OPEN ACCESS
Indo-JC
Indonesian Journal on Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 5, Issue. 3, December 2020. pp. 63-78
doi:10.34818/indojc.2021.5.3.536

Study of Machine Learning Algorithm on Phonocardiogram Signals for Detecting of Coronary Artery Disease

Satria Mandala [*1], Miftah Pramudyo[2], Ardian Rizal[3], Maurice Fikry[1]

[1]*Human Centric Engineering, Telkom University & Fakutltas Informatika, Universitas Telkom*
Jl. Telekomunikasi no. 1 Terusan Buah Batu Dayeuh Kolot Kabupaten Bandung Indonesia

[2]*Fakultas Kedokteran, Universitas Padjajaran*
Jl. Raya Bandung Sumedang KM.21, Hegarmanah, Kec. Jatinangor, Kabupaten Sumedang, Jawa Barat 45363

[3]*Fakultas Kedokteran, Universitas Brawijaya*
Jl. Veteran, Lowokwaru, Malang 65145, Jawa Timur, Indonesia

[*]satriamandala@telkomuniversity.ac.id, miftah.pramudyo@unpad.ac.id, drardianrizal@ ub.ac.id, mauricefikry@gmail.com

Abstract

Several methods of detecting coronary artery disease (CAD) have been developed, but they are expensive and generally use an invasive catheterization method. This research provides a solution to this problem by developing an inexpensive and non-invasive digital stethoscope for detecting CAD. To prove the effectiveness of this device, twenty-one subjects consisting of 11 CAD patients and 10 healthy people from Hasan Sadikin Hospital Bandung were selected as validation test participants. In addition, auscultation was carried out at four different locations around their chests, such as the aorta, pulmonary, tricuspid, and mitral. Then the phonocardiogram data taken from the stethoscope were analyzed using machine learning. To obtain optimal detection accuracy, several types of kernels such as radial basis function kernel (RBF), polynomial kernel and linear kernel of Support Vector Machine (SVM) have been analyzed. The experimental results show that the linear kernel outperforms compared to others; it provides a detection accuracy around 66%. Followed by RBF is 56% and Polynomial is 46%. In addition, the observation of phonocardiogram signals around the aorta is highly correlated with CAD, giving an average detection accuracy for the kernel of 66%; followed by 44% tricuspid and 43% pulmonary.

Keywords: Coronary Artery Disease (CAD), Heart Diseases, Machine Learning, Stethoscope Digital

Abstrak

Beberapa metode untuk mendeteksi penyakit arteri koroner (CAD) telah dikembangkan, tetapi harganya mahal dan umumnya menggunakan metode kateterisasi invasif. Penelitian ini memberikan solusi untuk masalah ini dengan mengembangkan stetoskop digital yang murah dan non-invasif untuk mendeteksi CAD. Untuk membuktikan keefektifan alat ini, dipilih dua puluh satu subjek yang terdiri dari 11 pasien CAD dan 10 orang sehat dari RS Hasan Sadikin Bandung sebagai peserta uji validasi. Selain itu, auskultasi dilakukan di empat lokasi berbeda di sekitar dada mereka, yaitu aorta, pulmonal, trikuspid, dan mitral. Kemudian data fonokardiogram yang diambil dari stetoskop dianalisis menggunakan pembelajaran mesin. Untuk mendapatkan akurasi deteksi yang optimal,

Received on January, 2020. Accepted on December, 2020

beberapa jenis kernel seperti radial basis function kernel (RBF), polynomial kernel dan linear kernel of Support Vector Machine (SVM) telah dianalisis. Hasil percobaan menunjukkan bahwa kernel linier berkinerja lebih baik dibandingkan dengan yang lain; ini memberikan akurasi deteksi sekitar 66%. Disusul RBF 56% dan Polinomial 46%. Selain itu, pengamatan sinyal fonokardiogram di sekitar aorta sangat berkorelasi dengan CAD, memberikan akurasi deteksi rata-rata untuk kernel 66%; Kemudian diikuti trikuspid 44% dan paru 43%.

Kata Kunci: *Coronary Artery Disease,* Jantung, *Machine Learning*, Stetoskop

I. INTRODUCTION

There are many types of diseases that are experienced by various groups of people, but it is difficult to identify the symptoms of the diseases. One type of the diseases is coronary heart disease or known also as coronary artery disease (CAD). CAD is an abnormality in the heart due to accumulation of plaque on the interior walls of the arteries. This plaque blocks blood flow to reach the heart muscles which may lead to a condition called as ischemia. In the worst-case ischemia leads to heart muscle cell death and heart attack [1, 2]. CAD has caused the death of thousands of people in Europe, with a mortality rate of 20% [3]. To overcome this situation, 59% of patients performed an examination of the body's condition using coronary angiography technology. Unfortunately, this method requires an expensive cost. The results of the angiograph evaluation is also sometimes less accurate in determining CAD. On the other hand, 50% of patients who do not undergo the examination for CAD have a sudden death condition [3].

Data from the Ministry of Health of the Republic of Indonesia [4], in 2013, states that the prevalence of CAD in Indonesia is around 0.5% (883,000) of the total population. This number continues to increase from year to year; and the prevalence has reached 1.5% (3,975 million) in 2018 [5]. Based on the doctor's diagnosis [4], the number of coronary heart sufferers in 2013 was much greater than the prevalence, i.e., 1.5% or an estimated 2,650,340 people. It is estimated that the highest number of coronary heart disease sufferers is in West Java Province, as many as 160,812 people (0.5%). Meanwhile, North Maluku Province had the lowest number of sufferers, around 1,436 people (0.2%).

The development of information technology in the current digital era should be able to minimize CAD. This can be achieved by conducting early screening at a lower cost. Addressing these challenges, several researchers have developed inexpensive, noninvasive early detection of CAD. Pouladian, et al. [6] developed the Arterio-Oscillo-Graph technique to detect CAD and evaluated it in 26 patients out of 29 test participants. Another researcher, K. Lei [7] used a machine learning to diagnose potential CAD based on patient medical records. Weighted Naïve Bayes was the proposed method chosen to detect CAD and produced a detection accuracy around 79%. This less-than-optimal accuracy result may have a negative impact because accurate diagnosis and analysis is needed in the medical world. Furthermore, collecting medical record data on patients takes a very long time. If the patient suffers from many diseases, selecting medical record data that can be used to detect CAD becomes difficult. In addition, manual recording of patient medical records in several doctors' clinics adds to the complexity of the tracing.

This research provides a solution to all the above problems, by developing an inexpensive machine learning-based CAD detector to shorten the time of diagnosis. The developed CAD detector is an enhanced version of a digital stethoscope with CAD detection capabilities. The machine learning method chosen for the CAD detector is the best algorithm from the Support Vector Machine (SVM), which is acquired by experimenting with several types of kernel on SVM, such as radial basis function kernel (RBF), polynomial kernel and linear kernel. Henceforth, this article also contains Related Works, Research Methodology, Experiments Results, Discussion, and Conclusion.

II. Related works

A. Acoustic Stethoscope Development

A stethoscope was first invented in 1818 by a French doctor named René Laennec [8]. The stethoscope, which was originally developed from a simple wooden tube, has now developed into a sophisticated case and known as an acoustic stethoscope. The device commonly used for detecting heart healthy through a heart sound or phonocardiogram (PCG) signal. A normal heartbeat sound is at a frequency of 10-200 Hz. The normal heart sounds can be classified into 4 type of sounds, namely: S1 (first heart sound), S2 (second heart sound), S3 (third heart sound) and S4 (fourth heart sound). The S1 sound is produced by closing the atrioventricular (AV) valve, the S2 sound is produced due to the closure of the aortic valve and pulmonic valve. The S3 sound occurs at the beginning of the diastole and the S4 sound at the end of the diastole [8]. The results of auscultation using an acoustic stethoscope may have several drawbacks [9, 10], such as: 1. the diagnosis results depend on the sensitivity of the hearing and the doctor's experience 2. the heart sound signal heard by doctor cannot be stored as a medical history of patient. Since a decade ago, various attempts have been performed to develop a digital stethoscope to overcome the above problems; and the rest of this section will explain these efforts.

Wang, et al. [11] developed a digital stethoscope to remotely analyze patient's heart sounds (home monitoring). The stethoscope works offline to record phonocardiogram signal data from the stethoscope to Universal Serial Bus (USB) device. The data is then analyzed on a computer using a single degree of freedom (SDOF) analytical model to obtain a cardiac sound characteristic waveform (CSCW). Furthermore, four characteristic parameters, known as T1, T2, T11 and T12, associated with the first heart sound (S1) and the second heart sound (S2) are determined by measuring the time interval between the points traversed in the CSCW and a threshold value (THV) to differentiate normal and abnormal heart sounds. Evaluation on 40 normal and 20 abnormal heart sounds from 10 healthy volunteers and 5 patients with heart problems were performed. The analysis results show that the proposed system can distinguish between normal and abnormal heart sounds with an accuracy of 92.5%.

Research on digital stethoscopes was also carried out by Shin, et al. [12] which increased the stethoscope's ability to detect murmurs. A murmur is a condition when the blood flows through the heart or blood vessels around the heart due to an abnormality caused by turbulence.

Research on digital stethoscopes by [13-15] focuses on utilizing of wireless technology to transmit stethoscope data. The wireless stethoscope can be used for monitoring patient’s conditions without limiting their mobility. Heart sounds recorded using a wireless digital stethoscope are then compared with the recorded heart sounds from a wired digital stethoscope. The result of experiments show that performance of the proposed wireless stethoscope is similar to the wired stethoscope [14].

B. Stethoscope for Detecting CAD

Schmidt, et al. [16] added a digital signal processing unit to an electronic stethoscope in order to be used for diagnosing coronary artery disease (CAD). Nine types of features from five overlapping frequency bands were obtained and analyzed using 435 records from 133 subjects to identify CAD. Using the quadratic discriminant function, several features are combined into a CAD score. The area under the operating characteristic of the recipient for CAD score was 0.73 (95% CI: 0.69-0.78). From these results it is concluded that there is a potential that heart sounds can be utilized for diagnosing CAD. However, further improvements are needed for clinical relevance. Research by [16] has triggered several studies of CAD detection based on phonocardiogram signals using machine learning, such as in [1, 2, 7, 17-22].

Dragomir, et al. [2] investigated acoustic features of CAD in coronary patient subjects who had a single coronary occlusion before and after stent placement. Acoustic characteristics are achieved by calculating fast Fourier transform (FFT) ratio of the acoustic signal, which is the total power above 150 Hz divided by the total

power below 150 Hz. In addition, an estimate of the entropy value is also calculated to assess the differences caused by the stent placement procedure. Using the power ratio and Classifier Linear Discriminant Analysis (LDA) features, the sensitivity and specificity values achieved were 82% and 64%. Meanwhile, using entropy estimates, the sensitivity and specificity were 82% and 55%. These results suggest that the effect of stent placement on coronary occlusion can be monitored using an electronic stethoscope.

Maknickas and Maknickas [17] used Deep Convolutional Neural Networks and mel-frequency spectrum coefficients to recognize normal and abnormal phonocardiogram signals from the human heart. The data used in the experiment is cardiac sound data from PhysioNet.org. The proposed approach yields an accuracy of 84.15%. The sensitivity and specificity were 80.63% and 87.66%.

Samanta, et al. [18] proposed a multi-channel PCG system to differentiate CAD and normal subjects. This system does not require additional reference signals, such as an electrocardiogram (ECG) signal. The PCG sensors are attached to four different places on the chest. Then the 5 features in the time and frequency domain are extracted from the PCG signal. Binary classification based on Artificial Neural Network (ANN) is applied to the features acquired. The classification results are evaluated on each channel and show that the proposed sub-band-based spectral features perform well on clean and noise data. The accuracy achieved was 82.57% with a combination of signals from the tricuspid, mitral, and midaxillary regions. This multi-channel PCG system provides an accuracy increase of more than 4% compared to the best performance acquired by a single channel system.

Sridhar, et al. [20] attempted to observe CAD patterns using Discrete Wavelet Transform (DWT) and machine learning. The accuracy achieved by the study is above 95%. However, the data used to design a machine learning model were only 7 CAD patients and 40 normal subjects. This makes the study lacks in the objectivity of the diagnosis of CAD itself, because the data used are imbalanced. The effect of the imbalance data causes the CAD detector to be inaccurate (bias) in predicting the minority class (CAD patients) compared to the majority class (normal subjects) [23].

AlHosani, et al. [21] focused on developing machine learning for CAD detection in larger samples than Sridhar, et al. [20]. Twenty-nine CAD subjects were invited to participate in the study; and an electrocardiogram (EKG) is used as a tool to detect CAD. The accuracy achieved in the study was 93%. Cüvitoğlu and Işik [22] also focused on CAD detection using machine learning and yielded an accuracy of more than 80%. In that study, Cüvitoğlu and Işik [22] used 216 sample data.

Lei, et al. [7] used 496 CAD data from a hospital in China in 2016 to conduct a CAD diagnosis using Naïve Bayes. Although using a large number of datasets and data taken directly at the hospital, this study only resulted in an accuracy of 79%. This is not optimal because good accuracy is needed in the treatment of coronary artery disease diagnosis.

III. RESEARCH METHODOLOGY

A. Data

This study used data grabbed from 10 doctors participating in the Program Pendidikan Dokter Spesialis (PPDS), Faculty of Medicine, Padjajaran University Bandung as normal data from healthy subjects. Meanwhile, 11 CAD patients at Hasan Sadikin Hospital Bandung (in the Cardiac Center division) were marked as CAD data. To ensure the validity of the data generated from test participants (both healthy and CAD subjects), all subjects are required to do a medical check-up. Furthermore, data from participants were picked up using a stethoscope that had been developed based on a predetermined research protocol. These data are numerical phonocardiogram data taken 4 times for each subject at different places around chest, i.e., in the aorta, pulmonary, tricuspid, and mitral valve areas for 8 seconds.

.

B. Phonocardiogram Signal Processing Methods for CAD Detection

Figure 1 illustrates the method used for processing the phonocardiogram signals using machine learning to detect CAD. As shown in Figure 1, there are three steps for detecting CAD, i.e., denoising the PCG signal, performing the extraction process, and finally applying classification.

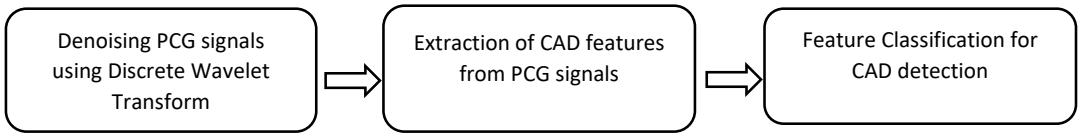

Figure 1. *Phonocardiogram* data processing to detect CAD

Denoising of the PCG signal in this study was carried out using DWT. In more detail, the denoising method used is adaptive thresholding as described in [24]. Using the method, both threshold value and wavelet coefficient at the selected decomposition level are determined adaptively. According to Jain and Tiwari [24] there are three steps that need to be performed for disseminating the PCG signal as follows: decomposing the signal, determining the threshold value and modifying the wavelet coefficient based on the selected threshold value, and finally reconstructing the signal.

DWT is also used in this research for CAD feature extraction process by decreasing the wavelet coefficient of the phonocardiogram signal. There are 5 statistical values picked up from the resulting wavelet coefficients, i.e., mean, median, standard deviation, skewness, and kurtosis. These five statistical values were chosen with the consideration that they could represent the change in high-frequency energy in the diastolic murmur as stated by Zhao [25].

This study uses a support vector machine (SVM) as a method to classify CAD and normal phonocardiogram signals. In many implementations, it has been proven that SVM gives stable prediction results and effectively used on small data samples compared to Artificial Neural Network (ANN) [26]. SVM is more stable than ANN because the prediction given by SVM is based on global optimal, while ANN is based on local optimal. On the other hand, SVM classifier focuses on the training case which is located at the corner of the class descriptors to get the optimal hyperplane. This hyperplane is also effective for training with small samples [27]. Hyperplanes can be represented in the feature space using kernel functions (dot products between mapped pairs of input points x_i) [28]. In this research, several kernels of SVM, such as: radial basis function kernel (RBF), polynomial kernel, and linear kernel, are studied to get the best predictions of CAD detection. In addition, for the distribution of training and testing data, the 5-fold cross validation principle is used, and the average of each fold literacy will be used as the final value of the measured performance.

C. Metrics

This study measures the detection performance of CAD using a metric of accuracy, sensitivity, and specificity. Accuracy is used to determine the closeness of the predicted value from the system to the actual value. Sensitivity is used to determine how successful the system is to detect CAD in the CAD subject population. Specificity is used to determine the accuracy of normal detection in healthy subject populations [8]. The formula for calculating the CAD detection performance used in this study can be found in equations 1 to 3. Meanwhile, confusion metrics for predicting the accuracy of CAD detection are presented in Table 1.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Specificity = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

TABLE I
Confusion Matrix

	Positive (Predicted Class)	*Negative (Predicted Class)*
Positive (Actual Class)	TP (*True Positive*)	FN (*False Negative*)
Negative (Actual Class)	FP (*False Positive*)	TN (*True Negative*)

TP is true positive, which means the amount of CAD subject data that was successfully identified. TN is true negative, which means the number of data on healthy people has been detected correctly. Then, FP is false positive which means the amount of CAD subject data is misidentified. The FN is false negative, which means the number of data on healthy people is incorrectly recognized.

IV. Experiment Results

A. Dataset Visualization

The following is an example of visualization of phonocardiogram data from CAD subjects from four different places around chest: aorta, pulmonary, tricuspid, and mitral (as identified in section III, Research Methodology).

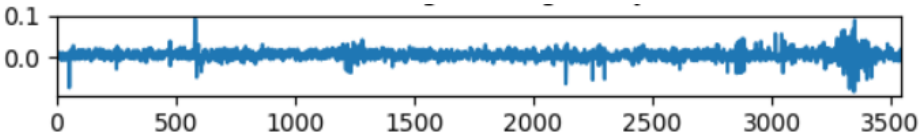

Figure 2. *PCG signal taken in Aorta*

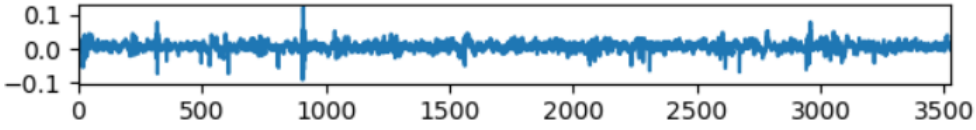

Figure 3. *PCG signal taken in Pulmonary*

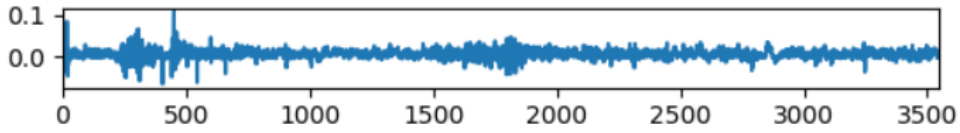

Figure 4. *PCG signal taken in Tricuspid*

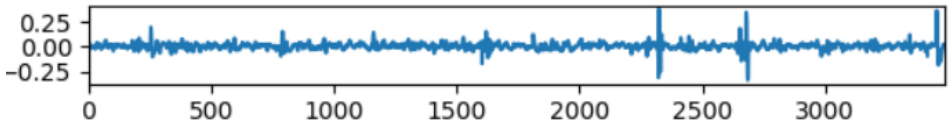

Figure 5. *PCG signal taken in Mitral*

As shown in Figures 2 to 5, the mitral phonocardiogram signal tends to have the lowest sound intensity when compared to the aorta, pulmonary, and tricuspid levels. However, it is difficult to determine signal abnormalities using a visual method. For this reason, these signals will be processed using a support vector machine (SVM).

B. Feature Extraction Results

TABLE II shows the results of feature extraction in the aorta using the DWT method as described in the Research Methodology section. While the feature statistics for all data in TABLE II are as follows, minimum value: -1.77, maximum: 1.27, standard deviation: 0.57, median: 0, skewness: -0.3 and kurtosis is 1.36. The meaning of "C" label in TABLE II is phonocardiogram data from CAD subjects, while "N" label is from healthy subjects. For the results of feature extraction on the mitral, pulmonary, and tricuspid can be seen in the Appendix.

TABLE II
PHONOCARDIOGRAM FEATURE EXTRACTION ON AORTA

Subjects	mean	stdev	median	skewness	kurtosis	Label
1	0.04	0.06	0.01	1.20	-0.24	C
2	-0.07	0.09	-0.03	-1.32	0.01	N
3	-0.01	0.02	0.00	-0.18	-1.22	N
4	-0.04	0.05	-0.01	-0.39	-1.77	C
5	0.21	0.28	0.06	1.12	-0.35	C
6	-0.01	0.02	0.00	-0.86	-0.70	N
7	-0.09	0.12	-0.03	-1.31	-0.04	N
8	0.05	0.05	0.02	1.27	-0.08	C
9	-0.08	0.11	-0.04	-1.22	-0.15	N
10	-0.26	0.27	-0.11	-0.87	-0.79	C
11	0.02	0.03	0.01	1.06	-0.22	N
12	-0.20	0.28	-0.04	-1.23	-0.17	C
13	0.28	0.36	0.10	1.22	-0.17	C
14	-0.05	0.08	0.00	-0.99	-0.62	N
15	0.28	0.35	0.09	1.08	-0.43	C
16	-0.11	0.12	-0.04	-1.02	-0.53	N
17	-0.08	0.08	-0.03	-0.58	-1.41	C
18	0.02	0.03	0.00	1.02	-0.38	N
19	0.08	0.09	0.05	1.14	-0.26	C
20	0.08	0.32	0.12	-0.31	-1.38	N
21	0.07	0.07	0.06	0.59	-0.98	C

C. SVM Classification Results

The experimental results of CAD detection performance using a support vector machine with RBF, polynomial and linear kernels can be seen in TABLE III to TABLE IV and are also presented in Figure 2 to Figure 4. The tables contain also CAD detection in the area around chest such as: aorta, tricuspid, pulmonary and mitral.

TABLE III
ACCURACY OF CAD DETECTION

Place of phonocardiogram data acquisition	SVM-Linear	SVM-Polynomial	SVM-RBF
Aorta	66%	46%	56%
Tricuspid	44%	43%	44%
Pulmonary	43%	43%	43%
Mitral	34%	38%	34%
Average	**47%**	**43%**	**44%**

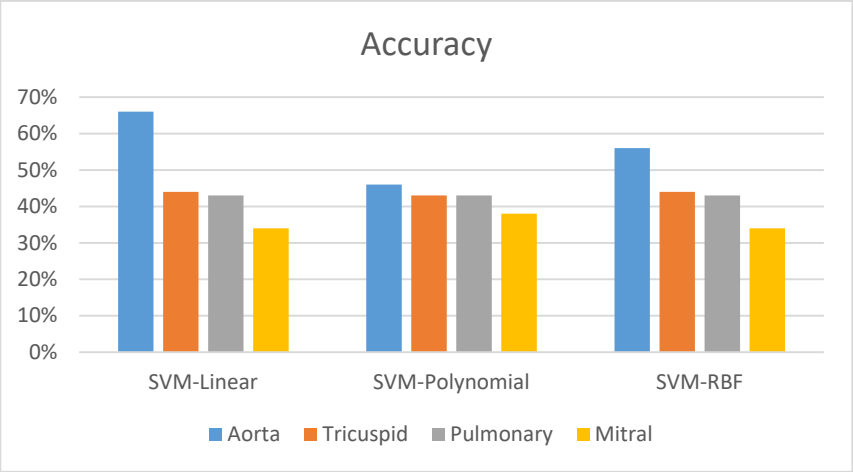

Figure 2. Accuracy of CAD Detection

TABLE III and Figure 2 show that the highest accuracy (around 66%) is obtained when the classifier uses a linear kernel, and the data is grabbed from aorta. On the other hand, the lowest accuracy of 34% is acquired when the kernel of the classifier is RBF and the data is mitral. The same situation (lowest accuracy value) is also achieved when the classifier uses a linear kernel, and the data is retrieved from mitral. From TABLE III and Figure 2, it can also be observed that the accuracy trend based on the three kernels has decreased simultaneously when data collection is in aorta, tricuspid, pulmonary and mitral regions respectively. When using linear, polynomial and RBF kernels, the lowest and highest accuracy ranges were 32%, 8% and 22%, respectively. The average accuracy of the three kernels from the highest to the lowest was 47% for linear kernel, 44% for RBF kernel and 43% for polynomial kernel.

TABLE IV and Figure 3 are the specificity of CAD detection using the SVM classifier. The maximum specificity value is 46.4%, and it is obtained when the SVM classifier uses a linear kernel and the data is from the aorta. Conversely, the lowest value is generated by the Polynomial kernel when data is in mitral area. The data distribution pattern on the specificity results is similar to the accuracy. Data decreased in the three kernels when data collection was carried out on aorta, tricuspid, pulmonary and mitral, respectively. The highest value ranges for specificity of linear, polynomial and RBF kernels were 26.4%, 12% and 31.8%, respectively. The average of specificity based on the three kernels from the highest to the lowest was 20% linear kernel, 15% RBF kernel, and 13% polynomial kernel.

TABLE IV
SPECIFICITY OF CAD DETECTION

Place of phonocardiogram data acquisition	SVM-Linear	SVM-Polynomial	SVM-RBF
Aorta	46.40%	25%	46.80%
Tricuspid	20%	15%	16.80%
Pulmonary	20%	15%	15%
Mitral	20%	13%	15%
Average	**20%**	**13%**	**15%**

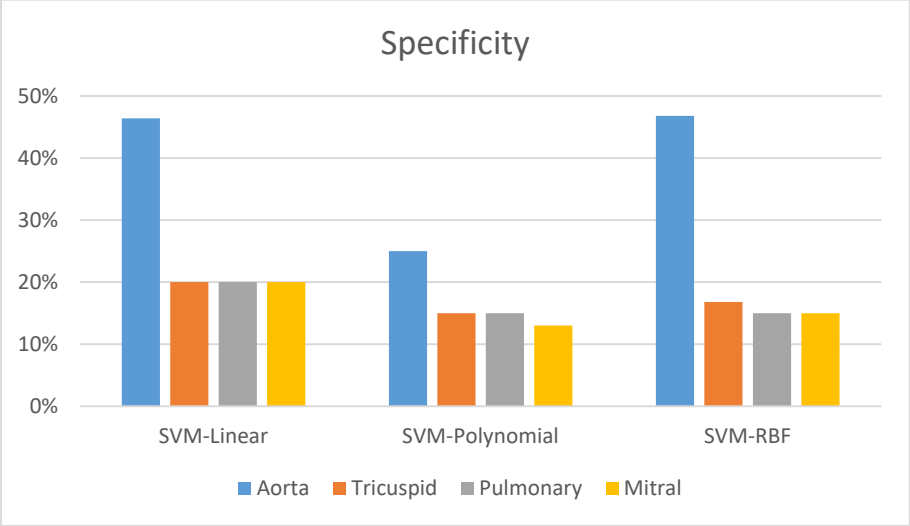

Figure 3. Specificity of CAD Detection

TABLE V and Figure 4 are sensitivity of CAD detection using the SVM classifier. The maximum specificity value is 80%, which is achieved when the SVM classifier uses a linear kernel and the data is in aorta. In contrast, the lowest sensitivity value, at 25%, was generated by the other three kernels: linear, polynomial and RBF when data were in mitral area. The pattern of data distribution still follows the pattern of accuracy and specificity results. Data decreased in the three kernels when data collection was carried out on aorta, tricuspid, pulmonary and mitral, respectively. The highest value ranges for detection specificity based on linear, polynomial and RBF kernels were 55%, 15% and 42%, respectively. The average sensitivity of the three kernels from the highest to the lowest was 48% linear kernels, 46% RBF kernels and 35% polynomial kernels.

TABLE V
SENSITIVITY OF CAD DETECTION

Place of phonocardiogram data acquisition	SVM-Linear	SVM-Polynomial	SVM-RBF
Aorta	80%	40%	67%
Tricuspid	48.60%	38%	55%
Pulmonary	38%	38%	38%
Mitral	25%	25%	25%
Average	**48%**	**35%**	**46%**

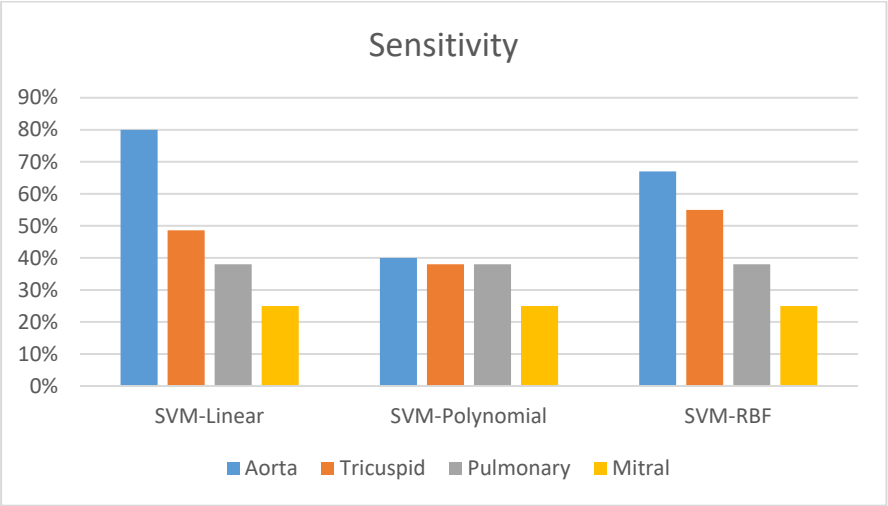

Figure 4. Sensitivity of CAD Detection

V. DISCUSSIONS

Although the results of CAD detection based on phonocardiogram signals still produce unsatisfactory accuracy, sensitivity, and specificity; however, there are several important notes that can be analyzed from the results as follows:

1. SVM uses kernel functions to systematically find support vector classifiers in a higher dimension. Several kernel functions, such as linear, polynomial and RBF, were studied in this study to achieve optimal detection results. The experimental results show that the linear kernel performs better on detecting CAD followed by RBF and polynomial on the three metrics used, i.e., accuracy, sensitivity and specificity. More details on the results can be seen in section IV C "SVM Classification Results". This fact shows that CAD features extracted from phonocardiogram signals are more suitable to be identified using linear kernels than both RBF kernels and polynomial kernels. Amiri, et al. [29] also had similar results when experimenting with the development of PCG data-based newborn cardiac monitoring.
2. The detection results of CAD in the area around aorta show that this area produces the highest detection performance compared to other areas such as tricuspid, pulmonary and mitral. This is because aorta is closest to the coronary arteries [30]. The location closer to the coronary arteries directly affects the sound of the phonocardiogram so that abnormalities on the phonocardiogram are easier to be identified.

VI. CONCLUSION

All research objectives to study of machine learning algorithm on phonocardiogram signals for detecting coronary artery disease have been achieved. It can be seen from section 4, Experimental Results, that the linear kernel outperforms compared to other kernels, such as RBF and polynomial kernel in supporting SVM to identify CAD. The highest detection accuracy of the linear kernel is 66%, followed by RBF 56% and polynomial 46%. For sensitivity, the linear kernels yield 80%, RBF kernels 67% and polynomial kernels 40%. Then for specificity, linear kernels yielded 46.4%, RBF kernels 46.8% and polynomial kernels 25%. On the other hand, this study also found that the selection of the CAD data collection location also had an effect on detection performance. The experimental results show that aorta is location that produces the highest detection performance, followed by tricuspid, pulmonary and mitral. This is due to the location of aorta closest to the coronary arteries, so the resulting phonocardiogram sound is easier to detect than other areas. The highest

detection accuracy in aorta was 66%, while the highest sensitivity at the same area was 80% and finally the highest specificity in aorta was 46.4%. In general, it can be concluded that the stethoscope phonocardiogram signal can be used to detect CAD, although studies using larger data should be carried out. Furthermore, other classifiers should also be explored for better CAD detection performance.

Acknowledgment

The authors deeply thank and acknowledge Kemenristek/BRIN (Ministry of Research and Technology, Republic of Indonesia) in supporting this research through INSINAS 2020 program, under contract no. 111 /INS-l/PPK/E4/2020.

REFERENCES

[1] J. L. Thomas, S. Winther, R. F. Wilson, and M. Bøttcher, "A novel approach to diagnosing coronary artery disease: acoustic detection of coronary turbulence," *The International Journal of Cardiovascular Imaging,* journal article vol. 33, no. 1, pp. 129-136, January 01 2017.

[2] A. Dragomir *et al.*, "Acoustic Detection of Coronary Occlusions before and after Stent Placement Using an Electronic Stethoscope," *Entropy,* vol. 18, no. 8, p. 281, 2016.

[3] S. Schmidt, "Detection of Coronary Artery Disease with an Electric Stethoscope," Aalborg University, Aalborg, Denmark, Denmark, 2011.

[4] KemenkesRepublikIndonesia. (2014, 24 November 2020). *Situasi Kesehatan Jantung*. Available: https://www.kemkes.go.id/resources/download/pusdatin/infodatin/infodatin-jantung.pdf

[5] R. L. Sulistyawati, "Kemenkes: Kelelahan Bukan Penyebab, tetapi Pemicu Kematian," in *Republika*, ed. Indonesia, 2019.

[6] M. Pouladian, M. R. H. Golpayegani, A. A. Tehrani-Fard, and M. Bubvay-Nejad, "Noninvasive detection of coronary artery disease by arterio-oscillography," *IEEE Transactions on Biomedical Engineering,* vol. 52, no. 4, pp. 743-747, 2005.

[7] K. Lei, L. Zhang, Y. Shen, X. Huang, and J. Wu, "Syndromes diagnostic model for coronary artery disease (CAD): An improved naïve Bayesian classification model based on attribute relevancy," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017, pp. 897-902.

[8] S. Swarup and A. N. Makaryus, "Digital stethoscope: technology update," (in eng), *Medical devices (Auckland, N.Z.),* vol. 11, pp. 29-36, 2018.

[9] S. Leng, R. S. Tan, K. T. C. Chai, C. Wang, D. Ghista, and L. Zhong, "The electronic stethoscope," (in eng), *Biomedical engineering online,* vol. 14, pp. 66-66, 2015.

[10] J. E. Suseno and M. Burhanudin, "The signal processing of heart sound from digital stethoscope for identification of heart condition using wavelet transform and neural network," in *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, 2017, pp. 153-158.

[11] H. Wang, J. Chen, Y. Hu, Z. Jiang, and C. Samjin, "Heart Sound Measurement and Analysis System with Digital Stethoscope," in *2009 2nd International Conference on Biomedical Engineering and Informatics*, 2009, pp. 1-5.

[12] J. Y. Shin, S. W. Lim, Y. C. Kim, S. J. Kim, E. J. Cha, and T. S. Lee, "Portable digital esophageal stethoscope system," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 1844-1847.

[13] J. Reynolds, J. Forero, J. Botero, V. Leguizamón, L. Ramírez, and C. Lozano, "Electronic Stethoscope With Wireless Communication to a Smart-phone, Including a Signal Filtering and Segmentation Algorithm of Digital Phonocardiography Signals," Singapore, 2017, pp. 553-556: Springer Singapore.

[14] W. Y. Shi, J. Mays, and J. Chiao, "Wireless stethoscope for recording heart and lung sound," in *2016 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireleSS)*, 2016, pp. 1-4.

[15] S. Szot, A. Levin, A. Ragazzi, and T. Ning, "A Wireless Digital Stethoscope Design," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, 2018, pp. 74-78.

[16] S. E. Schmidt, C. Holst-Hansen, J. Hansen, E. Toft, and J. J. Struijk, "Acoustic Features for the Identification of Coronary Artery Disease," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 11, pp. 2611-2619, 2015.

[17] V. Maknickas and A. Maknickas, "Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiological Measurement,* vol. 38, no. 8, pp. 1671-1684, 2017/07/31 2017.

[18] P. Samanta, A. Pathak, K. Mandana, and G. Saha, "Classification of coronary artery diseased and normal subjects using multi-channel phonocardiogram signal," *Biocybernetics and Biomedical Engineering,* vol. 39, no. 2, pp. 426-443, 2019/04/01/ 2019.

[19] S. Vernekar, 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, pp. 1141-1144.

[20] C. Sridhar, U. R. Acharya, H. Fujita, and G. M. Bairy, "Automated diagnosis of Coronary Artery Disease using nonlinear features extracted from ECG signals," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 000545-000549.

[21] A. AlHosani, S. AlShizawi, S. AlAli, H. Saleh, T. Assaf, and T. Stouraitis, "Automatic detection of coronary artery disease (CAD) in an ECG signal," in *2017 24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2017, pp. 414-418.

[22] A. Cüvitoğlu and Z. Işik, "Classification of CAD dataset by using principal component analysis and machine learning approaches," in *2018 5th International Conference on Electrical and Electronic Engineering (ICEEE)*, 2018, pp. 340-343.

[23] W. Zheng and M. Jin, "The Effects of Class Imbalance and Training Data Size on Classifier Learning: An Empirical Study," *SN Computer Science,* vol. 1, no. 2, p. 71, 2020/02/25 2020.

[24] P. K. Jain and A. K. Tiwari, "An adaptive thresholding method for the wavelet based denoising of phonocardiogram signal," *Biomedical Signal Processing and Control,* vol. 38, pp. 388-399, 2017/09/01/ 2017.

[25] Z. Zhao, "Noninvasive Diagnosis of Coronary Artery Disease Based on Instantaneous Frequency of Diastolic Murmurs and SVM," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 5651-5654.

[26] R. G. Brereton and G. R. Lloyda, "Support Vector Machines for classification and regression " *Analyst,* no. 2, 2010.

[27] M. Mohebbi and H. Ghassemian, "Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal," *Computer Methods and Programs in Biomedicine,* vol. 105, no. 1, pp. 40-49, 2012/01/01/ 2012.

[28] K. Veropoulos, I. Campbell, and N. Cristianini, "Controlling the Sensitivity of Support Vector Machines," in *IJCAI 1999*, 1999.

[29] A. M. Amiri, M. Abtahi, N. Constant, and K. Mankodiya, "Mobile Phonocardiogram Diagnosis in Newborns Using Support Vector Machine," (in eng), *Healthcare (Basel, Switzerland),* vol. 5, no. 1, p. 16, 2017.

[30] S. Y. Ho, "Structure and anatomy of the aortic root," *European Journal of Echocardiography,* vol. 10, no. 1, pp. i3-i10, 2009.

APPENDIX

Following are the results of feature extraction of phonocardiogram data on mitral, pulmonary and tricuspid.

TABLE VI
PHONOCARDIOGRAM FEATURE EXTRACTION ON MITRAL

Subject	mean	stdev	median	skewness	kurtosis	Label
1	0.14	0.15	0.07	0.99	-0.56	C
2	-0.08	0.12	-0.01	-1.21	-0.21	N
3	0.20	0.25	0.07	1.19	-0.23	N
4	0.01	0.02	0.01	0.83	-0.67	C
5	0.09	0.11	0.03	1.09	-0.35	C
6	-0.02	0.03	-0.02	0.38	-1.31	N
7	-0.17	0.20	-0.13	-1.09	-0.29	N
8	0.04	0.06	0.01	0.97	-0.47	C
9	-0.03	0.02	-0.02	-0.90	-0.49	N
10	0.05	0.06	0.01	0.95	-0.69	C
11	0.08	0.11	0.02	1.24	-0.17	N
12	0.00	0.04	-0.01	0.45	-1.30	C
13	-0.59	1.08	0.02	-1.35	0.02	C
14	0.10	0.11	0.05	1.10	-0.36	N
15	0.03	0.04	0.01	1.19	-0.22	C
16	0.02	0.04	0.01	1.32	0.03	N
17	-0.09	0.11	-0.03	-0.83	-0.82	C
18	-0.09	0.14	-0.02	-1.47	0.21	N
19	-0.01	0.05	-0.01	-0.48	-0.68	C
20	-0.17	0.22	-0.07	-0.99	-0.55	N
21	-0.05	0.06	-0.02	-1.04	-0.51	C

TABLE VII
PHONOCARDIOGRAM FEATURE EXTRACTION ON PULMONARY

Subject	mean	stdev	median	skewness	kurtosis	Label
1	0.01	0.02	0.02	0.11	-1.35	C
2	0.01	0.03	0.01	-0.08	-0.78	N
3	-0.23	0.29	-0.09	-1.23	-0.15	N
4	0.07	0.10	0.00	1.28	-0.10	C
5	0.06	0.08	0.03	0.84	-0.66	C
6	-0.07	0.08	-0.03	-1.13	-0.33	N

7	0.02	0.02	0.01	0.97	-0.56	N
8	0.01	0.01	0.00	1.24	-0.03	C
9	-0.07	0.08	-0.04	-1.12	-0.29	N
10	-0.12	0.12	-0.07	-0.94	-0.60	C
11	0.04	0.05	0.02	1.28	-0.08	N
12	0.07	0.08	0.02	0.73	-1.08	C
13	-0.27	0.37	-0.07	-1.20	-0.22	C
14	0.23	0.27	0.05	1.13	-0.36	N
15	0.10	0.14	0.03	1.21	-0.17	C
16	-0.10	0.16	0.00	-0.60	-1.36	N
17	-0.02	0.05	-0.01	-0.42	-1.18	C
18	0.17	0.19	0.10	1.11	-0.29	N
19	0.00	0.02	-0.01	0.45	-0.94	C
20	-0.12	0.16	-0.05	-1.34	0.02	N
21	-0.15	0.20	0.00	-0.70	-1.20	C

TABLE VIII
PHONOCARDIOGRAM FEATURE EXTRACTION ON TRICUSPID

Subject	mean	stdev	median	skewness	kurtosis	Label
1	-0.12	0.16	-0.02	-1.21	-0.23	C
2	-0.14	0.19	-0.04	-1.14	-0.32	N
3	-0.02	0.01	-0.01	-1.22	-0.04	N
4	0.20	0.24	0.08	1.22	-0.17	C
5	0.00	0.01	0.00	-0.29	-1.78	C
6	-0.07	0.10	-0.02	-1.16	-0.29	N
7	0.06	0.10	0.02	1.17	-0.15	N
8	-0.02	0.03	0.00	-0.43	-1.74	C
9	-0.08	0.11	-0.04	-1.06	-0.39	N
10	0.02	0.02	0.01	0.54	-1.55	C
11	0.04	0.06	0.01	1.10	-0.36	N
12	-0.08	0.11	-0.02	-1.25	-0.15	C
13	0.01	0.02	0.00	1.03	-0.56	C
14	0.00	0.02	0.00	0.00	-1.52	N
15	-0.02	0.03	-0.02	-1.01	-0.39	C
16	-0.17	0.22	-0.05	-1.19	-0.24	N
17	0.03	0.03	0.02	0.60	-1.11	C
18	0.01	0.02	0.00	0.79	-0.88	N
19	-0.01	0.03	0.00	-0.98	-0.38	C
20	-0.06	0.11	-0.01	-1.47	0.21	N
21	-0.09	0.08	-0.07	-0.25	-1.70	C