Machine Learning with Applications 7 (2022) 100206

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Heart sound classification using signal processing and machine learning algorithms

Yasser Zeinali [a], Seyed Taghi Akhavan Niaki [b,*]

[a] Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran
[b] Department of Industrial Engineering, Sharif University of Technology, PO Box 11155-9414 Azadi Ave., Tehran 1458889694, Iran

ARTICLE INFO

Keywords:
Heart sound
Signal processing algorithms
Machine learning algorithms
Dimensional reduction algorithms
Feature selection
Gradient boosting classifier

ABSTRACT

According to global statistics and the world health organization (WHO), about 17.5 million people die each year from cardiovascular disease. In this paper, the heart sounds gathered by a stethoscope are analyzed to diagnose several diseases caused by heart failure. This research's primary process is to identify and classify the data related to the heart sounds categorized in four general groups of S_1 to S_4. The sounds S_1 and S_2 are considered as the heart's normal sounds, and the sounds S_3 and S_4 are the abnormal sounds of the heart (heart murmurs), each expressing a specific type of heart disease. In this regard, the desired features are first extracted after retrieving the data by signal processing algorithms. In the next step, feature selection algorithms are used to select the compelling features to reduce the problem's dimensions and obtain the optimal answer faster. While the existing algorithms in the literature classify the sound into two groups of normal and abnormal, in the final section, some of the most popular classification algorithms are utilized to classify the type of sound into three classes of normal, S_3 and S_4 categories. The proposed methodology obtained an accuracy rate of 87.5% and 95% for multiclass data (3 classes) and 98% for binary classification (normal vs. abnormal) problems.

1. Introduction

This research aims at using artificial intelligence algorithms to diagnose heart failure by classifying heart sounds. According to studies conducted at several large hospitals, heart specialists utilize different medical tests to diagnose heart disease accurately. However, heart sound diagnosis using a stethoscope is very difficult due to the hospitals' noisy environment (Leng et al., 2015). As such, physicians do not hear the heart sound very well and need some relevant tests to analyze patients' conditions.

Each of the two sides of a human heart has two chambers called the ventricles and the atrium, connected by some valves. The heart cycle refers to all the heart events from the beginning of one beat to the beginning of the next beat. This cycle is divided into two parts, which have two modes of contraction and rest called systole and diastole. While normal sounds S_1 and S_2 are received by a healthy heart, any disturbance, including narrowing and widening of the valves, can cause the heart to no longer pump efficiently and generate abnormal sounds S_3 and S_4 (Bonow et al., 2011). The heart's first sound (S_1) comes from the opening and closing of the tricuspid and mitral valves. The second heart sound (S_2) is due to the tracheal valves' opening and closing. The other two sounds are abnormal sounds caused by various reasons, such as congenital cavities between the left and right ventricles, sagging and clogging, or the valves' calcification. An example of the heart sounds signals classified in the above four categories are shown in Fig. 1. This paper aims to classify heart sounds into normal, abnormal type 3, and abnormal type 4.

The problem of classifying heart sounds is directly related to signal processing algorithms. Due to the nature of the raw data collected in audio with a sampling frequency of 2000 kHz, the data are first reviewed and preprocessed in this study. This means that the noise in the sounds is first passed through special filters based on their frequency and amplitude so that the noise can be eliminated and the sound quality can be improved. The second and more important goal is extracting features, which is done by implementing signal processing algorithms. In the next step, a dataset is obtained to form the classification algorithms' input using the extracted features.

The significant difference between this research and previous works available in the literature is considering three heart sound classes instead of two. More specifically, while in previous studies, the data were classified into two modes of normal and abnormal sounds, in the current research, the heart sounds are classified into three classes of normal, abnormal due to the third heart sound, and abnormal due to the fourth heart sound. As each of the abnormal sounds can cause different heart problems, this approach can analyze the sounds better to

* Corresponding author.
E-mail addresses: yasser.zeinali@gmail.com (Y. Zeinali), Niaki@Sharif.edu (S.T.A. Niaki).

https://doi.org/10.1016/j.mlwa.2021.100206
Received 30 March 2021; Received in revised form 29 October 2021; Accepted 31 October 2021
Available online 12 November 2021

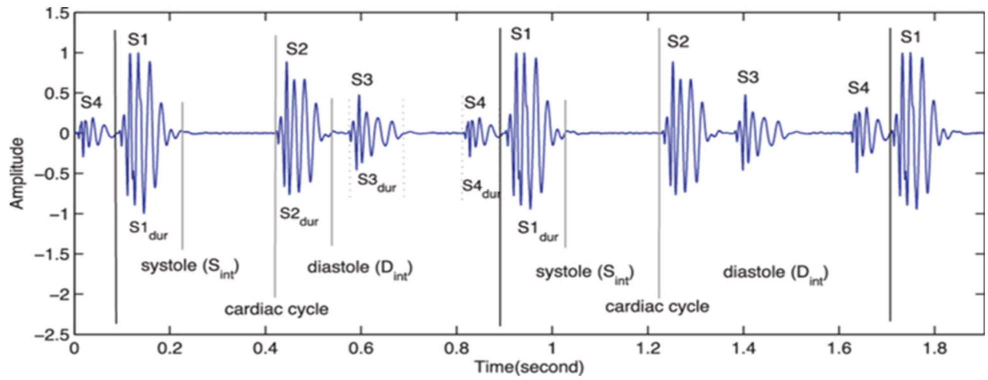

Fig. 1. Types of heart sounds.

make better decisions. Another contribution of this paper is to improve the accuracy of previous works on heart sound classifications.

The rest of this paper is organized as follows. Section 2 will review what has been done so far by machine learning algorithms to classify heart sounds and diseases. Section 3 explains the proposed methodology and its implementation in detail. Comparison and sensitivity analyses are performed in Section 4 to determine the strengths and weaknesses of each method. Finally, the best method is introduced, and future works are presented in Section 5.

2. Literature review

Long before the relatively new advent of artificial intelligence algorithms to classify heart sounds, various statistical studies have been performed in healthcare analytics. However, due to the great importance and the special view created by both engineers and physicians on this issue, significant progress has been made during a short period. In this section, the activities performed in heart sounds classification and heart disease diagnosis are explicitly examined. Then, data mining techniques and machine learning algorithms in various areas of health are briefly reviewed to explore the role of artificial intelligence in decision support systems.

Here, we gathered some of the research in heart sound classification in chronological order. Jia et al. (2012) tried to classify normal and abnormal sounds using fuzzy neural networks, which were initially extracted by features such as DWT and Shannon Entropy. The research used actual data collected by the researchers themselves. Deng and Han (2016) utilized the auto-correlation of sounds in the PASCAL dataset (one of the well-known datasets in the field of heart sound classification) and employed the discrete wavelet transform (DWT) technique for the feature extraction part. They also achieved a high accuracy using the SVM algorithm. However, the authors did not consider the use of signal filters to clean the dataset. Zabihi et al. (2016) conducted research that utilized an ensemble neural network algorithm to classify normal and abnormal heart sounds. In this study, the PhysioNet dataset was used, for which 91.5 percent accuracy was achieved. Besides, Potes et al. (2016) used deep learning to perform classification using AdaBoost and CNN algorithms, recording an accuracy rate of 94.24%. In this study, a sample consisting of 665 abnormal and 2575 normal sounds was used.

In 2017, research led to two articles' publication (Zhang et al., 2017a, 2017b). In the first article, the authors tried to classify heart sounds using spectrograms and a regression-based method, where the PASCAL dataset and the SVM algorithm were used. In the second article, the PhysioNet dataset was examined and analyzed, for which the feature extraction process was performed by the tensor decomposition method. They classified the dataset using the SVM algorithm. Another research conducted by Arabasadi et al. (2017) was about utilizing a hybrid neural network-Genetic algorithm. They proposed a method to improve the performance of the neural network by enhancing its initials weights. Using such methodology, they achieved accuracy, sensitivity, and specificity rates of 93.85%, 97%, and 92%, respectively. In another work, Dominguez-Morales et al. (2017) first classified the normal and abnormal sounds, and then, using electrical circuits and sensors, an automatic sound type detection system was invented. They used the deep convolutional neural network (DCNN) and recorded an accuracy of about 97%. In this study, the heart sounds were converted to sonogram shapes and placed as the input of the deep learning algorithms. Khateeb and Usman (2017) utilized kNN and Naïve Bayesian (NBs) algorithms on heart disease data, including 303 samples with 14 features. The researchers divided the test results into six categories, with the best results at 79.2% for kNN and 66.6% for NBs. Hamidi et al. (2018) employed curve fitting as a tool to extract features and a kNN algorithm to classify the normal and abnormal sounds of the heart, reaching an accuracy rate of 92%.

Recently, Noman et al. (2019) provided a framework for automatic heart sound detection based on neural network algorithms. The best performance was related to the combination of 1D-CNN and 2D-CNN with accuracy and sensitivity of 89.22% and 89.94%, respectively. Moreover, Zhang et al. (2019) used the long short-term memory (LSTM) network to classify sounds with an accuracy of 94.66 percent. They used the PhysioNet dataset as well. Kui et al. (2021) also utilized the dynamic frame length method to extract log Mel-frequency spectral coefficients (MFSC) features from the heart sound signal based on the heart cycle. Afterward, the convolution neural network (CNN) was used to classify the MFSC features. Finally, the majority voting algorithm was used to get the optimal classification results. Researchers know that patients do not know the device to understand the abnormal heart sounds; therefore, they proposed a system that enables remote patient monitoring by integrating advanced wireless communications with a customized low-cost stethoscope. A smartphone application also facilitates recording, processing, visualizing, listening to, and classifying heart sounds for patients and specialists. They built their classification model using the Mel-Frequency Cepstral Coefficient and Hidden Markov Model and tested the application in a hospital environment. The smartphone application correctly detected 92.68% of abnormal heart conditions in clinical trials at UT Southwestern Hospital (Thiyagaraja et al., 2018). Another research that utilized a new algorithm to classify the heart sounds was (Bilal, 2021). He proposed a model based on Local Binary Pattern (LBP) and Local Ternary (LTP) Pattern features and deep learning. He then used these methods to extract features from the heart sounds and feed them to a Dimensional Convolutional Neural Network (1D-CNN) to complete the classification process. Experiments were done with the help of two popular datasets that were used in the context to determine the efficiency of distinct techniques. These datasets are PASCAL and PhysioNet 2016. He obtained 91.66% and 91.78% accuracy rates for the datasets, respectively.

One of the most common diseases, especially among women, is breast cancer, which has been the subject of many articles and researches for a long time. Chen and Yang (2012) analyzed a breast cancer dataset from 97 patients using support vector machine (SVM) algorithms and a combination of genetic and SVM algorithms. Joshi and Mehta (2018) used the well-known kth nearest neighbor (kNN) algorithm to investigate breast cancer data's function. These data included 569 samples with 32 features. They used principal component analysis (PCA) and linear discriminant analysis (LDA) to reduce the data dimension. The result showed that kNN with LDA- reduction technique was better than kNN without dimension reduction and kNN with PCA. Their results were accurate 97.06%, 95.29%, and 95.88% of the time, respectively. Septiani et al. (2017) employed the kNN algorithm and tested 670 patients with nine features to achieve an accuracy of about 98 percent. Researchers are also trying to diagnose cancer by examining medical photographs and image processing algorithms. Kaymak et al. (2017) used 176 breast cancer images of benign and malignant cancers. They utilized a recurrent neural network technique and a neural network with the radial kernel to achieve 59% and 70% accuracy.

In addition to breast cancer, many diseases, including lung cancer, have been studied and analyzed by researchers. Kaucha et al. (2017) performed image processing on lung images using the K-means clustering approach in the image segmentation stage and finally using the SVM algorithm to diagnose cancer. They reached a 95.16% accuracy rate. Recently, Mousavi et al. (2021) proposed an intelligent classification algorithm comprising a fuzzy rule-based approach, a harmony search (HS) algorithm, and a heuristic to classify medical datasets intelligently. They used nine well-known medical datasets to evaluate the efficiency of their proposed approach. Maleki et al. (2021) used the kNN algorithm with an optimized k on a dataset, in which a genetic algorithm extracted its essential features. In the end, their result illustrated that their implemented technique gave 100% accuracy in diagnosing the stage of disease in lung cancer patients.

The above brief review is just a few previous activities on implementing artificial intelligence algorithms to predict, diagnose, or classify diseases. Obviously, by the use of more accurate data, more critical and better results can be achieved. This requires more cooperation between the health care sector and the engineering departments. In the next section, the performances of these algorithms in the field of heart sounds classification are examined to gain a deeper understanding of the subject.

3. Methodology and implementation

The proposed methodology's general framework is shown in Fig. 2, for which each box will be explained in this section separately. Before doing this, the dataset is introduced first.

3.1. The dataset

The dataset used in this study is part of the PhysioNet and PASCAL challenge data collected carefully by a physician (https://physionet.org/physiobank/database – http://www.peterjbentley.com/heartchallenge). It contains 650 samples, of which 104 patients with abnormal and 52 patients with normal heart sounds are selected to create a dataset consisting of 156 heart sound samples (Bentley et al., 2011; Liu et al., 2016).

3.2. Data-preprocessing

As the heart sounds in the dataset are generally recorded in noisy environments with a relatively large number of patients and medical staff, a heart sound specialist preprocessed them to obtain better quality data before using them in the experiment. It should be noted that the heart works in the vicinity of the human lung and is naturally accompanied by the sound of the lungs or human respiration, which causes a high-frequency sub-sound to be present. Fig. 3 is an illustration of a noisy heart sound. In this regard, one needs to eliminate the noises using different signal filters, such as the Savitzky–Golay filter (Potes et al., 2016) used in the current research.

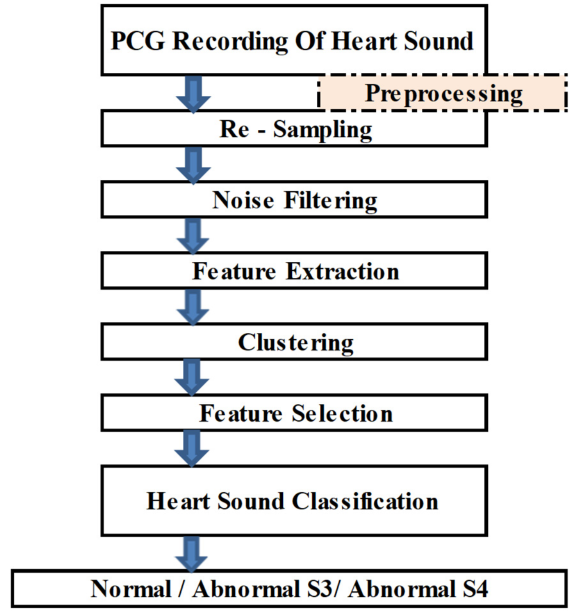

Fig. 2. The general framework of the proposed methodology.

3.3. Feature extraction

A feature is a measurable characteristic of a phenomenon that is observed. As the phenomenon possesses different features, when they are extracted together, a feature vector is obtained. For instance, if x_1 denotes the weight, x_2 is the height, …, and x_n is the sex of a person, then the feature vector is denoted by $\boldsymbol{X} = [x_1, x_2, \ldots, x_n]$. To extract features from the heart sounds, since the average human heart cycle is 0.8 s, each of the four heart sounds described previously is divided into five separate ranges, each with an average duration of 0.16 s. Consequently, the dataset created consists of features obtained in 5 different ranges. The feature extraction procedure of the current work is explained using Fig. 4. Each of the boxes shown in this figure is explained in the following subsections. Note here that another dataset that contains only the Mel-Frequency Cepstral Coefficients (MFCC) (Hasan et al., 2004) is also created to compare the performance of the proposed approach to the ones available in the literature. This will be discussed later.

3.3.1. Statistical features

The statistical features include standard deviation, skewness, and kurtosis. They are used to examine how the data is distributed.

3.3.2. Signal features

The signal features include amplitude and dominant frequencies. Amplitude is the maximum displacement or distance made by a point on a wave measured from its equilibrium position. Besides, as the lowest frequency component is known as the fundamental frequency, the dominant frequency is the fundamental frequency with the highest amplitude (Giron-Sierra, 2016). For instance, the fundamental frequency of about 450 Hz in Fig. 5 is the dominant frequency.

3.3.3. Wavelet features

A wavelet series represents a real- or complex-valued function by a particular orthonormal series generated by a wavelet. Wavelet transformation is one of the most important mathematical transformations used in various fields of science. The main idea of the wavelet transformation is to overcome the weaknesses and limitations of the Fourier transformation. Unlike the Fourier transformation, this transformation can be

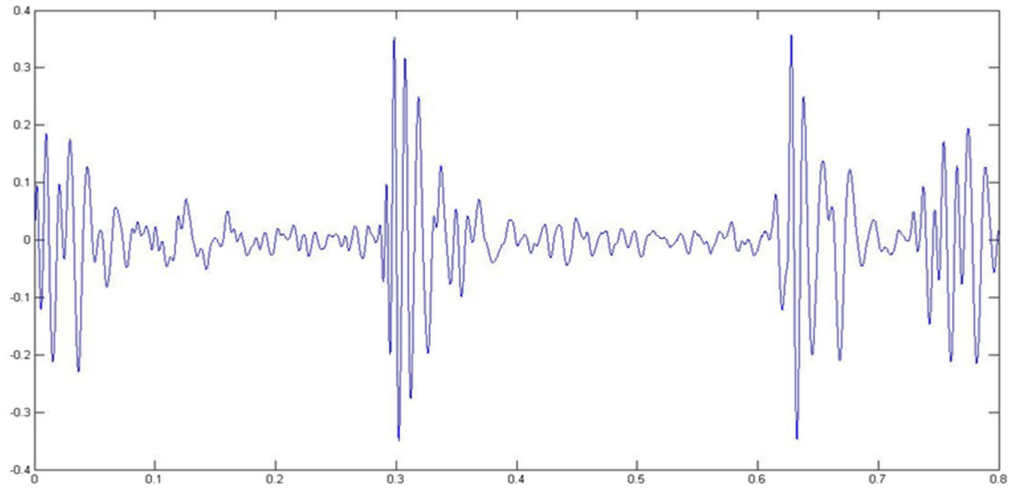

Fig. 3. Noisy heart sound.

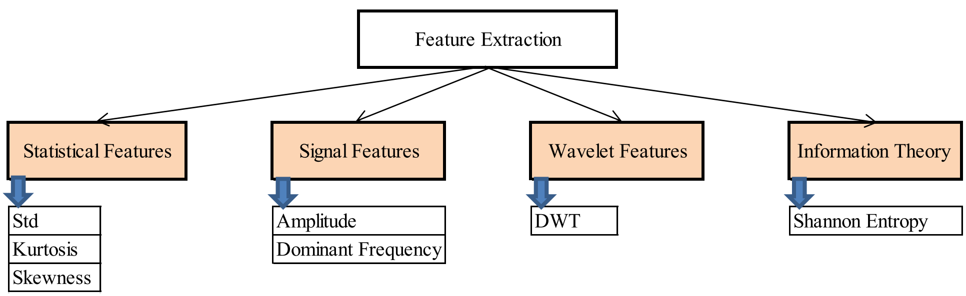

Fig. 4. Feature extraction process.

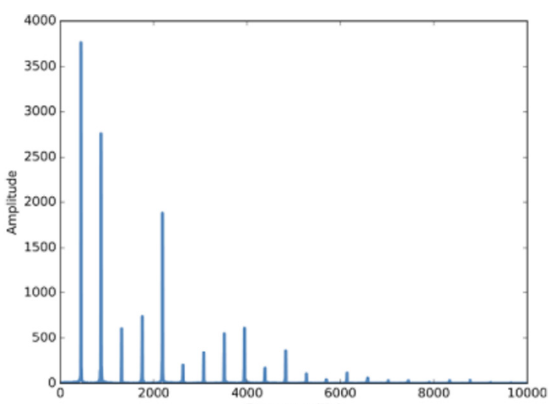

Fig. 5. Amplitude and frequency of a signal.

used for non-stationary signals (Akansu et al., 2010). Many of the real-world signals have non-stationary nature. In practical projects, when one tries to process ECG signals, stock market data, sensors data, and so on, she/he is more likely to encounter non-stationary signals of some dynamic systems. An excellent solution for processing non-stationary signals is the use of wavelet transformations instead of Fourier transformations. In this study, the discrete wavelet transform (DWT) approach is utilized for the heart sound data that is non-stationary.

The discrete wavelet transform (DWT) is a wavelet transform for which the wavelets are discretely sampled. The most commonly used set of DWTs was formulated by the Belgian mathematician Ingrid Daubechies in 1988. This formulation is based on utilizing recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. This transformation is employed in the current work to extract wavelet features from the processed heart sounds.

3.3.4. Information theory

Information theory is the mathematical treatment of the concepts, parameters, and rules governing messages transmission through communication systems (Smelser & Baltes, 2001). Although there are many different concepts and techniques related to this area, the key measure is entropy. The entropy of a discrete random variable X with the mass probability function $p(x)$ is denoted by $H(x)$ and is defined as:

$$H(x) = E(I(x)) = -log_b(p(x)) \tag{1}$$

While the logarithmic basis b in Eq. (1) can be 2, e, or 10, which calculates the entropy unit in a bit, nat, and Hartley, respectively, the most common basis to measure information is shown in Eq. (2) that obtains the information in the bit or Shannon unit (Cover & Thomas, 2012).

The Mel Frequency Cepstral Coefficients (MFCCs) are widely used in automatic speech and speaker recognition. They were introduced by Davis and Mermelstein in the 1980s and have been state-of-the-art ever since (Hasan et al., 2004). The formulae to convert the frequency to Mel scale are

$$M(f) = 1125\ln(1 + {}^{f}/_{700}) \tag{2}$$

$$\Delta f_{mel} = {}^{mel(f_{max})}/_{L} \tag{3}$$

$$c(l) = 700\left[10^{\frac{l\Delta f_{mel}}{2595}} - 1\right] l = 1, 2, \ldots, L. \tag{4}$$

In the above equations, f is the frequency given in Hertz, Δf_{mel} is determined according to the upper limit frequency f_{max} and the number of filters L, and $c(l)$ is the center frequency of the filter.

3.4. The feature selection algorithms

Feature selection algorithms are used to deal with high-dimensional data. These algorithms can be defined as the process of identifying relevant features and eliminating unrelated and repetitive features to observe a subset of traits that best describe the problem. Selecting a subset of data can reduce the data size and storage space required, affecting the processing time.

Feature selection algorithms have a wide range, including statistical-based, artificial intelligence-based, and meta-heuristic algorithms. Statistical-based algorithms such as forward selection, backward selection, hybrid models, and even the use of P-value are among the widely-used algorithms. In the field of artificial intelligence-based algorithms, one can name genetic algorithms (GA), particle swarm optimization (PSO), simulated annealing (SA), etc. One of the most advanced algorithms for feature selection is GA. GA is a stochastic method for function optimization based on the mechanics of natural genetics and biological evolution. Some advantages of genetic algorithms are the following:

- They outperform traditional feature selection methods.
- GAs handle datasets with many features.
- They do not need specific knowledge about the problems being studied.

After creating a set of data obtained from the heart's sounds in 156 rows and 65 columns with a total of 10 140 data, in the current research, the desired features are selected using the well-known GA due to its performance and applicability. The loss function used in this algorithm to be minimized is the weighted sum of the miss-classification ratio (mcr) and the number of selected features (n_f) shown in Eq. (5).

$$Min\, Z = w_1 * mcr + w_2 * n_f. \tag{5}$$

Using this objective function, GA tries to find the best combination with the minimum number of features that minimize both the cost and the misclassification rate. Here, the stopping criterion to end the iterations is chosen to be a predefined number of iterations.

If we divide the equation by w_1, we will have:

$$Min\, Z = mcr + {}^{w_2}/_{w_1} * n_f. \tag{6}$$

Assuming ${}^{w_2}/_{w_1} = W$. Therefore, the objective function becomes:

$$Min\, Z = mcr + W * n_f. \tag{7}$$

Now, this W can be defined as:

$$W \propto mcr \rightarrow W = \beta * mcr \rightarrow Min\, Z = mcr + \beta * mcr * n_f \tag{8}$$

Finally, the objective function will be like this:

$$Min\, Z = mcr(1 + \beta * n_f). \tag{9}$$

β can be defined as a penalty for having an additional feature $(0 \leq \beta \leq 1)$.

3.5. The dimension reduction algorithms

The performance of machine learning algorithms degrades when there are too many input variables. Having a large number of dimensions in the feature space implies a substantial space volume, and in turn, the points in space (rows of data) often represent a small and non-representative sample. This problem that can dramatically impact machine learning algorithms' performance is known as the "curse of dimensionality". In this study, two algorithms, namely the principal components analysis (PCA) and the linear discriminant analysis (LDA), reduce the data dimension.

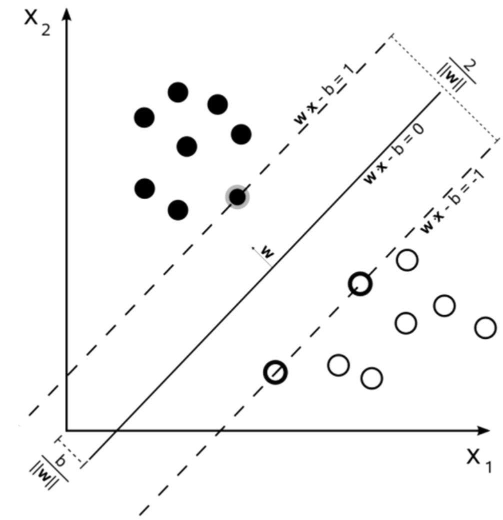

Fig. 6. Classifying by SVM.

PCA is a simple and efficient linear transformation method. The primary purpose of PCA analysis is to recognize patterns in data by identifying the variables' correlations. If there is a strong correlation, efforts to reduce the dimensions will be significant. In general, PCA finds the maximum direction of variance in high-dimensional data and plots it in sub-dimensions with fewer dimensions to retain most of the information.

LDA aims to project a dataset onto a lower-dimensional space with good class separability to avoid overfitting and reduce computational costs. The general LDA approach is very similar to a PCA, with the difference that in addition to finding the component axes that maximize the variables' variance, finding the axes that maximize the separation between multiple classes is also aimed.

3.6. Machine learning algorithms

The last step of this study is to cluster and classify the dataset. To this aim, while three well-known clustering algorithms, including K-means, DBscan, and hierarchical clustering algorithms, are used for clustering, support vector machines classifier (SVC), gradient boosting classifier (GBC), and random forest classifier (RFC) algorithms are utilized in the current research to classify the sounds. The classifiers are discussed in the following subsections.

3.6.1. Support vector machine algorithm

The support vector machine (SVM) is one of the most powerful machine-learning algorithms. It is a supervised machine learning algorithm that can be used for both classification and regression problems (Huang et al., 2006). This algorithm first takes the data to a higher spatial dimension so that it can create a distance between them by one or more hyperplanes. This can generally be used to draw multiple lines between data as decision lines; among them, the line that reduces the risk of categorization is found. Fig. 6 depicts this approach.

Suppose that the data used in the model to learn is: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where the y_i are either 1 or −1, each indicating the class to which the point x_i belongs. We want to find the "maximum-margin hyperplane" that divides the group of points x_i for which $y_i = 1$ from the group of points for which $y_i = -1$, which is defined so that the distance between the hyperplane and the nearest point from either group is maximized. A hyperplane can be written as:

$$W^T x - b = 0 \tag{10}$$

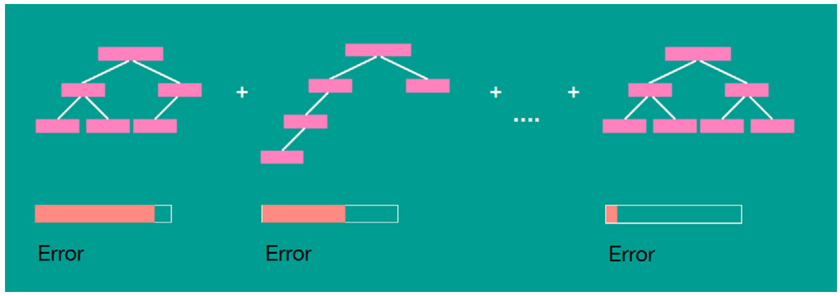

Fig. 7. Updating a model using a gradient boosting method.

where W is the vector to the hyperplane. In this equation, the parameter $\frac{b}{\|W\|}$ determines the offset of the hyperplane from the origin along the normal vector W.

3.6.2. Gradient boosting algorithm

Gradient boosting algorithm (GBA) is an ensemble algorithm that provides a high-efficient prediction or classification when dealing with large amounts of data. The algorithm combines the predictions of several basic estimators to improve performance. It includes several weak or medium estimators that together create a strong classifier or regressor. The algorithm possesses a boosting-based reinforcement that seeks to minimize the prediction error and, therefore, minimizes this error by adding new models (Mason et al., 1999). GBA can optimize different loss functions and provide several hyper-parameter tuning options that make the function very flexible.

Suppose that the data used in the model to learn is as follows: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ And the goal of learning is to minimize a loss function called $L(y, F(x))$ defined as

$$F = argmin\ \mathrm{E}_{x,y}[L(y, F(x))]. \tag{11}$$

This process is performed in iterations to find the final model as:

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + F_0 \tag{12}$$

Here h_is are the models selected from a group of models called H, for example. This set can be a collection of decision trees, where the first model is a fixed number called F_0 selected as follows:

$$F_0 = argmin \sum_{i=1}^{n} L(y_i, \gamma). \tag{13}$$

Fig. 7 illustrates how the model is updated using the gradient boosting method.

3.6.3. Random forest algorithm

The random forests algorithm (RFA) or random decision forests algorithm is created by aggregating a multitude of decision trees. This algorithm is an ensemble learning method for classification, regression, and other tasks. Random forest works very well with high-dimensional data since it considers subsets of data. That is why it is faster than decision trees in the training phase, in which hundreds of features can be easily handled. As the name implies, this algorithm generates a forest as a group of decision trees randomly. The forest's construction is often performed using the Bagging method, which combines learning models to improve the overall performance (Breiman, 2001). The random forest algorithm adds randomness to the model as the trees grow. This leads to more variety and ultimately a better model. One of the most important advantages of this algorithm is the lack of overfitting, the main problem of many algorithms in this field. The general framework of the RFA is depicted in Fig. 8.

The training algorithm for random forests applies the general bootstrap aggregating or bagging technique to tree learners. Suppose that the data used in the model to learn is as follows: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

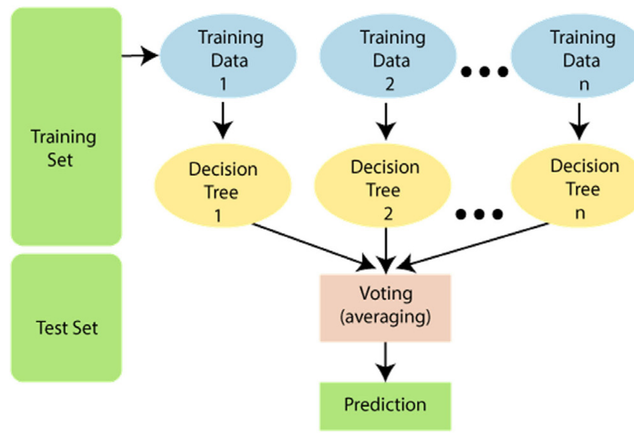

Fig. 8. The general framework of the random forest algorithm.

Bagging repeatedly selects a random sample with replacement of the training set and fits trees to these samples:

For $r = 1, 2, \ldots, R$ different random samples:

1. We provide samples with replacement (n training examples from X, Y; call these X_r, Y_r)
2. Train a classification or regression tree T_r on X_r, Y_r.

After training, predictions for the test set x_{test} can be made by averaging the predictions (or by taking the majority vote in the case of classification trees) from all the individual regression trees on x_{test} :

$$\hat{T} = {}^{1}/_{R} \sum_{r=1}^{R} T_r(x_{test}) \tag{14}$$

3.7. Performance criteria

The performance of a classifier can be measured through some factors, all of them gained from a confusion matrix. The confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. Table 1 demonstrates a confusion matrix for which the terms used are defined as follows.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{15}$$

$$Sensitivity = {}^{TP}/_{(TP+FN)} \tag{16}$$

$$Specificity = {}^{TN}/_{(TN+FP)} \tag{17}$$

$$Precision = {}^{TP}/_{(TP+FP)} \tag{18}$$

$$F = 2 * {}^{(Precision*Sensitivity)}/_{(Precision+Sensitivity)} \tag{19}$$

4. Feature extraction procedure

After explaining various steps involved in using the algorithms in the previous sections, the algorithms are implemented on a heart sound example by going through the steps in this section. The first three features, which are statistical features related to data distribution (standard deviation, skewness, and kurtosis) of one sample, are as follows:
Standard Deviation = [0.040656173, 0.043414002, 0.052018143, 0.052094222, 0.055029247]
Skewness = [0.288743681, 0.184201601, 0.035266239, 0.104826682, 0.117545241]
Kurtosis = [6.092877232, 4.431667889, 5.159243178, 5.221625844, 5.170539205]

In the next step, the dominant frequencies of all audio windows are examined. To this aim, some of the highest amplitudes of a frequency as signal peaks are first extracted as follows:
Amplitude_Section_1 = [0.015862133, 0.024666118, 0.027113745]
Amplitude_Section_2 = [0.038273228, 0.028721747, 0.007619529]
Amplitude_Section_3 = [0.022935687, 0.012337156, 0.028096329]
Amplitude_Section_4 = [0.017190622, 0.00700349, 0.018721624]
Amplitude_Section_5 = [0.011881425, 0.021728186, 0.016036953]

Then, the frequencies of each of the peaks specified above are the following dominant frequencies that will be used as inputs for the machine learning algorithms.
Dominant Fequency_Section_1 = [18, 11, 23]
Dominant Fequency_Section_2 = [11, 18, 12]
Dominant Fequency_Section_3 = [13, 20, 14]
Dominant Fequency_Section_4 = [13, 24, 8]
Dominant Fequency_Section_5 = [30, 8, 24]

The next part of the extraction process is the features related to wavelet transformations. In this study, after testing different wavelets, the Daubechies wavelet is chosen to extract the features. The wavelet coefficients in the sample signal are as follows:
Wavelet Coef_Section_1 = [0.026039559, 0.029882816, 0.043024967]
Wavelet Coef_Section_2 = [0.044132106, 0.051816736, $4.13 * 10^{-7}$]
Wavelet Coef_Section_3 = [$9.92 * 10^{-7}$, $2.40 * 10^{-6}$, $1.02 * 10^{-6}$]
Wavelet Coef_Section_4 = [$1.72 * 10^{-5}$, $8.02 * 10^{-8}$, $1.25 * 10^{-7}$]
Wavelet Coef_Section_5 = [$1.23 * 10^{-7}$, $1.40 * 10^{-7}$, $1.88 * 10^{-7}$]

The last step in the feature extraction section is devoted to Shannon Entropy, for which the coefficients are:
Shannon Entropy_Section_1 = [150.100672514795]
Shannon Entropy_Section_2 = [174.271610942798]
Shannon Entropy_Section_3 = [228.377602886118]
Shannon Entropy_Section_4 = [227.738726560246]
Shannon Entropy_Section_5 = [249.600061896928]

As mentioned previously, another dataset is also created using the features extracted by MFCC, for which the coefficients for the sample audio are as follows:

Mel-Frequency Cepstral Coefficients =

[−487.6704903 81.35004869 66.328044 47.3091901 30.41498369
19.94921547 16.82941564 18.71784134 21.63196296 22.11459924
18.83711975 12.89583605 6.818045198 2.972303966 2.303917322
4.023118871 6.282943959 7.341897362 6.500594951 4.317696586
2.083883542 0.957989067 1.306706764 2.599891964 3.846936426
4.245269912 3.627619682 2.467131116 1.495832325 1.217654846
1.625837376 2.273160383 2.609815599 2.34730993 1.617275924
0.847753671 0.464150304 0.619766938 1.121765609 1.580701382]

According to the extracted features from heart sounds and the division of each sample into 5 parts (0.8 s for each part), a total of 65 columns of features as a set of statistical features, features related to amplitude and frequency, and features included in wavelet and Shannon entropy is found for each sound. In the next step, the abnormal sounds are clustered into S_3, and S_4 sounds using the K-means, DBscan, and hierarchical clustering algorithms. Fig. 9 shows the flowchart of the proposed clustering approach.

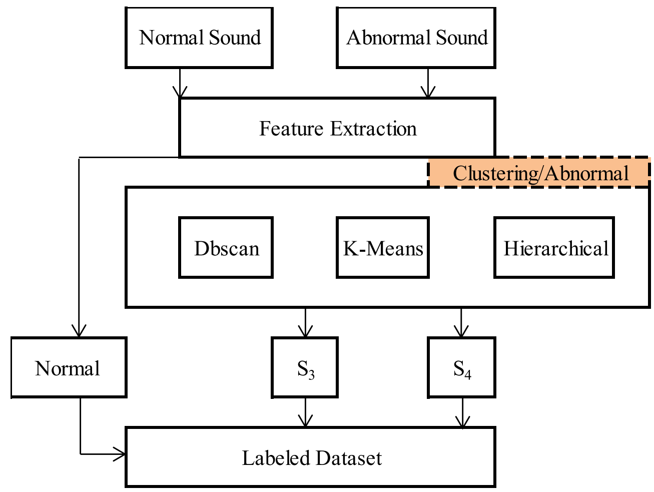

Fig. 9. The clustering approach.

Table 1
The confusion matrix.

		True Class: Positive	True Class: Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Table 2
The confusion matrix of the gradient boosting algorithm.

		Predicted classes		
		Normal	Abnormal S_3	Abnormal S_4
Real classes	Normal	4	1	1
	Abnormal S_3	0	5	0
	Abnormal S_4	0	0	5

5. Results

After creating the labeled dataset, the classification algorithms were utilized before and after executing the dimension reduction and the feature selection algorithms. Before implementing the dimension reduction and the feature selection algorithms, the algorithms' outcomes, including the confusion matrices and the performance measures in terms of precision, recall, and F1-score, were shown in Tables 2–7. As seen in these tables, the gradient boosting algorithm performed the best in terms of all performance measures. Besides, the best accuracy was achieved by using the gradient boosting algorithm with an accuracy of 87.5%, the number of estimators for this algorithm was 1000, and the learning rate was 0.01. This algorithm classified 14 out of 16 possible sounds correctly according to its confusion matrix. Meanwhile, the random forest (with 800 trees) and the support vector machine (with radius basis function kernel) algorithms had 81.25% and 75% accuracy.

The algorithms were also implemented after utilizing the principal component analysis and linear discriminant analysis to reduce dimensions. The first step in implementing PCA was the determination of the number of components. The variance shares of the dataset explained by the components are shown in Fig. 10. A good result was achieved by

Table 3
The performance of the gradient boosting algorithm.

	Performance criteria			
	Precision	Recall	F1-Score	Time
Normal	1	0.67	0.8	1.1032 s
Abnormal S_3	0.83	1	0.91	
Abnormal S_4	0.83	1	0.91	
Accuracy	87.50%			

Table 4
The confusion matrix of the random forest algorithm.

		Predicted classes		
		Normal	Abnormal S_3	Abnormal S_4
Real classes	Normal	4	1	1
	Abnormal S_3	0	5	0
	Abnormal S_4	0	1	4

Table 5
The performance of the random forest algorithm.

	Performance criteria			
	Precision	Recall	F1-Score	Time
Normal	1	0.67	0.8	1.0451 s
Abnormal S_3	0.71	1	0.83	
Abnormal S_4	0.8	0.8	0.8	
Accuracy	81.25%			

Table 6
The confusion matrix of the support vector machine algorithm.

		Predicted classes		
		Normal	Abnormal S_3	Abnormal S_4
Real classes	Normal	2	1	3
	Abnormal S_3	0	5	0
	Abnormal S_4	0	1	4

Table 7
The performance of the support vector machine algorithm.

	Performance criteria			Time
	Precision	Recall	F1-Score	
Normal	1	0.5	0.67	0.7554 s
Abnormal S_3	0.71	1	0.83	
Abnormal S_4	0.67	0.8	0.73	
Accuracy		75.00%		

considering 36 components that explain 99% of the dataset variance. As the maximum number of components in LDA is equal to the number of classes, which is 3, three components were used to implement this algorithm. In the case of using GA, after implementing the algorithm, it gave us 21 features. We then deployed the ML algorithms to analyze the results on each of the datasets we obtained after dimension reduction algorithms. The shape of these datasets were (156, 36), (156, 3), and (156, 21) based on PCA, LDA, and GA, respectively.

The next step was to implement the ML algorithms, including gradient boosting classifier (GBC), support vector classifier (SVC), and random forest classifier (RFC) alongside the GA that selects appropriate features. The outcomes (accuracy) are shown in Table 8. The results in Table 8 revealed that the best performance was achieved when the RFC or the SVC algorithm alongside GA with 78% accuracy was employed. Moreover, the outcomes showed that the accuracy decreases sharply when the dimension reduction algorithms (PCA or LDA) are used.

Finally, the Mel Frequency Cepstral Coefficients (MFCC) dataset was used to compare the outcomes with the ones obtained in the literature in classifying normal and abnormal sounds (for 2 and 3 classes) of the heart. Table 9 contains the implementation results.

For the heart sound classified into three classes, the outcomes in Table 9 showed that the best accuracy was 95% when the gradient boosting algorithm was used to classify the sound. In the two-class classification, the best result was 98% in terms of accuracy. Once again, note in this table that the dimension reduction approach using PCA, LDA, and GA not only did not add to the accuracy of the three classification algorithms but also decreased them. The purpose of using dimension reduction algorithms was to show the procedure and results theoretically.

6. Conclusion and future work

Artificial intelligence in healthcare management is promising, especially in diagnosing different diseases such as cardiovascular problems. Many scientists and researchers have worked in this field and achieved accurate results in many cases in recent years. This article utilized signal processing and machine learning algorithms to classify heart sounds into different classes. While in previous works, researchers classified heart sounds into two categories (normal and abnormal), in the current study, we used the available techniques in data analytics to divide heart sounds into three classes (normal, abnormal of the third type, and abnormal of the fourth type). This enables heart specialists to detect cardiovascular disease more accurately. We should mention that the analysis was based on stationary and non-stationary signals for normal and abnormal heart sounds data. Due to the nature of the raw data collected with a sampling frequency of 2000 kHz, the data were first reviewed and preprocessed under the supervision of a specialist. The analysis was made by extracting various heart sound information using statistical, discrete wavelet, and information theory features. Then we reduced the dimension of the dataset obtained by PCA, LDA, and GA by selecting appropriate features and classifying the sound using classification algorithms such as GBC, RFC, and SVC. The analysis and the codes related to the feature extractions and GA were written in MATLAB. We also utilized Python packages (Pandas, Numpy, Matplotlib, Scikitlearn, Scipy, etc.) for all the classification algorithms and data analysis. We also provided extensive comparative analysis to come up with the best technique in our experiments. The comparisons were made by performing all the processes and procedures on sets of preprocessed heart data.

Although various types of digital stethoscopes are available to the medical community, none can analyze sounds and differentiate the disease. As such, examining and analyzing heart sounds with the help of artificial intelligence algorithms can be a basis to build a device in the future to help heart specialists to make better decisions. The concepts of the Internet of Things (IoT) can also be combined with the concepts mentioned in this study to provide a basis for creating such a device in the future.

CRediT authorship contribution statement

Yasser Zeinali: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Validation. **Seyed Taghi Akhavan Niaki:** Conceptualization, Methodology, Visualization, Supervision, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

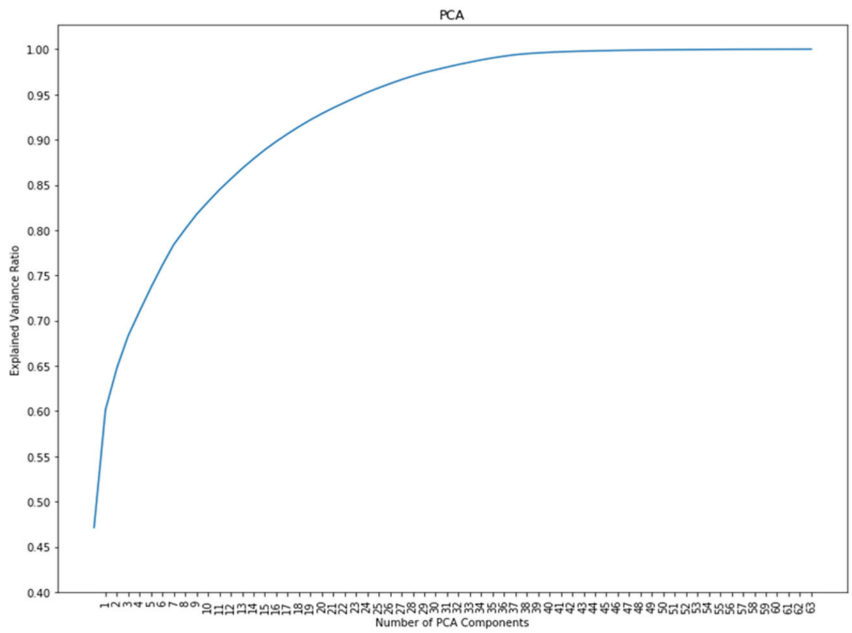

Fig. 10. Explained variance ratio of the PCA components.

Table 8
The accuracy of ML algorithms.

Main feature extraction	ML algorithms performance		
Experiments	GBC/Time	RFC/Time	SVC/Time
3 Classes + PCA	50.00% – 0.9728 s	61.10% – 0.8954 s	61.10% – 0.6179 s
3 Classes + LDA	56.00% – 0.5164 s	56.00% – 0.5262 s	56.00% – 0.3457 s
3 Classes + GA	67.00% – 1.0587 s	78.00% – 0.9130 s	78.00% – 0.7306 s

Table 9
The outcomes of the ML algorithms on the MFCC dataset.

MFCC feature selection	ML algorithms performance		
Experiments	GBC/Time	RFC/Time	SVC/Time
3 Classes	95.00% – 1.4874 s	88.00% – 1.349 s	88.00% – 1.1076 s
3 Classes + PCA	74.00% – 1.2854 s	77.00% – 1.3181 s	91.00% – 0.9856 s
3 Classes + LDA	68.00% – 0.7815 s	70.00% – 0.8701 s	84.00% – 0.4334 s
3 Classes + GA	74.00% – 1.2561 s	78.25% – 1.1027 s	81.20% – 0.9945 s
2 Classes	98.00% – 0.9667 s	94.00% – 0.9812 s	82.00% – 0.9671 s
2 Classes + PCA	88.50% – 0.8671 s	91.00% – 0.8812 s	75.00% – 0.7345 s
2 Classes + LDA	78.50% – 0.534 s	76.00% – 0.899 s	72.20% – 0.7123 s
2 Classes + GA	80.00% – 0.8566 s	78.00% – 0.8255 s	76.50% – 0.7277 s

References

Akansu, A., Serdijn, W., & Selesnick, I. (2010). Wavelet transforms in signal processing: a review of emerging applications. *Physical Communication*, *3*, 1–18. http://dx.doi.org/10.1016/j.phycom.2009.07.001.

Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. *Computer Methods and Programs in Biomedicine*, *141*, 19–26. http://dx.doi.org/10.1016/j.cmpb.2017.01.004.

Bentley, P., Nordehn, G., Coimbra, M., Mannor, S., & Getz, R. (2011). Classifying heart sounds challenge. Retrieved from Classifying Heart Sounds Challenge: http://www.peterjbentley.com/heartchallenge.

Bilal, E. M. (2021). Heart sounds classification using convolutional neural network with 1D-local binary pattern and 1D-local ternary pattern features. *Applied Acoustics*, *180*, Article 108152. http://dx.doi.org/10.1016/j.apacoust.2021.108152.

Bonow, R. O., Mann, D., Zipes, D., & Libby, P. (2011). *Braunwald's heart disease: A textbook of cardiovascular medicine, single volume*. Elsevier Science.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Chen, A. H., & Yang, C. (2012). The improvement of breast cancer prognosis accuracy from integrated gene expression and clinical data. *Expert Systems with Applications*, *39*, 4785–4795. http://dx.doi.org/10.1016/j.eswa.2011.09.144.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Deng, S.-W., & Han, J.-Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Generation Computer Systems*, *60*, 13–21. http://dx.doi.org/10.1016/j.future.2016.01.010.

Dominguez-Morales, J. P., Jimenez-Fernandez, A. F., Dominguez-Morales, M. J., & Jimenez-Moreno, G. (2017). Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. *IEEE Transactions on Biomedical Circuits and Systems*, *12*, 24–34. http://dx.doi.org/10.1109/TBCAS.2017.2751545.

Giron-Sierra, J. M. (2016). *Model-based actions and sparse representation: vol. 3, Digital signal processing with matlab examples*. Springer.

Hamidi, M., Ghassemian, H., & Imani, M. (2018). Classification of heart sound signal using curve fitting and fractal dimension. *Biomedical Signal Processing and Control*, *39*, 351–359. http://dx.doi.org/10.1016/j.bspc.2017.08.002.

Hasan, R., Jamil, M., Rabbani, G., & Rahman, S. (2004). Speaker identification using mel frequency cepstral coefficients. In *The proceedings of the 3rd international conference on electrical & computer engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh*.

Huang, T.-M., Kecman, V., & Kopriva, I. (2006). *Kernel based algorithms for mining huge data sets (Vol. 1)*. Springer.

Jia, L., Song, D., Tao, L., & Lu, Y. (2012). Heart sounds classification with a fuzzy neural network method with structure learning. In J. Wang, G. G. Yen, & M. M. Polycarpou (Eds.), *Lecture notes in computer science: vol 7368, Advances in neural networks – ISNN 2012. ISNN 2012*. Berlin, Heidelberg: Springer, http://dx.doi.org/10.1007/978-3-642-31362-2_15.

Joshi, A., & Mehta, A. (2018). Analysis of k-nearest neighbor technique for breast cancer disease classification. *International Journal of Recent Scientific Research*, *9*, 26126–26130.

Kaucha, D. P., Prasad, P. W. C., Alsadoon, A., Elchouemi, A., & Sreedharan, S. (2017). Early detection of lung cancer using SVM classifier in biomedical image processing. In *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)* (pp. 3143–3148). IEEE, http://dx.doi.org/10.1109/ICPCSI.2017.8392305.

Kaymak, S., Helwan, A., & Uzun, D. (2017). Breast cancer image classification using artificial neural networks. *Procedia Computer Science*, *120*, 126–131. http://dx.doi.org/10.1016/j.procs.2017.11.219.

Khateeb, N., & Usman, M. (2017). Efficient heart disease prediction system using K-nearest neighbor classification technique. In *Proceedings of the international conference on big data and internet of thing* (pp. 21–26). http://dx.doi.org/10.1145/3175684.3175703.

Kui, H., Pan, J., Zong, R., Yang, H., & Wang, W. (2021). Heart sound classification based on log Mel-frequency spectral coefficients features and convolutional neural networks. *Biomedical Signal Processing and Control*, *69*, Article 102893. http://dx.doi.org/10.1016/j.bspc.2021.102893.

Leng, S., San Tan, R., Chai, K. T. C., Wang, C., Ghista, D., & Zhong, L. (2015). The electronic stethoscope. *Biomedical Engineering Online*, *14*, 1–37. http://dx.doi.org/10.1186/s12938-015-0056-y.

Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., & … Clifford, G. D. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, *37*(12), 2181. http://dx.doi.org/10.1088/0967-3334/37/12/2181.

Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. *Expert Systems with Applications*, *164*, Article 113981. http://dx.doi.org/10.1016/j.eswa.2020.113981.

Mason, L., Baxter, J., Bartlett, P., & Frean, M. (1999). Boosting algorithms as gradient descent. *Advances in Neural Information Processing Systems*, *12*, 512–518, https://dl.acm.org/doi/10.5555/3009657.3009730.

Mousavi, S. M., Abdullah, S., Niaki, S. T. A., & Banihashemi, S. (2021). An intelligent hybrid classification algorithm integrating fuzzy rule-based extraction and harmony search optimization: Medical diagnosis applications. *Knowledge-Based Systems*, *220*, Article 106943. http://dx.doi.org/10.1016/j.knosys.2021.106943.

Noman, F., Ting, C.-M., Salleh, S.-H., & Ombao, H. (2019). Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1318–1322). IEEE, http://dx.doi.org/10.1109/ICASSP.2019.8682668.

Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016). Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In *2016 computing in cardiology conference (CinC)* (pp. 621–624). IEEE.

Septiani, N. W. P., Wulan, R., & Lestari, M. (2017). Breast cancer detection using data mining classification methods. *Proceeding ICMETA*, *1*(1).

Smelser, N. J., & Baltes, P. B. (2001). *International encyclopedia of the social & behavioral sciences (Vol. 11)*. Amsterdam: Elsevier.

Thiyagaraja, S. R., Dantu, R., Shrestha, P. L., Chitnis, A., Thompson, M. A., Anumandla, P. T., & … Dantu, S. (2018). A novel heart-mobile interface for detection and classification of heart sounds. *Biomedical Signal Processing and Control*, *45*, 313–324. http://dx.doi.org/10.1016/j.bspc.2018.05.008.

Zabihi, M., Rad, A. B., Kiranyaz, S., Gabbouj, M., & Katsaggelos, A. K. (2016). Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. In *2016 computing in cardiology conference (CinC)* (pp. 613–616). IEEE.

Zhang, W., Han, J., & Deng, S. (2017a). Heart sound classification based on scaled spectrogram and partial least squares regression. *Biomedical Signal Processing and Control*, *32*, 20–28. http://dx.doi.org/10.1016/j.bspc.2016.10.004.

Zhang, W., Han, J., & Deng, S. (2017b). Heart sound classification based on scaled spectrogram and tensor decomposition. *Expert Systems with Applications*, *84*, 220–231. http://dx.doi.org/10.1016/j.eswa.2017.05.014.

Zhang, W., Han, J., & Deng, S. (2019). Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation. *Biomedical Signal Processing and Control*, *53*, Article 101560. http://dx.doi.org/10.1016/j.bspc.2019.101560.