ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

Phonocardiogram Analysis for Cardiovascular Disease Screening Using K-Star Algorithm

N. Rashidian [1,*]

[1] Department of Physiotherapy, Rehabilitation Building, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Article History:
Received 24 January 2024
Received in revised form 12 March 2024
Accepted 4 June 2024
Available online 5 June 2024

Keywords:
Phonocardiogram Analysis, Screening, Cardiovascular Disease, K-Star Algorithm

ABSTRACT

Cardiovascular disease stands out as one of the most prevalent health issues among the population. The accurate diagnosis and effective treatment of this disease are of paramount importance. The primary objective of this research is to propose a novel model for the automatic classification of heart sounds, specifically targeting the analysis of phonocardiograms to aid in the screening and diagnosis of cardiovascular disease. In this study, a dataset consisting of 942 samples, recorded heart sounds, each characterized by 23 features. The K-Star algorithm was employed for the classification of heart sounds. The K-Star algorithm is a model-based learning method that utilizes entropy theory as a distance measure. This approach maximizes the extraction of information from available data, offering a consistent methodology for managing both symbolic features and missing values effectively. The algorithm calculates the distance between two samples by considering the complexity of transforming one sample into another. The Waka tool was employed to implement this algorithm. Through the utilization of the K-Star algorithm, the accuracy of phonocardiogram analysis for cardiovascular disease screening was significantly enhanced, achieving a notable accuracy rate of 80.8917%. This research contributes to the development of a reliable and efficient tool for the automatic classification of heart sounds, aiding in the early detection and screening of cardiovascular diseases.

1. INTRODUCTION

Cardiovascular diseases (CVD) are among the most dangerous types of illnesses due to the vital role the heart plays in supplying blood to all organs, delivering essential nutrients and oxygen to tissues, and removing waste products such as carbon dioxide. Proper blood circulation, a result of the heart's effective functioning, also helps in transporting hormones and cells throughout the body [1]. Despite being a leading cause of death, early detection of cardiovascular diseases can significantly reduce the risk factors associated with these conditions. Common diagnostic methods for cardiovascular diseases include phonocardiograms and heart sound auscultation [12].

The development of heart sound monitoring equipment for phonocardiography dates back to the 1930s and 1940s, with standardization beginning in the 1950s. The first international conference on phonocardiography took place in

[*] Corresponding Author: Nasrin.Rashidian@Yahoo.Com
Department of Physiotherapy, Rehabilitation Building, Shiraz University of Medical Sciences, Shiraz, Iran

http://dx.doi.org/10.47176/TMI.2024.82

Paris in 1950. Notably, a phonocardiogram system developed by Beckman Instruments was used during at least one of the Gemini space missions (1965-1966) to monitor astronauts' heart rates in flight. In 1970, John Keefer patented the phonocardiogram simulator while working as a US government employee. The original patent described the device as one that simulates human heart sounds through electrical voltage [13].

A phonocardiogram is a diagnostic technique that records the sounds and murmurs produced by the contracting heart, including those from the heart valves and associated blood vessels. The device works with either a chest microphone or a miniature sensor placed at the tip of a small tubular instrument, which is inserted into one of the heart chambers through the blood vessels. Phonocardiography is typically used alongside body sound auscultation with a stethoscope and, when combined with electrocardiography (ECG) and pulse rate measurements, offers significant diagnostic value [14-15].

The phonocardiogram device eliminates subjective interpretation of heart sounds, enabling a more objective evaluation of these sounds in relation to electrical and mechanical events in the cardiac cycle. In clinical practice, it is common to record other heart-related variables simultaneously with the phonocardiogram, including ECG, carotid artery pulses, and heart sound pulses. These are measured from the chest using a microphone system with a frequency response ranging from 0.1 to 100 Hz. Cardiologists then assess the phonocardiogram results based on waveform changes and various timing parameters [16]. The use of computer-based systems for automatic classification of heart sound signals can significantly enhance diagnostic accuracy, while also reducing the need for frequent patient visits to medical centers [17].

According to the World Health Organization's 2017 report, 46% of deaths in Iran were attributed to heart diseases. This percentage is likely to have increased in recent years, especially after the COVID-19 pandemic. The American Heart Association's report on cardiovascular diseases and strokes indicates that in 2022, approximately 19 million people worldwide died from CVDs, marking an 18.7% increase from 2010 [18].

This article explores the use of the K-Star algorithm to assess the accuracy of phonocardiogram analysis for cardiovascular disease screening. We implemented the algorithm using the Weka tool, demonstrating its potential to enhance the accuracy of phonocardiogram-based cardiovascular screening.

The article is structured as follows: the second section presents the methods, the third section details the proposed method, the fourth section discusses the research findings, and the fifth section offers conclusions.

2. METHODS

Various data mining methods have been applied to phonocardiogram analysis to enhance the screening of cardiovascular diseases. Below are some of the most prominent methods used for this purpose:

2.1. New Method of Phonocardiogram Analysis for Screening Cardiovascular Diseases Using Deep Learning Models

In this study, phonocardiogram (PCG) signals were screened using 2062 labeled PCG signals. New features were extracted and utilized in six different models, including Random Forest, Logistic Regression, Linear Discriminant Analysis, Decision Tree, K-Nearest Neighbors, and Deep Neural Network. Each model was evaluated using the k-fold cross-validation method ($k = 10$). The test data were applied to the mentioned models, and three key indicators—accuracy, sensitivity, and specificity—were calculated. The aim was to develop a new strategy for screening and differentiating heart patients from healthy individuals using PCG. The results revealed that the Deep Neural Network model exhibited the highest sensitivity value, with a sensitivity of 96.4 ± 0.14, specificity of 96.4 ± 0.14, and accuracy of 93.4 ± 0.11 [19].

2.2. Classification of Normal and Abnormal Heart Sounds Using Time-Frequency Analysis of PCG Signals

This research focused on classifying heart sounds by extracting time-frequency domain features from PCG signals. In the classification stage, a combination of two classifiers Adaboost and Convolutional Neural Network (CNN) was used. The proposed method was evaluated using the Leave-One-Out method and implemented on the

2016 PhysioNet Challenge database. The results demonstrated the superior performance of the proposed solution over the best existing method in the 2016 PhysioNet Challenge, achieving a sensitivity of 93.27% and specificity of 81.96%. In comparison, the best method in the 2016 PhysioNet Challenge achieved a sensitivity of 98.48% and specificity of 80.36% [17].

3. PROPOSED METHOD

The proposed method for phonocardiogram analysis aims to screen for cardiovascular diseases using the K-Star algorithm, which is outlined briefly in this section.

3.1. K-Star Algorithm

The K-Star algorithm is a cluster analysis method primarily designed to partition N observations into K clusters, ensuring that each observation is assigned to the cluster whose mean is the closest. The K-Star algorithm can be described as a model-based learning method that utilizes entropy theory as a distance measure. These methods enhance the possibility of extracting valuable information from available data by providing a consistent approach to manage the real properties of symbolic features and missing values.

In this algorithm, the distance between two samples is measured in terms of the complexity required to transform one sample into another. Specifically, the K-Star algorithm uses an entropy distance function to measure the interpeak distance and identify the most similar samples from the dataset. Suppose that A and B are the samples being considered. Then, P∗ can be described as the probability of each path from A to B. The relationship P is defined as follows [20].

$$P^*(b|a) = \sum_{t\in p:t(a)=b}^{N} p(t) \tag{1}$$

Where T Represents the Value of T (T Is a Set of Data Transformations) And P Is a Probability Function. Considering That P* Has the Following Conditions [20]:

$$\sum_{b}^{N} P^*(b|a) = 1, \;\; 0 \leq P^*(b|a) \leq 1 \tag{2}$$

According To the Above Relationships, The K-Star Function Is Expressed as Follows [20]:

$$K^*(b|a) = -\operatorname{Log} P^*(b|a) \tag{3}$$

That

$$K^*(b|a) \geq 0, \;\; k^*(b|a) + k^*(c|b) \geq K^*(c|a) \tag{4}$$

That The Above Statements Represent Whole Numbers and Are Rewritten as Below for Continuous Numbers[20].

$$P^*(b|a) = p^*(b|a) = P^*(i) = \frac{s}{\sqrt{2s-s^2}}\left(\frac{1-\sqrt{2s-s^2}}{1-s}\right), i = |a-b| \tag{5}$$

$$K^*(b|a) = K^*(i) = \frac{1}{2}\operatorname{Log}\left(\sqrt{2s-s^2} - \operatorname{Log}(s) + i[log(1-s) - \operatorname{Log}\left(1-\sqrt{2s-s^2}\right)\right] \tag{6}$$

Where S is the model parameter and varies between zero and one. With these records, the most appropriate sample for the desired data can be selected by using the calculated probability values [20].

4. RESEARCH FINDINGS

In this section, the implementation method and evaluation criteria are discussed.

4.1. Implementation

To implement the proposed method, we use the Weka tool, a popular data mining tool. The implementation is divided into five key steps:

Step 1: Data Preparation

In this step, we collect a dataset that includes 942 samples (heart sound recordings), and each sample consists of 23 features extracted from the dataset available at PhysioNet. We then load this dataset into the Weka software for further analysis.

Step 2: Data Sampling

Data sampling is performed to ensure a balanced representation of both normal and abnormal heart sounds within the dataset, allowing for accurate model training and validation.

Step 3: Adding the K-Star Algorithm

In this step, we incorporate the K-Star algorithm, which is utilized to cluster and classify the phonocardiogram signals effectively, as previously outlined.

Step 4: Data Classification

Finally, we perform the classification of heart sound signals using the K-Star algorithm, based on the features and the clustering results from the previous steps.

Table 1. Characteristics of Phonocardiogram for the Screening of Cardiovascular Disease [21].

Attributes	Description (Data Type)	Possible Values
Age	Age Category (String)	Neonate Infant Child Adolescent Young Adult
Sex	Reported Sex (String)	Female Male
Height	Height In Centimeters (Number)	> 0
Weight	Weight In Kilograms (Number)	> 0
Pregnancy Status	Did The Subject Report Being Pregnant at the Time of the Examination? (Boolean)	True False
Additional Id	The Second Record Identifier for Subjects That Participated to Both Screening Campaigns (String)	Subject Identifier
Campaign	Campaign Attended by the Subject (String)	Cc2014 Cc2015
Murmur	Indicates If a Murmur Is Present, Absent or Unidentifiable for the Annotator (String)	Present Absent Unknown
Murmur Locations	Auscultation Locations Where At Least One Murmur Has Been Observed (String)	Any Combination of the Following Abbreviations Separated by Plus (+) Signs: Pv, Tv, Av, Mv, And Phc
Most Audible Location	Auscultation Location Where Murmurs Sounded More Intense for the Annotator (String)	Pv Tv Av Mv Phc
Systolic Murmur Timing	Timing of the Murmur in the Systolic Period (String)	Early-Systolic Holosystolic Late-Systolic Mid-Systolic
Systolic Murmur Shape	Shape of the Murmur in The Systolic Period (String)	Crescendo Decrescendo Diamond Plateau
Systolic Murmur Pitch	Pitch of the Murmur in the Systolic Period (String)	Low Medium High

Systolic Murmur Grading	Grading of the Murmur in the Systolic Period According to the Levine's Scale (String)	I/Vi Ii/Vi Iii/Vi
Systolic Murmur Quality	Quality of the Murmur in the Systolic Period (String)	Blowing Harsh Musical
Diastolic Murmur Timing	Timing of the Murmur in the Diastolic Period (String)	Early-Diastolic Holodiastolic Mid-Diastolic
Diastolic Murmur Shape	Shape of the Murmur in the Diastolic Period (String)	Decrescendo Plateau
Diastolic Murmur Pitch	Pitch of the Murmur In the Diastolic Period (String)	Low Medium High
Diastolic Murmur Grading	Grading of the Murmur in The Diastolic Period According to the Levine Scale (String)	I/Iv Ii/Iv Iii/Iv
Diastolic Murmur Quality	Quality of the Murmur in the Diastolic Period (String)	Blowing Harsh
Outcome	The Expert Cardiologist's Overall Diagnosis	Normal Abnormal
Patient Id	Patient Type Id	Minimum Maximum Mean Stddev
Recording Locations	Recording Type Locations	Av+Pv+Tv+Mv Av+Mv Pv+Tv+Mv Av Mv Av+Pv+Mv Pv+Mv Av+Av+Pv+Pv+Tv+Mv Av+Pv+Tv Av+Pv+Tv+Mv+Mv Pv Pv+Tv+Tv Av+Pv+Pv+Tv+Tv+Mv Pv+Tv Av+Mv+Mv Av+Pv+Mv+Phc+Phc Av+Pv Tv Av+Av+Mv Av+Pv+Tv+Tv+Mv Av+Av+Mv+Mv Av+Av+Pv+Tv+Mv Av+Tv+Mv Av+Pv+Tv+Mv+Phc Av+Av+Av+Mv Av+Av+Pv+Tv+Mv+Mv Tv+Mv Tv+Mv+Phc
Campaign	Campaign	Cc2015 Cc2014

4.1.1. The Second Stage of Data Sampling

In this stage, we select the number of training and test datasets. The parameters used are Seed=1 and Fold=10, ensuring that the data is properly divided for validation and testing purposes.

4.1.2. The Third Step of the K-Star Algorithm:

In this step, we apply the K-Star algorithm. By clicking on the K-Star algorithm control in the Weka tool, we implement the algorithm, which uses entropy theory as a distance measure to classify the heart sound data.

4.1.3. The Fourth Stage of Data Classification:

During this stage, we estimate the classification accuracy based on the percentage of test samples or test datasets that are correctly classified. The results of this classification are presented in Table 2.

Table 2. The Results of Implementing The K-Star Algorithm

Number	Evaluation criteria	K-Star
1	Correctly The Number of Correctly Classified Samples	**80.8917%** **762**
2	Incorrectly Number of Misclassified Samples	**19.1083%** **180**
3	Kappa	**0.1092%**
4	Mean Absolute Error	**0.0033%**
5	Root Mean squared error	**0.0478%**
6	Relative Absolute Error	**62.4276%**
7	Root Relative Squared Error	**105.332%**
8	Total Number of Instances	**942**

4.2. Evaluation Criteria

According to Table 2, the number of samples that are correctly classified is 762 samples (recorded sound). The detection accuracy obtained by the proposed method is 80.8917%. Equation (7) has been used to calculate the detection accuracy:

$$Accuracy = \frac{TP^{\dagger}+TN^{\ddagger}}{TP^{\S}+TN^{**}+FP+FN} \tag{7}$$

According to Table 2, the number of samples that are incorrectly classified is 180 samples. The level of classification inaccuracy is equal to 19.1083%, which is calculated based on Equation (8):

$$Error = 100 - \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

5. CONCLUSION

Heart diseases are the leading cause of death in Iran and many countries worldwide, including the United States. The term "heart disease" refers to a range of conditions affecting the heart. In Iran, the death rate due to heart diseases is alarmingly high, rivaling and, in some cases, surpassing the death rate caused by road accidents.

In this study, we applied the K-Star algorithm for diagnosing cardiovascular diseases through phonocardiogram analysis. Using this method, we achieved an accuracy rate of 80.8917%. This result highlights the potential of phonocardiogram-based screening for cardiovascular diseases. Moving forward, we hope to enhance the accuracy of cardiovascular disease diagnosis by combining machine learning algorithms for more effective and precise screening.

Declaration

We acknowledge that we used ChatGPT to enhance the academic writing of our manuscript while ensuring the originality and integrity of our work.

Transparency Statement

† - True Positive
‡ - False Positive
§ - False Negative
** - True Negative

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFERENCES

[1] Virani, S., Alonso, Á., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... & Tsao, C. W. (2020). Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation, 141, e139–e151. https://doi.org/10.1161/CIR.0000000000000757

[2] Aviny, H. R., Ghasemi, M., & Fazlazad, M. (2023). Cardiovascular Disease Diagnosis Using the Combination of Principal Component Analysis Algorithm and Regression Tree. Transactions on Machine Intelligence, 6(2), 114–125.

[3] Darabi, F., Ezzati, F., & Mohammadhosseini, T. (2023). Assessing the Risk of Cardiovascular Diseases in Women and Men in Primary Health Care: A Case Study in Ardabil, Iran. Journal of the Iranian Institute for Health Sciences Research, 22(6), 699–708. https://doi.org/10.61186/payesh.22.6.699

[4] Amiri, M., Khosh Akhlagh, H., Sajadian, P. S., & Rezayi Jomaloyi, H. (2023). Comparing the Effectiveness of Acceptance and Commitment-Based Treatment and Transdiagnostic Treatment on Sexual Function and Type D Personality Components in People with Cardiovascular Disease. Knowledge and Research in Applied Psychology, 24(3), 18 pages.

[5] Sharghian Dizaji, V., & Sarbaz, Y. (2023). New Methods of Using the Photoplethysmogram Signal to Estimate the Body's Vital Signs and Identify Cardiovascular and Respiratory Diseases. The First National Conference of New Achievements in Electrical Engineering, Computer Engineering and Medical Engineering, 10 pages.

[6] Shtaban, S., Sid Esfehani, M. M., & Shataban, S. (2023). Management of Health Behaviors and Risk Factors with a Focus on Behavioral Control of Cardiovascular Disease Risk. Quarterly Journal of Behavioral Studies in Management, 14(34), 26 pages.

[7] Vahdani, F., Abedini, S., Mohseni, S., & Nikparvar, M. (2016). Nutritional Behaviors of Cardiovascular Disease Prevention in Women Referred to Health Care Centers in Minab Town. Journal of Preventive Medicine, 3(4), 59–66.

[8] Zanjanian, M., Mousaei, M., & Ghasemi, Z. (2022). Behavioral Style of Economic Status of Cardiovascular Patients in Javad-Al-Aemeh Heart Hospital of Mashhad. Journal of Social Welfare, 22(86). https://doi.org/10.32598/refahj.22.86.3913.1

[9] Rezapour Gatabi, N., Rahmani Firoozjah, A., Abbasi Asfajir, A. A., & Bijani, A. (2020). Evaluation of Lifestyle Related to Cardiovascular Health of Patients Referred to Hospitals Under the Auspices of Babol University of Medical Sciences. Journal of Social Welfare, 20(77), 301–327.

[10] Jahani, M. A., Barzgar, M., Abbasi, M., Yazdani Charati, J., & Mahmodi, G. (2021). Analyzing the Results of

Cardiovascular Disease Risk Assessment Program. Journal of Babol University of Medical Sciences, 236–243.

[11] Askari, E., Barzevi Some, S., Muetamid, S., & Farkh Bakht Fomeni, M. (2023). Investigating Smart Systems for Diagnosing Heart Disease Using Ultrasound Images Based on Deep Learning. The Second International Conference on Electrical, Computer, Mechanical and Artificial Intelligence Index Researches, 10 pages.

[12] Homeca. (n.d.). Types of Heart Disease. https://www.homeca.ir/mag/physical-health/cardiovascular/types-of-heart-disease

[13] Wikipedia. (n.d.). Phonocardiogram. https://en.wikipedia.org/wiki/Phonocardiogram

[14] Amiri, A. M., Abtahi, M., Constant, N., & Mankodiya, K. (2017). Mobile Phonocardiogram Diagnosis in Newborns Using Support Vector Machine. Healthcare, 5(1), 16. https://doi.org/10.3390/healthcare5010016

[15] Encyclopædia Britannica. (n.d.). Phonocardiography. www.britannica.com/science/phonocardiography

[16] Marzoratti, A., & Evans, T. M. (2022). Measurement of Interpersonal Physiological Synchrony in Dyads: A Review of Timing Parameters Used in the Literature. Cognitive, Affective, & Behavioral Neuroscience, 22, 1215–1230. https://doi.org/10.3758/s13415-022-01011-1

[17] Hashedi, H., Azmi, Q., & Zarjam, P. (2019). Classification of Normal and Abnormal Heart Sounds Using Time-Frequency Analysis of PCG Signals. Electrical Engineering (Technical Faculty of Tabriz University), 50(2), 657–668.

[18] Heidenreich, P. A., Bozkurt, B., Aguilar, D., Allen, L. A., Byun, J. J., Colvin, M., ... & Yancy, C. W. (2022). 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. https://doi.org/10.1161/CIR.0000000000001063

[19] Mohammadi, E., Kermani, S., Zawareh, M. N., Zare, A., Roudsari, H. A., Samii Nasab, M., & Sanei, H. (2022). A New Method of Phonocardiogram Analysis to Screen Part of Cardiovascular Diseases Using Deep Learning Model. Journal of Isfahan Medical College, 40(661), 109–114.

[20] Rahnma Qahfarkhi, N., Zarei, H., & Ahmadi, F. (2023). Application of K-Star Algorithm in Groundwater Level Prediction (Case Study: Aspas Plain). Iran Irrigation and Drainage Journal, 17(2), 305–320.

[21] Oliveira, J., Renna, F., Costa, P., Nogueira, M., Oliveira, A. C., Elola, A., Ferreira, C., Jorge, A., Bahrami Rad, A., Reyna, M., Sameni, R., Clifford, G., & Coimbra, M. (2022). The CirCor DigiScope Phonocardiogram Dataset (version 1.0.3). PhysioNet. RRID:SCR_007345. https://doi.org/10.13026/tshs-mw03