



Classification of Normal and Murmur Hearts Sound using the Fractal Method

Ratna Komalasari¹, Achmad Rizal², Fiky Yosef Suratman³

¹School of Electrical Engineering, Telkom University, Indonesia, ratnak1602@student.telkomuniversity.ac.id

²School of Electrical Engineering, Telkom University, Indonesia, achmadrizal@telkomuniversity.ac.id

³School of Electrical Engineering, Telkom University, Indonesia, fysuratman@telkomuniversity.ac.id

ABSTRACT

Heart sounds contain important information related to heart health. Normal heart sounds produce sound patterns that are different from abnormal ones. Various digital signal processing methods have been used to differentiate these heart sound signals and the most usual methods are wavelet analysis, entropy analysis, or a combination of both. In this study, the fractal dimension was used to classify normal heart sounds and murmur sounds by using the box-counting fractal dimension (BCFD), Katz fractal dimension (KFD), Sevcik fractal dimension (SFD), and Higuchi fractal dimension (HFD) as the heart sound features. The highest accuracy reached 100% using SFD as a feature and KNN as a classifier. These results were tested on 50 normal heart sounds and 50 heart sounds murmurs with 3-fold cross-validation. The results showed that choosing the right fractal dimension can distinguish normal heart sounds and murmurs.

Key words : classification, fractal dimension, heart sound, murmur.

1. INTRODUCTION

Cardiac abnormalities can be detected using several techniques such as electrocardiography, echocardiography, or phonocardiography [1]. In the electrocardiography technique, electrodes are used to detect the heart's electrical activity, while in echocardiography, an ultrasonic device is used to observe heart activity visually. In phonocardiography, heart sounds are heard using a stethoscope to see any heart abnormalities due to the heart valves' opening and closing [2]. One of the abnormalities in heart sound is known as murmur [3]. The process of listening to heart sounds is called auscultation, where heart sounds are assessed based on the expertise and experience of doctors [4]; however, the auscultation process is still subjective due to the expertise and experience of doctors,

Various signal processing methods have been developed to assist doctors in assessing heart sounds using a computer. The wavelet method is one of the most widely used methods [5]. It provides simplicity to decompose the signal into several sub bands. Some abnormalities in heart sounds can be observed in specific frequency components that can easily be observed using wavelet analysis. Safara et al. proposed the wavelet packet decomposition used to detect murmurs in heart sounds [6]; meanwhile, Rizal and Wijayanto proposed the development of entropy wavelet packet using multilevel wavelet packet entropy MWPE [7].

One of the methods of biological signal analysis frequently used is the fractal method. Fractal analysis has been shown to provide promising results in the classification of various biological signals such as lung sounds [8], electrocardiogram signal (ECG) [9], and electroencephalogram signal (EEG) [10]. The fractal dimension can distinguish the signal complexity in the normal signal and the pathological signal. The pathological signal changes the signal's complexity due to a decrease in function in the associated organ [11].

This research proposes a classification method for heart sounds using fractal dimensions. The fractal method used was box-counting method (BCFD) [12], Katz method (Katz fractal dimension, KFD) [13], Sevcik method (Sevcik fractal dimension, SFD) [14], and Higuchi method (Higuchi fractal dimension, HFD) [15]. In contrast to research combining wavelet transforms and fractal dimensions [16], in this study, the fractal dimensions were calculated directly on the sound signal of the heart. Thus, it could be immediately known what fractal dimensions could distinguish normal heart sounds and murmurs. It is hoped that this method can provide a high enough accuracy.

2. MATERIALS AND METHODS

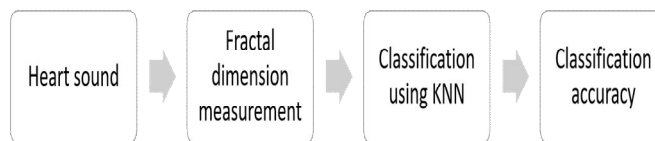


Figure 1: Proposed system

Figure 1 in general shows the method proposed in this study. In the heart sound, the fractal dimensions were measured, which were used as features. Furthermore, the classification was carried out using K-NN and accuracy was calculated from the amount of data that had been correctly classified, and divided by the total data. The details of each process are described in the following subsections.

2.1 Heart Sound Dataset

Heart sound is a low-frequency signal generated by the heart valves. The S1 and S2 are main sounds heard in a heartbeat cycle. The S1 and S2 refer to first and second heart sounds. Disruption of blood flow from the heart valve produced murmurs, a high-frequency sound in hear sound [17]. In this study, we classified heart murmurs and normal heart sounds using the fractal dimension as feature extraction method and KNN as the classifier. The input data consisted of 50 normal heart sounds and 50 murmur sounds from the website physionet.org [18][19]. Figure 2 shows the images of heart sounds.

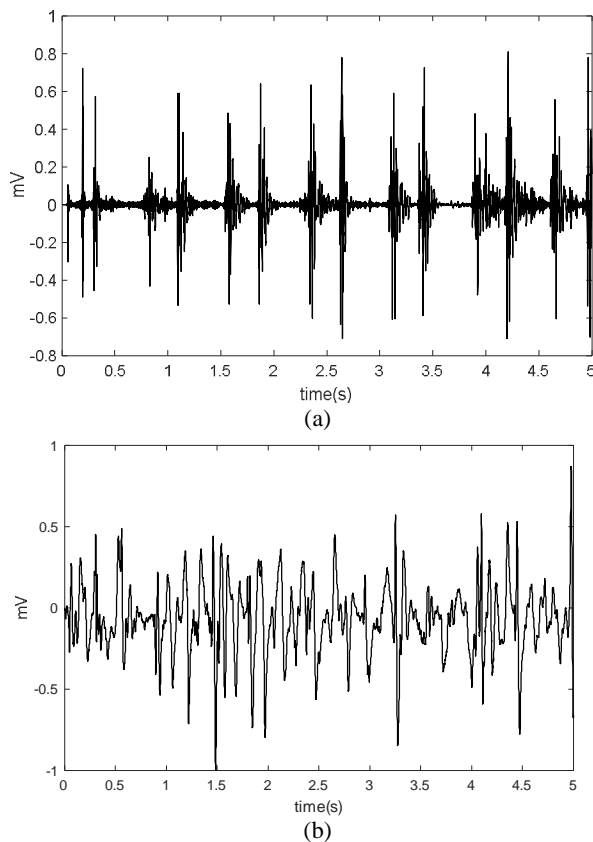


Figure 2: (a) Normal heart sound (b) Murmur

2.2 Feature extraction using Fractal Dimensions

The fractal dimension is defined as a measure of the emergence of self-similarity, which is a signal pattern that repeats itself on different scales [20]. Fractal dimension value

is not an integer like the one in the Euclidean dimension in which 1 is for line, 2 is for a plane, or 3 is for space. For 1-dimensional signals, the fractal dimensions are $1 \leq FD < 2$, where the more the complex a signal is, the closer the value will be to 2 [14]. In this study, four methods of measuring fractal dimensions were used, as described below.

A. Box Counting Method

The box-counting fractal dimension (BCFD) method is motivated by the curve's nature in filling a box-shaped space [20]. It calculates how many boxes in a specific size need to flap a curve. When the box size approaches zero, the entire curve will be closed by the box, and it can mathematically be written as (1).

$$D_B = \lim_{r \rightarrow 0} \frac{\log N(r)}{\log(1/r)} \tag{1}$$

where $N(r)$ is the number of squares of size r needed to cover the entire curve. In practical terms, the BCFD method counting the number of squares needed to cover curves of various box sizes to estimate fractal dimensions. Then, the D_B value is calculated by looking at a straight line on the log-log plot of $N(r)$ against r . It can mathematically be expressed as in (2).

$$\log N(r) = D_B \log \left(\frac{1}{r}\right) + C \tag{2}$$

The value of C is constant, while D_B is the slope value of the log-log $N(r)$ graph concerning r . This method often called the grid method requires a long computation time.

B. The Katz Method

The Katz fractal dimension (KFD) of a curve in a series along N is defined as in (3) [13].

$$KFD = \frac{\log_{10}(L_c)}{\log_{10}\left(\frac{d}{L_c}\right) + \log_{10}(N_c)} \tag{3}$$

where L_c is the total length of the curve as presented by (4).

$$L_c = \sum_{i=1}^{N-1} dist(i, i + 1) \tag{4}$$

$Dist(i, i + 1)$ represents the distance between two consecutive points. The d value in Equation (d) shows the farthest distance or diameter on the curve. This value can be found by calculating the starting point of the curve with the point farthest from the curve's starting point. The value of d can be calculated using (5).

$$d = \max[dist(1, i)], \quad i = 2, \dots, N \tag{5}$$

C. Sevcik Fractal Dimension

The calculation of Sevcik fractal dimensions (SFD) on a curve along N is displayed as shown in (6) [14].

$$SFD = 1 + \frac{\ln(Lc)}{\ln[2 \cdot (N-1)]} \tag{6}$$

Lc in (6) is the total length stated in (2). Another variation of the Sevcik method is the normalization process on the x-axis and y-axis before calculating the LC and SVD. The process of normalizing the x-axis pad is as shown in (7).

$$x_i^* = \frac{x_i}{x_{max}}, \quad i = 1, \dots, N \tag{7}$$

x_i 's value is the original value on the x-axis, while x_{max} is the maximum value of x_i . Normalization on the y-axis is expressed as in (8).

$$y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}}, \quad i = 1, \dots, N \tag{8}$$

y_i is the initial value of y, y_{min} is the minimum y value and y_{max} is the maximum value of y.

D. The Higuchi Method

The Higuchi fractal dimension (HFD) is one of the fractal dimension measurement algorithms often used in biomedical signals [15]. The advantages of this method are related to its high accuracy and efficiency for measuring fractal dimensions. If a signal is with a sample size of N, then a series of signals with the length of k can be formed with different resolutions as shown in (9).

$$X_k^m: x(m), x(m+k), x(m+2k), \dots, x\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right) \tag{9}$$

In (9), m indicates the indicator start of signal ($m=1, 2, \dots, k$), and length of curve X_k^m , $L_m(k)$ is defined as (10).

$$L_m(k) = \frac{\sum_{i=k}^{\lfloor \frac{N-m}{k} \rfloor} |x(m+ik) - x(m+(i-1)k)| \cdot (N-1)}{(\lfloor \frac{N-m}{k} \rfloor)k} \tag{10}$$

Notation $\lfloor a \rfloor$ mean floor(a), which $(\lfloor \frac{N-m}{k} \rfloor)k$ is the normalization factor. From equation (10), the length of the curve for every interval k is presented in (11).

$$L(k) = \sum_{m=1}^k L_m(k) \tag{11}$$

Fractal dimension was obtained from the slop between plot $\ln(L(k))$ towards $\ln(1/k)$. The value was obtained from the relation of $L(k) \propto k^{-D}$ in which the fractal dimension of Higuchi (HFD)= D.

E. K-Nearest Neighbor

K-nearest neighbor (K-NN) is a classification method that uses distance in determining the classification results based on the largest training data class with the closest distance to test data [21]. The steps of the K-Nearest Neighbor (KNN) algorithm are presented as follows:

1. Determine the parameter value of K; here, K is the number of closest neighbors
2. Calculate the distance of the test data with all existing data in the training data
3. Sort the distances and determine the nearest neighbor from the minimum distance to K data
4. Determine the category of the closest neighbor.

The distance used in this study was the Euclidean distance and City block distance as stated in (12) and (13).

$$D_e(x_i, y_i) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \tag{12}$$

$$D_c(x_i, y_i) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_p - y_p| \tag{13}$$

The accuracy test was tested using N-fold cross-validation. This research used N = 3, where the input data were then divided into three datasets. Two data set was used as training data, and one data sets were used as training data. The test was carried out three times so that each data set was used as test data. The accuracy obtained was the average accuracy of the three tests.

3. RESULTS AND DISCUSSION

Figure 3 shows a box-plot of the fractal dimensions of normal heart sounds and murmurs. It appeared that the fractal dimensions of normal heart sounds and murmurs tended to coincide for BC, KFD, and HFD. Meanwhile, the SFD for normal heart sounds and murmurs seemed totally separated. With this result, it could be predicted that the KFD accuracy would be lower than the other fractal dimensions; while, SFD would produce the highest accuracy. The four fractal dimensions' performance was tested using KNN with several K values and two distance calculation methods.

Table 1 and Table 2 present the classification results using the KNN with various K values and distance measurements. Those two tables display the results not much different from those shown in Figure 3. Accuracy was calculated using (14).

$$Acc(\%) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{14}$$

TP is a heart sound murmur recognized as a murmur, whereas TN is the amount of normal heart sound data recognized as a normal heart sound. TP + TN + FP + FN equals the total amount of data.

KFD produced the highest accuracy of 61% with K = 5 and city block distance. HFD provided the highest accuracy of 64% with K = 5 for all distance measurements. BC yielded the highest accuracy of 87% using K = 5 and the Euclidean distance. SFD provided 100% accuracy for various conditions. The composite of fractal dimension did not affect accuracy because SFD had a dominant contribution.

Theoretically, the fractal dimensions of normal heart sounds are higher for having a higher signal density compared to heart murmurs. These results are shown in Figure 2 (a), (c), and (d). Meanwhile, the KFD measurement resulted in a normal heart sound FD value against a heart murmur. From equation (5), KFD sought the farthest distance from the beginning of the curve. The heart murmur produced the farthest distance higher than the normal heart sound.

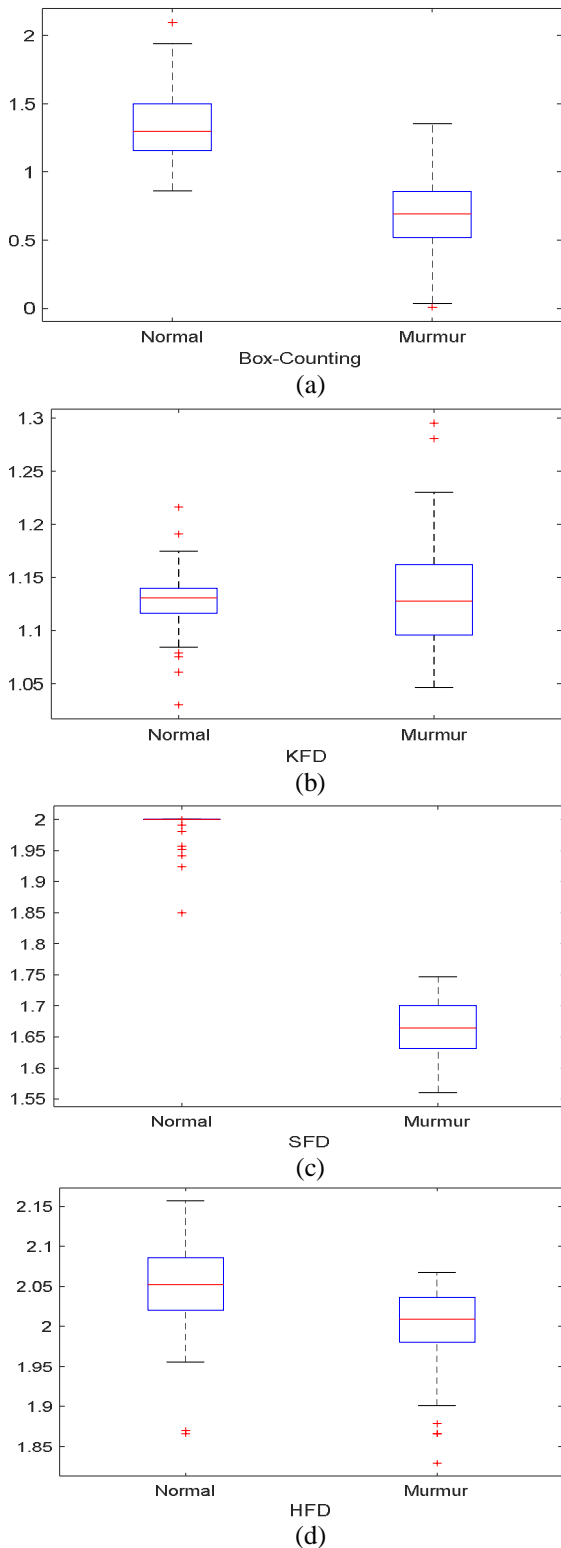


Figure 3: Box-plot of (a) BCFD (b) KFD, (c) SFD, (d) HFD

In a previous study, a multilevel wavelet packet entropy was used to classify normal and abnormal heart sounds [7]. The accuracy of 96.7% was produced using three features with the random forest as the classifier. Meanwhile, Safara et al. used wavelet packet entropy to classify heart murmurs [6]. 96.94% accuracy was generated using the BayesNet classifier. The previous research differences include the number of data classes used, the number of features used, and the signal segmentation process—in comparison to the proposed methods with previous research as shown in Table 3.

Table 1: Accuracy (%) using Euclidean Distance

KNN K	Fractal Dimension				
	BC	KFD	SFD	HFD	Composite
1	81	51	100	52	100
3	84	54	100	62	100
5	87	55	99	64	100

Table 2: Accuracy (%) using City Block Distance

KNN K	Fractal Dimension				
	BC	KFD	SFD	HFD	Composite
1	82	53	100	49	100
3	85	57	100	57	100
5	84	61	100	64	100

Table 3 shows that entropy is one of the most appropriate feature extraction methods used to analyze heart sounds. It is based on the fact from the different complexity between normal heart sounds and murmurs. One of the parameters for measuring signal complexity is entropy [26]. Meanwhile, the fractal dimension is a measure of the signal's self-similarity, where different signals will produce a number of different fractal dimensions. The fractal dimension has been proven to be widely used in the biological signal analysis [27][28]. In this study, this property was used where the selection of measurement methods with the appropriate fractal dimensions resulted in high accuracy.

The advantages of the proposed method are related to its simplicity and a few features with high accuracy [29][30]. Meanwhile, the proposed method's weakness is that the fractal dimension is highly dependent on the components in the signal. The presence of noise in the signal has the potential to change the FD value and affect the resulting characteristics. This research has just been tested on a dataset with a limited amount of data. Testing on a larger dataset is expected to provide consistency in the accuracy resulted.

4. CONCLUSION

This research proposes a classification method for normal heart sounds and murmurs using the fractal dimension. One

fractal dimension has been shown to differentiate between normal heart sound signals and murmur sounds. Of the four fractal dimensions measurement, the Sevcik fractal dimension produced 100% accuracy using KNN as the classifier. These results could be obtained in testing using 50 normal heart sound data and 50 heart murmurs. Testing

larger data with a more significant number of data classes becomes a challenge in future studies. In addition, testing for resistance to noise will become the next challenge. Heart sound is a signal that repeats itself every time; shifting data due to data segmentation needs to be addressed in further research.

Table 3: Comparison of several previous studies

Author	dataset	Feature extraction/reduction	Classifier	Results
Safara et al. [6]	50 Normal, 300 Murmur	Wavelet packet entropy	Decision tree, BayesNet, MLP, SVM, KNN	Accuracy 96.94% using BayesNet
Rizal & Wijayanto [7]	Normal, mitral prolapse (MP), mitral regurgitation (MR), and continuous murmur (CM).	Multilevel wavelet packet entropy	Random forest	Accuracy of 96.67%
Hamidi et al. [22]	https://www.physionet.org/challenge/2016/#challenge-data and other 5 dataset	Curve fitting, modified Mel spectrum using fractal dimension	KNN	Accuracy of 92%, 81% and 98% for three dataset
Juniati et al. [23]	40 data of normal heart sound, 40 data of extra systole, and 40 data of murmurs	HFD of DWT Approximation coefficient	KNN, fuzzy C-mean	Accuracy of 86.67% using KNN Accuracy of &8.56 using Fuzzy C-mean
Singh & Chema [24]	30 normal, 30 murmur	10 time domain features and 23 frequency domain features	BayesNet, NaiveBayes, SGD, LogitBost	Accuracy of 93.33% using Naïve Bayes
Langley & Murray [25]	2048 normal, 630 abnormal	Spectral amplitude and wavelet entropy	Decision tree	Accuracy of 79%
Proposed method	50 normal, 50 murmur	Fractal dimension	KNN	Accuracy of 100% using SFD

REFERENCES

1. A. Rizal, *Instrumentasi Biomedis*, First edit. Yogyakarta: Graha Ilmu, 2014.
2. D. Kristomo, A. Rizal, and A. Kusjani, **Seleksi ciri suara jantung pada domain waktu dan frekuensi**, in *Seminar Riset Teknologi Informasi (SRITI)*, 2016, vol. 6, pp. 332–337.
3. F. Safara, S. Doraisamy, A. Azman, A. Jantan, and A. R. Abdullah Ramaiah, **Multi-level basis selection of wavelet packet decomposition tree for heart sound classification**, *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1407–1414, 2013.
4. A. Ganguly and M. Sharma, **Detection of pathological heart murmurs by feature extraction of phonocardiogram signals**, *J. Appl. Adv. Res.*, vol. 2, no. 4, pp. 200–205, 2017.
5. S. Choi, Y. Shin, and H. K. Park, **Selection of wavelet packet measures for insufficiency murmur identification**, *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4264–4271, 2011.
6. F. Safara, S. Doraisamy, A. Azman, A. Jantan, and S. Ranga, **Wavelet packet entropy for heart murmurs classification**, *Adv. Bioinformatics*, vol. 2012, 2012.
7. A. Rizal and I. Wijayanto, **Phonocardiogram Classification using Multilevel Wavelet Packet Entropy and Random Forest**, in *the 6th International Conference on Science and Technology (ICST 2020)*, 2020.
8. A. Rizal, R. Hidayat, and H. A. Nugroho, **Fractal Dimension for Lung Sound Classification in Multiscale Scheme**, *J. Comput. Sci.*, vol. 14, no. 8, pp. 1081–1096, 2018.
9. H. M. Mercy Cleetus and D. Singh, **Multifractal application on electrocardiogram**, *J. Med. Eng. Technol.*, vol. 38, no. 1, pp. 55–61, 2014.
10. I. Wijayanto, A. Rizal, and A. Humairani, **Seizure Detection Based on EEG Signals Using Katz Fractal and SVM Classifiers**, in *2019 5th International Conference on Science in Information Technology (ICSITech)*, 2019, pp. 78–82.

11. M. Costa, A. L. Goldberger, and C. K. Peng, **Multiscale entropy analysis of biological signals**, *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 71, pp. 1–18, 2005.
12. R. Lopes and N. Betrouni, **Fractal and multifractal analysis: A review**, *Med. Image Anal.*, vol. 13, no. 4, pp. 634–649, 2009.
13. M. J. Katz, **Fractals and The Analysis of Waveforms**, *Comput. Biol. Med.*, vol. 18, no. 3, pp. 145–56, 1988.
14. C. Sevcik, **A procedure to Estimate the Fractal Dimension of Waveforms**, *Complex. Int. [online]*, vol. 5, 1998.
15. T. Higuchi, **Approach to an irregular time series on the basis of the fractal theory**, *Phys. D Nonlinear Phenom.*, vol. 31, no. 2, pp. 277–83, 1988.
16. D. Juniati, C. Khotimah, D. E. K. Wardani, and K. Budayasa, **Fractal dimension to classify the heart sound recordings with KNN and fuzzy c-mean clustering methods**, *J. Phys. Conf. Ser.*, vol. 953, no. 1, 2018.
17. F. Safara and A. R. A. Ramaiah, **RenyiBS: Renyi entropy basis selection from wavelet packet decomposition tree for phonocardiogram classification**, *J. Supercomput.*, no. 0123456789, Aug. 2020.
18. A. L. Goldberger *et al.*, **PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals**, *Circulation*, vol. 101, pp. e215–e220, 2000.
19. C. Liu *et al.*, **An open access database for the evaluation of heart sound algorithms**, *Physiol. Meas.*, vol. 37, no. December, pp. 2181–2213, 2016.
20. B. B. Mandelbrot, *The Fractal Geometry of Nature*. New York: W. H. Freeman and Company, 1983.
21. G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, **KNN model-based approach in classification**, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, pp. 986–996, 2003.
22. M. Hamidi, H. Ghassemian, and M. Imani, **Classification of heart sound signal using curve fitting and fractal dimension**, *Biomed. Signal Process. Control*, vol. 39, pp. 351–359, 2018.
23. D. Juniati, C. Khotimah, D. E. K. Wardani, and K. Budayasa, **Fractal dimension to classify the heart sound recordings with KNN and fuzzy c-mean clustering methods**, *J. Phys. Conf. Ser.*, vol. 953, no. 1, 2018.
24. M. Singh and A. Cheema, **Heart Sounds Classification using Feature Extraction of Phonocardiography Signal**, *Int. J. Comput. Appl.*, vol. 77, no. 4, pp. 13–17, 2013.
25. P. Langley and A. Murray, **Heart sound classification from unsegmented phonocardiograms**, *Physiol. Meas.*, vol. 38, no. 8, pp. 1658–1670, Jul. 2017.
26. A. Rizal, R. Hidayat, and H. A. Nugroho, **Entropy measurement as features extraction in automatic lung sound classification**, in 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), 2017, pp. 93–97.
27. O. Bikou, A. Delides, A. Drougou, A. Nonni, E. Patsouris, and K. Pavlakis, **Fractal dimension as a diagnostic tool of complex endometrial hyperplasia and well-differentiated endometrioid carcinoma**, *In Vivo (Brooklyn)*, vol. 30, no. 5, pp. 681–690, 2016.
28. S. Bohdan *et al.*, **The electromagnetic waves scattering evaluation on the composite material fractal structure with radioisotope elements**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 5, pp. 2272–2276, 2019.
29. M. T. Viega, E. Marvin, R. Jayadi, and T. Mauritsius, **Heart disease prediction system using data mining classification techniques: Naïve bayes, knn, and decision tree**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3028–3035, 2020.
30. S. Moturi, **Classification Model for Prediction of Heart Disease using Correlation Coefficient Technique**, *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 2116–2123, Apr. 2020.