

Improved pre-test likelihood estimation of coronary artery disease using phonocardiography

Bjarke Skogstad Larsen ^{1,*}, Simon Winther ^{2,3}, Louise Nissen ^{2,3},
Axel Diederichsen ⁴, Morten Bøttcher ^{2,3}, Matthias Renker ⁵,
Johannes Jan Struijk ¹, Mads Græsbøll Christensen ⁶,
and Samuel Emil Schmidt ¹

¹Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 7, 9220, Aalborg, Denmark; ²Department of Cardiology, Gødstrup Hospital, Herning, Denmark; ³Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; ⁴Department of Cardiology, Odense University Hospital, Odense, Denmark; ⁵Department of Cardiology, Kerckhoff Heart and Thorax Center, Bad Nauheim, Germany; and ⁶Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark

Received 13 April 2022; revised 22 August 2022; accepted 19 September 2022; online publish-ahead-of-print 10 October 2022

Aims

Current early risk stratification of coronary artery disease (CAD) consists of pre-test probability scoring such as the 2019 ESC guidelines on chronic coronary syndromes (ESC2019), which has low specificity and thus rule-out capacity. A newer clinical risk factor model (risk factor-weighted clinical likelihood, RF-CL) showed significantly improved rule-out capacity over the ESC2019 model. The aim of the current study was to investigate if the addition of acoustic features to the RF-CL model could improve the rule-out potential of the best performing clinical risk factor models.

Methods and results

Four studies with heart sound recordings from 2222 patients were pooled and distributed into two data sets: training and test. From a feature bank of 40 acoustic features, a forward-selection technique was used to select three features that were added to the RF-CL model. Using a cutoff of 5% predicted risk of CAD, the developed acoustic-weighted clinical likelihood (A-CL) model showed significantly ($P < 0.05$) higher specificity of 48.6% than the RF-CL model (specificity of 41.5%) and ESC 2019 model (specificity of 6.9%) while having the same sensitivity of 84.9% as the RF-CL model. Area under the curve of the receiver operating characteristic for the three models was 72.5% for ESC2019, 76.7% for RF-CL, and 79.5% for A-CL.

Conclusion

The proposed A-CL model offers significantly improved rule-out capacity over the ESC2019 model and showed better overall performance than the RF-CL model. The addition of acoustic features to the RF-CL model was shown to significantly improve early risk stratification of symptomatic patients suspected of having stable CAD.

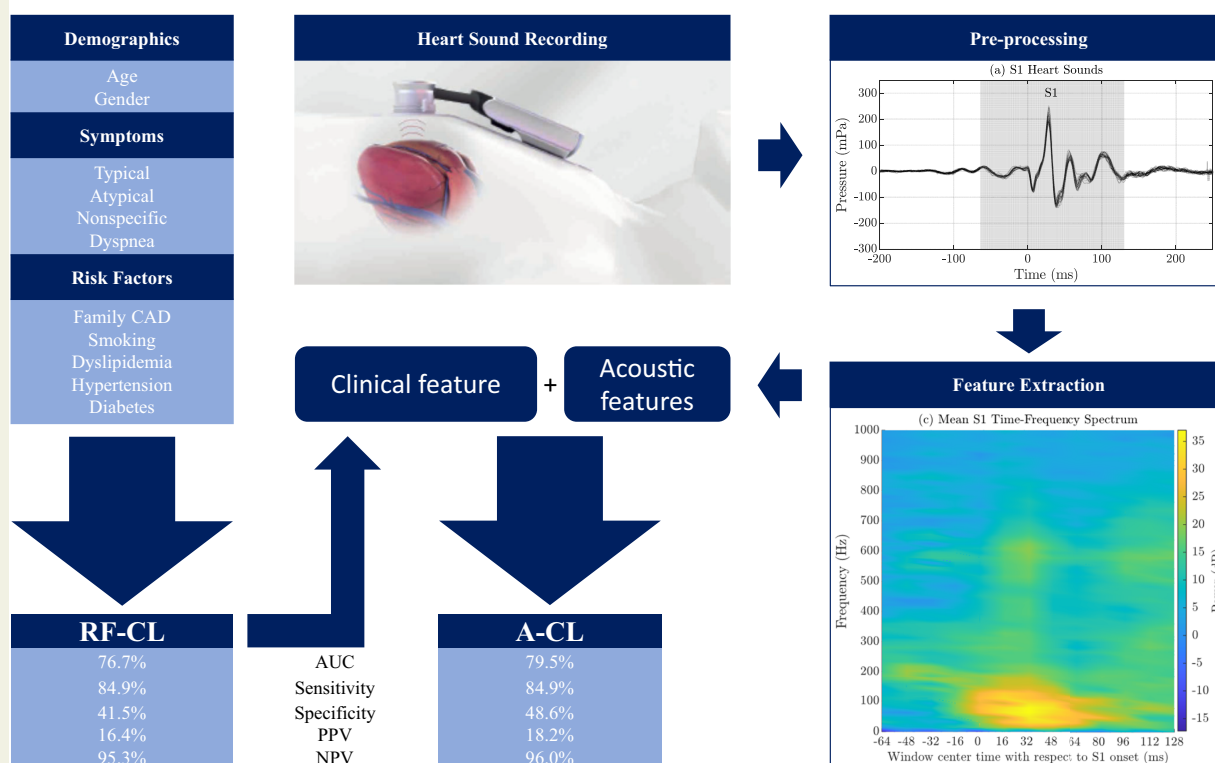
* Corresponding author. Tel: +45 26810869, Email: bjarke.skogstad@acarix.com

© The Author(s) 2022. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Graphical Abstract

Aims and objectives: to investigate if an advanced clinical risk model for early rule-out of patients with suspected coronary artery disease can be improved with acoustic features.



Keywords

Coronary artery disease • Chronic coronary syndrome • Heart sounds • Phonocardiography • Clinical rule-out • Pre-test probability

Introduction

Coronary artery disease (CAD) has long been a leading cause of death worldwide,¹ and as such it is likely one of the top concerns for doctors when examining patients with symptoms suggestive of the disease. This raised awareness and concern might be one of the reasons why there is a low prevalence (6–12%) of positive diagnostic test results among *de novo* symptomatic patients referred for further investigation of CAD.^{2–4} As a result, the remaining 88–94% (non-CAD) patients carry both unnecessary additional costs of healthcare services as well as stress, time, and sometimes risk of complications during invasive testing.

There are a variety of diagnostic methods for investigation of patients with suspected CAD, and the diagnostic pathway differs across countries. However, in general when a patient first experiences non-acute symptoms, they visit their general practitioner or outpatient clinics for initial investigation. Here they are typically assessed for risk of CAD based on a variety of pre-test probability (PTP) scores as well as the patient's anamnesis. One such PTP score is defined in the 2019 ESC guidelines on chronic coronary syndromes (ESC2019).⁵

Although this score is an improvement over previous scores such as the Diamond–Forrester score,⁶ it still has a weak rule-out capability, which is reflected by its high sensitivity and concurrent low specificity.^{7,8}

Winther *et al.*⁹ developed two extended models of the PTP score recommended by the 2019 ESC guidelines by incorporating the number of additional risk factors, and secondly by also adding coronary artery calcium score (CACS). With a cutoff value of 5% for predicted CAD, the ESC2019 PTP model based on age, sex, and symptoms had a relatively low specificity of 12.1% and generally overestimated the probability of CAD. Including the number of risks factors to create the risk factor-weighted clinical likelihood (RF-CL) model increased the specificity to 41.5% in the international validation cohort. When further adding CACS to create the CACS-weighted clinical likelihood (CACS-CL) model, the specificity increased to 59.3%.

Both these models significantly increase the rule-out potential of CAD compared with the 2019 ESC guidelines. The RF-CL model only requires testing if some factors (blood pressure, diabetes, and dyslipidaemia) are not known and can generally be utilized at the first point of examination. Though the CACS-CL model has an impressive rule-out potential of CAD, it also requires more expensive and

specialized testing equipment to obtain CACS, which generally is not available at an initial investigation. For this reason, it is more likely that the RF-CL model could be widely used for early rule-out of CAD.

Whereas CACS is not a feasible addition to an initial investigation for CAD, heart sound analysis promises a low-cost, non-invasive, and widely available method for further improving early rule-out of CAD. This is due to the presence of diastolic murmurs arising from turbulent blood flow following a coronary stenosis, as first reported by Dock and Zonerach¹⁰ in 1967. As most coronary stenoses are not audible by auscultation at the chest surface, heart sound analysis is a necessary step to detect these diastolic murmurs.

Phonocardiography (PCG) as a method for detecting CAD was first proposed by Semmlow *et al.*¹¹ in 1983 who identified that the presence of CAD was associated with increased magnitude of the frequency band 120–200 Hz. These findings have subsequently been confirmed by several other studies,^{12–17} though there is some disagreement on the exact frequency band.

The field has since expanded to cover time, frequency, time–frequency, and non-linear domains using a multitude of filtering and feature extraction techniques. Akay *et al.*¹⁸ compared features extracted using Fast Fourier Transform, auto-regressive and auto-regressive moving average models, and minimum norm. Zhidong¹⁹ estimated the instantaneous frequency of the diastolic period using the Hilbert Huang Transform to observe differences in patients before and after angioplasty. Schmidt *et al.*¹⁴ evaluated nine different feature classes for their potential to classify CAD. A more recent development has been the application of advanced machine-learning techniques such as convolutional neural networks and deep learning which have shown very promising performances, though the results have generally been based on limited data sets.^{20,21}

Studies in the field have largely focused on examining the diastolic period of the PCG, as it is in this period the blood flow through the coronary arteries is maximal and thus should be associated with the clearest murmur from turbulent flow following a stenosis. However, some newer studies have investigated other segments of the heart sound, either by segment-wise analysis (S1, S2, systole, or diastole) or by evaluating the entire heartbeat through neural networks.^{20–23} These studies have demonstrated that valuable information related to CAD is available outside the diastolic period.

Acoustic features have previously been used and combined with clinical risk factors to create a rule-out method for CAD.^{4,24–27} However, the increased performance of newer clinical risk factor models motivates further examination of the potential of heart sound analysis to improve these models. The aim of this article was to explore the possibility to improve the rule-out potential of the current clinical standard for risk stratification of patients with CAD symptoms (ESC2019⁵) beyond the performance of the RF-CL model developed by Winther *et al.* by using PCG as an addition to clinical risk factors.

Methods

Data

A pooled data set was created from four studies: AdoptCAD ($n = 255$)¹⁵ (ClinicalTrials.gov number NCT01564628); Dan-NICAD ($n = 1675$)^{4,28} (ClinicalTrials.gov number NCT02264717); BIO-CAC ($n = 661$)^{29,30} (ClinicalTrials.gov number NCT02913144); and VALIDATE ($n = 226$)³¹

(drks.de number DRKS00010492). The prevalence of CAD patients varied highly between the four studies, which reflects the different aims of the studies as well as where in the diagnostic pathway patients were recruited.

AdoptCAD and Dan-NICAD were approved by the Regional Committees on Health Research Ethics for Central Denmark, and BIO-CAC was approved by the Regional Committees on Health Research Ethics for Southern Denmark. VALIDATE was approved by Justus Liebig University Giessen Ethics Committee for the Medical Department in Germany. All studies were conducted according to the Helsinki Declaration, and written informed consent was obtained from all patients.

The procedure for recording heart sounds was the same for all studies. Following a resting period of 5 min, heart sounds were recorded from patients in supine position using an Acarix CADScor device at the left fourth intercostal space during four breath-hold periods of 8 s each. Patients would hold their breath during recording to eliminate the possibility of breathing noise contaminating the heart sounds. The device records with a sampling frequency of 8000 Hz and a resolution of 16 bits.

Patients from the four studies were only included if they had a heart sound recording which passed pre-qualification and at least five fully annotated heart beats. Pre-qualification consisted of the following criteria: background noise level is below 65 dB, heart sound level is above 60 dB, EMC noise is below 65 dB, heart sounds are recorded and segmented, and heart beats come at regular intervals. They were then classified into one of two diagnostic categories: CAD and Other. Patients with an invasive coronary angiographic (ICA) identified stenosis with at least 50% diameter reduction were classified as CAD; the remaining patients were classified as Other.

The pooled data set was divided into a training data set (80%) and a test data set (20%) using stratified randomization with,³² balancing for study (AdoptCAD, Dan-NICAD, BIO-CAC, and VALIDATE), gender (Male, Female), and diagnosis (CAD, Other). This stratification allowed for the creation of two comparable data sets even though the individual studies have highly different prevalence of CAD as shown in [Supplementary material online, Table S1\(a\)](#) and (b). Model training was done exclusively using the training data set, whereas the test data set was only used for final verification of the trained model.

Existing models

The PTP model recommended by ESC2019 is based on the Diamond–Forrester approach using age, sex, and symptoms as predictive variables. The PTP scores were determined using Table 5 in Knuuti *et al.*⁵ and will be referred to as the ESC2019 model going forward.

The RF-CL model developed by Winther *et al.*⁹ is a logistic regression model with predictive variables for age, sex, and symptoms as well as the number (0–5) of the following risk factors: family history of CAD, smoking, dyslipidaemia, hypertension, and diabetes.

The RF-CL model is reproduced below in Equation (1):

$$\begin{aligned}
 s = & 1.6128 * \text{sex} + 0.08440 * \text{age} + 2.7112 * \text{symp_typical} \\
 & - 0.4675 * \text{symp_non_angina} + 1.4940 \\
 & * \text{nb_rf} - 0.0187 * \text{age} * \text{symp_typical} - 0.0131 * \text{age} * \text{nb_rf} - 0.2799 \\
 & * \text{symp_typical} * \text{nb_rf} - 0.2091 * \text{sex} * \text{nb_rf}
 \end{aligned}
 \tag{1}$$

Sex is 1 when male and otherwise 0, age is the subject age, symp_typical is 1 if the subject has typical symptoms and 0 otherwise, symp_non_angina is 1 if the subject has non-anginal symptoms and 0 otherwise, and nb_rf is the number of previously mentioned risk factors (0–5). Risk factors and symptoms were defined in the same way as detailed in Winther *et al.*⁹

Acoustic features

A feature bank consisting of 40 acoustic features was created using 8 previously developed features,^{4,14} and 32 newly developed features based on time–frequency components of S1 and S2 segments which were previously shown in Larsen *et al.*²² to have statistical significance in distinguishing between CAD and non-CAD patients.

In developing the new features, heart sound recordings underwent pre-processing steps as detailed in Larsen *et al.*,²² which included filtering, noise cancellation, segmentation, annotation, and alignment to S1 and S2 for extraction of time–frequency features of the first and second heart sound, respectively. Time–frequency spectra were obtained for both S1 and S2 heart sounds by estimating the power spectral density (PSD) using short-time Fourier transform with 64 ms windows and 16 ms steps with recordings aligned to the onset of S1 and S2, respectively. The PSDs were combined to form a time–frequency resolution (TFR) of the respective heart sounds as shown in *Figure 1*, and the logarithm of the mean of all heart beats for each subject was used for further analysis. The TFR window centres ranged from –64 to 128 ms for S1 and –64 to 96 ms for S2 with a frequency range of 0–1000 Hz. A longer window was chosen for the S1 than S2 as S1 is typically of longer duration.

Time–frequency candidate features were extracted by taking the mean of the TFR within windows around components of interest which have shown statistically significant power in distinguishing between CAD and non-CAD patients.²²

The full list of features investigated in this study can be seen in [Supplementary material online, Tables S2 and S3](#).

Feature selection

The RF-CL model was used as the starting point for the acoustic-weighted clinical likelihood (A-CL) model developed in this study. Thus, acoustic candidate features were evaluated for their capacity to improve the performance of the RF-CL model, meaning that though acoustic features had clear independent classification power, they were omitted in the feature selection process if this independent performance did not translate into model improvement.

A baseline predictive variable (RF_{CL}) was computed from the RF-CL model to be used for updating the baseline model with acoustic features and was calculated using the coefficients and predictive variables for the RF-CL formula in Winther *et al.*⁹ as shown in Equation (1). A logistic regression model with the RF_{CL} feature was the starting point for adding acoustic features. From this starting point, features were selected iteratively using a modified forward-selection algorithm with five steps in each iteration:

- Steps 1 and 2: Add the feature resulting in the greatest mean area under the curve (AUC) using 5×5 -fold cross-validation on the training data set.
- Steps 3 and 4: Remove the feature with the highest P -value above 0.1.
- Step 5: Terminate if model has not changed since previous iteration, otherwise perform another iteration starting from Step 1.

When adding features to the model in Steps 1 and 2, candidate features were tested exhaustively in each step. Five five-fold cross-validations were performed in both steps, meaning that for each of the five cross-validations, the training data set was divided randomly into five groups. Four of these groups were used for model training and the last group was used for evaluating model performance. The feature resulting in the best average model performance of the five five-fold cross-validations was selected for each of Steps 1 and 2. The AUC of the receiver operating characteristic (ROC) was used as performance measure when evaluating features' capacity to improve the RF-CL model.

Likewise, in Steps 3 and 4, features were removed one at a time, each time evaluating the P -value of the features in the model. Only the feature

with the highest P -value above 0.1 was removed. If no feature had a P -value above 0.1, no feature was removed.

Feature selection was done exclusively using the training data set.

Statistical analysis

Variables are expressed as mean (\pm standard deviation) and categorical variables are reported as frequencies (percentages). The AUC of the ROC curve was calculated for each of the models and compared with the method described by DeLong *et al.*³³ Additionally, partial AUC (pAUC) comparison was made for sensitivity (SE) > 0.7 using a bootstrap test as implemented by Robin *et al.*³⁴ for two correlated ROC curves. Performance values for sensitivity, specificity, positive, and negative predictive values (PPVs and NPVs) were calculated with cutoff of 5%. All performance values are presented with 95% confidence intervals.

Patients were classified into one of three risk groups with cutoffs in parentheses: low ($\leq 5\%$), intermediate ($> 5\text{--}15\%$), and high ($> 15\%$). Net reclassification improvement (NRI) was calculated using the method described by Pencina *et al.*³⁵ for reclassification from ESC2019 to RF-CL and A-CL, respectively, using the three risk groups. NRI is a measure of how well a model correctly reclassifies subjects compared with another model with a score ranging from –1 to 1, where a negative score suggests an incorrect reclassification and a positive score suggests a successful reclassification. Finally, the χ^2 test was used for comparison of proportions. Statistical analyses were performed on the test data set using MATLAB R2021a (MathWorks, USA) except for the partial AUC comparison, which was done in R version 4.1.0.

Results

Out of the initial 2817 patients available from the four pooled studies, 533 (18.9%) patients were excluded because either no heart sound recording was available or the heart sound recording failed pre-analysis qualification of the Acarix CADScor heart sound processing framework. Additional 62 (2.2%) patients were excluded because their heart sound recordings had fewer than five fully annotated heartbeats. This resulted in a pooled data set of 2222 patients who were included in the current analysis as summarized in *Table 1* with patient demographics, risk factors, and symptoms for CAD as summarized in *Table 2*.

Feature selection

The feature selection process finalized after four iterations with the end of the fourth and third iterations having the same selected features. The final model formula for the A-CL model was as follows in Equation (2):

$$s = \frac{1}{1 + \exp(-(-8.70 + 1.14 RF_{CL} + 2.20 \text{ SampEn} - 0.0634 S1_4 + 0.100 S1_{21}))} \quad (2)$$

The three acoustic features included in the A-CL model are:

- SampEn: Sample entropy as described in Winther *et al.*⁴
- $S1_4$: Mean power of time–frequency components: window centre at 16 ms after the onset of S1 in the frequency range 400–1000 Hz.
- $S1_{21}$: Mean power of time–frequency components: window centre at –16, 0, and 16 ms relative to the onset of S1 in the frequency range 800–1000 Hz.

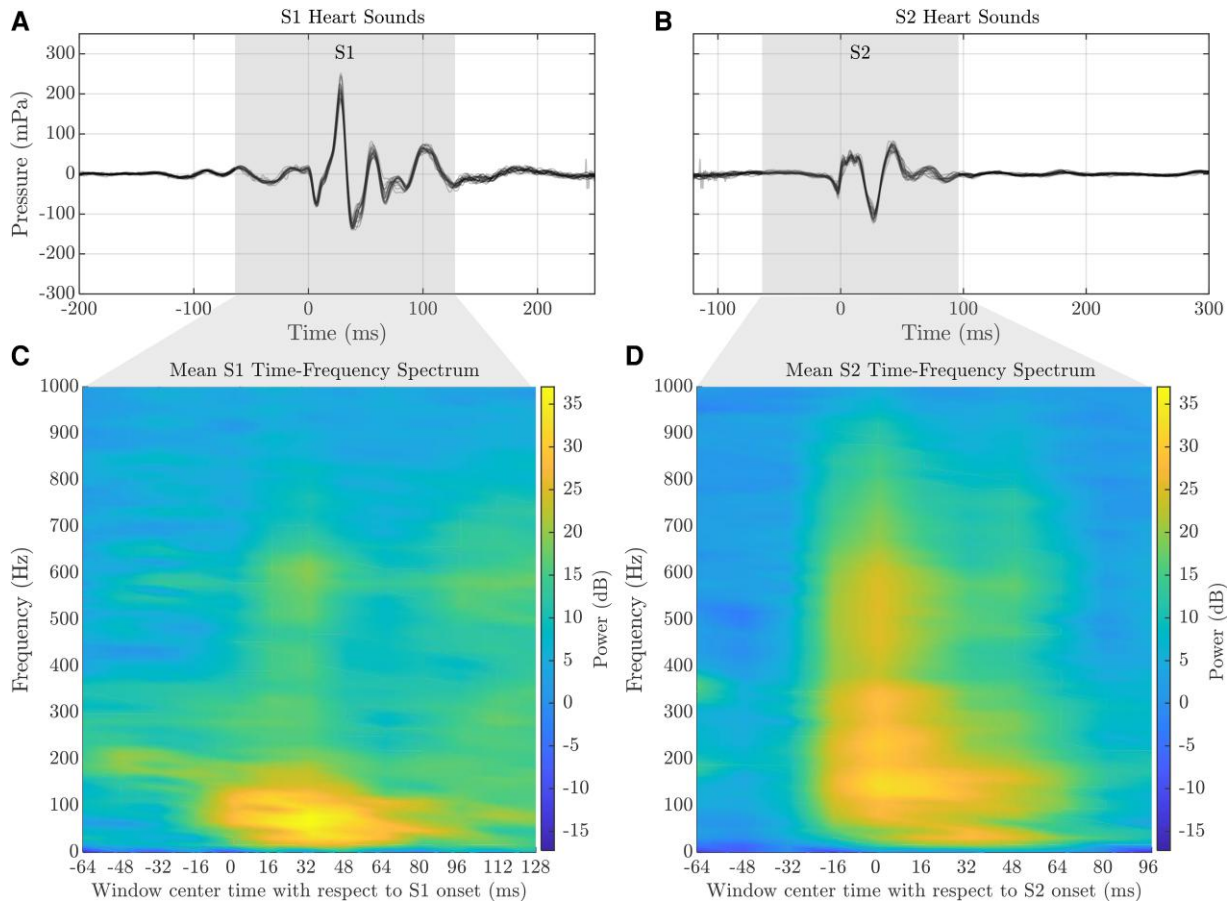


Figure 1 Example of how a heart sound recording from one subject was first segmented into heartbeats and aligned to the onset of S1 and S2 for the respective segments. Subsequently, mean time–frequency spectra were extracted for of the two segments, and the mean spectrum for each subject was used for further analysis and feature extraction. Note that the timings in the figure are with respect to the onset of S1 for figures (A) and (C) and with respect to the onset of S2 for figures (B) and (D). Additionally, spectra in (C) and (D) were obtained after whitening filtering.

Model performance

The performance of the three models on the training and test data sets are shown in the ROC curves in [Figure 2\(A and B\)](#), and the diagnostic accuracy with a cutoff of 5% as evaluated on the test data set is shown in [Figure 3](#). The AUCs for the three models using the test data set were 72.5% (CI 64.4–80.6%) (ESC2019), 76.7% (CI 69.0–84.5%) (RF-CL), and 79.5% (CI 72.0%–86.9%) (A-CL), showing increasing AUC with the addition of new factors for both RF-CL and A-CL models. In both the training and test data sets, the A-CL model outperformed the RF-CL model by 1.6 percentage points and 2.9 percentage points, respectively, and the improvement is situated similarly for the two ROC curves: $SE > 0.7$.

Evaluating model performance on the test data set, both the RF-CL and A-CL models showed significantly ($P < 0.05$) higher AUC than the ESC2019 model. The A-CL model did not show significantly ($P > 0.05$) higher AUC than the RF-CL model; however, pAUC comparison for $SE > 0.7$ showed significantly ($P < 0.05$) higher pAUC for the A-CL model compared with the RF-CL model. See [Supplementary material](#)

[online, Tables S4 and S5](#) for further details on AUC and pAUC comparisons.

[Figure 3](#) shows the diagnostic accuracy of the three models based on a cutoff of 5% as evaluated on the test data set. The results reaffirm the low rule-out power of the ESC2019 model with a specificity of 6.9%, though with a sensitivity of 100%. Both the RF-CL and A-CL models showed significantly ($P < 0.05$) higher specificity of 41.5% and 48.6%, respectively, with the A-CL additionally showing significantly ($P < 0.05$) higher specificity over the RF-CL model. The RF-CL and A-CL models had the same sensitivity of 84.9% but with the A-CL model having superior PPV and NPV.

Risk classification

As illustrated in [Figure 4](#), the A-CL model reclassified a higher portion of patients to the low- and high-risk groups, leaving fewer patients in the intermediate risk group. Additionally, the prevalence of CAD in the Low-risk group was somewhat lower; from 4.7% for the RF-CL model to 4.0% for the A-CL model.

[Table 3](#) shows the risk classification for the three models as distributed on the diagnosis category as well as the portion of patients ruled

Table 1 Summary of pooled data set

Study	Subjects			Diagnosis		Gender		Age	
	Original	Excluded	Included	CAD	Other	Female	Male	Mean	SD
AdoptCAD	255	63 (24.7)	192 (75.3)	53 (27.6)	139 (72.4)	83 (43.2)	109 (56.8)	61.1	10.8
Dan-NICAD	1675	285 (17.0)	1390 (83.0)	137 (9.9)	1253 (90.1)	716 (51.5)	674 (48.5)	57.0	8.7
BIO-CAC	661	193 (29.2)	468 (70.8)	1 (0.2)	467 (99.8)	251 (53.6)	217 (46.4)	60.0	5.0
VALIDATE	226	54 (23.9)	172 (76.1)	66 (38.4)	106 (61.6)	69 (40.1)	103 (59.9)	64.2	10.5
Total	2817	595 (21.1)	2222 (78.9)	257	1965	1119	1103	58.5	8.7

The data set consists of data pooled from four studies: AdoptCAD, Dan-NICAD, BIO-CAC, and VALIDATE. The number of original subjects is the number of subjects that were included in the respective study before any exclusions. Statistics for diagnosis, gender, and age are for patients included in the pooled data set and thus further analysis.

Table 2 Summary of patient demographics, risk factors, and symptoms

	All (n = 2222)	Training (n = 1776)	Test (n = 446)
Characteristics			
Male	1103 (49.6%)	882 (49.7%)	221 (49.6%)
Age, years (mean ± SD)	58.5 ± 8.74	58.5 ± 8.75	58.5 ± 8.68
<40	5 (0.225%)	5 (0.282%)	0 (0%)
40–50	349 (15.7%)	279 (15.7%)	70 (15.7%)
50–60	845 (38%)	658 (37%)	187 (41.9%)
60–70	818 (36.8%)	677 (38.1%)	141 (31.6%)
≥70	205 (9.23%)	157 (8.84%)	48 (10.8%)
Body mass index, kg/m ² (mean ± SD)	27 ± 4.30	27 ± 4.39	26.7 ± 3.93
Risk factors and symptoms			
Family history of CAD	687 (30.9%)	551 (31.0%)	136 (30.5%)
Smoking			
Never	1064 (47.9%)	835 (47.0%)	229 (51.3%)
Former	790 (35.6%)	642 (36.1%)	148 (33.2%)
Active	368 (16.6%)	299 (16.8%)	69 (15.5%)
Dyslipidaemia	1530 (68.9%)	1228 (69.1%)	302 (67.7%)
Hypertension	1308 (58.9%)	1045 (58.8%)	263 (59.0%)
Diabetes	142 (6.39%)	107 (6.02%)	35 (7.85%)
Cardiac symptoms at referral			
Typical chest pain	601 (27.0%)	470 (26.5%)	131 (29.4%)
Atypical chest pain	605 (27.2%)	490 (27.6%)	115 (25.8%)
Nonspecific chest pain	739 (33.3%)	592 (33.3%)	147 (33.0%)
Dyspnoea	277 (12.5%)	224 (12.6%)	53 (11.9%)
Prevalence of CAD			
CAD	257 (11.6%)	204 (11.5%)	53 (11.9%)

out based on a 5% cutoff in the training (a) and test (b) data sets. It is clear that the risk classification of the training and test data set are similar, though the proportion of CAD patients in the low-risk category is somewhat higher for the test data set.

The rule-out portions on the test data set of the RF-CL and A-CL models were significantly ($P < 0.05$) higher than the ESC2019 model, but the rule-out portion of the A-CL model was not significantly ($P = 0.0571$) higher than the RF-CL model. However, the same comparison on the training data set showed a significant ($P < 0.05$) higher rule-out for the A-CL model due to similar performance on a much larger data set.

Both the RF-CL and A-CL models showed substantial reclassification of patients to a lower likelihood category of CAD compared with the ESC2019 model, with NRI scores of 0.23 (RF-CL) and 0.33 (A-CL). See [Supplementary material online, Table S6](#) for further details on NRI evaluation.

Discussion

The 2019 ESC guidelines on chronic coronary syndromes⁵ recommend a PTP table based on age, sex, and symptoms. Furthermore, the guidelines suggest that risk factors of CAD can be used as

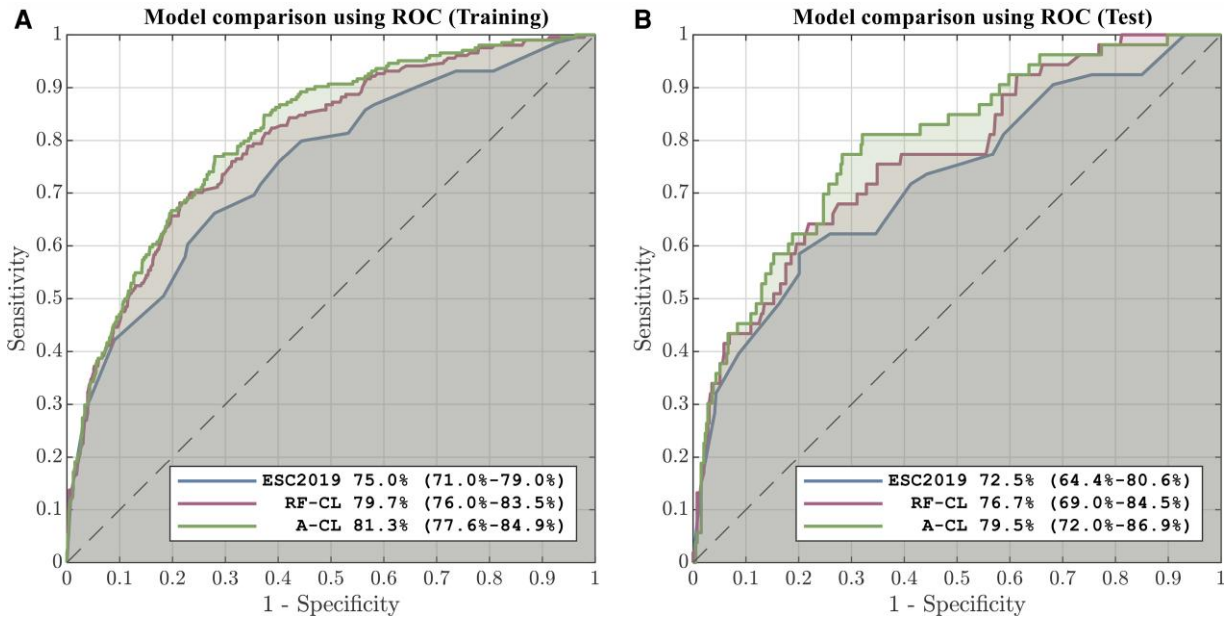


Figure 2 Receiver operating characteristic curve comparison of the ESC2019, risk factor-weighted clinical likelihood, and acoustic-weighted clinical likelihood models evaluated on the training (A) and test (B) data sets. The numbers in the legends are the area under the curve for each model and in parentheses the 95% confidence intervals.

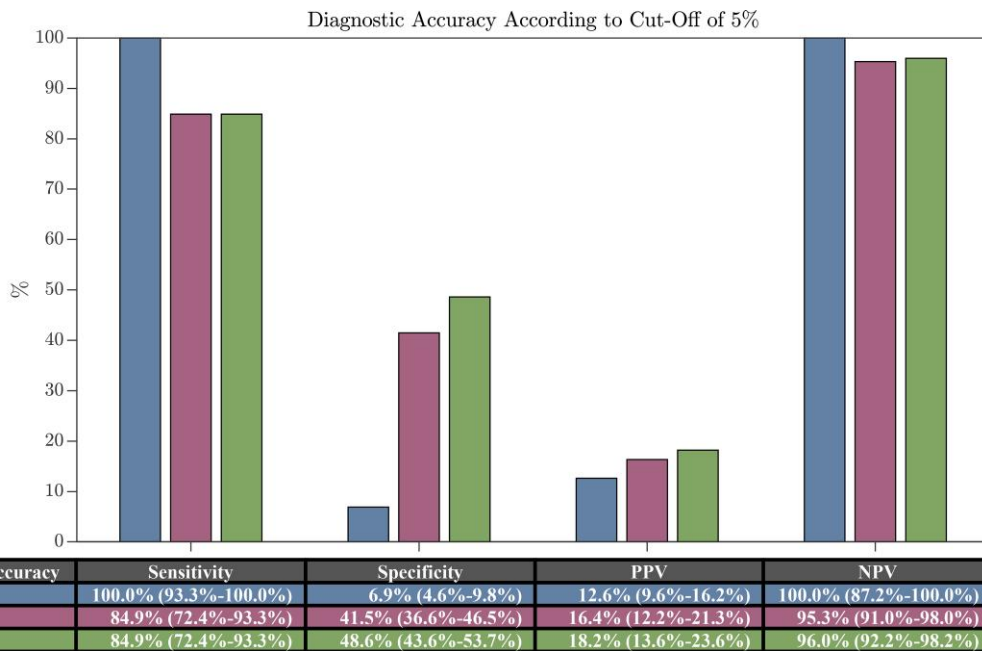


Figure 3 Diagnostic accuracy evaluated on the test data set for sensitivity, specificity, and positive and negative predictive values with a clinical likelihood cutoff of 5%.

modifiers to the PTP estimate without specifying how these risk factors should be weighed. This motivated Winther et al.⁹ to develop a model (RF-CL) that took these risk factors into account, resulting in significantly improved specificity. A second model (CACS-CL) was

suggested, which in addition to the risk factors included a CACS, resulting in even better model performance. However, acquiring CACS requires specialized and expensive equipment which is usually not available early in the diagnostic pathway. Conversely, the proposed

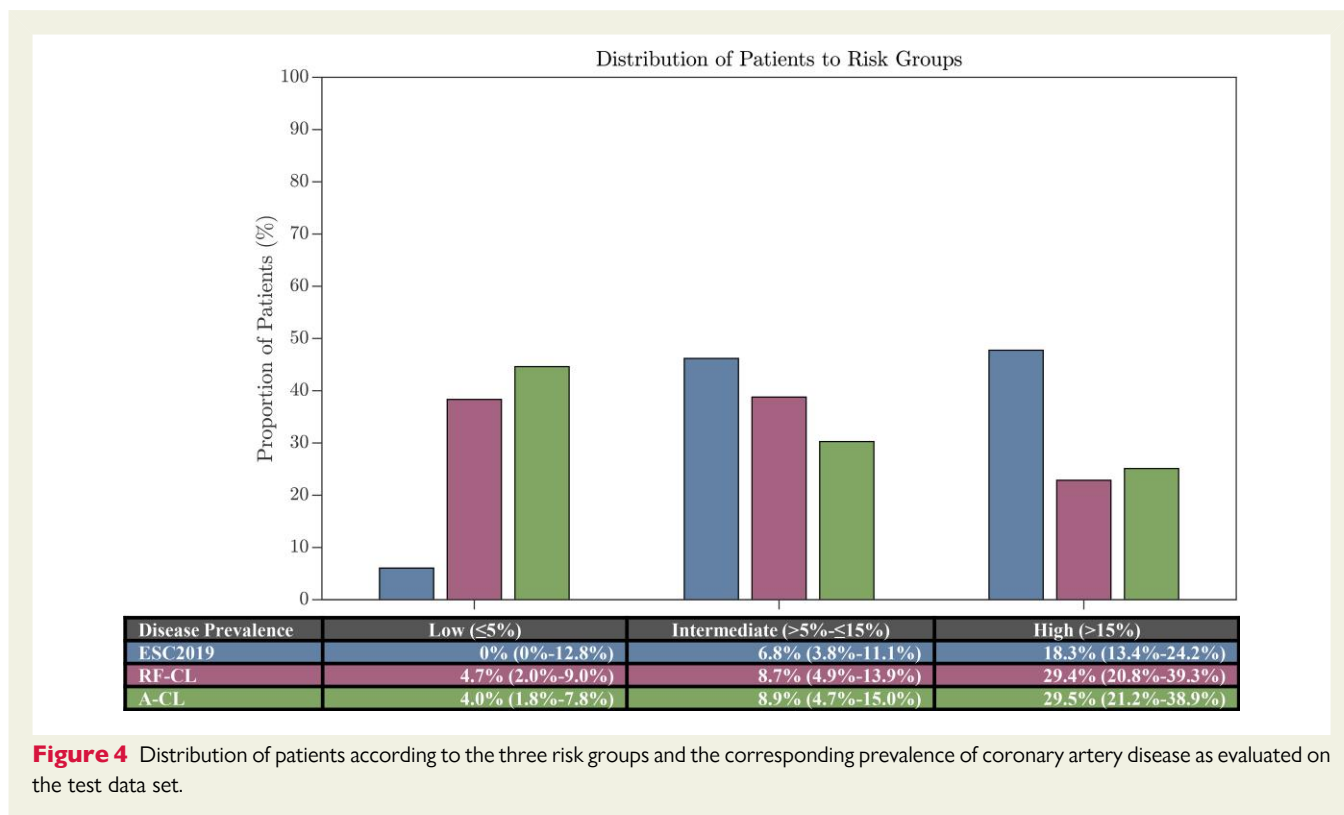


Figure 4 Distribution of patients according to the three risk groups and the corresponding prevalence of coronary artery disease as evaluated on the test data set.

Table 3 Risk classification and rule out

Risk Classification	CAD			Other			All Rule-out ($\leq 5\%$)
	Low	Intermediate	High	Low	Intermediate	High	
<i>Training</i>							
ESC2019	3 (1%)	38 (19%)	163 (80%)	116 (7%)	759 (48%)	697 (44%)	119 (7%)
RF-CL	16 (8%)	64 (31%)	124 (61%)	664 (42%)	628 (40%)	280 (18%)	680 (38%)
A-CL	19 (9%)	49 (24%)	136 (67%)	764 (49%)	494 (31%)	314 (20%)	783 (44%)
<i>Test</i>							
ESC2019	0 (0%)	14 (26%)	39 (74%)	27 (7%)	192 (49%)	174 (44%)	27 (6%)
RF-CL	8 (15%)	15 (28%)	30 (57%)	163 (41%)	158 (40%)	72 (18%)	171 (38%)
A-CL	8 (15%)	12 (23%)	33 (62%)	191 (49%)	123 (31%)	79 (20%)	199 (45%)

addition of acoustic features to the RF-CL model can be applied with portable low-cost equipment as a point-of-care device that only requires 10 min of testing.

We developed an improved clinical likelihood estimation model for CAD by adding acoustic features to the RF-CL model⁹ which resulted in a significantly higher specificity at 5% risk threshold. Sensitivity was the same for the original RF-CL model and the acoustic model. The acoustic features that improved the clinical risk factor model and were included in the developed model (A-CL) were diastolic sample entropy and two high-frequency features near the onset of the S1 heart sound. Though other acoustic features showed high discriminatory power either alone or in concert with other acoustic features, the selected features were those that improved

the RF-CL model the most based on AUC performance on the training data set.

Sample entropy was previously used in the Acarix CAD-score,⁴ and other studies have also used this measure of complexity for detection of cardiovascular occlusions using PCG. For example, Zhang *et al.*³⁶ investigated several different complexity measures for their capacity to distinguish patients with varying degrees of coronary artery stenosis. In the current study, sample entropy demonstrated discriminatory power beyond what is contained in the clinical risk factors of the RF-CL model. Turbulence in the coronary arteries is expected to increase complexity of the heart sound which will affect sample entropy as a measure of complexity.

The two high-frequency S1 features are comparable with the study in which Makaryus et al.²⁶ used the Cardiac Sonospectrographic Analyzer (SonoMedica) for analysis of the 400–2700 Hz frequency spectrum to detect the presence of micro-bruits. However, where the current study identifies a relatively narrow time-window using recordings from a single site on the chest surface, the study by Makaryus et al. utilized the entire heartbeat using heart sounds recorded sequentially from multiple sites.

The applied feature selection method with 5 × 5-fold cross-validation was successful in selecting features using the training data set which performed equally well in the test data set. Sensitivity was notably but not significantly lower in the test data set; however, this was the case for both the RF-CL and A-CL model, suggesting that this was likely due to patients with outlier clinical risk factors randomly being distributed to the test data set to a higher degree. With just 53 CAD patients in the test data set, fluctuations of a few patients can change the results considerably.

The added acoustic features moderately increased performance by around 2 percentage points AUC with most of the performance increase in the upper part of the ROC curve (SE > 7). Considering that the model is aimed towards ruling out CAD early in the diagnostic pathway, this upper part of the ROC curve is of particular interest in that it allows for ruling out with high sensitivity. This means a higher rule-out capacity with the same sensitivity.

Though the rule-out proportion for the A-CL model (45%) in the test data set was not significantly higher when compared with the RF-CL model (38%), there was a significant improvement of rule-out proportion in the full data set. Overall, the A-CL model demonstrated better performance than the RF-CL model, and both offer significantly better rule-out proportions than the ESC2019 model.

The improvement seen in the A-CL model (49% specificity) is not competitive with the CACS-CL model⁹ (59% specificity); however, the A-CL model offers a different use-case for early rule-out of patients suspected of CAD using a point-of-care device. The proposed addition of acoustic features to the clinical risk factor model can be implemented with a low-cost device before patients undergo more expensive and/or invasive testing.

Study limitations

Analyses in this study are retrospective and based on data pooled from four studies and thus the data set might not be representative of the clinical workflow. The data set included study data from asymptomatic subjects in the BIO-CAC study as well as patients referred for ICA in the VALIDATE and AdoptCAD studies, neither of which are typical for patients referred for non-invasive testing.

Furthermore, patients with arrhythmia were excluded from the studies that make up the data set of the current study.

Conclusion

The results of this article showed that the addition of acoustic features provides significant improvements to the high-sensitivity part of the ROC curve of an existing highly performing clinical likelihood model. Additionally, the developed A-CL model yielded substantial reclassification of patients to low likelihood ($\leq 5\%$) of CAD, thus allowing more patients to be ruled out at an earlier stage. This

demonstrates the efficacy of using heart sound analysis as an addition to clinical risk factors when risk stratifying patients suspected of having CAD.

Lead author biography



Bjarke Skogstad Larsen is an industrial PhD student at Aalborg University in a collaboration Acarix A/S and with a grant from the Innovation Fund Denmark. His thesis work encompasses heart sound analysis for the detection of coronary artery disease.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health* online.

Acknowledgements

The authors would like to thank Acarix A/S for providing access to the heart sound recordings used in this study.

Funding

This work was carried out as part of an Industrial PhD grant (5016–00179B) awarded to Bjarke Skogstad Larsen and funded by Innovation Fund Denmark.

Conflict of interest: B.S.L. is an industrial PhD student at Acarix A/S and is also a minor shareholder in the company. S.E.S. is a minor shareholder in Acarix A/S and is also employed at the company as a part-time consultant.

Data availability

The data used in this article are not publicly available because they are patient health data. Making the data publicly available would compromise patients' privacy. If other investigators are interested in performing additional analyses, requests can be made to the corresponding author and analyses will be performed in collaboration with Acarix.

References

1. World Health Organization. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000–2016. World Health Organization 2018.
2. Therning C, Galatius S, Heitmann M, Højberg S, Sørum C, Bech J, Husum D, Dominguez H, Sehestedt T, Hermann T, Reeh J, Simonsen L, Prescott E. Low diagnostic yield of non-invasive testing in patients with suspected coronary artery disease: results from a large unselected hospital-based sample. *Eur Hear J Qual Care Clin Outcomes* 2018;**4**:301–308.
3. Douglas PS, Hoffmann U, Patel MR, Mark DB, Al-Khalidi HR, Cavanaugh B, Cole J, Dolor RJ, Fordyce CB, Huang M, Khan MA, Kosinski AS, Krucoff MW, Malhotra V, Picard MH, Udelson JE, Velazquez EJ, Yow E, Cooper LS, Lee KL. Outcomes of anatomical versus functional testing for coronary artery disease. *N Engl J Med* 2015;**372**: 1291–1300.

4. Winther S, Nissen L, Schmidt SE, Westra JS, Rasmussen LD, Knudsen LL, Madsen LH, Johansen JK, Larsen BS, Struijk JJ, Frost L, Holm NR, Christiansen EH, Bøtker HE, Bøttcher M. Diagnostic performance of an acoustic-based system for coronary artery disease risk stratification. *Heart* 2018;**104**:928–935.
5. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, Prescott E, Storey RF, Deaton C, Cuisset T, Agewall S, Dickstein K, Edvardsen T, Escaned J, Gersh BJ, Svtil P, Gilard M, Hasdai D, Hatala R, Mahfoud F, Masip J, Muneretto C, Valgimigli M, Achenbach S, Bax JJ, ESC Scientific Document Group. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes: the task force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J* 2020;**41**:407–477.
6. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med* 1979;**300**:1350–1358.
7. Winther S, Schmidt SE, Rasmussen LD, Juárez Orozco LE, Steffensen FH, Bøtker HE, Knuuti J, Bøttcher M. Validation of the European Society of Cardiology pre-test probability model for obstructive coronary artery disease. *Eur Heart J* 2021;**42**:1401–1411.
8. Bing R, Singh T, Dweck MR, Mills NL, Williams MC, Adamson PD, Newby DE. Validation of European Society of Cardiology pre-test probabilities for obstructive coronary artery disease in suspected stable angina. *Eur Hear J Qual Care Clin Outcomes* 2020;**6**:293–300.
9. Winther S, Schmidt SE, Mayrhofer T, Bøtker HE, Hoffmann U, Douglas PS, Wijns W, Bax J, Nissen L, Lynggaard V, Christiansen JJ, Saraste A, Bøttcher M, Knuuti J. Incorporating coronary calcification into pre-test assessment of the likelihood of coronary artery disease. *J Am Coll Cardiol* 2020;**76**:2421–2432.
10. Dock W, Zonerach S. A diastolic murmur arising in a stenosed coronary artery. *Am J Med* 1967;**42**:617–619.
11. Semmlow J, Welkowitz W, Kostis J, Mackenzie JW. Coronary artery disease—correlates between diastolic auditory characteristics and coronary artery stenoses. *IEEE Trans Biomed Eng* 1983;**30**:136–139.
12. Gauthier D, Akay YM, Paden RG, Pavlicek W, Fortuin FD, Sweeney JK, et al. Spectral analysis of heart sounds associated with coronary occlusions. In: 2007 6th International Special Topic Conference on Information Technology Applications in Biomedicine, Tokyo, Japan. IEEE; 2007. p. 49–52
13. Schmidt SE, Toft E, Holst-Hansen C, Struijk JJ. Noise and the detection of coronary artery disease with an electronic stethoscope. In: 2010 5th Cairo International Biomedical Engineering Conference. IEEE; 2010. p. 53–56.
14. Schmidt SE, Holst-Hansen C, Hansen J, Toft E, Struijk JJ. Acoustic features for the identification of coronary artery disease. *IEEE Trans Biomed Eng* 2015;**62**:2611–2619.
15. Winther S, Schmidt SE, Holm NR, Toft E, Struijk JJ, Bøtker HE, Bøttcher M. Diagnosing coronary artery disease by sound analysis from coronary stenosis induced turbulent blood flow: diagnostic performance in patients with stable angina pectoris. *Int J Cardiovasc Imaging* 2016;**32**:235–245.
16. Schmidt SE, Hansen J, Zimmermann H, Hammershøi D, Toft E, Struijk JJ. Coronary artery disease and low frequency heart sound signatures. In: Murray A (ed.), *2011 Computing in cardiology*. Hangzhou: IEEE; 2011. p. 481–484.
17. Dragomir A, Post A, Akay YM, Jneid H, Paniagua D, Denktas A, Bozkurt B, Akay M. Acoustic detection of coronary occlusions before and after stent placement using an electronic stethoscope. *Entropy* 2016;**18**:281.
18. Akay YM, Akay M, Welkowitz W, Semmlow JL, Kostis JB. Noninvasive acoustical detection of coronary artery disease: a comparative study of signal processing methods. *IEEE Trans Biomed Eng* 1993;**40**:571–578.
19. Zhidong Z. Instantaneous frequency analysis of diastolic murmurs for coronary artery disease. In: 2005 International Conference on Neural Networks and Brain. IEEE; 2005. p. 1097–1100.
20. Pathak A, Mandana K, Saha G. Ensembled transfer learning and multiple kernel learning for phonocardiogram based atherosclerotic coronary artery disease detection. *IEEE J Biomed Health Inform* 2022;**26**:2804–2813.
21. Li H, Wang X, Liu C, Zeng Q, Zheng Y, Chu X, Yao L, Wang J, Karmakar C. A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection. *Comput Biol Med* 2020;**120**:103733.
22. Larsen BS, Winther S, Nissen L, Diederichsen A, Bøttcher M, Jan Struijk J, Christensen MG, Schmidt SE. Spectral analysis of heart sounds associated with coronary artery disease. *Physiol Meas* 2021;**42**:105013.
23. Li H, Zhang G, Shao G, Wang A, Gu Y, Tian Z, Zhang Q, Shi P. Improvement of the accuracy in the identification of coronary artery disease combining heart sound features. *Biomed Res Int* 2022;**2022**:1–16.
24. Schmidt SE, Winther S, Larsen BS, Groenhoef MH, Nissen L, Westra J, Frost L, Holm NR, Micklely H, Steffensen FH, Lambrechtsen J, Nørskov MS, Struijk JJ, Diederichsen ACP, Bøttcher M. Coronary artery disease risk reclassification by a new acoustic-based score. *Int J Cardiovasc Imaging* 2019;**35**:2019–2028.
25. Winther S, Nissen L, Schmidt SE, Westra J, Andersen IT, Nyegaard M, Madsen LH, Knudsen LL, Urbonaviciene G, Larsen BS, Struijk JJ, Frost L, Holm NR, Christiansen EH, Bøtker HE, Bøttcher M. Advanced heart sound analysis as a new prognostic marker in stable coronary artery disease. *Eur Hear J Digit Heal* 2021;**2**:279–289.
26. Makaryus AN, Makaryus JN, Figgatt A, Mulholland D, Kushner H, Semmlow JL, Mieres J, Taylor AJ. Utility of an advanced digital electronic stethoscope in the diagnosis of coronary artery disease compared with coronary computed tomographic angiography. *Am J Cardiol* 2013;**111**:786–792.
27. Azimpour F, Caldwell E, Tawfik P, Duval S, Wilson RF. Audible coronary artery stenosis. *Am J Med* 2016;**129**:515–521.e3.
28. Nissen L, Winther S, Isaksen C, Ejlersen JA, Brix L, Urbonaviciene G, et al. Danish study of non-invasive testing in coronary artery disease (Dan-NICAD): study protocol for a randomised controlled trial. *Trials* 2016;**17**:262.
29. Grønhoj MH, Gerke O, Micklely H, Steffensen FH, Lambrechtsen J, Sand NPR, Rasmussen LM, Olsen MH, Hallas J, Diederichsen ACP. External validity of a cardiovascular screening including a coronary artery calcium examination in middle-aged individuals from the general population. *Eur J Prev Cardiol* 2018;**25**:1156–1166.
30. Diederichsen SZ, Grønhoj MH, Micklely H, Gerke O, Steffensen FH, Lambrechtsen J, Sand NPR, Rasmussen LM, Olsen MH, Diederichsen ACP. CT-detected growth of coronary artery calcification in asymptomatic middle-aged subjects and association with 15 biomarkers. *JACC Cardiovasc Imaging* 2017;**10**:858–866.
31. Renker M, Kriechbaum SD, Schmidt SE, Larsen BS, Wolter JS, Dörr O, Fischer-Rasokat U, Kim W, Liebetrau C, Bøttcher M, Nef H, Bauer T, Hamm CW. Prospective validation of an acoustic-based system for the detection of obstructive coronary artery disease in a high-prevalence population. *Heart Vessels* 2021;**36**:1132–1140.
32. Larsen BS. Stratified Block Randomization. 2021. (<https://se.mathworks.com/matlab-central/fileexchange/96058-stratified-block-randomization>).
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;**44**:837.
34. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma* 2011;**12**:1–8.
35. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;**27**:157–172.
36. Zhang H, Wang X, Liu C, Li Y, Liu Y, Jiao Y, Liu T, Dong H, Wang J. Discrimination of patients with varying degrees of coronary artery stenosis by ECG and PCG signals based on entropy. *Entropy* 2021;**23**:823.