNet architectures through various techniques, including vanilla fine-tuning, co-tuning, stochastic normalization, and a combination of co-tuning and stochastic normalization. To enhance the system's robustness, the researchers introduce spectrum correction and flipping data augmentation methods. Four ResNet models—ResNet18, ResNet34, ResNet50, and ResNet101—are implemented to assess their accuracy levels. Notably, ResNet101 achieved the highest accuracy of 90% in the 3-class classification task, which involved differentiating between Healthy, Chronic, and Non-Chronic categories. Furthermore, the study also encompassed sound type classification, specifically distinguishing between Wheezes, Crackles, and Subtle sounds. However, the researchers found that the intended accuracy levels were not attained due to the utilization of pre-trained deep learning models.

Fatih Demir et al. (2020) work on the classification of lung sound signals using two deep learning-based approaches. Initially, the lung sound signals were transformed into spectrogram images using a time-frequency method. In the first approach, a pre-trained deep convolutional neural network (CNN) model was employed for feature extraction. Subsequently, a support vector machine (SVM) classifier was utilized to classify the lung sounds based on the extracted features. The second approach involved fine-tuning the pre-trained deep CNN model using the spectrogram images specifically

for lung sound classification. The VGG16 architecture was utilized as the CNN model. The accuracy achieved with VGG16 was 65%, while the SVM classifier obtained an accuracy of 63%. The classification tasks performed included distinguishing between different sound types, namely Wheezes, Crackles, and Subtle. Additionally, a 3-class classification was conducted, categorizing the lung sounds as Healthy, Chronic, and Non-Chronic. The training and testing of the models were conducted using the acknowledged ICBHI dataset. Overall, this study demonstrates the utilization of deep learning techniques for lung sound classification. Two approaches were explored, one involving a pre-trained CNN model with an SVM classifier, and the other incorporating fine-tuning of the pre-trained CNN model. The achieved accuracies provide insights into the effectiveness of these approaches, while the specific classification tasks highlight the capability of the models to differentiate between different lung sound patterns. The use of the ICBHI dataset ensures the credibility and reliability of the training and testing process.

Elmar Messner et al. (2021) present a novel approach to address multi-channel lung sound classification. The proposed framework introduces a frame-wise classification method, leveraging a convolutional recurrent neural network (RNN) to analyze complete breathing cycles within multi-channel lung sound recordings. To extract relevant features, spectrogram representations are derived from the lung sound recordings and evaluated using different deep neural network architectures for binary classification, specifically focusing on distinguishing between healthy and pathological samples. However, it is crucial to highlight that this study solely focused on binary classification, categorizing the lung sound recordings as either Healthy or Unhealthy. The dataset employed in this research comprises a combination of self-collected and non-acknowledged data. It is worth noting that the dataset is imbalanced, with a limited number of unhealthy patients compared to the healthy samples. As a result, the reported accuracy of 92% should be interpreted with caution due to the potential biases introduced by the imbalanced dataset. Despite the limitations regarding the dataset, the

proposed framework and approach for multi-channel lung sound classification are innovative and hold promise. By employing a convolutional RNN and utilizing spectrogram features, the study addresses the complexity of lung sound analysis and offers insights into the potential for automated detection and classification of pathological conditions. The frame-wise classification method allows for a more granular analysis of breathing cycles, potentially capturing crucial details that can aid in the identification of respiratory abnormalities.

L. Pham et al. (2021) emphasized the classification of anomalies within respiratory cycles and detect diseases using respiratory sound recordings. The proposed framework consists of a front-end feature extraction process, which transforms the input sound data into a spectrogram representation. Subsequently, a back-end deep learning network is employed to classify the extracted spectrogram features into different categories, representing respiratory anomaly cycles or specific diseases. To strike a balance between model performance and complexity, a Teacher-Student scheme is applied, showing potential for real-time applications. The study incorporates two tasks: task 1 focuses on sound classification, while task 2 centers around disease classification. To determine the highest accuracy achieved, two types of classification were implemented. In the 2-class classification, an accuracy of 93% was attained, while the 3-class classification achieved an accuracy of 86%. The models were trained and tested using the acknowledged ICBHI dataset. In summary, this paper introduces a framework for classifying respiratory anomalies and detecting diseases based on respiratory sound recordings. The process involves front-end feature extraction and back-end deep learning models. By employing a Teacher-Student scheme, a balance is achieved between model performance and complexity, rendering the framework suitable for real-time applications. The study encompasses sound classification and disease classification tasks, with notable accuracies reported for both 2-class and 3-class classifications. The utilization of the ICBHI dataset adds credibility to the training and testing processes.

J. Acharya et al. (2020) introduces a deep CNN-RNN model designed to classify respiratory sounds using Mel-spectrograms. Additionally, a patient-specific model tuning strategy is implemented, which involves screening respiratory patients and building individualized classification models using limited patient data to enhance anomaly detection reliability. The proposed model achieves a state-of-the-art score on the ICBHI'17 dataset, demonstrating its effectiveness in respiratory sound classification. Furthermore, the study explores the ability of deep learning models to learn domain-specific knowledge by pre-training them with breathing data, resulting in significantly improved performance compared to generalized models. The CNN-RNN architecture is utilized for training and testing the data, with the inclusion of VGG16 and MobileNet models. The ICBHI dataset is employed in this research. The highest recorded accuracy achieved by the model is 66.31%. In summary, this work presents a deep CNN-RNN model for respiratory sound classification, leveraging Mel-spectrograms. The inclusion of a patient-specific model tuning strategy contributes to reliable anomaly detection. The study highlights the model's state-of-the-art performance on the ICBHI'17 dataset and emphasizes the benefits of pre-training deep learning models with domain-specific data. The utilization of the CNN-RNN architecture, along with VGG16 and MobileNet models, demonstrates the potential for improved classification results. However, it is important to consider the context and limitations of the dataset used when interpreting the recorded accuracy.

Kim, Y. et al. (2021) concluded that the deep learning-based classification approach proved effective in accurately classifying respiratory sounds. By incorporating transfer learning, which combines pre-trained image feature extraction from respiratory sound data with a convolutional neural network (CNN) classifier, the accuracy of the classification process was significantly improved. However, challenges persist in analyzing mixed abnormal sounds and filtering out noise. Nonetheless, recent advancements in analytical algorithms and recording technologies are expected to expedite progress in respiratory sound analysis. It is worth noting that the obtained

accuracy of 60.3% is approximately similar to the accuracy achieved without the utilization of neural networks. However, the significance lies in the potential breakthrough of detecting pneumonia in an unsupervised manner, as currently, only experienced pulmonologists are able to accurately detect it.

Improving the model's accuracy could have a significant impact on pneumonia detection. The study employed Mel Spectrogram as the extracted feature, and training and testing were conducted using the VGG16 model. Overall, the research signifies the efficacy of deep learning-based approaches in classifying respiratory sounds and highlights the potential for advancements in unsupervised pneumonia detection. By leveraging Mel Spectrogram and the VGG16 model, this study contributes to the ongoing efforts to enhance respiratory sound analysis and improve pneumonia detection in a manner that is accessible and beneficial to a broader range of healthcare settings.

CHAPTER 3

PROPOSED WORK

Respiratory anomaly classification, is separated into two sub-tasks. The first aims to classify 2 different stages (Healthy, Unhealthy). The second is to classify the eight types of classes (Asthma, Bronchiectasis, Bronchiolitis, COPD, LRTI, URTI, Pneumonia, and Healthy). Figure 3.1 represents the architecture of the proposed work in classifying pulmonary diseases.

The several steps in the execution are:

- Audio Pre-processing
- Feature Extraction
- Data Augmentation
- Model Training

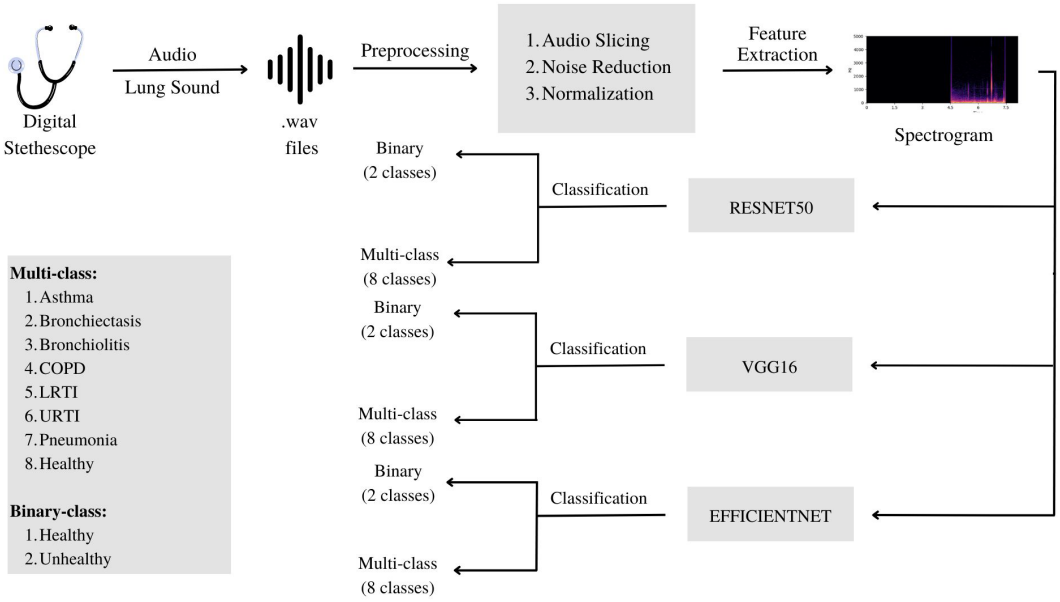

Figure 3.1 Architecture of the System

3.1 AUDIO PREPROCESSING

In the pre-processing stage of the dataset, an important step is performed to ensure the removal of noises and blanks in the audio files. This step is known as audio splitting, and its purpose is to extract the relevant portions of the audio while discarding any unwanted parts. By applying this technique, the researchers are able to reduce the time and size of the audio files to approximately 6 seconds, resulting in a more manageable dataset for further analysis. Moreover, this pre-processing step has been found to contribute to an increase in accuracy in subsequent classification tasks.

One specific aspect addressed during the pre-processing stage is the variation in sampling frequencies among the audio samples in the dataset. To address this issue, all of the signals are down-sampled to a common frequency of 4 kHz. This down-sampling process involves reducing the sampling rate of the audio signals while preserving the relevant information. In the case of wheeze and crackle signals, which typically occur within the frequency range of 0-2 kHz, down-sampling the audio samples to 4 kHz is considered appropriate since it does not result in the loss of important frequency components.

By down-sampling the audio signals to a common frequency of 4 kHz, the researchers ensure consistency and compatibility among the samples in the dataset. This facilitates subsequent analysis and classification tasks by providing a standardized format for the audio data. It also allows for easier comparison and extraction of relevant features from the signals.

The decision to down-sample the audio samples to 4 kHz is based on the understanding that the wheeze and crackle signals of interest are primarily contained within the lower frequency range. Therefore, reducing the sampling rate to 4 kHz does

not result in the loss of critical information for the classification tasks related to wheeze and crackle detection.

3.2 FEATURE EXTRACTION

The After the audio splitting, the .wav files are converted into Mel-Spectrogram images for disease prediction. The Mel-scale is a perceptual scale of pitches. Mel spectrograms describe the audio signal in the Mel scale over time. The lung sounds were converted to its Mel spectrograms to uncover pitch patterns informative to the domain experts. A hop length of 0.01s and a window length of 0.02s were used. It has been observed that the human ear acts as a filter when it focuses on only specific frequency components. The relationship between the Mel-scale and linear-scale frequency is defined as follows:

$$Mel = 2695 \times log_{10}\left(1 + \frac{f}{700}\right) \tag{3.1}$$

For feature extraction we have used Mel-frequency spectrogram with a window size of 60ms with 50% overlap. Each breathing cycle is converted to a 2D image where rows correspond to frequencies in Mel scale and columns correspond to time (window) and each value represent log amplitude value of the signal corresponding to that frequency and time window.

The Short Time Fourier Transform (STFT) is used for T-F image construction. Because lung sounds are recorded at different frequencies, the window sizes that should be used for the STFT are different. Window sizes are chosen between 0.01 and 0.025 times of the sampling frequency because they would better reveal the lung sound characteristics. After the T-F images are constructed, they are resized to 224×224 because of being suitable with deep feature extraction and transfer learning.

Four types of spectrogram:

1. Log-Mel spectrogram,
2. Gammatone filter bank (Gamma) spectrogram,
3. Stacked Mel-Frequency Cepstral Coefficients (MFCC), and
4. Rectangular Constant Q Transform (CQT) spectrogram.

The log-Mel spectrograms are normalized with zero mean and unit variance. Then these spectrograms are duplicated into three channels to match the input size of the pre-trained ResNet model for the ALSC task.

However, for the RDC task of the ICBHI dataset, we convert the spectrogram, which is considered as a grey image, into a RGB colour image and enlarge the image to twice the size using linear interpolation. These techniques are commonly used.

The figures in 3.2 display the Mel Spectrogram images corresponding to various respiratory conditions, including Asthma, Bronchiectasis, Bronchiolitis, COPD, URTI, LRTI, Pneumonia, and Healthy data.

These spectrograms provide visual representations of the frequency content of the respiratory sounds for each condition, offering insights into their unique characteristics and patterns.

Analyzing these spectrograms can aid in the diagnosis and differentiation of these respiratory conditions based on their distinct spectral features.

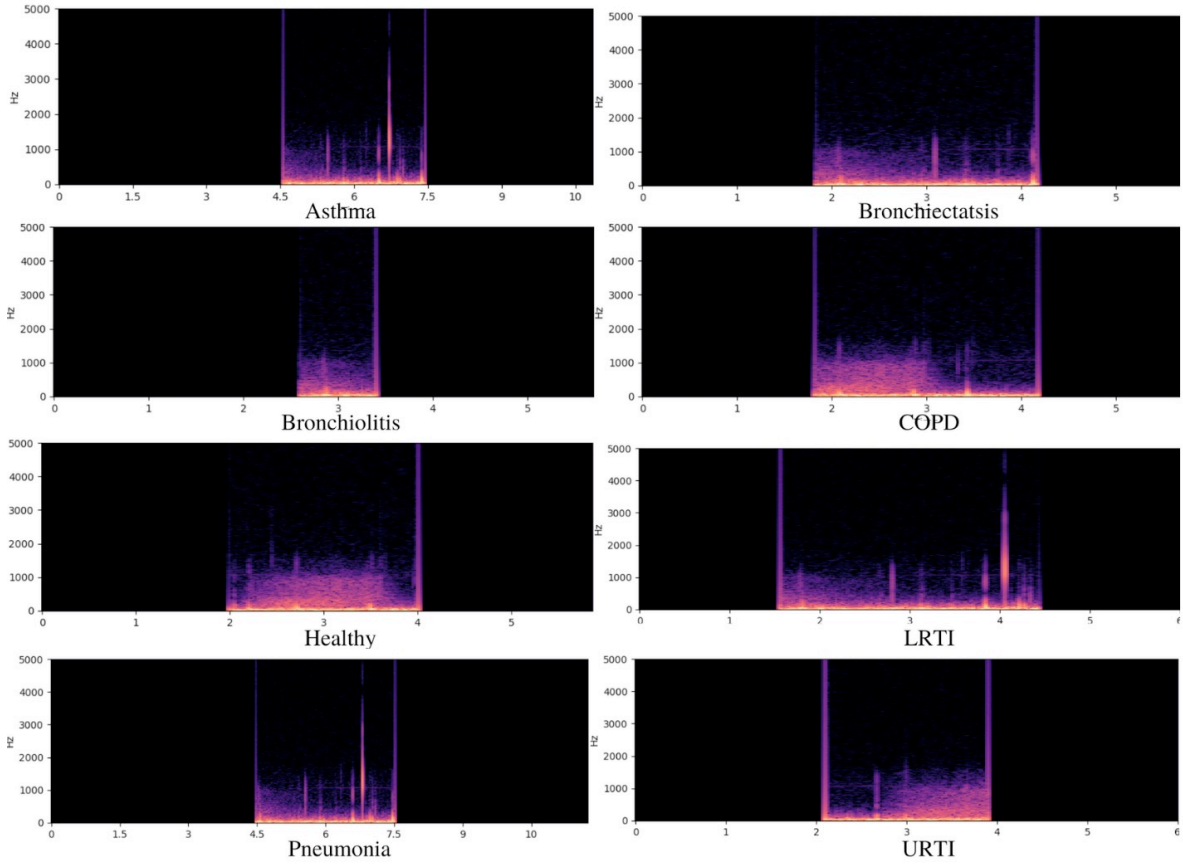

Figure 3.2 Mel-spectrograms of 8 classes

3.3 DATA AUGMENTATION

To enhance the performance of deep learning models, the availability of a large and diverse training dataset is crucial. However, in many domains, including medical image analysis, access to such extensive datasets is limited. Moreover, class imbalances within available databases pose an additional challenge. In the context of the ICBHI Dataset, the researchers employ a technique called data augmentation to address these issues and improve the quantity and quality of the training data for deep learning models.

Data augmentation involves generating additional training samples by applying various transformations to the existing data. In the case of the ICBHI Dataset, the researchers focus on augmenting the wheeze and crackle classes. They use a method

called time stretching to double the number of segments in these classes, effectively increasing the size of the training set. This technique modifies the temporal characteristics of the audio samples without altering the essential characteristics of the wheeze and crackle signals.

To further augment the training set, the researchers apply additional data augmentation methods with predefined probabilities. These methods include volume adjusting, noise addition, pitch adjusting, and speed adjusting. By randomly applying these transformations, the researchers introduce further variations to the training data, making the model more robust to different instances and enhancing its generalization capability.

In addition to these augmentation techniques, the researchers apply Variable Time and Length Padding (VTLP) to expand the dataset for all classes in both tasks. VTLP modifies the temporal characteristics of the audio signals by stretching or compressing them, thus increasing the diversity of the training data. This augmentation technique helps to alleviate the limitations imposed by a small dataset and class imbalances, enabling the model to learn more effectively from the available samples.

Furthermore, to enrich the feature representation of the dataset, the researchers double the log-mel features by adding the flipped log-mel features along the frequency axis. This technique enhances the discriminatory power of the features and contributes to the improved performance of the deep learning model in tasks such as adventitious lung sound classification and crackle detection.

By employing various data augmentation techniques, including time stretching, volume adjusting, noise addition, pitch adjusting, speed adjusting, VTLP, and feature flipping, the researchers effectively expand the training dataset and enhance its diversity. These augmentation methods help to mitigate the challenges posed by limited

data availability and class imbalances. Ultimately, they improve the performance of the deep learning models on the ICBHI Dataset and enable more accurate and robust classification of adventitious lung sounds and crackles.

3.4 MODEL TRAINING

Convolutional Neural Networks (CNNs) are feed-forward neural networks that process data with a recognized grid-like architecture, such as time series (1-dimensional) and image data (2-dimensional). They are frequently used in audio processing and computer vision. Convolutional layers, pooling layers, and fully-connected layers are the three types of layers that make up CNNs. We exclusively employ convolutional and fully-connected layers in this paper.

In Transfer Learning (TL), which is contemporary trend in deep learning, the layers of a pre-trained network are shared or conveyed to other networks for fine-tuning or features extraction. Initially, the training of a CNN model is carried out by a large dataset. After this process, training of pre-trained model is conducted once more by a smaller dataset to get fine tuning for developing estimated performance of CNN model. TL which provides satisfying tuning is more enduring than CNN model training from scratch. While initial layers represent features such as curves, colour blobs, edges in CNN architecture, abstract and specific features are provided by final layers.

In Deep Feature Extraction (DFE), which also performs on principle of transfer learning in place of training a pre-trained CNN model, the related feature vectors are extracted by using activation layers of CNN models. While the previous layers' activations present low-level image features such as edges, later or deeper layers present explicitly higher-level features for recognition of image. For instance, the activations of first and second fully connected layers provide feature representation in ImageNets. The first stage consists of batch-normalization, convolution and max-pool layers. The batch

normalization layer scales the input images over each batch to stabilize the training. In the 2D convolution layer the input is convolved with 2D kernels to produce abstract feature maps. Each convolution layer is followed by Rectified Linear activation functions (ReLU). The max-pool layer selects the maximum values from a pixel neighbourhood which reduces the overall network parameters and results in shift-invariance.

To benchmark the performance of our proposed model, we compare it to two standard CNN models, VGG-16 and ResNet. Since our dataset size is limited even after data augmentation, it can cause overfitting if we train these models from scratch on our dataset. Hence, we used ImageNet trained weights instead and replaced the dense layers of these models with an architecture similar to the fully connected and SoftMax layers of our proposed EfficientNet architecture. Then the models are trained with a small learning rate.

CHAPTER 4

SYSTEM REQUIREMENTS

The hardware and software requirements are as follows:

4.1 HARDWARE REQUIREMENTS:

Random Access Memory	:	8GB
Processors	:	Intel Core i3 Processor or above
Microphone	:	50 Hz - 15 kHz

4.2 SOFTWARE REQUIREMENTS:

Operating Systems	:	Microsoft Windows 7 or later, MacOS Montenary
Python Version	:	3.8.3
Development Tools	:	Anaconda Framework
Compatible Tools	:	Microsoft Visual Studio Code
Python Libraries	:	TensorFlow, Librosa, Pandas,NumPy

CHAPTER 5

IMPLEMENTATION MODULES

5.1 DATASET DESCRIPTION

The experiments in this project utilized the ICBHI 2017 Dataset, which is a comprehensive collection of respiratory sound recordings specifically designed for scientific research in the field of respiratory diseases. The dataset consists of 920 audio samples from 126 subjects, all of which have been carefully annotated by medical experts. This dataset is notable for being one of the largest publicly available databases with detailed annotations.

The audio samples in the dataset were classified into four categories based on the characteristics of the respiratory cycles:

1. Normal: These samples exhibit subtle sounds with no abnormalities or variations.
2. Wheezes: These samples contain continuous high-pitched sounds, typically associated with chronic respiratory conditions.
3. Crackles: These samples feature discontinuous sounds of shorter duration, often indicating non-chronic respiratory conditions.
4. Both (Wheezes and Crackles): These samples show the presence of both wheezes and crackles, suggesting the possibility of both chronic and non-chronic respiratory diseases.

The database contains a total of 6,898 respiratory cycles spanning 5.5 hours of audio. Among these cycles, 1,864 contain crackles, 886 contain wheezes, 506 contain both wheezes and crackles, while the remaining cycles are classified as normal.

The recordings were captured using various equipment such as:

1. AKG C417 L Microphone,
2. 3M Littmann Classic II SE Stethoscope,
3. 3M Littmann 3200 Electronic Stethoscope, and
4. WelchAllyn Meditron Master Elite Electronic Stethoscope

During the data collection process in hospitals located in Portugal and Greece, observations were made at various chest locations to gather relevant information. These locations included the trachea, anterior left/right, posterior left/right, and lateral left/right. By examining these specific areas of the chest, healthcare professionals can obtain a comprehensive understanding of the respiratory sounds and potentially identify any abnormal findings or patterns that may indicate underlying respiratory conditions. The data collected from these chest locations can contribute to accurate diagnoses and effective treatment planning for patients in these hospitals.

The data was collected from patients aged 19 years and above, including both inpatient and outpatient cases. The following Table 5.1 consists of the number of audio recordings for each classification type in Binary Classification.

Table 5.1 Binary Classification Audio Set Description

Classification Type	Number of audio recordings
Healthy	322
Unhealthy	6607

Table 5.2 Multiclass Classification Audio Set Description

Classification Type	Number of audio recordings
Asthma	6
Bronchiectasis	104
Bronchiolitis	160
COPD	5747
Healthy	322
LRTI	32
Pneumonia	285
URTI	243

Table 5.2 presents the distribution of audio recordings across different classification types in a multiclass classification setting. Each classification type represents a specific respiratory condition, and the table displays the corresponding number of audio recordings available for each condition. This information is crucial for training and evaluating machine learning models designed for multiclass classification tasks, as it ensures that an adequate representation of each classification type is present in the dataset.

5.2 DATA PREPROCESSING

Data preprocessing was performed to prepare the dataset for pre-trained deep learning models. Techniques such as audio slicing, noise reduction, and normalization were applied to ensure compatibility and optimal performance.

5.2.1 Audio Slicing

Audio slicing is a fundamental technique in deep learning datasets that deal with audio data. It involves dividing longer audio recordings into smaller segments or chunks

to facilitate the training and processing of the data within deep learning models. The purpose of audio slicing is to improve computational efficiency and reduce memory requirements during training by breaking down the audio into manageable units.

In many applications of deep learning, audio files can vary significantly in length. Working with long audio recordings directly within deep learning models can be challenging due to computational constraints and memory limitations. Audio slicing addresses this issue by dividing the audio into shorter segments, making it easier to process and analyze within the model.

The process of audio slicing involves selecting a suitable segment length and dividing the audio recording into multiple segments of equal or variable duration. The segment length is typically determined based on the specific requirements of the deep learning model and the nature of the audio data. By breaking the audio into smaller segments, it becomes more feasible to handle and manipulate the data effectively.

Audio slicing offers several benefits in deep learning applications. Firstly, it improves computational efficiency by allowing the model to process smaller chunks of data at a time. This enables faster training and inference times, especially when working with large audio datasets. Secondly, it reduces memory requirements since the model only needs to hold a portion of the audio data in memory at any given time, rather than the entire recording. This is particularly important for models with limited memory capacity.

Additionally, audio slicing enables better alignment between the audio data and the corresponding labels or annotations. By dividing the audio into segments, it becomes easier to associate specific sections of the audio with their corresponding target outputs, such as classification labels or timestamps. This alignment is crucial for training deep learning models effectively and accurately.

5.2.2 Noise Reduction

In the realm of deep learning sound datasets, noise reduction holds significant importance as a crucial preprocessing step. It entails the elimination or mitigation of undesirable background noise or interference present in audio recordings. The objective is to enable deep learning models to concentrate on the relevant signal and enhance the accuracy of sound classification or analysis tasks. A range of noise reduction techniques can be employed, including spectral subtraction, Wiener filtering, and deep learning-based denoising algorithms. These methods aim to improve the signal-to-noise ratio, thereby augmenting the overall quality of the audio data. As a result, deep learning model training and evaluation become more dependable, robust, and effective.

Background noise can stem from various sources, such as environmental factors, recording equipment, or transmission artifacts. It can obscure the desired audio signal, making it challenging for deep learning models to discern the pertinent features for accurate classification or analysis. Noise reduction techniques help mitigate this challenge by extracting and isolating the signal from the background noise.

Spectral subtraction is a widely used technique for noise reduction in audio processing. It operates by estimating the spectral profile of the noise and subtracting it from the spectrogram of the noisy signal. This approach exploits the assumption that the noise is stationary and can be modeled as a background additive component in the frequency domain. By subtracting the estimated noise spectrum from the original signal, the clean signal can be reconstructed.

Wiener filtering is another popular approach for noise reduction. It utilizes statistical properties of the signal and noise to estimate the clean signal. The Wiener filter aims to minimize the mean square error between the estimated clean signal and

the original signal. It achieves this by adapting its filtering characteristics based on the frequency content of the signal and the noise.

Deep learning-based denoising algorithms have emerged as powerful techniques for noise reduction in recent years. These algorithms employ deep neural networks to learn the mapping between noisy and clean audio signals. By training on a large dataset of noisy and corresponding clean signals, the neural network learns to extract meaningful features and effectively denoise the input. Deep learning models, such as autoencoders or convolutional neural networks, have shown promising results in noise reduction tasks.

By applying noise reduction techniques in the preprocessing stage of deep learning sound datasets, the models can benefit from cleaner and more reliable training data. The removal or reduction of background noise enhances the discriminative power of the models, allowing them to focus on the relevant audio features. This, in turn, leads to improved accuracy in sound classification, speech recognition, or other audio-related tasks.

5.2.3 Normalization

Normalization is a fundamental data preprocessing technique widely employed in deep learning datasets. Its primary objective is to rescale the input data to a standard range, typically between 0 and 1 or -1 and 1. By doing so, normalization ensures that each feature in the dataset possesses a similar scale, preventing certain features from dominating the learning process due to their larger magnitudes. This normalization process plays a crucial role in promoting stable and efficient training of deep learning models.

The need for normalization arises because features in the input data can have varying scales and ranges. Some features may have values that span a wide range, while others may have values confined to a much smaller range. This discrepancy in scales can adversely affect the learning process within deep learning models. For example, features with larger magnitudes may contribute more significantly to the model's overall learning process, overpowering the impact of features with smaller magnitudes. As a result, the model may fail to learn the underlying patterns and relationships present in the data accurately.

Normalization tackles this issue by transforming the data to a standardized range, effectively eliminating the influence of different scales. One common method of normalization is z-score normalization, also known as standardization. In this approach, each data point is transformed by subtracting the mean of the feature and dividing it by the standard deviation. This process centers the data around a mean of zero and scales it to have a standard deviation of one. As a result, the transformed data has a similar scale across all features, enabling the model to treat each feature equally during training.

Another approach to normalization is min-max scaling, where the data is scaled to a specified range, typically between 0 and 1. The minimum value of the feature is subtracted from each data point, and the result is divided by the range of the feature (the difference between the maximum and minimum values). This transformation ensures that the data is scaled proportionally within the specified range, making it suitable for training deep learning models.

Normalization offers several benefits in deep learning. Firstly, it helps accelerate the convergence of the model during training. With normalized data, the model can update its parameters more efficiently, as the gradients computed during backpropagation are less prone to exploding or vanishing. This leads to faster convergence and more stable learning dynamics.

Normalization also aids in improving the generalization capability of the model. By bringing all features to a similar scale, normalization prevents certain features from dominating the learning process solely based on their magnitudes. This allows the model to pay equal attention to all features, ensuring that important patterns and relationships are learned from the entire dataset.

Normalization enhances the model's robustness to variations in input data. When new data is encountered during inference, normalization ensures that the input is transformed to a consistent scale, aligning with the normalization applied during training. This consistency facilitates accurate predictions on unseen data.

5.3 FEATURE EXTRACTION

Feature extraction plays a vital role in data preprocessing, particularly in machine learning and deep learning tasks. Its primary objective is to transform raw input data into a set of meaningful features that effectively capture the relevant information necessary for analysis or modeling. By extracting informative and discriminative features, the dimensionality of the data can be reduced, facilitating more efficient and effective analysis.

The process of feature extraction involves selecting or creating a set of variables, known as features, that best represent the underlying patterns or characteristics of the data. These features should encapsulate the essential information required for the specific task at hand. For example, in image processing, features can be extracted using algorithms that detect edges, corners, textures, or other visual patterns that are relevant to the analysis. In natural language processing, features can include word frequencies, n-grams, or semantic representations that capture the linguistic aspects of the text.

Feature extraction can be performed either manually by domain experts with a deep understanding of the data and its relevance to the task or automatically using various techniques. Automated feature extraction methods leverage algorithms and mathematical transformations to identify and extract relevant features from the data automatically. Techniques such as principal component analysis (PCA), wavelet transforms, or deep learning-based methods can be employed for this purpose. These methods analyze the data, identify patterns, and extract features that possess high information content and discriminatory power.

The extracted features serve as input to machine learning or deep learning algorithms, enabling them to learn patterns and make predictions or classifications. By providing meaningful and informative features, the models can better understand the underlying structure and relationships within the data, leading to improved performance in the task at hand. Effective feature extraction is crucial for enhancing model accuracy, reducing computational complexity, and promoting interpretability.

Proper feature extraction can significantly impact the success of a machine learning or deep learning task. When relevant and meaningful features are extracted, the models can focus on the most informative aspects of the data, effectively reducing noise and irrelevant information. This simplifies the learning process and allows the models to make more accurate and reliable predictions or classifications. Additionally, feature extraction aids in reducing the dimensionality of the data, which can be beneficial in scenarios where high-dimensional data poses computational challenges. By representing the data in a lower-dimensional feature space, the models can handle the data more efficiently and effectively.

Furthermore, feature extraction enhances interpretability by providing a more understandable representation of the data. Instead of working with the raw input data, which may be complex and difficult to interpret, the models operate on a set of extracted

features that have meaningful semantics. This allows researchers and practitioners to gain insights into the important factors influencing the model's decision-making process, making it easier to interpret and explain the model's behavior.

5.3.1 Mel Spectrogram

The Mel spectrogram is a commonly used feature extraction technique in audio signal processing, specifically for tasks related to speech and music analysis. It is based on the Mel scale, which is a perceptual scale of pitches that approximates the human auditory system's response to different frequencies.

The Mel spectrogram is derived from the traditional spectrogram, which represents the magnitude of frequencies over time. However, the Mel spectrogram applies a nonlinear transformation to the frequency axis to better match the human perception of sound. This transformation is achieved by dividing the frequency range into a set of Mel filter banks, which are triangular-shaped filters with overlapping frequency ranges. The energy within each filter bank is summed, resulting in a Mel-scale representation.

The Mel spectrogram provides a more meaningful representation of audio signals by emphasizing important frequency regions while suppressing less relevant ones. It is particularly useful for capturing spectral characteristics that are relevant for human perception, such as formants in speech or harmonic content in music. The Mel spectrogram is often used as input for various machine learning models, including deep learning architectures, to analyze and classify audio signals in applications such as speech recognition, music genre classification, and sound event detection.

5.4 MODEL TRAINING

Model training is a fundamental and critical step in the field of deep learning, where a neural network model learns from labeled training data to make predictions or classify new, unseen data. During the training process, the model iteratively adjusts its internal parameters based on the input data and the corresponding expected outputs, optimizing them to minimize the discrepancy between predicted and actual values.

The significance of model training in deep learning stems from several key reasons. Firstly, it enables the model to learn complex patterns and relationships in the data. Deep learning models, with their ability to learn hierarchical representations, can capture intricate features and non-linear dependencies in the data, making them well-suited for tasks such as image recognition, Natural language processing, and speech recognition. Through training, the model can extract meaningful representations and acquire the knowledge necessary to perform accurate predictions on unseen examples.

Another important aspect of model training is transfer learning. Transfer learning leverages the knowledge learned from a pre-trained model on a large dataset and applies it to a new task or domain. By starting with a model that has already learned general features and patterns, the training process for the specific task can benefit from this initial knowledge. Transfer learning reduces the need for large amounts of labeled data and decreases training time, making it particularly useful when working with limited resources or time constraints.

Furthermore, model training facilitates the process of fine-tuning. Fine-tuning involves taking a pre-trained model and further optimizing its parameters using a smaller, domain-specific dataset. This approach allows the model to adapt to the specific characteristics and nuances of the new data, improving its performance on the target task. Fine-tuning is especially valuable when the available labeled data for the specific

task is limited, as it allows the model to leverage the general knowledge it has acquired during pre-training.

The process of model training involves an iterative optimization algorithm known as backpropagation. This algorithm computes the gradients of the model's parameters with respect to a defined loss function, which quantifies the discrepancy between the predicted outputs and the true labels. By iteratively updating the model's parameters in the opposite direction of the gradients, the model gradually improves its predictions over time. This optimization process typically involves the use of optimization techniques such as Stochastic Gradient Descent (SGD) or its variants, which help efficiently navigate the high-dimensional parameter space and find the optimal set of weights for the model.

In addition to the technical aspects, model training also requires careful consideration of hyperparameters. Hyperparameters, such as learning rate, batch size, and regularization parameters, influence the training dynamics and the final performance of the model. Tuning these hyperparameters is crucial to achieve optimal training results and avoid issues such as overfitting or underfitting.

5.4.1 VGG16

VGG16, short for Visual Geometry Group 16 as shown in Figure 5.1, is a Convolutional Neural Network (CNN) architecture that has been widely used in computer vision tasks, including image classification and object recognition. It was developed by the Visual Geometry Group at the University of Oxford.

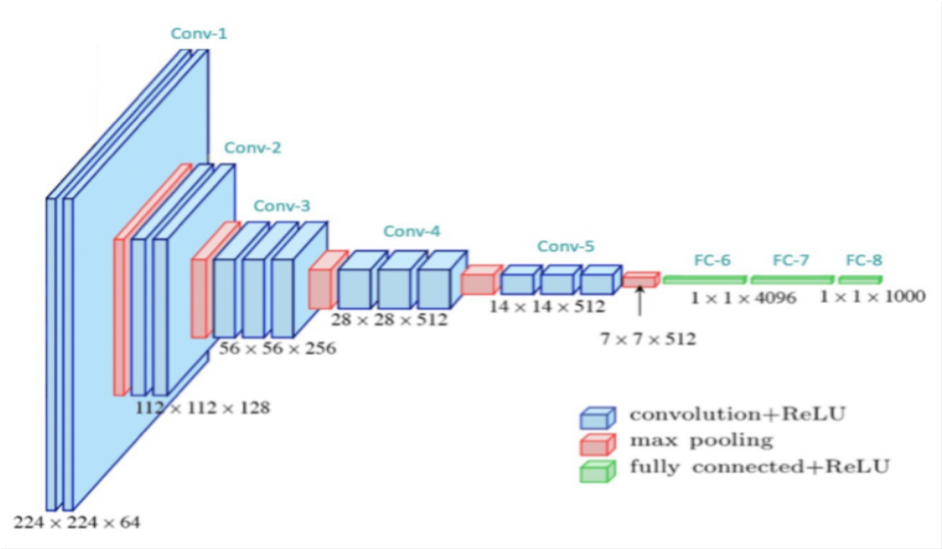

Figure 5.1 VGG16 Architecture

The VGG16 architecture is known for its simplicity and effectiveness. It consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. The convolutional layers use small 3x3 filters with a stride of 1, and they are stacked on top of each other. This repetitive stacking of layers helps capture increasingly complex features in the input image.

One of the key contributions of VGG16 is its deep architecture, which allows for the learning of highly abstract features from images. By having more layers, VGG16 can learn hierarchical representations of images, starting from low-level features like edges and textures and gradually progressing to higher-level features such as shapes and objects. VGG16 has achieved remarkable performance on various image classification benchmarks, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014. It has become a popular choice as a pretrained model for transfer learning, where the learned weights from the model trained on large-scale datasets can be used as a starting point for training on new, smaller datasets with similar tasks. This allows for faster convergence and improved performance on new tasks.

5.4.2 ResNet50

ResNet-50 is a popular deep convolutional neural network architecture that belongs to the ResNet (Residual Network) family. It was introduced by Microsoft Research in 2015 and has since become widely used in various computer vision tasks, including image classification, object detection, and image recognition. The "50" in ResNet-50 represents the number of layers in the network, indicating its depth. ResNet-50 is deeper than its predecessors, such as ResNet-18 and ResNet-34, which allows it to capture more complex features and learn hierarchical representations of images. The architecture incorporates skip connections or shortcut connections that introduce residual learning, enabling the network to overcome the degradation problem that occurs when networks become too deep. Figure 5.2 represents the diagram of ResNet50.

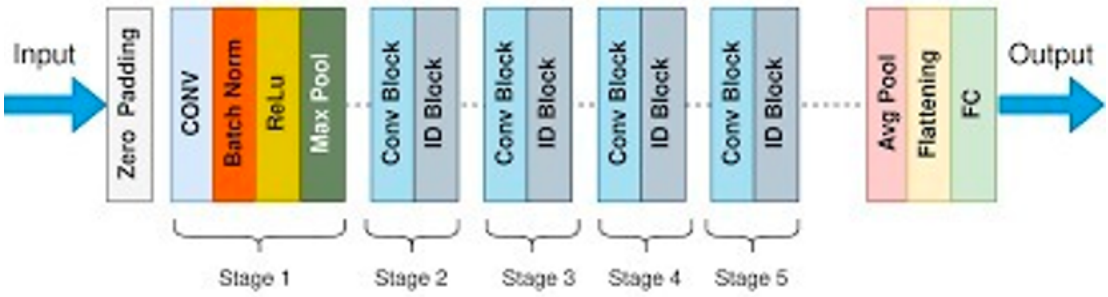

Figure 5.2 ResNet50 Architecture

ResNet-50 comprises convolutional layers, pooling layers, fully connected layers, and residual blocks. The residual blocks are the key component of ResNet-50, which contain multiple convolutional layers with identity shortcuts. These shortcuts allow the network to learn residual functions, enabling easier optimization and alleviating the vanishing gradient problem. By leveraging the ResNet-50 architecture, deep learning models can achieve impressive accuracy on large-scale image classification tasks. The pre-trained ResNet-50 model, trained on large datasets like ImageNet, can be fine-tuned or used as a feature extractor for various computer vision applications, providing a powerful tool for image analysis and recognition tasks.

5.4.3 EfficientNet-B0

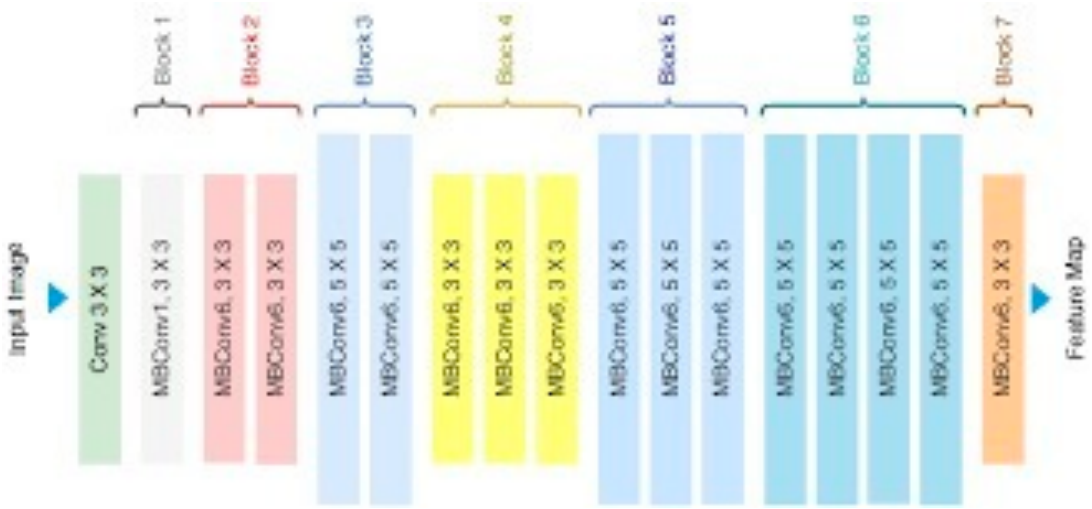

Figure 5.3 EfficientNet Architecture

EfficientNet-B0 is a convolutional neural network architecture as shown in Figure 5.3, that has gained attention for its superior performance and efficiency in deep learning tasks. It is part of the EfficientNet family of models, which were developed to provide a scalable and well-balanced solution for image classification tasks. EfficientNet-B0 is characterized by a compound scaling method that uniformly scales the network's depth, width, and resolution. This scaling strategy allows the model to achieve high accuracy while minimizing computational resources and parameters.

The "B0" designation refers to the baseline model, which is designed to strike a balance between model size and performance. One key feature of EfficientNet-B0 is its use of a Mobile inverted Bottleneck Convolution (MBConv) block, which consists of depth-wise convolutions, expansion convolutions, and squeeze-and-excitation operations. These operations effectively capture and model spatial and channel dependencies within the data, enabling the network to learn complex patterns and features.

EfficientNet-B0 has demonstrated impressive performance across various benchmark datasets, outperforming previous state-of-the-art models while being significantly more computationally efficient. Its efficient design makes it suitable for deployment on resource-constrained devices, such as mobile phones or embedded systems, without sacrificing accuracy. Researchers and practitioners often use EfficientNet-B0 as a starting point for more advanced EfficientNet models by scaling up the network's depth, width, and resolution, allowing for even better performance in image classification tasks.

5.5 VALIDATION

Validation refers to a subset of the data that is used to assess the performance and tune the hyperparameters of a machine learning model during the training phase. When splitting a dataset, typically three subsets are created: a training set, a validation set, and a test set. The training set is used to train the model, the test set is used to evaluate the final performance of the trained model, and the validation set is used to fine-tune the model and make decisions regarding hyperparameter selection, model architecture, and feature selection. During training, the model is iteratively optimized using the training set, and its performance on the validation set is evaluated after each iteration. This allows for monitoring the model's performance and making adjustments to improve its generalization and avoid overfitting. The validation set serves as an unbiased evaluation metric during model development. It helps in selecting the best-performing model based on its performance on unseen data. By assessing the model on the validation set, one can make informed decisions about modifying the model, such as adjusting hyperparameters or selecting different features. It is important to note that the validation set should be independent of the training and test sets to ensure an unbiased evaluation of the model's performance. This helps in accurately estimating the model's performance on new, unseen data.

5.5.1 K-Fold Validation

K-Fold validation is a cross-validation technique used to assess the performance and generalization of a machine learning model. It involves dividing the dataset into k equal-sized folds or subsets. The model is trained and evaluated k times, with each fold serving as both the validation set and the training set.

K-Fold validation helps in obtaining a more reliable estimate of the model's performance by reducing the impact of the dataset's partitioning. It also allows for a more thorough evaluation of the model's ability to generalize to unseen data, as each instance in the dataset is used for both training and validation.

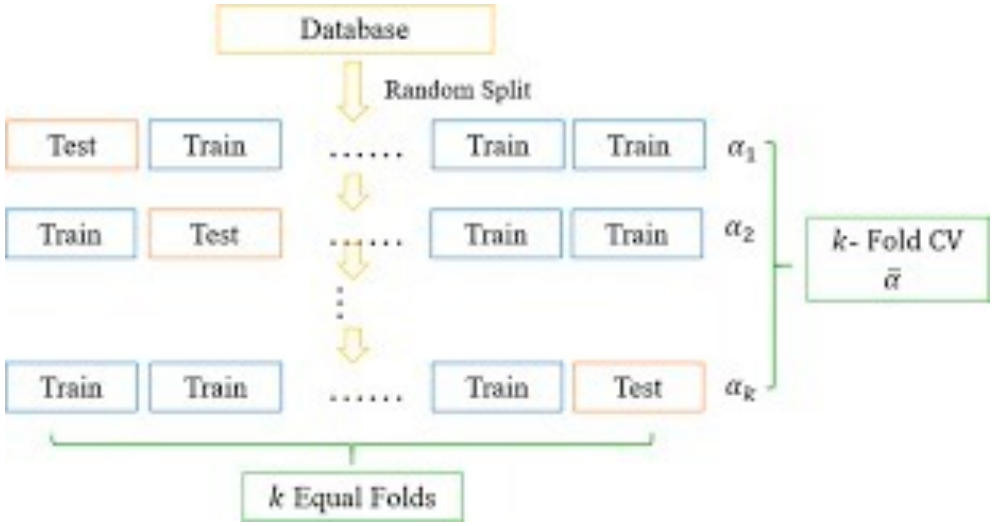

Figure 5.4 Architecture of K-Fold Validation

By performing k-fold validation, one can gain insights into the model's stability and robustness, identify potential issues like overfitting or underfitting, and make informed decisions about model selection and parameter tuning. The architecture is shown in Figure 5.4

CHAPTER 6

SNAPSHOTS OF MODULES

6.1 VGG16

6.1.1 Binary Classification

The Figures 6.1 & 6.2 represent the relationship between Accuracy and Validation Accuracy, and Confusion Matrix for Binary Classification of VGG16.

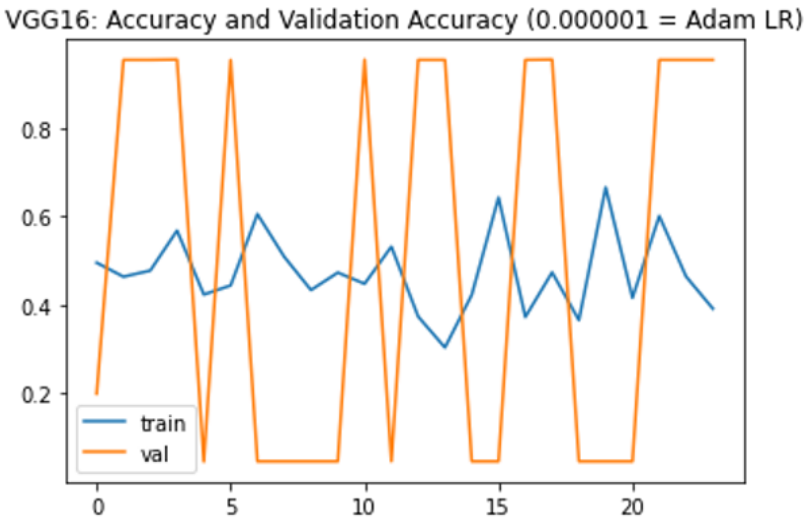

Figure 6.1 Accuracy graph for VGG16 Binary Classification

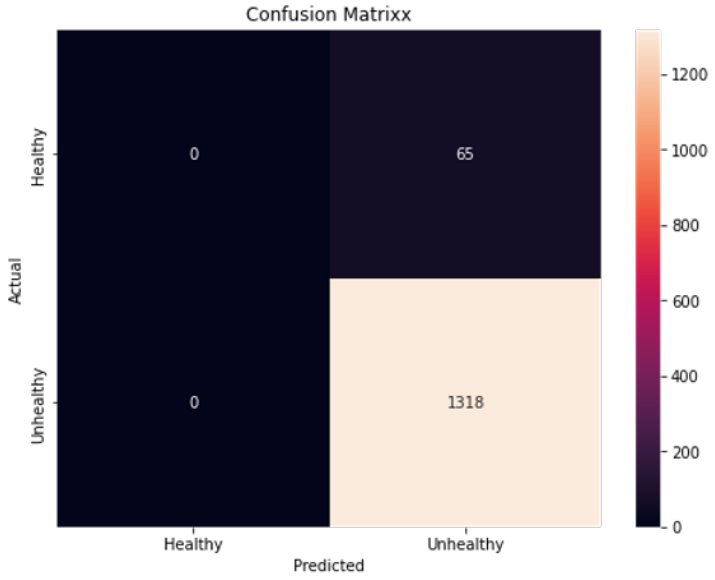

Figure 6.2 Confusion Matrix for VGG16 Binary Classification

6.1.2 Multi-class Classification

The Figures 6.3 & 6.4 represent the relationship between Accuracy and Validation Accuracy, and Confusion Matrix for Multi-class Classification of VGG16.

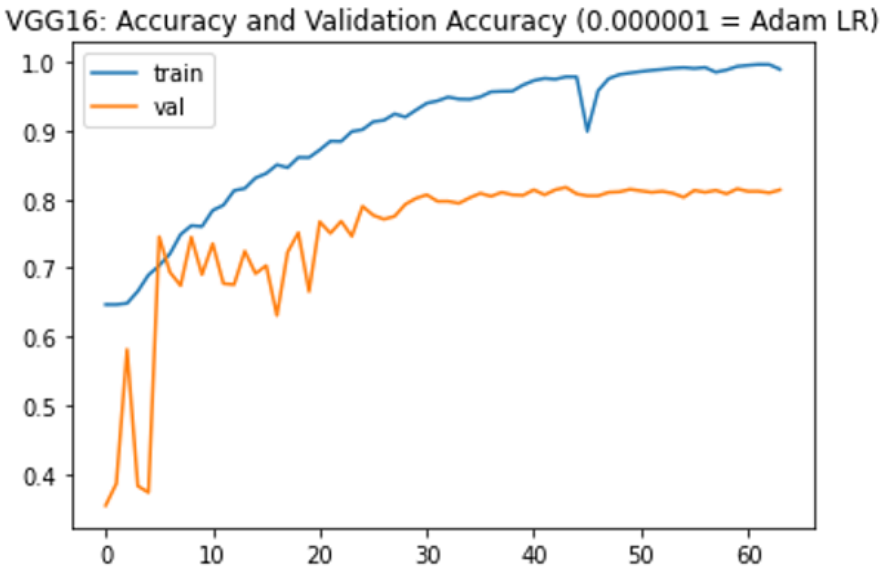

Figure 6.3 Accuracy graph for VGG16 Multi-class Classification

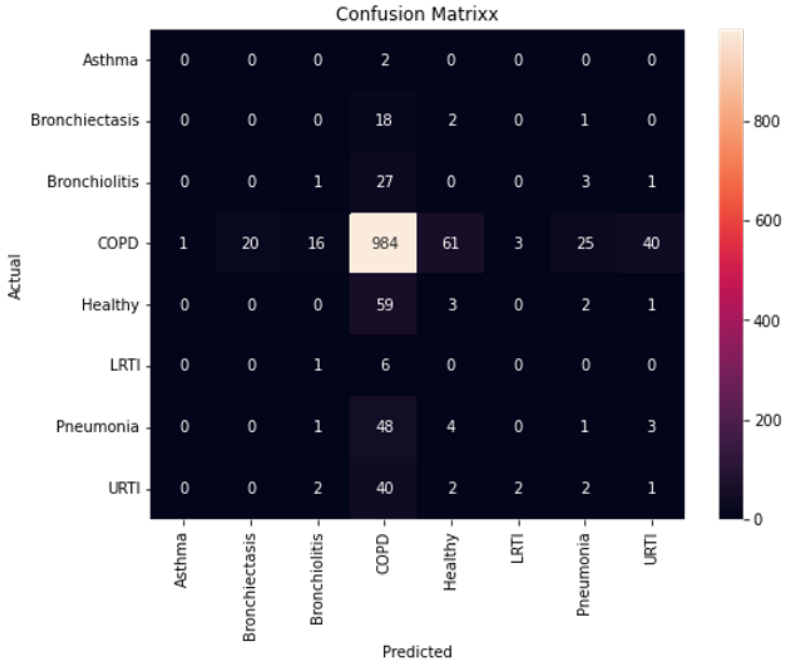

Figure 6.4 Confusion Matrix for VGG16 Multi-class Classification

6.2 RESNET50

6.2.1 Binary Classification

The Figures 6.5 & 6.6 represent the relationship between Accuracy and Validation Accuracy, and Confusion Matrix for Binary Classification of ResNet50.

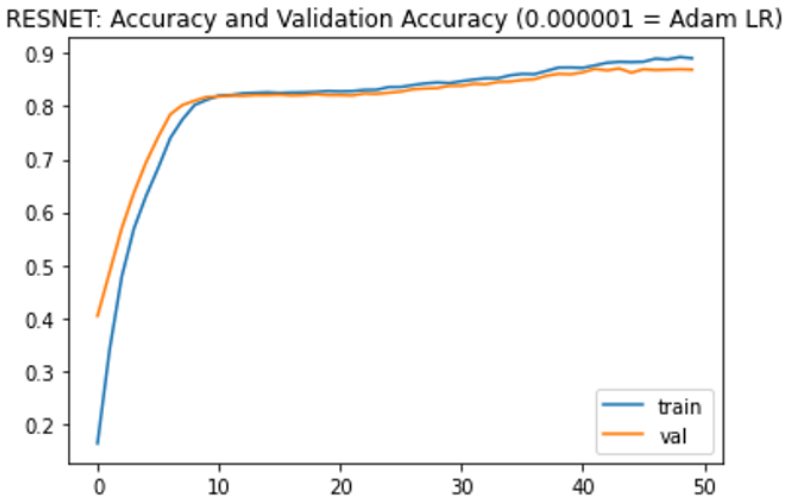

Figure 6.5 Accuracy graph for ResNet50 Binary Classification

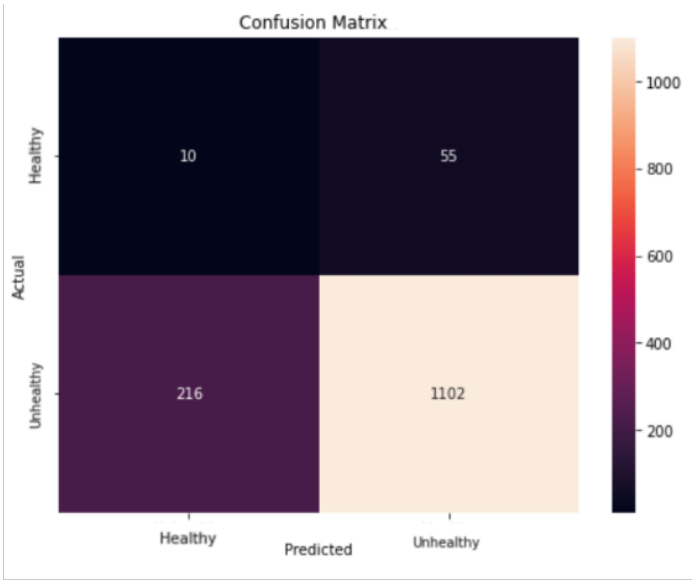

Figure 6.6 Confusion Matrix for ResNet50 Binary Classification

6.2.2 Multi-class Classification

The Figures 6.7 & 6.8 represent the relationship between Accuracy and Validation Accuracy, and Confusion Matrix for Binary Classification of ResNet50.

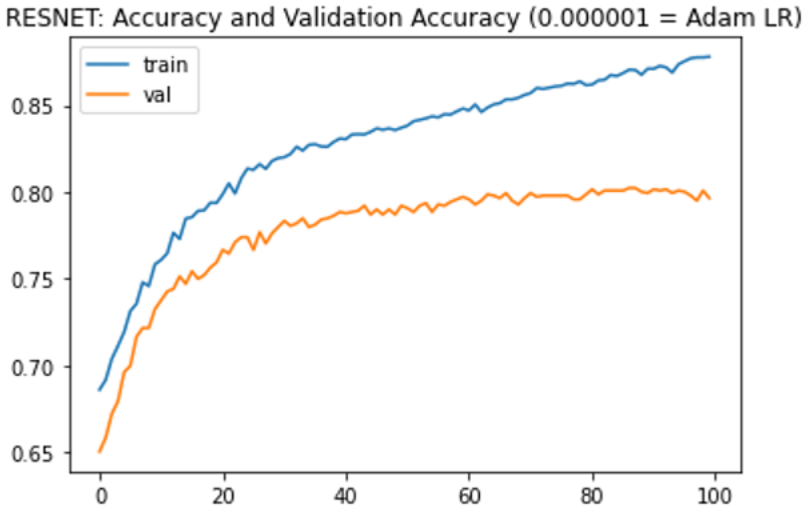

Figure 6.7 Accuracy graph for ResNet50 Multi-class Classification

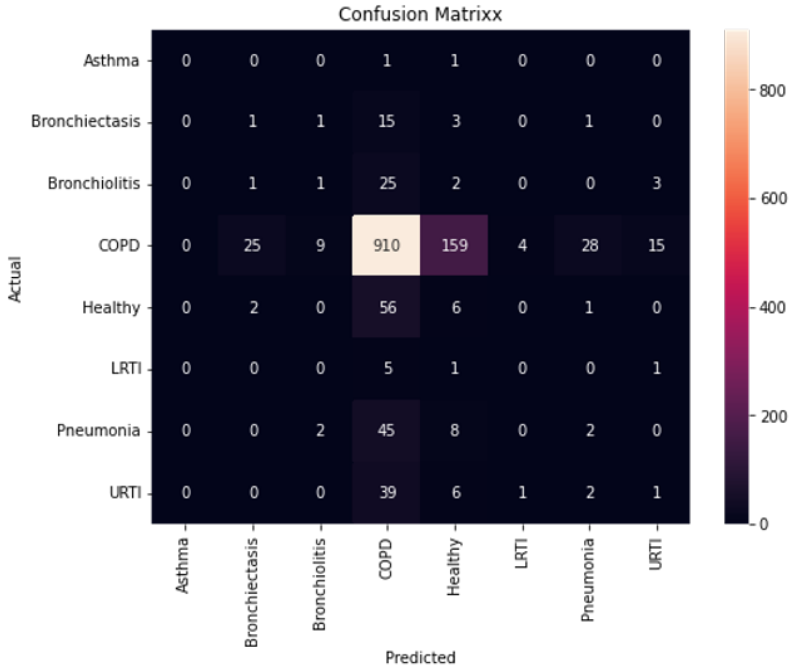

Figure 6.8 Confusion Matrix for ResNet50 Multi-class Classification

6.3 EFFICIENTNET-B0

6.3.1 Binary Classification

The figures 6.9 & 6.10 represent the relationship between Accuracy and Validation Accuracy, Confusion Matrix for Binary Classification of EfficientNet-B0.

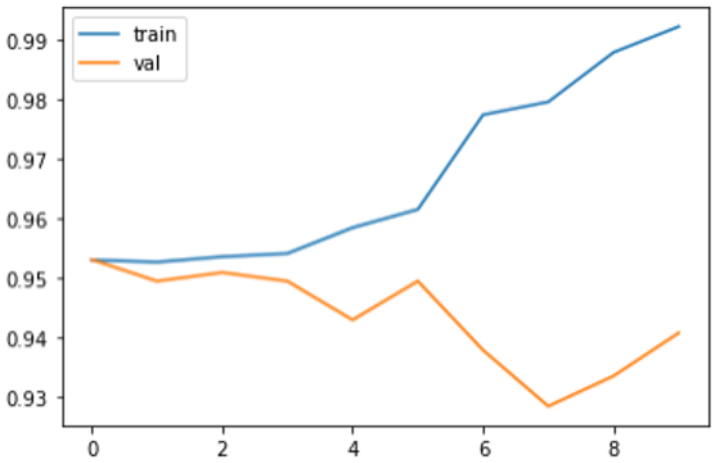

Figure 6.9 Accuracy graph for EfficientNet-B0 Binary Classification

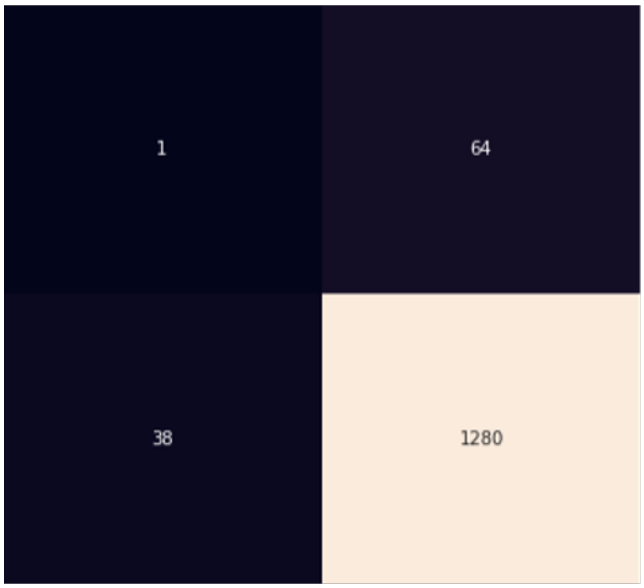

Figure 6.10 Confusion Matrix for EfficientNet-B0 Binary Classification

6.3.2 Multi-class Classification

The figures 6.11 & 6.12 represent the relationship between Accuracy and Validation Accuracy, and Confusion Matrix for Multi-class Classification of EfficientNet-B0.

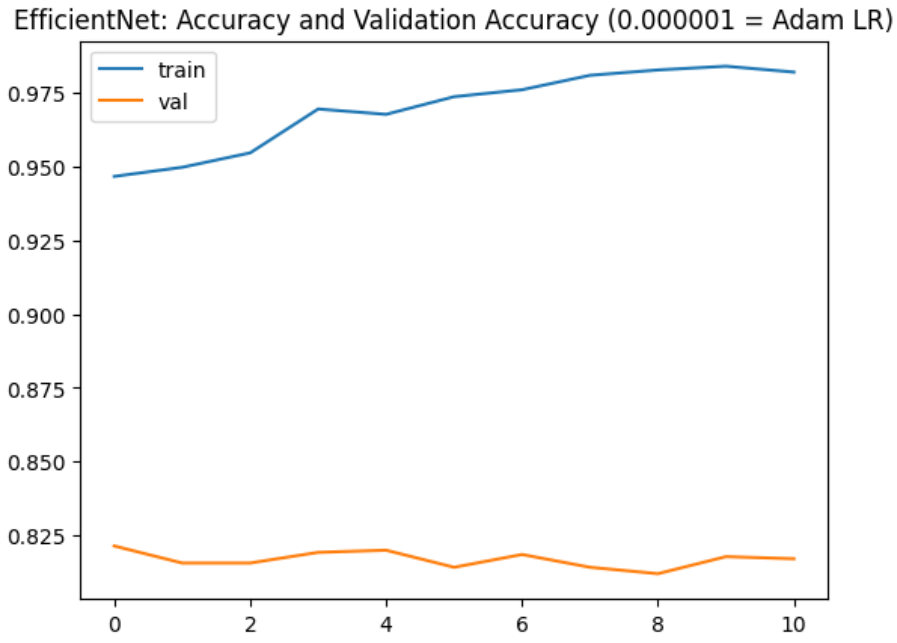

Figure 6.11 Accuracy graph for EfficientNet-B0 Multi-class Classification

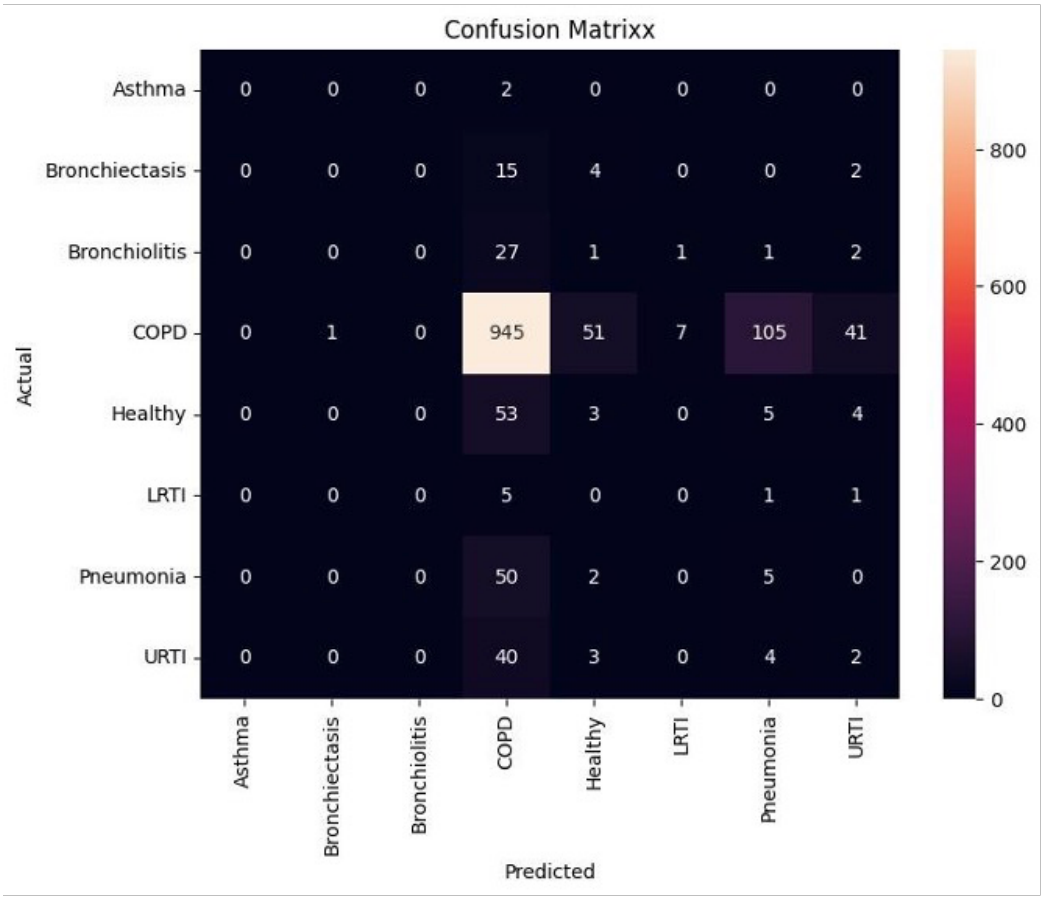

Figure 6.12 Confusion Matrix for EfficientNet-B0 Multi-class Classification

6.4 COMPARISON OF NEURAL NETWORK MODELS

Table 6.1 presents a comparison of the accuracies obtained by three deep learning models (VGG-16, ResNet-50, and EfficientNet-B0) for both binary and multi-class classification tasks. The "Model Name" column specifies the name of the model, while the "Classification Type" column indicates the type of classification being performed. The "Accuracy" column represents the corresponding accuracy achieved by each model in the specified classification task.

Table 6.1 Accuracy Comparison of Various Models

Model Name	Classification Type	Accuracy
VGG16	Binary	95%
VGG16	Multi-class (8 Class)	72%
ResNet50	Binary	80%
ResNet50	Multi-class (8 Class)	67%
EfficientNet-B0	Binary	94.07%
EfficientNet-B0	Multi-class (8 Class)	80.33%

CHAPTER 7

CONCLUSION AND FUTURE WORK

This research work proposes an efficient solution for the Detection of Pulmonary Diseases using Deep Learning techniques. VGG16, Resnet50, and EfficientNet-B0 were used for the detection of pulmonary disease for binary and multiclass (8-class) classification. It is inferred that binary classification works with an accuracy of 94% for EfficientNet-B0 whereas, for multiclass classification, the accuracy is 75%. The research stands unique as the multiclass classification of EfficientNet-B0 for Pulmonary Disease classification is the first of its kind. Audio Splitting, Feature Extraction, and Normalization were implemented in the pre-processing stage and fitted in pre-trained VGG16, Resnet50, and EfficientNet-B0 models. Also, K-Fold Validation for multi-class classification was carried out and procured with an accuracy of 80.33%.

This research can be further developed by creating an application for real-time implementation. The integration of an electronic stethoscope and the application can be a huge breakthrough in producing an efficient, cost-effective, non-invasive, sustainable, and time-saving method for the detection of pulmonary diseases. A separate own dataset with balanced data to train the model can also be prepared for increasing the accuracy.

REFERENCES

1. Acharya. J and Basu. A (2020) 'Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning' IEEE Transactions on Biomedical Circuits and Systems, Vol. 14, pp. 535-544.

2. Elmar Messner, Melanie Fediuk, Paul Swatek, Stefan Scheidl, Freyja-Maria Smolle-Jüttner, Horst Olschewski, Franz Pernkopf (2020) 'Multi-channel lung sound classification with convolutional recurrent neural networks Computers in Biology and Medicine' Science Direct, Vol. 122, pp. 834-846

3. Fatih Demir, Abdulkadir Sengur, Varun Bajaj (2020) 'Convolutional Neural Networks based efficient approach for classification of lung diseases' SpringerLink on Health Information Science and Systems, Vol. 8, pp. 1343-1355.

4. Kim, Y., Hyon, Y., Jung, S.S. et al. (2021) 'Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning' Nature-Scientific Reports, Vol. 11, pp. 2502-2512.

5. Nguyen. T and Pernkopf. F (2022) 'Lung Sound Classification Using Co-Tuning and Stochastic Normalization' IEEE Transactions on Biomedical Engineering, Vol. 69, pp. 2872-2882.

6. Pham. L, Phan. H, Palaniappan. R, Mertins. A and McLoughlin. I (2021) 'CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection' IEEE Journal of Biomedical and Health Informatics, Vol. 25, pp. 2938-2947.

7. Dataset link: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge