الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR- ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR - ANNABA

جامعة باجي مختار- عنابة

Faculté: Sciences de l'ingéniorat
Département: Electronique

Année : 2021

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat 3ème Cycle

Intitulé

Aide au Diagnostic et Identification des pathologies pulmonaires par l'analyse et traitement des signaux respiratoires

Assistance in the diagnosis and identification of pulmonary pathologies by analysis and processing of respiratory signals

Option : Multimédia et communications numériques

Par : NEILI Zakaria

DIRECTEUR DE THÈSE : FEZARI Mohamed — Pr. Université de Annaba
CO-ENCADREUR : REDJATI Abdelghani — Dr. Université de Annaba

DEVANT Le JURY

PRESIDENT :	HAMDI Rachid	Pr.	Université de Annaba
EXAMINATEURS :	BABOURI Abdesselam	Pr.	Université de Guelma
	BOUGHAZI Mohamed	Pr.	Université de Annaba

ACKNOWLEDGMENT

I have received a great deal of support and assistance throughout writing this thesis. In the name of Almighty **Allah** for guiding me, helping me and giving me the strength to overcome difficulties and to continue working during these long years. I would say that without the help of Allah this challenging journey was impossible.

I would like to express my unlimited appreciation to my supervisor **Prof. FEZARI Mohamed** for his motivation for working, and sincere willingness to solve problems that have made this work to be completed within the expected time. I would also like to thank my co-supervisor **Dr. REDJATI Abdeghani** for his meaningful support during this work.

My sincere thanks go out to **Prof. Kenneth Sundaraj** from UTeM university in Malayasia, whose advice and his continuous inspirations for working and publishing papers were extremely valuable for this thesis. I am grateful for all I learned from him and for the feedback he provided to improve my work.

I would like to thank the president of the jury, **Prof. HAMDI Rachid**, from ANNABA university, who does me the honour of chairing a jury. I also thank the jury members, **Prof. BABOURI Abdesselam**, from GUELMA University, **Prof. BOUGHAZI Mohamed**, from ANNABA University, who have been kind enough to judge this work.

Many thanks to all my LASA laboratory colleague: **MANSRI Islem**, **ALIOUAT Ahcen**, **KEBIR abdeljalil** and **MEKAHLIA mohammed saddek** which encourage me to do work hardly and I thank you very much my friends.

I owe a big thanks to **Professor Noureddine DOGHMANE**, our former laboratory director, for his passion for scientific exploration which will always be my inspiration in the future, and I owe thanks to **Professor Messaoud RAMDANI**, our current laboratory director.

DEDICATION

I dedicate this modest work to my family. A special feeling of gratitude to my loving parents, thank you both for making me a man which could reach the stars and chase his dreams. I give to them very special thanks, which they encouraged and push me to continue doing this work and they have never left my side since I started the study journey.

Contents

ACKNOWLEDGMENT iii
Contents v
LIST OF TABLES viii
LIST OF ABBREVIATIONS xi
ملخص xiii
ABSTRACT xiv
Keywords xiv
RÉSUMÉ xv
Mots clés xv
CHAPTER 1 1
INTRODUCTION 1
1.1. Research Background 1
1.2. Motivation for the work 1
1.3. Problem Statement 2
1.4. Objectives 2
1.5. Thesis Scope 5
1.6. Thesis Organization 6
CHAPTER 2 9
LITERATURE REVIEW 9
2.1. Human Respiratory System Physiology 9
2.2. Breathing system diseases and the Causes 10
2.3. Corona Virus Disease 2019 (COVID-19) 11
2.4. Analysis of Computerized breathing sounds 12
2.4.1. Characteristics and Types of Breathing Sounds 12
2.4.2. Breathing Sound Data 14
2.4.3. Lung Sounds Data Acquisition 14
2.4.3.1. Devices and Sensors for lung sound Data Collection 15
2.4.3.2. Lung sounds data collection techniques 15
2.4.4. Pre-processing of Breathing Sounds 16
2.4.5. Feature Extraction methods in Breathing Sound Analysis 17
2.5. Artificial Intelligence 17
2.6. Machine Learning Algorithms in Breathing Sound Analysis 18
2.7. Deep Learning Algorithms in Breathing Sound Analysis 19
2.8. Research Gap 23
CHAPTER 3 26
METHODOLOGY 26

3.1. Methodology ... 26
3.1.1. Part 1: Machine Learning Algorithms in Breathing Sounds Classification ... 26
3.1.1.1 Database ... 27
3.1.1.2 Breath Sounds Pre-processing ... 28
3.1.1.3 Empirical mode decomposition ... 28
3.1.1.4 Features Extraction ... 30
3.1.1.5 Statistical analysis ... 30
3.1.1.6 Classification ... 31
3.1.1.7 Extreme Learning Machine ... 31
3.1.1.8 k-nearest neighbour ... 34
3.1.2. Part 2: Deep Learning Algorithms in Breathing Sounds Classification ... 35
3.1.2.1. Scenario (i) - symptoms-based ... 38
3.1.2.2. Scenario (ii) - conditions-based ... 39
3.1.2.3. Scenario (iii) – diseases-based ... 40
3.1.2.4. Gammatone Filter Bank ... 41
3.1.2.5. Deep learning models ... 42
3.1.2.6. Proposed Method ... 48
3.1.2.7. Datasets ... 57
3.1.2.8. Computing Platform ... 60
3.1.2.9. Imaging Software ... 61
3.1.2.10. Performance evaluation criteria ... 61
CHAPTER 4 ... 64
RESULTS AND DISCUSSION ... 64
4.1. Part 1: Performance of Machine Learning Algorithms in Breathing Sounds Classification ... 64
4.2. Part 2: Performance of Deep Learning Algorithms in Breathing Sounds Classification 68
4.2.1. Scenario (i) – Symptoms-based ... 68
4.2.1.1. Results ... 68
4.2.1.2. Discussion ... 75
4.2.2. Scenario (ii) – Conditions-based ... 78
4.2.2.1. Results ... 78
4.2.2.2. Discussion ... 86
4.2.3. Scenario (iii) – Diseases-based ... 89
4.2.3.1. Results ... 89
4.2.3.2. Discussion ... 94
Conclusion ... 98
REFERENCES ... 99

[88] FIZ, José Antonio, JANÉ, Raimon, SALVATELLA, David, et al. Analysis of tracheal sounds during forced exhalation in asthma patients and normal subjects: bronchodilator response effect. Chest, 1999, vol. 116, no 3, p. 633-638. 103

LIST OF PUBLICATIONS 104

LIST OF TABLES

NO **PAGE**

2.1 Time-frequency representation (TFR) in computer-based lung sound classification systems. 23
3.1 The results of binary classification using X-ray images data. 52
3.2 Results taken from [63]. 53
3.4 Example of recording sounds characteristics from the ICBHI database. 57
3.5 Dataset summary (Conditions-based). 59
3.6 Dataset summary (Diseases-based) 60
3.7 Hyperparameters settings for trained VGG16, ResNet-50, AlexNet and GoogLeNet models. 61
4.1 Statistical analysis of features extracted from breath sounds 65
4.2 Classification performance of (Activity, PE) from IMFs of BS signals in multiclass classification stage 66
4.3 Classification performance of (Activity, PE) from IMFs of BS signals in binary classification stage 67
4.4 VGG16 performance of Gammatonegrams cycles classification. 70
4.5 Precision, Recall and F1-Score comparison between two Experiments. 70
4.6 ResNet-50 performance of Gammatonegrams cycles classification. 71
4.7 Precision, Recall and F1-Score comparison between two Experiments (ResNet-50) 72
4.8 GoogLeNet performance of Gammatonegrams cycles classification. 73
4.9 Precision, Recall and F1-Score comparison between two Experiments (GoogLeNet) 75
4.10 A comparison result between proposed and fixed methods using different CNN models (Symptoms-based). 76
4.11 VGG16 performance of Gammatonegrams cycles classification. 80
4.12 Precision, Recall and F1-Score comparisons. 81
4.13 ResNet-50 performance of Gammatonegrams cycles classification. 82
4.14 Precision, Recall and F1-Score comparisons (ResNet-50). 83
4.15 AlexNet performance of Gammatonegrams cycles classification. 83
4.16 Precision, Recall and F1-Score comparisons (AlexNet). 84
4.17 Comparison results between the proposed and fixed methods using different CNN models. 85
4.18 VGG16 performance of Gammatonegrams recordings classification ....
without and with Multi ep-Batch method. 90
4.19 Precision, Recall and F1-Score comparison between two Experiments (VGG16) 92
4.20 AlexNet performance of Gammatonegrams recordings classification 93
without and with Multi ep-Batch method. 93
4.21 Precision, Recall and F1-Score comparison between two Experiments. 94
4.22 A comparison result between proposed and fixed methods using diff erent CNN models. 95

LIST OF FIGURES

NO **PAGE**

2.1 Human Respiratory System [7] ... 10
2.2 General process steps for computerized breathing sound analysis ... 12
3.1 Two principal stage of the breath sounds signal classification (a) multiclass classification (b) binary classification. ... 27
3.2 SLFN: additive hidden nodes. ... 32
3.3 Block diagram of the proposed Multi ep-Batch, VGG16, ResNet-50 and GoogLeNet based symptoms-classification. ... 39
3.4 Flowchart of the proposed VGG16, ResNet-50 and AlexNet based conditions-classification. ... 40
3.5 Flowchart of the proposed VGG16 and AlexNet, (a) without the Multi ep-Batch method and (b) with the Multi ep-Batch method. ... 41
3.6 Visual Geometry Group 16 (VGG16) based CNN architecture. ... 44
3.7 Residual network based on CNN architecture. ... 45
3.8 Structure of the identity and conv blocks in the ResNet architecture. ... 45
3.9 The overview of our ResNet architecture used in this work. ... 46
3.10 Conv2 and Conv5 blocks of the ResNet-50 architecture. ... 47
3.11 AlexNet based CNN architecture. ... 48
3.12 The training and test process using Fixed method (standard). ... 49
3.13 The training and test process using our proposed Multi ep-Batch ... 51
3.14 The procedure followed for getting the optimal weight during learning phases to every model. ... 54
3.15 Distribution of the cycles data for normal, crackle, wheeze. ... 58
4.1 Gammatonegram feature maps of (a) crackle, (b) normal and (c) wheeze sounds. ... 69
4.2 Confusion matrixes for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 70
4.3 Confusion matrixes for the ResNet-50 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 72
4.4 Confusion matrixes for the GoogLeNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 74
4.5 Accuracy and Learning time distribution for both fixed (standard) and multi ep-batch methods based on VGG16, ResNet-50 and GoogLeNet architectures. ... 76
4.6 Feature maps based-Gammatonegram applied for (a) Healthy, (b) Chronic and (c) Non-Chronic pulmonary conditions. ... 79
4.7 Confusion matrices for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 80
4.8 Confusion matrices for the ResNet-50 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 82
4.9 Confusion matrices for the AlexNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch. ... 84

4.10 Accuracy distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16, ResNet-50 and AlexNet architectures.. 85

4.11 Training time distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16, ResNet-50 and AlexNet architectures.. 86

4.12 Feature map based-Gammatonegram applied for Healthy recording... 90

4.13 Confusion matrixes for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.. 91

4.14 Confusion matrixes for the AlexNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.. 94

4.15 Accuracy distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16 and AlexNet architectures.. 95

4.16 Training time distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16 and AlexNet architectures.. 96

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
ARNN	Attractor Recurrent Neural Network
Adam	Adaptive Moment Estimation
BiGRU	Bidirectional Gated Recurrent Unit neural network
BS	Breathing Sounds
CORSA	Computerized Respiratory Sound Analysis
CNN	Convolutional Neural Network
CHF	Congestive Heart Failure
COPD	Chronic Obstructive Pulmonary Disease
CWT	Continuous Wavelet Transforms
ELM	Extreme learning machine
EMD	Empirical Mode decomposition
ERB	Equivalent Rectangular Bandwidth
FF	Fuzzy Functions
GMM	Gaussian mixture model
GPUs	Graphic Processing Units
HHT	Hilbert-Huang Transform
HD	Hjorth descriptors
ICBHI	International Conference on Biomedical and Health Informatics
IPF	Idiopathic Pulmonary Fibrosis
IMFs	Intrinsic Mode Functions
ILSVRC14	ImageNet Large-Scale Visual Recognition Challenge 2014
K-nn	K-nearest neighbour

LRTI	Lower Respiratory Tract Infection
MARS	Marburg Respiratory Sounds
MFCC	Mel-Frequency Cepstral Coefficient
Multi ep-Batch	Multi epochs and Batch sizes
MLP	Multilayer Perceptron
OST	Optimized Stockwell Transform
PE	Permutation Entropy
PPV	Positive Predicted Value
RBF	Radial Basis Function
RQA	Recurrent Quantification Analysis
ResNets	Deep Residual Networks
RNN	Recurrent Neural Network
R.A.L.E	Respiration Acoustic Laboratory Environment
SVM	Support Vector Machine
SD	Standard Deviation
SLFNs	Single hidden Layer Feedforward Neural Networks
STFT	Short-Time Fourier Transform
SGD	Stochastic Gradient Descent
TPUs	Tensor Processing Units
TFR	Time Frequency Representation
VGG16	Visual Geometry Group 16
URTI	Upper Respiratory Tract Infection
WHO	World Health Organization

ملخص

تحتوي أصوات التنفس على معلومات مفيدة يمكن أن تساعد الأطباء في تشخيص أمراض الرئة بطريقة غير مكلفة. هذه الأطروحة هي دراسة بحثية تهدف إلى تطوير نظام قائم على تصنيف أمراض الرئة باستخدام أصوات التنفس. بناءً على هذه الأصوات ، يمكننا إنشاء العديد من الخوارزميات لتطوير نظام تصنيف آلي يمكن استخدامه لتصنيف أمراض الرئة. تحتوي الطرق التقليدية التي يستخدمها الباحثون عند تنفيذ هذه الأنظمة الألية خطوتين رئيسيتين - استخراج الميزات وتصنيف الفئات. في السنوات الأخير، ظل الإهتمام بمجال تصنيف أصوات الرئة من خلال استخدام الشبكات العصبية العميقة كبيرا، والتي أثبتت فعاليتها في تدريب مجموعة البيانات الكبيرة. في هذا العمل ، قمنا بتنفيذ جزأين رئيسيين : الجزء التجريبي يعتمد على خوارزميات التعلم الآلي التقليدية والآخر مخصص لاستخدام تقنيات التعلم العميق.

أولاً ، قمنا بمقارنة قدرة خوارزمية (ELM) وخوارزمية (K-nn) في تصنيف أصوات الرئة الغير طبيعية والطبيعية ، باستخدام تقنيات EMD و Hjorth descriptors (Activity) و Permutation Entropy (PE) والتي تم استخراجها كميزات. وجدت الدراسة أن الجمع بين السمات Hjorth descriptors (Activity) و Permutation Entropy (PE) أنتجت دقة 90.71٪ ، 95٪ باستخدام ELM و K-nn على التوالي في التصنيف الثنائي ، و 83.57٪ ، 86.42٪ في التصنيف متعدد الطبقات.

ثانيًا ، قمنا بإقتراح طريقة جديدة تتضمن استخدام Gammatonegrams كمدخل للشبكة العصبية العميقة واستخدام Batchs وEpochs كقيم متعددة (Multi ep-Batch) أثناء خطوات التدريب والاختبار لتصنيف أصوات الرئة. أجرينا تجارب لأربع شبكات CNN شهيرة معززة بأسلوب Multi ep- Batch وأيضا باستخدام ثلاث قواعد بيانات فرعية من قاعدة بيانات كبيرة إسمها ICBHI. تم معالجة التسجيلات الرقمية لأصوات الرئة للحصول على صور gammatonegram التي يتم إرسالها كمدخلات لشبكات CNN الأربعة. أشارت نتائج التصنيف إلى أن الطريقة المقترحة هي أفضل استراتيجية أداء لعملية التدريب والاختبار. بالتالي ، فإن النتائج تقدم دليلًا أوليًا على أن طريقة التعلم Multi ep-Batch المقترحة يمكن أن تحسن بشكل كبير من وقت التدريب والدقة باستخدام شبكات عصبية عميقة مختلفة دون الحاجة إلى العديد من موارد الأجهزة.

الكلمات المفتاحية

تعلم الألة , التعلم العميق , تحليل الوضع التجريبي , واصفات Hjorth ,ميزة الأنتروبي , تصنيف أصوات الرئة , متعدد الحجوم و المرات , تمثيل للصوت في الزمن و التردد.

ABSTRACT

Breathing sounds contain prominent information that can aid doctors to diagnose pulmonary pathologies in a non-invasive way. This thesis aimed at developing a classification system of pulmonary pathologies using breathing sound signals. Based on these sounds, we can establish many algorithms to develop an automatic classification system that could be used to categorize lung diseases. The traditional methods used by researchers when implementing these systems involve two main steps – feature extraction and pattern classification. In recent years, the topic of interest in the field of breathing sound classification focuses on the use of deep neural networks, which have been proven to be effective for training large datasets. In this work, we implemented two main parts of experimental one based on traditional machine learning algorithms and the other one is focused on using deep learning techniques.

Firstly, we compared the ability of the extreme learning machine (ELM) and k-nearest neighbour (K-nn) machine learning algorithms in the classification of adventitious and normal breath sounds, using EMD decomposition techniques and the Hjorth descriptors (Activity) and Permutation Entropy (PE) were extracted as features. The study has found that the combination of features (activity and PE) yielded an accuracy of 90.71%, 95% using ELM and K-nn respectively in binary classification, and 83.57%, 86.42% in multiclass classification.

Secondly, we proposed a novel method that involves the use of Gammatonegrams as input to the deep neural network and the use of multi epochs and batch sizes (Multi ep-Batch) during the training and testing steps of the lung sounds classification. We have performed experiments with four popular CNNs architectures enhanced with the Multi ep-Batch method using three sub-data from the ICBHI database. Digital recordings of lung sound are processed to obtain gammatonegram images that are fed as an input to the four CNNs architectures. The classification results, indicated that the proposed method it's the best performing strategy for the training and testing process. Hence, the results provide initial evidence that the proposed Multi ep-Batch learning method can significantly improve the training time and accuracy using different deep neural networks without the need for many hardware resources.

Keywords

Machine Learning; Deep learning; Empirical Mode Decomposition; Hjorth descriptors; Permutation Entropy; Lung sounds classification; Multi ep-batch; Gammatonegrams.

RÉSUMÉ

Les sons respiratoires contiennent des informations importantes qui peuvent aider les médecins à diagnostiquer les pathologies pulmonaires de manière non-invasive. Cette thèse vise à développer un système de classification des pathologies pulmonaires à l'aide de signaux respiratoires. Sur la base de ces signaux, nous pouvons établir de nombreux algorithmes pour développer un système de classification automatique qui pourrait être utilisé pour catégoriser les maladies pulmonaires. Les méthodes traditionnelles utilisées par les chercheurs lors de la mise en œuvre de ces systèmes impliquent deux étapes principales: l'extraction des caractéristiques et la classification des motifs. Ces dernières années, le sujet d'intérêt dans le domaine de la classification des sons respiratoires se concentre sur l'utilisation de réseaux de neurones profonds, qui se sont avérés efficaces pour entraîner de grands ensembles de données. Dans ce travail, nous avons mis en œuvre deux parties principals. Une partie expérimentale basée sur des algorithmes d'apprentissage automatique traditionnels et l'autre est axée sur l'utilisation de techniques d'apprentissage en profondeur.

Tout d'abord, nous avons comparé la capacité des algorithmes d'apprentissage automatique de la machine d'apprentissage extrême (ELM) et du k-plus proche voisin (K-nn) dans la classification des bruits respiratoires fortuits et normaux, en utilisant des techniques de décomposition EMD et les descripteurs de Hjorth (activité) et permutation de l'entropie (PE) a été extraite en tant que caractéristiques. L'étude a révélé que la combinaison des caractéristiques (activité et PE) donnait une précision de 90,71%, 95% en utilisant respectivement ELM et K-nn en classification binaire, et 83,57%, 86,42% en classification multiclasse.

Puis, nous avons proposé une nouvelle méthode qui implique l'utilisation de Gammatonegrams comme entrée du réseau neuronal profond et l'utilisation de plusieurs epochs et batch sizes (Multi ep-Batch) pendant les étapes de l'entraînement et de test de la classification des sons pulmonaires. Nous avons effectué des expériences avec quatre architectures CNN populaires améliorées avec la méthode Multi ep-Batch en utilisant trois sous-données de la base de données ICBHI. Les enregistrements numériques du son pulmonaire sont traités pour obtenir des images gammatonegram qui sont alimentées en entrée des quatre architectures CNN. Les résultats de la classification indiquent que la méthode proposée est la stratégie la plus performante pour le processus de l'entraînement et de test. Par conséquent, les résultats fournissent une première preuve que la méthode d'apprentissage Multi ep-Batch proposée peut considérablement améliorer le temps et la précision de l'entraînement en utilisant différents réseaux de neurones profonds sans avoir besoin de nombreuses ressources matérielles.

Mots clés

Apprentissage automatique; l'apprentissage en profondeur; Décomposition en mode empirique; Descripteurs de Hjorth; Entropie de permutation; Classification des sons pulmonaires; multi ep-batch; gammatonégrammes.

CHAPTER 1

INTRODUCTION

1.1. Research Background

listening to the lung sounds using a traditional stethoscope is a clinical procedure which called auscultation and was developed firstly in 1816 [1]. In order to diagnosing the lung diseases in pulmonary medicine field, the physicians listen carefully to the lung sounds in a human body by using a stethoscope. These respiratory sounds contain a prominent and powerful information regarding pulmonary conditions [1]. Based on [2], lung sounds are closely related to pulmonary diseases. Using a stethoscope is non-invasive and economical way to diagnosis the patients, however, to classify the pulmonary pathologies the doctors need to detect some symptoms which related to a specific disease, and due to breathing manoeuvre it's difficult to monitoring the patients continuously, also cannot be saved their lung sounds to maintain their history.

Furthermore, the doctor's auscultation methods depend on the experience of professional to detect abnormal in underline disease. The relation between the patient and doctor auscultation plays a circular role to make a decision concerning the pulmonary conditions of the patients. Besides, cannot be done at home the personal management of lung sounds. A computerized respiratory sound analysis started from 1980 to address these drawbacks.

1.2. Motivation for the work

Pulmonary diseases are the third largest cause of death in the world [3], and five major pulmonary diseases were mentioned by the World Health Organization (WHO) [4], namely acute lower respiratory tract infection, asthma, lung cancer, tuberculosis, and chronic obstructive pulmonary disease. During the past decade, major attempts have been made towards the development of a system for the automated classification of abnormal lung sounds. The self-management and self-monitoring of the pulmonary pathologies important and are becoming more necessary. To do so, an electronic stethoscope as possible solution would be capable to record and computerized the respiratory sound data, and using these data with a machine learning-based classification algorithm to detect different pulmonary diseases classes. A rapid diagnosis and non-expensive are the advantages when using the computerised respiratory sound analysis. Also, it can offer an alternative tool for medical professionals. The distinguishing between a specific pulmonary pathology suffered by a patient is a diagnosis process which could eliminate the other worsening of the patient's condition. This research study for goal is to develop a computerized respiratory sound to classify the pulmonary pathology using the lung sounds.

1.3. Problem Statement

A several problems related to the pulmonary pathology's classification from lung sounds.

The non-stationary nature of the respiratory sound signals due the variation of the lung volume making the identification and the classification difficulties, which required to use an advanced signal processing techniques which has the same nature such as EMD method.

Recently, the researchers have focused on using machine and deep learning algorithms as a classifier without modifying and developing it to improve the learning time and accuracy. Also, there is no other work in the literature that used the deep learning techniques and breathing sounds, using the original files (raw data) without any sliced cycles (segmentation), pre-processing (resampling, remove artefacts and other nose such heart sounds) and feature extraction techniques when they using the. By doing so, we inherited the following two challenges – 1) the identification process which automatically includes noise and other artefacts in the input data, and – 2) the difficulty of choosing appropriate features.

1.4. Objectives

This research aims to develop an automatic classification system for pulmonary pathology using breathing sounds, advanced signal processing techniques and machine learning techniques. Although the current research studies in the literature, this work focuses on different features of lung sound signals in an effort to extract prominent parameters to detect and identify the important information from the lung sounds, for the classification of pulmonary sounds, as well as, using the recent algorithms in deep learning to computerized the whole classification system with high accuracy and shorter computational time. The objectives of our implementation are formulated as follows.

i. **To extract and test a suitable feature for lung sounds classification**

The performance of the machine learning algorithms as classifiers were required to using the prominent and informative features to distinguishing between lung sounds signals. A statistical analysis of these features should be performed after the features extracted to visualize and observe the difference between the features for each class (Normal bronchial, Wheeze, Crackle, Pleural rub, Stridor), and to select the prominent features. The classification accuracy also must test and validate it by using the classification system on two main classification problem: in multiclass classification case and in binary classification case. These show the ability of machine learning algorithms in any test conditions such as (database, methods of analyses the breath signals, and features used).

ii. **To propose a new learning method for training and testing phases in deep neural network**

In the classification of lung sounds; limitations in the time duration required for training and the accuracy results from testing a deep neural network could be addressed through the changing the hyperparameters such as (batch sizes, learning rate, optimized algorithms and epochs) during the learning phases. The batch size and epoch parameters have diverse values varying from a minimum to a maximum and can, therefore, be adjusted toward optimized learning time and accuracy results. This provides an avenue for performance improvement in the classification processes when the data are inputted without augmentation and when the network is trained from scratch by using deep neural network models. As well, it could be used in any deep learning application as long as the method is incorporated into the learning stages.

iii. To evaluate the proposed learning method in classification of pulmonary pathologies symptoms

In order to diagnose the pulmonary pathologies, it requires an automatic system capable of learning and making decisions. In this sub-work, we evaluate the proposed learning method, based on breathing cycles, for the classification of three types of breathing sounds – normal, crackles and wheeze. These sounds were obtained from the ICBHI scientific dataset consisting of noisy breathing sounds. They were transformed, from the 1D time domain into the 2D time-frequency domain, as an image using the gammatonegram algorithm. This proposed method needs to validated using non-pre-processed breathing sound signals, which contain other sounds such as heart sounds and other artefacts.

iv. To evaluate the proposed learning method in in classification of pulmonary conditions

The application of a new accelerated learning method for deep neural networks should be validated through the classification of other three kinds of pulmonary conditions – healthy, chronic and non-chronic, using breathing cycles based gammatonegrams as input data. The ICBHI dataset of noisy lung sounds served as the benchmark in this sub-work. The VGG16, ResNet-50 and AlexNet architectures will be used in this sub-work to obtaine a classification accuracy with the proposed learning approach.

v. **To evaluate the proposed learning method in in classification of pulmonary diseases**

The development of automatic diagnosis of pulmonary diseases is an important task in this research study. In this work, an application of the new accelerated learning method for deep neural networks will be investigated, for classification of six class of pulmonary diseases – healthy, Bronchiolitis, Bronchiectasis, COPD, Pneumonia and URTI, by using entire recordings of lung sounds based gammatonegrams as input images data, these recordings from the benchmarking ICBHI dataset. Both the VGG16 and AlexNet architectures will investigated in this sub-study using the Multi ep-Batch proposed approach. The robustness in our methodology is exemplified through the results that will obtained without applying any segmentation, preprocessing and without performing feature extraction to the raw data. Moreover, with the results, we can conclude that using any complex CNN architecture with our proposed method appears to gain from more improvements in terms of accuracy and training time.

1.5. Thesis Scope

1) Based on the detail of the literature review and in order to overcome the shortcomings of previous research studies, our protocol was designed to develop a system which can be used for classification of the pulmonary symptoms, diseases and conditions. Data has been used from R.A.L.E and ICBHI respiratory database.
2) Various features (Hjorth descriptor and permutation entropy) and advanced signal processing methods to characterize the breathing sounds are combined and explored in this thesis.

3) A new learning method called Multi ep-Batch aimed to automatically incorporate various values of batch sizes and number of epochs during the training and testing phases was designed and implemented.
4) A time-frequency based input called gammatonegram has been used for different pulmonary state classification using deep learning architectures. For classification four popular CNNs architectures enhanced with the Multi ep-Batch method VGG16, ResNet-50, AlexNet, and GoogLeNet have been used to classify three categories symptomes, conditions and diseases pulmonary state of patients.
5) All experimental of signal processing have been done by MATLAB® (version 2019a), and experimental of deep learning algorithms have been done by python as language and Google Collaboratory as cloud computing system GPU-based.

1.6. Thesis Organization

This dissertation explores the subject of different pulmonary state detection using computerized breathing sounds. Two different machine learning techniques were explored and a novel learning method was proposed. The first machine learning techniques is based on traditional algorithms, when the input sounds signal decomposed by EMD (Empirical Mode Decomposition) into IMF (Intrinsic Mode Function) and the features vector is composed of the two parameters (Activity and Permutation entropy) were extracted from each IMF, after that the extracted features vectors fed into two classifiers namely ELM (Extrem Learning Machine) and K-nn for classifying the lung sounds. The second machine learning focused on recent deep learning algorithms, when the input respiratory sounds are transformed into 2D images (time-frequency representation) namely gammatonegram by using gammatone filter bank, and these images fed into four deep learning models for classifying the different scenarios, with and without the proposed learning method. The research study carried out are presented in four chapters in this dissertation.

Chapter 1 consist of an introduction to the pulmonary pathology and a background on the subject of interest. The motivation of this research study, problem statement, objectives, and finally the scope of this dissertation are presented.

Chapter 2 presents the literature overview on pulmonary pathology and the causes of breathing illness. A review of the traditional and recent machine learning methods and the scope of the study are also explained in this chapter.

Chapter 3 describes the different methodologies implemented to classify the breathing sounds for pulmonary diseases.

Chapter 4 presents the results and discussion obtained from two different parts -machine learning and deep learning algorithms- in lung sound classification using the proposed methods which were discussed in this chapter.

CHAPTER 2

LITERATURE REVIEW

The current chapter reviews the research literature and the method, drawbacks and benefits of existing methods used for lung sounds classifications by using computerized lung sounds. An overview of human respiratory system, breathing system diseases and the causes, analysis of computerized breathing sounds, characteristics and types of breathing sounds, breathing sound data, pre-processing of breathing sounds, feature extraction methods in breathing sound analysis, machine learning algorithms in breathing sound analysis, deep learning algorithms in breathing sound analysis and research gap have been presented in this chapter.

2.1. Human Respiratory System Physiology

Getting rid of carbon dioxide by supplying oxygen to blood cells for the exchange of gasses is the main function of which the breathing system does in the human body. Figure 2.1 shows the two main parts for Respiratory system, the first represent the lower respiratory track and the second is upper respiratory track. The lower respiratory tract involves organs such as Alveoli, Bronchi and Bronchioles [5]. The upper respiratory tract comprises of nasal, nasal cavity, pharynx, glottis, epiglottis, larynx and trachea. The Inhalation and exhalation process of the human breathing system doing by getting in the oxygen and get out the carbon dioxide. One inhalation and exhalation represent one cycle of breathing sound. In the pulmonary alveoli, the exchange of oxygen and carbon dioxide happens. Besides, the second process involves warming, filtering and humidifying the inhaled air [6].

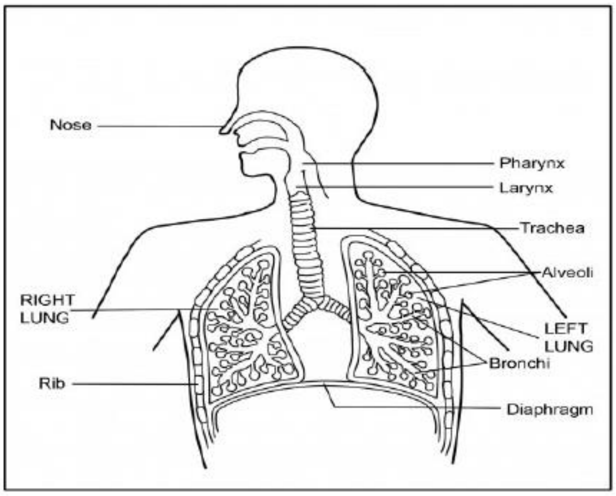


Figure 2.1: Human Respiratory System [7]

2.2. Breathing system diseases and the Causes

Breathing system diseases are mainly classified into two principal parts namely restrictive lung pathology and obstructive lung pathology. Both restrictive and obstructive pulmonary disease also called parenchymal lung disease and airway obstruction disease respectively. A difficulty breathing as a symptom comes from Both of these pulmonary diseases. The stiffness of the pulmonary tissue is the reason for pulmonary restrictive, which lead to difficulty to expand the pulmonary while inhaling air in the lungs. The volume of the pulmonary is reduced in the restrictive diseases [8]. The narrow of the pulmonary airways is the reason for pulmonary obstructive, which lead to difficulty exhale all the air in the pulmonary while inhaling air in the lungs. Idiopathic Pulmonary Fibrosis (IPF), Congestive Heart Failure (CHF), Pulmonary fibrosis and Pneumonia are the commonly pulmonary pathologies related to the restrictive lung disease. Cystic fibrosis, Bronchiectasis, chronic bronchitis, Asthma, emphysema and chronic obstructive pulmonary disease (COPD) are the commonly pulmonary pathologies related to the obstructive lung disease.

2.3. Corona Virus Disease 2019 (COVID-19)

In December 2019, exactly in Wuhan City from China, a severe acute respiratory syndrome coronavirus appears and affect the people by pneumonia disease. Officially, the World Health Organization (WHO) in Feb. 11, 2020 gave a name to this virus and called it COVID-19. The famous clinical manifestations for this virus are as follows: dry cough, fatigue and fever. The infected people with covid-19 are now the major sources of pathologies which is transmitted via direct contact and cough droplets, from the time of outbreak, the scientific community started to detect and identify the cause of this covid-19 and have conducted all the measures. However, recent research studies have been performed for an automatic classification and diagnosis of COVID-19. Hence, this sub-section aims to summarize the latest research which conducted in biomedical filed especially in our thesis context (lung diseases). Various studied using machine and deep learning techniques for COVID-19 diagnosis and classification have been performed. PENG et all [79], provide an empirical view of machine learning used for COVID-19 data and they gave a opportunity for researchers and professionals to understand better the trade-offs involved for build the models in this area. The authors in [80] reviews a deep learning application in image analysis for medical domain, as well as, they giving insights of contributions to COVID- 19 using pulmonary imaging task. In [81] authors classify COVID-19, normal, and pneumonia patients from chest X-ray images, and for the automatic diagnosis of COVID-19 they used Optimized CNN architecture which involved optimized feature extraction and classification components. In [82] an intelligence computer-aided model was proposed that can detect positive COVID-19 cases automatically would aid the clinical applications, the authors used X-ray images data with a deep learning CNN-based. EZZAT et all [83] proposed a GSA-DenseNet121-COVID-19 using an optimization algorithm and the approach based on a hybrid CNN architecture. A DenseNet121 and gravitational search algorithm was used as CNN network and optimization algorithm respectively. The authors in [84] reviews the role of artificial intelligence and machine learning algorithms in the arena of forecasting, contact tracing, screening, predicting and drug development for SARS-CoV-2. In [85] for the detecting the COVID-19 disease the authors proposed Convolutional CapsNet using chest X-ray images data. This novel method was proposed to conduct accurate and fast diagnosis for COVID-19 diseases: - binary classification (COVID-19, and No-Findings) - and multi-class classification (COVID-19, and No-Findings, and Pneumonia).

2.4. Analysis of Computerized breathing sounds

The listening of breathing sounds and diagnosing the pulmonary condition is the main steps included in the breathing sounds analysis. In order to the purpose of patient lung auscultation, René Théophile Hyacinth Laennec a physician who comes from french was invented in 1816 a stethoscope. As we can say, the most common tool which remains used by medical physicians for auscultation purpose the stethoscope. Nowadays, technological development led scientists to perform research studies on computerized breathing sound analysis. In the literature and start from 1980's the computerized breathing sound analysis begin to appear. By taking signal processing algorithms and machine learning algorithms together could analyse the computerized breathing sounds. The general process steps for computerized breathing sound analysis are shown in Figure 2.2.

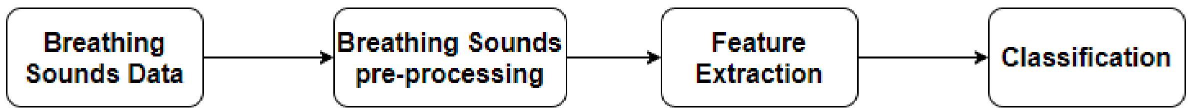


Figure 2.2: General process steps for computerized breathing sound analysis

As we can see in Figure 2.2 the computerized breathing sound analysis involved four main modules namely; breathing sounds data, breathing sound pre-processing, feature extraction and finally pulmonary pathologies classification. The next sections will review briefly the previous studies reported in the literature for each module.

2.4.1. Characteristics and Types of Breathing Sounds

During the breathing phases, the breathing sounds originate in the wide airways. Where, in the airway walls the turbulence and air velocity induced vibrations. Then, through the lung tissue these vibrations will be transmitted through the lung tissue to the surface which can be heard readily using a stethoscope. These breathing sounds contain various characteristics for normal and abnormal respiratory state. The breathing sounds which heard on the chest wall as well as on the trachea are caused by airflow turbulent in the passage of air and lungs during the inhalation and exhalation of the respiratory system. These breathing sounds are non-linear and non-stationary signals nature [74]. Several pulmonary pathologies conditions have a similar symptom it could be heard over the breathing sounds, for which advanced digital signal processing methods should be implemented to distinguish it. Breathing sounds divides into three main groups: normal abnormal and adventitious breathing sounds. The bandwidth of normal breathing sounds related to the area of the sound recording: trachea or chest or wall. The normal breathing sounds has an interval of frequency between 150 to 1000Hz over the chest wall [75], [76]. The normal breathing sounds has an interval of frequency between 150 to 2000Hz over the trachea [75]. The breathing sounds has a dominant frequency between the range of 150 to 2000 Hz, the heart sounds have also a dominant frequency which less than 150 Hz [77]. The obstructive pulmonary pathologies have a dominant frequency rang less than 400 Hz, also the restrictive pulmonary pathologies have a dominant frequency rang sated between 200 and 2000 Hz. The rhonchi and wheeze are the sounds categorized as symptoms of obstructive pulmonary pathological conditions where its duration is greater than 250 ms, whereas the crackles are the sound categorized as a symptom of the restrictive pulmonary pathological conditions its duration less than 100 ms. Based on aforementioned characteristics of the breathing sound, the researchers have the opportunities to explore the possibility of classifying pulmonary pathology using robust advanced digital signal processing techniques and machine learning algorithms.

The following section will discuss on breathing sound data, pre-processing of breathing sounds, feature extraction methods in breathing sound analysis, machine learning algorithms in breathing sound analysis, deep learning algorithms in breathing sound analysis.

2.4.2. Breathing Sound Data

Lung sound recording has a crucial role in breathing sound analysis. In computerized breathing sound analysis, the most common sensors used for lung sounds acquisition are electret microphones, accelerometers, piezoelectric microphones and contact microphones which could be reach a range from 0 and 2000 Hz of frequencies [9].

In breathing sound analysis, there are some electret microphones used such as ECM 77B (Sony), EMT 25C (Siemens) and ECM 44 (Sony). Also, for breathing sounds recording and analysis there are some multichannel devices which is commercially available, such as Welch Allyn, WISE, PCP, Thinklabs and Littmann. Using these sophisticated stethoscopes devices, the cleaning of the noise and other artifacts such heart sounds from the breathing sounds it will be easy after the lung sounds recordings. The computerized respiratory sound analysis (CORSA) is the only standard developed by the European respiratory society [10] for the placement of the sensors correctly. Besides, CORSA offer a method and protocol for collect clinical data properly which involves the location of the sensor to be placed in order to acquire breathing sounds, recording duration and subject's posture.

There are also a few standard databases used from researchers such as R.A.L.E repository [11], ICBHI Scientific Challenge database [12] and Marburg Respiratory Sounds (MARS) [13]. The ICBHI is the only free and available database and the only commercially is the R.A.L.E database. However, R.A.L.E repository has a few numbers of lung sounds recordings. Besides, ICBHI the only large and challenging lung sounds database which is currently available.

2.4.3. Lung Sounds Data Acquisition

Lung sounds data collection consists different stages such as: environmental conditions for lung sounds data collection, as well as, the devices and sensors used for lung sounds data collection, we will review these stages in this current sub-section.

2.4.3.1. Devices and Sensors for lung sound Data Collection

Recently, the automatic sound analysis is the great interest for biomedical audio-based applications. We will discuss the reasons in this subsection. First, it should overcome, at least in part, the limits of human hearing. Indeed, studies have been performed to test the ability of the human ear to detect crackles in an auscultatory signal. The method used consisted of superimposing artificial crackles on a real signal. In general, based on the literature there are many microphones could be used for lung sounds data collection. There are two main methods when these microphones are used: first, kinematic and the second one is acoustic. Regardless of the methods, a condenser or piezoelectric sensor were used for convert the mechanical vibrations into electric signals. Microphones Condenser-based are attached to the skin through air couplers while microphones piezoelectric-based are attached directly to surface of the skin, for lung sounds data collection [86]. Piezoelectric contact-microphones have been used for a few studied [87], [88], [89]. Air-coupled microphones have been used in [90]. The MARS and RALE respiratory sounds databases used for data collected an air-coupled microphone (ECM77) and accelerometer (EMT-25C) respectively.

2.4.3.2. Lung sounds data collection techniques

Recording of lung sounds in a suitable and proper way is an important step which precedes the signal analysis phase. Typically, the collect sound chain includes the following elements [91]:

- **Sound capture:**

Location of the microphone is very important; in fact, the ribcage acts as an attenuator and a low pass filter [92]. Besides, there are a different methods and tools for sounds capturing:

First, the most commonly used method, is using a single microphone. The sensor is generally an electret microphone, the sampling frequency is most often that used for codecs which used in telephony (8kHz), an analog / digital conversion with a resolution of 16bits. Others use an accelerometer, less sensitive to ambient noise, but less efficient than the electret microphone.

Second method is the use of multiple microphones and 3D representation. This technique allows the identification of the areas at the origin of the sounds.

Finally, the last method is the emission of a sound and analysis of its propagation. This method, described in the reference [93], consists in emitting a sound from a loudspeaker introduced into the patient's mouth. It is based on an analysis of the signal characteristics which propagated through the airways and chest cavity.

- **Signal amplification**
- **Filtering and sampling**

Filtering is concerns of remove the artifacts and other noise. The “cleaning” of lung sounds must also take into account for reduction the ambient noise. This processing can be carried out in two different ways: remove the noise by adaptive filtering and remove the noise by wavelet packets (method of Donoho…).

- **Reduction of heart noise**

Heart sounds can introduce disturbances when analysing lung sounds. The spectrum of heart sounds is between 20 and 100 Hz. But a 100 Hz high pass filter may not be a relevant solution because the majority of lung components are also located in this region. Different methods have therefore been tested: wavelets, adaptive filtering with recursive least squares algorithm, time / frequency filtering and reconstruction, AR / MA (autoregressive / moving average) estimation in time / frequency with wavelet coefficients, independent component analysis, entropy methods.

- **Lung sound recording**

Cheetham et al. [94] underlines the important points concerning the sound recordings digitization which are: - the sampling frequency; - the filtering and -the signal to noise ratio introduced by the analog-to-digital conversion.

2.4.4. Pre-processing of Breathing Sounds

Remove and reduction of artifacts such as heart sounds and other noise from the lung sound signals is the key steps in computerized breathing sound analysis which called Pre-processing step. The researchers have been applied the pre-processing step using Butterworth filter to filtering the lung sounds by implementing a low pass filter with the cut off 1600-3000 Hz and a high pass filter with the cut off frequency 30-150 Hz. A breathing sound filtering has provided as a standard guideline in CORSA.

2.4.5. Feature Extraction methods in Breathing Sound Analysis

In the breathing sounds analysis, a different feature extraction method has been used by the researchers in this field. An overview of these methods was discussed within this section. The feature extraction step plays a crucial role in the differentiate and distinguishing between pulmonary diseases from breathing sounds which is the process of characterize the distinctive characteristics from signal. However, from the breathing sound signals can be extracted the features in three domains (the time domain, the frequency domain and the time-frequency domain). The most widely factures extraction techniques used in computed breathing sound analysis are the Mel-frequency cepstral coefficient (MFCC), entropy, autoregressive model, wavelet features based and spectral features. There are two kinds of representation concerning the features extraction methods which can be classified into parametric representation and non-parametric representation.

2.5. Artificial Intelligence

The goal of artificial intelligence (AI) is to design systems capable to reproducing the behavior of humans in their reasoning activities. AI is an import topic that has become omnipresent in economic, social and scientific discussions. Nowadays, the development of this technology in the health sector is a key project around the world. AI has seen spectacular progress in recent years, and his was made possible mainly by two factors: first is the rise of Big Data with the increase in the volume and quality of data collected, as well as the increase in the capacities of storage of these, the second factor is the increase in the computing power of the processors. These developments have allowed “Machine Learning” and “Deep Learning” to emerge from the labs. AI is now a technology with concrete health applications which already in existence. The development of AI in the healthcare sector will lead to reflection on future developments in our healthcare system. This development of AI is generating ambivalent feelings among healthcare professionals. It is both a source of fascination and hope, but also a source of fear among these professionals.

2.6. Machine Learning Algorithms in Breathing Sound Analysis

Over the past decade, many studies have attempted to develop automated detection systems for lung sound classification based on handcrafted feature extraction using machine learning techniques. One of the most popular feature extraction technique used in audio classification involves the integration of Mel-frequency cepstral coefficients (MFCC) into the Gaussian mixture model (GMM) for the recognition of respiratory sounds [14] in which the achieved sensitivity was 0.88% and specificity was 0.99%. A previous study [15] compared extreme learning machine (ELM) and k-nearest neighbor (K-NN) algorithms for lung sound classification which obtained 90.71%, 95.00% accuracy in the binary classification and 83.57%, 86.42% in the multiclass classification using ELM and K-NN respectively. In [16], the researchers used artificial neural network (ANN)-based MFCC features to classify crackles, wheezes, and normal sounds and achieved a sensitivity of 87% and as high as 80% for specificity. Whereas in [17], the frequency characteristics of crackles were extracted using time-frequency and time-scale analysis, and crackle and non-crackle sounds were classified based on the K-NN, support vector machine (SVM), and multilayer perceptron algorithms. In this work, the classification achievement performance with SVM was 81.10%, higher than the 71.55% and 78.50% achieved using MLP and K-NN classifiers respectively. Researchers in [18] had normal and continuous adventitious sounds classified using SVM, based on sample entropy histogram distortion, recursively measured instantaneous kurtosis and auto-regressive averaging features. The obtained mean accuracies were 97.7% and 98.8% for inspiratory and expiratory segments respectively. The authors in [19], classified abnormal lung sounds using a novel attractor recurrent neural network (ARNN) based on fuzzy functions (FFs-ARNN) and subsequently performed recurrent quantification analysis (RQA) to evaluate the efficacy of their system and classification accuracy of 91% was achieved using FFs-ARNN with sequences of RQA features. In [78] A statistical analysis and comparative study were conducted of several parameters such as, Pitch, Harmonic to Nois Ratio (HNR) and Amplitude Perturbation for various cases in respiratory sound signals.

2.7. Deep Learning Algorithms in Breathing Sound Analysis

However, in recent years, many related studies have shifted to using deep learning techniques for a variety of applications such as analysing quantization of space [20], early diagnosis medical support systems [21], and environmental sound classification [22]. Because deep neural network algorithms have addressed many problems in several research applications, they might have a similarly important impact in the analysis of digitized lung sounds. Given this possibility, the authors in [23] proposed a novel method using deep residual networks (ResNets) based on an optimized extracted set of S-transform (OST) features for the classification of wheezes, crackles, and normal sounds. This study used different fixed values of batch sizes and iterations and achieved accuracy, sensitivity and specificity up to 98.79%, 96.27% and 100.00% respectively. Similarly, in [24], a convolutional neural network, based on spectrograms as features, was utilized to develop an algorithm for breathing phase detection in lung sound recordings, and their algorithm achieved an average sensitivity of 97% and an average specificity of 84% by using a fixed epoch and batch size in the learning stages (training and testing process). In [25], the authors implemented a system for lung sound classification and compared three machine learning algorithms – two were based on handcrafted feature extraction and trained using SVM, K-NN and GMM, and the other was based on CNN. In this work, the authors managed to marginally increase the accuracy from 95.10% to 95.56% and they used different fixed batch sizes and iterations numbers. The work in [26], proposed a deep CNN-RNN hybrid model for the classification of four-class respiratory sounds based on Mel spectrogram features of breathing cycles. Their model achieved a score of 66.31% and when the model was retrained with patient-specific data, it achieved a score of 71.81%, both conducted using a leave-one-out validation approach and the standard method (fixed batch size and epoch) in the training and testing process. Researchers of the work in [27] began experimenting on transfer learning, where the VGGish network was combined with the bidirectional gated recurrent unit neural network (VGGish-BiGRU) for lung sound recognition, the hyperparameters such as fixed epoch and batch size are jointly optimized by a large number of experiments. The authors in [28] employed two types of machine learning algorithms for lung sound classification – one was focused on the use of SVM based on MFCC features, and the second was based on a CNN with spectrogram images as features and fixed batch size and epoch (standard method) in learning process were used, the highest and lowest accuracy results of their experiments were 86% and 62% for both CNN and SVM respectively.

To carry out correctly, the novelty and validity of our results a justification of the latest and earlier important publications which have been conducted in our subject must be mentioned in this manuscript. Demir et al [29] propose a new pretrained (CNN) model based on spectrogram for lung sounds classification, and their algorithm achieved an accuracy of 71.15% with a max-epochs of 12. Jayalakshmy et al [30], propose a pre-trained optimized Alexnet architecture for predicting respiratory disorders using EMD method and scalogram as Time Frequency Representation (TFR), in this study an improved accuracy of 83.78 % is achieved using different IMFs of EMD, the authors used different fixed epochs 2,4,8,16, 20 and fixed batch size of 10. Garcia et al [31] used CNN to classify the respiratory sounds into healthy, chronic, and non-chronic pulmonary condition by using Mel spectrogram, and their algorithm achieved up to 0.993 F-Score, the authors used the standard method (fixed epoch and batch size). Rocha et al [32] the authors studied the influence of event duration on automatic adventitious respiratory sounds classification, and classifier namely linear discriminant analysis, support vector machines, boosted trees, and convolutional neural networks were used and their experiments achieved a maximum accuracy of 96.9%, in this study to train all the deep learning models, a total of 30 epochs and a batch size of 16 were used. Shuvo et al [33] proposed a CNN architecture to classify respiratory diseases (ternary chronic and six-class) using the same ICBHI 2017 lung sound dataset. They used empirical mode decomposition (EMD), and continuous wavelet transforms (CWT). This study achieved an accuracy of 99.20% for ternary chronic classification and 99.05% for the six-class pathological classification. In both classification schemes, a fixed batch size of 6 has been taken for training and validation. Demir et al [34] used a pre-trained deep convolutional neural networks model for extract features and a support vector machine classifier used in classification. A second approach used, in which a pre-trained deep CNN model was fine-tuned with spectrogram images for the classification of the lung sounds, and accuracy of 65.5% and 63.09% respectively, was achieved in this study, the authors used the standard method (fixed epoch and batch size). A summary of the literature is presented in Table 2.1.

Work	Dataset	Number of classes (C)	Classification Method	Training and Testing process (Epoch & Batch size)	Outcome
Demir *et al* [29]	ICBHI [12]	C = 4	Spectrogram + Linear Discriminant Analysis (LDA)	**Fixed values** **(Epoch & Batch size)**	The classification accuracy using the deep feature with CNN & LDA classifier with RSE method was 71.15 %.
Jayalakshmy *et al* [30]	R.A.L.E [11]	C = 4	Scalogram + AlexNet (pretrained model)	**Fixed values**	An accuracy of 83.78 % is achieved by the virtue of scalogram representation of various IMFs of EMD.
Garcia *et al* [31]	ICBHI	C = 6	Melspectrogram + Convolutional Neural Network (CNN)	**Fixed values**	They achieved up to 0.993 F-Score in the three-label classification and 0.990 F-Score in the six-class classification.

Chen *et al* [23]	ICBHI	C = 3	OPTIMIZED S-TRANSFORM (OST) + DEEP RESIDUAL NETWORKS (RESNET-50)	**Fixed values**	The classification accuracy obtained using Optimized S-transform with ResNet-50 up to 98.79%.
Bardou *et al* [25]	R.A.L.E	C = 7	Spectrogram + SVM, k-NN, GMM and CNN	**Fixed values (Epoch & Batch size)**	The classification accuracy was successfully increased from 95.10% to 95.56% using ensembling through summing up the output of Softmax activation of four CNN models.
Acharya *et al* [26]	ICBHI	C = 4	Melspectrogram + Hybrid CNN-RNN model	**Fixed values**	A score of 66.31 % on the four-class classification of breathing was achieved. Also a score of 71.81 % for leave-one-out validation.
Shi *et al* [27]	Self-collected lung sound data	C = 3	Melspectrogram + VGG & RNN	**Fixed values**	87.41% total accuracy was reported. Asthma, pneumonia and normal were 83.33%, 86.75% and 91.94% respectively

					3) singular respiratory sound type 80% and 80% respectively. 4) audio type classification with all sound 62% and 62% respectively
Rocha *et al* [32]	ICBHI	C = 5	Spectrogram, Mel spectrogram, and scalogram + Linear	**Fixed values**	The best accuracy of 96.9%3 was archived in the Class task with fixed durations and an accuracy of

			discriminant analysis, SVM, boosted trees, and CNN		81.8% on the more realistic 3 Class task with variable durations.
Shuvo *et al* [33]	ICBHI	C = 6	Scalogram + CNN	**Fixed values**	Accuracy scores of 98.92% for three-class chronic classification and 98.70% for six-class pathological classification were achieved
Demir *et al* [34]	ICBHI	C = 4	Spectrogram + CNN and SVM	**Fixed values**	The accuracies for the first and the second proposed methods were 65.5% and 63.09%, respectively.

Table 2.1: Time-frequency representation (TFR) in computer-based lung sound classification systems.

2.8. Research Gap

Based on the aforementioned studies in the literature review, there are a several questions concerning to the pulmonary pathologies classification which raises our attention. All the research studies reported in the literature have been used fixed learning system deep learning based in terms of hyperparameters (Batch size, epochs, difference in the learning rate used and difference in the optimized algorithms for each work). In the feature's extraction concerning traditional machine learning classification, the researchers have implemented the popular and common parameters of respiratory sounds. They don't try to explore another feature. The non-stationary signal processing techniques are considered to characterize the breathing signals and hence the not all sophisticated signal processing techniques have been implemented by the previous researcher.

In the deep learning approaches, researchers have focused on particular non-configurable deep networks and not developed a new learning method which can accelerate the training and testing phases. All research studies on the literature have worked on the challenging dataset such as (ICBHI), before classification, using deep learning they did: slice audio recording, resample, pre-processing and features extraction. There is no work have used the direct with entire Recordings from the dataset without any steps are mentioned. The previous researcher has been reported the accuracy results in lung sounds classification using deep learning, was also found to be low, that means the automatic classification of berthing pathology will needed to more efforts. The authors in the literature performed three steps of pre-processing to the ICBHI sound segments – frame composition, feature extraction, and feature normalization. These works when taken together indicate that while pre-processing and feature extraction aids the classification process, it nevertheless negates the original purpose of deep learning. Besides, it is important to subject the findings of past and present studies to critical scrutiny in terms of the benchmarked input physiological data. In the pulmonary diseases classification phase, only a few researcher studies have used all the pulmonary disease categories.

CHAPTER 3

METHODOLOGY

This chapter describes the different methodologies implemented to classify the breathing sounds for pulmonary diseases. An efficient classification system of lung sounds for detect various pulmonary pathology requires a reliable input features and classifier. The dataset used in this research study are described in this chapter. The methodology implemented for the slicing of breathing sounds are also discussed. A different feature was presented. Machine learning and deep learning algorithms used in this work are then presented into two different main part of experiments.

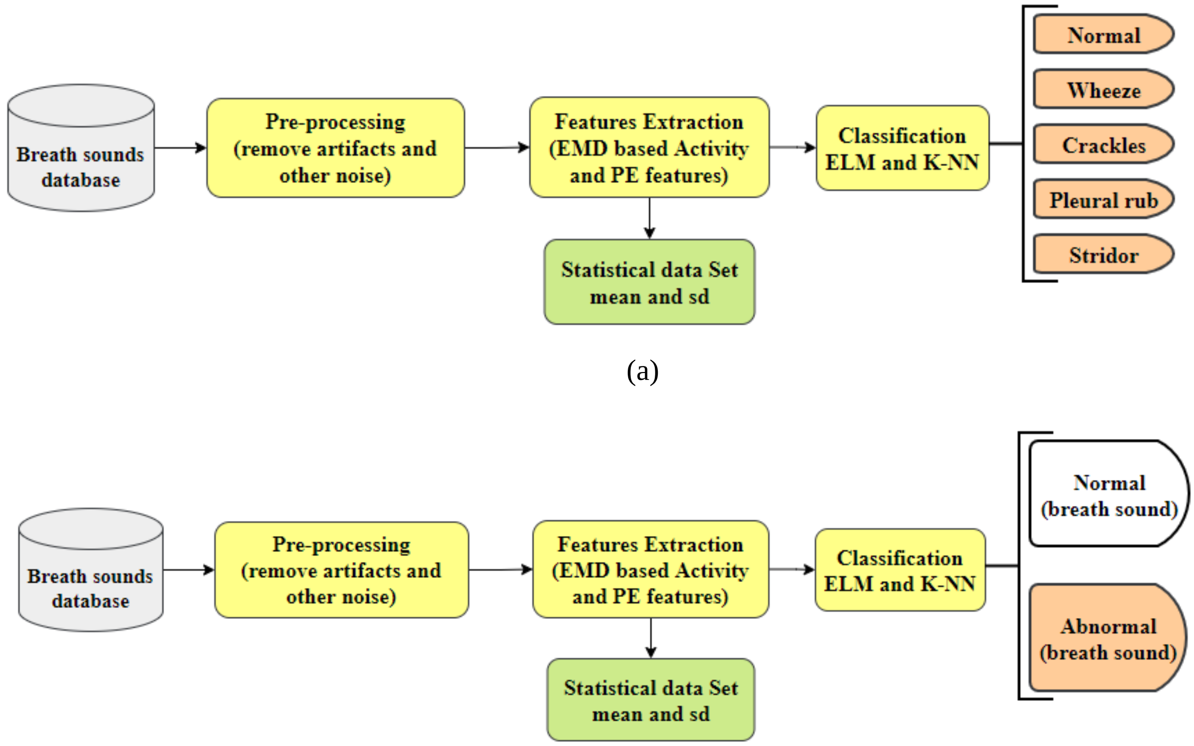
3.1. Methodology

To develop our classification system, two principal parts (methodologies) were implemented to exploit the machine learning and deep learning algorithms for pulmonary diseases detection using breathing sounds. The first part focused on machine learning and second part on deep learning. The lung sounds used in this work are given from ICBHI challenging database and R.A.L.E Lung Sounds database.

3.1.1. Part 1: Machine Learning Algorithms in Breathing Sounds Classification

In this part, we compare the effectiveness of the extreme learning machine (ELM) and k-nearest neighbour (K-nn) machine learning algorithms in the adventitious and normal breath sounds classification. To do so, the breath sounds signal was decomposed by using empirical mode decomposition (EMD) technique, which is rarely used in the breath sounds analysis. After getting the Intrinsic Mode Functions (IMFs) using EMD on the signals, the Permutation Entropy (PE) and Hjorth descriptors (Activity) features were extracted from each IMFs and combined for classification stage.

This part was divided into two main stages namely (Multiclass classification stage, Binary classification stage). The four steps proposed for both stage study namely (database, pre-processing, feature extraction and classification) are presented in Figure 3.1.



(a)

(b)

Figure 3.1: Two principal stage of the breath sounds signal classification (a) multiclass classification (b) binary classification.

3.1.1.1 Database

In this part the database of breath sounds signals used for analysis are the R.A.L.E (Respiration Acoustic Laboratory Environment) Lung Sounds, is the only commercially available database, is an educational program to help doctors and researchers in respiratory signals processing area, offer more than 50 breath sounds were recorded using a contact accelerometer (Siemens-EMT25C) covering normal and abnormal respiratory sounds [35] are sampled at 10240Hz. As this database (R.A.L.E) has a few data, therefore to ensure the credibility of this comparative study we used another data were collected from the internet:

- The Auscultation Assistant, 2015 [36]
- Arnall, 2015 [37]
- The CD of the book [38]

In all a 75 breath sounds divided into five classes (Normal bronchial, Wheeze, Crackle, Pleural rub, Stridor) were used in our study, each sound is an effect of particular disease such as Wheeze indicate that the patient suffering from asthma and COPD (Chronic Obstructive Pulmonary diseases), crackle indicate pneumonia or lung cancer.

3.1.1.2 Breath Sounds Pre-processing

Breath sounds signals are subject to several artefacts such as heart sounds and noise which simulate real-life conditions. The breath sounds signals (R.A.L.E) that have been filtered by a high-pass filter with 7.5 Hz by 1st order Butterworth to remove DC offset, and a low-pass filter at 2.5 Hz by 8th order Butterworth, and we apply a mean and amplitude normalization to reduce the effect of heart sounds. Finally, all samples are down sampled to 8000 Hz sampling frequency according to CORSA (computerized respiratory sound analysis) [39], in this study, the 16-bit resolution and one respiratory cycle are used.

3.1.1.3 Empirical mode decomposition

The Empirical Mode Decomposition (EMD) method is a new adaptive signal time-frequency processing method proposed by NE Huang in 1998 by NASA and others [40]. It is especially suitable for nonlinearity, analysis and processing of non-stationary signals. The Hilbert transform transforms the well-known Hilbert-Huang Transform (HHT).

EMD is actually a method of decomposing signals. It is consistent with the core idea of Fourier transform and wavelet transform. Everyone wants to decompose the signal into a superposition of independent components, only the Fourier transform and the wavelet transform it is required to have a basic function, but EMD completely abandons the constraint of the basis function, and only performs signal decomposition based on the time scale feature of the data itself, and has adaptability. Since no basis function is required, EMD can be used for almost any type of signal decomposition, especially for the decomposition of nonlinear, non-stationary signals [41]. The purpose of EMD is to decompose the signal into a superposition of multiple intrinsic mode functions (IMFs). In addition, the IMF must satisfy the following two conditions (the function must have the same number of local extreme points and zero crossings within the entire time range, and at any point in time, the envelope of the local maximum the envelope of the (upper envelope) and the local minimum (lower envelope) must be zero on average).

The EMD method is based on:

The signal has at least two extreme points, one maximum and one minimum.

The characteristic time scale is defined by the time between the two extreme points.

If the data lacks extreme points but has deformation points, the extreme points can be obtained by data differentiation once or several times, and then the decomposition results are obtained by integration.

The algorithm flow is as follows:

1) Identify all extrema of x(t)

2) Interpolate between minima (resp. maxima), ending up with some envelope emin(t) (resp. emax(t))

3) Compute the mean m(t) = (emin(t)+emax(t))/2

4) Extract the detail d(t) = x(t) − m(t)

5) Iterate on the residual m(t)

3.1.1.4 Features Extraction

A helpful feature for express a biomedical signal namely Hjorth descriptors (HD) divided into three main parameters as follows:

Activity: is the most useful parameters in biological signals, simply its variance of the signal represents the energy:

$$\mathrm{Actv} = \sigma_0{}^2 \tag{1}$$

Mobility: Mobility is given by:

$$\mathrm{Mob} = \sigma_1{}^2 / \sigma_0{}^2 \tag{2}$$

Complexity: gives a computational value for the shape of the signal:

$$Comp = \sqrt[2]{\left(\frac{\sigma^2{}_{m+1}}{\sigma^2{}_m} - \frac{\sigma^2{}_m}{\sigma^2{}_{m-1}}\right)} \tag{3}$$

Permutation Entropy: Bandt and Pompe are investigated the (PE) Permutation entropy to measure the complexity of the non-linearity and non-stationary nature in time series signals [42]. the Shannon entropy is calculated in PE for the different symbol in the signal and can be calculated as follows:

$$\mathrm{PEn} = \left(\sum_{i=1}^{m} p_i * \log(p_i)\right)/\ln(m) \tag{4}$$

3.1.1.5 Statistical analysis

In this study, a statistical analysis of mean and standard deviation was used to test the significance of the activity and PE features. SD and Mean are expressed respectively as follows:

$$\sigma = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(x_j - \mu\right)^2} \tag{5}$$

$$\bar{x} = (\sum x_j)/M \quad (6)$$

Where:

x_j : each value of the dataset.

M : the total number of data points.

3.1.1.6 Classification

In this study, two classifiers were used for two classification types (multiclass classification, binary classification), one is the extreme learning machine (ELM) and the other is a k-nearest neighbour (K-NN). detailed of these classifiers are presented in the next section

3.1.1.7 Extreme Learning Machine

Huang et al. [43] propose an algorithm for solving a single hidden layer neural network which is an extreme learning machine (ELM).

The biggest feature of ELM is that traditional neural networks, especially concerning a single hidden layer feedforward neural networks (SLFNs), are faster than traditional learning algorithms while guaranteeing learning accuracy [44].

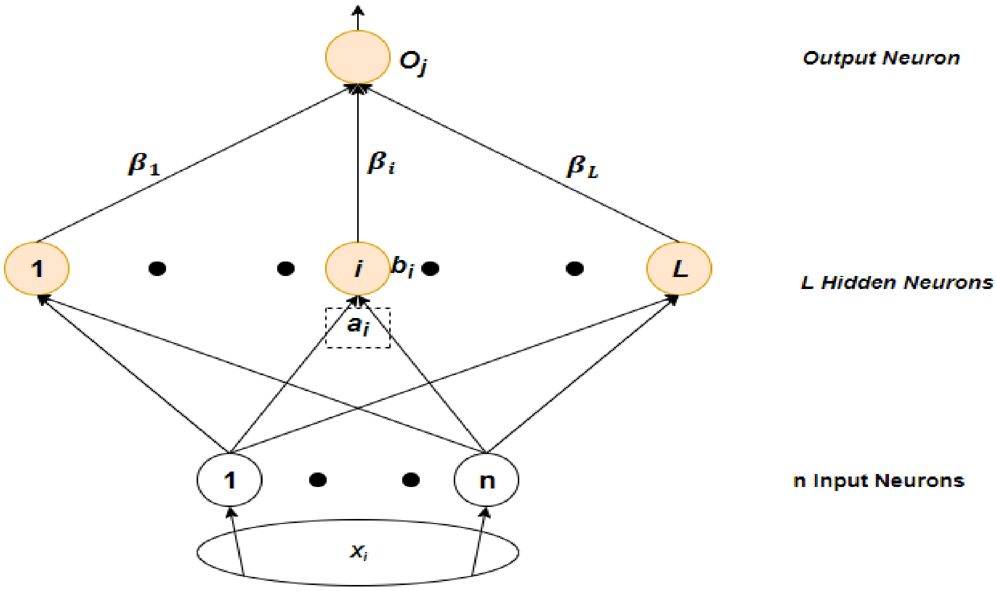


Figure 3.2: SLFN: additive hidden nodes.

For a single hidden layer neural network shown in Figure 3.2, assume that there is N an arbitrary sample (X_i, t_i) of which [45]:

$$X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n \text{ , } t_i = [t_{i1}, t_{i2}, \dots, t_{in}]^T \in R^m \quad (7)$$

For a L single hidden layer neural network with a hidden layer node, it can be expressed as

$$\sum_{i=1}^{L} \beta_i g(W_i . X_j + b_i) = o_j, j = 1, \dots, N \quad (8)$$

Among them $g(x)$, the activation function, which:

$W_i = [w_{i,1}, w_{i,2}, \dots, w_{i,n}]^T$ is the input weight and β_i the output weight, b_i is the offset of the first hidden layer unit. $W_i . X_j$ Representation W_i and X_j inner product.

The goal of a single hidden layer neural network learning is to minimize the error of the output, which can be expressed as:

$$\sum_{j=1}^{N} \left\| o_j - t_j \right\| = 0 \tag{9}$$

That exists β_i , W_i and b_i so that

$$\sum_{i=1}^{L} \beta_i g(W_i \,. X_j + b_i) = t_j, j = 1, \dots, N \tag{10}$$

Can be expressed as a matrix

$$H\beta = T \tag{11}$$

Among them it H is the output of the hidden layer node, which β is the output weight and T is the expected output.

$$H = (W_1, \dots, W_L, b_1, \dots, b_l, X_1, \dots, X_L) \tag{12}$$

$$= \begin{bmatrix} g(W_1.X_1 + b_1) & \dots & g(W_L\,.X_1 + b_L) \\ \vdots & \dots & \vdots \\ g(W_1.X_N + b_1) & \dots & g(W_L\,.X_N + b_L) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m,} \qquad T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N \times m,} \tag{13}$$

In order to be able to train a single hidden layer neural network, we hope to get W_1 , b_i and β_i to make

$$\left\| H\left(\widehat{W}_i, \hat{b}_i\right)\hat{\beta}_i - T \right\| = \min_{w,b,\beta} \| H\,(W_i, b_i)\beta_i - T \| \tag{14}$$

Where $i = \mathbf{1}, \dots, L$ this is equivalent to minimizing the loss function

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i g(W_i . X_j + b_i) - t_j \right)^2 \tag{15}$$

Some traditional algorithms based on the gradient descent method can be used to solve such problems, but the basic gradient-based learning algorithm needs to adjust all parameters during the iterative process. In the ELM algorithm, once the input weight W_i and the bias of the hidden layer b_i are randomly determined, the output matrix of the hidden layer H is uniquely determined. The training single hidden layer neural network can be transformed into a linear system $H\beta = T$ and the output weight β can be determined

$$\hat{\beta} = H^{\dagger} T \tag{16}$$

Among them $H^{\dagger}$is H the Moore-Penrose generalized inverse of the matrix. And it can be proved that $\hat{\beta}$ the norm of the solution obtained is minimal and unique [45].

3.1.1.8 k-nearest neighbour

The K-Nearest Neighbors (K-NN) algorithm is a classification algorithm and one of the easiest to understand machine learning algorithms. In 1968, it was proposed by Cover and Hart [46].

The simplest and mundane classifier may be the kind of memorable classifier, remember all the training data, for the new data, it matches the training data directly, if there is training data of the same attribute, use it directly, come as a classification of new data.

The k-NN algorithm finds the k records closest to the new data from the training set and then determines the category of the new data based on their primary classification.

The algorithm involves three main factors:

(1) The training set.

(2) The distance or similar measure. In this study a Euclidian distance has been used:

$$d(x,y) = \sum_{i=1}^{N} \sqrt{x_i^2 - y_i^2} \qquad (17)$$

(3) The size of k.

In the validation stage, the dataset X is divided into a training set Y (training set) and a test set Z (test set), for the case that the sample size is insufficient such as in our study, and in order to full use of all data set to test the algorithms effect, database X is randomly divided into k packets, one of which is used as a test set each time, and the remaining k-1 packets are trained as a training set, by using k-fold cross-validation method [47].

3.1.2. Part 2: Deep Learning Algorithms in Breathing Sounds Classification

Breathing sounds contain prominent information that can aid doctors diagnose pulmonary pathologies in a non-invasive way. Based on these sounds, we can establish many algorithms to develop an automatic classification system that could be used to categorize lung diseases. The traditional methods used in the first part for this chapter when implementing these systems involve two main steps – feature extraction and pattern classification. In recent years, the topic of interest in the field of breathing sound classification focuses on the use of deep neural networks, which have been proven to be effective for training large datasets. Deep learning has been widely used in a variety of applications, such as in biomedical engineering [48]. The important part that distinguishes deep learning from traditional machine learning algorithms is that from the data, there is no handcrafted feature extraction part since the deep learning model learns a prominent feature over the data in the training phase. Most patients who suffer from the chronic obstructive pulmonary disease (COPD) or asthma can be diagnosed by detecting associated symptomatic sounds such as wheezes. Whereas, patients with lung fibrosis, pneumonia and obstructive lung diseases, including COPD and chronic bronchitis [49], experience other associated symptomatic sounds such as crackles. Given this, a proper database of digitally acquired continuous and discontinuous adventitious breathing sounds, coupled with a computerized decision support system, could help us distinguish several pathologies linked to the pulmonary organ, and it is expected that the accuracy of their classification into several types of sounds should ideally be high. However, the pre-processing step, which aims to clean and prepare the data, and the feature extraction steps require robust algorithms to deal with noise and other artefacts. To address these limitations, recent algorithms appear to be established based on those developed in the field of machine learning. Large amounts of data require the use of powerful techniques, which include deep learning algorithms. In the field of biomedical engineering, with the technological advancement related to machine learning and deep learning algorithms, many neural network architectures have been proposed for the development of automatic systems for lung sound classification. The convolutional neural network (CNN) is a promising architecture for addressing this task, the most common deep learning architecture, and usually contains three important layers, namely a convolutional layer, a pooling layer, and a fully connected layer. Several network architectures have been developed based on CNN layers, and these include Visual Geometry Group 16 (VGG16) [50], deep residual networks (ResNet) [51] and GoogLeNet [52]. Initially, CNN was designed for 2D image classification. Hence in the biomedical field, where researchers very often acquire time-varying physiological signals, to adapt the CNN for 1D signal classification, many researchers have used these 1D signals and applied the short-time Fourier transform (STFT) in the time domain to transform these signals into spectrograms outputs, which is a 2D representation in the time-frequency domain. A CNN with many layers, such as deep neural networks, require a long training time. To speed up this process, more computational power (e.g., more CPU/GPU) is needed, also can be a complex task because of some challenges, such as (i) a long learning (training and testing phases) time, (ii) the deep learning models often have limited accuracy due to the fixed values chosen of hyperparameters such as (epoch and batch size), (iii) the high computational burden due to the amount of input data. Over years since starting applications based on deep learning, researchers have never developed such kind of method to explore the potential of optimized learning method in different domains.

Contribution of this part

We propose a novel Multi ep-Batch method that applies the use of multiple epochs and batch sizes during training and testing phases for deep neural networks, to speed up the training process and improve the accuracy in terms of both efficiency and effectiveness, for the lung sounds classification. We evaluate our proposed method in different scenarios using three different sub-data taken from the ICBHI database:

scenario (i) symptoms-based, *scenario (ii)* conditions-based and *scenario (iii)* diseases-based. In the first *scenario (i)* the normal, crackle and wheeze lung sounds were classified using VGG16, ResNet-50 and GoogLeNet architectures with and without Multi ep-Batch method, in the second *scenario (ii)* a VGG16, ResNet-50 and AlexNet [53] were used for the classification of healthy, chronic (COPD, Bronchiectasis and Asthma) and nonchronic (Pneumonia, URTI, Bronchiolitis and LRTI) pulmonary conditions with and without Multi ep-Batch method, in the last *scenario (iii)*, a six-class of pulmonary diseases namely (healthy, COPD, Bronchiectasis, Pneumonia, URTI, and Bronchiolitis) were classified using VGG16 and AlexNet architectures with and without Multi ep-Batch method. By This Multi ep-Batch method provides an avenue for performance improvement in the classification processes when the data are inputted without augmentation and when the network is trained from scratch by using deep neural network models. As well, it could be used in any deep learning application as long as the method is incorporated into the training and testing stages. To the best of our knowledge, a Multi ep-Batch method has not been previously reported and examined in the literature for deep neural networks. The objectives of our study are to exploiting the training and testing process based on deep learning for lung sound classification, were the following: (1) to assess the VGG16, ResNet-50, AlexNet and GoogLeNet models for classification with and without the Multi ep-Batch method, (2) to evaluate the performance of our approach in terms of training time and accuracy rate and (3) to discuss and elucidate the aspect of learning system which are a necessary stage in deep learning. To do so, we systematically carried out a set of experiments using different scenarios mentioned using four existing CNNs architectures in three sub-data organized from the ICBHI database.

This section aims at explaining the three different scenarios. As introduced, a significant amount of computational time and data are required when a deep network trained from scratch. The behaviour of the training and testing process plays an essential role in many problems of deep learning applications, therefore the learning system in any deep neural network is a challenging task. Hence, it is common to use a fixed learning method which involves the use of fixed values of epoch and batch size, although all research study in the literature used fixed values for the training and testing process, still, need new exploitation of these values to get an effective learning process. We describe our new learning system, in the next three different subsections to prove the results concerning the classification with and without our Multi ep-Batch proposed method. Section 3.1.2.1 describes the scenario (i) symptoms-based classification. Section 3.1.2.2 presents the scenario (ii) conditions-based classification. Finally, Section 3.1.2.3 explains scenario (iii) diseases-based classification.

3.1.2.1. Scenario (i) - symptoms-based

In this implementation (symptoms-based), we used three CNN-based architectures to study our Multi ep-Batch method, for the classification of three lung sounds, namely normal sounds, crackles, and wheezes. A block diagram of the proposed Multi ep-Batch method, VGG16-, ResNet-50 and GoogLeNet-based classification is illustrated in Figure 3.3. First, we segmented all audio recordings used in this scenario into cycles. We then processed these cycles consisting of the lung sounds using a Gammatone filter bank transformation and converted the output into a 2D representation to obtain images. We subsequently fed these images rescaled to a size of 224 × 224 pixels in RGB format into the VGG16, ResNet-50 and GoogLeNet deep neural networks. Second, we classified the normal, crackle and wheeze breathing sounds with and without multi ep-batch method.

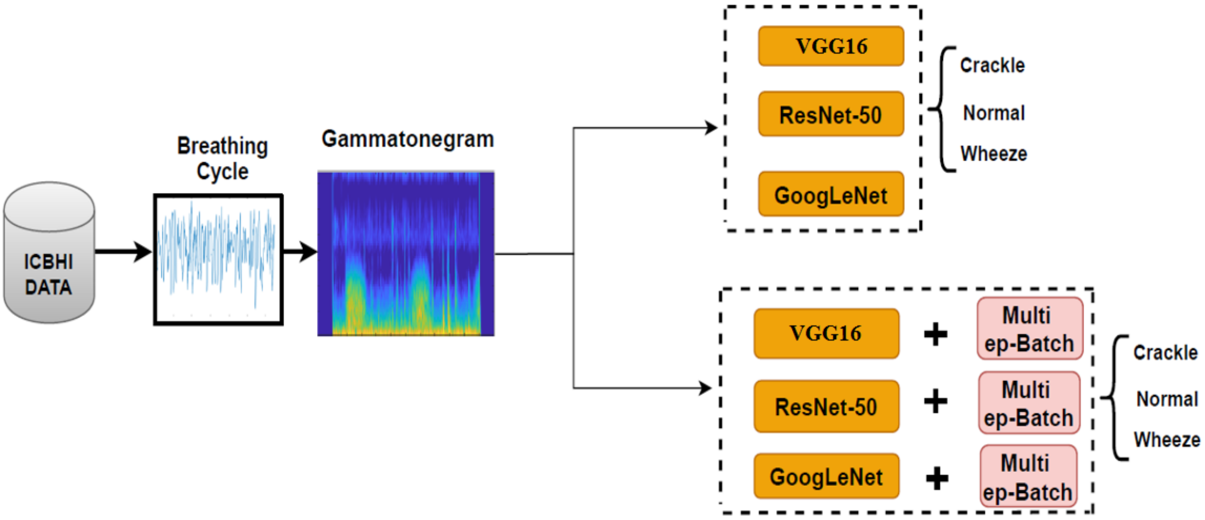


Figure 3.3: Block diagram of the proposed Multi ep-Batch, VGG16, ResNet-50 and GoogLeNet based symptoms-classification.

3.1.2.2. Scenario (ii) - conditions-based

In this scenario (conditions-based), we implemented a Visual Geometry Group 16 (VGG16), residual networks (ResNets) and AlexNet architectures. We apply these models with and without the Multi ep-Batch (multi epochs and batch size) learning mexthod, for classifying three types of pulmonary conditions – healthy, chronic (COPD, Bronchiectasis and Asthma) and non-chronic (Pneumonia, LRTI, URTI, and Bronchiolitis). We divided all the audio recordings into cycles and processed these cycles of each pulmonary condition using a gammatone filter bank, and the output was then converted into a two-dimensional representation as images. In the first stage, we applied VGG16, ResNet-50 and AlexNet without Multi ep-Batch for the classification of pulmonary conditions into healthy, chronic and non-chronic, and in the second stage, we performed the same analysis with Multi ep-Batch to see the effect of our proposed method. The flowchart of the proposed Multi ep-Batch, VGG16, ResNet-50 and AlexNet based classification is shown in Figure 3.4.

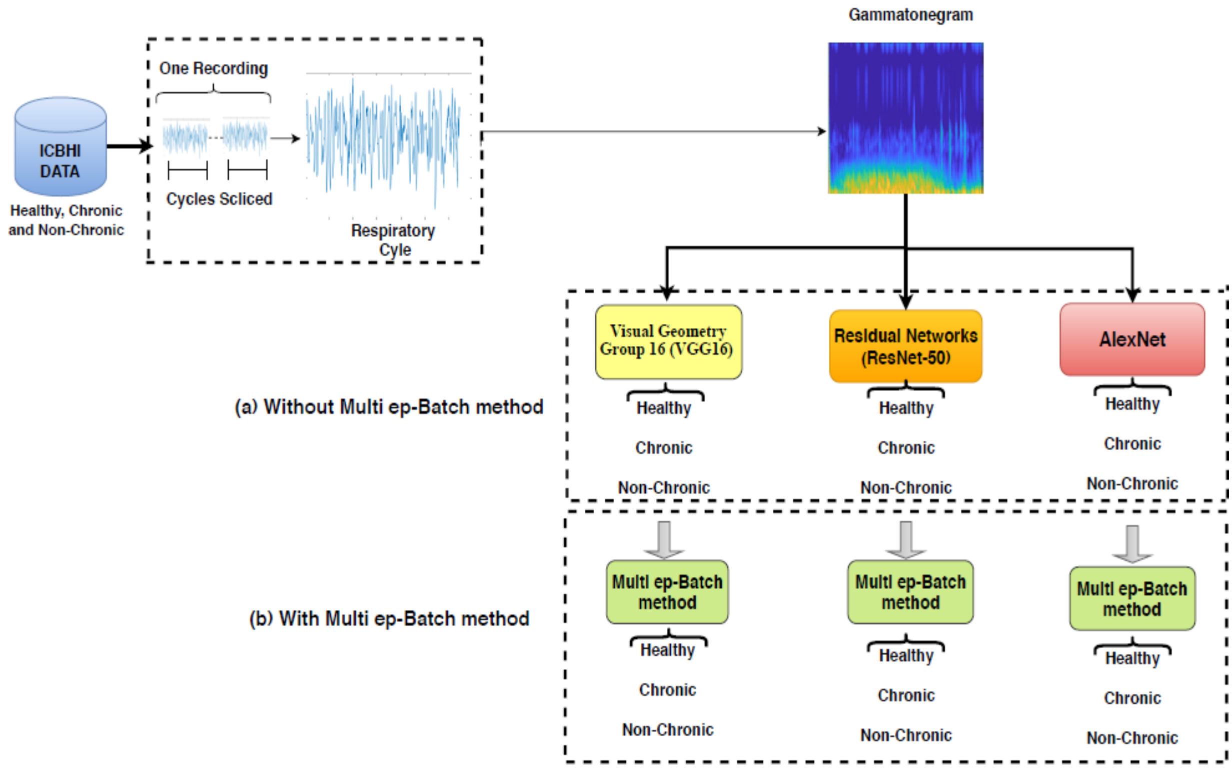


Figure 3.4: Flowchart of the proposed VGG16, ResNet-50 and AlexNet based conditions-classification.

3.1.2.3. Scenario (iii) – diseases-based

In this scenario (diseases-based), we used the Visual Geometry Group 16 (VGG16) and AlexNet models with and without the Multi ep-Batch method, for the classification of pulmonary diseases into six-class namely (healthy, COPD, Bronchiectasis, Pneumonia, URTI, and Bronchiolitis). Using a gammatone filter bank, we processed directly all the entire audio recordings samples of each pulmonary disease from the ICBHI database, then the output was fed into VGG16 and AlexNet without Multi ep-Batch method. Secondly, we performed the same analysis with the Multi ep-Batch method. The flowchart of the proposed Multi ep-Batch, VGG16 and AlexNet based classification are shown in Figure 3.5.

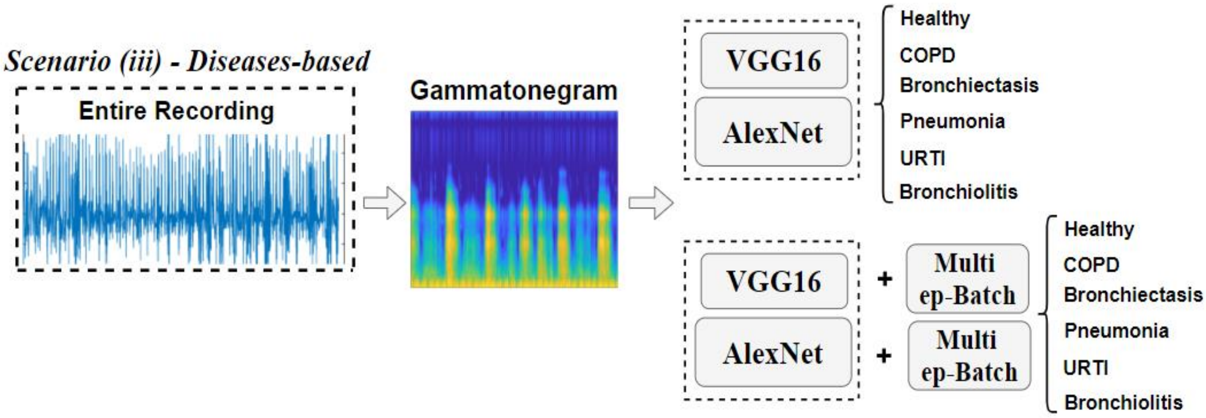


Figure 3.5: Flowchart of the proposed VGG16 and AlexNet, (a) without the Multi ep-Batch method and (b) with the Multi ep-Batch method.

3.1.2.4. Gammatone Filter Bank

The Gammatone filter bank was initially proposed by Roy Patterson and colleagues [54] and was designed to model the human auditory system. A gammatone analysis can be easily used to generate a time-frequency representation, which can be used as an image. For respiratory sounds, most researchers [25, 28] have used STFT. In this study, a gammatonegram was used to represent the input audio cycles as a 2D representation. To the best of our knowledge, breath sound signals have a non-stationary nature, which complicates their classification into some classes. Hence, the large variations in the non-stationary signal can be captured using a filter with a higher time resolution. Therefore, in this study, we exploited the bandpass filters of Gammatones and found that increases in the central frequency were associated with increases in the filter bandwidth.

A Gammatone filter has an impulse response produced by gamma (statistical distribution) and sinusoidal carrier tones, which can be described as follows:

$$f_i(t) = \begin{cases} at^{n-1}e^{-2\pi b_i t}\cos(2\pi C_f t + \emptyset) & ,t \geq 0 \\ 0 & ,t < 0 \end{cases} \tag{18}$$

where,

C_f is the central frequency of the Gammatone filter;

$\emptyset$ is the phase of the filter;

a is a constant for controlling the gain;

n is the order of the filter; and

b_i is used to determine the bandwidth of the bandpass Gammatone filter.

In [55], Glasberg and Moore developed the following equation for the computation of equivalent rectangular bandwidth (ERB):

$$ERB(f) = 24.7(4.37C_f + 1) \tag{19}$$

where,

$ERB(f)$ is in units of Hz; and

C_f is in units of kHz.

This formula was originally introduced by Greenwood [56] to explain the relationship between the central frequency and the variation in the critical bandwidth. These equations became digitally applicable with the availability of an efficient implementation of the Patterson-Holdsworth Auditory Filter Bank by Malcolm Slaney [57], which inevitably led to the popularity of this approach. Subsequently, the code available in [58] includes the creation of a gammatonegram in the main routine, which takes a waveform and other parameters and returns a spectrogram-like time-frequency matrix.

3.1.2.5. Deep learning models

VGG16

The standard structure of the CNN is usually composed of three layers, namely, a convolutional layer, a pooling layer and a fully connected layer. In recent years, many researchers focused on the design of these layers to obtain different architectures for classifying breathing sounds into several classes based on converting audio signals from the time domain into 2D representations (images) in the time-frequency domain, denoted as spectrograms. Similarly, our work is based on the CNN model which utilizes the VGG16 network that was first proposed by [50]. Figure 3.6 shows the detailed architecture of the VGG16 network. Our system, based on the VGG16 network, is divided into three principal stages. During the first stage, we converted all breathing sound segments into gammatonegram feature map images. Subsequently, all the images were resized with a fixed size of 224 × 224 × 3, which is the required input image size to the VGG16 network. The second stage is described as follows:

Block 1 [Green] : (2 × convolutional layer [Con2D] + Rectified Linear Unit Activation Function [ReLU] + 64 × filters with a kernel size of 3 × 3) + max pooling layer with a size of 2 × 2 and a stride size of 2 × 2.

Block 2 [Blue] : (2 × convolutional layer [Con2D] + Rectified Linear Unit Activation Function [ReLU] + 128 × filters with a kernel size of 3 × 3) + max pooling layer with a size of 2 × 2 and a stride size of 2 × 2.

Block 3 [Gray] : (3 × convolutional layer [Con2D] + Rectified Linear Unit Activation Function [ReLU] + 256 × filters with a kernel size of 3 × 3) + max pooling layer with a size of 2 × 2 and a stride size of 2 × 2.

Block 4 [Pink] : (3 × convolutional layer [Con2D] + Rectified Linear Unit Activation Function [ReLU] + 512 × filters with a kernel size of 3 × 3) + max pooling layer with a size of 2 × 2 and a stride size of 2 × 2.

Block 5[Red] : (3 × convolutional layer [Con2D] + Rectified Linear Unit Activation Function [ReLU] + 512 × filters with a kernel size of 3 × 3) + max pooling layer with a size of 2 × 2 and a stride size of 2 × 2.

The last stage focuses on classification, which involves using the outputs from Block 5 which contains the processed features [(7 × 7 × 512) + 2 × fully connected layer (1× 1 × 4096)]. After the data is processed through all the layers, the activation function (SoftMax) allows the classification of healthy, chronic and non-chronic pulmonary conditions as example.

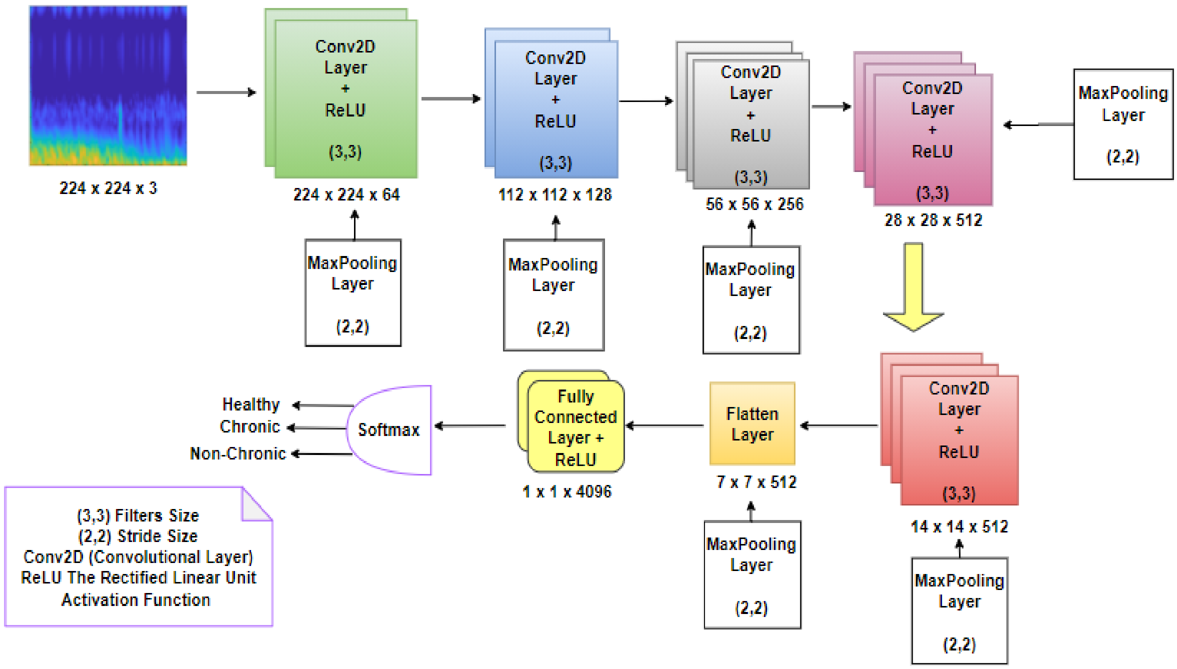


Figure 3.6: Visual Geometry Group 16 (VGG16) based CNN architecture.

ResNet-50

The ResNet [51] architecture has two types of layers – conv block and identity block, and these serve as shortcuts in the residual blocks and are included in the order, as shown in Figure 3.7.

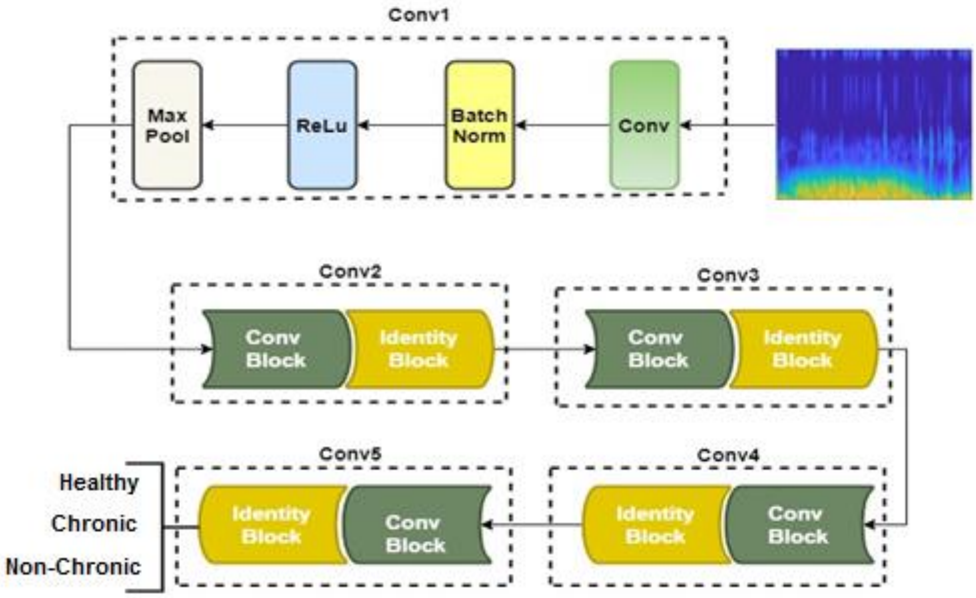
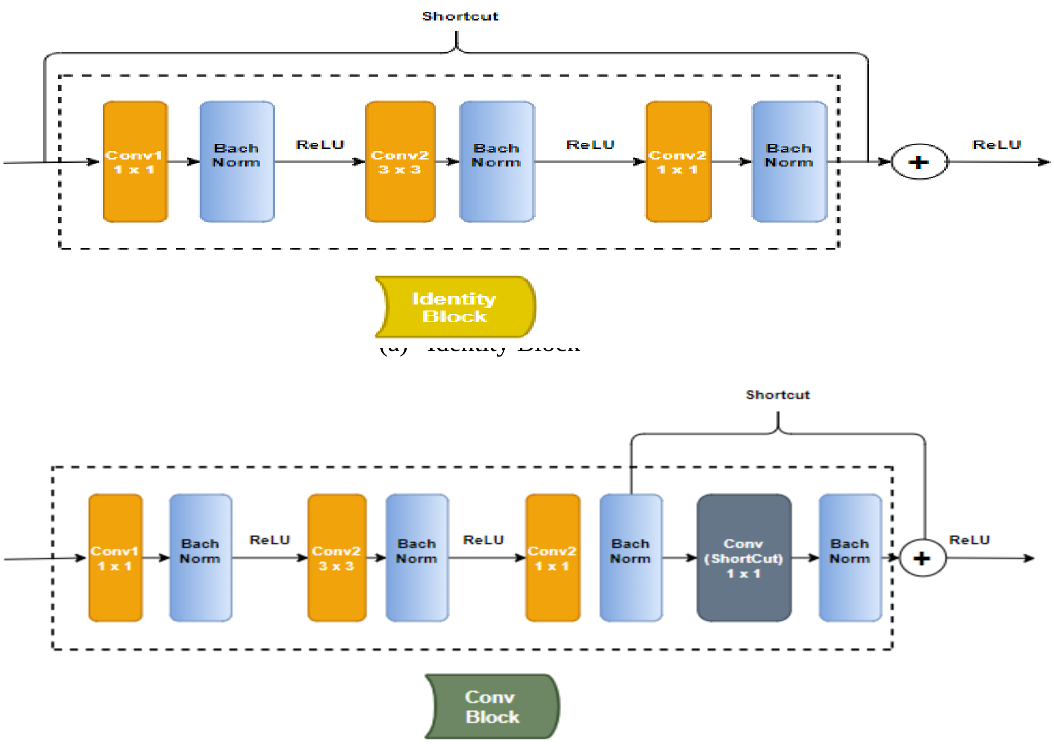


Figure 3.7: Residual network based on CNN architecture.



(b) Conv Block

Figure 3.8: Structure of the identity and conv blocks in the ResNet architecture.

Figure 3.8 (a) and (b) present the structures of the identity and conv blocks, respectively. A stack of three layers was used for each residual block. The 1×1, 3×3, and 1×1 layers are three convolutions layers. The 1×1 layers focus on first reducing and then increasing the dimensions, and the 3×3 layer has smaller input and output dimensions.

In our classification process, we focused on three main steps. First, we converted all respiratory sound segments (breathing cycles) into gammatonegram feature maps as an image.

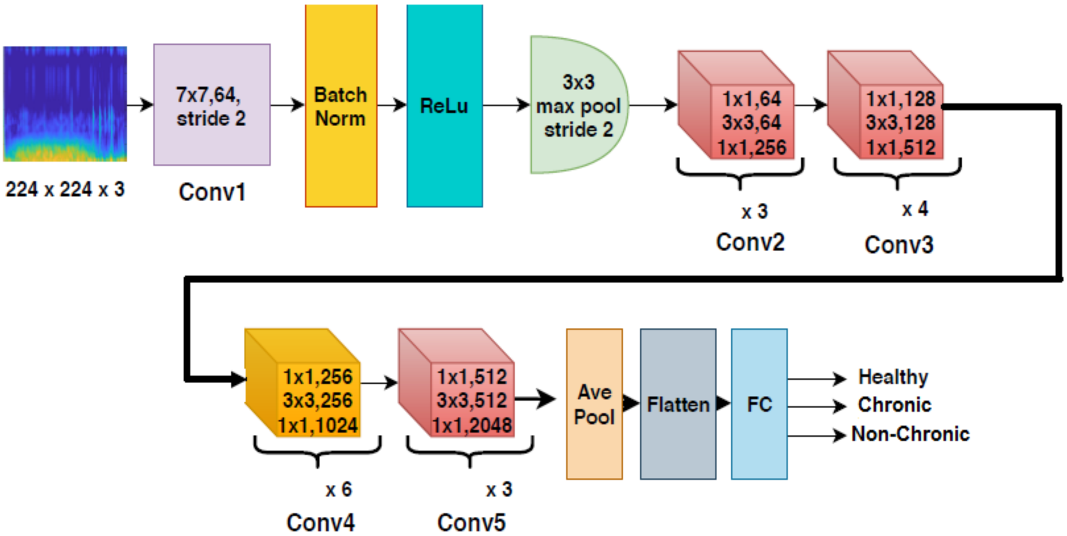


Figure 3.9: The overview of our ResNet architecture used in this work.

Second, as both networks require an input shape of size 224 × 224 × 3 pixels, we performed the resizing. Third, we fed all of our resized image data into ResNet-50. Figure 3.9 shows the ResNet-50 network. The input image is then passed through the ResNet-50 structure as shown in Figure 3.9 and the healthy, chronic and non-chronic labels are used as the output classes as exemple. In all the architectures, the term '×3' indicates one conv block and two identity block, and the rest of the terms have similar definitions. The network was divided into the following three main stages:

STAGE-1: The input feature map of three channel RGB images are pre-processed through convolution, batch-normalization, ReLU and maxpooling.

STAGE-2: As shown in Figure 3.9, the conv2,3,4, and 5 blocks are orderly. In contrast, the conv2 and conv5 blocks are similar, as shown in Figure 3.10.

STAGE-3: The output from stage 2 forms 7 × 7 × 2048 features, and these are then pooled to obtain an average of 1 × 1 × 2048 and then flattened to 2048. Subsequently, the healthy, chronic and non-chronic conditions as example. can be classified by processing the full connection (FC) layer and softmax activation.

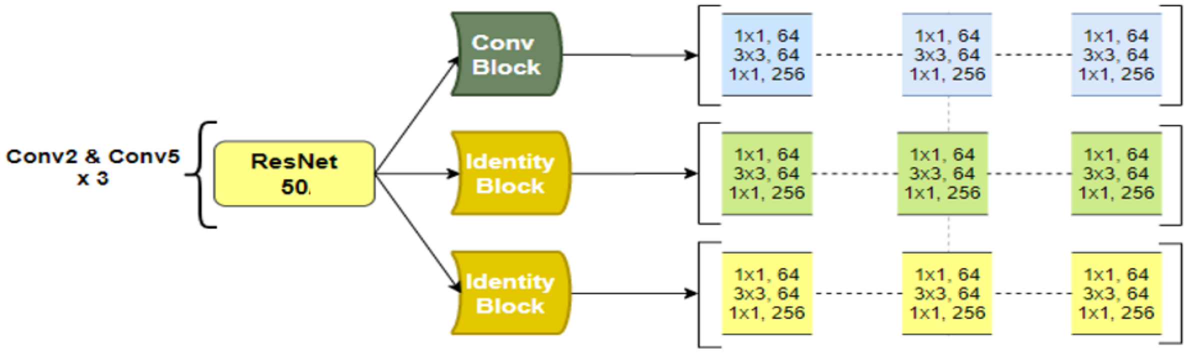


Figure 3.10: Conv2 and Conv5 blocks of the ResNet-50 architecture.

GoogLeNet

The GoogLeNet model was introduced in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) [52] as a deep learning model. This network has nine modules, each one called an inception. Each inception has a different-sized convolution and max-pooling layers. GoogLeNet has 22 layers, indicating good performance on image classification. Basically, the inception modules work as multiple convolution filters that are applied to similar inputs. After that, the results will be concatenated, allowing the network to get the advantage of multilevel feature extraction and to cover a wide area, by saving a good resolution for small information on the images. Further details of these three deep learning networks can be found in [59].

AlexNet

The AlexNet architecture composed of five convolutional layers followed by maximum pooling layers, three fully connected layers followed by drop out layers, and a 6-class Softmax classifier in the end [53]. Based on using AlexNet architectures the image data of the gammatonegrams were classified into healthy, chronic (COPD, Bronchiectasis and Asthma) and non-chronic (Pneumonia, LRTI, URTI, and Bronchiolitis) breathing sounds as exemple. Figure 3.11 shows the detailed architecture of the AlexNet network.

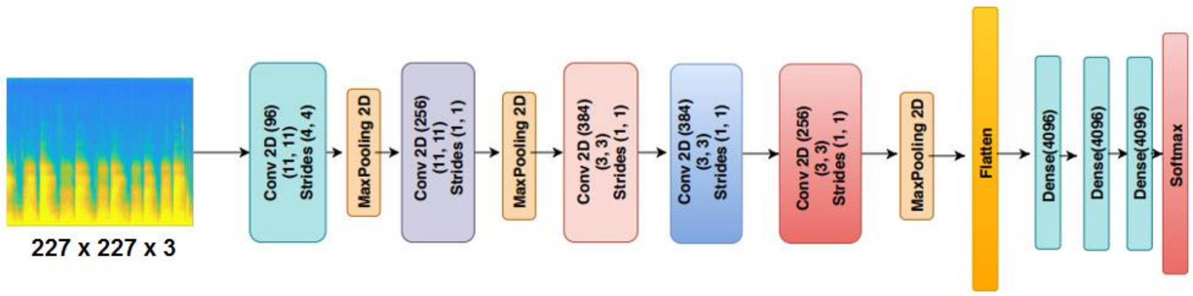


Figure 3.11: AlexNet based CNN architecture.

3.1.2.6. Proposed Method

In recent years, many researchers in the field of computer vision-based on deep learning have used various approaches for data computation based on different types of models that have been proposed on CNN. However, the analysis of very large data is often associated with time- and accuracy-related problems, especially for those working on biomedical studies and those of prediction of earlier diagnosis of Sever's diseases when the accuracy plays an important role in their applications. First, to reduce the time required for training using large amounts of data, powerful hardware (specifically more GPU accelerators) is needed, which can be sometimes costly for researchers. Second, to achieve high accuracy, researchers need to select appropriate hyperparameters, such as batch size, the number of epochs, and the learning rate after several evaluations. Taken together, these processes appear to be some form of black art. Hence, to address these problems, we have proposed a new learning and testing method that aids and accelerates the training process toward increasing accuracy. One of the most important parts in a biomedical classification system that is based on a deep neural network is the training and testing process. A common strategy used in the biomedical literature to select parameters such as batch size and epochs is to set it as a fixed value. Besides, in the previous work, there is no useful method for how to select these parameters. Hence, there is a need for studies in deep learning that discuss the learning phases (the training and testing process) and offer new perspectives in the biomedical field. This part of thesis provides a new method for the training and testing process in deep neural networks for classification of breathing sounds; we consider it as an effective method for processing the learning stages for improving accuracy and speed up the training of the system.

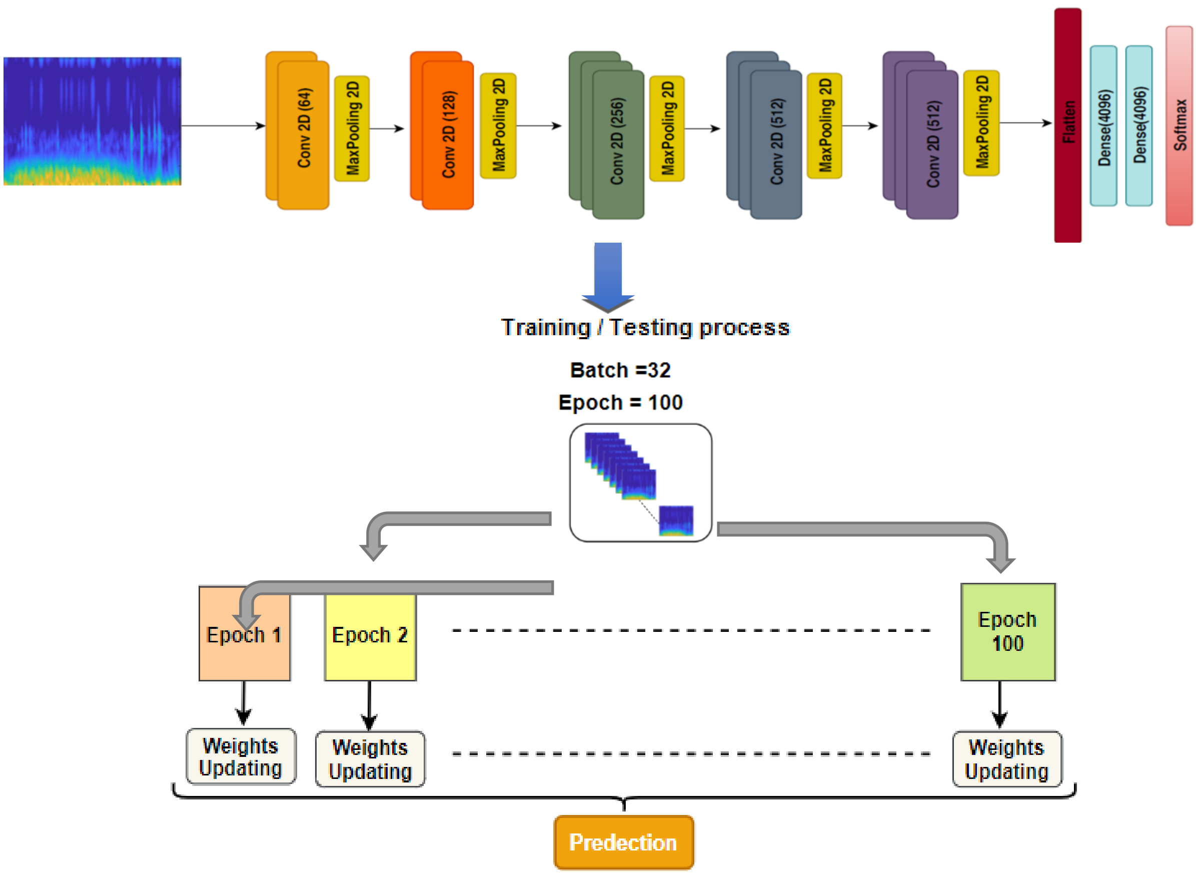


Figure 3.12: The training and test process using Fixed method (standard).

The most common critical factor used in the training and testing stages for the success of a deep neural network model-based classification system is hyperparameters such as the batch size and epoch. After several experiments conducted by researchers to get the final prominent fixed batch size and epoch values for the training and testing process, still needed improvements to achieve high performance. Therefore, in this work, multiple values of these parameters changing at the same time during the training and testing process are set according to our method. In this subsection, we clarify the process that can improve the quality of the classification system. Because we used multiple batch sizes and epoch values together in our proposed method, we thought it would be appropriate to name it Multi ep-Batch; thus, our method will be referred to by this name.

The selection of batch and epoch parameters depends on the performance achieved by it. There are many available deeply CNN architectures that have a large number of layers. The number of learnable parameters was augmented by using a complex deep learning architecture significantly. Furthermore, the training and testing time required is higher for more complex networks such as ResNet-50, VGG16, and GoogLeNet. The performance of such networks, which are based on CNN, extremely depends on the hyperparameters. By varying the batch size, the epoch value, the filter size, dropout, and so on, classification accuracy can also be varied. Good classification accuracy could be achieved with configuration turned during the training and testing process, with varying in hyperparameters such as the batch size and epochs. In [60], Samuel L. Smith et al and other members of the Google Brain team increased the batch size during training and obtained equivalent testing accuracies based on the same number of training epochs but fewer parameter updates, which led to greater parallelism and shorter training times. Keskar et al. [61] stated that “the lack of generalization ability is because large-batch methods tend to converge to sharp minimizers of the training function.” During learning, a deep neural network with connected neurons between different layers is trained to identify optimal weights and biases to improve the performance of the system with better classification accuracy after many epochs (the weights are updated at the end of each training epoch) and adjustments to its hyperparameters. In the literature, based on deep learning, there is no specific resource that examines the learning (training and testing) process in biomedical applications and offers suggestions. Motivated by this fact, a configurable learning process (training and testing) with a varying the number of batch sizes and epochs was designed. The configurable process consists of three values of batch size and three values of the epoch.

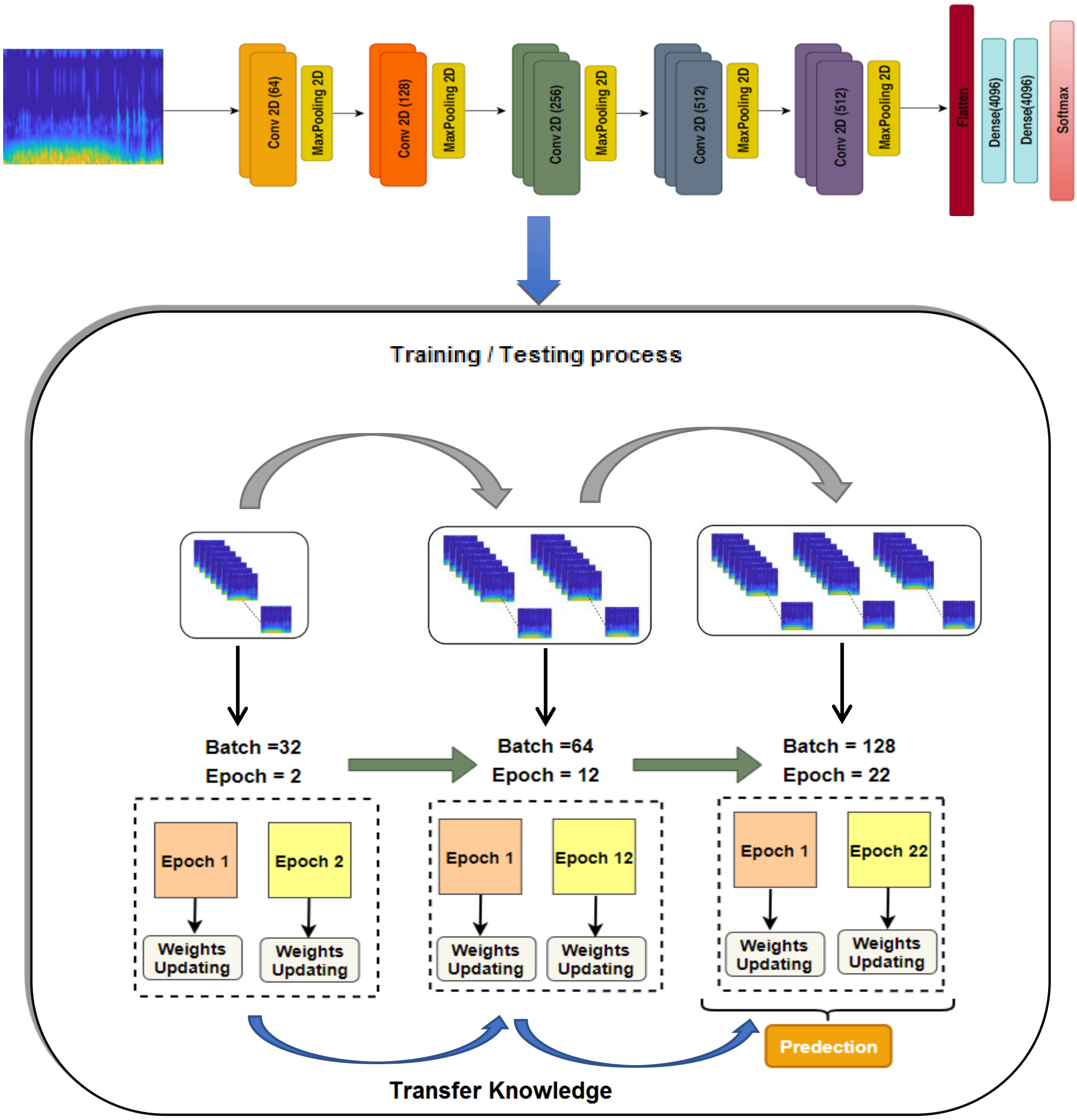


Figure 3.13: The training and test process using our proposed Multi ep-Batch

The configuration of the proposed method can be modified according to the amount of training and testing data. By recognizing the possible applications suitable for this method, it could be used in all topic that is based on deep learning, especially in training and testing phases. More knowledge is required on the topic; therefore, this study can be performed in any learning process, because it is an intervention in the learning stages. Our approach and the standard one can be distinguished as two methods: The first one is the fixed method (standard), as illustrated in Figure 3.12 with an example of VGG16, and the other adopts the Multi ep-batch method, as illustrated in Figure 3.13. To clarify further the behaviour and effect of our proposed Multi ep-batch method before we apply it on our multi-class classification case, we only consider an additional interpretation by using a different medical image dataset and different types of classification; this is not our concern, but it serves as an additional understanding of the significant role of our proposed method. We applied our Multi ep-batch on X-ray scan image datasets [62] from pediatric patients who are from 1 to 5 years of age, which consisted of two classes, namely normal and pneumonia pulmonary conditions, and a total of 5,856 chest X-ray images that were collected and labelled were used. We took data as 3,883 pneumonia images and 1,349 normal images to use a binary classification system. Table 3.1 presents the results before and after applying our proposed method; the accuracy and time were affected positively. This is the case of different data, a deep learning model and a binary classification system; as we can see, the first results give a very promising outlook for the other classification system, as in this study's multi-class system, we could say that our method positively affected the learning stages of the deep learning models.

Data	Method	Batch	Epochs	Learning rate	Accuracy	Training Time
X-ray images [62]	VGG16	32	100	0.0001	85.03 %	149 min
X-ray images [62]	VGG16 + **Multi ep-Batch**	**[32, 64, 128]**	**[2, 12, 22]**	0.0001	**93.75 %**	**51 min**

Table 3.1: The results of binary classification using X-ray images data.

Besides, Yang et al. [63] repeated experiments with different batch sizes and epochs, and obtained results as shown in Table 3.2. From these results, we can observe that variations in these two parameters play an important role in the performance of deep neural networks.

Dataset	Model	Batch	Epochs	Accuracy
ImageNet	ResNet-50 v1.5 [52]	256	90	75.9%
MNIST	LeNet [64]	256	30	99.2%
CIFAR-10	ResNet-50 v1 [52]	128	200	93.9%

Table 3.2: Results taken from [63].

In our experiments, based on previous studies conducted in this area, we investigated the training time and accuracy using our method, which is denoted as Multi ep-Batch and found that it yielded better results than the current state-of-the-art methods when applied to the International Conference on Biomedical and Health Informatics (ICBHI) database [12]. One of the supervised learning task problems observed when using CNN is the loss function, which can mathematically be described as follows [59]:

$$L_f = \frac{1}{|X|} \sum_{x \in X} c(x, w) \tag{20}$$

where,

L_f is the loss function;

w is the weight;

X is the labelled training set; and

$c(x, w)$ is the loss computed from samples $x \in X$ and their labels y.

In recent works, many researchers who used CNN applied stochastic gradient descent (SGD) and its variants for training and then optimized the loss function L_f through iterative steps as follows [65]:

$$w_{t+1} = w_t - \eta \frac{1}{|B_s|} \sum_{x \in B_s} \nabla c(x, w_t) \quad (21)$$

where,

B_s is the batch sampled from X; $|B_s|$ is the batch size;

η is the learning rate; t is the iteration index; and η is the learning rate for iteration i.

A single gradient computation divided by the weight corresponds to a single iteration of the SGD-based training. One epoch corresponds to $X/\ |B_s|$ iterations in the training loop and it represents a single pass of the full data. Figure 3.14 shows the process followed to get the optimal weights corresponding during the learning phases to every model.

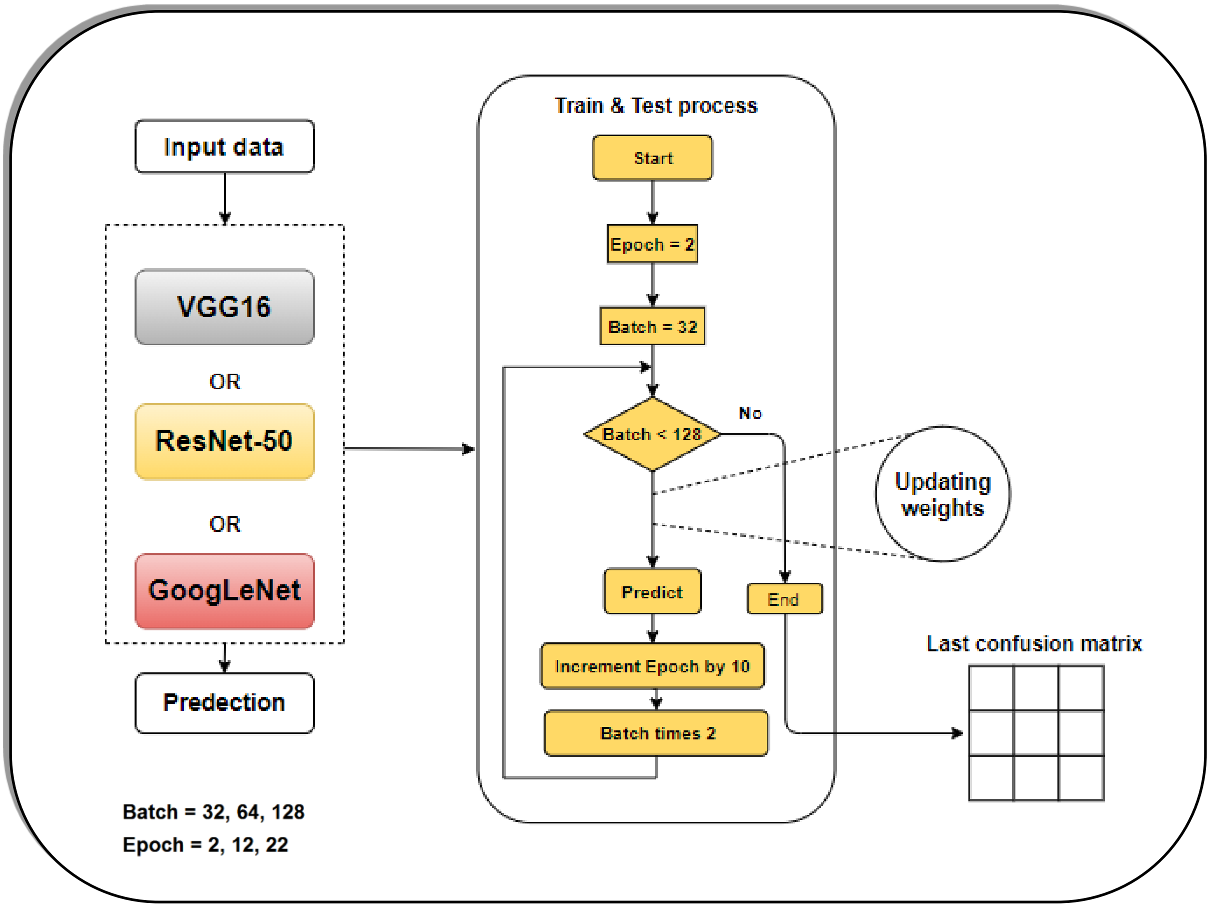


Figure 3.14: The procedure followed for getting the optimal weight during learning phases to every model.

We can perform consecutive calls for model fitting. Hence, in this study 3, 5 and 3 calls (model. fit) were used for updating and saving the weights during the Multi ep-Batch training approach in *scenario (i)*, *scenario (ii)* and *scenario (iii)* respectively. The weights are saved only if an improvement in the accuracy was obtained with the test data during the training step using the Checkpointing function in the Keras framework [66].

We can calculate the number of iterations for one epoch using the following equations:

$$\mathrm{i} \leftarrow (\mathrm{D}/\mathrm{B_s}) \tag{22}$$

$$\mathrm{i_T} \leftarrow (\mathrm{D_{Tr}}/\mathrm{B_s}) \tag{23}$$

$$\mathrm{i_V} \leftarrow (\mathrm{D_{Ts}}/\mathrm{B_s}) \tag{24}$$

$$\mathrm{E} = (\mathrm{i_T}, \mathrm{i_V}) \tag{25}$$

where,

i is the number of total iterations; D is the global database (all image samples); B_s is the batch size; E is the epoch; i_T is the training iteration in one epoch; $\boldsymbol{i_V}$ is the testing iteration in one epoch; and D_{Tr} and D_{Ts} are the training and testing data, respectively. To illustrate the application of the aforementioned formulas, we will apply them to the ICBHI dataset experimented in this study (Conditions-based). For the modified training process using the Multi ep-Batch method as shown in Figure 3.13, we obtain:

$\mathrm{D} = 7080$

$\mathrm{B_s} = [32, 64, 128, 256, 512]$

$\mathrm{D_{Tr}}$ = 5664 (3 classes: healthy, chronic and non-chronic pulmonary conditions)

$\mathrm{D_{Ts}}$ = 1416 (3 classes: healthy, chronic and non-chronic pulmonary conditions)

In other words, we started the training process with $B_s = 32$ and by using equations (22), (23) and (24) we get the following variables – (i = 382, i_T = 177, and i_V = 44). Each epoch E has two sets of iterations (train and test iterations). Therefore, our Multi ep-Batch method involves multiple E (from 2 to 42) and multiple B_s (from 32 to 512) values. Each time the batch is augmented, 10 epochs are added, and the number of iteration decreases signifying quicker convergence. All weights are updated based on the different numbers of batches and epochs during all iterations, as shown in Figure 3.14. Using the standard training process as shown in Figure 3.12 we implemented the same process as aforementioned but fixed both hyper-parameters as $B_s = 32$ and E = 100. To benchmark and evaluate the efficiency of our proposed method, we carried out the experiments in three different scenarios using different sub-data, and we organized the investigation into three main experiments A, B and C as shown in Table 3.3. The last column (Multi ep-Batch Method) indicates that our proposed solution is enabled or disabled (/).

Scenario	Sub-data	Deep learning model	Experiment	Multi ep-Batch Method
Scenario (i)	Symptoms-based	VGG16	A (1)	/
		VGG16	A (2)	Enable
		ResNet-50	B (1)	/
		ResNet-50	B (2).	Enable
		GoogLeNet	C (1)	/
		GoogLeNet	C (2)	Enable
Scenario (ii)	Conditions-based	VGG16	A (1)	/
		VGG16	A (2)	Enable
		ResNet-50	B (1)	/
		ResNet-50	B (2)	Enable
		AlexNet	C (1)	/
		AlexNet	C (2)	Enable
Scenario (iii)	Diseases-based	VGG16	A (1)	/
		VGG16	A (2)	Enable
		AlexNet	B (1)	/
		AlexNet	B (2)	Enable

(/): disable

Table 3.3: Three different scenarios using different sub-data.

3.1.2.7. Datasets

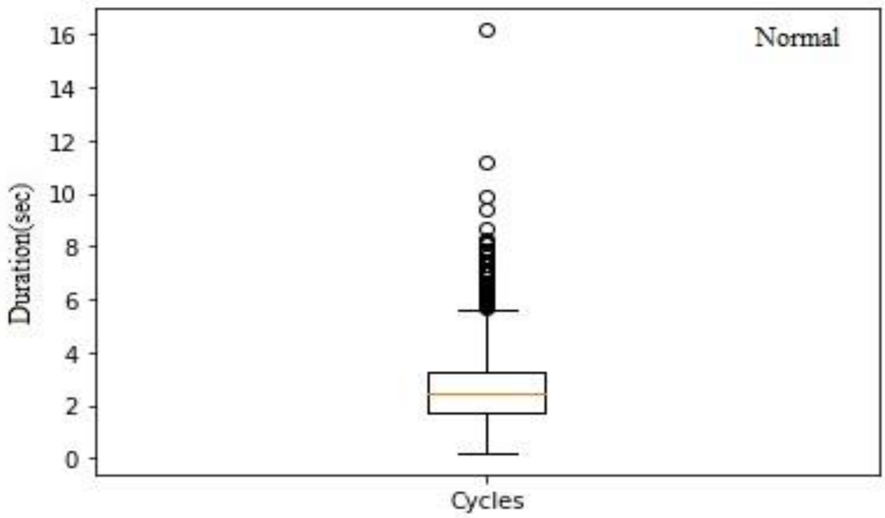
The lung sounds used in this study are sounds included in the ICBHI Scientific Challenge database [12], which contains 920 audio recordings of breathing sounds from healthy and diseased individuals. Table 3.4 provides the breakdown of the information per recording in the database, which includes the patient number, chest location, mode of acquisition, recording equipment, type of breath sounds (per cycle), diagnosis and a total number of cycles in the recording. The audio recordings ranged in duration from 10 to 90 s with sampling frequencies from 4.0 to 44.1 kHz. We resampled all audio used on three sub-data to 8000 Hz to ensure consistency among the data. Finally, each sub-data was randomly divided into 80% for training and 20% for testing for every type of lung sound.

Patient number	Chest location	Acquisition mode	Recording equipment	Breathing sound type	Diagnosis	Cycle numbers
101	Anterior left	Single-channel	Welch Allyn Meditron Master Elite electronic stethoscope	Normal sounds	Healthy	12 cycles
122	Trachea	Multi-channel	3M Littmann Classic II SE stethoscope	Crackles	Pneumonia	10 cycles
107	Anterior right	Multi-channel	AKG C417L microphone	Crackles	COPD	8 cycles

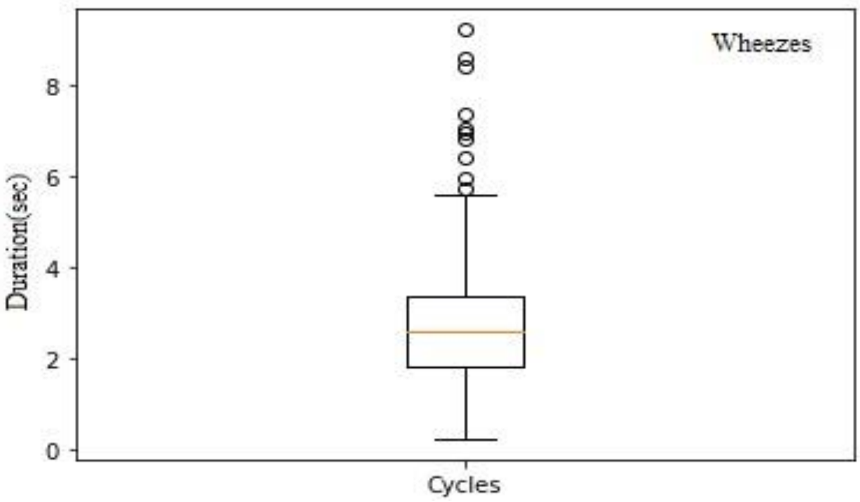
Table 3.4: Example of recording sounds characteristics from the ICBHI database.

Sub-data (Symptoms-based)

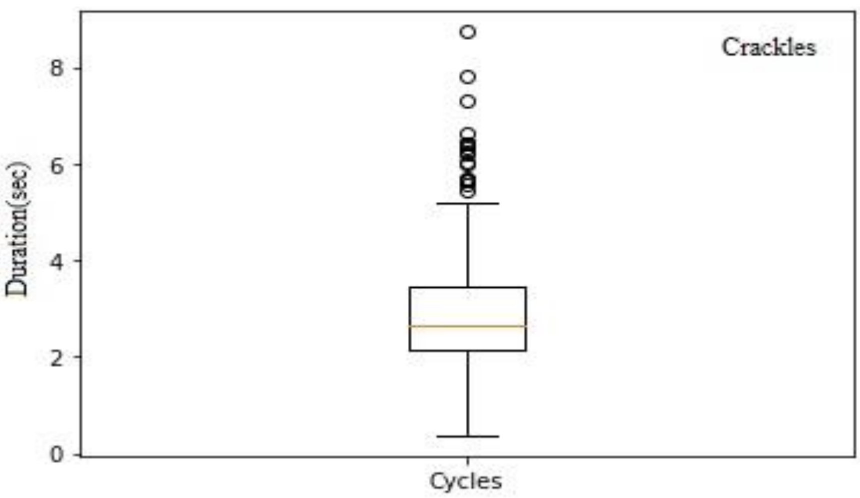
All the recordings were divided into a total of 1864, 886, and 3642 cycles, which contained the three different types of sounds, namely crackles, wheezes and normal sounds respectively.



(a)



(b)



(c)

Figure 3.15: Distribution of the cycles data for normal, crackle, wheeze.

Sub-data (Conditions-based)

In total, the number of cycles used in this sub-data was set to 1416 cycles, and these included 374 cycles from chronic (264, 104 and 6 from COPD, Bronchiectasis and Asthma, respectively) cases, 720 cycles from non-chronic (285, 243, 160 and 32 from Pneumonia, URTI, Bronchiolitis and LRTI, respectively) cases and 322 cycles from healthy cases. The key step to increase the cardinality of the audio data for lung sound classification tasks is data augmentation, which used in the training of deep learning networks. Typically, to overcome the problem of overfitting two ways can be done to perform the data augmentation – one is performed on the 1D time-series data (audio data) and the other performed on the 2D gammatonegram images (image data). In our work, we opted for the second method. The gammatonegram images were augmented with 4 different strategies, based on simple transformations [67] – horizontal flipping, vertical flipping, horizontal plus vertical flipping, and added noise. A summary of the sliced recorded audio files and the final number of augmented gammatonegram images corresponding to each disease class is presented in Table 3.5.

Diseases	No. of Cycles	No. of Cycles selected	No. of Augmented Images
Healthy	322	322	1610
COPD	5746	264	1320
Bronchiectasis	104	104	520
Asthma	6	6	30
Pneumonia	285	285	1425
LRTI	32	32	160
URTI	243	243	1215
Bronchiolitis	160	160	800

Table 3.5: Dataset summary (Conditions-based)

Sub-data (Diseases-based)

The number of recordings used in this sub-data was 203, and these included (79, 35, 37, 23, 16, and 13 from COPD, Healthy, Pneumonia, Bronchiectasis, Bronchiolitis and URTI respectively) cases. In this sub-work, the gammatonegram images were augmented with 4 different strategies as mentioned before (condition-based). A summary of the subject, recordings audio files and final augmented gammatonegram images with corresponding diseases classes are presented in Table 3.6. As we can see, from the table is the number of the recording of COPD represents the highest number of all the recording numbers, which means the COPD recordings number make an unbalanced distribution. To balance the data, we reduced the COPD samples to 79.

Diseases	No. of Subjects	No. of Recordings	No. of Recordings selected	No. of Augmented Images
Healthy	26	35	35	175
COPD	**64**	**793**	**79**	**79**
Bronchiectasis	7	16	16	80
Pneumonia	6	37	37	185
URTI	14	23	23	115
Bronchiolitis	6	13	13	65

Table 3.6: Dataset summary (Diseases-based)

3.1.2.8. Computing Platform

Google Colaboratory [68], which is based on the Jupyter notebook, was used in this work for all the training experiments. This notebook is an open-source solution for running and sharing code written in the Python programming language and provides user-friendly tools for data integration, libraries, and visualization [69]. The objective of this platform is to distribute machine learning education and research [70]. This platform can be run using highly powerful hardware – parallel tensor processing units (TPUs) and graphic processing units (GPUs), with accelerated training of a maximum of 12 hours per user per session. This platform has an NVIDIA Tesla T4 with a GPU of 12 GB, and we obtained permission to upload the data from Google drive and save our training session. In this experiment, we built our VGG16, ResNet-50, AlexNet and GoogLeNet model from Keras on top of TensorFlow, which is an API designed to support deep neural network architectures [66].

3.1.2.9. Imaging Software

The software used in this experiment to generate the gammatonegram images of lung sounds is MATLAB 2019. The generated gammatonegrams were saved as images in JPG format and fed into our deep neural networks.

3.1.2.10. Performance evaluation criteria

In this study, an optimization algorithm named Adaptive Moment Estimation (Adam) was examined to update the network weights during training with VGG16, ResNet-50, AlexNet and GoogLeNet CNN models. The experimental settings for modelling the network are listed below in Table 3.7.

Settings. No	Hyperparameters	Values
1	Learning rate	0.00001
2	Numbers of Epochs	[2, 12, 22, 32, 42]
3	Batch Size	[32, 64, 128, 256, 512]
4	Optimizer	Adam

Table 3.7: Hyperparameters settings for trained VGG16, ResNet-50, AlexNet and GoogLeNet models.

In this study, the experimental results were evaluated using four metrics – overall accuracy, precision, recall or sensitivity and F1 score. The overall accuracy measures the number of correctly classified normal and abnormal samples corresponding to all test samples. Precision is defined as the positive predicted value (PPV) that provides the results relevant to an accurate classification. Recall or sensitivity is defined as true positive classified segments, divided by the total number of positive segments. The F1 score is based on the harmonic average of specificity and sensitivity, and the definitions are given in (26), (27), (28) and (29) respectively as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{26}$$

$$Precision = \frac{TP}{TP + FP} \tag{27}$$

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN} \tag{28}$$

$$F - score = 2 * \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{29}$$

Where:

TP: means true positive;
TN: means true negative;
FP: means false positive;
FN: means false negative.

CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents the results of the two different parts such as machine learning and deep learning algorithms in lung sound classification using the proposed methods which were discussed in Chapter 3. The performance of machine learning and deep learning for classification system in detecting the lung sounds and diseases are presented in Part 1 and Part 2 respectively. This chapter also compares the results from various machine learning and deep learning algorithms to find out the optimum system for lung sounds classification.

4.1. Part 1: Performance of Machine Learning Algorithms in Breathing Sounds Classification

In this part, MATLAB R2013b have been used to performe the all experiments on a pc with a configuration of Intel CPU Core i5, 4 GB RAM, and Windows 10 operating system. In [71] the authors apply Hjorth descriptors as features and find that the activity feature is the best feature compared with mobility and complexity as shown in Equations (1),(2) and (3). Therefore, in our work, the activity feature was exploited for enhanced this study with combined it with the permutation entropy feature shown in Equation (4), and formed a features vectors to fed into two machine learning algorithms namely ELM and K-NN, to compare them in the classification of breath sounds signals.

The EMD decomposes BS signals into a set of IMFs. The features (Activity and Permutation Entropy) were extracted from each IMF and tested using a statistical measure of (mean and standard deviation SD described in Equations (1) and (2)) as tabulated in Table 4.1.

Breath sounds	Activity Mean ± standard deviation	Permutation entropy Mean ± standard deviation
Normal bronchial	0.001387 ± 0.004387	0.56497 ± 0.202819
Wheeze	0.005829 ± 0.000468	0.56792 ± 0.209888
Crackle	0.000645 ± 0.000765	0.586679 ± 0.216714
Pleural rub	0.002608 ± 0.000513	0.56675 ± 0.223923
Stridor	0.001594 ± 0.001472	0.563352±0.215641

Table 4.1: Statistical analysis of features extracted from breath sounds

From Table 4.1 we inferred that there is significant discrimination in the activity and PE features of different classes. Can be observed a mean and SD are different from each class in activity features, but in PE features a little different between classes. From this, we can combine theme to test and compare the classification accuracy of both K-nn and ELM classifiers. These features have been formed as follows:

Features = [Activity, PE].

In order to verify the reliability of the outcome of the classifiers, the k-fold cross-validation was used. After several tests to choose the k value, we found that k=10 is promised value, therefore it has been used in this study.

In the literature review many researchers based on activity or entropy features extraction, nevertheless, this study has combined both activities and PE features for observed the ability of both ELM and K-nn to classify different BS signals.

In Table 4.2 the classification stage is described, and give the classification performance of features (Activity, PE) extracted from IMFs using ELM with RBF Kernel, Polynomial Kernel and K-nn with distance euclidian which is described in equation (17) , and 1 to 10 number of neighbours.

Classifier	K-Fold	K neighbours	Kernel	Average accuracy (%)
ELM (Activity, PE)	10	/	RBF	**83.57**
ELM (Activity, PE)	10	/	Polynomial	77.86
K-nn (Activity, PE)	10	1	/	**86.42**
K-nn (Activity, PE)	10	2	/	80.71
K-nn (Activity, PE)	10	3	/	82.14
K-nn (Activity, PE)	10	4	/	80.00
K-nn (Activity, PE)	10	5	/	82.14
K-nn (Activity, PE)	10	6	/	80.00
K-nn (Activity, PE)	10	7	/	81.42
K-nn (Activity, PE)	10	8	/	82.14
K-nn (Activity, PE)	10	9	/	81.00
K-nn (Activity, PE)	10	10	/	77.14

Table 4.2: Classification performance of (Activity, PE) from IMFs of BS signals in multiclass classification stage

As shown in Tables 4.2, The ELM with RBF Kernel and K-nn with 1 neighbour gave the higher classification accuracy of 83.57% and 86.42% respectively. The ELM by RBF kernel is better than ELM with Polynomial kernel in multiclass classification case, and k-nn by 1 neighbour is better than rest neighbours. We can say that the ability of the K-nn is higher than ELM in the classification of the Breath sounds signals into several classes (Normal bronchial, Wheeze, Crackle, Pleural rub, Stridor).

We can be seen in Table 4.3, the accuracy found from k-nn is 95% by 6-8-10 neighbours and from ELM with Polynomial Kernel is 90.71% better than RBF kernel in binary classification case.

Classifier	K-Fold	K neighbours	Kernel	Average accuracy (%)
ELM (Activity, PE)	10	/	RBF	89.29
ELM (Activity, PE)	10	/	Polynomial	**90.71**
K-nn (Activity, PE)	10	1	/	93.57
K-nn (Activity, PE)	10	2	/	92.14
K-nn (Activity, PE)	10	3	/	94.29
K-nn (Activity, PE)	10	4	/	92.86
K-nn (Activity, PE)	10	5	/	94.28
K-nn (Activity, PE)	10	6	/	**95.00**
K-nn (Activity, PE)	10	7	/	94.29
K-nn (Activity, PE)	10	8	/	**95.00**
K-nn (Activity, PE)	10	9	/	93.00
K-nn (Activity, PE)	10	10	/	**95.00**

Table 4.3: Classification performance of (Activity, PE) from IMFs of BS signals in binary classification stage

However, according to these results, we can say that, this comparative study shows that the capability of the k-nn classifier is higher compared with that of the ELM classifier in the classification of breath sounds signals from our test conditions. the ability of the k-nn is higher than ELM in the classification of the Breath sounds signals into binary and multiclass classification cases.

In this part, the performance of the ELM and K-nn classifiers were compared using the Hjorth descriptors (Activity) and Permutation Entropy (PE) features in distinguishing between breath sounds signals with combination these features (Activity, PE). The features extracted were analyzed statistically by calculating a mean and standard deviation to observe the difference between them for each class (Normal bronchial, Wheeze, Crackle, Pleural rub, Stridor). The classification accuracy in multiclass classification case of the ELM and k-nn classifiers is 83.57% and 86.42% respectively, and in binary classification case, the accuracy is 90.71% , 95% respectively. These show that the ability of k-nn in our test conditions (database, methods of analyses the breath signals, and features used) is higher than the ELM classifier in multiclass and binary classification. In future work, the EMD methods will be compared with another method for further analysis of breath sounds signals using a large database.

4.2. Part 2: Performance of Deep Learning Algorithms in Breathing Sounds Classification

In this part, we implemented our novel method that involves the use of Gammatonegrams as input to the deep neural networks and the use of multi epochs and batch sizes (Multi ep-Batch) during the training and testing steps in three different scenario.

4.2.1. Scenario (i) – Symptoms-based

4.2.1.1. Results

The ability of the proposed Multi ep-Batch method applied on VGG16, ResNet-50 and GoogLeNet architectures, to accurately classify breathing sounds such as normal, crackles and wheezes sounds were compared to the fixed (standard) method. Figure 4.1 shows the three Gammatonegram feature maps of the normal, wheezes and crackles sounds.

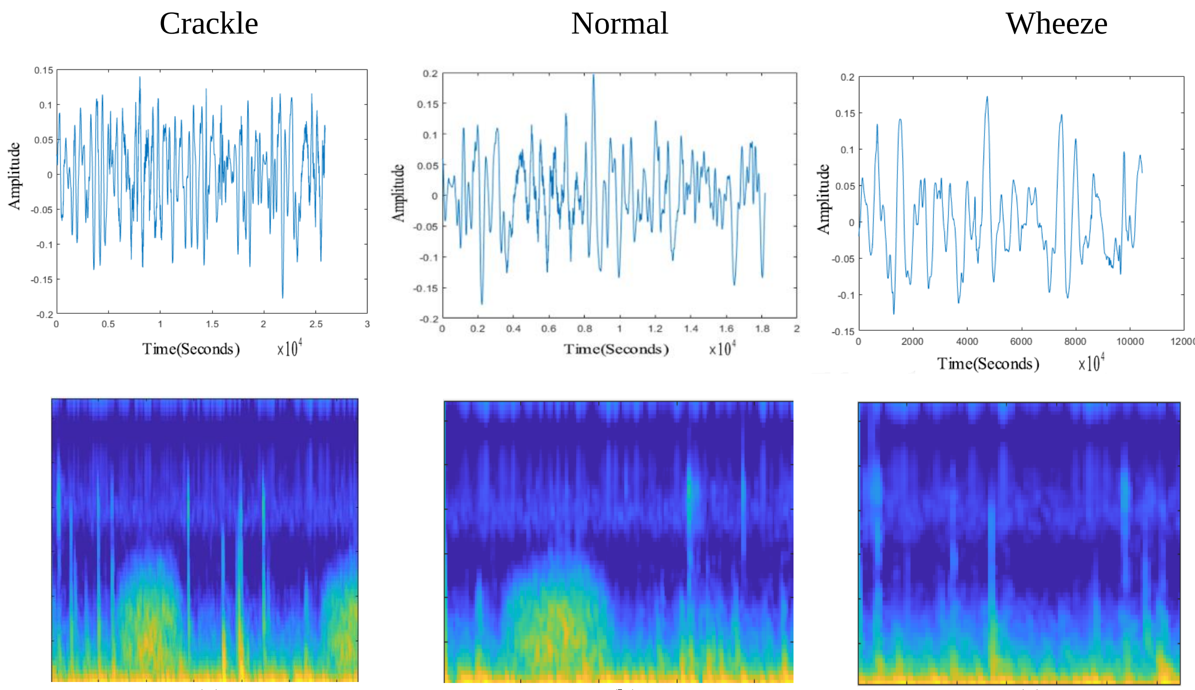
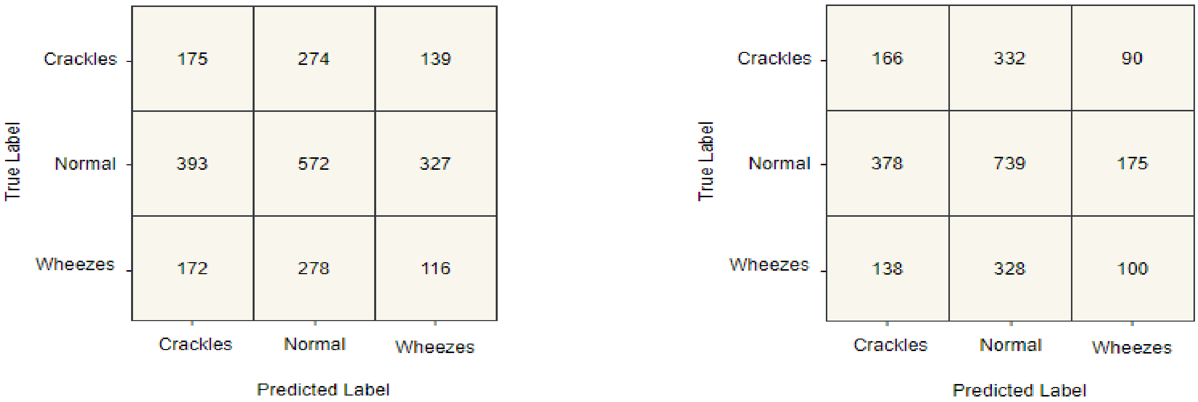


Figure 4.1: Gammatonegram feature maps of (a) crackle, (b) normal and (c) wheeze sounds.

We started executing our experiments on Google Colaboratory, which involved feeding the gammatonegram cycle images into VGG16 model, the accuracy results and timings of experiments are shown in Table 4.4 (experiment set A), for their classification into normal, crackle and wheeze breathing sounds without the Multi ep-Batch method. To evaluate the results obtained after several experiments, we selected a prominent value of 0.00001 for the learning rate in our study and used this value for all other subsequent experiments. As shown in Table 4.4, the same fixed batch size and epoch parameter yielded an accuracy of 62.50% over a 295 min training time for all three classes (A1). We then implemented the same process as in A1, but our proposed Multi ep-Batch method option turned on in experiment (A2) yielded an accuracy of 75.00%. We observed improvements in both time and accuracy compared to A1. As detailed in the table, there is an improvement in the obtained accuracy, and a reduction in training time when compared to the fixed batch-epoch approach (standard). To further explore the classification performance of the VGG16 learning method with and without Multi ep-Batch, we generated the corresponding confusion matrix as shown in Figure 4.2.

Experiment	Batch Size	Epoch	Accuracy %	Training Time
A (1) VGG16	32	100	62.50 %	295 min
A (2) VGG16+ **Our Method**	[32, 64, 128]	[2, 12, 22]	**75.00 %**	**100 min**

Table 4.4: VGG16 performance of Gammatonegrams cycles classification.



(a) (b)

Figure 4.2: Confusion matrixes for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Experiment	Pulmonary Condition	Precision	Recall	F1-score
A (1) VGG16	Crackles	0.21	0.26	0.23
A (1) VGG16	Normal	0.51	0.50	0.51
A (1) VGG16	Wheezes	0.23	0.19	0.21
A (2) VGG16+ Multi ep-Batch	Crackles	0.21	0.24	0.23
A (2) VGG16+ Multi ep-Batch	Normal	0.53	0.51	0.52
A (2) VGG16+ Multi ep-Batch	Wheezes	0.26	0.25	0.26

Table 4.5: Precision, Recall and F1-Score comparison between two Experiments

Table 4.5 summarizes the quality of the network through the performance parameters. From the table, it can be seen that the precision, recall and F1 score do not suffer any severe degradation when the Multi ep-Batch is used as a learning method. The overall values of precision, recall and F1-score in the experiments provide preliminary evidence that the Multi ep-Batch learning method can make the VGG16 network discriminate a multiclass problem. Therefore, it can be inferred that the classification of breathing sounds is possible using the proposed Multi ep-Batch learning method in conjunction with the VGG16 network. This system yields a maximum accuracy of 75.00%. We then executed the same set of experiments again, but instead of using the VGG16 model, we implemented another deep learning network – ResNet-50 (experiment set B). The results are also shown in Table 4.6. In experiment (B1), with the same fixed batch size and epoch hyper-parameters, an accuracy of 62.29% was obtained with a training time of 254 minutes, which indicates that the use of a more complex model from scratch makes the learning process resulting in a long training time in both B and A. This observation can be explained by the fact that all the layers in the ResNet-50 model obtained from scratch were trained from new images to create new weights. When the Multi ep-Batch method turned on (B2), we obtained accuracy results higher to that obtained with (B1), and also within a short training time. In this second round of experiments, we obtained the best accuracy results of 71.09% with an overall improvement in accuracy of about 14% and training time reduction of 44%.

Experiment	Batch Size	Epoch	Accuracy %	Training Time
B (1) ResNet-50	32	100	62.29 %	254 min
B (2) ResNet-50 + **Our Method**	[32, 64, 128]	[2, 12, 22]	**71.09 %**	**142 min**

Table 4.6: ResNet-50 performance of Gammatonegrams cycles classification.

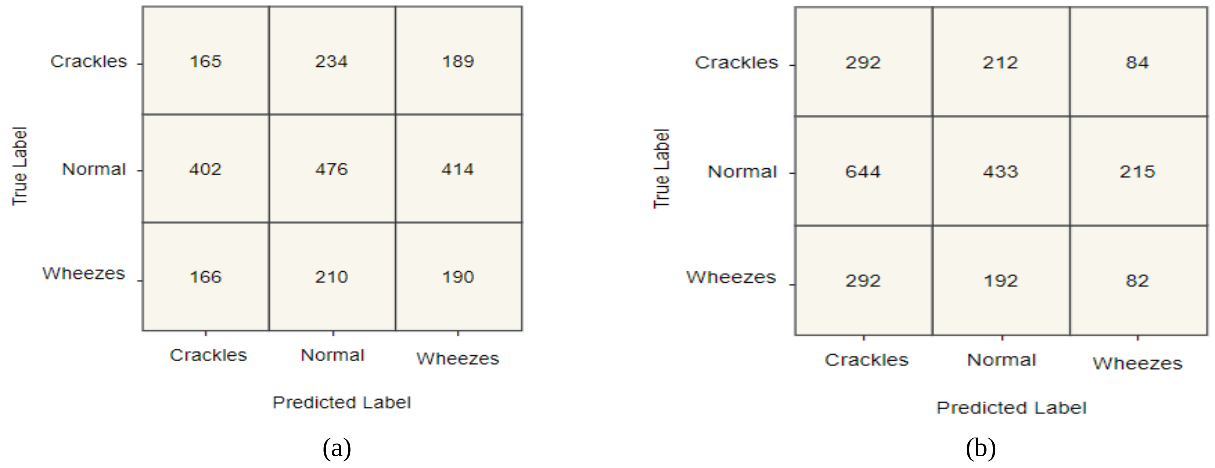


Figure 4.3: Confusion matrixes for the ResNet-50 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

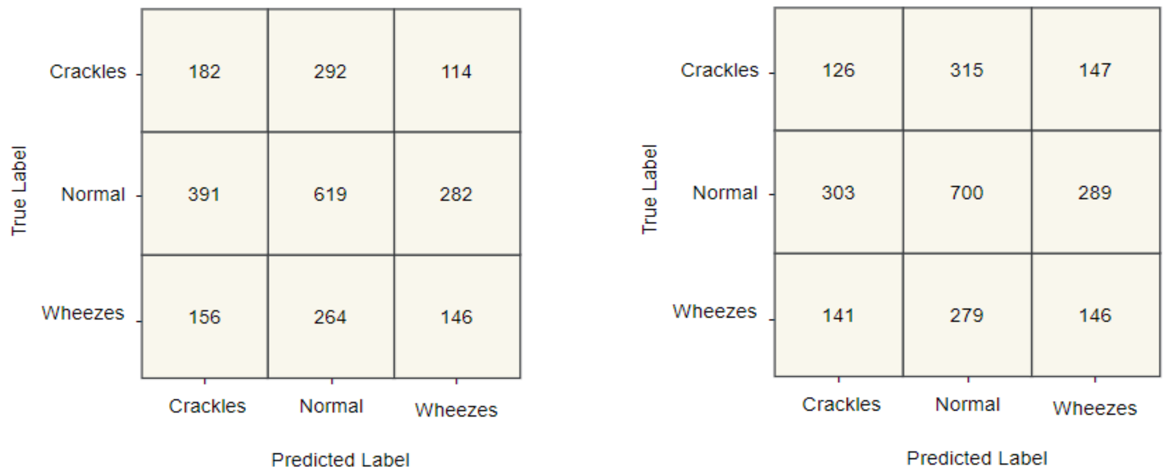
Experiment	**Pulmonary Condition**	**Precision**	**Recall**	**F1-score**
B (1) ResNet-50	Crackles	0.23	0.28	0.25
B (1) ResNet-50	Normal	0.52	0.37	0.43
B (1) ResNet-50	Wheezes	0.24	0.34	0.28
B (2) ResNet-50 + Multi ep-Batch	Crackles	0.24	0.50	0.32
B (2) ResNet-50 + Multi ep-Batch	Normal	0.52	0.34	0.41
B (2) ResNet-50 + Multi ep-Batch	Wheezes	0.22	0.14	0.17

Table 4.7: Precision, Recall and F1-Score comparison between two Experiments (ResNet-50)

To get more insight into the performance of the Multi ep-Batch method, performance parameters of the ResNet-50 network are summarized in Table 4.7. The values (>0.5) of precision, in both B experiment sets (B1 and B2) for the normal class, provide preliminary evidence that when one uses the Multi ep-Batch learning method, there is no severe degradation, and vice versa; the value (0.50) of recall was achieved when our proposed method was turned on (B2) for the crackles class, but with the fixed method as we can see the value (<0.5) of this class; this means that our proposed method can make the ResNet-50 network discriminate better between the normal and abnormal samples. Hence, we can say that the proposed Multi ep-Batch learning method in conjunction with ResNet-50 outperformed the fixed method. Using this approach, the system yields a maximum accuracy of 71.09% in the network. To further clarify the effect of our Multi ep-Batch learning method on the CNN deep learning architectures, we evaluated the performance of another well-known and popular network, namely GoogLeNet. We obtained an accuracy of 63.69%, as we can see in Table 4.8, by using the fixed method (C1) for the classification and a training time of 222 min. After that, we performed the same process but by using our Multi ep-Batch method (C2). As detailed in the table, there is an improvement in the obtained accuracy, as well as a 3 times reduction in training time, when compared to the fixed method (standard). Furthermore, Figure 4.4 illustrates the confusion matrix of the proposed method (with and without the Multi ep-Batch learning method).

Experiment	**Batch Size**	**Epoch**	**Accuracy %**	**Training Time**
C (1) GoogLeNet	32	100	63.69 %	222 min
C (2) GoogLeNet + **Our Method**	[32, 64, 128]	[2, 12, 22]	**68.06 %**	**62 min**

Table 4.8: GoogLeNet performance of Gammatonegrams cycles classification.



(a)

(b)

Figure 4.4: Confusion matrixes for the GoogLeNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Table 4.9 illustrates the performance results of the GoogLeNet network. The parameters listed in Table 4.9 also indicate that our approach generalizes well to different deep learning architectures. Based on the values (>0.5) of precision, recall and F1 score in experiment sets (C2) for the normal class, the tone can say that our proposed method successfully affected the training and testing stages of GoogLeNet. This system yields a maximum accuracy of 68.06%. The primary importance of this study is that this Multi ep-Batch method can achieve considerable accuracy and learning time without any modification to the network. Also, it can be concluded that the proposed method can now be utilized in any deep learning classification system as long as it is a method that is incorporated into the training and testing stages for any deep learning applications.

Experiment	Pulmonary Condition	Precision	Recall	F1-score
C (1) GoogLeNet	Crackles	0.25	0.31	0.28
C (1) GoogLeNet	Normal	0.53	0.48	0.50
C (1) GoogLeNet	Wheezes	0.27	0.26	0.26
C (2) GoogLeNet + Multi ep-Batch	Crackles	0.22	0.21	0.22
C (2) GoogLeNet + Multi ep-Batch	Normal	0.54	0.54	0.54
C (2) GoogLeNet + Multi ep-Batch	Wheezes	0.25	0.26	0.25

Table 4.9: Precision, Recall and F1-Score comparison between two Experiments (GoogLeNet)

4.2.1.2. Discussion

Computer-aided classification of lung sounds can be used in expedited diagnoses of various pulmonary diseases. In this work, we propose changing the batch size and epoch values during the training and testing phases to perform the multi-class classification of three lung sounds. As imperative to all deep learning frameworks, the learning (training and testing) process stage aims at providing a better updating of the network weights and biases. Thus, optimized hyperparameters are essential to the implementation of an effective classification system and improvement of the model's accuracy. To highlight the advantages and for further comparison and understanding of the effect of the proposed method, the accuracy and training time for three networks architectures, VGG16, ResNet-50 and GoogLeNet, are compared based on the fixed and Multi ep-Batch methods as an additional view. As we can see in Figure 4.5, the performance of the proposed Multi ep-Batch method in conjunction with these three networks is found to be superior. GoogLeNet gives an accuracy of 63.69% for the fixed method, whereas VGG16 and ResNet-50 give 62.50% and 62.29%, respectively. In each case, our proposed Multi ep-Batch method was affected positively in both accuracy and training time. A summary of the comparison between the results is presented in Table 4.10.

Methods	VGG16		ResNet-50		GoogLeNet	
	Accuracy %	**Training Time**	**Accuracy %**	**Training Time**	**Accuracy %**	**Training Time**
Fixed method (standard)	62.50 %	295 min	62.29 %	254 min	63.69 %	222 min
Multi ep-Batch method	**75.00 %**	**100 min**	**71.09 %**	**142 min**	**68.06 %**	**62 min**

Table 4.10: A comparison result between proposed and fixed methods using different CNN models (Symptoms-based).

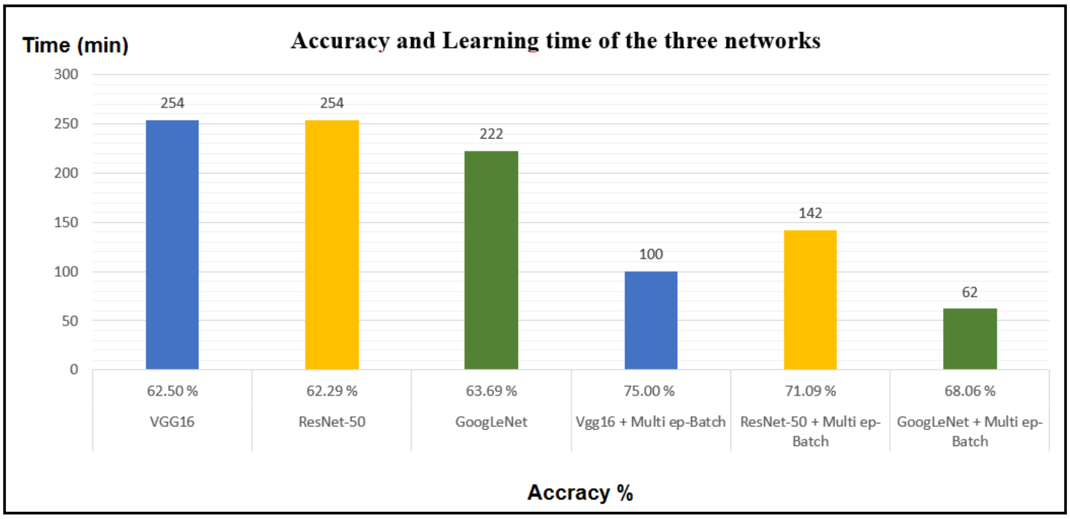


Figure 4.5: Accuracy and Learning time distribution for both fixed (standard) and multi ep-batch methods based on VGG16, ResNet-50 and GoogLeNet architectures.

First, our proposed method in conjunction with the three deep network models outperforms all other models based on the fixed method and obtains a maximum accuracy of 75.00% as shown in Figure 4.5. Second, VGG-16, ResNet-50, and GoogLeNet achieved accuracies of 62.50%, 62.29%, and 63.69%; this signifies that a fixed method based on CNNs could be employed for lung sound classification, but not as effective as an enhancement with the Multi ep-Batch method. Finally, varying hyperparameter (batch and epoch) values during learning (training and testing) phases hold promise as a classification system for lung sounds data.

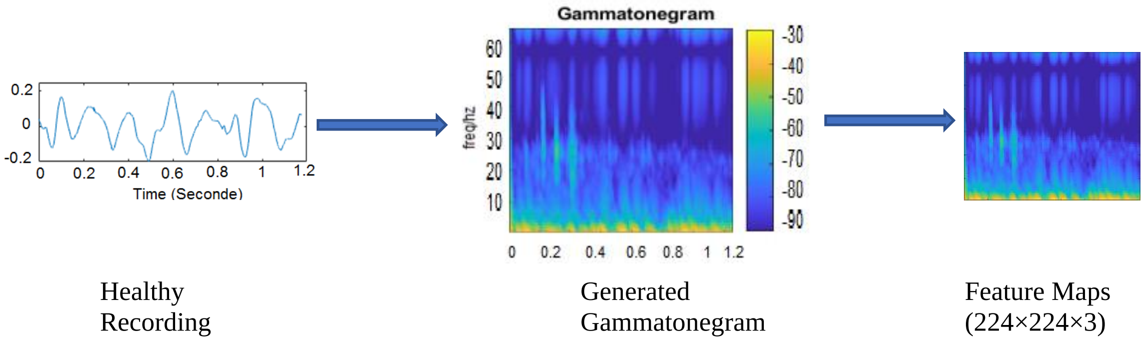
To evaluate a deep learning model, previous studies have implemented the training and testing phases with fixed batch size and epoch hyperparameters values. Each study has used different values of batch size and epochs. These hyperparameters values are then used to find and evaluate the performance of the model. However, the main problem with this method is that because of the variety of the values of the hyperparameters, fine-tuning it could appear as a form of black art. Our approach explores a possible solution to this problem using the Multi ep-Batch learning method, where a set of batch size and epochs values can be automatically incorporated during the training and testing phases. From the achieved results, we found that ResNet-50 and GoogLeNet VGG16 with Multi ep-Batch learning method outperformed these three models with a fixed method and has promising results in both time and accuracy. Our experiments indicate that the use of Gammatonegram and the Multi ep-Batch method improves the classification accuracy of breathing sounds, such as wheezes, crackles and normal sounds, and reduces the training time compared with those obtained without this method. While many studies have conducted lung sound classification, some studies have focused on the ICBHI database, which is used in this study and currently considered the only publicly available exhaustive and challenging database covering, albeit somewhat imbalanced, a wide range of lung sounds and diseases. In [26], using the aforementioned database, the authors proposed a deep CNN-RNN model for respiratory sound classification based on Mel spectrograms and achieved an initial overall classification score of 66.31%. They eventually obtained a score of 71.81% using a model retrained with patient-specific data. The work in [34] for lung sound classification, also using the ICBHI database, used two approaches – the use of a pretrained CNN model to extract features based on SVM as the classifier achieved an accuracy of 65.5%, and the use of a CNN model obtained after transfer learning with fine-tuning spectrogram data yielded an accuracy of 63.09%. While these cannot be directly compared with our work, our results demonstrate that the use of breathing cycle-based gammatonegrams coupled with the

Multi ep-Batch learning method can produce a comparable performance against existing methods. The results from our experiments indicate 1) the Multi ep-Batch method to be an optimized resolution learning method, regardless of data augmentation and 2) the classification accuracy and training time are significantly affected positively by the Multi ep-Batch method.

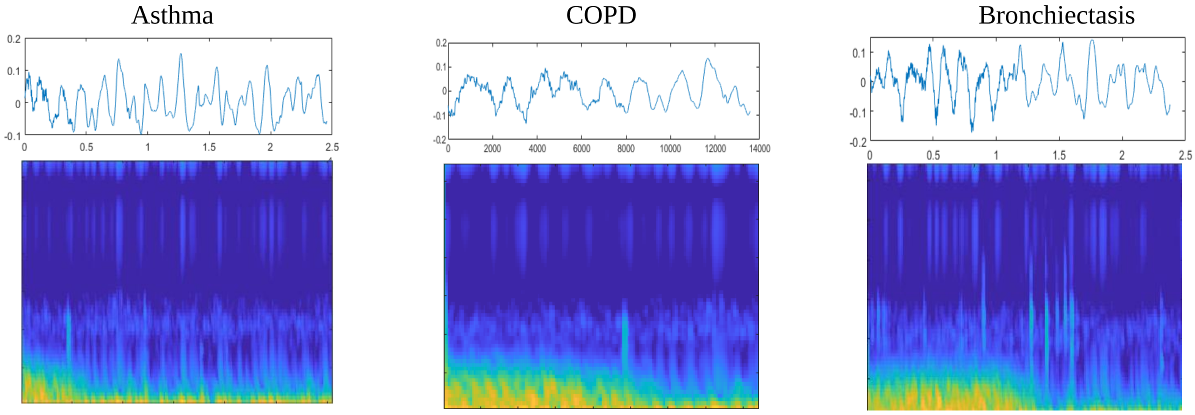
4.2.2. Scenario (ii) – Conditions-based

4.2.2.1. Results

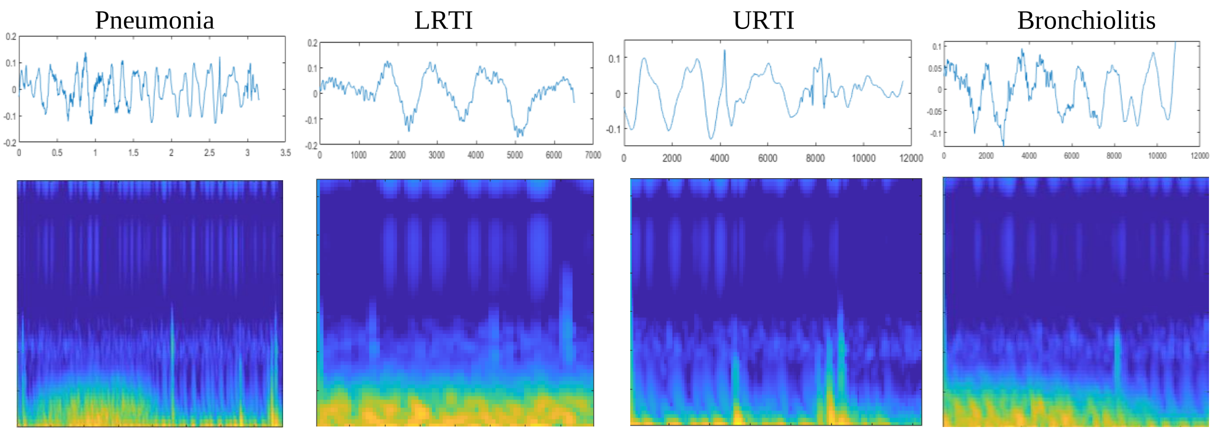
In this scenario, the generated gammatongrams images fed into VGG16, ResNet-50 and AlexNet architectures. Figure 4.6, shows an example of the feature map based on gammatone-grams obtained for pulmonary conditions.



(a) Gammatone-based like image feature map of an audio waveform (Healthy).



(b) Gammatone-based spectrogram-like image feature map of an audio waveform (Chronic).



(c) Gammatone-based spectrogram-like image feature map of an audio waveform (Non-Chronic).

Figure 4.6: Feature maps based-Gammatonegram applied for (a) Healthy, (b) Chronic and (c) Non-Chronic pulmonary conditions.

Expirement	Batch Size	Epoch	Accuracy %	Training Time
A (1) VGG16	32	100	67.97 %	190 min
A (2) VGG16 + **Our Method**	[32, 64, 128, 256, 512]	[2, 12, 22, 32, 42]	**70.31 %**	**62 min**

Table 4.11: VGG16 performance of Gammatonegrams cycles classification.

Table 4.11 shows that by using fixed values for the batch size and number of epochs in experiment (A1), we obtained an accuracy of 67.97% for the classification and a training time of 190 min. We then executed the same process but with the Multi ep-Batch method option turned on in experiment (A2). As detailed in the table, there is an improvement in the obtained accuracy, and a 3 times reduction in training time when compared to the fixed batch-epoch approach (standard). To further scrutinize the classification performance of the VGG16 network with and without the Multi ep-Batch learning method, we generated the corresponding confusion matrix as shown in Figure 4.7.

(a) (b)

Figure 4.7: Confusion matrices for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Table 4.12 summarizes the quality of the network through the performance parameters. From the table, it can be seen that the precision, recall and F1 score do not suffer any severe degradation when the Multi ep-Batch is used as a learning method. The overall values of precision, recall and F1-score in the experiments provide preliminary evidence that the Multi ep-Batch learning method can make the VGG16 network discriminate a multiclass problem. Therefore, it can be inferred that the classification of pulmonary conditions is possible using the proposed Multi ep-Batch learning method in conjunction with the VGG16 network. This system yields a maximum accuracy of 70.31%.

Experiment	**Pulmonary Condition**	**Precision**	**Recall**	**F1-score**
A (1) VGG16	Healthy	0.26	0.29	0.28
A (1) VGG16	Chronic	0.24	0.29	0.26
A (1) VGG16	Non-Chronic	0.51	0.43	0.47
A (2) VGG16 + Multi ep-Batch	Healthy	0.26	0.28	0.27
A (2) VGG16 + Multi ep-Batch	Chronic	0.23	0.21	0.22
A (2) VGG16 + Multi ep-Batch	Non-Chronic	0.52	0.52	0.52

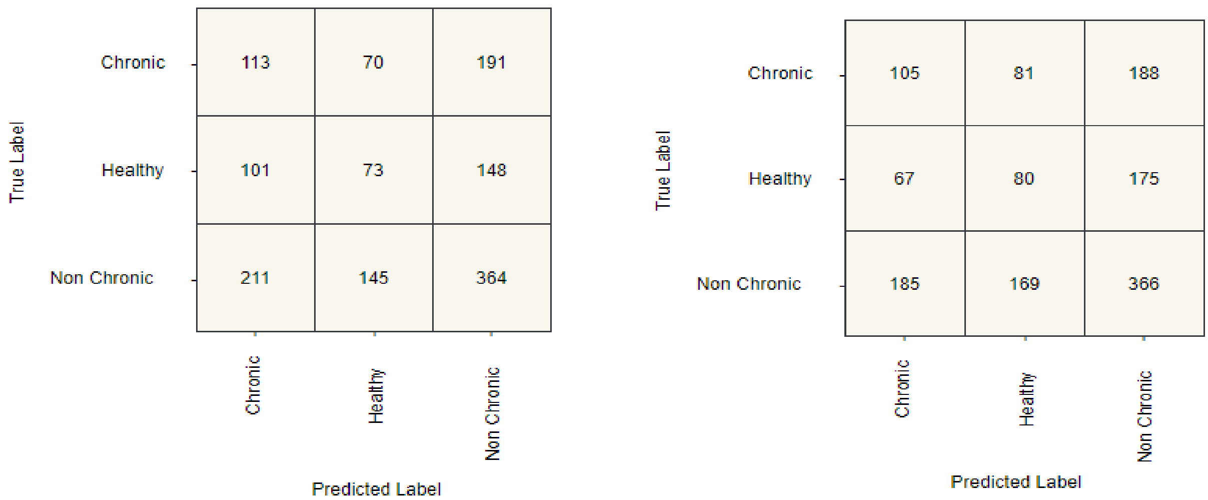
Table 4.12: Precision, Recall and F1-Score comparisons.

To investigate the effect of our approach on a different commonly CNN architecture, we implemented the same experiments as previously to another more complex deep learning network – ResNet-50. The results are shown in Table 4.13. While we observed a 7% reduction in performance in terms of accuracy results in experiment (B1) compared with to the previous network in experiment (A1), the training times were almost comparable. It appears that the VGG16 network performs better than the ResNet-50 architecture for our intended application. This time, when we repeated the experiment with the Multi ep-Batch method turned on in experiment (B2), we noticed a 10% increment in the accuracy rates (70.31%) and similar to the previous network, a 3 times reduction in training time (54 mins). Incidentally, the Multi ep-Batch method appears just as or increasingly advantageous in more complex CNN architectures.

Expirement	Batch Size	Epoch	Accuracy %	Training
B (1) ResNet-50	32	100	60.80 %	178 min
B (2) ResNet-50 + **Our Method**	[32, 64, 128, 256, 512]	[2, 12, 22, 32, 42]	**70.31 %**	**54 min**

Table 4.13: ResNet-50 performance of Gammatonegrams cycles classification.

Figure 4.8 shows the corresponding confusion matrix for the classification results of the ResNet-50 architecture with and without our Multi ep-Batch learning method. For this complex architecture, Table 4.14, shows the resulting performance parameters of ResNet-50 with and without the Multi ep-Batch learning method. Again, we observe no degradation in the precision, recall and F1-score when Multi ep-Batch is used. Similar to VGG16, the ResNet-50 architecture also yields a maximum accuracy of 70.31%.



(a) (b)

Figure 4.8: Confusion matrices for the ResNet-50 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Experiment	Pulmonary Condition	Precision	Recall	F1-score
B (1) ResNet-50	Healthy	0.27	0.30	0.28
B (1) ResNet-50	Chronic	0.25	0.23	0.24
B (1) ResNet-50	Non-Chronic	0.52	0.51	0.51
B (2) ResNet-50 + Multi ep-Batch	Healthy	0.29	0.28	0.29
B (2) ResNet-50 + Multi ep-Batch	Chronic	0.24	0.25	0.25
B (2) ResNet-50 + Multi ep-Batch	Non-Chronic	0.50	0.51	0.51

Table 4.14: Precision, Recall and F1-Score comparisons (ResNet-50).

For further investigate the effect of our Multi ep-Batch learning method on the CNN deep learning architectures, we benchmarked the performance to another network – AlexNet. Table 4.15 shows these results. By using the fixed method, (C1) we obtained an accuracy of 60.16% for the classification and a training time of 140 min. When we performed the same process, but by using our Multi ep-Batch method (C2), there is an improvement in the obtained accuracy as well as a 3 times reduction in training time. Figure 4.9 shows the confusion matrix of the proposed method (with and without the Multi ep-Batch learning method).

Experiment	Batch Size	Epoch	Accuracy %	Training Time
C (1) AlexNet	32	100	60.16 %	140 min
C (2) AlexNet + **Our Method**	[32, 64, 128, 256, 512]	[2, 12, 22, 32, 42]	**64.06 %**	**51 min**

Table 4.15: AlexNet performance of Gammatonegrams cycles classification.

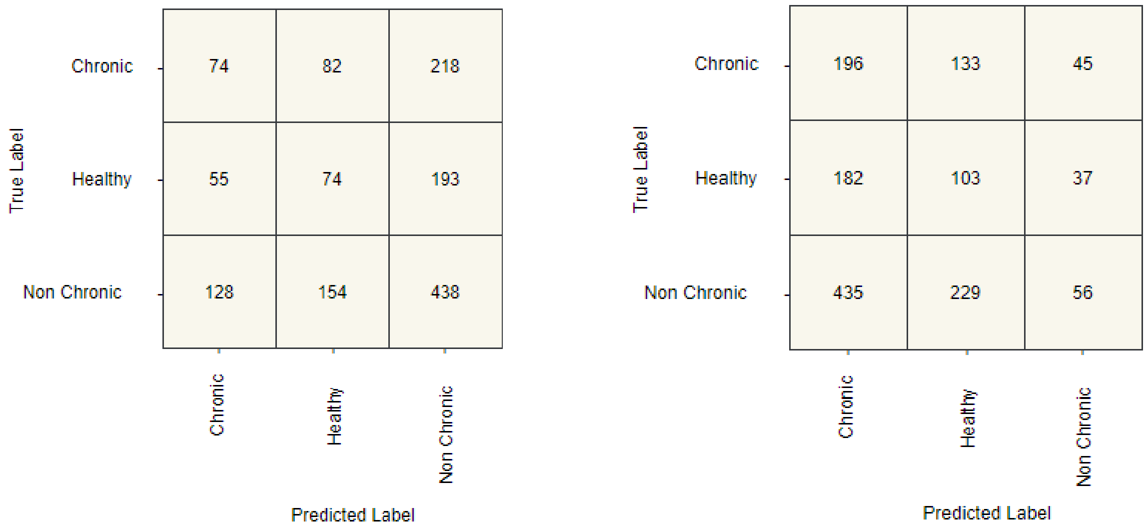


Figure 4.9: Confusion matrices for the AlexNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Experiment	**Pulmonary Condition**	**Precision**	**Recall**	**F1-score**
B (1) AlexNet	Healthy	0.29	0.20	0.23
B (1) AlexNet	Chronic	0.24	0.23	0.23
B (1) AlexNet	Non-Chronic	0.52	0.61	0.56
B (2) AlexNet + Multi ep-Batch	Healthy	0.24	0.52	0.33
B (2) AlexNet + Multi ep-Batch	Chronic	0.22	0.32	0.26
B (2) AlexNet + Multi ep-Batch	Non-Chronic	0.41	0.08	0.13

Table 4.16: Precision, Recall and F1-Score comparisons (AlexNet).

Table 4.16 provides an insight into the quality of the network through the performance parameters – precision, recall and F1-score. We can see that the Multi ep-batch method in conjunction with AlexNet does not suffer from any severe degradation in performance. Based on the overall values of precision, recall and F1-scores, we find that the Multi ep-Batch learning method exhibits the best performance in the multiclass problem. Therefore, it can be inferred that the classification of pulmonary conditions is possible using the proposed Multi ep-Batch learning method in conjunction with the AlexNet network. This system yields a maximum accuracy of 64.06%.

To highlight the overall advantages and understand the effects of the proposed method, we can compare the accuracy and training time for three networks architectures – VGG16, AlexNet and ResNet-50 based on the fixed and multi ep-batch methods. As we can see from Table 4.17, Figure 4.10 and Figure 4.11, superior performance is observed with the proposed multi ep-batch approach in conjunction with these networks. The VGG16 CNN gives an accuracy of 67.97 % for the fixed method while AlexNet and ResNet-50 give 60.16% and 60.80% respectively. In each case, our proposed multi ep-batch method produced positive effects in both accuracy and training time. It can be observed that regardless of the used network models, when they were accelerated using our proposed approach, the accuracy improved and the training time decreased.

Methods	**VGG16**		**AlexNet**		**ResNet-50**	
	Accuracy %	**Training Time**	**Accuracy %**	**Training Time**	**Accuracy %**	**Training Time**
Fixed method (standard)	67.97 %	190 min	60.16 %	140 min	60.80 %	178 min
Multi ep-Batch method	**70.31 %**	**62 min**	**64.06 %**	**51 min**	**70.31 %**	**54 min**

Table 4.17: Comparison results between the proposed and fixed methods using different CNN models.

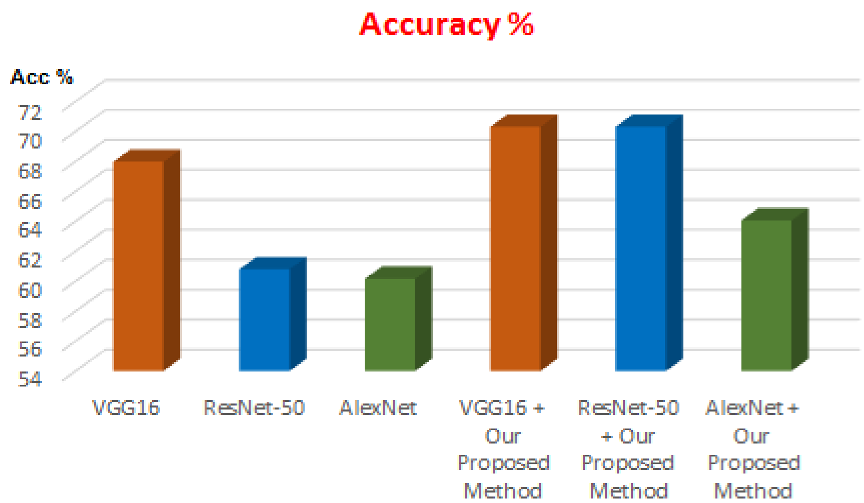


Figure 4.10: Accuracy distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16, ResNet-50 and AlexNet architectures.

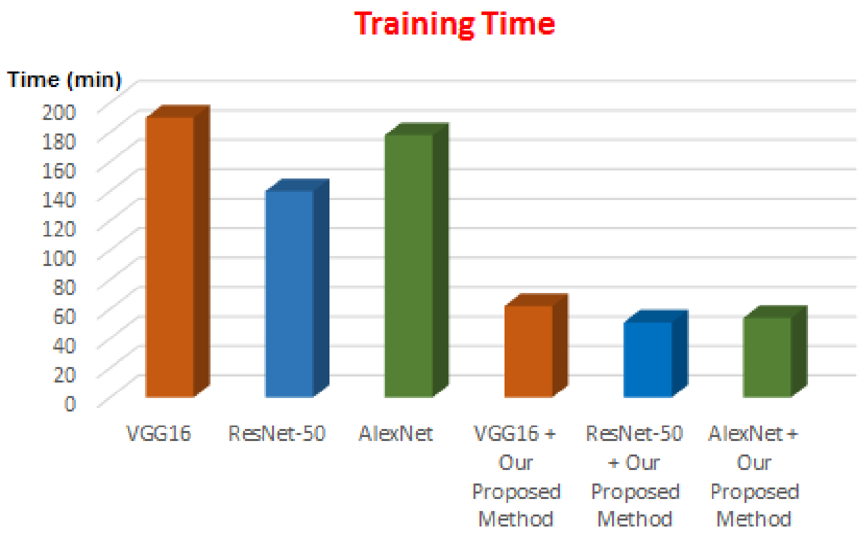


Figure 4.11: Training time distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16, ResNet-50 and AlexNet architectures.

The advantages and limitations of the proposed method are listed as follows.

Advantages:

1) Robustness in terms of its scope with other deep learning models

3) Tunability as per the amount of data

4) Use in any deep learning application as long as the method is incorporated into the training and testing stages

Limitations:

1) Use of 2D images for classification

3) Use with a deep neural network

4.2.2.2. Discussion

From the observed results, we conclude that the accuracy rates and the training times are greatly affected by the Multi ep-Batch method. 3 well know CNN architectures – VGG16, ResNet-50 and AlexNet which are widely used for image classification (Gammatonegrams in our case) have been used to evaluate our new Multi ep-Batch learning method. We found that the performance of VGG16, ResNet-50 and AlexNet with Multi ep-Batch learning method outperformed their standard implementation, and produced promising results in both time and accuracy. In the classification of pulmonary conditions, when the Multi ep-Batch method is used, the training time to develop a model improved (reduction in time) by an average 66% and the accuracy results improved by an average 10% when compared to the standard approach. In recent years, all the studies have implemented the training phase with fixed hyperparameters values, such as batch size and epoch, to evaluate a deep learning model. While each study has used different values of batch size and epochs, the results per fixed hyperparameter values are often used to evaluate and optimize the performance of the deep neural network model. However, the common problem faced by the researchers is that given the variety of hyperparameters and its corresponding range of values, the fine-tuning process more than often appears as a form of black art. Our approach explores a possible solution to this problem using the Multi ep-Batch learning method, where a set of varying batch size and epochs values can be directly incorporated during the training and testing phases.

Several other works in the literature can be used to compare and justify the purpose of this work. Shuvo et al. [33] proposed a CNN architecture to classify respiratory diseases (ternary chronic and six-class) using the same ICBHI 2017 lung sound dataset. They used data preprocessing, empirical mode decomposition (EMD), and continuous wavelet transforms (CWT) as part of their features extraction step. In this study, two of the disease classes in ICBHI – Asthma and LRTI were not considered. The accuracy scores achieved was 99.20% for ternary chronic classification and 99.05% for the six-class pathological classification. A similar investigation was done by García-Ordás et al. [31] who proposed a CNN to classify the respiratory diseases of the ICBHI dataset. While the authors also rejected the Asthma and LRTI diseases, they decided to conduct 2 experiments – 3 class classification (healthy, chronic, and non-chronic) and a 6 class classification (the remaining 6 diseases in the ICBHI dataset). The authors also used several preprocessing steps in their work to facilitate feature extraction. The authors got results up to 0.993 as F-Score in the three-class classification and a 0.990 F-Score in the six-class classification. In [72], Wu and Li proposed a framework combining the random forest classifier and the Empirical Mode Decomposition (EMD) feature extraction technique to classify six pulmonary diseases (healthy, bronchiectasis, bronchiolitis, COPD, pneumonia, and URTI) using the ICBHI database. The authors also ignored the LRTI and asthma of pulmonary diseases. However, their classification system was machine learning-based with preprocessing and features extraction. The classifier's best performance for accuracy was 88%. Finally, a recent work by Perna and Tagarelli in [73] made use of deep neural network architectures, namely recurrent neural networks (RNNs) for the prediction of chronic diseases (COPD, bronchiectasis, asthma) and non-chronic diseases (URTI, LRTI, pneumonia, and bronchiolitis) as a multi-class problem, using the ICBHI dataset. The authors performed three steps of preprocessing to the ICBHI sound segments – frame composition, feature extraction, and feature normalization.

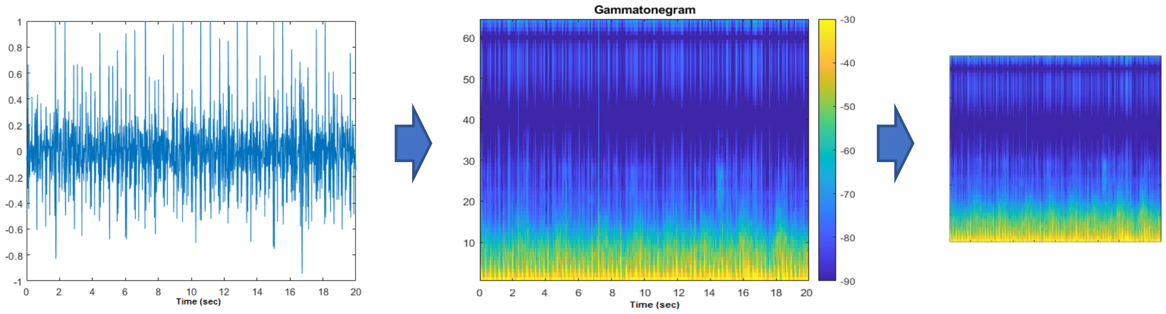
These works when taken together indicate that while preprocessing and feature extraction aids the classification process, it nevertheless negates the original purpose of deep learning. In addition, it is important to subject the findings of past and present studies to critical scrutiny in terms of the benchmarked input physiological data. Therefore to evaluate the robustness of our findings, we tested our proposed method with a challenging medical data such as ICBHI. In our work, audio cycles from three classes (pulmonary condition-based) which contain all diseases in the ICBHI scientific challenge respiratory sound database were used without any preprocessing to observe the robustness of our approach towards varying recording environments and specifications. By doing so, we inherited the following two challenges among others – (1) the identification process automatically includes noise and other artefacts in the input data, and (2) the difficulty of choosing appropriate hyperparameters such as (batch size and epochs). In contrast to the aforementioned works, we have assessed the performance of 3 different CNN networks with both fixed and multiple batch sizes and epochs values for the classification of pulmonary conditions into three classes (healthy, chronic, and non-chronic). In the fixed method, we used a single value for both the batch size and epochs. For the Multi ep-Batch method, we used multiple values of batch size and epochs, which we observed increased the fit of the model during the training phase and subsequently improved the accuracy during the testing phase. The results from our experiments indicate 1) the Multi ep-Batch method to be an optimized learning method, regardless of the deep neural network model, and 2) the classification accuracy and training time are significantly improved by the Multi ep-Batch method. To the best of our knowledge, there is no other work in the literature that has attempted to use all the 8 ICBHI diseases. With the Multi ep-Batch method, we are the first to address the classification problem in the challenging ICBHI dataset which contains data with high noise levels that simulate real-life data collection conditions by using the original files (raw data) without any preprocessing and feature extraction.

Our attempt yielded a maximum accuracy of 70.31% for both VGG16 and ResNet-50 using the Multi ep-Batch learning method for the classification of pulmonary conditions using gammatonegram based input data. It is worthy to note that direct comparison is not applicable due to audio data variations in the employed preprocessing, features extraction and classification techniques. However, the classification results obtained in this work were numerically comparable to the results obtained from the fixed method (standard) used by all other researchers.

4.2.3. Scenario (iii) – Diseases-based

4.2.3.1. Results

Figure 4.12 shows an example of the feature map based-Gammatonegram applied for a healthy sample which saved as JPEG images to fed it as input into our VGG16 and AlexNet-50 architectures, the same process was performed for the rest diseases (Bronchiolitis, Bronchiectasis, COPD, Pneumonia and URTI).



Healthy Recording | Generated Gammatonegram | Feature Maps (224×224×3)

Figure 4.12: Feature map based-Gammatonegram applied for Healthy recording.

From Table 4.18, it can be seen that by using the VGG16 CNN model classifier and data augmentation in conjunction with the Multi ep-Batch method (A2), shows the best accuracy, 81.25% as well as with faster learning time compared to the (A1). However, the difference between the corresponding accuracy obtained by using fixed values for the batch size and number of epochs in experiment (A1) and the multi values for the batch size and number of epochs in experiment (A2) is around 15% improvements for the classification and a training time also decreased in (A2).

Expirement	**Batch Size**	**Epoch**	**Accuracy %**	**Training Time**
A (1) VGG16 + DA	32	100	65.63 %	21 min
A (2) VGG16 + DA + Our Method	**[32, 64, 128]**	**[2, 12, 22]**	**81.25 %**	**14 min**

Table 4.18: VGG16 performance of Gammatonegrams recordings classification without and with Multi ep-Batch method.

Hence, we can say, our proposal method with VGG16 and data augmentation demonstrate that by varying the hyperparameters (batch and epoch) values during learning (tarin and test) phases, that this approach yields comparable learning performance to fixed method. Also, this proposal method could solve the issue of “which appropriate batch and epoch hyperparameters values we need to choose for getting good model performance?”. To further scrutinize the classification performance the corresponding confusion matrices for both the (A1) and (A2) results are illustrated in Figure 4.13.

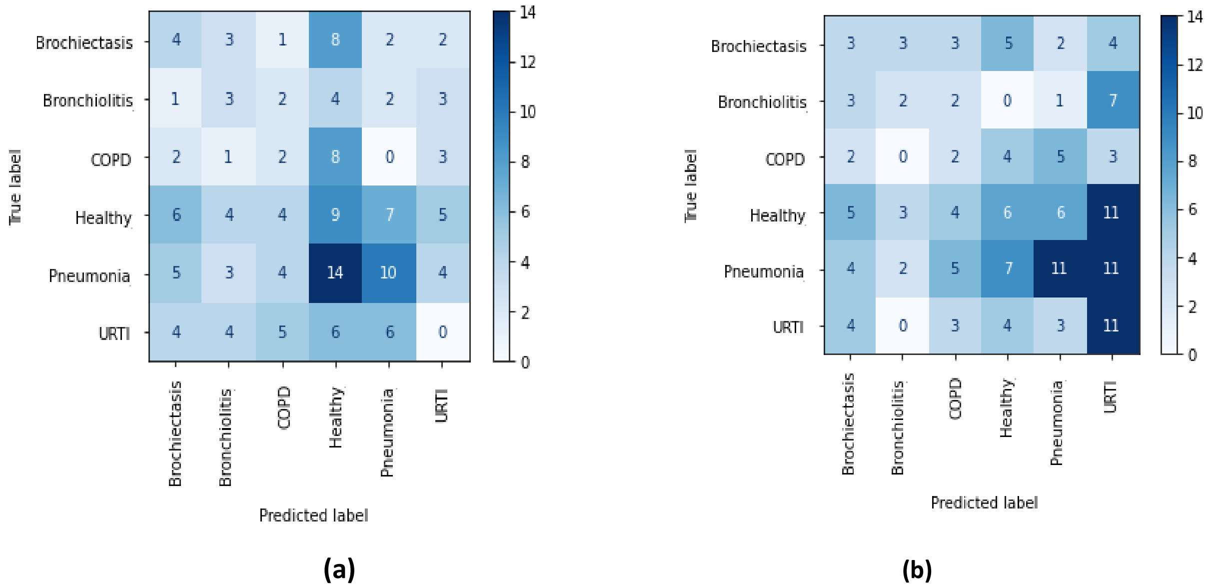


Figure 4.13: Confusion matrixes for the VGG16 and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

As we can see in the Table 4.19 which summarizes the quality of the network through the performance parameters. It can be seen that the precision, recall and F1 score do not suffer any severe degradation when the Multi ep-Batch is used as a learning method. The overall values of precision, recall and F1-score in the experiments provide preliminary evidence that the Multi ep-Batch method has the capability to make the VGG16 network discriminate a multiclass problem. From the same table also, it is observed that in (A1) the precision, recall and F1-score of URTI case the model can’t be predicted any samples, but when we use the VGG16 with our proposed Multi ep-Batch method it make the model to predicted this diseases as well as from the confusion matrix it can observed clearly between Figure 4.13 (a) and (b) in URTI there are 11 samples predicted correctly with our method but no sample detected without it. Therefore, it can be inferred that the classification of pulmonary conditions is possible using the proposed Multi ep-Batch method in conjunction with the VGG16 network. This system yields a maximum accuracy of 81.25%.

Experiment	**Pulmonary Condition**	**Precision**	**Recall**	**F1-score**
A (1) VGG16 + DA	Bronchiectasis	0.18	0.20	0.19
A (1) VGG16 + DA	Bronchiolitis	0.17	0.20	0.18
A (1) VGG16 + DA	COPD	0.11	0.12	0.12
A (1) VGG16+ DA	Healthy	0.18	0.26	0.21
A (1) VGG16 + DA	Pneumonia	0.37	025	0.30
A (1) VGG16 + DA	**URTI**	**0.00**	**0.00**	**0.00**
A (2) VGG16 + DA + Multi ep-Batch	Bronchiectasis	0.14	0.15	0.15
A (2) VGG16 + DA + Multi ep-Batch	Bronchiolitis	0.20	0.13	0.16
A (2) VGG16 + DA + Multi ep-Batch	COPD	0.11	0.12	0.11
A (2) VGG16 + DA + Multi ep-Batch	Healthy	0.23	0.17	0.20
A (2) VGG16 + DA + Multi ep-Batch	Pneumonia	0.39	0.28	0.32
A (2) VGG16 + DA + Multi ep-Batch	**URTI**	**0.23**	**0.44**	**0.31**

Table 4.19: Precision, Recall and F1-Score comparison between two Experiments (VGG16)

As we did in the experiments (A), we implemented the same experiments, in order to check out the effect of the multi ep-batch method on a different CNN network architecture, a deep learning network – AlexNet has been carried out. The results for the experiments can be shown in Table 4.20. As we can see, when comparing the results of the previous network in experiment (A1) with (B1) there is a 10% reduction in performance in terms of accuracy. It appears that the VGG16 network performs better than the AlexNet architecture for our intended application.

This time, when we repeated the experiment with the Multi ep-Batch method turned on in experiment (B2), we noticed a 10% increment in the accuracy rates (65.63%) and also to the previous network with 15%, as well as times reduction in training time.

Expirement	**Batch Size**	**Epoch**	**Accuracy %**	**Training Time**
B (1) AlexNet + DA	32	100	54.69 %	17 min
B (2) AlexNet + DA + **Our Method**	[32, 64, 128]	[2, 12, 22]	**65.63 %**	**14 min**

Table 4.20: AlexNet performance of Gammatonegrams recordings classification without and with Multi ep-Batch method.

In Figure 4.14, the confusion matrix obtained for the classification results of the AlexNet architecture shown with and without our Multi ep-Batch method. The predicted parameters such precision, recall, and F1 score are shown in Table 4.21. Again, From the table it is observed that no degradation in the precision, recall and F1-score when Multi ep-Batch is used. The AlexNet architecture also yields a maximum accuracy of 65.63%. Hence, it can be inferred that the VGG16 and AlexNet using data augmentation in conjunction with the Multi ep-Batch proposed method performs better in the classification of pulmonary diseases.

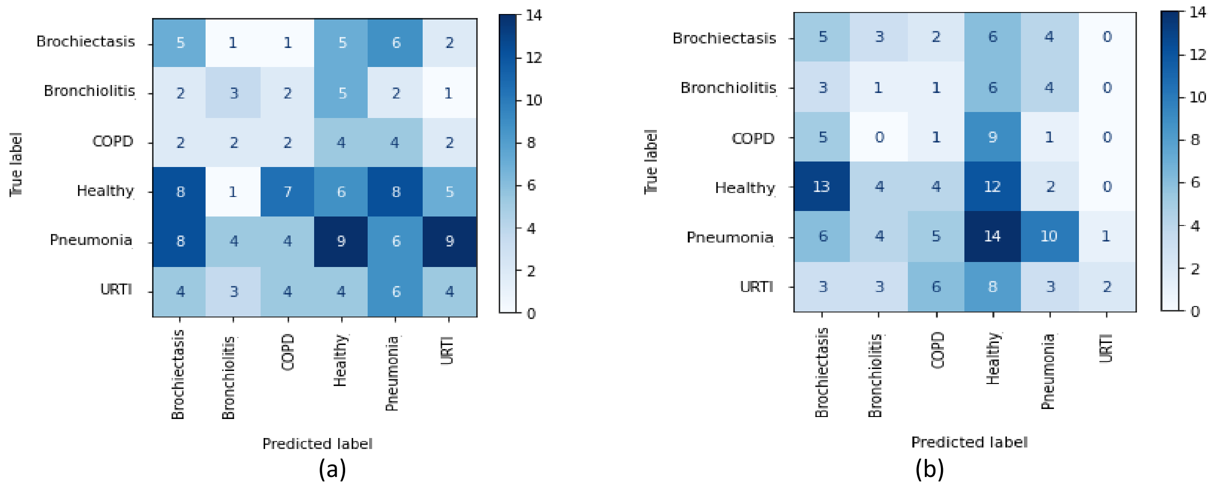


Figure 4.14: Confusion matrixes for the AlexNet and Gammatonegrams classification: (a) without Multi ep-Batch (b) with Multi ep-Batch.

Experiment	Pulmonary Condition	Precision	Recall	F1-score
B (1) AlexNet + DA	Bronchiectasis	0.17	0.25	0.20
B (1) AlexNet + DA	Bronchiolitis	0.21	0.20	0.21
B (1) AlexNet + DA	COPD	0.10	0.12	0.11
B (1) AlexNet + DA	Healthy	0.18	0.17	0.18
B (1) AlexNet + DA	Pneumonia	0.19	015	0.17
B (1) AlexNet + DA	URTI	0.17	0.16	0.17
B (2) AlexNet + DA + Multi ep-Batch	Bronchiectasis	0.14	0.25	0.18
B (2) AlexNet + DA + Multi ep-Batch	Bronchiolitis	0.07	0.07	0.07
B (2) AlexNet + DA + Multi ep-Batch	COPD	0.05	0.06	0.06
B (2) AlexNet + DA + Multi ep-Batch	Healthy	0.22	0.34	0.27
B (2) AlexNet + DA + Multi ep-Batch	Pneumonia	0.42	0.25	0.31
B (2) AlexNet + DA + Multi ep-Batch	URTI	0.67	0.08	0.14

Table 4.21: Precision, Recall and F1-Score comparison between two Experiments

4.2.3.2. Discussion

To highlight the advantages and for further comparison and understanding the effect of the proposed method, the accuracy and training time for two networks architectures VGG16 and

AlexNet is compared based on the fixed and multi ep-batch methods. As we can see from Table 4.22, Figure 4.15 and Figure 4.16 the performance of the proposed multi ep-batch in conjunction with these networks is found to be superior. The VGG16 CNN gives an accuracy of 65.63 % for the fixed method while the accuracy of Alexnet gives 54.69 %. In each case, our proposed multi ep-batch method was affected positively in both accuracy and training time.

Methods	**VGG16**		**AlexNet**	
	Accuracy %	**Training Time**	**Accuracy %**	**Training Time**
Fixed method (standard)	65.63 %	21 min	54.69 %	17 min
Multi ep-Batch method	**81.25 %**	**14 min**	**65.63 %**	**14 min**

Table 4.22: A comparison result between proposed and fixed methods using different CNN models.

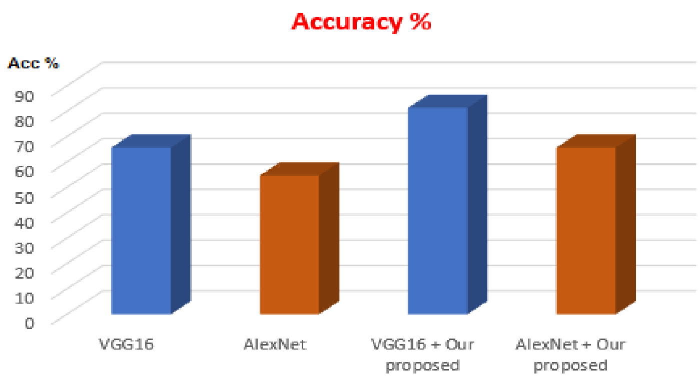


Figure 4.15: Accuracy distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16 and AlexNet architectures.

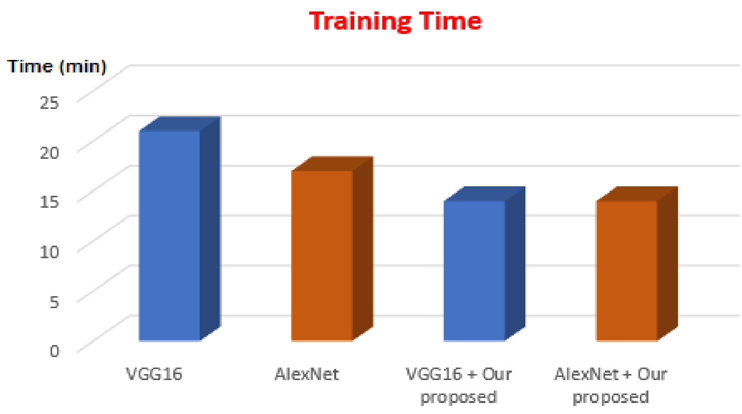


Figure 4.16: Training time distribution for both fixed (standard) and multi ep-batch methods in conjunction with VGG16 and AlexNet architectures.

Figure 14 and Figure 15 illustrate the same information shown in Table 21, but we present it as a column chart for both accuracy and training time of fixed and multi ep-batch methods in conjunction with VGG16 and AlexNet architectures, as an additional view as we thought this could be aid the researchers to read and compare different experiments. It can be observed that regardless of the Network models when we used our proposed approach the accuracy improved as well as the training time also decreased which refer that the Multi ep-Batch could solve the problem of "which hyper-parameters are suitable for a specific deep learning model?". Based on the observed results of this study, we conclude that using the Multi ep-Batch method in conjunction with 2 well know CNN architectures – VGG16 and AlexNet, the accuracy rates and the training times have greatly affected positively. The results indicate that both the performance of VGG16 and AlexNet with Multi ep-Batch learning method outperformed their standard implementation, and achieved a promising result in both time and accuracy. when used Multi ep-Batch method in the classification of pulmonary diseases, the training time to develop a model improved (reduction in time) by an average of 34% and the accuracy results improved by an average of 15% when compared to the standard approach. Several other works in the literature can be used to compare and justify the purpose of this work such as [36, 34, 52, 53]. These works when taken together indicate that, while preprocessing (slicing, resampling, remove artefacts and other nose such heart sounds) and feature extraction aids the classification process, it nevertheless negates the original purpose of deep learning techniques which is to learn from raw data. Furthermore, these processes require a lot of computation time which could appear as a bottleneck in the implementation of a real-time classification system. Besides, the use of raw data without any manipulation mimics practical conditions where much of the challenges remain unresolved to researchers. In contrast to the aforementioned works, we have assessed the performance of 2 different CNN networks with both fixed and multiple batch sizes and epochs values for the classification of pulmonary diseases into six classes healthy, Bronchiolitis, Bronchiectasis, COPD, Pneumonia and URTI. To the best of our knowledge, there is no other work in the literature that used 6 ICBHI diseases, using the original files (raw data) without any sliced cycles (segmentation), preprocessing (resampling, remove artefacts and other nose such heart sounds) and feature extraction techniques. By doing so, we inherited the following two challenges – 1) the identification process which automatically includes noise and other artefacts in the input data, and – 2) the difficulty of choosing appropriate features. Our attempt to work toward these objectives with the raw data yielded a maximum accuracy of 81.25% for both VGG16 and AlexNet using the Multi ep-Batch learning method for the classification of the pulmonary disease by using gammatonegram based input data. According to the results, it could be postulated that multi ep-batch learning method provides benefits not just to the accuracy of a classification system, but also makes the classification system more robust to noisy data. In fact, no need to compare our numerical results with the literature as long as we have developed a new learning system that can be deployed in any deep learning application also in any domain based on deep learning.

Conclusion

In this thesis two principal experimental parts were conducted as follows:

Part 1: The ELM and K-nn classifiers were compared using the Hjorth descriptors (Activity) and Permutation Entropy (PE) features in distinguishing between breath sounds signals with combination these features (Activity, PE). The features extracted were analyzed statistically by calculating a mean and standard deviation to observe the difference between them for each class (Normal bronchial, Wheeze, Crackle, Pleural rub, Stridor). The classification accuracy in multiclass classification case of the ELM and k-nn classifiers is 83.57% and 86.42% respectively, and in binary classification case, the accuracy is 90.71% , 95% respectively. These show that the ability of k-nn in our test conditions (database, methods of analyses the breath signals, and features used) is higher than the ELM classifier in multiclass and binary classification.

Part 2: In this part, we proposed a Multi ep-Batch method, based on lung sounds, for the classification of three types of data – symptoms-based, conditions-based and diseases-based. These sub-data were obtained from the ICBHI scientific dataset consisting of noisy breathing sounds. Their sounds were transformed, from the 1D time domain into the 2D time-frequency domain, as an image using the gammatonegram algorithm. The experimental was divided into three main scenarios, the results obtained point that Multi ep-Batch tends to be the best learning process in different situations. This proposed method was successfully validated using non-pre-processed lung sound signals, which contain other sounds such as heart sounds and other artefacts. Our results showed that the method appears robust for the classification of pulmonary sounds under difficult conditions.

The presented conclusions open new opportunities towards a better learning system (training and testing process), which is still needed for any deep learning applications such as in our case lung sounds classification. We also believe that this learning system could also apply to other applications outside of biomedical topics.

As future work, we would like to perform experiments using the Multi ep-Batch method for a system based on severity disease detection such as cancer for further understanding the generalization power of our new learning system, also we intend to deploy in the embedded system our classification system for real-time application.

REFERENCES

[1] OUD, M. Lung function interpolation by means of neural-network-supported analysis of respiration sounds. *Medical esngineering & physics*, 2003, vol. 25, no 4, p. 309-316.

[2] SKALICKY, D., LOPOT, F., KOUCKY, V., *et al.* Respiratory sounds as a source of information in asthma diagnosis. *Lékař a technika-Clinician and Technology*, 2017, vol. 47, no 2, p. 56-59.

[3] "The Global Impact of Respiratory Disease – Second Edition — CHEST Physician," 2017, [Online]. Available: https://www.mdedge.com/chestphysician/article/140055/society-news/global-impact-respiratory-disease-second-edition.

[4] CRUZ, Alvaro A. *Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach*. World Health Organization, 2007.

[5] WEST, John Burnard. *Respiratory physiology: the essentials*. Lippincott Williams & Wilkins, 2012.

[6] TU, Jiyuan, INTHAVONG, Kiao, et AHMADI, Goodarz. *Computational fluid and particle dynamics in the human respiratory system*. Springer Science & Business Media, 2012.

[7] Custers, E. What is the Respiratory System | Diagram and Function. Retrieved 26-03-2015, 2015

[8] LAVENEZIANA, Pierantonio. Qualitative aspects of exertional dyspnea in patients with restrictive lung disease. *Multidisciplinary respiratory medicine*, 2010, vol. 5, no 3, p. 1-5.

[9] PASTERKAMP, Hans, KRAMAN, Steve S., et WODICKA, George R. Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 1997, vol. 156, no 3, p. 974-987.

[10] SOVIJÄRVI, A., VANDERSCHOOT, J., et EARIS, J. Standardization of computerized respiratory sound analysis. *Crit Care Med*, 1997, vol. 156, p. 974-987.

[11] PASTERKAMP, H. RALE: A computer-assisted instructional package. *Respir Care*, 1990, vol. 35, p. 1006.

[12] ROCHA, B. M., FILOS, Dimitris, MENDES, L., *et al.* A respiratory sound database for the development of automated classification. In : *International Conference on Biomedical and Health Informatics*. Springer, Singapore, 2017. p. 33-37.

[13] GROSS, Volker, HADJILEONTIADIS, L. J., PENZEL, Thomas, *et al.* Multimedia database" marburg respiratory sounds (mars)",". In : *Proc. 25th Annual Int Engineering in Medicine and Biology Society Conf. of the IEEE*. 2003. p. 456-457.

[14] LIN, Bor-Shing et LIN, Bor-Shyh. Automatic wheezing detection using speech recognition technique. *Journal of Medical and Biological Engineering*, 2016, vol. 36, no 4, p. 545-554.

[15] NEILI, Z., FEZARI, M., et REDJATI, A. ELM and K-nn machine learning in classification of Breath sounds signals. *International Journal of Electrical & Computer Engineering (2088-8708)*, 2020, vol. 10.

[16] ORJUELA-CAÑÓN, Alvaro D., GÓMEZ-CAJAS, Diego F., et JIMÉNEZ-MORENO, Robinson. Artificial neural networks for acoustic lung signals classification. In : *Iberoamerican Congress on Pattern Recognition*. Springer, Cham, 2014. p. 214-221.

[17] SERBES, Gorkem, SAKAR, C. Okan, KAHYA, Yasemin P., *et al.* Pulmonary crackle detection using time–frequency and time–scale analysis. *Digital Signal Processing*, 2013, vol. 23, no 3, p. 1012-1021.

[18] JIN, Feng, SATTAR, Farook, et GOH, Daniel YT. New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. *Neurocomputing*, 2014, vol. 123, p. 362-371.

[19] KHODABAKHSHI, Mohammad Bagher et MORADI, Mohammad Hassan. The attractor recurrent neural network based on fuzzy functions: An effective model for the classification of lung abnormalities. *Computers in biology and medicine*, 2017, vol. 84, p. 124-136.

[20] ALTAN, Gokhan, KUTLU, Yakup, PEKMEZCI, Adnan Özhan, *et al.* Deep learning with 3D-second order difference plot on respiratory sounds. *Biomedical Signal Processing and Control*, 2018, vol. 45, p. 58-69.

[21] ALTAN, Gokhan, KUTLU, Yakup, et ALLAHVERDI, Novruz. Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, 2019, vol. 24, no 5, p. 1344-1350.

[22] DEMIR, Fatih, ABDULLAH, Daban Abdulsalam, et SENGUR, Abdulkadir. A new deep CNN model for environmental sound classification. *IEEE Access*, 2020, vol. 8, p. 66529-66537.

[23] CHEN, Hai, YUAN, Xiaochen, PEI, Zhiyuan, *et al.* Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, 2019, vol. 7, p. 32845-32852.

[24] JÁCOME, Cristina, RAVN, Johan, HOLSBØ, Einar, *et al.* Convolutional neural network for breathing phase detection in lung sounds. *Sensors*, 2019, vol. 19, no 8, p. 1798.

[25] BARDOU, Dalal, ZHANG, Kun, et AHMAD, Sayed Mohammad. Lung sounds classification using convolutional neural networks. *Artificial intelligence in medicine*, 2018, vol. 88, p. 58-69.

[26] ACHARYA, Jyotibdha et BASU, Arindam. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE transactions on biomedical circuits and systems*, 2020, vol. 14, no 3, p. 535-544.

[27] SHI, Lukui, DU, Kang, ZHANG, Chaozong, *et al.* Lung sound recognition algorithm based on vggish-bigru. *IEEE Access*, 2019, vol. 7, p. 139438-139449.

[28] AYKANAT, Murat, KILIÇ, Özkan, KURT, Bahar, *et al.* Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017, vol. 2017, no 1, p. 1-9.

[29] DEMIR, Fatih, ISMAEL, Aras Masood, et SENGUR, Abdulkadir. Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure. *IEEE Access*, 2020, vol. 8, p. 105376-105383.

[30] JAYALAKSHMY, S. et SUDHA, Gnanou Florence. Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks. *Artificial intelligence in medicine*, 2020, vol. 103, p. 101809.

[31] GARCÍA-ORDÁS, María Teresa, BENÍTEZ-ANDRADES, José Alberto, GARCÍA-RODRÍGUEZ, Isaías, *et al.* Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. *Sensors*, 2020, vol. 20, no 4, p. 1214.

[32] ROCHA, Bruno Machado, PESSOA, Diogo, MARQUES, Alda, *et al.* Automatic Classification of Adventitious Respiratory Sounds: A (Un) Solved Problem?. *Sensors*, 2021, vol. 21, no 1, p. 57.

[33] SHUVO, Samiul Based, ALI, Shams Nafisa, SWAPNIL, Soham Irtiza, *et al.* A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram. *IEEE Journal of Biomedical and Health Informatics*, 2020.

[34] DEMIR, Fatih, SENGUR, Abdulkadir, et BAJAJ, Varun. Convolutional neural networks based efficient approach for classification of lung diseases. *Health information science and systems*, 2020, vol. 8, no 1, p. 1-8.

[35] PASTERKAMP, H., KRAMAN, S. S., et WODICKA, G. R. State of the art advances beyond the stethoscope. *Am. J. Respir. Crit. Care Med*, 1997, vol. 156, no 3, p. 974-987.

[36] The Auscultation Assistant, 2015, [Online]. Available : https://www.med.ucla.edu/wilkes/inex.htm

[37] D.Arnall, Pulmonary breath sounds, 2015.

[38] R.Wilkins, *et al.*, "Lung Sounds: A Practical Guide with Audio CD .2nd Ed," *Mosby.* February 1996.

[39] SOVIJÄRVI, A., VANDERSCHOOT, J., et EARIS, J. Standardization of computerized respiratory sound analysis. *Crit Care Med*, 1997, vol. 156, p. 974-987.

[40] HUANG, Norden E., SHEN, Zheng, LONG, Steven R., *et al.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 1998, vol. 454, no 1971, p. 903-995.

[41] RILLING, Gabriel, FLANDRIN, Patrick, GONCALVES, Paulo, *et al.* On empirical mode decomposition and its algorithms. In : *IEEE-EURASIP workshop on nonlinear signal and image processing*. NSIP-03, Grado (I), 2003. p. 8-11.

[42] BANDT, Christoph et POMPE, Bernd. Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 2002, vol. 88, no 17, p. 174102.

[43] HUANG, Guang-Bin, ZHU, Qin-Yu, et SIEW, Chee-Kheong. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, vol. 70, no 1-3, p. 489-501.

[44] HUANG, Guang-Bin, WANG, Dian Hui, et LAN, Yuan. Extreme learning machines: a survey. *International journal of machine learning and cybernetics*, 2011, vol. 2, no 2, p. 107-122.

[45] LIAO, Shizhong et FENG, Chang. Meta-ELM: ELM with ELM hidden nodes. *Neurocomputing*, 2014, vol. 128, p. 81-87.

[46] COVER, Thomas et HART, Peter. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 1967, vol. 13, no 1, p. 21-27.

[47] ANDERSEN, Tim et MARTINEZ, Tony. Cross validation and MLP architecture selection. In : *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*. IEEE, 1999. p. 1614-1619.

[48] LITJENS, Geert, KOOI, Thijs, BEJNORDI, Babak Ehteshami, *et al.* A survey on deep learning in medical image analysis. *Medical image analysis*, 2017, vol. 42, p. 60-88.

[49] PRAMONO, Renard Xaviero Adhi, BOWYER, Stuart, et RODRIGUEZ-VILLEGAS, Esther. Automatic adventitious respiratory sound analysis: A systematic review. *PloS one*, 2017, vol. 12, no 5, p. e0177926.

[50] SIMONYAN, Karen et ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[51] HE, Kaiming, ZHANG, Xiangyu, REN, Shaoqing, *et al.* Deep residual learning for image recognition. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.

[52] SZEGEDY, Christian, LIU, Wei, JIA, Yangqing, *et al.* Going deeper with convolutions. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p. 1-9.

[53] ALOM, Md Zahangir, TAHA, Tarek M., YAKOPCIC, Christopher, *et al.* The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018.

[54] PATTERSON, Roy D., ROBINSON, K. E. N., HOLDSWORTH, John, *et al.* Complex sounds and auditory images. In : *Auditory physiology and perception*. Pergamon, 1992. p. 429-446.

[55] GLASBERG, Brian R. et MOORE, Brian CJ. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 1990, vol. 47, no 1-2, p. 103-138.

[56] GREENWOOD, Donald D. Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*, 1961, vol. 33, no 10, p. 1344-1356.

[57] SLANEY, Malcolm. Auditory toolbox. *Interval Research Corporation, Tech. Rep*, 1998, vol. 10, no 1998, p. 1194.

[58] D.P.W. Ellis, “Gammatone-like spectrograms”, 2009, [Online]. Available : http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/

[59] GOYAL, Priya, DOLLÁR, Piotr, GIRSHICK, Ross, *et al.* Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017.

[60] SMITH, Samuel L., KINDERMANS, Pieter-Jan, YING, Chris, *et al.* Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*, 2017.

[61] KESKAR, Nitish Shirish, MUDIGERE, Dheevatsa, NOCEDAL, Jorge, *et al.* On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016.

[62] KERMANY, Daniel, ZHANG, Kang, GOLDBAUM, Michael, *et al.* Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. *Mendeley data*, 2018, vol. 2, no 2.

[63] YOU, Yang, WANG, Yuhui, ZHANG, Huan, *et al.* The Limit of the Batch Size. *arXiv preprint arXiv:2006.08517*, 2020.

[64] LECUN, Yann, *et al.* LeNet-5, convolutional neural networks. *URL: http://yann. lecun. com/exdb/lenet*, 2015, vol. 20, no 5, p. 14.

[65] BOTTOU, Léon. Online learning and stochastic approximations. *On-line learning in neural networks*, 1998, vol. 17, no 9, p. 142.

[66] F. Chollet et al., ‘‘Keras,’’ 2015. [Online]. Available: https://keras.io.

[67] SHORTEN, Connor et KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. *Journal of Big Data*, 2019, vol. 6, no 1, p. 1-48.

[68] CARNEIRO, Tiago, DA NÓBREGA, Raul Victor Medeiros, NEPOMUCENO, Thiago, *et al.* Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, 2018, vol. 6, p. 61677-61685.

[69] PÉREZ, Fernando et GRANGER, Brian E. IPython: a system for interactive scientific computing. *Computing in science & engineering*, 2007, vol. 9, no 3, p. 21-29.

[70] Google, “Colaboratory: frequently asked questions,” 2018, [Access: 6-21-2018]. [Online]. Available: https://research.google.com/colaboratory/faq.html

[71] RIZAL, Achmad, HIDAYAT, Risanuri, et NUGROHO, Hanung Adi. Lung sound classification using empirical mode decomposition and the Hjorth descriptor. *American Journal of Applied Sciences*, 2017, vol. 14, no 1, p. 166-173.

[72] WU, Liqun et LI, Ling. Investigating into segmentation methods for diagnosis of respiratory diseases using adventitious respiratory sounds. In : *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020. p. 768-771.

[73] PERNA, Diego et TAGARELLI, Andrea. Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In : *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019. p. 50-55.

[74] KANDASWAMY, Arumugam, KUMAR, C. Sathish, RAMANATHAN, Rm Pl, et al. Neural classification of lung sounds using wavelet coefficients. *Computers in biology and medicine*, 2004, vol. 34, no 6, p. 523-537.

[75] PASTERKAMP, Hans, KRAMAN, Steve S., et WODICKA, George R. Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 1997, vol. 156, no 3, p. 974-987.

[76] WELSBY, P. D., PARRY, G., et SMITH, D. The stethoscope: some preliminary investigations. *Postgraduate medical journal*, 2003, vol. 79, no 938, p. 695-698.

[77] CHAUHAN, Sunita, WANG, Ping, LIM, Chu Sing, et al. A computer-aided MFCC-based HMM system for automatic auscultation. *Computers in biology and medicine*, 2008, vol. 38, no 2, p. 221-233.

[78] NEILI, Zakaria, FEZARI, Mohamed, et ABDEGHANI, REDJATI. Analysis of acoustic parameters from respiratory signal in copd and pneumonia patients. In : 2018 International Conference on Signal, Image, Vision and their Applications (SIVA). IEEE, 2018. p. 1-4.

[79] PENG, Yaohao et NAGATA, Mateus Hiro. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos, Solitons & Fractals*, 2020, vol. 139, p. 110055.

[80] FARHAT, Hanan, SAKR, George E., et KILANY, Rima. Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Machine vision and applications*, 2020, vol. 31, no 6, p. 1-42.
[81] GOEL, Tripti, MURUGAN, R., MIRJALILI, Seyedali, *et al.* OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. *Applied Intelligence*, 2021, vol. 51, no 3, p. 1351-1366.

[82] NOUR, Majid, CÖMERT, Zafer, et POLAT, Kemal. A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Applied Soft Computing*, 2020, vol. 97, p. 106580.

[83] EZZAT, Dalia, HASSANIEN, Aboul Ella, et ELLA, Hassan Aboul. An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. *Applied Soft Computing*, 2021, vol. 98, p. 106742.

[84] LALMUANAWMA, Samuel, HUSSAIN, Jamal, et CHHAKCHHUAK, Lalrinfela. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals*, 2020, vol. 139, p. 110059.

[85] TORAMAN, Suat, ALAKUS, Talha Burak, et TURKOGLU, Ibrahim. Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. *Chaos, Solitons & Fractals*, 2020, vol. 140, p. 110122.

[86] VANNUCCINI, L., EARIS, J. E., HELISTO, P., et al. Capturing and preprocessing of respiratory sounds. *European Respiratory Review*, 2000, vol. 10, no 77, p. 616-620.
[87] FIZ, José A., JANÉ, Raimon, HOMS, Antonio, et al. Detection of wheezing during maximal forced exhalation in patients with obstructed airways. *Chest*, 2002, vol. 122, no 1, p. 186-191.

[88] FIZ, José Antonio, JANÉ, Raimon, SALVATELLA, David, et al. Analysis of tracheal sounds during forced exhalation in asthma patients and normal subjects: bronchodilator response effect. *Chest*, 1999, vol. 116, no 3, p. 633-638.

[89] HOMS-CORBERA, Antoni, FIZ, José Antonio, MORERA, José, et al. Time-frequency detection and analysis of wheezes during forced exhalation. *IEEE Transactions on Biomedical Engineering*, 2004, vol. 51, no 1, p. 182-186.

[90] TAPLIDOU, Styliani A., HADJILEONTIADIS, Leontios J., KITSAS, Ilias K., et al. On applying continuous wavelet transform in wheeze analysis. In : The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2004. p. 3832-3835.

[91] BAHOURA, Mohammed. Analyse des signaux acoustiques respiratoires: contribution à la detection automatique des sibilants par paquets d'ondelettes. 1999. Thèse de doctorat. Rouen.

[92] KRAMAN, Steve S., WODICKA, George R., OH, Yuns, *et al.* Measurement of respiratory acoustic signals: effect of microphone air cavity width, shape, and venting. *Chest*, 1995, vol. 108, no 4, p. 1004-1008.

[93] HUSSEIN, M. et RICHARD, S. Acoustic detection of respiratory conditions. *United States Patent US6443907*, 2002.

[94] CHEETHAM, B. M. G., CHARBONNEAU, G., GIORDANO, A., *et al.* Digitization of data for respiratory sound recordings. *European Respiratory Review*, 2000, vol. 10, no 77, p. 621-624.

LIST OF PUBLICATIONS

The publications and communications relating to this thesis are as follows:

Publications:

- NEILI, Z., FEZARI, M., et REDJATI, A. ELM and K-nn machine learning in classification of Breath sounds signals. *International Journal of Electrical & Computer Engineering (2088-8708)*, 2020, vol. 10.

Communications:

International:

- NEILI, Zakaria, FEZARI, Mohamed, et ABDEGHANI, REDJATI. Analysis of acoustic parameters from respiratory signal in copd and pneumonia patients. In : 2018 International Conference on Signal, Image, Vision and their Applications (SIVA). IEEE, 2018. p. 1-4.
- NEILI, Zakaria, FEZARI, Mohamed, et ABDEGHANI, REDJATI. Classification of breath sounds signals using K-nn machine learning based on energy and entropy features. In: 2019 International Conference on Embedded Systems in Telecommunications and Instrumentation (ICESTI). Annaba, Algeria, October, 2019.

National:

- NEILI, Zakaria, FEZARI, Mohamed, et ABDEGHANI, REDJATI. Classification of Breath Sound Using Machine Learning Algorithms. In: 2019 Algerian Thematic School on Signal Processing & its Applications (ATSSPA). Annaba, Algeria, Jun, 2019.