

RESEARCH ARTICLE

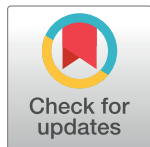
Enhanced heart sound classification using Mel frequency cepstral coefficients and comparative analysis of single vs. ensemble classifier strategies

Mehdi Hosseinzadeh^{1,2}, Amir Haider³, Mazhar Hussain Malik^{4*}, Mohammad Adeli⁵, Olfa Mzoughi⁶, Entesar Gemeay⁷, Mokhtar Mohammadi⁸, Hamid Alinejad-Rokny^{9,10}, Parisa Khoshvaght¹¹, Thantrira Porntaveetus^{12*}, Amir Masoud Rahmani^{13*}

1 School of Computer Science, Duy Tan University, Da Nang, Vietnam, **2** Jadara University Research Center, Jadara University, Irbid, Jordan, **3** Department of AI and Robotics, Sejong University, Seoul, Republic of Korea, **4** School of Computing and Creative Technologies College of Arts, Technology and Environment (CATE) University of the West of England Frenchay Campus, Bristol, United Kingdom, **5** Department of Biomedical Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran, **6** Department of Computer Sciences, College of Computer Engineering & Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia, **7** Department of Computer Engineering, Computer and Information Technology College, Taif University, Taif, Saudi Arabia, **8** Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Kurdistan Region, Iraq, **9** UNSW BioMedical Machine Learning Lab (BML), The Graduate School of Biomedical Engineering, UNSW Sydney, NSW, Australia, **10** Tyree Institute of Health Engineering (IHealthE), UNSW Sydney, Kensington, NSW, Australia, **11** DTU AI and Data Science Hub (DAIDASH), Duy Tan University, Da Nang, Vietnam, **12** Faculty of Dentistry, Department of Physiology, Center of Excellence in Genomics and Precision Dentistry, Clinical Research Center, Geriatric Dentistry and Special Patients Care International Program, Chulalongkorn University, Bangkok, Thailand, **13** Future Technology Research Center, National Yunlin University of Science and Technology, Yunlin, Taiwan

* These authors contributed equally to this work.

* mazhar.malik@uwe.ac.uk (MHM); thantrira.p@chula.ac.th (TP); rahmania@yuntech.edu.tw (AMR)



OPEN ACCESS

Citation: Hosseinzadeh M, Haider A, Malik MH, Adeli M, Mzoughi O, Gemeay E, et al. (2024) Enhanced heart sound classification using Mel frequency cepstral coefficients and comparative analysis of single vs. ensemble classifier strategies. PLoS ONE 19(12): e0316645. <https://doi.org/10.1371/journal.pone.0316645>

Editor: Xizhe Zhang, Nanjing Medical University, CHINA

Received: July 1, 2024

Accepted: December 14, 2024

Published: December 31, 2024

Copyright: © 2024 Hosseinzadeh et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

This paper seeks to enhance the performance of Mel Frequency Cepstral Coefficients (MFCCs) for detecting abnormal heart sounds. Heart sounds are first pre-processed to remove noise and then segmented into S1, systole, S2, and diastole intervals, with thirteen MFCCs estimated from each segment, yielding 52 MFCCs per beat. Finally, MFCCs are used for heart sound classification. For that purpose, a single classifier and an innovative ensemble classifier strategy are presented and compared. In the single classifier strategy, the MFCCs from nine consecutive beats are averaged to classify heart sounds by a single classifier (either a support vector machine (SVM), the k nearest neighbors (kNN), or a decision tree (DT)). Conversely, the ensemble classifier strategy employs nine classifiers (either nine SVMs, nine kNN classifiers, or nine DTs) to individually assess beats as normal or abnormal, with the overall classification based on the majority vote. Both methods were tested on a publicly available phonocardiogram database. The heart sound classification accuracy was 91.95% for the SVM, 91.9% for the kNN, and 87.33% for the DT in the single classifier strategy. Also, the accuracy was 93.59% for the SVM, 91.84% for the kNN, and 92.22% for the DT in the ensemble classifier strategy. Overall, the results demonstrated that MFCCs were more effective than other features, including time, time-frequency, and

statistical features, evaluated in similar studies. In addition, the ensemble classifier strategy improved the accuracies of the DT and the SVM by 4.89% and 1.64%, implying that the averaging of MFCCs across multiple phonocardiogram beats in the single classifier strategy degraded the important cues that are required for detecting the abnormal heart sounds, and therefore should be avoided.

1 Introduction

The mechanical activities of the heart and blood flow generate heart sounds [1]. The graphical representation of the heart sounds is usually called a phonocardiogram (PCG). A phonocardiogram typically comprises four components: S1 sound, systole, S2 sound, and diastole. However, there can be other sounds in a PCG [1].

Phonocardiograms can be used to develop assistive intelligent systems to detect cardiovascular diseases [2–5]. In general, such a system is composed of four steps: preprocessing, segmentation, feature extraction, and classification [5,6]. These steps are briefly reviewed below, but an exhaustive review of these techniques can be found here [5,7].

Preprocessing usually involves removing undesirable noises, artifacts, and spikes. Some of the techniques that have been used in other studies for PCG preprocessing include the normalization of PCGs to have zero mean [8], low-pass filtering [9,10], high-pass filtering from 10 Hz and normalization [11], band-pass filtering from 40 Hz to 400 Hz [12], band-pass filtering from 20 or 25 Hz to 400 Hz [12–14], band-pass filtering from 5 Hz to 700 Hz [15], band-pass filtering from 2 Hz to 100 Hz, and discrete wavelet transform (DWT) [16]. Sometimes, no preprocessing is applied [17].

Segmentation aims to find the S1, systole, S2, and diastole intervals in a PCG signal. Schmidt et al. (2010) used a duration-dependent hidden Markov model (DHMM) for PCG segmentation [18]. This method was extended by Springer et al. (2015) [19] using hidden semi-Markov models (HSMM) and logistic regression. This method has been adopted in many studies [3,12,13]. In [15], Mel-Scaled Wavelet Transform (MSWT) and dynamic thresholding were used for PCG segmentation while durations of systole and diastole were analyzed to find S1 and S2 in [10]. The PCG segmentation [11] was based on Shannon energy, envelope smoothing, and peak finding. Similarly, Jaros et al. (2023) [16] applied 3rd-order Shannon energy, envelope detection by low-pass filtering, thresholding, and the k-means algorithm. The PCG segmentation method proposed by Alonso-Arévalo et al. (2021) was based on spectral change detection and genetic algorithms [20]. Many of the PCG segmentation methods are reviewed here [21]. It is also necessary to mention that no segmentation strategies were used in some previous studies [8,9,17].

Quantitative features are extracted from the PCG segments in the third step of PCG processing. Various types of PCG features have been used in different applications. These types include time-domain features [3,10–13,15], spectral features [8,11–13], time-frequency features [3,10–12,15], time-scale features [8,11,13], Mel Frequency Cepstral Coefficients [3,10,11,15,22], and features estimated by convolutional neural networks (CNNs) [9]. Some feature selection strategies, such as linear discriminant analysis [10], correlation-based feature selection [13], and genetic algorithms, [11] have been used for dimensionality reduction.

The fourth step of PCG processing involves training and testing a classification model to detect underlying diseases [5,6]. Classification methods that have been used in applications of PCGs include convolutional neural networks (CNNs) [17,23], artificial neural networks

(ANNs) [3,9,10,15], deep neural networks [12], the k nearest neighbors algorithm (kNN) [3,11,15], decision trees (DT) [3,8], long short-term memory (LSTM) networks [3], ensemble classifiers [3,13], support vector machines (SVMs) [15], and hidden Markov models (HMMs) [24].

This study aimed to investigate the performance of MFCCs in discriminating abnormal PCGs from normal ones. MFCCs were first used in speech processing applications [25], but later in other applications such as PCG processing. MFCCs were selected because they are weakly-correlated and highly discriminating features of audio signals, providing compact spectral representations successfully used in speech processing applications [26]. Despite that, they have performed modestly in some PCG processing applications [3,10]. Therefore, the main goal of this study was to enhance the performance of MFCCs for PCG classification into normal/abnormal classes. For that purpose, two classification strategies are presented: 1) a single-classifier strategy, which takes as input the average MFCCs from multiple PCG beats, and 2) an innovative ensemble-classifier strategy comprising of an ensemble of 9 classifiers, each of which takes as input the MFCCs from a different PCG beat.

The rest of this article is organized as follows: Section 2 describes the PCG database used for evaluation of the proposed method, MFCC estimation, and the two classification strategies. Section 3 presents the results in detail. Section 4 discusses the results and compares them with similar studies. Section 5 presents this study's conclusions.

2 Materials and methods

2.1 Heart sounds database

The PhysioNET CinC 2016 PCG database [27,28] was used to evaluate the PCG classification methods proposed in this paper. The signals of this database have been collected from healthy subjects and patients with such heart diseases as heart valve defects and coronary artery problems. The researchers that contributed to the PhysioNET CinC 2016 PCG database include Syed (2003) [29] and Syed et al. (2007) [30] at Massachusetts Institute of Technology, Schmidt et al. (2010, 2015) [18,31] at Aalborg University, Papadaniil and Hadjileontiadis (2014) [32] at Aristotle University of Thessaloniki, Naseri et al. (2013) [33] at K. N. Toosi University of Technology, Moukadem et al. (2013) [34] at University of Haute Alsace, Tang et al. (2010) [35] at Dalian University, Samieinasab and Sameni (2015) [36] at Shiraz University, and Skejby Sygehus Hospital, Denmark [27].

The potential confounding variables such as the sample size, age, gender, nationality, and the individual state at the recording time are not a concern as the database contains a large number of signals from subjects with different age groups/genders/nationalities, in different states (rest or exercise), in different places (hospital or home), and with the signals obtained using different recording stethoscopes, leading to the validity, unbiasedness, and generalizability of the results.

Overall, the signals lasted from 5 to 120 seconds and were resampled to a rate of 2000 samples/s [27]. The PhysioNET CinC 2016 PCG database contains 6 datasets (labeled A to F) that contain 3153 signals (2488 from the healthy subjects and 665 from the patients). Dataset A (including 490 signals) was used to train the segmentation model introduced in section 2.2.2. and datasets B to F were used to evaluate our single-classifier and ensemble classifier strategies described in section 2.2.4. There are a total of 2744 signals in datasets B to F. Two hundred sixty-two of them that are labeled as "uncertain" are noisy signals and, therefore, were ignored. There remained 2482 signals (including 296 signals from the patient class and 2186 from the healthy class). Only signals with at least 9 PCG beats were used in this study. Among the remaining 2482 signals, 2137 PCGs (including 218 abnormal and 1919 normal PCGs) met this criterion and, therefore, were used to evaluate our single-classifier and ensemble-classifier strategies as described in section 2.2.4.

2.2 The proposed method for the classification of heart sounds

The proposed method for heart sound classification includes four stages: pre-processing, segmentation, feature extraction, and classification. These stages are explained in detail as follows.

2.2.1 Preprocessing. All the signals are first resampled from 2000 Hz to 1000 Hz to reduce the computational cost. They are then preprocessed using a band-pass filter from 25 to 400 Hz.

2.2.2 Segmentation. The preprocessed signals are segmented using the sophisticated supervised method proposed by Springer et al. (2015) [19]. This method includes a feature extraction step followed by a four-class classifier (S1, systole, S2, and diastole). The features were extracted using the Hilbert envelope, power spectrum density, and wavelet transform. The classifier assigns labels 1, 2, 3, and 4 to the PCG samples that belong to S1, systole, S2, and diastole, respectively.

2.2.3 Feature extraction. In this step, MFCCs are estimated for PCG signals. The MFCCs, which represent the short-term power spectrum of an audio signal, might be similar to the principles of structural-energy dynamics observed in friction processes [37]. The Mel scale is used to approximate the performance of the auditory system, which uses a non-linear frequency scale instead of a linear one. Estimation of MFCCs involves the following steps [38]:

1. The PCG signal is divided into frames.
2. Discrete Fourier Transform (DFT) of each PCG frame is computed.
3. The power spectrum of each frame is estimated using a Mel filter bank.
4. The logarithm of the power coefficients is calculated.
5. The discrete cosine transformation of the log power coefficients is computed.

To extract the MFCCs, we first divide the PCG signal into 24 ms frames. Adjacent frames overlap by 18 ms. Next, each frame is multiplied by a Hamming window, and then the 64-point DFT coefficients of the windowed frame are computed. Assuming that $X_i[k]$ denotes the DFT coefficients of the i^{th} windowed frame of the PCG signal, we estimate the power spectrum of this frame as $|X_i[k]|^2/N$. Then we calculate the power of this frame within all Mel bands. For that purpose, we use $M = 20$ triangular filters [10]. The frequency response of the m^{th} filter, $H_m(k)$, is defined as follows [38]:

$$H_m(k) = \begin{cases} 0, & k < k_f(m) \\ \frac{k - k_f(m)}{k_f(m+1) - k_f(m)}, & k_f(m) \leq k < k_f(m+1) \\ \frac{k_f(m+2) - k}{k_f(m+2) - k_f(m+1)}, & k_f(m+1) \leq k \leq k_f(m+2) \\ 0, & k > k_f(m+2) \end{cases}, \quad 0 \leq m \leq M-1 \quad (1)$$

Where the variable $k_f(m)$ is the index for the center frequency of the m^{th} triangular filter. A total of cap M plus 2 frequencies are required to design M filters. The relationship between the frequency in the Mel scale (f_{mel}) and the frequency in Hertz is calculated from the following equation [38]:

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2)$$

The minimum and maximum Mel frequencies are calculated for $f_{min} = 0$ and $f_{max} = 400$ Hz using Eq (2). Afterwards, we find $M+2$ equally-spaced Mel frequencies from the minimum to

the maximum Mel frequencies. The obtained $M+2$ Mel frequencies are converted back to the Hz scale using the inverse of Eq (2). With the $M+2$ frequencies required for designing M triangular filters available, the k_f index for the j^{th} frequency is calculated as:

$$k_f(j) = \left\lceil \frac{f(j)}{f_s} N \right\rceil, \quad 0 \leq j \leq M+1 \quad (3)$$

Where $f(j)$ is one of the $M+2$ designed frequencies in Hz, f_s is the sampling frequency, and N is the number of DFT coefficients. Now, the frequency response of these filters can be obtained using Eq (1).

The power of the i^{th} frame in the m^{th} Mel band, $P_i[m]$, is estimated as:

$$P_i[m] = 10 \log_{10} \left(\frac{1}{N} \sum_{k=0}^{\frac{N}{2}} |X_i[k]|^2 |H_m[k]| \right), \quad 0 \leq m \leq M-1 \quad (4)$$

Finally, the MFCCs $C_i[k']$ of the i^{th} frame are computed from $P_i[m]$ by the type II discrete cosine transformation (DCT-II) [38]:

$$C_i[k'] = \sum_{m=0}^{M-1} P_i[m] \cos \left(\frac{\pi}{M} \left(m + \frac{1}{2} \right) k' \right), \quad 0 \leq k' \leq M-1 \quad (5)$$

The number of MFCCs is usually between 12 and 20 [3,11]. In a nutshell, for a 24-ms frame of the PCG signal, 20 MFCCs are calculated, but only the first 13 coefficients were used in this research. As shown in Fig 1, by averaging the MFCCs obtained for all the frames belonging to the S1 sound, 13 features are obtained. Similarly, 13 features are calculated for each of the other segments, i.e. systole, S2, and diastole. Therefore, 52 MFCC features are obtained for a given PCG beat.

2.2.4 Classification. This research used two classification strategies to discriminate normal (healthy) PCGs from abnormal (pathological) ones: a single-classifier and an innovative ensemble-classifier. In the single-classifier strategy (Fig 2A), the MFCCs for the first 9 PCG beats are first averaged and then fed to a classifier. Since there are 52 features per beat, averaging results in 52 mean MFCCs (Fig 2A), based on which the classifier determines whether the PCG signal is normal or abnormal. Three classifier types were used in this strategy: k nearest neighbors (kNN), support vector machine (SVM), and decision tree (DT).

In the ensemble-classifier strategy (Fig 2B), the 52 MFCCs of a given beat are fed to a distinct classifier. Since only the first 9 beats of a PCG signal are used, there are 9 different classifiers. Each classifier decides whether its respective input beat is normal or abnormal. In the end, if more normal beats are predicted by the 9 classifiers altogether, the PCG is decided to be normal, otherwise abnormal (Fig 2B). All nine classifiers are the same type (i.e., kNN, SVM, or DT). For the kNN, k was considered 1, 3, 5, and 7. Also, linear, Gaussian, and polynomial kernels were considered for the support vector machine.

As explained in section 2.1, we selected 218 abnormal and 1919 normal PCG signals with at least 9 beats to evaluate the proposed methods. The first 9 beats were used for MFCC estimation. To balance the dataset, 218 normal PCGs were randomly selected and used alongside the 218 abnormal ones to train and test the classifiers.

Ten-fold cross-validation was used to evaluate both classification strategies. During each fold, we calculated four parameters of accuracy (Acc), sensitivity (Se), specificity (Sp), and

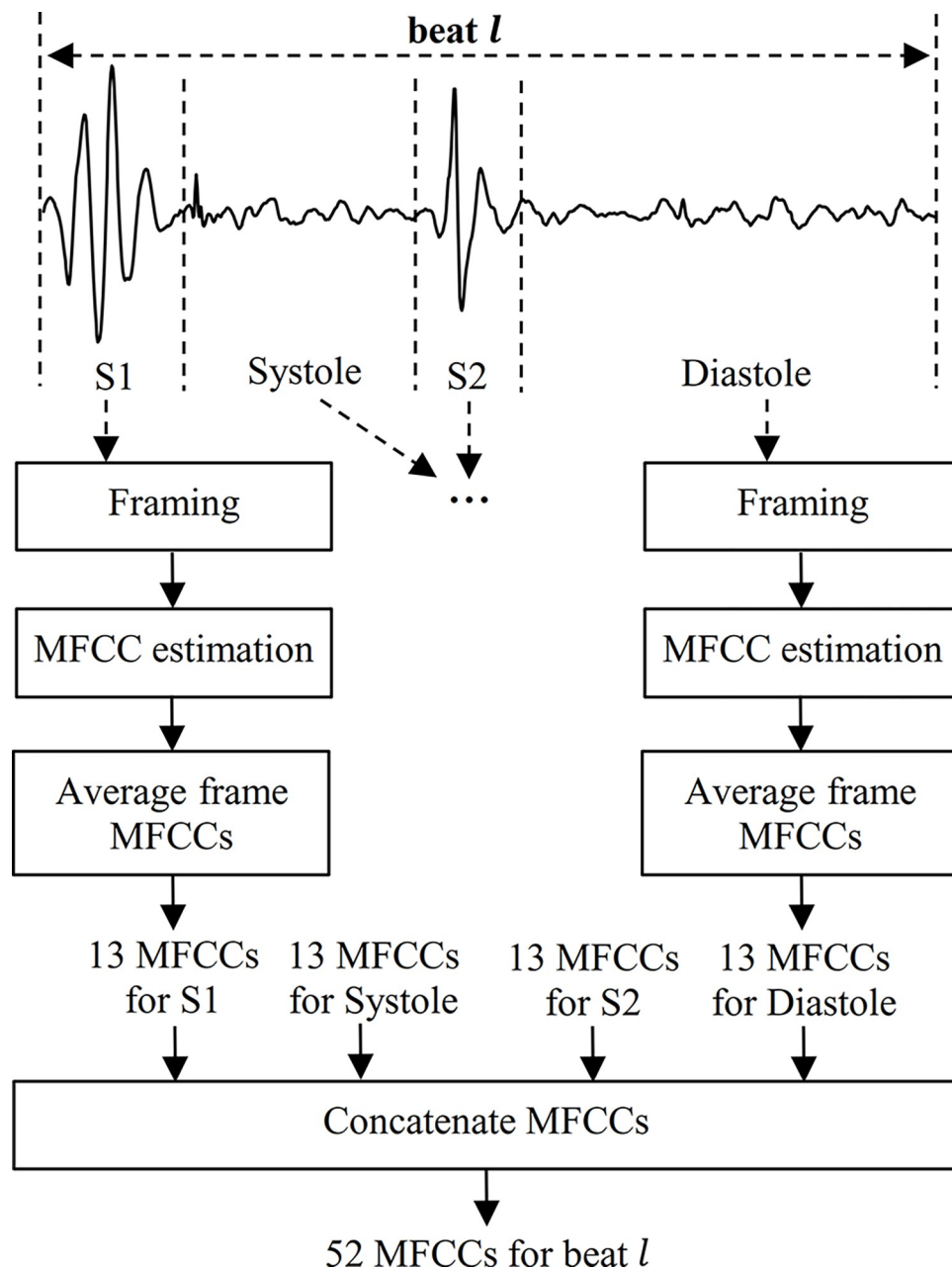


Fig 1. Feature extraction. 13 MFCCs are extracted from each interval (i.e., S1, systole, S2, and diastole) of a given PCG beat, summing up to 52 MFCCs for that beat.

<https://doi.org/10.1371/journal.pone.0316645.g001>

modified accuracy ($MAcc$) by:

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (6)$$

$$Se = \frac{TP}{TP + FN} \quad (7)$$

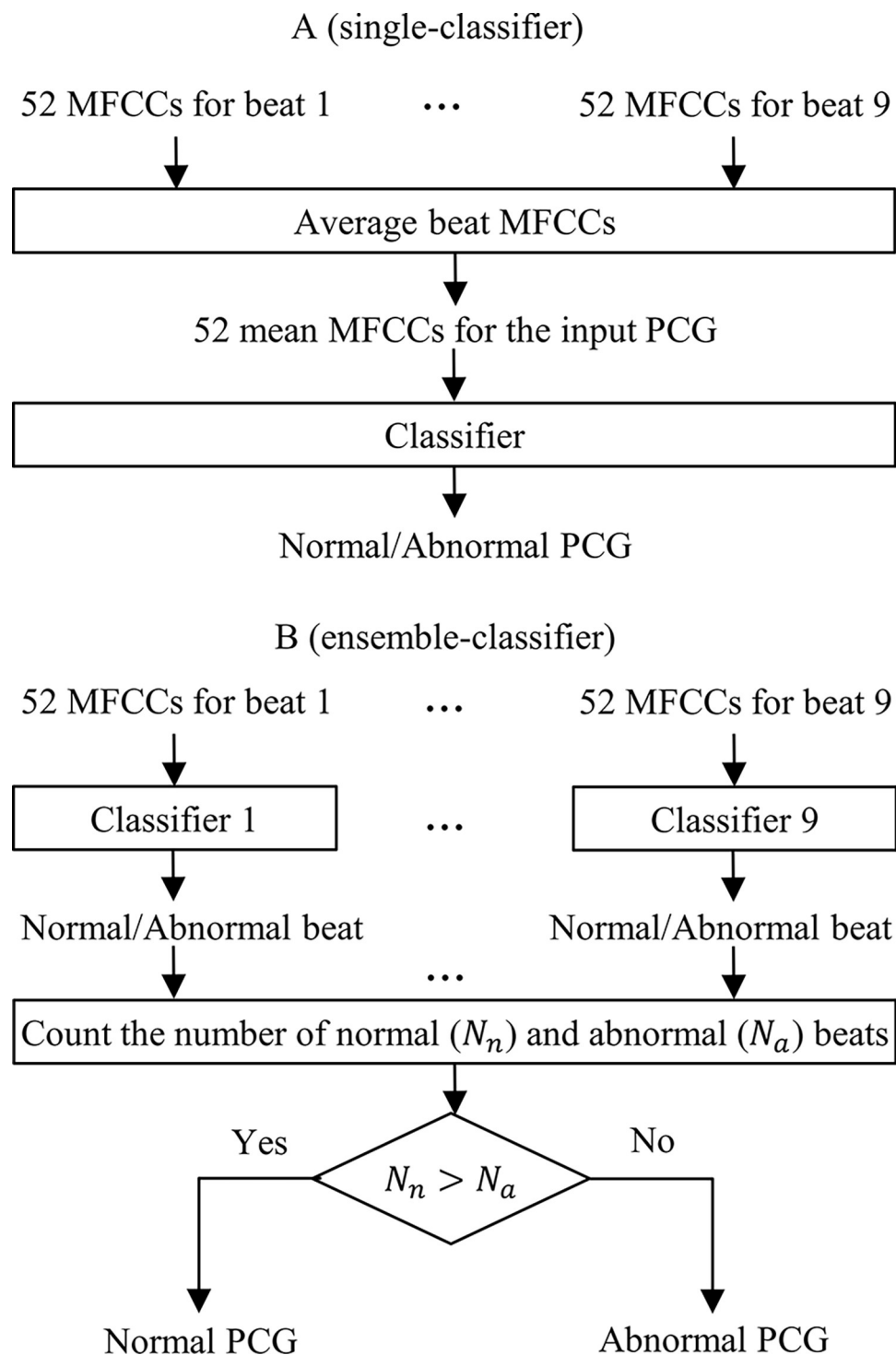


Fig 2. Classification strategies. A: Single-classifier strategy: The MFCCs extracted from the first 9 beats of a PCG are averaged and fed to a single classifier, B: Ensemble-classifier strategy: Nine classifiers are used separately to distinguish normal beats from abnormal ones. In the end, if the number of the normal beats is more than the abnormal beats, the PCG signal is decided to be normal; otherwise, it is abnormal.

<https://doi.org/10.1371/journal.pone.0316645.g002>

$$Sp = \frac{TN}{FP + TN} \quad (8)$$

$$MAcc = \frac{Se + Sp}{2} \quad (9)$$

Where TP is the number of patients who were correctly classified as patients, FN is the number of patients who were wrongly classified as healthy, FP is the number of healthy subjects who were wrongly classified as patients, and TN is the number of healthy people who were correctly classified as healthy subjects. The average parameters across the 10 folds were calculated in the end. The 10-fold cross-validation was repeated 50 times. Each time, a random set of 218 normal signals was selected and concatenated with the 218 abnormal PCGs. The average results across the 50 runs are reported in section 3.

It should be noted that all analyses, including feature extraction and classification strategies, were implemented using MATLAB programming language.

3 Results

Fig 3 shows the results for segmenting a PCG signal. In the staircase graph of Fig 3, levels 1, 2, 3, and 4 define the S1 sound intervals, the systole intervals, the S2 sound intervals, and the diastole intervals, respectively. As the segmentation strategy we used is a sophisticated strategy proposed and evaluated by Springer et al. (2015) [19], we did not evaluate its performance.

3.1. Results for the single-classifier strategy

The single-classifier was trained and tested using the 52 mean MFCCs extracted from the PCGs, as shown in Fig 2A. For the kNN classifier, $k = 3$ obtained better results than 1, 5, and 7 values. For that reason, only the results for the case that $k = 3$ are presented. Also, the polynomial kernel had better results for the SVM classifier than the linear and Gaussian kernels. For

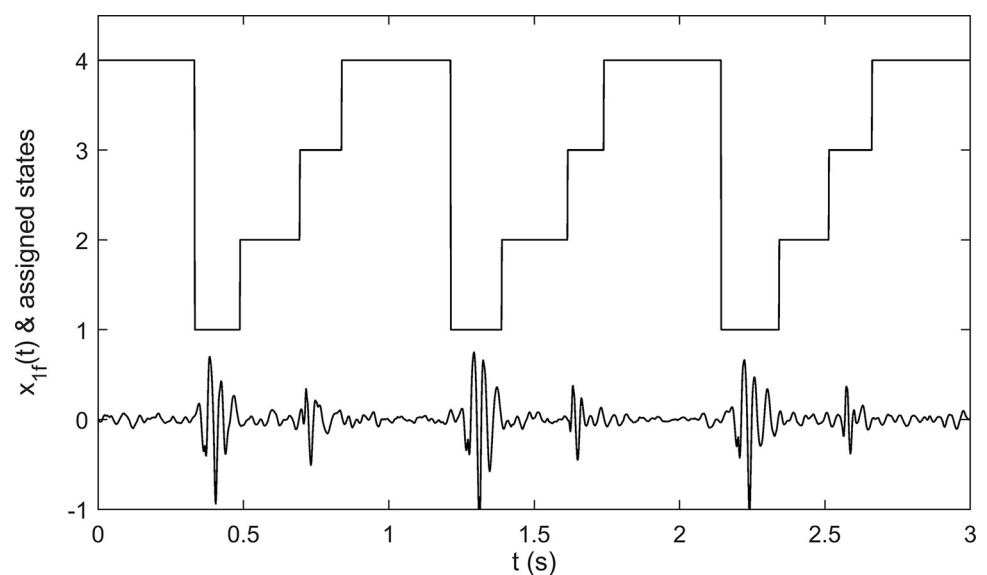


Fig 3. Segmentation of a PCG signal. The staircase graph shows the segmentation results of the plotted PCG signal: Level 1 shows S1 intervals, level 2 shows systole intervals, level 3 shows S2 intervals, and level 4 shows diastole intervals.

<https://doi.org/10.1371/journal.pone.0316645.g003>

Table 1. The results for the single-classifier strategy for the three types of classifiers: kNN ($k = 3$), SVM with a polynomial kernel, and DT.

Classifier	Acc (%)	Se (%)	Sp (%)	MAcc (%)
kNN	91.9	91.41	92.48	91.95
SVM	91.95	92.78	91.14	91.96
DT	87.33	86.72	88.03	87.34

<https://doi.org/10.1371/journal.pone.0316645.t001>

that reason, only the results of the SVM with the polynomial kernel are presented. The accuracy, sensitivity, and specificity parameters of the three classifiers (kNN, SVM, and DT) are presented in Table 1. Both the support vector machine ($Acc = 91.95\%$, $Se = 92.78\%$, $Sp = 91.14\%$) and the kNN ($Acc = 91.9\%$, $Se = 91.41\%$, $Sp = 92.48\%$) has outperformed the decision tree ($Acc = 87.33\%$, $Se = 86.72\%$, $Sp = 88.03\%$). There was no statistically significant difference between the accuracy of the SVM and the kNN algorithm. However, the SVM had a higher sensitivity than the kNN, while the kNN had a higher specificity. The 95% confidence intervals for the sensitivity of the SVM and the kNN were [92.31, 93.25] and [91.11, 91.71], respectively. Also, the 95% confidence intervals for the sensitivity of the SVM and the kNN were [90.67, 91.61] and [92.03, 92.93], respectively.

3.2 Results for the ensemble-classifier strategy

Similar to the single-classifier strategy, for the kNN classifier, $k = 3$ obtained better results than 1, 5, and 7 values. For that reason, only the results for the case that $k = 3$ are presented. Also, the polynomial kernel had better results for the SVM classifier than the linear and Gaussian kernels. Therefore, only the results of the SVM with a polynomial kernel are presented. The accuracy, sensitivity, and specificity parameters of the three classifiers (kNN, SVM, and DT) are presented in Table 2. The SVM had the highest accuracy (93.59%) and sensitivity (95.4%) while the kNN had the highest specificity (93.02%). The 95% confidence intervals for the accuracy of the kNN, SVM, and DT were [91.57, 92.11], [93.27, 93.91], [91.89, 92.55], respectively. The 95% confidence intervals for the sensitivity of the kNN, SVM, and DT were [90.63, 91.23], [95.05, 95.75], [93.64, 94.42], respectively. The 95% confidence intervals for the specificity of the kNN, SVM, and DT were [92.62, 93.42], [91.4, 92.22], [90.06, 91], respectively.

3.3 Comparison of single-classifier and ensemble-classifier strategies

As shown in Fig 4, when the classifier type was either DT or SVM, the ensemble-classifier strategy achieved a higher accuracy than the single-classifier strategy. In Fig 4, the error bars represent the 95% confidence intervals for the classification accuracy. On average, DT's accuracy improved by 4.89%, while SVM's improved by 1.64%. There was no statistically significant difference between the accuracy of the single- and ensemble-classifier strategies when the kNN was used.

As shown in Fig 5, when the classifier type was either DT or SVM, the ensemble-classifier strategy achieved a higher sensitivity than the single-classifier strategy. In Fig 5, the error bars

Table 2. The results for the ensemble classification strategy for three types of classifiers: kNN ($k = 3$), SVM with a polynomial kernel, and DT.

Classifier	Acc (%)	Se (%)	Sp (%)	MAcc (%)
kNN	91.84	90.93	93.02	91.98
SVM	93.59	95.40	91.81	93.61
DT	92.22	94.03	90.53	92.28

<https://doi.org/10.1371/journal.pone.0316645.t002>

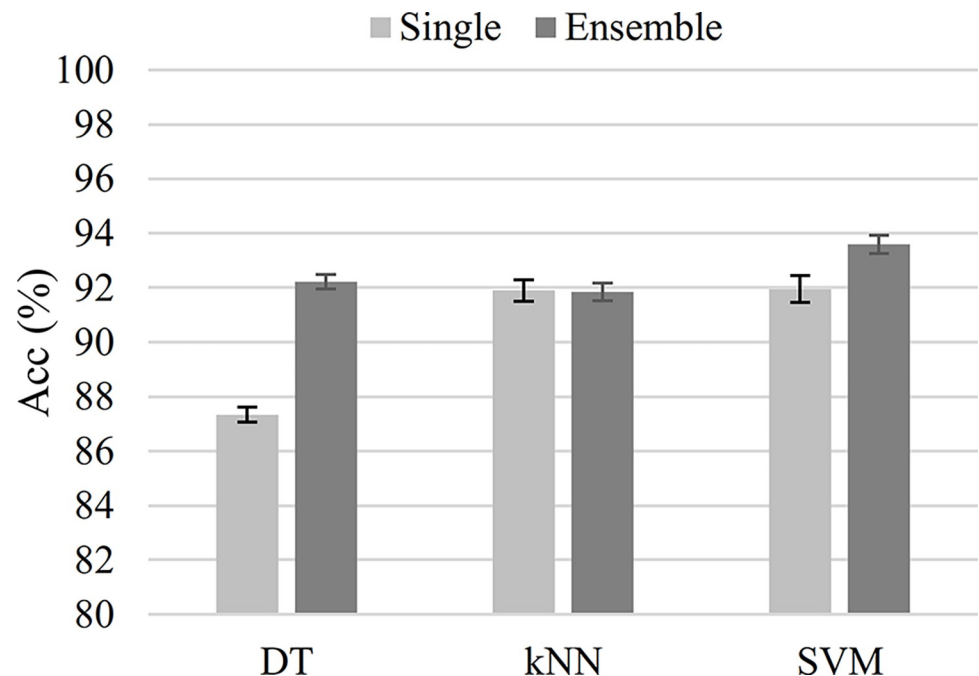


Fig 4. The classification accuracy for the single- (light gray) and the ensemble classifiers (dark gray). The ensemble classifier has outperformed the single classifier for both the decision tree (DT) and the SVM classifier types. The error bars represent the 95% confidence interval for the classification accuracy.

<https://doi.org/10.1371/journal.pone.0316645.g004>

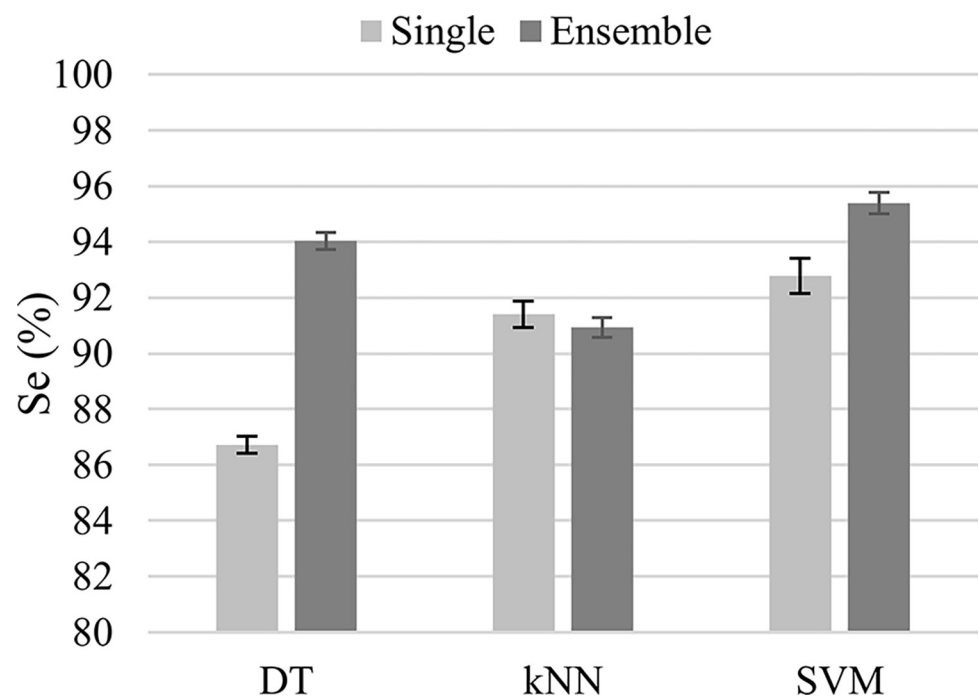


Fig 5. The classification sensitivity for the single- (light gray) and the ensemble classifiers (dark gray). The ensemble classifier has outperformed the single classifier for both the decision tree (DT) and the SVM classifier types. The error bars represent the 95% confidence interval for the classification sensitivity.

<https://doi.org/10.1371/journal.pone.0316645.g005>

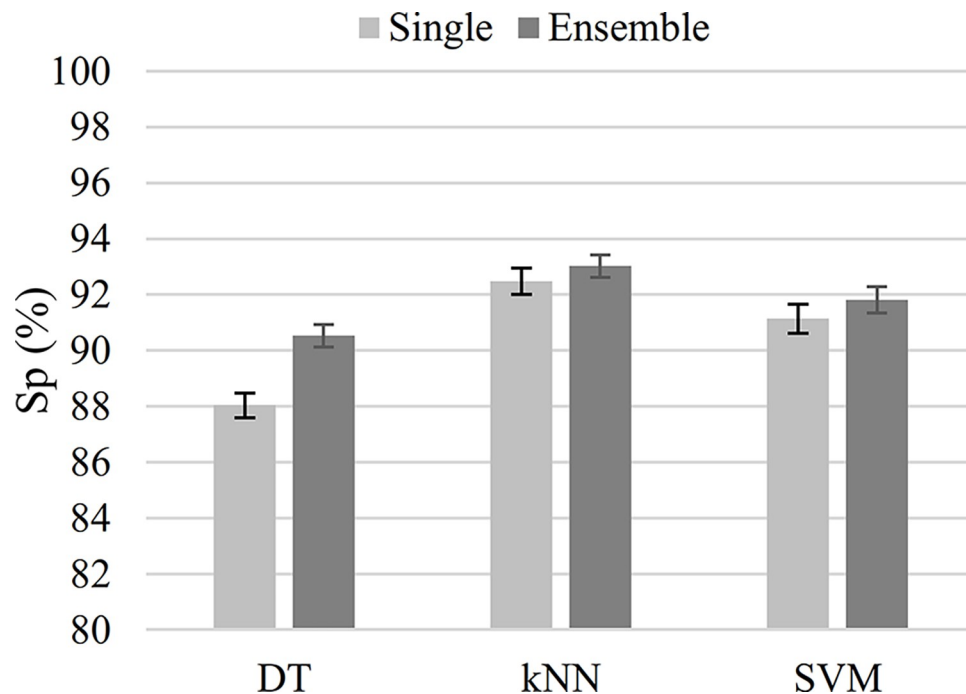


Fig 6. The classification specificity for the single- (light gray) and the ensemble classifiers (dark gray). The ensemble classifier has outperformed the single classifier only for the decision tree (DT) classifier type. The error bars represent the 95% confidence interval for the classification specificity.

<https://doi.org/10.1371/journal.pone.0316645.g006>

represent the 95% confidence intervals for the classification sensitivity. On average, DT's sensitivity improved by 7.31%, while SVM's sensitivity improved by 2.62%. There was no statistically significant difference between the sensitivity of the single- and ensemble-classifier strategies when the kNN was used.

As shown in Fig 6, when the classifier type was DT, the ensemble-classifier strategy achieved a higher specificity than the single-classifier strategy. In Fig 6, the error bars represent the 95% confidence intervals for the classification specificity. On average, DT's specificity improved by 2.5%. There was no statistically significant difference between the specificity of the single- and ensemble-classifier strategies when the kNN and SVM were used.

4 Discussion

The well-known MFCCs have been used in applications such as audio processing [38]. They have also been used for PCG processing [3,10,11,15]. In this study, two classification strategies were designed to investigate the performance of the MFCCs in detecting abnormal PCGs. In the single-classifier strategy, both the SVM ($Acc = 91.95\%$) and the kNN ($Acc = 91.9\%$) obtained higher classification accuracies than the DT ($Acc = 87.33\%$). Still, there was no significant difference between the kNN and SVM.

In the ensemble-classifier strategy, the SVM ($Acc = 93.59\%$) produced a higher classification accuracy than the kNN ($Acc = 91.84\%$) and DT ($Acc = 92.22\%$). Still, there was no significant difference between the kNN and DT. The results suggest that MFCC averaging in the single classifier strategy removed the discriminating cues required for abnormal PCG detection, and therefore should be avoided in similar applications.

When comparing the two classification strategies, the accuracy of the SVM and DT improved by 1.64% and 4.89% in the ensemble-classifier strategy, while the accuracy of the

kNN did not change significantly. Overall, the accuracy, sensitivity, and specificity of the DT increased considerably from the single-classifier strategy to the ensemble-classifier strategy. Also, the SVM classifier in the ensemble classifier achieved the highest classification accuracy (93.59%). One reason that the ensemble classifier outperformed the single-classifier could be because the averaging of MFCCs across multiple PCG beats in the latter strategy removes the individual beat differences, which are actually very important cues for detecting the abnormal PCGs. Another reason could be that averaging requires a very accurate timing for the PCG segments S1, systole, S2, and diastole with their respective segments in other beats. This accurate timing is almost impossible in practice as phonocardiograms are non-stationary. In a nutshell, the results suggest that the ensemble-classifier strategy is more efficient than the single-classifier strategy.

Table 3 compares the results of the current study with a few similar studies, which have used the PhysioNET CinC 2016 database (section 2.1) and/or MFCCs. The classification accuracies presented in Table 3 were either directly reported or estimated from the data reported in the respective articles. In the approach taken by Khan et al. (2020) [3], MFCCs were estimated from unsegmented signals, leading to a lower accuracy of 80.68%. Unlike that, our method benefits from the segmentation of phonocardiograms into distinct heart sound intervals (S1, systole, S2, and diastole), which likely contributes to our higher accuracies of 91.95% and 93.59% for single and ensemble classifiers, respectively. Though MFCCs were estimated from segmented signals in the method proposed by Milani et al. (2021) [10], S1 and S2 segmentation was based on a simple method of systole and diastole detection, leading to a low accuracy of 83.33%. The concatenation of MFCCs with time domain features increased the accuracy from 83.33% to 93.33%, comparable to the accuracy of 93.59% achieved by our ensemble classifier, increasing the complexity of their model while leaving its specificity ($Sp = 88.24\%$) much lower than that of our ensemble-classifier ($Se = 91.81\%$).

Table 3. Comparison of the results of this study with similar studies.

Study	Preprocessing	Segmentation	Database	Features	Classifier	Results (%)
[3]	BPF	[19]	[27,28]	T, T-F	ensemble of 100 DTs	Acc: 91.23
[3]	BPF	N/A	[27,28]	MFCCs	LSTM	Acc: 80.68
[10]	LPF	systole and diastole durations	[27,28]	MFCCs	ANN	Acc: 83.33
				T, MFCCs	ANN	Acc: 93.33
[8]	Zero-mean	N/A	[27,28]	SA & WE	DT	Acc: 79.33
[13]	BPF	[19]	[27,28]	T, F, W, ST	ensemble + voting (20)	Acc: 86.58
[12]	BPF	[19]	[27,28]	T, F, T-F,	DNN	Acc: 92.6
[9]	Savitzky–Golay LPF	N/A	[27,28]	1-D CNN	MLP	Acc: 85.65
					MLP	Acc: 85.65
[17]	PIFS	N/A	[27,28]	N/A	CNN	MAcc: 85
[11]	HPF & Amplitude Normalization	Shannon energy, envelope smoothing, peak finding	Private	T, F, T-F, MFCCs	kNN	Acc: 98.78
[15]	BPF	Mel-Scaled Wavelet Transform	Private	T, ST, MFCCs	MLP	Acc: 99.91
Current study	BPF	[19]	[27,28]	MFCCs	Single-classifier	MAcc: 91.96% Acc: 91.95%
					Ensemble-classifier	MAcc: 93.61% Acc: 93.59%

LPF: Low-pass filtering, HPF: High-pass filtering, BPF: Band-pass filtering, PIFS: Partitioned iterated function systems, MLP: Multi-layer Perceptron, DNN: Deep neural factor, T: Time-domain feature, F: Frequency-domain features, T-F: Time-frequency features, W: Wavelet features, SA: Spectral amplitude, WE: Wavelet entropy, ST: Statistical features.

<https://doi.org/10.1371/journal.pone.0316645.t003>

Features extracted in time, frequency, and time-frequency domains have also been used in applications of PCG processing [3,8,12,13]. Langley and Murray (2017) [8] used Spectral amplitude and wavelet entropy (2 features) to classify unsegmented PCGs, leading to an accuracy of 79.33%, much lower than our single-classifier (Acc = 91.95%) and ensemble-classifier strategy (Acc = 93.59%), once again confirming that signal segmentation is essential for PCG classification. Unlike the method proposed by Langley and Murray (2017) [8], the method proposed by Khan et al. (2020) [3] classified segmented phonocardiograms using time and time-frequency [3], reaching an accuracy of 91.23%. Similarly, Homsy and Warrick (2017) [13] used time, frequency, statistical, and wavelet features for the classification of segmented phonocardiograms. They reached an accuracy of (Acc = 86.58%). Sotaquirá et al. (2018) [12] also used similar features to classify normal/abnormal PCG cycles. They achieved an accuracy of (Acc = 92.6%) using deep learning. This significantly improves Homsy and Warrick (2017) [13], yet is lower than our ensemble classifier (Acc = 93.59%). The superiority of our method (esp. our ensemble classifier), which uses MFCCs, over the methods using time and time-frequency features implies that MFCCs provide a better representation for phonocardiograms than time and time-frequency features.

Some studies investigated the combination of MFCCs with time and time-frequency features to improve classification accuracy. For instance, concatenation of MFCCs with time features increased the classification accuracy from 83.33% to 93.33% in [10]. In [15], concatenating MFCCs with time features and statistical features led to an increased accuracy of 99.91%. Similarly, in [11], concatenating MFCCs with time features, and frequency features, time-frequency features, and wavelet features increased the classification accuracy of tricuspid regurgitation severity using PCGS to 98.78%. Though the PCG databases used in the last two studies differed from the PhysioNET CinC 2016 database, the findings suggest that concatenation of MFCCs with other features might be more effective for PCG applications. However, this improved accuracy comes at the cost of increasing the complexity of the systems.

In a different approach taken by Krishnan et al. (2020) [9], a 1-D convolutional network was proposed for feature extraction from unsegmented PCGs, reaching an accuracy of 85.65%, a very low sensitivity of 57.78%, and a specificity of 92.98%. In another experiment, they applied an MLP with 4 hidden layers, increasing the sensitivity from 57.78% to 86.73% while the accuracy remained unchanged. Similarly, Riccio et al. (2023) [17] reached a modified accuracy of 85% using a convolutional neural network. They used Partitioned Iterated Function Systems (PIFS) to generate 2D color images from 1D PCGs. These images were used as input for the CNN. The results of these two studies are much lower than our single-classifier (Acc = 91.95%) and ensemble-classifier strategy (Acc = 93.59%). This could be because a) they used unsegmented PCGs and/or b) the features extracted by deep neural nets are less efficient than MFCCs.

Fig 7 compares the accuracy, sensitivity, and specificity of the aforementioned methods (cited in Table 3) with our single and ensemble classifiers. Only six of these studies, which a) used the PhysioNET CinC 2016 database [27,28], and b) reported the accuracy, sensitivity, and specificity directly (or it was possible to estimate them from the reported data), were included in Fig 7. As can be seen in Fig 7, our ensemble classifier has the highest accuracy (Acc = 93.59%) and the second highest sensitivity (Se = 95.4%). The method proposed in [10] has the highest sensitivity (Se = 100%), but it should be emphasized that its specificity (Sp = 88.24%) is lower than that of our ensemble classifier (Sp = 91.84%). Finally, our ensemble classifier has the third highest specificity (Sp = 91.81%). The highest specificities were achieved by [3] (Sp = 97.04%) and [12] (Sp = 93.8%), but both have lower sensitivities (Se = 78.81%, and 91.3%, respectively) than our ensemble classifier (Se = 95.40%). Overall, it seems that our ensemble classifier has outperformed other studies.

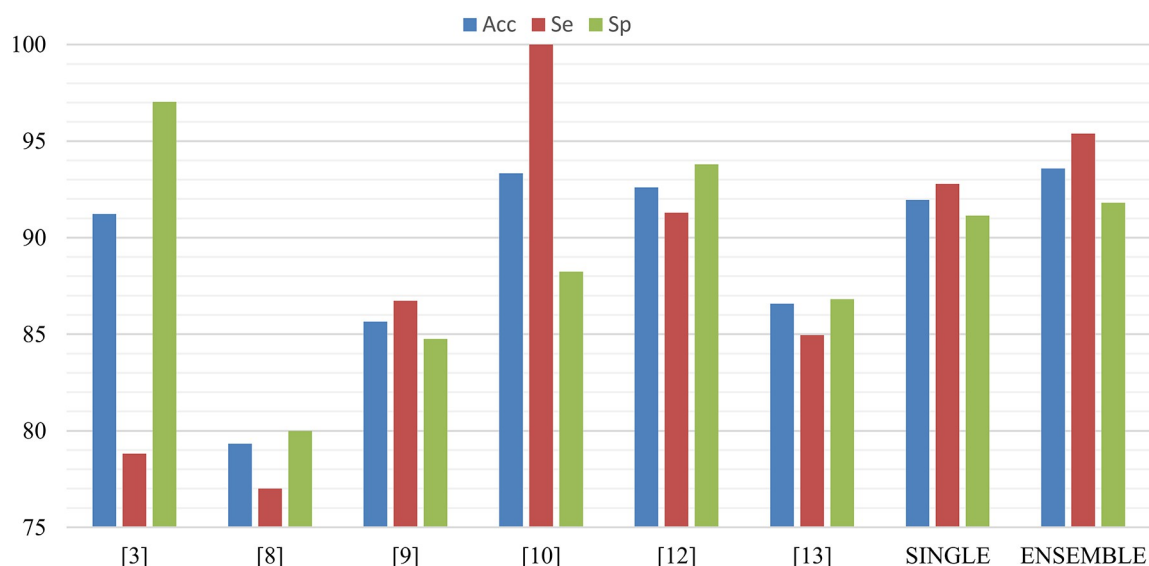


Fig 7. Comparison of the accuracy, sensitivity, and specificity of our ensemble and single classifiers with the some of the methods cited in Table 3.

<https://doi.org/10.1371/journal.pone.0316645.g007>

Though our ensemble-classifier achieved a high accuracy for PCG classification, our method has a number of limitations. First, as the segmentation algorithm we used was state-of-the-art, we did not evaluate the segmentation step. Second, the MFCCs capture only the spectral properties of the heart sounds, while as shown in previous studies [10,11,15], temporal features can contribute to the classification performance. Therefore, it is necessary to incorporate the temporal features of the heart sounds into our model in future. Third, since MFCCs were effective for phonocardiogram classification, it should be investigated whether they can effectively be used to develop a supervised segmentation algorithm. If so, the complexity of the proposed strategies will be reduced significantly. Last, although our results confirm that our ensemble classifier is very efficient for binary classification problems to discriminate abnormal phonocardiograms from normal ones, it is still necessary to evaluate it in multi-class classification problems to detect cardiovascular diseases.

5 Conclusion

The performance of MFCCs for detecting abnormal PCGs was evaluated using two classification strategies, i.e., a single-classifier strategy and an innovative ensemble-classifier strategy. In the single-classifier strategy, the MFCCs extracted from different PCG beats are first averaged, and the mean MFCCs are then used to classify PCGs. However, in the ensemble-classifier strategy, MFCCs are used by an ensemble of 9 classifiers to classify PCG beats into normal/abnormal beats. In the end, if most beats are classified as normal, the PCG is considered normal; otherwise, it is abnormal. Both strategies were tested on a publicly available database of PCG signals. The results showed that MFCCs were more effective than other features, including time, time-frequency, and statistical features, evaluated in similar studies and the ensemble-classifier strategy could classify PCGs with a higher accuracy, implying that the averaging of MFCCs should be avoided in similar studies.

Acknowledgments

The authors would like to acknowledge Deanship of Graduate Studies and Scientific Research, Taif University for funding this work. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1446).

Author Contributions

Conceptualization: Amir Haider, Amir Masoud Rahmani.

Data curation: Mohammad Adeli, Olfa Mzoughi, Hamid Alinejad-Rokny.

Investigation: Parisa Khoshvaght.

Methodology: Entesar Gemeay, Mokhtar Mohammadi.

Project administration: Mehdi Hosseinzadeh.

Resources: Mazhar Hussain Malik.

Validation: Mazhar Hussain Malik, Thantrira Porntaveetus.

Writing – original draft: Mehdi Hosseinzadeh, Amir Haider, Mohammad Adeli, Olfa Mzoughi, Entesar Gemeay, Mokhtar Mohammadi, Hamid Alinejad-Rokny, Parisa Khoshvaght, Amir Masoud Rahmani.

Writing – review & editing: Mazhar Hussain Malik, Thantrira Porntaveetus.

References

1. Webster J.G., Encyclopedia of medical devices and instrumentation. 2nd ed. Vol. 5. 1990, Hoboken, New Jersey: John Wiley & Sons, Inc.
2. Deperlioglu O., Classification of phonocardiograms with convolutional neural networks. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 2018. 9(2): p. 22–33.
3. Khan F.A., Abid A., and Khan M.S., Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. *Physiological measurement*, 2020. 41(5): p. 055006. <https://doi.org/10.1088/1361-6579/ab8770> PMID: 32259811
4. Han W., Yang Z., Lu J., and Xie S., Supervised threshold-based heart sound classification algorithm. *Physiological Measurement*, 2018. 39(11): p. 115011. <https://doi.org/10.1088/1361-6579/aae7fa> PMID: 30500785
5. Altaf A., Mahdin H., Alive A.M., Ninggal M.I.H., Altaf A., and Javid I., Systematic Review for Phonocardiography Classification Based on Machine Learning. *International Journal of Advanced Computer Science and Applications*, 2023. 14(8).
6. Chowdhury T.H., Poudel K.N., and Hu Y., Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals. *Ieee Access*, 2020. 8: p. 160882–160890.
7. Dwivedi A.K., Imtiaz S.A., and Rodriguez-Villegas E., Algorithms for automatic analysis and classification of heart sounds—a systematic review. *IEEE Access*, 2018. 7: p. 8316–8345.
8. Langley P. and Murray A., Heart sound classification from unsegmented phonocardiograms. *Physiological measurement*, 2017. 38(8): p. 1658. <https://doi.org/10.1088/1361-6579/aa724c> PMID: 28489019
9. Krishnan P.T., Balasubramanian P., and Umapathy S., Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Physical and Engineering Sciences in Medicine*, 2020. 43(2): p. 505–515. <https://doi.org/10.1007/s13246-020-00851-w> PMID: 32524434
10. Milani M., Abas P.E., De Silva L.C., and Nanayakkara N.D., Abnormal heart sound classification using phonocardiography signals. *Smart Health*, 2021. 21: p. 100194.
11. Rujoie A., Fallah A., Rashidi S., Khoshnood E.R., and Ala T.S., Classification and evaluation of the severity of tricuspid regurgitation using phonocardiogram. *Biomedical Signal Processing and Control*, 2020. 57: p. 101688.
12. Sotaquirá M., Alvear D., and Mondragon M., Phonocardiogram classification using deep neural networks and weighted probability comparisons. *Journal of medical engineering & technology*, 2018. 42(7): p. 510–517. <https://doi.org/10.1080/03091902.2019.1576789> PMID: 30773957
13. Homsy M.N. and Warrick P., Ensemble methods with outliers for phonocardiogram classification. *Physiological measurement*, 2017. 38(8): p. 1631. <https://doi.org/10.1088/1361-6579/aa7982> PMID: 28613208
14. Vernekar, S., S. Nair, D. Vijaysenan, and R. Ranjan. A novel approach for classification of normal/abnormal phonocardiogram recordings using temporal signal analysis and machine learning. in 2016 computing in cardiology conference (CinC). 2016. IEEE.

15. Ozkan, I. and A. Yilmaz. Performance of using Mel-Frequency Cepstrum Based Features in Nonlinear Classifiers for Phonocardiography Recordings. in 2023 31st European Signal Processing Conference (EUSIPCO). 2023. IEEE.
16. Jaros R., Koutny J., Ladrova M., and Martinek R., Novel phonocardiography system for heartbeat detection from various locations. *Scientific Reports*, 2023. 13(1): p. 14392. <https://doi.org/10.1038/s41598-023-41102-8> PMID: 37658080
17. Riccio D., Brancati N., Sannino G., Verde L., and Frucci M., CNN-based classification of phonocardiograms using fractal techniques. *Biomedical Signal Processing and Control*, 2023. 86: p. 105186.
18. Schmidt S.E., Holst-Hansen C., Graff C., Toft E., and Struijk J.J., Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological measurement*, 2010. 31(4): p. 513. <https://doi.org/10.1088/0967-3334/31/4/004> PMID: 20208091
19. Springer D.B., Tarassenko L., and Clifford G.D., Logistic regression-HSMM-based heart sound segmentation. *IEEE transactions on biomedical engineering*, 2015. 63(4): p. 822–832. <https://doi.org/10.1109/TBME.2015.2475278> PMID: 26340769
20. Alonso-Arévalo M.A., Cruz-Gutiérrez A., Ibarra-Hernández R.F., García-Canseco E., and Conte-Galván R., Robust heart sound segmentation based on spectral change detection and genetic algorithms. *Biomedical Signal Processing and Control*, 2021. 63: p. 102208.
21. Milani M.M., Abas P.E., and De Silva L.C., A critical review of heart sound signal segmentation algorithms. *Smart Health*, 2022. 24: p. 100283.
22. Hamidi M., Ghassemian H., and Imani M., Classification of heart sound signal using curve fitting and fractal dimension. *Biomedical Signal Processing and Control*, 2018. 39: p. 351–359.
23. Renna F., Oliveira J., and Coimbra M.T., Deep convolutional neural networks for heart sound segmentation. *IEEE journal of biomedical and health informatics*, 2019. 23(6): p. 2435–2445. <https://doi.org/10.1109/JBHI.2019.2894222> PMID: 30668487
24. Fahad H., Ghani Khan M.U., Saba T., Rehman A., and Iqbal S., Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM. *Microscopy research and technique*, 2018. 81(5): p. 449–457. <https://doi.org/10.1002/jemt.22998> PMID: 29363219
25. Davis S. and Mermelstein P., Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 1980. 28(4): p. 357–366.
26. Anusuya M. and Katti S., Front end analysis of speech recognition: a review. *International Journal of Speech Technology*, 2011. 14: p. 99–145.
27. Liu C., Springer D., Li Q., Moody B., Juan R.A., Chorro F.J., Castells F., Roig J.M., Silva I., and Johnson A.E., An open access database for the evaluation of heart sound algorithms. *Physiological measurement*, 2016. 37(12): p. 2181. <https://doi.org/10.1088/0967-3334/37/12/2181> PMID: 27869105
28. Goldberger A.L., Amaral L.A., Glass L., Hausdorff J.M., Ivanov P.C., Mark R.G., Mietus J.E., Moody G.B., Peng C.-K., and Stanley H.E., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 2000. 101(23): p. e215–e220. <https://doi.org/10.1161/01.cir.101.23.e215> PMID: 10851218
29. Syed Z.H., MIT automated auscultation system. 2003, Massachusetts Institute of Technology.
30. Syed Z., Leeds D., Curtis D., Nesta F., Levine R.A., and Gutttag J., A framework for the analysis of acoustical cardiac signals. *IEEE Transactions on Biomedical Engineering*, 2007. 54(4): p. 651–662. <https://doi.org/10.1109/TBME.2006.889189> PMID: 17405372
31. Schmidt S.E., Holst-Hansen C., Hansen J., Toft E., and Struijk J.J., Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering*, 2015. 62(11): p. 2611–2619. <https://doi.org/10.1109/TBME.2015.2432129> PMID: 25974927
32. Papadaniil C.D. and Hadjileontiadis L.J., Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. *IEEE journal of biomedical and health informatics*, 2014. 18(4): p. 1138–1152. <https://doi.org/10.1109/JBHI.2013.2294399> PMID: 25014929
33. Naseri H., Homaeinezhad M.R., and Pourkhajeh H., Noise/spike detection in phonocardiogram signal as a cyclic random process with non-stationary period interval. *Computers in biology and medicine*, 2013. 43(9): p. 1205–1213. <https://doi.org/10.1016/j.combiomed.2013.05.020> PMID: 23930815
34. Moukadem A., Dieterlen A., Hueber N., and Brandt C., A robust heart sounds segmentation module based on S-transform. *Biomedical Signal Processing and Control*, 2013. 8(3): p. 273–281.
35. Tang H., Li T., Park Y., and Qiu T., Separation of heart sound signal from noise in joint cycle frequency–time–frequency domains based on fuzzy detection. *IEEE Transactions on Biomedical Engineering*, 2010. 57(10): p. 2438–2447. <https://doi.org/10.1109/TBME.2010.2051225> PMID: 20542764
36. Samieinasab, M. and R. Sameni. Fetal phonocardiogram extraction using single channel blind source separation. in 2015 23rd Iranian Conference on Electrical Engineering. 2015. IEEE.

37. Fedorov S.V., The Mystery and Clarity of Leonardo da Vinci's Coefficient of Friction. *Journal of Materials*, 2023. 1(1): p. 8–20.
38. Abdul Z.K. and Al-Talabani A.K., Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 2022. 10: p. 122136–122158.