**REPUBLIC OF TURKEY**
**ERCIYES UNIVERSITY**
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**
**DEPARTMENT OF BIOMEDICAL ENGINEERING**

# ACCURATE CLASSIFICATION OF HEART SOUND SIGNALS FOR CARDIOVASCULAR DISEASE DIAGNOSIS USING WAVELET ANALYSIS WITH MACHINE LEARNING AND DEEP LEARNING METHODOLOGIES

**Prepared by**
**Afrah Elfatih Farah Malik**

**Supervisor**
**Prof. Dr. Mehmet Emin YÜKSEL**

**M.Sc. Thesis**

**August 2020**
**KAYSERI**

**REPUBLIC OF TURKEY**
**ERCIYES UNIVERSITY**
**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**
**DEPARTMENT OF BIOMEDICAL ENGINEERING**

# ACCURATE CLASSIFICATION OF HEART SOUND SIGNALS FOR CARDIOVASCULAR DISEASE DIAGNOSIS USING WAVELET ANALYSIS WITH MACHINE LEARNING AND DEEP LEARNING METHODOLOGIES

**Prepared by**
**Afrah Elfatih Farah Malik**

**Supervisor**
**Prof. Dr. Mehmet Emin YÜKSEL**

**M.Sc. Thesis**

**August 2020**
**KAYSERI**

## COMPLIANCE WITH SCIENTIFIC ETHICS

I hereby declare that all information in this document has been obtained and presented in accordance with the academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Afrah Elfatih Farah Malik

Signature: 

## COMPLIANCE WITH DIRECTIVES

The study named **"Accurate Classification Of Heart Sound Signals For Cardiovascular Disease Diagnosis Using Wavelet Analysis With Machine Learning And Deep Learning Methodologies"** presented as an **M.Sc.** thesis has been prepared in accordance with Erciyes University Graduate Thesis Proposal and Thesis Writing Directives.

Prepared by
Afrah Elfatih Farah Malik

Supervisor
Prof. Dr. Mehmet Emin YÜKSEL

Head, Department of Biomedical Engineering
Prof. Dr. Mahmut TOKMAKÇI

The study named **"Accurate Classification Of Heart Sound Signals For Cardiovascular Disease Diagnosis Using Wavelet Analysis With Machine Learning And Deep Learning Methodologies"** prepared by **Afrah Elfatih Farah Malik** under the supervision of **Prof. Dr. Mehmet Emin YÜKSEL** was accepted by the jury as an **M.Sc.** thesis in Erciyes University, Graduate School of Natural and Applied Science, Department of Biomedical Engineering .

../../2020

**JURY :**

Supervisor : Prof. Dr. Mehmet Emin YÜKSEL

Member : Prof. Dr. Mahmut TOKMAKÇI

Member : Doç. Dr. Ali ÖZEN

APPROVAL :

The acceptance of this thesis has been approved by the Institute Board with the decision number ..... and the date of .....

**..... / ..... /2020**
**Prof. Dr. Mehmet AKKURT**
**Director of the Institute**

## ACKNOWLEDGEMENT

Firstly, I would like to express my gratitude and sincere thanks to my supervisor, Prof. Dr. Mehmet Emin YÜKSEL, who gave me the opportunity to work as a part of his research group and continuously supported me through this thesis work. His expertise, ideas, advice and encouragement, greatly influenced my M.S.c graduate studies.

I also wish to extend special thanks to the Presidency For Turks Abroad And Related Communities, the coordinators of the Turkey Scholarships program for the opportunity to pursue a Master's degree in Turkey.

I am indebted to my colleagues Derek Kweku DEGBEDZUI and Sezin BARIN for their encouragement, support and criticism throughout my Master's degree studies.

Additionally, I would like to thank my family and friends who have always been there for me.

Afrah Elfatih Farah Malik
Kayseri, 2020

# MAKINE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİ İLE DALGACIK ANALİZİ KULLANILARAK KALP SES SİNYALLERİNİN KARDİYOVASKÜLER HASTALIK TANI AMAÇLI DOĞRU SINIFLANDIRILMASI

**Afrah Elfatih Farah Malik**
**Erciyes Üniversitesi, Fen Bilimleri Enstitüsü**
**Yüksek Lisans Tezi, Ağustos 2020**
**Danışman: Prof. Dr. Mehmet Emin YÜKSEL**

## ÖZET

Kardiyovasküler hastalıkların (KVH) erken tanısı, bu hastalıklardan kaynaklanan olumsuzlukların ve olası ölümlerin azaltılması açısından çok önemlidir. Kalp sesleri kalbin durumu ile ilgili değerli bilgiler içerir. Bu sebeple kalbin durumunun değerlendirilmesine yönelik ilk işlem genellikle uzman bir hekimin stetoskop yardımıyla kalp seslerini dinlediği ve bu sesleri yorumlayarak tanı koyduğu oskültasyon işlemidir. Ancak bu işlemin başarımı doğal olarak hekimin tecrübesine bağlıdır ki bu durum tanının öznel olmasına yol açar.

Makine öğrenmesi ve derin öğrenme yöntemleri kalp sesi analizi için nesnel, daha düşük maliyetli ve invasiv olmayan alternatif yöntemler sunmaktadır. Bu tez çalışmasında KVHların otomatik tanısına yönelik makine öğrenmesi ve derin öğrenme tabanlı yeni ve etkin yöntemler önerilmektedir.

Önerilen yöntemler temelde kalp seslerinin dalgacık skalogramlarının makine öğrenmesi ve derin öğrenme tabanlı yaklaşımlarla sınıflandırılmasına dayanmaktadır. Kalp sesi kayıtlarının skalogramlarını elde etmek için sürekli dalgacık dönüşümü seçilmiştir. Bunu takiben skalogram görüntüleri bir evrişimsel sinir ağı modelini ve dört farklı önceden eğitilmiş derin ağ modelini eğitmek için kullanılmıştır. Buna ek olarak, nesne sezmede popüler bir yöntem olan yönlü gradyan histogramı (HOG) skalogram görüntülerinden nitelik çıkarmak için kullanılmıştır. Çıkarılan HOG nitelikleri sınıflandırma deneyleri için makine öğrenmesi tabanlı modellere beslenmiştir.

Önerilen yöntemlerin performansları hem normal hem de dört farklı kalp patolojisine ait kalp sesleri içeren bir veri seti kullanılarak değerlendirilmiştir. Önerilen yöntemlerin her ikisinin de literatürde mevcut diğer otomatik kalp sesi sınıflandırma yöntemlerinden daha iyi başarım sunduğu görülmüştür.

**Anahtar Kelimeler:** Kalp sesleri, makine öğrenmesi, derin öğrenme, dalgacık analizi

# ACCURATE CLASSIFICATION OF HEART SOUND SIGNALS FOR CARDIOVASCULAR DISEASE DIAGNOSIS USING WAVELET ANALYSIS WITH MACHINE LEARNING AND DEEP LEARNING METHODOLOGIES

**Afrah Elfatih Farah Malik**
**Erciyes University, Graduate School of Natural and Applied Sciences**
**Master of Science Thesis, August 2020**
**Supervisor: Prof. Dr. Mehmet Emin YÜKSEL**

## ABSTRACT

The early detection of cardiovascular diseases (CVDs) is very important to reduce the occurrences of deaths or mitigate their implications. Heart sounds contain valuable information about the condition of the heart. For this reason, the first step in the evaluation of the condition of the heart is usually the practice of auscultation, in which a physician uses a stethoscope to listen to the heart sounds and interprets them to reach a medical diagnosis. However, the success of this process naturally depends on the expertise of the physician, which makes the diagnosis subjective.

Machine learning and deep learning methods offer objective, lower cost, and non-invasive alternatives for heart sound analysis. This thesis study proposes novel and efficient methods for automatic CVDs diagnosis based on machine learning and deep learning methodologies. The continuous wavelet transform is selected to obtain the scalogram of heart sound recordings. The scalogram images are then used to train a convolutional neural network model and four different pre-trained deep network models. Furthermore, the histogram of oriented gradient (HOG), a popular method for object detection, is employed to extract features from the scalogram images. Subsequently, the HOG features are fed to machine learning models for classification experiments.

The performances of the proposed methods are evaluated using a dataset that consists of normal heart sounds and heart sounds representing four different heart pathologies. It is seen that both of the proposed methods offer superior performance over other methods available in the literature for automatic heart sound classification.

**Keywords:** Heart sounds, Machine learning, Deep learning, Wavelet analysis.

**CONTENTS**

**ACCURATE CLASSIFICATION OF HEART SOUND SIGNALS FOR CARDIOVASCULAR DISEASE DIAGNOSIS USING WAVELET ANALYSIS WITH MACHINE LEARNING AND DEEP LEARNING METHODOLOGIES**

**CHAPTER 2**
**Method**



**CHAPTER 3**
**Results**

## ABBREVIATIONS

| | | |
|---|---|---|
| WHO | : | World Health Organization |
| ML | : | Machine learning |
| DL | : | Deep learning |
| 1D | : | 1-dimensional |
| 2D | : | 2-dimensional |
| S1 | : | First heart sound |
| S2 | : | Second heart sound |
| ECG | : | Electrocardiograph |
| DWT | : | Discrete wavelet transform |
| MLP | : | Multilayer perceptron |
| KNN | : | K nearest neighbors |
| SVM | : | Support vector machine |
| CVDs | : | Cardiovascular diseases |
| HS | : | Heart sound |
| AS | : | Aortic stenosis |
| MS | : | Mitral stenosis |
| MR | : | Mitral regurgitation |
| MVP | : | Mitral valve prolapse |
| N | : | Normal heart sound |
| CNN | : | Convolutional neural network |
| LSTM | : | Long short term memory |
| EMD | : | Empirical model decomposition |
| IMF | : | Intrinsic mode function |
| ANN | : | Artificial neural network |
| PCG | : | Phonocardiogram |
| CWT | : | Continuous wavelet transform |
| DWT | : | Discrete wavelet transform |
| HOG | : | Histogram of oriented gradient |
| TP | : | True positive |
| TN | : | True negative |

| | | |
|---|---|---|
| FP | : | False positive |
| FN | : | False negative |
| ILSVRC | : | ImageNet Large Scale Visual Recognition Challenge |

# LIST OF TABLES

# LIST OF FIGURES

# INTRODUCTION

Cardiovascular diseases (CVDs) account for the highest proportion of deaths worldwide [1]. Early diagnosis of heart disorders can reduce the mortality rate due to CVDs. There have been advancements in the diagnosis of CVDs. However, the death rates due to CVDs still shows an increasing trend. Therefore, it is urgent to address this problem. Normally, a digital stethoscope is utilized to record heart sounds, and the recorded signals are analyzed to make a diagnosis. Generally, an automated cardiac auscultation system consists of three steps: segmentation step, feature extraction step and classification step. In the segmentation stage, the signal is divided into representative segments. In the feature extraction stage, the features of the segments are computed using various methods. The obtained features are then fed into the classification stage, where a suitable classifier is selected to make a decision.

This study mainly focuses on computing the wavelet scalogram from the heart sound and the classification of the scalograms and histogram of oriented gradient features using deep learning and machine learning classifiers. Binary classification and multi-class classification are performed here, and the performances of the classifiers are validated using different performance metrics.

## Thesis Organization

This thesis is organized as follows:

- Chapter 1 presents background information, the definition of normal heart sounds and murmurs and valvular heart diseases. The available methods of cardiovascular disease diagnosis are discussed. The chapter also presents a review of the studies that investigated cardiovascular diseases diagnosis using ML and DL methods existing in the literature. It concludes by discussing the problems of current solutions and thesis objective.

- Chapter 2 gives detailed information about the method used in this study. It commences by discussing how the data are acquired and segmented. It then continues to present the computation of scalogram images and the extraction of the histogram of oriented gradient features. Following this, information regarding the employed machine learning and deep learning classifiers is present. In addition, this chapter discusses the validation method and the performance metrics utilized in this study.
- Chapter 3 reports the experimental results of DL models with scalogram images and HOG features with ML models.
- Chapter 4 discusses the experimental results and concludes with prospective future works of the thesis project.

# 1. CHAPTER

# General Information and Literature Review

## 1.1. Background

Cardiovascular diseases (CVDs) are a group of diseases that affect both the heart and blood vessels. Valvular heart diseases are a subdivision of CVDs involving heart valves. These diseases affect the four valves of the heart. These kinds of disorders could be diagnosed using heart sounds.

### 1.1.1. Heart Sounds

Heart sounds (HS) are sounds that are generated by the closure of the heart valves. Normally, two sounds are produced by the human heart: the first heart sound (S1) and the second heart sound (S2). Figure 1.1 depicts a normal heart sound. Aside from S1 and S2, there are other sounds including murmurs. The two groups of the murmur are normal and abnormal murmurs. While normal murmurs occur during non-resting conditions of healthy individuals, abnormal murmurs are an indicator of pathological heart condition generated from abnormal blood flow through the heart valves and vessels. The diagnosis of abnormal murmurs depends on their intensity, duration in the cardiac cycle, and the location which provides the best audible murmur [2]. The four types of CVDs addressed in this thesis project are aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR) and mitral valve prolapse (MVP). Further detailed information is presented in section 1.1.1.1.

#### 1.1.1.1. Addressed Cardiovascular Diseases

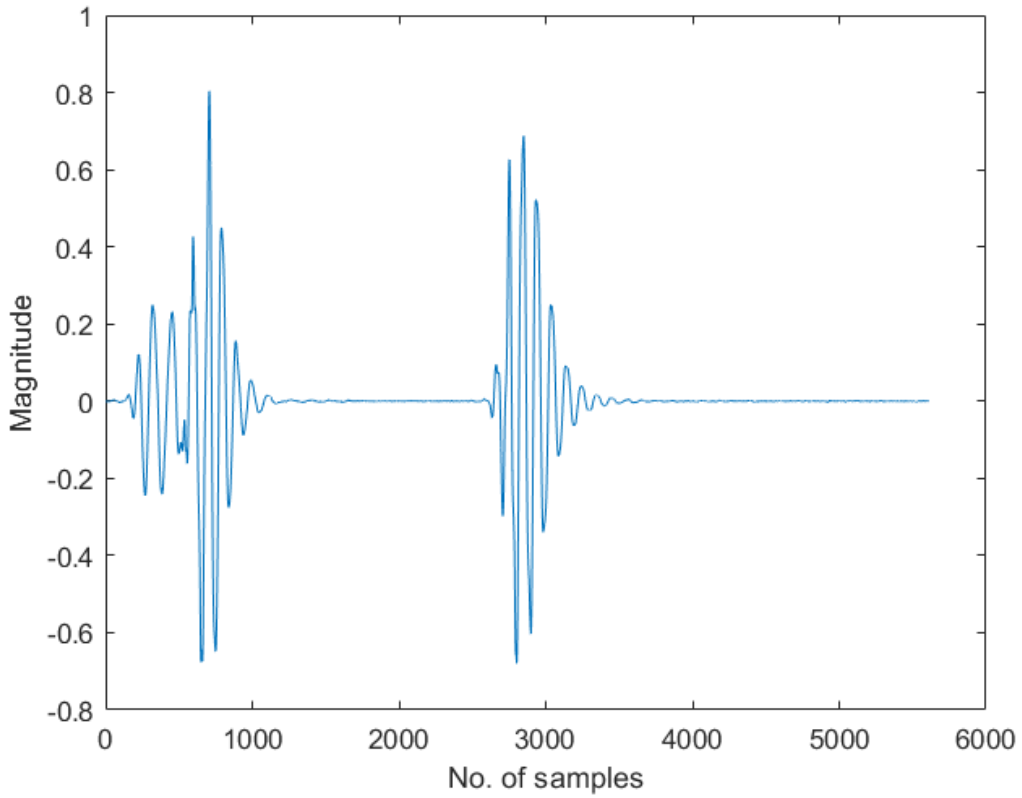



Figure 1.1. A phonocardiogram of normal heart sound.

#### 1.1.1.1.1. Aortic Stenosis

Aortic stenosis (AS) is the narrowing of the aortic valve (the valve located at the exit of the left ventricle, and permits a unidirectional flow of blood from the heart to the rest of the body). This condition restricts blood flow through the valve. AS is a degenerative disease, such that the condition worsens over time. At the early stages of the disease, the left ventricle tries to compensate and maintain a regular cardiac output by increasing the pressure and wall thickness, but at the advanced stages, this adaptation process of the left ventricle causes heart failure [3]. The murmur associated with AS is presented in Figure 1.2

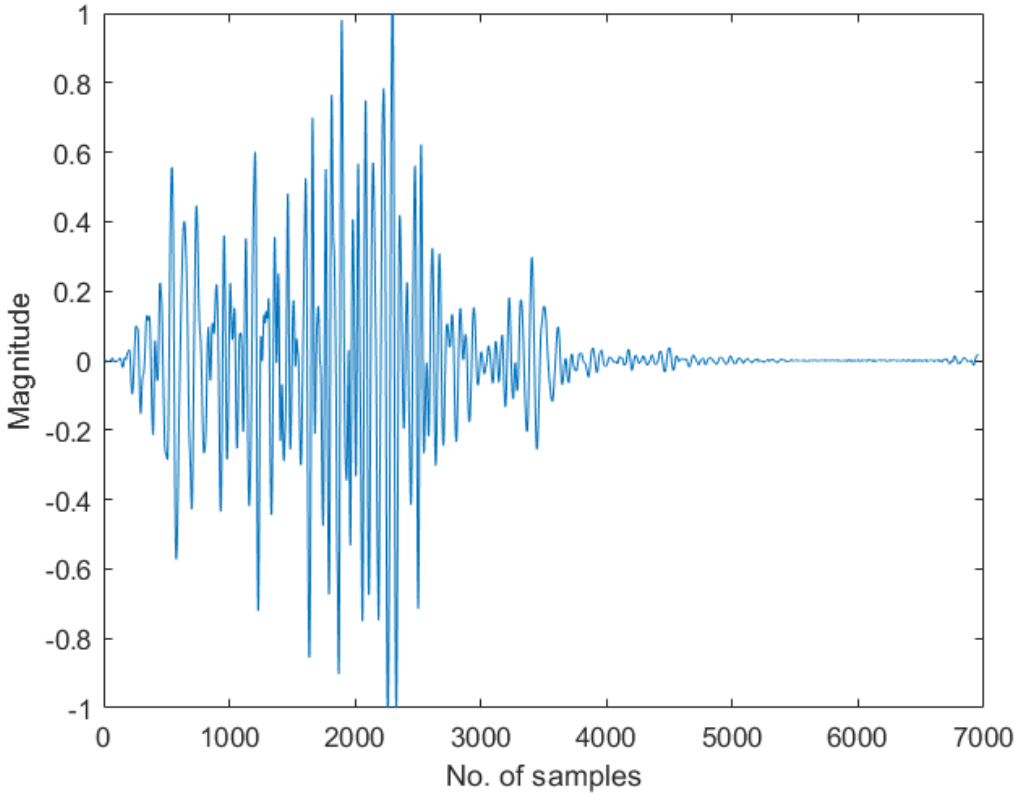



Figure 1.2. A phonocardiogram of aortic stenosis.

#### 1.1.1.1.2. Mitral Stenosis

Mitral stenosis (MS) is a mitral valve disease, where the mitral opening narrows and obstructs blood flow from the left atrium to the left ventricle. As this condition deteriorates over time, there is a reduction in the ventricular output resulting in a condition similar to ventricular failure [4]. Figure 1.3 shows the murmur associated with MS.

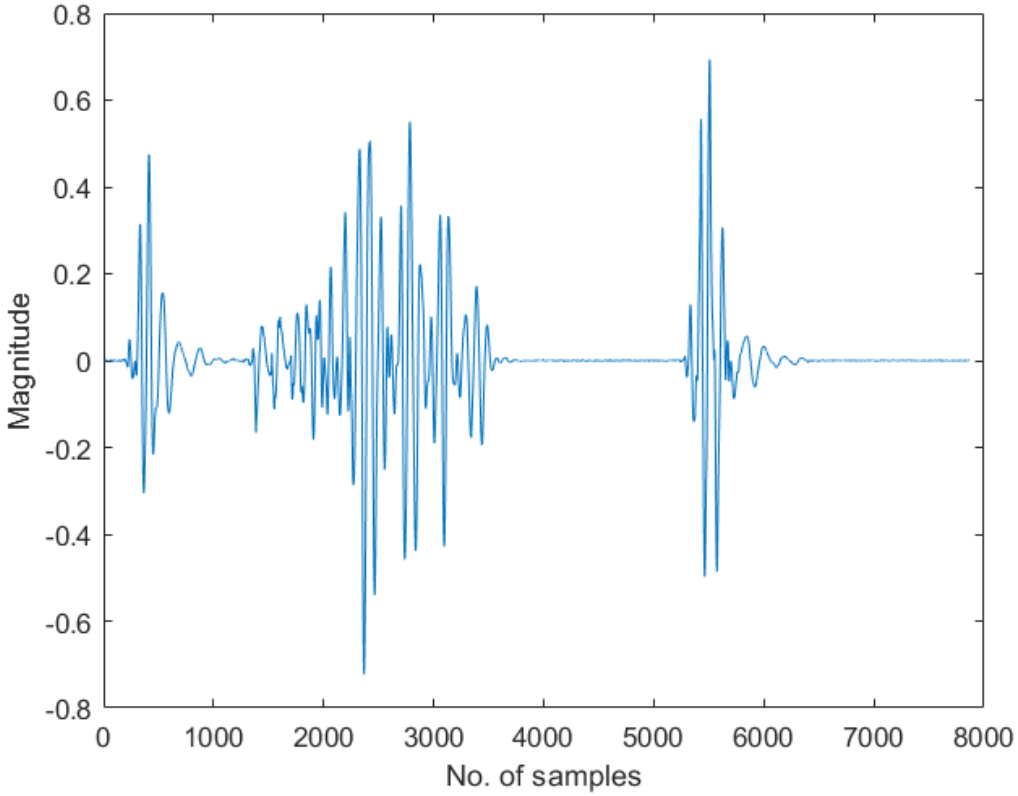



Figure 1.3. A phonocardiogram of mitral stenosis.

#### 1.1.1.1.3. Mitral Regurgitation

Mitral regurgitation (MR) is a valvular heart disease that results from the disruption in any parts of the mitral valve [5]. During systole, there is an abnormal regurgitation of blood flow from the left ventricle to the left atrium [5]. MR is characterized by a holosystolic murmur that is best heard at the apex. It begins at S1 and continues to S2. This kind of murmur is depicted in Figure 1.4.

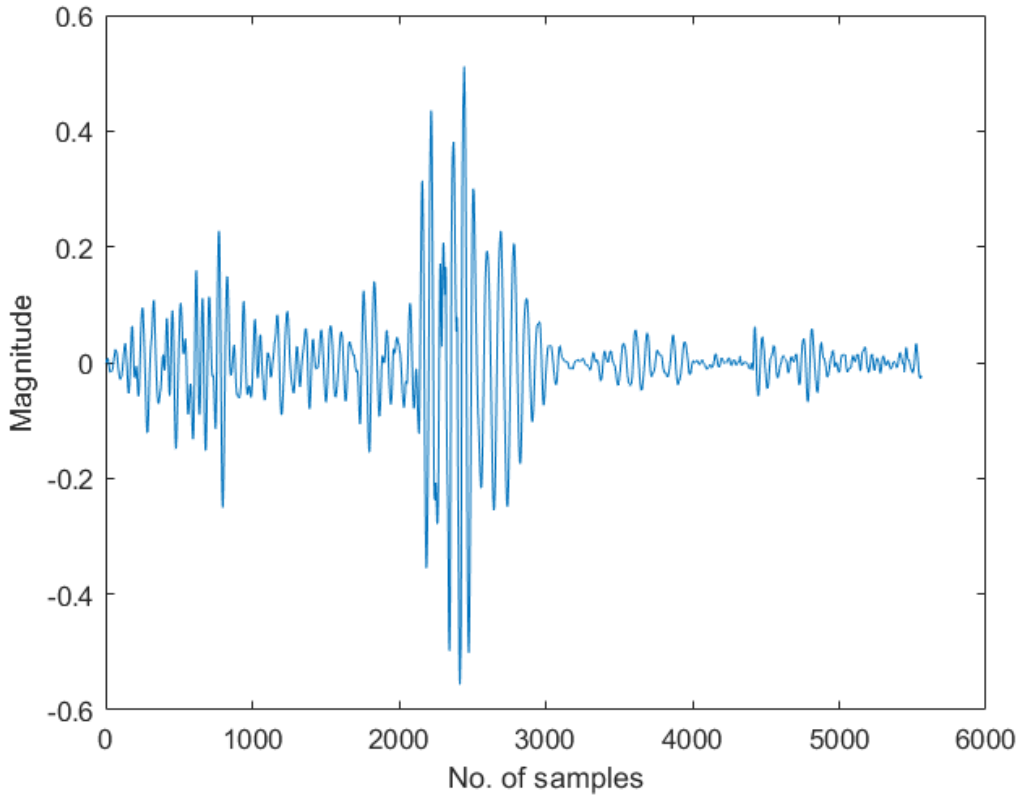



Figure 1.4. A phonocardiogram of mitral regurgitation.

#### 1.1.1.1.4. Mitral Valve Prolapse

Mitral valve prolapse (MVP) is a condition where one or both of the mitral valve leaflets prolapse into the left atrium during a late systole [6] and result in an abnormally reversed blood flow to the left atrium. The kind of murmur associated with MVP is shown in Figure 1.5

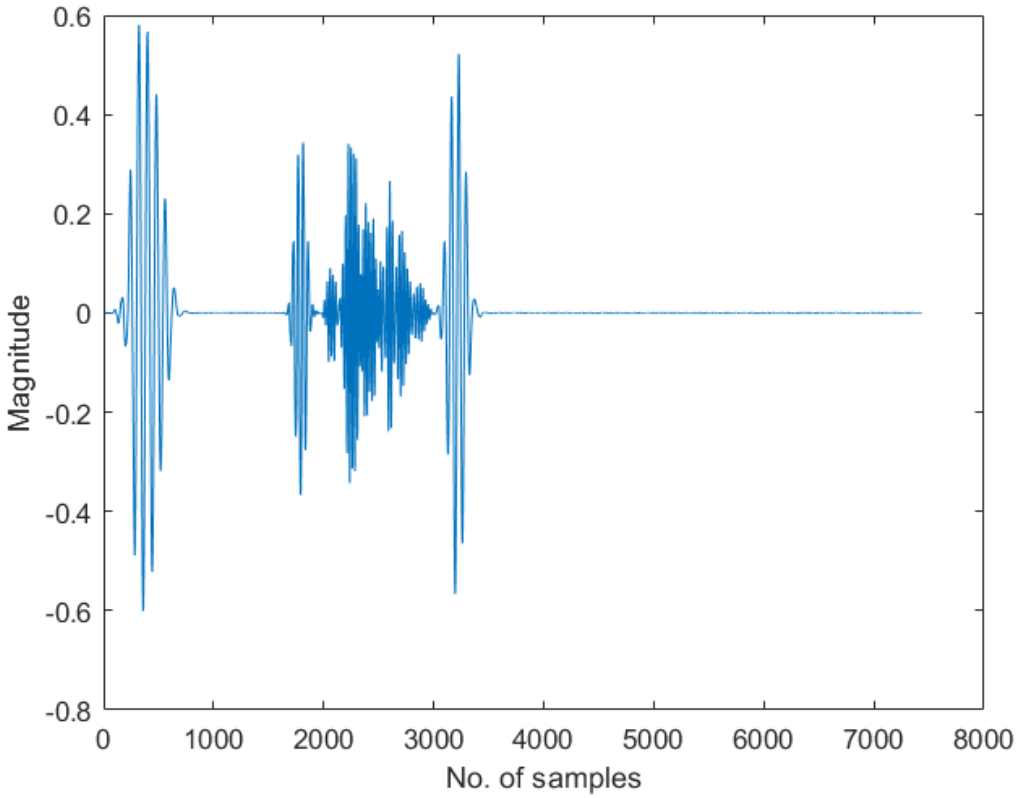



Figure 1.5. A phonocardiogram of mitral valve prolpase.

### 1.1.2. Cardiovascular Disease Diagnosis

The diagnosis of CVDs includes invasive and non-invasive methods. Traditionally, heart sounds are assessed by a physiologist through a procedure known as auscultation. In this procedure, a cardiologist listens to a heart sound (HS) using a stethoscope and makes a medical prognosis. Thus, auscultation, a non-invasive procedure, is a convenient method employed to listen to heart sounds. However, its dependence on a physician's expertise and skill is considered to be its main drawback [7–9]. On the other hand, a heart sound graph, known as a phonocardiograph, could be obtained for visual interpretation by a physician. Yet, the non-stationarity nature of heart sound signals adds to the difficulty of employing visual diagnosis by the physician. In order to alleviate the subjectivity of the heart sound diagnosis by a physician, there is ongoing research to automate HS diagnosis as it is seen as a better alternative. A plethora of published studies in the literature have investigated automatic HS signal classification in the few last decades.

## 1.2. Literature Review

Leung et. al. [10] segmented the recordings in an HS dataset using the QRS complex in the ECG signal. The dataset contains 21 normal and 35 abnormal recordings. The authors employed the trimmed mean spectrogram method to extract time-frequency features from the HS segments. A specificity of 94.4% and sensitivity of 97.3% were obtained using a probability neural network.

Turkoglu and Arslan [11] investigated heart sound classification using a dataset of 200 recordings with ten varieties of the heart sounds (including AS, MS among others). They employed wavelet decomposition as their feature extraction method. The extracted features were used to train a feed-forward backpropagation network with adaptive learning rate. They trained and tested their model with 100 records each and reported a 4% incorrect classification.

Zin et. al. [12] utilized integrated wavelet transform to distinguish between normal heart sounds and MR. They divided the HS signals into a fixed number of frames using the points of maximum energy in S1 and S2. Next, the computed the average energy of each frame. Using these features, they reported their best classification accuracy of 94.2% using a multilayer perceptron (MLP) network. The employed 125 instances for training and 52 instances for testing.

Ari and Saha [13] investigated heart sound classification using a dataset that included normal heart sounds and 12 variants of heart pathologies. The authors segmented the heart sound signals to separate cardiac cycles. Next, they employed empirical mode decomposition (EMD) and split the residue, the $2^{nd}$ and $3^{rd}$ intrinsic mode function (IMF) into separate sub-windows. Subsequently, they calculated the average energy for each sub-window and formed 25 long feature vector. The performance of their proposed method was evaluated using noise-free and noisy heart sounds with an artificial neural network. They compared the performance of EMD-based feature extraction method with the wavelet-based feature extraction method. According to the authors, the EMD based feature surpassed its wavelet counterpart and achieved a recognition rate of 99.038% using an ANN classifier.

Vepa [14] investigated the use of cepstral features to distinguish between normal heart sound, systolic murmur and diastolic murmur. The author used 3 different classifiers, namely k-nearest neighbor (KNN), MLP, and support vector machines (SVM) to compare the performance of the proposed feature with spectral and wavelet-based features. The author reported that the Cepstral features offered the best classification accuracy of 95.2% from 10-fold cross-validation experiments with an SVM classifier.

Safara et. al. [15] investigated a multi-level basis (MLBS) method for node selection in wavelet packet tree for heart sound classification. The authors calculated the relative energy of the wavelet packet tree. They compared MLBS with other node selection techniques including best basis selection, single-level basis selection and location discriminant basis selection. The MLBS experiments with an SVM classifier outperformed the other methods and achieved the best classification accuracy of 97.56%.

Randhawa and Singh [16] investigated heart sound classification using a dataset that comprised normal heart sounds, systolic murmur and diastolic murmur. They extracted several time, frequency and statistical domain features. Next, they employed Fisher's Discriminant Ratio for feature reduction. The selected features were fed to 3 different classifiers namely KNN, fuzzy KNN and ANN. Using 5-fold cross-validation experiments, the highest classification accuracy of 99.6% was achieved with KNN and fuzzy KNN.

To classify heart sound signals into normal, abnormal and unsure (cases that include a significant amount of noise), Nilanon et. al. [17] windowed heart sound signals to obtain 5-second segments and classified each recording based on the majority vote reported from the classification experiments for the segments. Using both spectrogram and Mel-frequency cepstrum coefficients features with CNN, the authors reported the best accuracy score of 0.813, sensitivity of 0.735 and specificity of 0.892.

Karar et.al. [18] proposed a method that involved the use of discrete wavelet transform to preprocess PCG signals and Lyapunov exponents to extract features from heartbeat cycles. They performed cross-validation experiments with a rule-based classifier. They investigated the classification of the heart sounds into normal, aortic valve stenosis, aortic insufficient, and ventricular septum defect and reported the best classification accuracy of

95.5%.

Deperlioglu [19] used 8-second segments of heart sound recordings to classify heart sounds into normal, murmur and extrasystole. The author employed a resized version of the PCG image with both artificial neural networks and CNNs. A total of 134 ,29 and 29 samples were used for training, validation and testing, respectively. According to the experimental results, the CNN model outperformed the ANN model and achieved the best classification accuracy of 97.7%.

Yaseen et.al. [2] created a heart sound dataset that comprised normal, AS, MS, MR and MVP (this dataset was selected for experiments in this thesis). The authors proposed a joint feature extraction method based on Mel-frequency cepstral coefficients and DWT. The authors tested the performance of their proposed method using SVM, deep neural network and a centroid based KNN. Using 5 fold cross-validation experiments, the SVM surpassed the other classifiers and recorded a classification accuracy of 97.9%.

Upretee and Yüksel [7] proposed a time-varying spectral feature method, known as centroid frequency, to investigate heart sound classification. They segmented the heart sound recordings into separate cardiac cycles. Next, they resampled the cardiac cycles to unify the number of samples per signal. Both KNN and SVM were used to evaluate the performance of the proposed feature vector. The best classification accuracies of 99.60% and 96.50% were achieved using KNN for binary and multi-class classification experiments, respectively.

Krishnan et. al. [20] investigated heart sound classification from raw signals using 1D CNN and MLP. The authors divided each signal into 6 s epochs. An MLP with 5 hidden layers outperformed 3 different 1D CNN models by recording the maximum accuracy of 0.8565 with a sensitivity of 0.8673, and specificity of 0.8475.

### 1.3. Problem Statement

The World Health Organization (WHO) reported that cardiovascular diseases (CVDs) were the leading cause of global death between the years 2000 and 2016 [21]. CVDs accounted for an estimated 31% of the global death cases in 2016. According to WHO,

more than 75% of the mortality caused by CVDs take place in the developing countries that lack the required health care facilities for early detection [1]. Hence, it is pressing to address this problem and find solutions to reduce the high mortalities rates associated with CVDs.

The proposed methodologies in the literature for automatic heart sound classification could be divided into two broad groups: machine learning (ML) based methods and deep learning (DL) based methods. The steps followed in ML-based methods are summarized in Figure 1.6. In these methods, the heart sounds in a dataset are processed by a suitable technique to extract several salient features for subsequent classification experiments. Thus, the extracted features are used to train an ML model. The efficiency of this approach, however, is highly dependent on the extracted features and the selected ML model. Even the optimization algorithm employed for training the model indirectly affects the model performance.

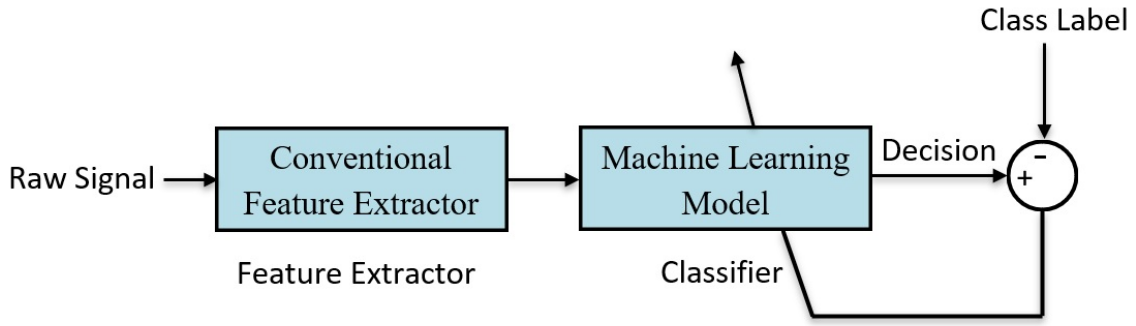



Figure 1.6. The training of machine learning models.

DL based methods could be employed to avoid the inevitable problems arising from the use of ML-based methods. In DL methods, there is no need for a separate feature extractor, as features are extracted automatically within the layers of the deep network. Figure 1.7 presents the block diagram of DL methods. Long short term memory (LSTM), for example, is a deep network that can directly process a raw signal. Convolutional neural network, on the other hand, requires the transformation of the signal into image representation methods (such as spectrogram) that are more convenient for CNN classification. In addition, transfer learning-based models, such as AlexNet, are more efficient and serve as a better alternative. The training process of these models is summarized in Figure 1.8

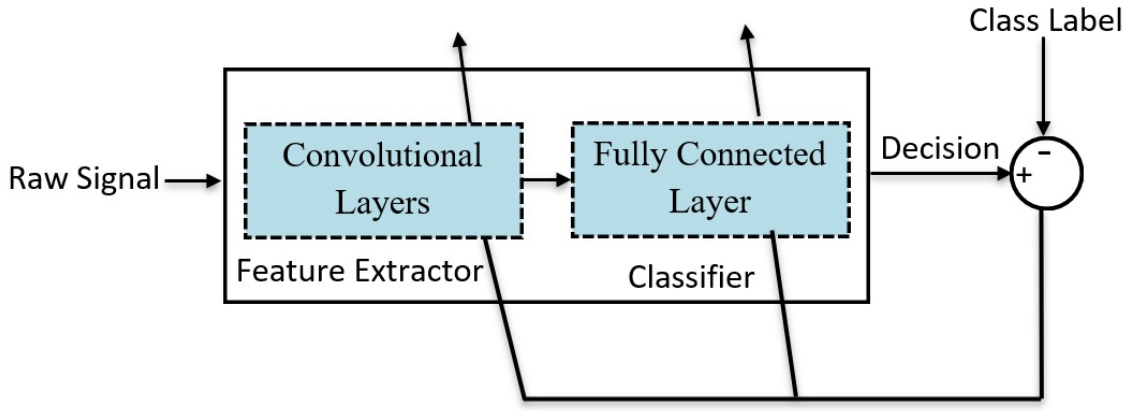



Figure 1.7. The training of deep learning models.

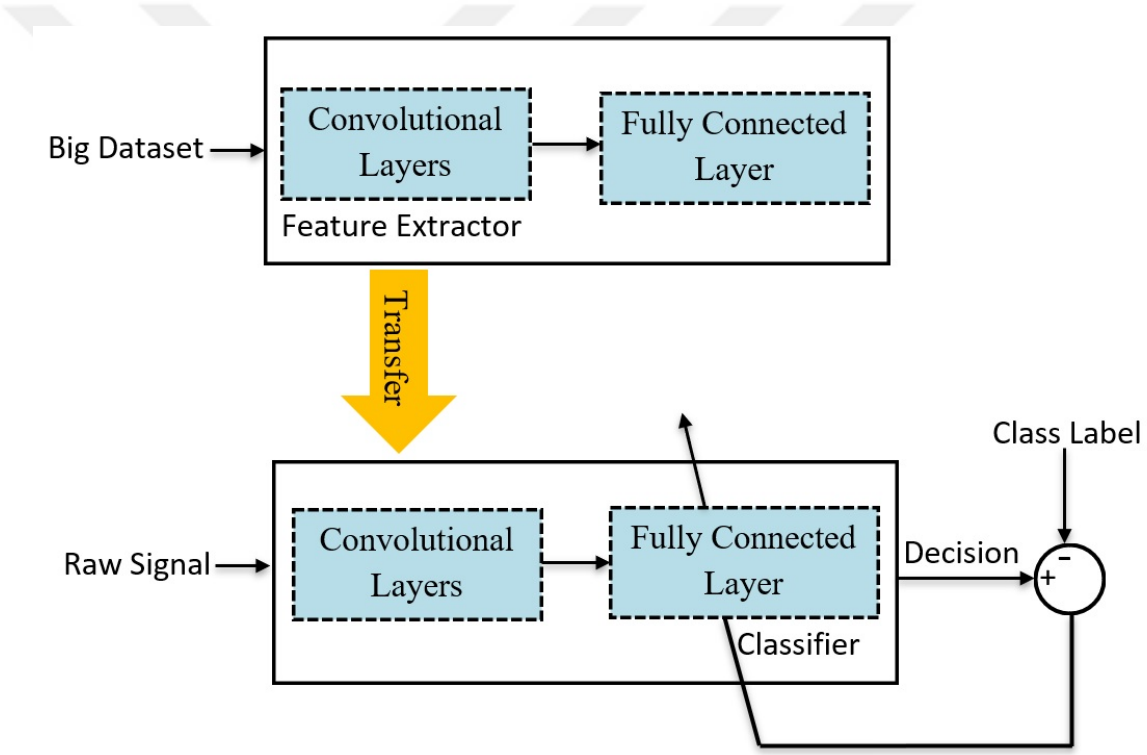



Figure 1.8. The training of pretrained deep learning models.

Majority of the methods proposed so far in the literature for automatic heart sound classification include the joint use of several features. Although the use of several features sometimes improves the model performance, it can also lead to performance deterioration and an increased computational complexity. Hence, there is a need for a highly efficient and a less complex method for heart sound classification with improved classification performance.

## 1.4. Thesis Objective

The objective of this thesis involves the investigation of the highly efficient wavelet analysis based scalogram for heart sound classification. The scalogram images are used with CNN and 4 different popular pre-trained models to confirm the robustness of the method for heart sound classification. Furthermore, the histogram of oriented gradient features are extracted from the scalogram images of the HS for subsequent experiments with ML models. Thus, this thesis aims to give a solution with improved classification accuracy to reduce the mortality caused by CVDs.

# 2. CHAPTER

# Method

## 2.1. Introduction

Figure 2.1 shows a summary of the proposed classification system. First, the heart sound (HS) signals are segmented to 3 individual cardiac cycles per signal. Secondly, continuous wavelet transform (CWT) is utilized to compute the scalogram of the separate segments. The time-frequency scalogram images are then fed to deep learning models for classification. In the alternative scenario, the histogram of oriented gradient (HOG) features are extracted from the scalogram images, and the feature vectors are fed to machine learning classifiers for decision making.

## 2.2. Data Acquisition

The data used in this study is obtained from an open-source dataset collected by Yaseen et. al. [2]. It consists of normal and abnormal heart sounds. The abnormal set is further subdivided into aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP). Each of the subcategories consists of 200 recordings. The data were sampled at a sampling frequency of 8000 Hz. After removing recordings with extreme noise, the authors in [2] filtered and standardized the data, hence no further preprocessing was implemented in this thesis project.

## 2.3. Data Segmentation

Segmentation is one of the principal steps in an automatic heart sound classification system. Different methods are used in the literature to segment heart sound recordings.

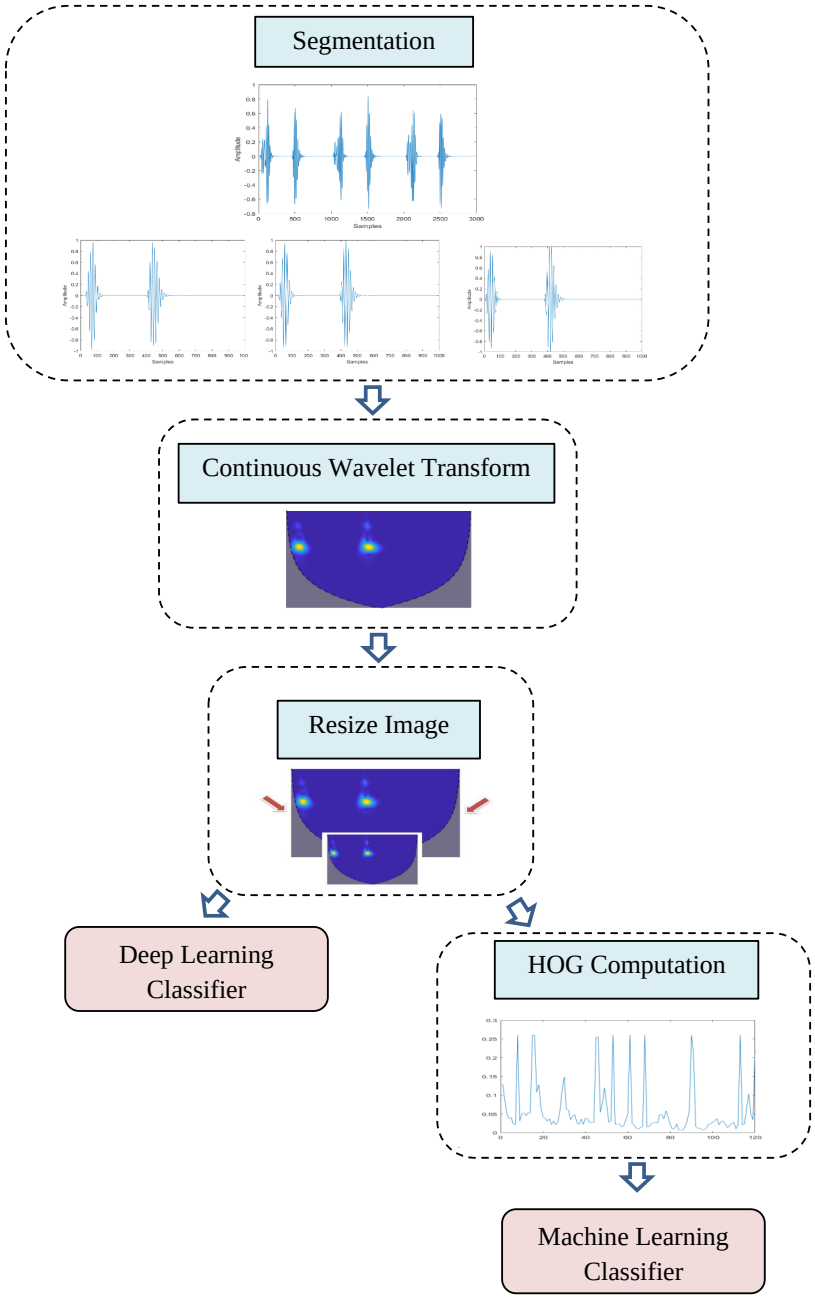



Figure 2.1. Proposed method for Heart sound signals classification

While some divided the heart sound into first and second heart sound (S1 and S2) [22], others split it into time frames [17, 19]. However, the division of heart sound recordings into fundamental heart sounds is not an accurate process because murmurs are present not only within S1 or S2 but they are also present in the period between them. Also, time-based segmentation is not an accurate process, since segments may not start at the same point. The majority of the studies seen in the literature segmented heart sound recording into cardiac cycles [18, 23]. The cardiac cycle is defined as the events that take place from the beginning of the S1 until the start of the next S1 [2], hence it contains all the information necessary for analysis. A single recording in the dataset consists of 3 cardiac cycles, hence segmenting the recordings into individual cardiac cycle increased the number of signals from originally 200 recordings per class to 600 recordings per class. Figure 2.2 shows the individual cardiac cycles acquired after segmentation. Details of the dataset before and after the segmentation process are presented in Table 2.1. It is worth pointing out that the segmentation had an additional advantage of increasing the size of the dataset which is crucial for deep learning methodologies [24].

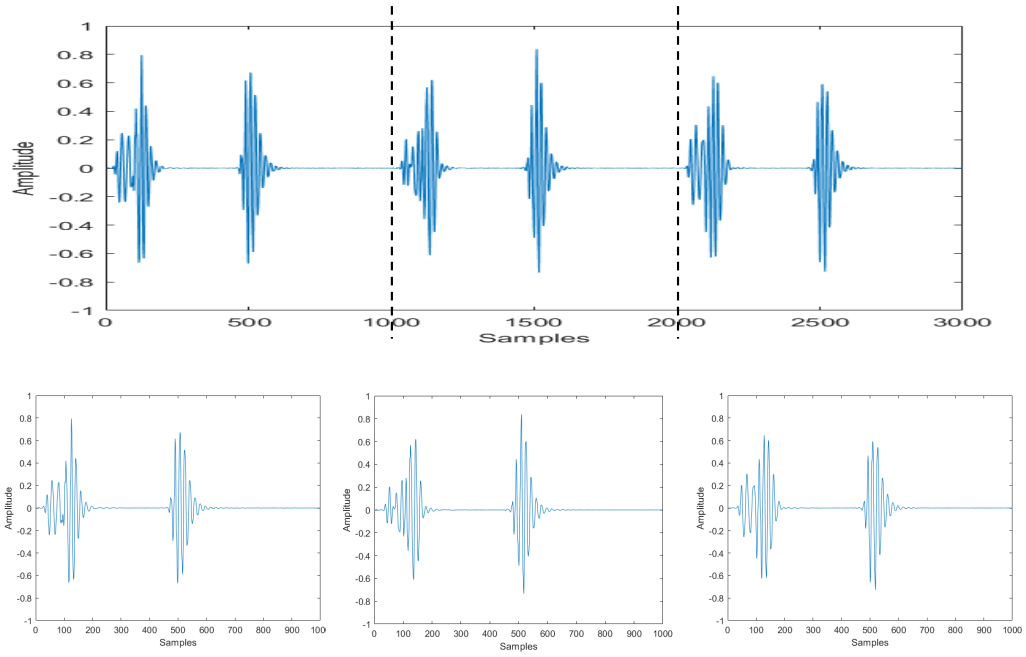



Figure 2.2. The segmentation of a heart sound signal into 3 separate cardiac cycles.

Table 2.1. The details of the dataset before and after the segmentation process.

| 2 class | 5 class | Pre Segmentation Number of recordings | Post segmentation Number of recordings |
|---|---|---|---|
| Abnormal | MVP | 200 | 600 |
| | MR | 200 | 600 |
| | AS | 200 | 600 |
| | MS | 200 | 600 |
| Normal | | 200 | 600 |
| Total | | 1000 | 3000 |

## 2.4. Feature Extraction

### 2.4.1. Continuous Wavelet Transform

The Fourier transform is considered as a very powerful tool for signal analysis in frequency domain; nevertheless, it reflects no information regarding the time domain of the signal. In the case of signals with frequency content varying with time, simultaneous analysis of the signal in both time and frequency domain becomes a must for successful representation of the signal. Short-Time Fourier Transform is a better substitute for Fourier transform since it can simultaneously provide time and frequency analysis of the signal. It employs a window to slice the signal and then applies Fourier transform on the slices [25]. However, Short-Time Fourier Transform does not represent the signal efficiently, because it uses windows of fixed sizes which is disadvantageous as a small sized window captures high frequency components and misses low frequency components, and vice versa. To capture both high and low frequency components of a signal, wavelet transform can be used. It employs wavelet functions instead of sinusoids used in the case of Fourier transform. The wavelet function has scale and shift parameters that allow the function to be adapted to low and high frequency components. [26–28]. Two types of wavelet transform exist: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The key difference between CWT and DWT is the possible values of the scale and shift parameters. While CWT can take continuous values, DWT can only take discrete values [25]. CWT is used in real-time applications. The

mathematical formula to calculate CWT is expressed as:

$$C_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi^*\big(\frac{t-b}{a}\big)\mathrm{d}t \tag{2.1}$$

where $f(t)$ is the time domain signal, $\psi^*(t)$ is the complex conjugate of wavelet function, $a$ is the scale and $b$ is the time location [26].

For the case of a discrete signal, the CWT is defined as the convolution of the signal with the scaled and translated version of the mother wavelet [29], which is mathematically expressed as:

$$W_f(a,b) = \sum_{n=1}^{N-1} f[n]\psi^*\big[\frac{(n-b)\delta t}{a}\big] \tag{2.2}$$

where $f[n]$ is the discrete signal. Morlet wavelet is seen as the most convenient for analysis compared to other wavelet families [26]. It offers the best time-frequency localization [30]. The generalized Morse wavelets, on the other hand, constitute a superfamily that contains all the other commonly used analytic wavelets. The frequency domain definition of Morse wavelets is given as. [31]:

$$\psi_{\gamma,\beta} = U(\omega)a_{\gamma,\beta}\omega^{\beta}e^{-\omega^{\gamma}} \tag{2.3}$$

where $U(\omega)$ is the unit step function, $a_{\gamma,\beta}$ is the normalization constant, $\beta$ and $\gamma$ are the controlling parameters of Morse wavelet shape [31].

Using $\gamma = 3$, Morse wavelet catches the essential idea of Morlet wavelet, and avoids its deficiencies.

The scalogram, the magnitude squared of continuous wavelet transform, is the amount of energy at a single scale and time [32] and computed using:

$$S(a,b) = |C_f(a,b)|^2 \tag{2.4}$$

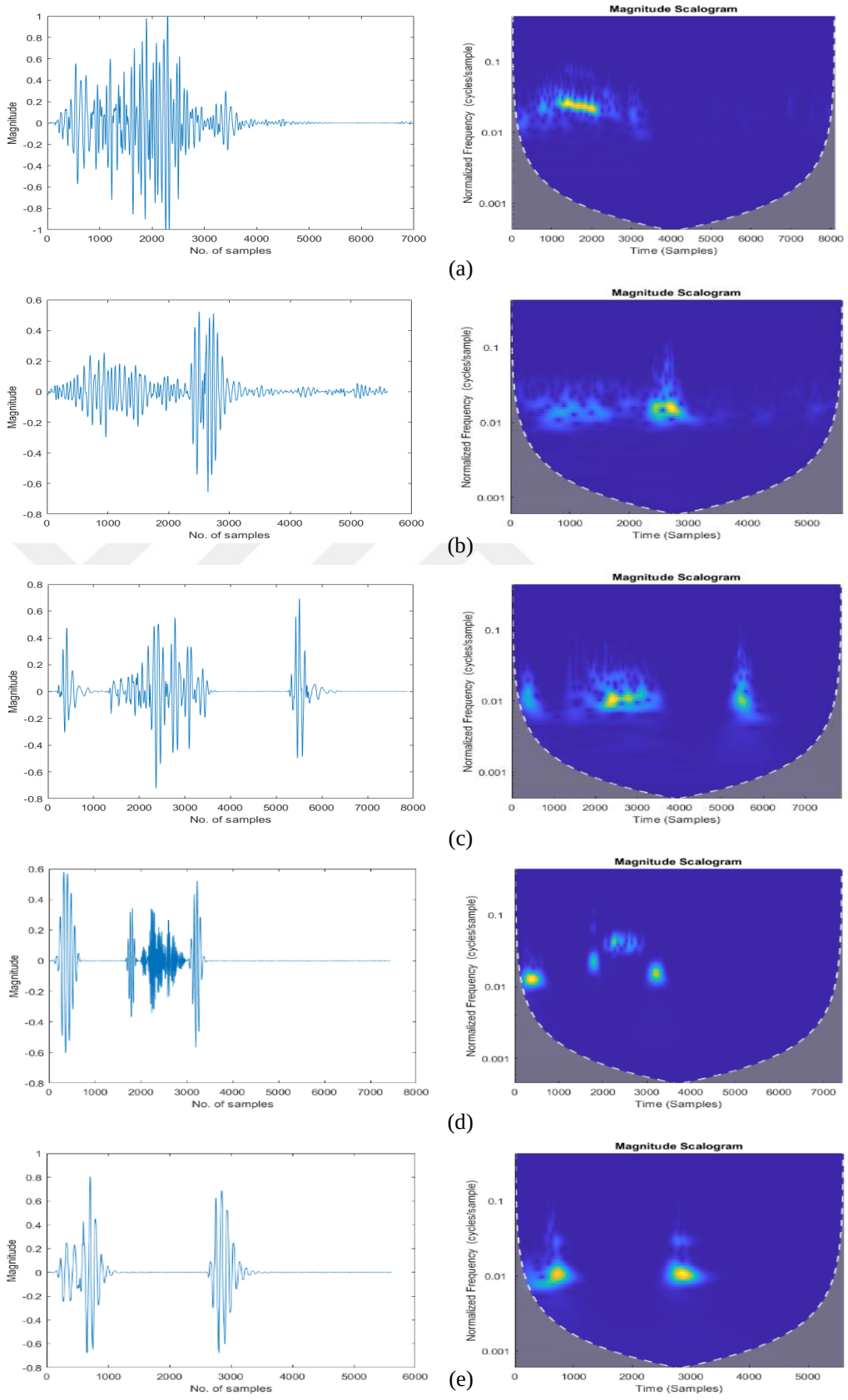



Figure 2.3. Heart sound signals and their corresponding scalogram images for: (a): AS; (b): MR; (c): MS; (d): MVP and (e): N.

Figure 2.3 shows the scalogram images for the different PCG signal types.

### 2.4.2. Histogram of Oriented Gradient

Histogram of Oriented Gradient (HOG) is a technique that has widely been used in object detection. The method counts the frequency of occurrences of gradient orientation (orientation histogram) in a certain region of the image. The HOG feature vector grasps all the useful information from the image and discard all the extraneous information. The feature extraction procedure includes several steps which are summarized in Figure 2.4.

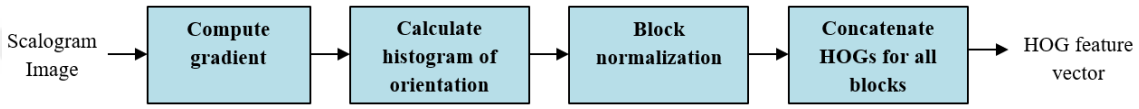



Figure 2.4. The HOG feature vector extraction process.

#### 2.4.2.1. Gradient Computation

Several masks (shown in Figure 2.5) could be utilized to calculate the horizontal and vertical gradients (Gx and Gy), including centered and uncentered first derivative masks, and 3x3 Sobel masks [33].

$$M_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad M_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

| -1 | 0 | 1 |
|---|---|---|

| -1 |
|---|
| 0 |
| 1 |

Figure 2.5. The horizontal and vertical Sobel masks, and first derivative centered masks [34]

Subsequently, the magnitude and direction of the gradient are calculated using: [35]:

$$G_{(x,y)} = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{2.5}$$

$$\theta(x,y) = arctan[G_y(x,y)/G_x(x,y)] \tag{2.6}$$

#### 2.4.2.2. Histogram of Gradient Computation

After computing the magnitude and direction of gradient, the image is divided into separate cells (e.g. 8x8 pixels) and the histogram of orientation is calculated for every cell. The magnitude of each pixel is assigned to its corresponding orientation bin. The orientation bins are equally spaced between 0 and 360 degrees (for the case of signed orientation) or 0 and 180 degrees (unsigned orientation) [33]. Figure 2.6 provides an overview of the assignment process. The signed orientation was used in this study because it has been empirically proven to be better than its counterpart for scalogram image classification.

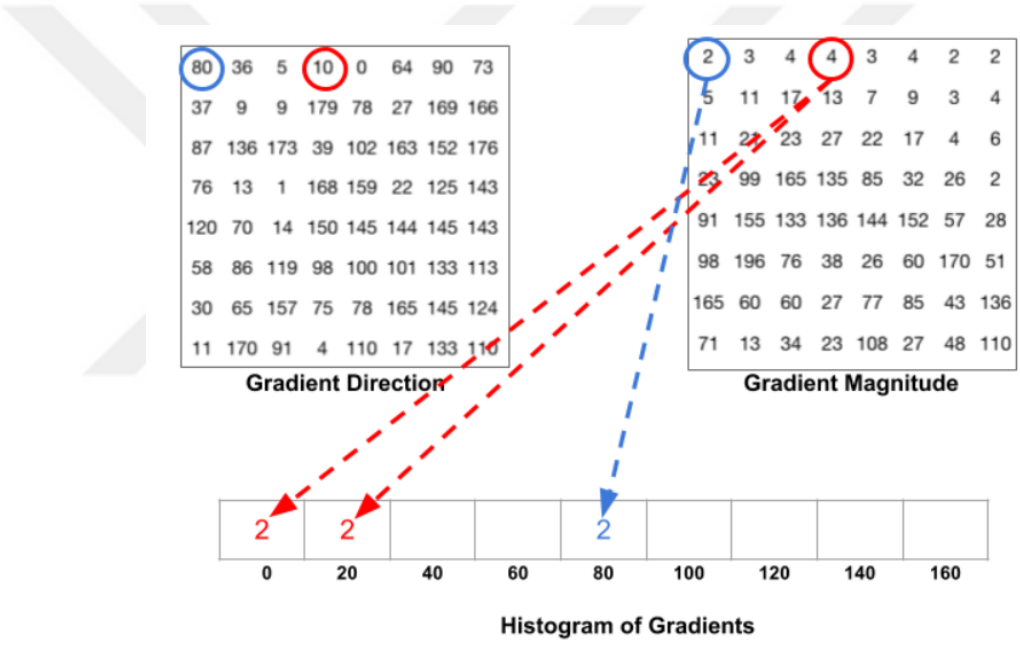



Figure 2.6. Assigning gradients to their corresponding bins [34]

#### 2.4.2.3. Block Normalization

Contrast normalization aims to remove the effect of differences of local illumination. Here, the cells are grouped into overlapping blocks (e.g. 2x2 cells) and normalization in executed on separate blocks. [33]

Four different methods could be used for block normalization where the normalization factors for them are:

- $L2 - norm : f = \frac{v}{\sqrt{||v||_2{}^2 + e^2}}$

- $L2-Hys$ is $L2-norm$ followed by clipping (limiting the maximum values of $v$ to 0.2) and renormalizing
- $L1-norm: f = \frac{v}{||v||_1+e}$
- $L1-sqrt: f = \sqrt{\frac{v}{||v||_1+e}}$

where $v$ is the non-normalized vector, $||v||_k$ is the vector's $k$-norm for $k$ = 1, 2 and $e$ is a small constant.

#### 2.4.2.4. HOG Feature Vector Computation

The HOG feature vector is produced by concatenating all components of the normalized cells into one vector. Figure 2.7 shows the resultant HOG feature vectors for the five heart sound data subcategories.

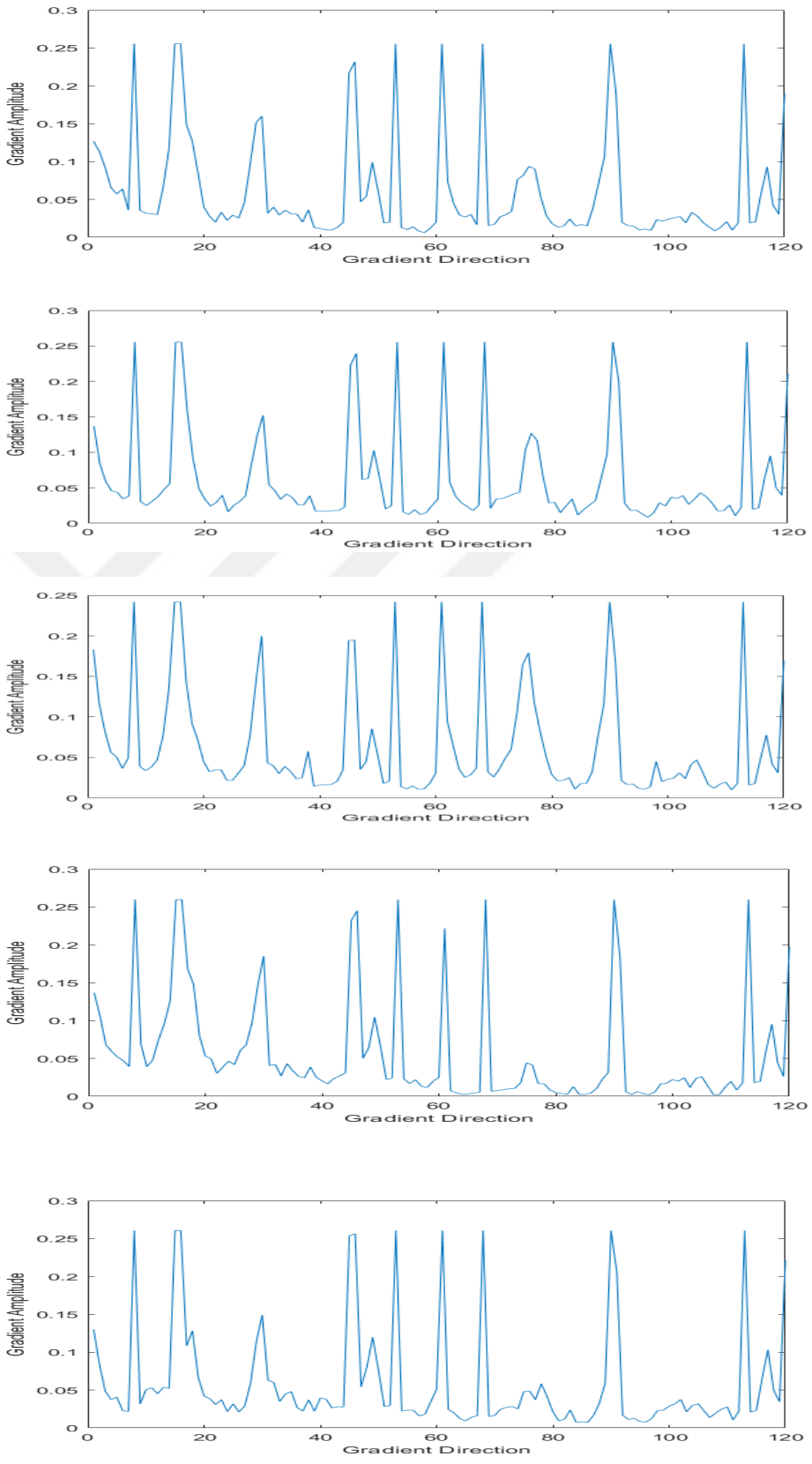



Figure 2.7. Resulted HOG feature vector for: AS; MR; MS; MVP and N

## 2.5. Image Resize

The initial size of the scalogram images is $656 \times 875$. Pretrained models, however, demand a specific size for their input images. For example, AlexNet requires input images of size $227 \times 227$ and ResNet requires input images of size of $224 \times 224$, therefore scalogram images were resized using the bicubic interpolation algorithm to match the input image size required by the selected pre-trained model. Furthermore, resizing the images reduce the computation complexity and thereby the time taken by the classification model for reaching a decision. For extracting HOG features, the image size was determined heuristically and verified experimentally. The size of $128 \times 128$ gave the best classification performance in comparison to the other experimented values.

## 2.6. Classification

### 2.6.1. Introduction

The classifiers used in this step include deep learning as well as machine learning classifiers. While the DL classifiers are used to classify the scalogram images, the ML classifiers are used to classify the HOG feature vectors which are extracted from the scalogram images.

- Classification with machine learning methodologies includes the following classifiers:
  - KNN.
  - SVM.
- Classification with deep learning methodologies includes the following classifiers:
  - CNN.
  - Transfer learning based models:
    - AlexNet.
    - VGG Net.

* GoogLeNet.
* ResNet.

### 2.6.2. Machine Learning Classifiers

#### 2.6.2.1. Support Vector Machines

Support vector machine (SVM) classifier employs what is called a kernel to transform a dataset that cannot be separated linearly into higher dimensional space. The kernels used include linear kernel, polynomial kernel, and exponential kernel. Table 2.2 presents the mathematical representation of various kernels.

Table 2.2. Different kernel functions in SVM [36].

| Name of the kernel | Mathematical formula |
|---|---|
| Linear | $k(x, y) = x^T.y$ |
| Polynomial | $k(x, y) = (x^T.y + 1)^p$ where $p$ is the polynomial degree |
| RBF (Gaussian) | $k(x, y) = exp(\frac{-\|x-y\|^2}{2\sigma^2})$ where $0 < \sigma < 1$ |

After transforming the data into the new hypothetical space, multiple hyperplanes – $p-1$ dimensional subspaces for $p$ dimensional data space – could be used to separate the data points. The main objective of SVM is to find the optimal hyperplane among them, which is also called maximal margin hyperplane. It is called the maximal margin hyperplane because it has the largest margin (the smallest perpendicular distance between the data points and the hyperplane). The data points can then be classified based on their location on the sides of this hyperplane [37]. The choice of the optimal hyperplane depends on a subset of the data elements known as the support vectors. These support vectors support the choice of the optimal hyperplane since they are the closest data points to the hyperplane [37]. The support vectors are shown in Figure 2.8.

#### 2.6.2.2. K Nearest Neighbors

K nearest neighbor (KNN) classifier assigns a single data point to a particular class according to its similarity with other data points in the training set. In other words,

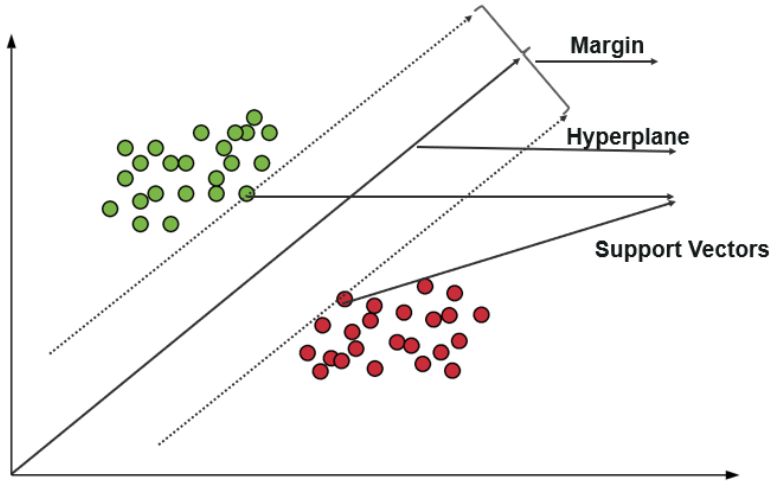



Figure 2.8. Classification using a support vector machine [38].

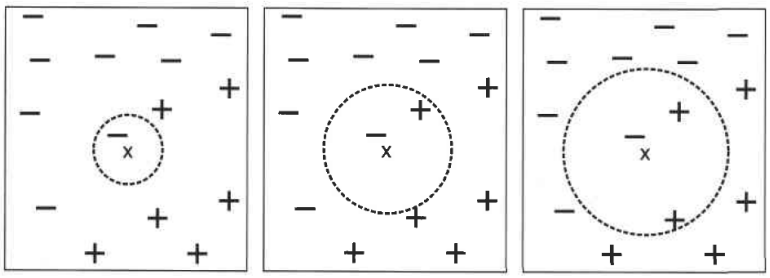


Figure 2.9. 1, 2, and 3 nearest neighbors in KNN [39].

test records are considered to belong to the same class of the training records similar to them (nearest neighbors). If the nearest neighbors belong to more than one class, the test record is assigned to the majority class of them. Figure 2.9 represents the classification process using different numbers of neighbors. For measuring the similarity KNN employs proximity measures. Table 2.3 presents some of the distance metrics used for measuring similarity between two points $A$ and $B$, where $A = (x_1, x_2, ..., x_n)$ and $B = (y_1, y_2, ..., y_n)$ .

Table 2.3. Distance metrics in KNN.

| Distance metric | Equation |
|---|---|
| Euclidean distance | $dist(A,B) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$ |
| Cosine similarity | dist(A, B) = $\frac{A.B}{\lvert A\rvert \times \lvert B\rvert}$ |
| Correlation distance | $dist(A,B) = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})\sum_{i=1}^{n}(y_i-\overline{y})}}$ |

### 2.6.3. Deep Learning Classifiers

#### 2.6.3.1. Convolutional Neural Networks

Convolutional Neural Network (CNN) is one of the most popular deep learning architectures. As its name implies, CNN performs a convolution process [40]. Convolutional layers, pooling layers and fully connected layers are the main constituents of CNN [41, 42].

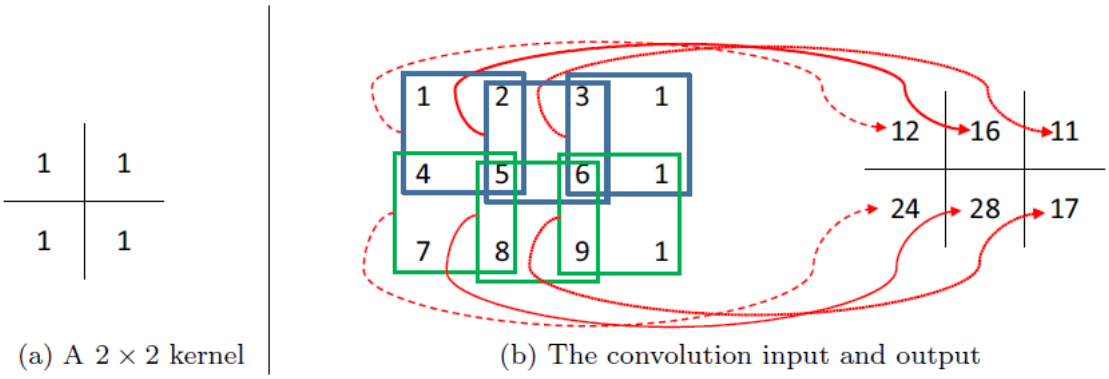
Convolutional Layer: This layer employs multiple kernels to perform the convolution operation with an input and generates feature maps as its output [42, 43]. Convolution operation is presented in Figure 2.10. To perform the operation, the kernel simply slides over the input and performs a dot product. An important parameter in the convolution process is the stride, which defines the number of pixels the kernel skips when sliding over the input. When stride (S) = 1, the kernel moves without skipping any pixel, but for every value S > 1, the kernel skips S-1 pixel(s) while moving [44]. Another important concept is padding, in which the input image is padded with 0 or a numerical value such that the output has the same size as the input [44].

Pooling Layer: It is also called the down-sampling layer. This layer performs pooling operation which is a process of mapping a region into a single number by replacing the region with its average (average pooling) or its maximum (max pooling) [44]. The pooling operation is depicted in Figure 2.11. The importance of this operation is that it prevents overfitting and reduces computational complexity [45].

Fully Connected Layer: By working like a conventional artificial neural network [41, 46, 47] this layer flattens the input 2D feature map and converts it to 1D feature vector [41,46]. The flattening operation is depicted in Figure 2.12.

#### 2.6.3.2. Transfer Learning Based Classifiers

The intuition behind transfer learning is adopting a model that was initially trained (pre-trained) on a larger dataset to solve a complicated task and using this pre-trained model to solve a relatively simpler problem with a small dataset [50] such that the knowledge acquired will facilitate the training process in the second implementation.




(a) A $2 \times 2$ kernel (b) The convolution input and output

Figure 2.10. Convolution Operation [44]

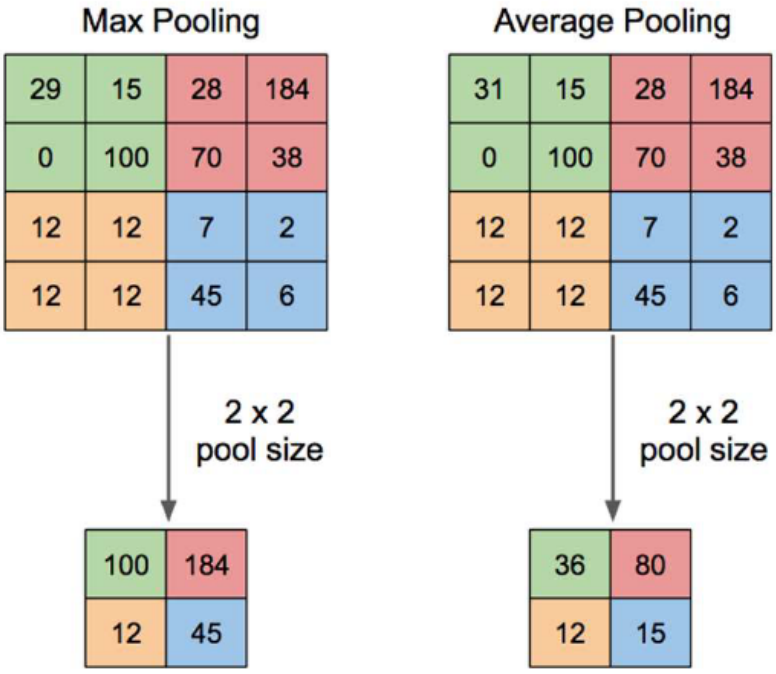



Figure 2.11. Pooling Operation [48]

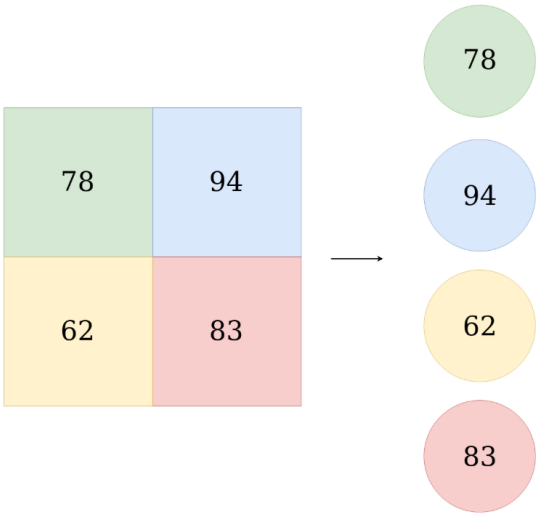



Figure 2.12. Flattening Operation [49]

It has been empirically proved that the utilization of transfer leaning based models considerably maximize the model performance and minimize the training cost [51].

#### 2.6.3.2.1. AlexNet

AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [52]. It is a transfer leaning-based CNN that consist of 5 convolutional layers and 3 fully connected layers.

#### 2.6.3.2.2. VGG Net

Increasing the depth of a network significantly increases performance. Therefore, the idea behind VGG Net is to build a model that is relatively deep compared to its previous AlexNet counterpart without increasing the total number of model parameters. To achieve this, VGG Net employs a single filter size of 3x3 in all of its convolutional layers. Three different VGG Net models with different number of layers exist, these are: VGG-11, VGG-16 and VGG-19 [52] and they have 11 layers, 16 layers and 19 layers respectively. **VGG-16 was used in this thesis study**. It is depicted in Figure 2.13.

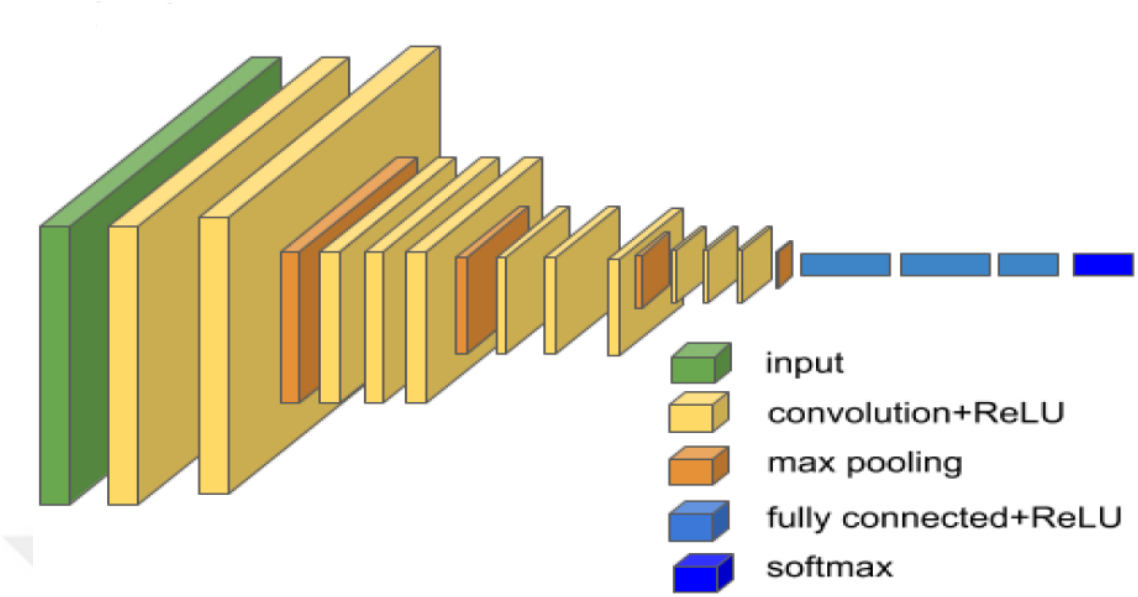



Figure 2.13. VGG-16 Architecture [53].

#### 2.6.3.2.3. GoogLeNet

GoogLeNet was the winner of the annual ILSVRC competition in 2014 [52]. It is made up of a number of concatenated Inception modules which is depicted in Figure 2.14 . It is significantly deep compared to its predecessors (AlexNet and VGG Net). It has 22 layers, nevertheless, it has less computational complexity and thereby shorter computational time. It achieves this by introducing what is called network-in-network (1x1 convolutional) architecture [52].

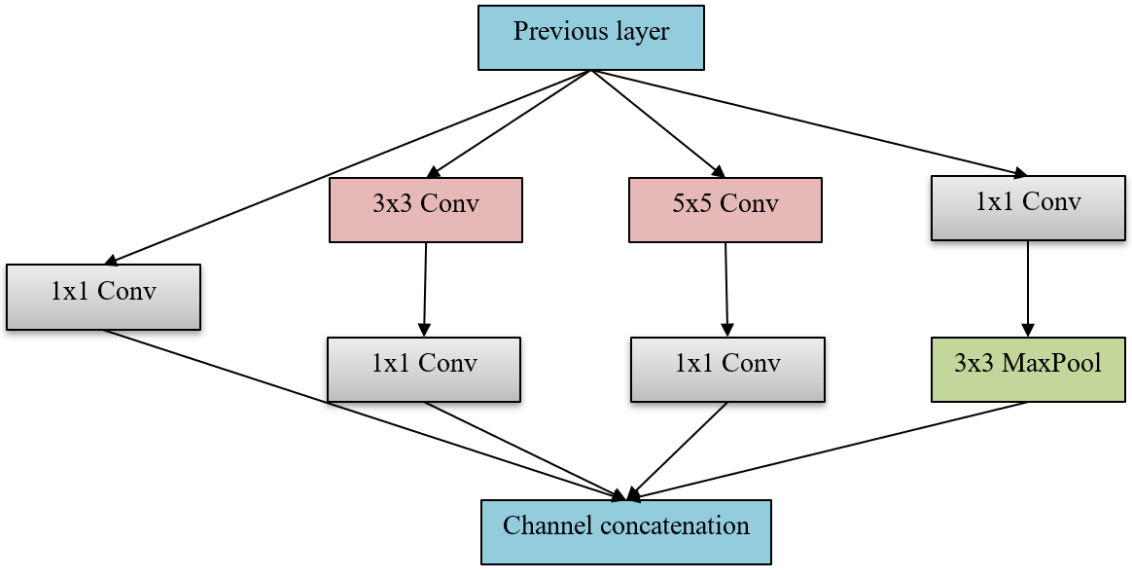



Figure 2.14. Inception module.

#### 2.6.3.2.4. ResNet

As the network gets deeper, it encounters vanishing and exploding gradient problems during training. As the depth of the network increases, the accuracy increases, but after a while it runs through a period of saturation followed by degradation. ResNet employs what is called a residual block which utilizes an identity shortcut to solve vanishing gradient problems by feeding the input to the output. It is presented in Figure 2.15. ResNet won the ILSVRC contest in the year 2015 [52].

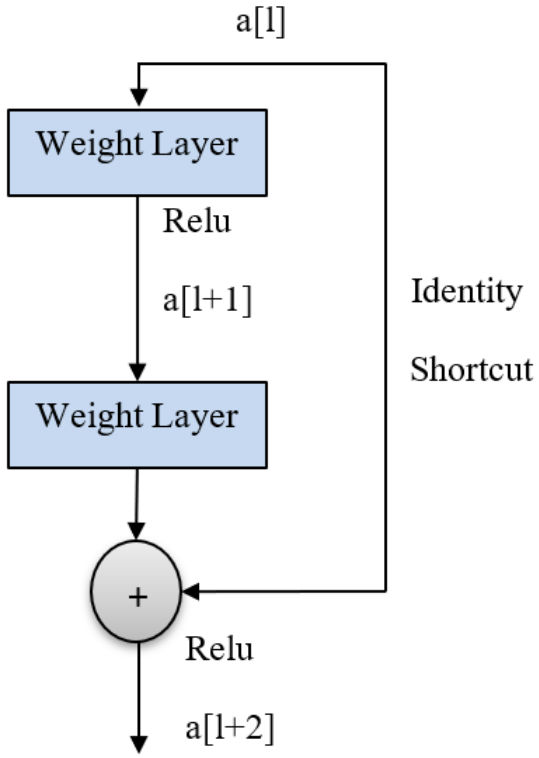



Figure 2.15. Residual block

## 2.7. Validation

Cross validation or k-fold cross-validation is a procedure used to evaluate classification models. It is used to estimate the performance of a model using unseen (test) data. As its name suggests, it works by randomly splitting that dataset into $k$ folds or groups and iteratively selecting one group for validation and the rest $k-1$ for training the model. The $5$ fold cross validation is summarized in Figure 2.16 The results of a k-fold cross-validation are often summarized with the mean and the standard deviation of the model validation scores.

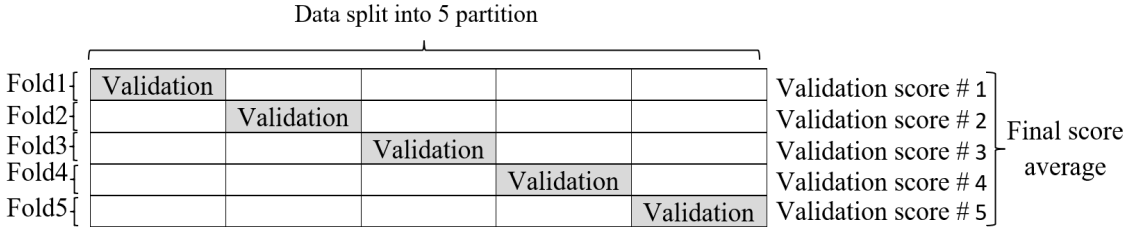



Figure 2.16. A 5 fold cross validation process.

## 2.8. Performance Evaluation

The confusion matrix is a table used to report the performance of a classification model. It contains a number of rows and columns equal to the number of classes. Confusion matrix for binary classification is depicted in Figure 2.17 It describes the predicted and actual values using the following terms:

| | | Actual class | |
|---|---|---|---|
| | | P | N |
| Predicted class | P | TP | FN |
| | N | FP | TN |

Figure 2.17. The confusion matrix for binary classification

True Positive (TP) represents the truly classified positive instances. True Negative (TN) represents the truly classified negative instances. False Positive (FP) represents negative instances misclassified as positive. False Negative (FN) represents positive instances misclassified as negative.

To assess the performance of a classification model, a number of metrics including accuracy, sensitivity, specificity and F1 score are used. Accuracy is the proportion of truly classified instances and is expressed mathematically as:

$$Accuracy = \frac{TP + TG}{TP + TN + FP + FN} \tag{2.7}$$

Sensitivity is the ratio of truly classified positive instances, and is computed as:

$$Sensitivity/Recall = \frac{TP}{TP + FN} \tag{2.8}$$

Precision is the ratio that is calculated by dividing the truly classified positive instances by the total number of predicted positive instances, and can be expressed mathematically

as:

$$Precision = \frac{TP}{TP + FP} \tag{2.9}$$

F1 score is the weighted mean of the precision and recall and is calculated as:

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{2.10}$$

# 3. CHAPTER

# Results

## 3.1. Introduction

The heart sound dataset collected by Yaseen et. al. [2] was used in all the experiments carried out in this thesis. Two classification problems were addressed: binary classification (normal and abnormal) with 600 data points for the normal class and 2400 data points for the abnormal class; and multi-class classification (classification into AS, MR, MS, MVP and N) with 600 instances per class. First, each recording in the dataset was segmented to 3 separate cardiac cycles. Next, the continuous wavelet transform was employed to compute the scalogram of the heart sound segments. The scalogram images were then resized and fed to deep learning classifiers. In another scenario, the HOG features were extracted from the resized scalogram images and the extracted feature vectors were examined with machine learning classifiers. Several different experiments were carried out to test the performance of the proposed methods.

## 3.2. Image Based Classification

Deep learning classifiers, such as CNN and transfer learning-based models offer high classification accuracies for 2D images. Hence, CNN and 4 different popular pre-trained models were employed to classify the scalogram images obtained for the heart sounds.

### 3.2.1. Selection of the CNN model

To find the best CNN model (optimal number of convolutional layers and fully connected layer) for the scalogram image classification, a number of experiments were carried out.

Table 3.1 presents the accuracies recorded during varying the number of convolutional layers (Conv) from 1 to 5 and setting the number of fully connected layers (FC) to one layer. It was observed that as the model gets deeper, the performance got better, hence the number of convolutional layers was set to 5 in all further experiments. Furthermore, the performance of the model was enhanced by adding 2 other FC layers. The final architecture of the CNN is presented in Table 3.2.

Table 3.1. The accuracy performance of the multi-class classification experiments using different number of convolutional (Conv) layers.

| No. of Conv Layers | Accuracy(%) |
|---|---|
| 1 | 97.00 |
| 2 | 98.43 |
| 3 | 98.83 |
| 5 | 99.07 |

Table 3.2. The optimal parameters of the convolutional neural network model.

| Layer | Filter Size | Number of ilters | Number of Neurons | Stride |
|---|---|---|---|---|
| Conv-1 | $11 \times 11$ | 96 | - | 4 |
| MaxPooling | - | - | - | 2 |
| Conv-2 | $5 \times 5$ | 256 | - | 1 |
| MaxPooling | - | - | - | 2 |
| Conv-3 | $3 \times 3$ | 384 | - | 1 |
| Conv-4 | $3 \times 3$ | 384 | - | 1 |
| Conv-5 | $3 \times 3$ | 256 | - | 1 |
| MaxPooling | - | - | - | 2 |
| Fc1 | - | - | 2048 | - |
| Fc2 | - | - | 1000 | - |
| Fc3 | - | - | 2 or 5 | - |

The performance metrics were evaluated using 5-fold cross-validation for all the deep learning experiments to ensure the reliability of the results. The experimental results for 30 simulations of the 5-fold cross-validation experiments with deep learning classifiers are presented in Tables 3.3 and 3.4 for binary and multi-class classifications, respectively. GoogLeNet achieved the best classification accuracy of 99.9978 $\pm$ 0.0084 % in the case of binary classification, while the ResNet model recorded the best classification accuracy of 99.7222 $\pm$ 0.0686 % for multi-class classification. By contrast, VGG Net achieved the worst accuracies of 97.7822% and 98.8711% in both cases. It is also worth noticing that while GoogLeNet showed the most stable behavior in the case of binary

classification with a least standard deviation of 0.0084, ResNet exhibited the most stable performance in the case of multi-class classification, with a minimum standard deviation of 0.0686. Conversely, VGG Net showed highly unstable performance in both cases with a maximum standard deviation of 2.0350 and 1.5307 for binary classification and multi-class classification, respectively.

Table 3.3. Experimental results for 30 simulations of DL classifiers for binary class classification.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| CNN | $99.8778 \pm 0.1540$ | $99.9000 \pm 0.1618$ | $99.7889 \pm 0.3062$ | $99.9235 \pm 0.0965$ |
| AlexNet | $99.9633 \pm 0.1015$ | $99.9736 \pm 0.0762$ | $99.9222 \pm 0.4260$ | $99.9771 \pm 0.0630$ |
| GoogLeNet | $99.9978 \pm 0.0084$ | $99.9972 \pm 0.0106$ | $100.0000 \pm 0.0000$ | $99.9986 \pm 0.0053$ |
| VGG | $97.7822 \pm 2.0350$ | $99.5569 \pm 0.6264$ | $90.6833 \pm 10.0421$ | $98.792 \pm 1.1396$ |
| ResNet | $99.9811 \pm 0.0258$ | $99.9764 \pm 0.0322$ | $100.0000 \pm 0.0000$ | $99.9882 \pm 0.0161$ |

Table 3.4. Experimental results for 30 simulations of DL classifiers for multi-class classification.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| CNN | $99.3956 \pm 0.2274$ | $99.3956 \pm 0.2274$ | $99.8489 \pm 0.0568$ | $99.3964 \pm 0.2262$ |
| AlexNet | $99.6400 \pm 0.3362$ | $99.6400 \pm 0.3362$ | $99.9100 \pm 0.0841$ | $9.6410 \pm 0.3327$ |
| GoogLeNet | $99.6700 \pm 0.0903$ | $99.6700 \pm 0.0903$ | $99.9175 \pm 0.0226$ | $99.6702 \pm 0.0901$ |
| VGG | $98.8711 \pm 1.5307$ | $98.8711 \pm 1.5307$ | $99.7178 \pm 0.3827$ | $98.8840 \pm 1.5262$ |
| ResNet | $99.7222 \pm 0.0686$ | $99.7222 \pm 0.0686$ | $99.9306 \pm 0.1714$ | $99.7237 \pm 0.0681$ |

### 3.2.2. HOG feature vector-based classification

#### 3.2.2.1. Selection of the HOG parameters

The multi-class classification problem was solely investigated using the HOG parameter selection because it is more difficult compared to its binary counterpart. Also, the SVM classifier with radial basis function (RBF) kernel was utilized in this experimental parameter selection process.

HOG feature vector has a number of user-defined parameters including cell size (CellSize) and the number of histogram bins (NumBins). Multiple experiments were carried out to find the optimal values for these parameters. The first set of experiments was performed to find the best CellSize. This is done by fixing the NumBins to a value of $9$ and changing

the CellSize from $4 \times 4$ gradually up to $64 \times 64$. Results of these experiments are presented in Table 3.5.

Table 3.5. Experimental results for multi-class classification by varying the cell size (CellSize) at fixed number of histogram bins (NumBins) of 9

| CellSize | Accuracy (%) |
|---|---|
| [4 4] | 92.9333 |
| [8 8] | 95.2667 |
| [16 16] | 96.8000 |
| [32 32] | 97.6667 |
| **[64 64]** | **97.8667** |

The size of the input image at this stage is $128 \times 128$. Using a CellSize of $64 \times 64$ (only $4$ cells per image), the best classification accuracy of $97.8667\%$ was reported. Hence, the CellSize was set to $64 \times 64$ in all of the subsequent experiments.

The second set of experiments was carried out to find the best NumBins by setting CellSize to $64 \times 64$ and varying the NumBins between 12 and 31 as seen in Table 3.6. It can be noticed that the accuracy rate increases as the NumBins increases until it reaches a point where no further increase in the accuracy is observed ($NumBins = 30$). Therefore, 30 was selected to be the best number of histogram bins to reduce the computational load.

Table 3.6. Experimental results for multi-class classification by varying the number of histogram bins (NumBins) with fixed cell size (CellSize) of [64 64].

| NumBins | Length of feature vector | Accuracy (%) |
|---|---|---|
| 3 | 12 | 92.03 |
| 4 | 16 | 94.77 |
| 6 | 24 | 96.53 |
| 9 | 36 | 97.87 |
| 12 | 48 | 97.80 |
| 18 | 72 | 97.93 |
| 21 | 84 | 98.20 |
| 24 | 96 | 98.20 |
| **30** | **120** | **98.27** |
| **31** | **124** | **98.27** |
| 32 | 128 | 98.13 |

The performance metrics were evaluated using 10-fold cross-validation for all the machines learning experiments. Experimental results for 30 simulations of 10-fold cross-validation with SVM variants are present in Tables 3.7 and 3.8 for binary and

multi-class classification, respectively. All classification accuracies using SVM classifiers exceeded 99% for binary classification and most of them exceeded 98% for multi-class classification. SVM-RBF recorded the best classification performance of 99.91 $\pm$ 0.02% for binary classification, while SVM-Cubic achieved the best performance of 98.14 $\pm$ 0.10% for multi-class classification. By contrast, SVM-Linear achieved the worst classification performances of 99.72 $\pm$ 0.06 % and 92.55 $\pm$ 0.16 % for binary classification and multi-class classification, respectively.

Table 3.7. Experimental results of binary classification for 30 simulations of SVM variants.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| SVM-RBF | 99.91 $\pm$ 0.02 | 99.93 $\pm$ 0.02 | 99.82 $\pm$ 0.05 | 99.94 $\pm$ 0.01 |
| SVM-Linear | 99.72 $\pm$ 0.06 | 99.80 $\pm$ 0.14 | 99.39 $\pm$ 0.28 | 99.82 $\pm$ 0.04 |
| SVM-Quadratic | 99.90 $\pm$ 0.02 | 99.93 $\pm$ 0.02 | 99.81 $\pm$ 0.07 | 99.94 $\pm$ 0.01 |
| SVM-Cubic | 99.90 $\pm$ 0.03 | 99.92 $\pm$ 0.02 | 99.78 $\pm$ 0.09 | 99.93 $\pm$ 0.02 |

Table 3.8. Experimental results of multi-class classification for 30 simulations of SVM variants.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| SVM-RBF | 98.10 $\pm$ 0.10 | 98.10 $\pm$ 0.10 | 99.52 $\pm$ 0.03 | 98.10 $\pm$ 0.10 |
| SVM-Linear | 92.55 $\pm$ 0.16 | 92.55 $\pm$ 0.16 | 98.14 $\pm$ 0.04 | 92.55 $\pm$ 0.16 |
| SVM-Quadratic | 98.13 $\pm$ 0.10 | 98.13 $\pm$ 0.10 | 99.53 $\pm$ 0.03 | 98.13 $\pm$ 0.10 |
| SVM-Cubic | 98.14 $\pm$ 0.10 | 98.14 $\pm$ 0.10 | 99.54 $\pm$ 0.02 | 98.14 $\pm$ 0.10 |

Tables 3.9 and 3.10 present the experimental results for 30 simulations of 10-fold cross-validation using the KNN variants for binary classification and multi-class classification, respectively. KNN-Cos achieved the highest classification performance of 99.76 $\pm$ 0.03% for binary classification and the lowest classification performance of 98.67 $\pm$ 0.09% for multi-class classification. By contrast, KNN-Cor recorded the lowest classification performance of 99.73 $\pm$ 0.05% for binary classification and the highest classification performance of 98.75 $\pm$ 0.11% for multi-class classification.

Also, all of the classification accuracies for the KNN variants were above 99% for binary classification and above 98% for multi-class classification.

Table 3.9. Experimental results of binary classification for 30 simulations of KNN variants.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| KNN-Spearman | $99.74 \pm 0.05$ | $99.80 \pm 0.05$ | $99.50 \pm 0.08$ | $99.84 \pm 0.03$ |
| KNN-Cos | $99.76 \pm 0.03$ | $99.78 \pm 0.04$ | $99.67 \pm 0.07$ | $99.85 \pm 0.02$ |
| KNN-Cor | $99.73 \pm 0.05$ | $99.75 \pm 0.05$ | $99.65 \pm 0.12$ | $99.83 \pm 0.03$ |

Table 3.10. Experimental results of multi-class classification for 30 simulations of KNN variants.

| Classifier | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1 score (%) |
|---|---|---|---|---|
| KNN-Spearman | $98.16 \pm 0.13$ | $98.16 \pm 0.13$ | $99.54 \pm 0.03$ | $98.16 \pm 0.13$ |
| KNN-Cos | $98.67 \pm 0.09$ | $98.67 \pm 0.09$ | $99.67 \pm 0.02$ | $98.67 \pm 0.09$ |
| KNN-Cor | $98.75 \pm 0.11$ | $98.75 \pm 0.11$ | $99.69 \pm 0.03$ | $98.74 \pm 0.11$ |

# 4. CHAPTER

# Discussion, Conclusion and Future Work

## 4.1. Discussions

In this chapter, the results recorded in this thesis are discussed and a conclusion with suggestions for prospective future works is provided.

Several different steps were implemented to perform binary classification as well as multi-class classification for heart sound signals. These steps included data segmentation, feature extraction and classification experiments with ML and DL methods.

The multi-class classification problem is more complicated than the binary classification, and this is reflected in the results achieved. As observed, the accuracies reported for binary classification always surpassed those reported for its multi-class counterpart.

All of the experiments were repeated for 30 simulations and the average results along with the standard deviation of the simulations are presented. While 10-fold cross-validation was used for ML methods, 5-fold cross-validation was utilized for DL experiments due to the high computational time with re-running the DL experiments with such large cross fold experiments.

Scalogram images were employed for classification experiments with DL classifiers while scalogram-HOG feature vectors were employed for classification experiments with the ML classifiers.

Majority of the accuracies recorded using DL methods were above 99% while the worst accuracy was above 97%, confirming the robustness of the proposed classification method for heart sound classification using scalogram images and deep learning models.

Regarding the ML methods, all classification performances were above 99% for binary classification and above 98% for multi-class classification.

However, the maximum classification performance offered by DL based method using heart sound scalograms surpassed the best classification performace offered by ML-based method using scalogram-HOG features. Nevertheless, the size of the input feature vectors to the DL models (i.e the scalogram images) varied between $227 \times 227$ in the case of AlexNet and $224 \times 224$ in the case of VGG Net, GoogLeNet and ResNet models, and the depth of DL models resulted in long computational time when compared with the time taken using the simple ML models with a feature vector of length $120$.

Regarding the ML methodologies, while SVM offered the highest classification performance for binary classification, the KNN offered the best classification performance for multi-class classification.

Table 4.1 presents a comparison of the results obtained in this study for heart sound classification with other studies that used the same dataset in the literature. As seen from the table, the proposed method of computing scalograms of heart sound signals and feeding them to DL classifiers and the alternative method of calculating the HOG features of the scalogram images for experiments using DL classifiers both provide better heart sound classification performance compared to those reported in the literature.

Table 4.1. The performance comparisons of the proposed method with other studies in the literature.

| Problem | Study | Method | Accuracy(%) |
|---|---|---|---|
| Binary | Upretee et al. (2019) | Centroid frequency + KNN | 99.60 |
| | **This study** | Scalogram + GoogLeNet | **99.88** |
| | **This study** | **Scalogram + HOG + SVM** | **99.91** |
| Multi-class | Upretee et al. (2019) | Centroid frequency + KNN | 96.50 |
| | Yaseen et al. (2018) | MFCC + DWT feature | 97.9 |
| | **This study** | Scalogram + ResNet | **99.40** |
| | **This study** | **Scalogram + HOG + KNN-Cor** | **98.75** |

## 4.2. Conclusion and Future works

Cardiovascular diseases (CVDs) are the first cause of human death. Hence, it is necessary to address this problem and find solutions to reduce the death cases. The early detection of CVDs can play an important role in reducing the deaths caused by CVDs. Machine learning and deep learning have been used extensively in the literature to automate the process of CVDs diagnosis. wavelet scalograms was proposed in [26] for epilepsy diagnosis. Hence it is used in this thesis for automatic CVDs diagnosis. The highest classification performances of 99.99% and 99.72% for binary classification and multi-class classification were recorded using the scalogram with the deep learning methods. Another method based on the histogram of oriented gradient features extracted for scalogram images was also proposed. This method also achieved better performance than the existing methods in the literature that used the same dataset. Since the heart sound classification performance recorded in this study is superior to others recorded in all of the reviewed studies in the literature, it can efficiently be used in heart sound signal classification for cardiovascular disease diagnosis.

Future works may consider exploring other signal-to-image techniques such as utilization of trajectory matrices.

## REFERENCES


1. , 2017 . Word Health Organization, Cardiovascular diseases (CVDs) Fact sheets.

2. Son, G.Y., Kwon, S., et al., 2018. Classification of heart sound signal using multiple features, **Applied Sciences**, **8**(12):2344.

3. Czarny, M.J. and Resar, J.R., 2014. Diagnosis and management of valvular aortic stenosis, **Clinical Medicine Insights: Cardiology**, **8**:CMC–S15716.

4. Carabello, B.A., 2005. Modern management of mitral stenosis, **Circulation**, **112**(3):432–437.

5. Enriquez-Sarano, M., Nkomo, V.T., and Michelena, H.I., 2009. Mitral regurgitation, Valvular Heart Disease, 221–246, Springer.

6. Hayek, E., Gring, C.N., and Griffin, B.P., 2005. Mitral valve prolapse, **The Lancet**, **365**(9458):507–518.

7. Upretee, P. and Yüksel, M.E., 2019. Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: preliminary results, *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, 1–4, IEEE.

8. Huiying, L., Sakari, L., and Iiro, H., 1997. A heart sound segmentation algorithm using wavelet decomposition and reconstruction, *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)*, volume 4, 1630–1633, IEEE.

9. Nogueira, D.M., Zarmehri, M.N., Ferreira, C.A., Jorge, A.M., and Antunes, L., 2019. Heart sounds classification using images from wavelet transformation, *EPIA Conference on Artificial Intelligence*, 311–322, Springer.

10. Leung, T., White, P., Collis, W., Brown, E., and Salmon, A., 2000. Classification of heart sounds using time-frequency method and artificial neural networks,

*Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*, volume 2, 988–991, IEEE.

11. Turkoglu, I. and Arslan, A., 2001. An intelligent pattern recognition system based on neural network and wavelet decomposition for interpretation of heart sounds, *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, 1747–1750, IEEE.

12. Zin, Z.M., Hussain-Salleh, S., and Sulaiman, M.D., 2003. Wavelet analysis and classification of mitral regurgitation and normal heart sounds based on artificial neural networks, *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings.*, volume 2, 619–620, IEEE.

13. Ari, S. and Saha, G., 2008. Classification of heart sounds using empirical mode decomposition based features, **International Journal of Medical Engineering and Informatics**, **1**(1):91–108.

14. Vepa, J., 2009. Classification of heart murmurs using cepstral features and support vector machines, *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2539–2542, IEEE.

15. Safara, F., Doraisamy, S., Azman, A., Jantan, A., and Ramaiah, A.R.A., 2013. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification, **Computers in Biology and Medicine**, **43**(10):1407–1414.

16. Randhawa, S.K. and Singh, M., 2015. Classification of heart sound signals using multi-modal features, **Procedia Computer Science**, **58**:165–171.

17. Nilanon, T., Yao, J., Hao, J., Purushotham, S., and Liu, Y., 2016. Normal/abnormal heart sound recordings classification using convolutional neural network, *2016 Computing in Cardiology Conference (CinC)*, 585–588, IEEE.

18. Karar, M.E., El-Khafif, S.H., and El-Brawany, M.A., 2017. Automated diagnosis of heart sounds using rule-based classification tree, **Journal of Medical Systems**, **41**(4):60.

19. Deperlioglu, O., 2018. Classification of phonocardiograms with convolutional neural networks, **BRAIN. Broad Research in Artificial Intelligence and Neuroscience**, **9**(2):22–33.

20. Krishnan, P.T., Balasubramanian, P., and Umapathy, S., 2020. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network, **Physical and Engineering Sciences in Medicine**, 1–11.

21. , 2018 . World Health Organization, The Top 10 Causes of Death.

22. Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., and Sricharan, K., 2016. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients, *2016 Computing in cardiology conference (CinC)*, 813–816, IEEE.

23. Hu, W., Lv, J., Liu, D., and Chen, Y., 2018. Unsupervised feature learning for heart sounds classification using autoencoder, *Journal of Physics: Conference Series*, volume 1004, IOP Publishing.

24. Xiao, B., Xu, Y., Bi, X., Zhang, J., and Ma, X., 2020. Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption, **Neurocomputing**, **392**:153–159.

25. Daubechies, I., 1992. Ten lectures on wavelets, Society for Industrial and Applied Mathematics.

26. Türk, Ö. and Özerdem, M.S., 2019. Epilepsy detection by using scalogram based convolutional neural network from EEG signals, **Brain sciences**, **9**(5):115.

27. Li, B. and Chen, X., 2014. Wavelet-based numerical analysis: a review and classification, **Finite Elements in Analysis and Design**, **81**:14–31.

28. Falamarzi, Y., Palizdan, N., Huang, Y.F., and Lee, T.S., 2014. Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs), **Agricultural Water Management**, **140**:26–36.

29. Torrence, C. and Compo, G.P., 1998. A practical guide to wavelet analysis, **Bulletin of the American Meteorological society**, **79**(1):61–78.

30. Unser, M. and Aldroubi, A., 1996. A review of wavelets in biomedical applications, **Proceedings of the IEEE**, **84**(4):626–638.

31. Lilly, J.M. and Olhede, S.C., 2012. Generalized Morse wavelets as a superfamily of analytic wavelets, **IEEE Transactions on Signal Processing**, **60**(11):6036–6041.

32. Olhede, S.C. and Walden, A.T., 2002. Generalized morse wavelets, **IEEE Transactions on Signal Processing**, **50**(11):2661–2670.

33. Dalal, N. and Triggs, B., 2005. Histograms of oriented gradients for human detection, *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, 886–893, IEEE.

34. Mallick, S., 2016 . Histogram of Oriented Gradients.

35. Gornale, S.S., Patravali, P.U., Marathe, K.S., and Hiremath, P.S., 2017. Determination of osteoarthritis using histogram of oriented gradients and multiclass SVM., **International Journal of Image, Graphics & Signal Processing**, **9**(12).

36. Sahoo, P., Behera, A.K., Pandia, M.K., Dash, C.S.K., and Dehuri, S., 2013. On the study of GRBF and polynomial kernel based support vector machine in web logs, *2013 1st International Conference on Emerging Trends and Applications in Computer Science*, 1–5, IEEE.

37. Gareth, J., Daniela, W., Trevor, H., and Robert, T., 2013. An introduction to statistical learning: with applications in R, Springer.

38. Waseem, M., 2019. A quick guide to learn support vector machine in python.

39. Tan, P.N., Steinbach, M., and Kumar, V., 2016. Introduction to data mining, Pearson Education India.

40. Ronao, C.A. and Cho, S.B., 2016. Human activity recognition with smartphone sensors using deep learning neural networks, **Expert systems with applications**, **59**:235–244.

41. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M.S., 2016. Deep learning for visual understanding: A review, **Neurocomputing**, **187**:27–48.

42. Xia, Y., Wulan, N., Wang, K., and Zhang, H., 2018. Detecting atrial fibrillation by deep convolutional neural networks, **Computers in Biology and Medicine**, **93**:84–92.

43. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., 2018. Recent advances in convolutional neural networks, **Pattern Recognition**, **77**:354–377.

44. Wu, J., 2017. Introduction to convolutional neural networks, **National Key Lab for Novel Software Technology. Nanjing University. China**, **5**:23.

45. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., and Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, **Computers in Biology and Medicine**, **100**:270–278.

46. Kao, W.C. and Wei, C.C., 2011. Automatic phonocardiograph signal analysis for detecting heart valve disorders, **Expert Systems with Applications**, **38**(6):6458–6468.

47. Lin, M., Chen, Q., and Yan, S., 2013. Network in network, **arXiv preprint arXiv:1312.4400**.

48. Yani, M. et al., 2019. Application of transfer learning using convolutional neural network method for early detection of Terry's nail, *Journal of Physics: Conference Series*, volume 1201, 012052, IOP Publishing.

49. Arunava, 2018. Convolutional Neural Network.

50. Wang, S.H., Xie, S., Chen, X., Guttery, D.S., Tang, C., Sun, J., and Zhang, Y.D., 2019. Alcoholism identification based on an AlexNet transfer learning model, **Frontiers in Psychiatry**, **10**:205.

51. Xiang, Q., Wang, X., Li, R., Zhang, G., Lai, J., and Hu, Q., 2019. Fruit image classification based on MobileNetV2 with transfer learning technique,

*Proceedings of the 3rd International Conference on Computer Science and Application Engineering*, 1–7.

52. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S., and Asari, V.K., 2018. The history began from AlexNet: A comprehensive survey on deep learning approaches, **arXiv preprint arXiv:1803.01164**.

53. Kazimi, B., Thiemann, F., Malek, K., Sester, M., and Khoshelham, K., 2018. Deep learning for archaeological object detection in airborne laser scanning data, *2nd Workshop On Computing Techniques For Spatio-Temporal Data in Archaeology And Cultural Heritage (COARCH 2018)*, volume 2230, 21–35.

# CURRICULUM VITAE

## PERSONAL INFORMATION


| | | |
|---|---|---|
| **Name, Surname** | : | Afrah Elfatih Farah Malik |
| **Nationality** | : | Sudan |
| **Date and place of birth** | : | 22/02/1995 - Gezira, Sudan |
| **Telephone** | : | +90 552 623 48 12 |
| **Email** | : | afrah39frah@gmail.com |
| **Address** | : | Department of Biomedical Engineering<br>Erciyes University, 38039, Melikgazi Kayseri Turkey |


## EDUCATION

| Level | Institution | Graduation Date |
|---|---|---|
| Bachelor | Sudan University of Science and Technology, Khartoum, Sudan | 2016 |
| Masters | Erciyes University, Kayseri, Turkey. | 2020 |

## PUBLICATIONS

- Malik, A. E. F., Sezin BARIN, Yüksel, M. E., Accurate Classification of Heart Sound Signals for Cardiovascular Disease Diagnosis by Wavelet Analysis and Convolutional Neural Network: Preliminary Results, *2020 28th Signal Processing and Communications Applications Conference (SIU)*. **(Accepted)**

## LANGUAGES

- Arabic: fluent (Native speaker)
- English: upper intermediate: e-YDS Feb. 2020 score: 88.75/100.
- Turkish: C1 Turkish Certificate 2017 (ERSEM, Erciyes University).