MEKATRONIKA
e-ISSN: 2637-0883
VOL. 2, ISSUE 2, 36 – 43
DOI: https://doi.org/10.15282/mekatronika.v2i2.6748

ORIGINAL ARTICLE

The Classification of Heart Murmurs: The Identification of Significant Time Domain Features

Cheng Wai Kit[1], Ismail Mohd Khairuddin[1*], Anwar P.P. Abdul Majeed[1] and Mohd Azraai Mohd Razman[1]

[1]Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang (UMP), 26600 Pekan, Pahang Darul Makmur, Malaysia

ABSTRACT – Phonocardiogram (PCG) is a type of acoustic signal collected from the heartbeat sound. PCG signals collected in the form of wave files and collected type of heart sound with a specific period. The difficulty of the binomial class in supervised machine learning is the steps-by-steps workflow. The consideration and decision make for every part are importantly stated so that misclassification will not occur. For the heart murmurs classification, data extraction has highly cared for it as we might have fault data consisting of outside signals. Before classifying murmurs in binomial, it will involve multiple features for selection that can have a better classification of the heart murmurs. Nevertheless, since classification performance is vital to conclude the results, models are needed to compare the research's output. The main objective of this study is to classify the signal of the murmur via time-domain based EEG signals. In this study, significant time-domain features were identified to determine the best features by using different feature selection methods. It continues with the classification with different models to compete for the highest accuracy as the performance for murmur classification. A set of Michigan Heart Sound and Murmur database (MHSDB) was provided by the University of Michigan Health System with chosen signals listening with the bell of the stethoscope at the apex area by left decubitus posture of the subjects. The PCG signals are pre-processed by segmentation of three seconds, downsampling eight thousand Hz and normalized to -1, +1. Features are extracted into ten features: Root Mean Square, Variance, Standard Deviation, Maximum, Minimum, Median, Skewness, Shape Factor, Kurtosis, and Mean. Two cross-validation methods applied, such as data splitting and k-fold cross-validation, were used to measure this study's data. Chi-Square and ANOVA technique practice to identify the significant features to improve the classification performance. The classification learners are enrolled and compared by Ada Boost, Random Forest (RF) and Support Vector Machine (SVM). The datasets will separate into a ratio of 70:30 and 80:20 for training and testing data, respectively. The chi-Square selection method was the best features selection method and 80:20 data splitting with better performance than the 70:30 ratio. The best classification accuracy for the models significantly come by SVM with all the categories with 100% except 70:30 test on test data with 97.2% only.

ARTICLE HISTORY
Received: 4th Nov 2020
Revised: 29th Nov 2020
Accepted: 13th Dec 2020

KEYWORDS
PCG
Heart Murmur
Machine Learning
Classification
Time Domain
Feature Selection

INTRODUCTION

Heart murmurs are a type of sound which is showing abnormalities of the cardio flowing system. This natural disease does not count on sex, age, race, or occupation. The murmurs have several types of sound, which are whooshing, tickling, or clicking sound beats. It lacks the normal sound of "lub dub". However, most of the murmurs are not life-threatening and need not be treated immediately. However, the worse condition for a heart murmurs effect can make the heart beat faster or slower than average person one and lead to death due to a heart attack. Commonly, this innocent murmur is hardly being treated; thus, it is usually resolved independently. Heart murmurs are defined as mystery as they are easily come and go and do not get permanently stay unless it was not being found and solved.

Phonocardiogram (PCG) is the type of acoustic signal collected from the heartbeat sound. It is helpful to determine the heart murmur's part by converting this heart sound to the array by converting the wave to code method. Auscultation is the collection of processes. To show that the heart sound cycle over a certain period, the transmitted software needs to be connected. Below are the categories of the normal heart sounds and heart murmurs with presence mode on the serial monitor with ideal results:

The extraction of characteristics is also the first stage in every artificial intelligence system. For example, the identification procedures given are the first stage in diagnosing mechanical/electrical systems utilising vibration data and illness monitoring using various biological signals. The time domain, frequency domain, or Time-Frequency (TF) domain will be the extracted features. Time-domain characteristics are simpler and more computationally feasible than transformed domain characteristics. Therefore they do not necessitate a sophisticated domain transformation [1], [2].

*CORRESPONDING AUTHOR |Ismail Mohd Khairuddin | ✉ ismailkhai@ump.edu.my

This study aims to identify the significant time-domain features that could yield the highest classification accuracy and determine the best classifiers compatible with the classification of heart murmurs. Several classifiers such as Ada Boost, Random Forest (RF) and Support Vector Machine (SVM) will be used to determine the classification accuracy of the results.

RELATED WORK

Heart murmurs are a large family with many different types. There are two forms of blood pressure: systolic and diastolic. Then there is Aortic stenosis (AS) classes, mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP) classes [1]. Moreover, two more hidden sounds of s3 and s4 are founded in murmurs categories. Minor murmurs identification to have more separation types of murmurs [3]. This research helps to find the dataset of the heart murmurs too. We can reference the classifier that is suitable for the murmurs type based on the related data. The murmurs could have higher accuracy when the tuned parameter is compared [4]. The dataset that could be complete and ideally choose are found in the University of Michigan Heart murmurs library and "Yaseen21khan" [3], [5].

PCG is a test collection from the electronic stethoscope with microphones and signals transmitters present on the monitor. PCG signals involve frequency and amplitude determination with the correlation with cycle-time. Commonly, heart murmurs having a higher amplitude in the segmentation of s1 and s2 compare with normal sounds [6]. Therefore, PCG signals need to have lots of pre-processing. For example, denoising, normalization, segmentation, filtering, resampling and sampling methods [7]. In addition, there are various steps for the PCG signals in time-domain features, including all statistical-based extraction [1]. PCG signals also need feature selection to avoid the outliers data to have more significant results [8].

[9] investigated the heart sound classification with feature extraction of Phonocardiography Signal. The dataset is using a phonocardiogram signal (PCG) with thirty normal and thirty murmur signals. The features extraction is using time domain, frequency domain, cepstrum and statistical features. The total numbers are twenty-three features. Therefore, all Ranker and Info Gain Attribute Evaluation was practised, and five optimal features are gained after that. The classification models used are Bayes Net, Naïve Bayes, Stochastic Gradient Descent (SGD), and Logit Boost, while the sampling method has five-fold cross-validation. The results were gained by 93.33% accuracy in Naïve Bayes.

[6] had researched the detection of cardiac abnormality, which is murmurs too from the PCG signals. The datasets are having sixty-four different recordings of heart sounds with five different pathological cases. All the collection is PCG signals. There are pre-processing and segmentation for the features extracted.

Moreover, the time-frequency domain is another step for continuity of Lagrange multipliers (LMS algorithm) in selecting features. Compared with proposed methods, Least Square Support Vector Machine (LS-SVM) and regular SVM are classified. The findings ended with proposed techniques giving 86.718% accuracy of recognition and 4.296% and 8.593% more than LS-SVM and regular SVM with the polynomial kernel.

METHODOLOGY

Data Acquisition

In the library of Michigan Heart Sound and Murmur database (MHSDB), there are four data sections in PCG signals collected by the University of Michigan Health System. A total of twenty-three heart sound in murmur and normal class with a total length of 1496.8s. [10]. The section consists of the Apex area, Aortic Area and Pulmonic Area, with supine, sitting, and left decubitus. The device for collections is with bell stethoscope and diaphragm stethoscope. Section two of Aortic Area was chosen to acquire the type of heart sound with left decubitus posture; bell stethoscope collects from apex area of the heart. Table 1 depicts the six signals data with sixty lengths above for each set listed the data name with the code and its length [11].

Table 1. List of the raw data collect from heart sound library of University of Michigan Health System

Ref. No	Name	Length (s)	Types
03	S4 gallop bell	75	Normal
05	S3 gallop bell	68	Normal
10	Systolic Click & Late Systolic	61	Murmur
11	S4 & Mid-Systolic Murmur	64	Murmur
12	S3 & Holo-Systolic Murmur	65	Murmur
13	Opening Snap & Diastolic	61	Murmur

Project Flowchart

The experiment starts with the data acquisition collected from Section 2 of the University of Michigan Health System [10]. The Signals are in PCG signals, and 120 instances were collected and group with 40 normal sounds and 80 murmurs sounds. Ten time-domain features are extracted by using MATLAB (2018b). Mean, Variance, Standard Deviation, Kurtosis, Shape Factor, Root Mean Square, Maximum, Minimum, Median and Skewness. All the data were export to an excel file. After that, input the particular file to Orange Data Mining to have features selection. Seven significant features were selected by the Chi-Square method, and eight significant features were selected by the Analysis of variance (ANOVA) method. The next step is to classify the identified features by various models. Finally, compare the classification accuracy to determine the best classifier among them.

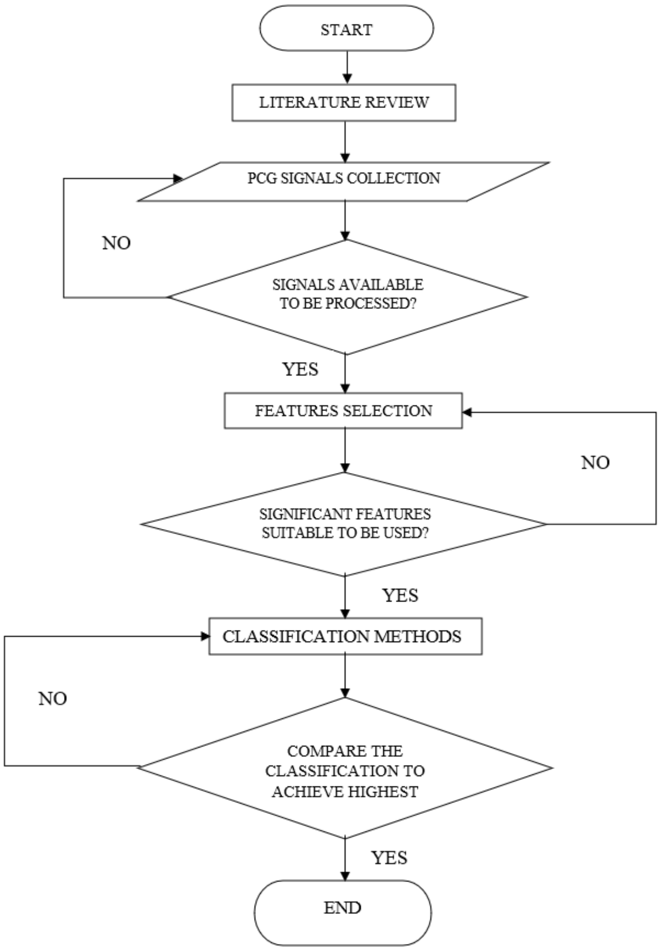

Figure 1. Process Flow Chart for research study methodology.

Signal Pre-processing, Feature Extraction and Selection

PCG signals are collected under a very calm and quiet surrounding to avoid noise. The part of denoising can be left out so that data is easily processed. Although denoising can be avoided, the full cycle of the PCG signals needed to be separated by part. We called it segmentation. In this section, each of the signals is cut into twenty samples for three seconds. Therefore, hundred and twenty instances were ready to use. Segmentation is for better convenience to display and testing [3]. Next, segmentation needed to be done carefully as heart murmurs signals have various types in sampling rate and resolution bit. Online Wav processed a sixteen-bit resolution with an 8000Hz sampling rate in mono audio format to achieve the same data size [6]. Hence, the data size in hundred and twenty instances have the same value of 24,000 for each. 24,000 data size is referring to the amplitude of three seconds, and each amplitude happens at 0.125ms. The reason to choose sixteen bits resolution and eight thousand Hz sampling rate is that there are no missing values compared to other parameters. When transferring a sample or other signal to a different system that records or saves at a different sample rate, it is occasionally required to downsample. In the past, downsampling was often performed on samplers in order to save memory. If a sample did not have much high-frequency content, it could be, for instance, downsampled from a 44.1kHz sampling rate to 8 kHz, cutting the amount of memory needed to store the sample. However, it is better to downsample the data in this dataset to achieve the desired target, although it may affect the original data gained. There is a last step for the data to have a good range of (-1, +1) with normalization that practice inside MATLAB apps (2018b). All the 120 instances will bring forward to the next section for features extraction. Ten time-domain features were then extracted and features selection will be added value to those data. It is recommended to have it impact or influence the performance of the results. Some irrelevant features may affect the results, and we need to ignore them or find the best ranked to have the steps for classifier only. Moreover, the training time for the classifier will reduce as we are not needed to have such values that are too big or too small. Therefore, features selection is also saving the time of the users while having the classification parts. All this part is the best preparation for the last important process, classification. Figure 2 illstrates the pre-process signal.

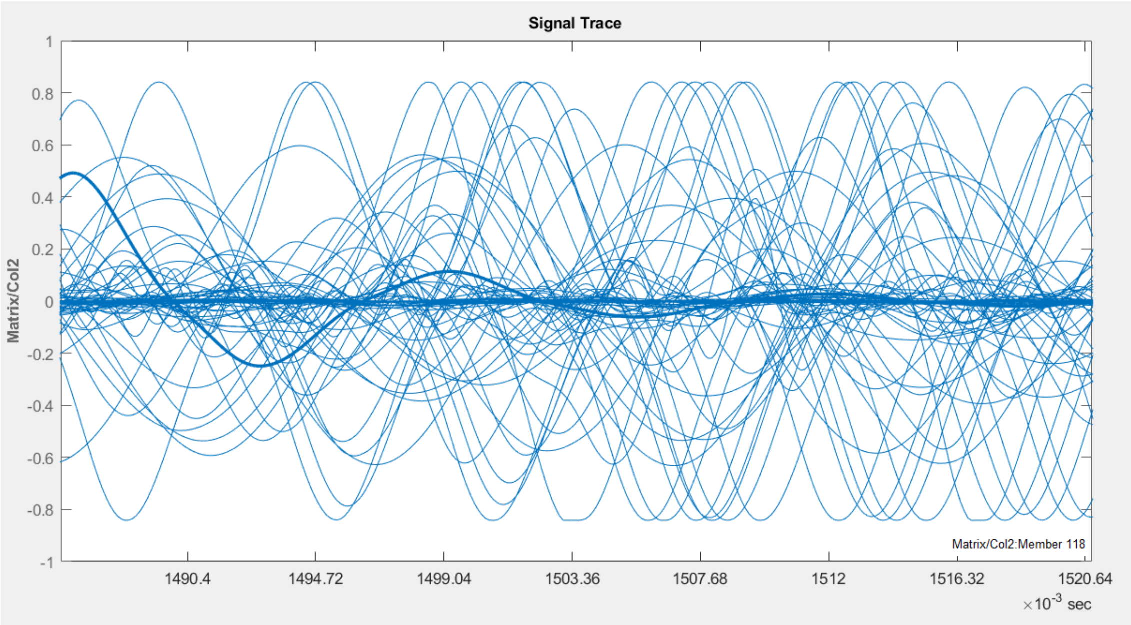

Figure 2. Pre-process PCG signal for a heart murmur

The Chi-Square test is a statistical test commonly used to examine the independence of categorical variables and the target variable in a dataset as in Equation 1, where *O* is the observed values, and *E is* denoted as the expected value. First, the link between features and the target variable is determined. Then, the rank score of the chi-square represents the dependency of the class or target. Thus, it is a good features selection to find out the value of PCG signals that are reliable on murmurs' class.

$$Chi - square(\mathcal{X}^2) = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

ANOVA is a statistical method that determines the significant difference in the means of two or more with a hypothesis. F-statistic will determine the ranking of the features, which the higher the value, the more important the feature is. Thus, F-statistic is the statistic that measures whether the means of different samples differ substantially or not. For example, suppose the value of F-statistic for the feature is enough to reject the null hypothesis in ANOVA thus. In that case, it will be included as an important feature, and the level of importance will depend on F-statistic values.

Classification Models

Ada Boost

AdaBoost is a machine learning algorithm that may be used to improve the performance of any other machine learning technique. It works well with students who are struggling. On a classification task, these are models that reach accuracy just above random chance. The model stumps are associated stage by stage, and each prior stage will be repaired. Models are added until the dataset is perfectly predicted or until the maximum number of models is reached.

Random Forest (RF)

Random Forest is a scalable and easy supervise learning method that produces the best results even when no hyper-parameters are changed. Random forests are a set of tree predictors in which each tree is based on a set of random vector values that are sampled independently for all trees in the forest and have the same distribution. The intensity and connection between the individual trees in a tree classifier forest determine the generalisation error.

For bagging, the hyperparameters of a random forest are comparable to those of a decision tree or a classifier. Forest at random Because the random forest classifier class may be simply employed, there is no need to combine a decision tree with a classifier for bagging. Furthermore, Random Forest increases the model's randomness while increasing the number of trees. In general, the more trees you have in your forest, the better your RF classification performance will be. Because it can identify the essential features from the training set, the RF algorithm is an excellent choice for feature engineering.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a technology that represents an algorithm for classification and regression analysis in supervised learning for data analysis. The classification will be carried out using this classifier by locating the hyper-plane or line between the two or more classes or categories. SVM will use the maximum margin to distinguish between the two classes. In other words, SVM is a classifier that is used to find the hyper-plane to achieve the best separation of features into different classes [12].

Its kernel trick is well-known in the SVM classifier. This kernel is a method for computing the dot product of two vectors for features in SVM. Generally speaking, the kernel defines two vectors: x and y [12]. Kernel-based SVM methods rely on a set of mathematical functions known as kernels. The kernel's job is to take data and convert it into the correct format.

Data Splitting Method

The basic sampling distribution to distinguish the dataset into two parts is the hold-out method, training and testing data. Testing data is considering the hold-out part for the testing and learn the model from training data. This method can save our time to evaluate the machine learning model. Since the evaluation of the models is different and depends on the separation ratio, it affects the results of CA, too [13]. In this research, 80:20 is used as the standard way for sampling separation 80% on the training datasets and 20% on the test datasets. An additional 70:30 ratio data is splitting to make a comparison on it.

k-Fold Cross-Validation

This method is the method to prevent overfitting. Overfitting is the error of the datasets of the test datasets has higher CA than the training datasets. The presence of the k-fold validation is to cover or overcome as the training dataset. Therefore, the results are considering the new training dataset to compare. It is not good to overfitting; the same goes for underfitting. It is recommended that the CA for both is between 0%-5% only. This study recommends using normal 10-fold cross-validation to have a standard value that gains the best Goodness of Fit evaluation.

RESULTS AND DISCUSSION

The features selection method consists of ANOVA and the Chi-Square method to improve the classification performance of all the classifiers. Chi-Square seems to be the better features selection method compared to ANOVA. It selects the seven best features from ten extracted features. The scores calculated from it is slightly the same as ANOVA, but the deducted features are not the same as ANOVA. Furthermore, the classification performance after features selection of Chi-Square is increased in percentages. Both of the selection methods were done by using Orange Data Mining (widget of rank). Three was eliminated from the ten features extracted: Minimum, Kurtosis, and Mean. Three of the features had the lowest scores that are two and below. Thus, we ignored it to obtain a better classification accuracy. The similarity of features reduction for ANOVA is to eliminate 'Mean' features. However, the Shape Factor feature's score in ANOVA is below one and is not selected for the classification. Figures 3 and 4 are the features selected listed for both Chi-Squares and ANOVA with their scores achieved.

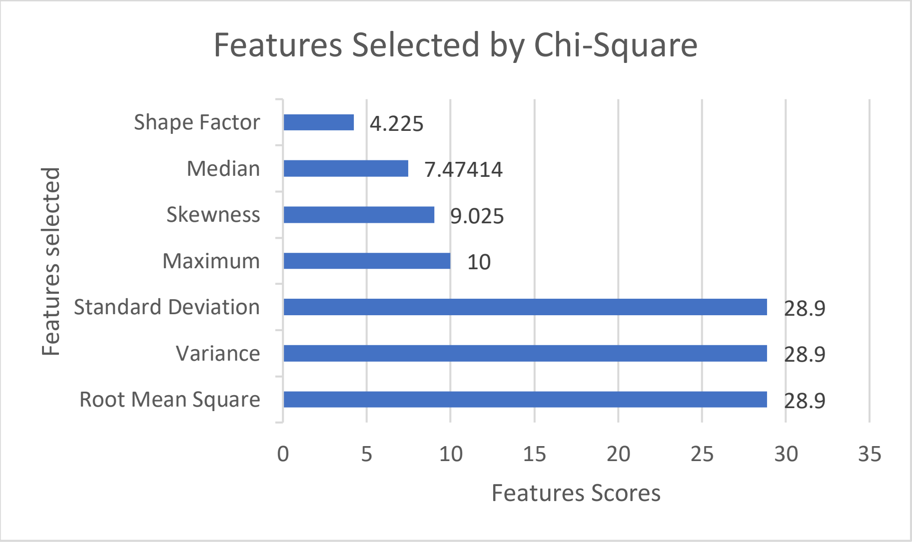

Figure 3. Seven best features selected by Chi-Square

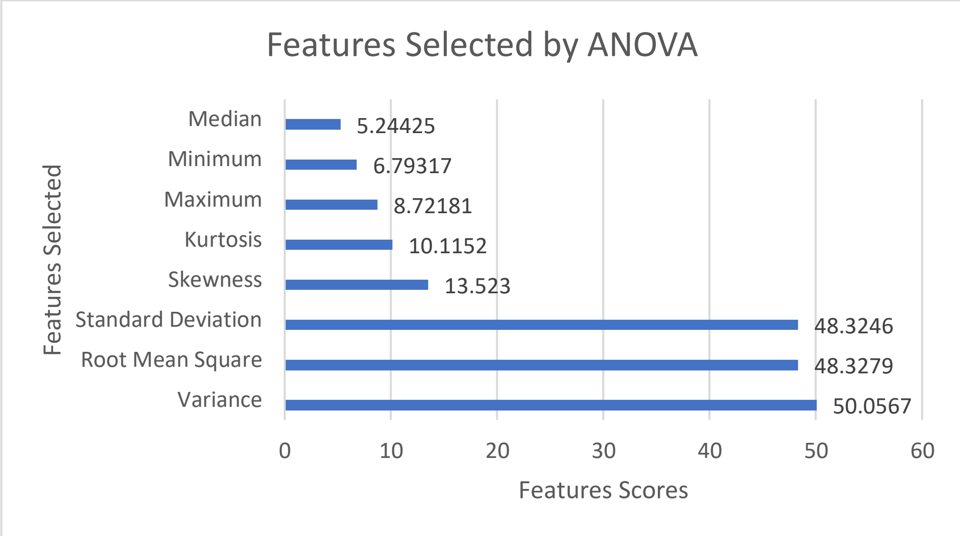

Figure 4. Eight best features selected by ANOVA

Throughout all the results in classification accuracy, Support Vector Machine (SVM) appeared to be the best classifier compared to other models rely on its high classification accuracy performance in 80:20 data splitting. It only has a slightly lower accuracy proportion at 70:30 data splitting with ANOVA features selection to overcome the overfitting problem. On the overfitting problem, Random Forest (RF) classifiers turn up an excellent example at the data splitting 70:30. It is gained by lowering overfitting error after the ten-fold cross-validation sampling approach replaces the test value on train data. In addition, Ada Boost rose to be the classification accuracy increase after selecting the Chi-Square method. The output reaches a maximum of 100% in all categories. Figure 5 and figure 6 will show all the model's overall performance with all the sampling methods in data splitting 70:30 and 80:20.

After analysing and discussing the results gained, the classification performance is astonishing, whether in original features, Chi-Square, and ANOVA with all the sampling methods. In addition, features selection on ANOVA and Chi-Square also slightly increases the training and testing time, which I have mentioned above. From the results gained, data splitting with 80:20 have higher classification accuracy compared to 70:30 as most of the models reach a maximum of 100% except AdaBoost with 10-fold cross-validation appeared to be 99%.

Nevertheless, Ada Boost models presented the improvement of the classifier after a Chi-Square features selection technique applies. The confusion matrix gives out visible results for which type of class had misclassification. There is mostly only one in either murmur class or normal class for the classifier that does not reach 100% accuracy. Other than that, all perform well in the classification of the murmur and normal. Overall, the misclassification error is lower than 5%, which is only one instance error counted. Table 2 illustrates the confusion matrices for SVM on Chi-Square.

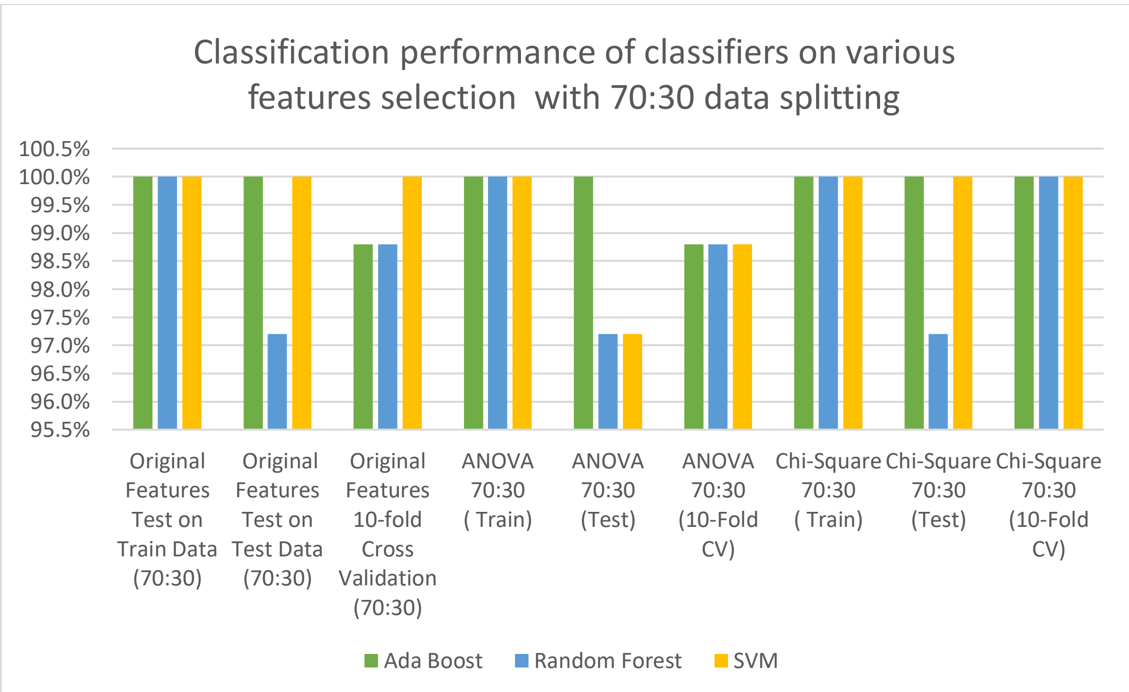

Figure 5. Classification performance of classifiers on various features selection with 70:30 data splitting

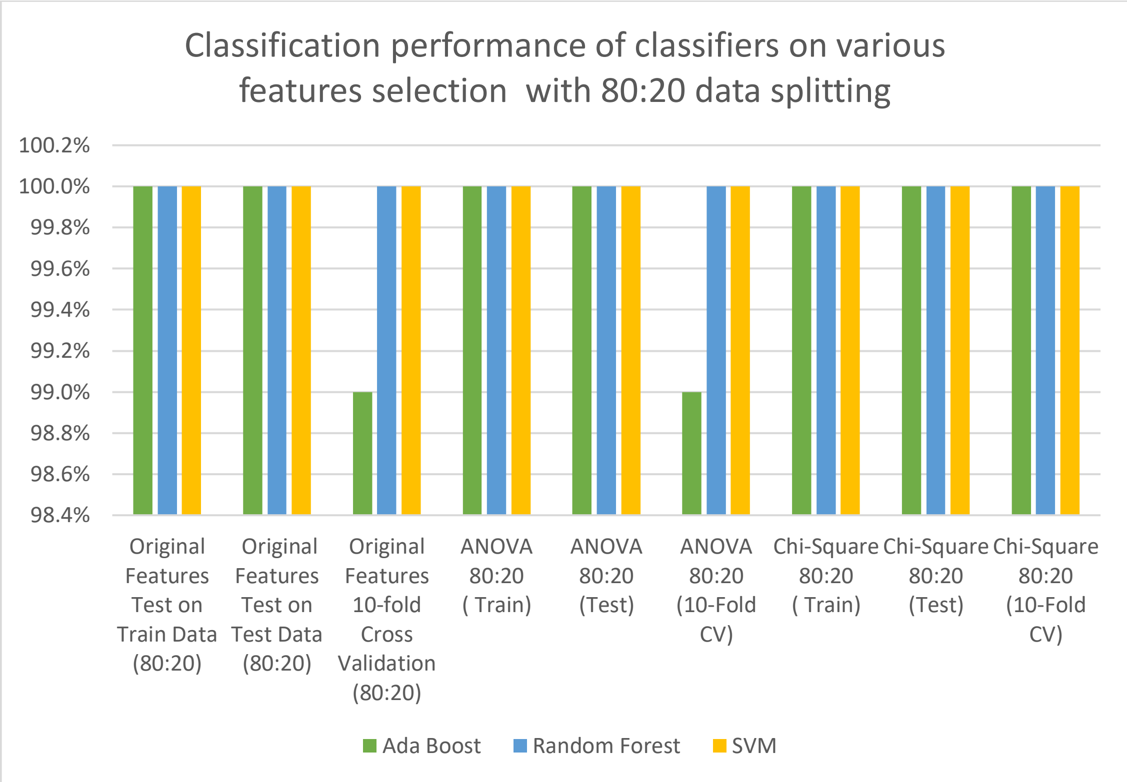

Figure 6. Classification performance of classifiers on various features selection with 80:20 data splitting

Table 2. Confusion Matrix for SVM on Chi-Square Features selection on 80:20 Training and Testing Set

		Predicted		
		Murmur	Normal	Σ
Actual	Murmur	56	0	56
	Normal	0	28	28
	Σ	56	28	84

		Predicted		
		Murmur	Normal	Σ
Actual	Murmur	24	0	24
	Normal	0	12	12
	Σ	24	12	36

CONCLUSION

The present study evaluated the identified significant statistical time-domain features towards classifying the onset of the heart murmur. The CA of machine learning machine models had show improvement with the implementation of the feature selection method to identified the seven best features from a total of 10 features extracted. On top of that, it is shown that SVM classifiers had achieved outstanding results compare to others selection models. In the future, this study can be improved by implementing different types of feature selection methods and other machine learning models in classifying heart murmur. Other than that, Hyperparameter tuning can be applied if the classification models focus on machine learning models as one of the suggestions to generate better CA in classifying heart murmur.

REFERENCES

[1] P. Careena, M. Mary Synthuja Jain Preetha, and P. Arun, "Research on murmur from time domain features of heart sounds," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1 SpecialIssue4, pp. 736–743, 2019.

[2] I. M. Khairuddin, S. N. I. Sidek, A. P. P. A. Majeed, and A. A. Puzi, "Classifying Motion Intention from EMG signal: A k-NN Approach," *2019 7th Int. Conf. Mechatronics Eng. ICOM 2019*, pp. 1–4, 2019.

[3] Q. Liu, X. Wu, and X. Ma, "An automatic segmentation method for heart sounds," *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–22, 2018.

[4] Y. Zheng, X. Guo, and X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2710–2721, Apr. 2015.

[5] Yaseen, G. Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, 2018.

[6] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, 2010.

[7] F. Safara, S. Doraisamy, A. Azman, A. Jantan, and S. Ranga, "Wavelet packet entropy for heart murmurs classification," *Adv. Bioinformatics*, vol. 2012, no. November, 2012.

[8] D. Kristomo, R. Hidayat, I. Soesanti, and A. Kusjani, "Heart sound feature extraction and classification using autoregressive power spectral density (AR-PSD) and statistics features," *AIP Conf. Proc.*, vol. 1755, pp. 1–8, 2016.

[9] M. Singh and A. Cheema, "Heart Sounds Classification using Feature Extraction of Phonocardiography Signal," 2013.

[10] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Nov. 2016.

[11] "Professional Skill Builder | Heart Sound & Murmur Library.".

[12] A. Kumar, L. Bi, J. Kim, and D. D. Feng, *Machine learning in medical imaging*. Elsevier Inc., 2020.

[13] A. Abdo Salman, I. Mohd Khairuddin, A. P.P. Abdul Majeed, and M. A. Mohd Razman, "The Diagnosis Of Diabetic Retinopathy By Means Of Transfer Learning And Fine-Tuned Dense Layer Pipeline", MEKATRONIKA, vol. 2, no. 1, pp. 68–72, Jun. 2020