

A Machine Learning Approach to Heart Murmur Detection and Classification

Alisa D. Levin, Anthony Ragazzi, Skyler L. Szot and Taikang Ning
Department of Engineering, Trinity College Hartford, Connecticut USA

Abstract—This paper presents a heart murmur detection and classification approach via machine learning. Over 450 clinical heart sound episodes collected from patients of different cardiac alterations were evaluated. We extracted heart sound and murmur features of diagnostic importance, as well as nontraditional features based on our expert domain knowledge, via AR modeling and the amplitude profile. A supervised machine learning approach was implemented with k-nearest neighbor (KNN) and support vector machine (SVM) models in order to evaluate the performance of our heart sound features for murmur classification. Our unique set of features resulted in a classification accuracy consistently exceeding 90% for both classifiers.

Keywords—heart sounds; heart murmurs; classification; supervised machine learning

I. INTRODUCTION

Heart disease is the number one cause of death worldwide. In fact, in 2016, 31% of deaths across the globe were due to cardiovascular disease [1]. The primary and most common tool for bedside diagnosis of possible cardiovascular alteration is the classical stethoscope, which was invented over two hundred years ago. With the aid of the stethoscope, doctors use their ears to detect abnormal heart murmurs that signify the presence of heart disease. However, cardiac auscultation is subjective and varies in accuracy depending on each individual physician's experience and hearing ability. Studies of primary care physicians have found that proficiency in cardiac auscultation for clinically important heart sounds and murmurs ranges only between 20 and 50 percent [12][13]. It is therefore no wonder that many patients referred to cardiologists for echocardiograms and further examinations are found to be healthy. This not only results in a negligible use of medical resources and unnecessary expenses, but also demonstrates the need for a more accurate method of cardiac auscultation [2].

Modern diagnostic medical equipment has witnessed great improvement with embedded microprocessors, from digital thermometers to blood pressure monitors, yet many medical professionals still rely on analog stethoscopes for cardiac auscultation despite their documented limitations. Automatic cardiac auscultation and heart murmur classification algorithms have been developed, but they are limited by the patterns that researchers can find to distinguish between different heart sound features and heart murmur types. With the recent explosion in online data and low-cost processing capabilities, it is only a logical step to apply machine learning

methods to automatic cardiac auscultation in order to facilitate more evidence-based diagnoses in healthcare [11]. Machine learning algorithms can distinguish features, and relationships between features, that human eyes and ears cannot recognize. Additionally, as a heart sound travels through a physician's analog stethoscope and into their ear, it becomes gradually distorted. To simulate this, some digital cardiac auscultation analyses modulate recorded heart sounds to resemble the distortion of sounds processed by the human ear. In this study, we instead used unaltered signals with the full frequency range in order to maximize the amount of data that could be useful in classifying different types of murmurs.

Machine learning has become an increasingly popular decision-making method in all industries—from marketing and commerce to science and education [11]. In recent years, it has been applied to biomedical classification and diagnostic algorithms for health issues such as skin disease, diabetic retinopathy, and breast cancer [8]-[10]. In our biomedical research, we are using machine learning for heart murmur classification. For example, a physician might describe a ventricular septal defect as having a mid-systolic, decrescendo mitral murmur with a “blowing” quality. Through feature extraction, these qualitative measures can be transformed into quantitative features such as pitch, heart sound duration, heart murmur duration, and onset time [6]. A machine learning analysis of these features allows us to find connections between the physician's description and their diagnosis. Furthermore, machine learning can identify connections between features and heart murmur classifications that are not known or used by physicians in heart murmur diagnoses.

Investigating further into trends of data-driven decision-making, this paper presents a method for heart murmur classification using supervised machine learning. Several intuitive parameters are used to describe a particular heart murmur (Outlined in Section II. A). These parameters can be extracted from a given heart sound recording using methods described in previous studies [6]. In addition to common heart sound parameters described by physicians, we extracted a number of other, nontraditional features based on our expert domain knowledge. We hypothesized that these features, imperceptible to even a highly trained human ear, would be advantageous to heart murmur classification. We propose a supervised machine learning approach using these parameters to identify and classify the different types of heart murmurs. For this

study, we trained two popular classification models, the k-nearest neighbor model and support vector machine model, and evaluated their performance in order to determine the usefulness of our feature set for identifying different types of murmurs. The remainder of this paper is organized as follows: Section II is devoted to murmur feature extraction and machine learning classification methods. In Section III, the validity of the proposed feature extraction methods is evaluated with heart sound episodes containing distinct types of murmurs—including early, mid-, late and holosystolic murmurs, and early and mid-diastolic murmurs—through the presentation and discussion of our results. A conclusion is provided in Section IV.

II. METHODS

A. Feature Extraction

The features of heart sounds and murmurs previously described are essential to an effective machine learning system. To extract relevant features that capture essential characteristics of heart sounds and murmurs in both the time and frequency domains, we adopted a signal segmentation approach by dividing digitized heart sound episodes into many short segments of 10-msec duration. Important features were extracted from each 10-msec segment, to allow for a profile description of complete cardiac cycles as well as short time variations.

To observe the heart sound and murmur intensity, an average magnitude value (ABV) index of each short segment was computed.

$$ABV_n = \frac{1}{N} \sum_{k=1}^N |x(k) - \mu_n| \quad (1)$$

where μ represents the mean value of the n^{th} 10-msec segment. We have found that ABV indices are effective in identifying

the occurrence of S1 and S2 and detecting murmurs when it was used together with other parameters to be described later. For example, Fig. 1 displays an early diastolic murmur signal (top trace) and its ABV indices (middle trace) of consecutive 10-msec segments. Local maxima of the ABV plot are recorded for their locations and associated values. With the assumption that systole is about 3/8 of the cardiac cycle and diastole is 5/8 of a cycle, the first and second heart sounds are labeled (bottom trace of Fig. 1).

The systole, the interval between the first and second heart sounds (S1 to S2), and the diastole (S2 to S1) were extracted, respectively. Without loss of generality, heart sound episodes used in our study all began with systole and followed by diastole. The boundaries of S1 and S2 were determined when the ABV was reduced below 10% of the heart sound. The existence of a murmur in systole and/or diastole was marked when continuous 10-msec segments showed an ABV more than 25% either the magnitude of S1 or S2 values. It should be noted that the detection threshold could be adjusted for varying sensitivity as needed.

Systolic murmurs and diastolic murmurs were detected and delineated separately, using a similar procedure. Once detected, the murmur onset and duration time were recorded. In addition, the average ABV and the average murmur frequency were also calculated. The murmur frequency was efficiently estimated using a second-order linear prediction AR model [4]. The 2nd order AR model is described below.

$$\begin{aligned} e^f &= x(k) - a_1x(k-1) - a_2x(k-2) \\ e_b &= x(k) - a_1x(k+1) - a_2x(k-2) \end{aligned} \quad (2)$$

The estimated AR model coefficients $\{a_1, a_2\}$ were used to capture the murmur pitch frequency by the following

$$pitch = \frac{f_s}{2\pi} \tan^{-1} \left(\frac{\sqrt{4a_2 - a_1^2}}{a_1} \right) \quad (3)$$

where f_s is the heart sound signal sampling frequency. The pitch frequency of each 10-msec segment of the detected murmur is calculated. The example of a detected diastolic murmur and pitch frequency by 10-msec segments is shown in Fig. 2.

Through our repeated trial and error and experience, we noticed some additional patterns that would be valuable for heart murmur recognition and classification. Inspired by our observations, we generated 16 unique features that are not currently used by physicians for heart murmur diagnosis. For example, we calculated features based on the ratio of average amplitude of systole and diastole to the average amplitude of

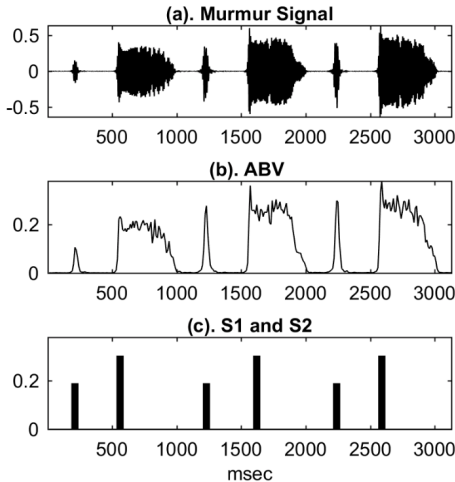


Fig. 1. Heart murmur signal, ABV indices, and detected S1 and S2.

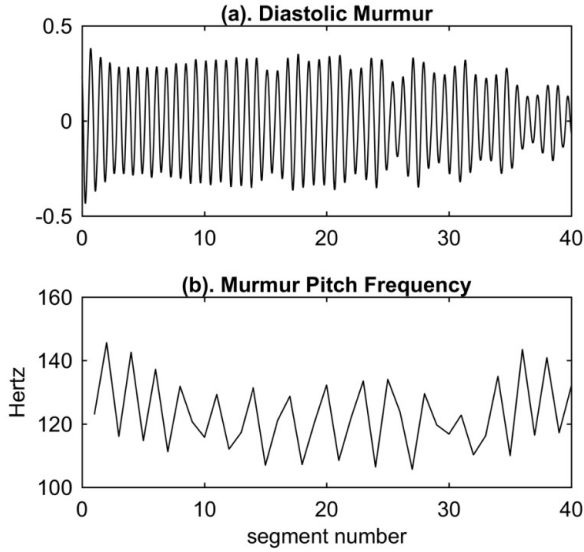


Fig. 2. Extracted diastolic murmur and pitch frequency.

S1. These features, we theorized, could be beneficial since such ratios are usually very small when no murmur is present.

We also generated a series of features using two threshold values for each period, 40% and 10% of S1 for systole and 40% and 10% of S2 for diastole. This created three regions: above the largest threshold, between the two thresholds, and below the smallest threshold. Each peak in systole and diastole was placed into one of these three regions. We found that murmurs are likely to have more peaks that exceed higher thresholds, but reasoned that it would be advantageous to measure this at different levels to accommodate variations in amplitude, and consequently detect quieter murmurs.

Additionally, we calculated features relating to the variance in peaks in both systole and diastole periods. These features reveal a rough estimate of frequency variance, which could be useful because murmurs typically have more consistent frequency. Finally, we noticed that heart sounds without murmurs have horizontal systole and diastole silhouettes, while those with murmurs have more irregular slopes. To capture this, we extracted features based on the derivatives of both systole and diastole silhouettes.

B. Machine Learning Classification

Once the heart sound features were successfully extracted from the heart sound recording, these data served as parameters for our machine learning models. We implemented a supervised machine learning approach to classify the heart sounds into seven categories: early, mid-, late and holosystolic murmurs, diastolic murmurs, a combination of systolic and diastolic murmurs, and normal heart sounds without murmurs, testing the validity of our extracted features. We applied the k-nearest neighbor (KNN) and support vector machine (SVM) classification algorithms and compared their performance using

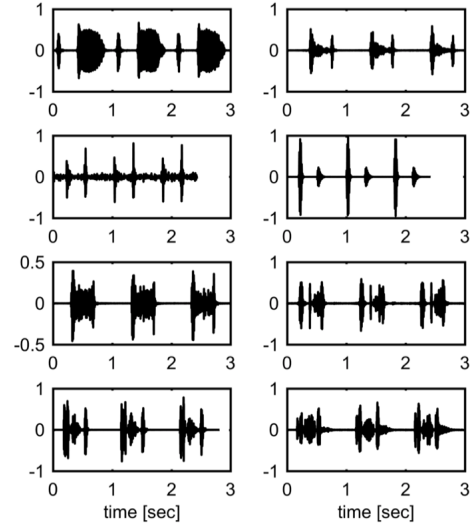


Fig. 3. Different heart sound and murmur examples.

80–20 and 90–10 splits through 10-fold cross-validation. We compared the accuracy of correctly classified heart sounds for KNN and SVM classification algorithms under these different testing scenarios.

III. RESULTS & DISCUSSION

To ensure a consistent evaluation of our approach, our classification accuracy score was calculated with 10-fold cross-validation on 453 clinically recorded heart sound episodes. Fig. 3 exemplifies a few short heart sound and murmur episodes that were analyzed in our study.

Since overfitting and underfitting are important concerns in machine learning, we addressed the issue by comparing the accuracy scores of each model between 80–20 split and 90–10 training and test set splits. Our KNN model had a classification accuracy of 90.11% with an 80–20 split and an accuracy of 91.43% with a 90–10 split. Similarly, our SVM model with a linear kernel had a classification accuracy of 92.09% with an 80–20 split and an accuracy of 94.51% with a 90–10 split (See Table I). The fact that the classification accuracy scores of KNN and SVM increased by only about 1.4% and 2.6% respectively with a 90–10 split indicates that the models are not sensitive due to our robust features.

After investigating misclassification patterns of our classifiers throughout multiple confusion matrices, we garnered the following observations:

- Holosystolic murmurs are most likely to be mistaken for mid-systolic or diastolic murmurs, and vice versa. The former misclassification is possibly due to the fact that holosystolic and mid-systolic murmurs often both extend for the vast duration of the systole.
- Late systolic murmurs are most likely to be mistaken for healthy heart sounds without murmurs present, perhaps

TABLE I
CLASSIFICATION RESULTS

Model	Train/Test Split	Accuracy (%) 10-fold cross-val.	Average Precision	Average Recall
KNN	80–20	90.11	0.863	0.869
	90–10	91.43	0.866	0.860
SVM	80–20	92.09	0.913	0.922
	90–10	94.51	0.913	0.924

because late systolic murmurs can be so brief that they blend into the S2 heart sound.

- Early systolic murmurs are most likely to be mistaken for holosystolic murmurs, perhaps because some early systolic murmurs extend quite far into the systole period.
- Heart sounds with both systolic and diastolic murmurs present are most likely to be mistaken for holosystolic murmurs. This is perhaps because the systolic murmurs in these heart sound episodes are often mid-systolic murmurs, which are frequently mistaken for holosystolic murmurs.

It is also important to note that we only classified six common types of murmurs along with healthy heart sounds without murmurs—and did not designate specific labels for heart sounds with clicks, S3s, S4s or other variations. These additional abnormalities are likely also factors in the murmur misclassifications described above.

Despite these classification challenges, our models scored well in both precision and recall. With a 90–10 split, we found that our KNN classifier had an average precision of 0.866 and an average recall of 0.860. Our SVM classifier, even better, had an average precision of 0.913 and an average recall of 0.924 (See Table I). These scores, which are considered more representative of classification performance than simple classification accuracy, show that our features lend themselves to effective heart sound analysis via machine learning.

IV. CONCLUSION

We have shown in the current research that heart murmurs classification can take advantage of machine learning when relevant heart sound and murmur features are developed. With carefully extracted heart sound features, supervised machine learning can be effectively utilized to classify heart murmurs. Additionally, through machine learning, we can explore and gain knowledge about the hidden relationships between different heart sound features and heart murmurs that are imperceptible to the human ear. This knowledge, in turn, allows us to adapt and improve heart sound and murmur classification. Our study, with the advantage of our added nontraditional heart sound features, has demonstrated a classification accuracy that consistently exceeds 90% in two different machine learning classifiers. The achieved results using our small heart sound dataset can be improved upon and

expanded to cover additional types of abnormal heart sounds in the future, when a larger dataset of heart murmur signals is available. Besides the need for a larger dataset, we maintain that relevant features are crucial to the success of classification; more useful heart sound and murmur features are necessary to improve our machine learning classification. The results of our study, along with this proposed further work, lead towards a promising future of reliable automatic cardiac auscultation using machine learning.

REFERENCES

- [1] Noncommunicable Diseases and Mental Health. *World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals*, (pp. 12). Geneva World Health Organization.
- [2] A. Gharehbaghi, M. Borgia, B. Sjöber, & P. Ask. "A Novel Method for Discrimination between Innocent and Pathological Heart Murmurs," *Medical Engineering and Physics*, 2015, 37(7), 674-682.
- [3] S. Szot, A. Levin, A. Ragazzi, and T. Ning. "A Wireless Digital Stethoscope Design," In Proc. 14th IEEE International Conference on Signal Processing, August 2018, pp.74-78.
- [4] T. Ning and J. D. Bronzino, "Autoregressive and bispectral analysis techniques: EEG applications," *Special Issue on Biomedical Signal Processing for IEEE Engineering in Medicine and Biology Magazine*, Vol. 9, No. 1, pp.47-50, March 1990.
- [5] A. Hamidah, R. Saputra, T. L. R. Mengko, R. Mengko, and B. Anggoro. "Effective Heart Sounds Detection Method Based on Signal's Characteristics," *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2016, 1-4.
- [6] T. Ning and K. S. Hsieh. "Automatic Heart Sounds Detection and Systolic Murmur Characterization Using Wavelet Transform and AR Modeling," *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013: 2555-558.
- [7] K. Chen, A. Mudvari, F. G. G. Barrera, L. Cheng, and T. Ning, "Heart Murmurs Clustering Using Machine Learning," In *Proc. 14th IEEE International Conference on Signal Processing*, August 2018, pp. 94-98.
- [8] P. R. Hegde, M. M. Shenoy, and B.H. Shekar. "Comparison of Machine Learning Algorithms for Skin Disease Classification Using Color and Texture Features," *Advances in Computing, Communications and Informatics (ICACCI), 2018 International Conference on*, 2018, 1825-828.
- [9] K. Bhatia, S. Arora, and R. Tomar. "Diagnosis of Diabetic Retinopathy Using Machine Learning Classification Algorithm," *Next Generation Computing Technologies (NGCT), 2nd International Conference on*, 2016, 347-51.
- [10] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari. "Breast Cancer Classification Using Machine Learning," *Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 2018, 1-4.
- [11] M. I. Jordan and T. M. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science*, 2015, 349(6245), 255-260.
- [12] S. Mangione and L. Z. Nieman. "Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency." *Jama*, 1997, 278(9), 717-722.
- [13] J. R. T. C. Roelandt. "The decline of our physical examination skills: is echocardiography to blame?" *European Heart Journal—Cardiovascular Imaging*, 2014, 15(3), 249-252.