Heart Murmur Classification using Machine Learning for Diagnostic Applications

Ashley FitzGerald '24
Professor Taikang Ning, PhD
Trinity College, Department of Engineering

PROBLEM DEFINITION

Valvular heart disease (VHD) is defined as irregular functioning of the atrial, pulmonary, tricuspid, or mitral valves. Structural damage is categorized as either stenosis (narrowing of the valve) or regurgitation (backward flow of blood). The disruption in blood flow causes subtle differences in the sound produced by the contraction of the heart.

In the US, 11.6 million people have VHD and there are 25,000 deaths each year. [1]

To diagnose VHD, medical professionals rely on audibly detecting heart murmurs with a stethoscope. However, this is subjective due to differences in auditory sensitivity, varying levels of experience and attentiveness among clinicians, fluctuations in environmental noise levels, and equipment quality. Moreover, the normal frequency range of heart sounds is 50 to 500 Hz, while typical human auditory sensitivity spans from 500 to 4000 Hz [2, 3]. As a result, the stethoscope has proven to have a limited specificity of only 44% when used by licensed medical professionals [4].

OBJECTIVE

This project aims to develop a signal processing and machine learning program for automatic evaluation of heart sounds to detect murmurs and provide diagnostic information.

DESIGN REQUIREMENTS

CONSTRAINTS

Data that has already been reviewed and diagnosed by a cardiologist is necessary for machine learning. The amount of data available through open-access online for heart sound murmur classification research is limited.

Database	File Count	Information Provided		Demographics
2022 CirCor DigiScope	5272	Normal/Abnormal	Pitch, shape, duration, quality	Adolescents
2016 PhysioNet Challenge	2435	Normal/Abnormal		Not Provided
Michigan Heart Sound and Murmur Library	919			
Cardiac Auscultation of Heart Murmurs				
PASCAL				

The 2022 CirCor DigiScope database is the first to have pitch, shape, duration, and quality – information used for diagnosis. However, the database only contains adolescent heart sounds. To create a representative program, more signals across age groups (adults, elderly) with diagnostic information is necessary.

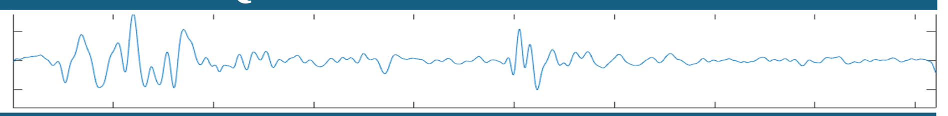
STANDARDS

FDA ISO 13485 MDSAP MEDICAL DEVICE SINGLE AUDIT PROGRAM IEEE

DESIGN METHODS

DATA ACQUISTION

CirCor DigiScope
2016 PhysioNet



PREPROCESSING

All signals were processed in MATLAB. This began with filtering, followed by an average magnitude index (AMI) function to determine the location of the heart sounds (S1, S2). This allowed for segmenting the entire signal into periods (systolic, diastolic). To prepare for feature extraction, one cardiac cycle corresponded to one data entry. The following graphs show the AMI and the detected location of S1 and S2.

FEATURE EXTRACTION

For each cardiac cycle, features were extracted at the early, mid, and late periods of both systole and diastole. A total of 163 features were tested across the following categories: time-domain statistics, periodogram, autoregressive model, AMI, and spectral kurtosis.

MACHINE LEARNING

Artificial intelligence (AI) is an umbrella term for algorithms designed for complex pattern recognition. Three of the major categories include deep learning, machine learning, and generative models. Machine learning is used largely for sorting data. The two methods are unsupervised (grouping) and supervised (classifying). Supervised is the ideal method for this project since the objective is to increase the accuracy of pattern recognition in heart sound signals to match the characteristics to a diagnosis.

DESIGN EVALUATION

MODEL SELECTION

Each supervised model considered was tested across the entire 18 classifications. The performance was averaged and then compared. The XGBoost model is the most efficient for this data set.

PERFORMANCE

The performance of single-class models was analyzed for selecting a model and during the development of the feature extraction program. A confusion matrix analysis allowed for comparing the accuracy of each individual classification to determine which were scoring lower so that new features could be derived. Moreover, a feature importance algorithm was used to quantify the weighted contribution of each feature for each classification's accuracy.

ITERATIVE PROCESS EXAMPLE

The harsh systolic and blowing diastolic classes had low confusion matrix results at 68.7% and 74.1%, respectively. From the feature importance algorithm, there were not any features with high (>0.5) weights, meaning that the model was not able identify patterns that strongly correlate to those categories. Through a reanalysis of the signals between those two classes, new features were derived with the hypotheses that they could assist differentiate the classes. After implementing the new features, the harsh classification increased by 18.7% and the blowing classification increased by 16.0%. An analysis of the feature importance results after the adjustment confirmed that the machine learning model identified this new feature as significant.

FINAL MODEL ACCURACY

Murmur Detection
97.56%

Systolic
96.60%

Pitch			Sound		
Low	Medium	High	Harsh	Musical	Blowing
98.07	100	93.73	90.95	95.83	88.10

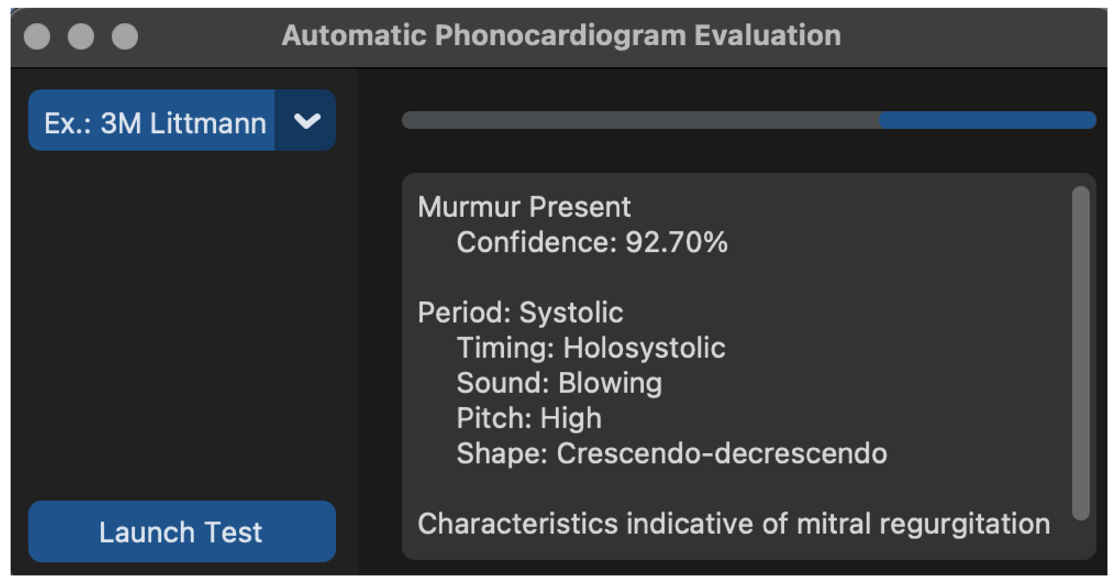
Shape		Duration		
Plateau	Crescendo-decrescendo	Holosystolic	Early	Mid
98.09	91.37	90.01	86.00	84.91

Diastolic
92.68%

Pitch	Sound		Timing
High	Harsh	Blowing	Early
100	100	99.60	99.6

DISCUSSION

FINAL DESIGN OF CLINICAL SOFTWARE PROGRAM



CONCLUSIONS

The implementation of the algorithm designed in this project into a clinical setting can be done through a software program that connects to a Bluetooth-enabled stethoscope. This would allow for a physician to have the option of using a conventual stethoscope or using the automatic evaluation. After a Bluetooth stethoscope is connected, the test is initiated, the heart sound is received, and then the program extracts the features and evaluates them on the machine learning model. The results are printed out if a murmur is detected, and if so, the probability that it is correct with the murmur classifications identified by the program. Then based on the characteristics, the program can provide a probable diagnosis.

To continue this project and improve it, more data is required. Certain murmurs require other classifications that were not accessible, such as the presence of an extra heart sound. Furthermore, certain classifications that were accessible require a higher count of recordings, such as diastolic murmurs. Diastolic murmurs accounted for only 2% of the murmurs in the entire data set. To make this project more applicable, the distribution of data must be improved.

CONTACT

Ashley FitzGerald '24
Student, Trinity College, Department of Engineering
ashley.fitzgerald@trincoll.edu
Taikang Ning
Professor, Trinity College, Department of Engineering
taikang.ning@trincoll.edu

REFERENCES

[1] Centers for Disease Control and Prevention. Valvular Heart Disease
[2] J. Oliviera, D. Noguiera, F. Renna, C. Ferriera, A.M. Jorge, and M. Coimbra, "Do we really need a segmentation step in heart sound classification alogirthms?," *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, pp. 286-289, Nov. 2021.
[3] R. A. Dobie and S. Van Hemel, Eds. *Hearing Loss: Determining Eligibility fot Social Security Benefits*. Washington, DC, USA: National Academies Press, 2002.
[4] C. Howell, A. Kaelin, B. Pennington, J. Sillaman, M. Bonk, and V. Rogers, "Novice accuracy in ultrasound and stethoscope study (NAUSS)," *J. Physician Assist. Educ.*, vol. 33, no. 3, pp. 229-233. Sep. 2022.