7-2021

Extraction and Assessment of Diagnosis-Relevant Features for Heart Murmur Classification [post-print]

Alisa Levin

Anthony Ragazzi

Skyler Szot

Taikang Ning
Trinity College, Hartford Connecticut, taikang.ning@trincoll.edu

Extraction and Assessment of Diagnosis-Relevant Features for Heart Murmur Classification

Alisa D. Levin[a,*], Anthony Ragazzi[a], Skyler L. Szot[a], Taikang Ning[a],

[a]*Department of Engineering, Trinity College*

Abstract

This paper presents a heart murmur detection and classification approach via machine learning. We extracted heart sound and murmur features that are of diagnostic importance and developed additional 16 features that are not perceivable by human ears but are valuable to improve murmur classification accuracy. We examined and compared the classification performance of supervised machine learning with k-nearest neighbor (KNN) and support vector machine (SVM) algorithms. We put together a test repertoire having more than 450 heart sound and murmur episodes to evaluate the performance of murmur classification using cross-validation of 80-20 and 90-10 splits. As clearly demonstrated in our evaluation, the specific set of features chosen in our study resulted in accurate classification consistently exceeding 90% for both classifiers.

Keywords: heart sounds, heart murmurs, classification, supervised machine learning

1. 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

*Corresponding author

Email addresses: alisa.levin@trincoll.edu (Alisa D. Levin), anthony.ragazzi@trincoll.edu (Anthony Ragazzi), skyler.szot@trincoll.edu (Skyler L. Szot), taikang.ning@trincoll.edu (Taikang Ning)

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 [2][3]. 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 [4].

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 [5], 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 [6]. 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 [6]. 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 [7]-[9]. One study, for example, proposed a feature extraction and support vector regression model to classify EEG spectral activity [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 [11]. 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.

There have been recent efforts to implement feature extraction and machine learning classification algorithms for automatic cardiac auscultation. One study presents a decision tree approach to classify heart sounds from the PhysioNet Computing in Cardiology (CinC) Challenge 2016 dataset as either normal or abnormal [12]. Another study proposes a pre-trained image classification convolutional neural network (CNN) approach to classify heart sounds from the same PhysioNet dataset into normal or abnormal [13]. Expanding from binary classification, another study presents a feature extraction and cardiac auscultation algorithm utilizing support vector machine (SVM), deep neural network (DNN) and centroid displacement based k-nearest neighbor to classify heart sounds into five different categories based on clinical diagnoses [14]. In this study, we aimed to classify heart sounds via machine learning into seven categories with a focus on expanding the feature extraction capabilities in this field.

Investigating further into trends of data-driven decision-making in healthcare, 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 [10]. In addition to common heart sound parameters described by physicians and used for automatic cardiac auscultation in previous studies [11], [15], [16], 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.

2. Material & Methods

2.1. Automatic Heart Sound Segmentation

To extract relevant features that capture essential characteristics of heart sounds and murmurs, we developed an automatic segmentation algorithm that can divide a cardiac cycle into the following segments: S1, systole, S2, and diastole. Within each identified segment, we extract useful bedside diagnosis features both in time and frequency domains. The success of a segmentation approach hinges on accurate detection of S1 and S2. We note that a cardiac cycle is delineated by its intensity variation in S1-systole-S2-diastole sequence. We developed an Average Magnitude Value (AMV) index to outline this intensity variation. To begin, an underlying heart sound signal of interest is divided into non-overlapped consecutive 10-msec slices; AMV index is calculated for each slice as follows.

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

where μ represents the mean value of the n$^{\text{th}}$ 10-mesc slice.

The choice of 10-msec duration for each slice leads to few advantages. Firstly, an underlying heart sound signal is much simplified and compactly represented by an AMV index profile that resembles intensity change of heart sound signals. Secondly, consecutive 10-msec slices can correctly capture even the short-term intensity variation in each segment of S1, systole, S2, and diastole. Slices less than 10-msec also can be used to achieve the same purpose, however, at more computation.

As seen in most phonocardiogram recordings, the first S1(lud) and the second S2(dub) heart sounds exhibit higher amplitude during a cardiac cycle. This observed high amplitude can be effectively captured by the suggested AMV index. Figure 1 exemplifies three types of heart sound signals: healthy, systolic murmur, and diastolic murmur. The AMV profiles in the middle trace of Fig.1 provide an effective means to delineate the intensity changes of original heart sound signals measured with a high sampling rate.

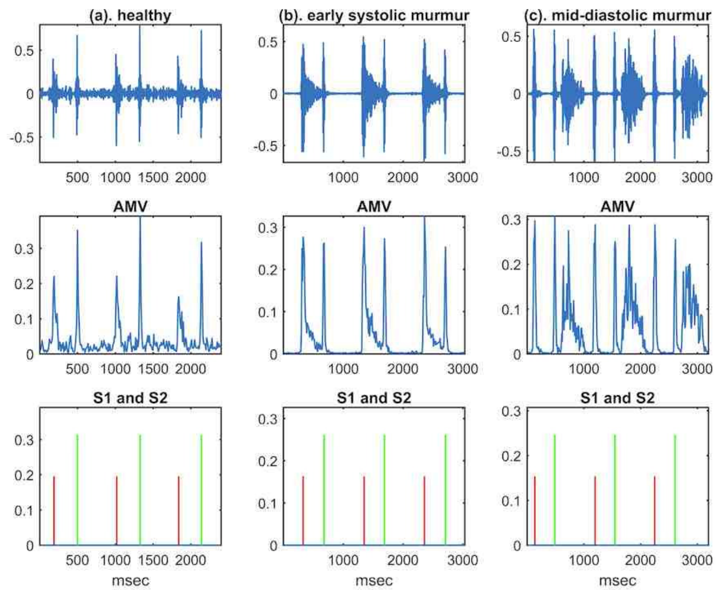

Figure 1: Three types of heart sound signals.

We have found that AMV indices are effective in identifying the possible occurrence of S1 and S2 from local maximal of AMV profile. Without loss of generality, we used the assumption that systole duration (from S1 to S2) is shorter than diastole duration (from S2 to S1). In addition, we imposed that the duration of either systole or diastole must satisfy prescribed range limits. With imposed thresholds and prescribed conditions, a detected local maximum in AMV profile could be counted as either S1 or S2. When S1 and S2 were identified (bottom trace in Fig.1), the boundaries of S1/S2 were determined when the AMV values dropped below 10% of S1/S2. With identified heart sounds, the systole segment is the interval from the boundary of S1 to the boundary of S2, and the diastole segment is the interval from the boundary of S2 to the boundary of S1 of the next cardiac cycle. Without loss of generality, heart sound episodes used in our study all began with systole and followed by diastole.

2.2. *Feature Extraction*

The features of heart sounds and murmurs previously described are essential to machine learning. To extract relevant features that capture essential characteristics of heart sounds and murmurs in both the time and frequency domains, we adopted a slice by slice analysis approach, where essential features were extracted from each 10-msec slice to provide a profile description cardiac cycles indicating short term variations.

We have found that AMV indices not only are effective in identifying S1 and S2 occurrence, they also are good indicators of detecting murmurs with additional conditions. For example, murmur is called when a sufficient number of consecutive AMV indices in the systole segment and/or diastole segment exhibiting large values above specified thresholds. More specifically, the existence of a murmur in systole and/or diastole was marked when continuous 10-msec segments showed an AMV 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 described separately using a similar procedure. Once detected, the murmur onset and duration time were recorded. In addition to time domain features thus derived from AMV index, we also extracted the average murmur pitch in the frequency domain. The murmur pitch was efficiently estimated using a second-order linear prediction AR model [17].

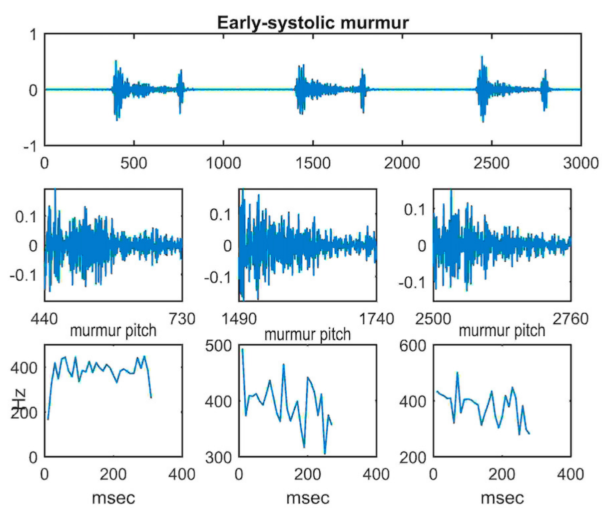

Figure 2: Extracted early systolic murmur and pitch frequency.

We used the forward-backward prediction AR model (2) and computed the optimal AR model coefficients with the least sum of squared prediction errors. The forward and backward prediction errors are given, respectively, below:

$$\begin{aligned} e^f &= x(k) - a_1 x(k-1) - a_2 x(k-2) \\ e^b &= x(k) - a_1 x(k+1) - a_2 x(k-2) \end{aligned} \tag{2}$$

The optimal AR model coefficients $\{a_1, a_2\}$ are estimated by minimizing the sum of squared forward and backward prediction errors

$$min \sum_k e^f(k)^2 + e^b(k)^2 \tag{3}$$

The AR coefficients can be effectively used to capture the murmur pitch frequency [17] by the following

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

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

3. Calculation

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 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 there were differences in the silhouettes of each heart sound signal. For example, the silhouettes of systole and diastole of healthy heart sounds are flat and constant. If there is a murmur present, on the other hand, the silhouette is sloped. To capture this, we extracted features based on the derivatives of both systole and diastole silhouettes. In total, we observed and extracted an additional 16 heart sound and murmur features in this study. These features provide useful signatures for classification and are shown below:

1. Ratio of average systole amplitude to average S1 amplitude
2. Ratio of average diastole amplitude to average S1 amplitude
3. Theorized presence of systolic murmur, determined by whether Feature 1 crosses an empirically determined threshold
4. Theorized presence of diastolic murmur, determined by whether Feature 2 crosses an empirically determined threshold
5. Sum of the absolute values of the derivatives of every point in the systole silhouette
6. Sum of the absolute values of the derivatives of every point in the diastole silhouette
7. Number of peaks in the systole
8. Number of peaks in the diastole
9. Variance in the time between peaks within the systole
10. Variance in the time between peaks within the diastole
11. Number of peaks in systole below a threshold 1a, which is 10% of the amplitude of S1
12. Number of peaks in systole above threshold 2a, which is 40% of the amplitude of S1
13. Number of peaks in systole in between threshold 1a and threshold 2a
14. Number of peaks in diastole below threshold 1b, which is 10% of the amplitude of S2
15. Number of peaks in diastole above threshold 2b, which is 40% of the amplitude of S2
16. Number of peaks in diastole in between threshold 1b and threshold 2b

3.1. 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 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.

3.2. Gini Gain

We utilized Gini Gain, which provides a quantification of which features provide the most information about the classification, to eliminate potentially redundant or unimportant features. First, we trained eight decision tree models with depths of 3–10 on our entire feature set, and identified the five most important features for each of them. Finally, we aggregated all of the unique features among the top five lists of all of our models, resulting in a total of eleven features: #1, 3, 5, 6, 7, 8, 9, 11, 12, and 14 from the feature list above, as well as the measured onset time of a diastolic murmur. We again applied the KNN and SVM classification algorithms to evaluate the efficacy of this refined feature set.

4. Results & Discussion

We have completed extensive tests to examine classification accuracy under changed conditions. For example, 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.

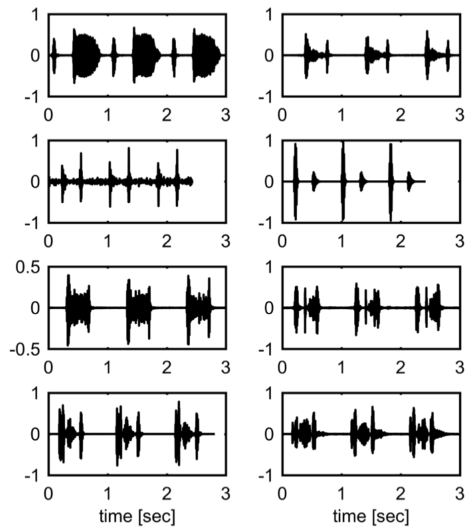

Figure 3: Different heart sound and murmur examples.

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 94.73% with a 90–10 split (See Table I). The fact that the classification accuracy scores of KNN and SVM increased by only about 1.5% and 2.9% respectively with a 90–10 split indicates that the models are not sensitive due to our robust features.

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	94.73	0.917	0.923

Table 1: Classification Results

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 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 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 important to note that in our test signals we grouped them into six common types of murmurs and normal heart sounds without murmurs—without designating additional labels for noted abnormalities, such as heart sounds with clicks, splitted S1 or S2, etc. These additional abnormalities are likely to contribute to murmur misclassification described above.

In spite of prescribed factors for potential misclassification, our method performed accurately 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 0.917 and an average recall of 0.923 (See Table I). Table II features an example of precision and recall scores of each heart sound category for our best performing model, SVM with a 90–10 split, which are then used to calculate the model's average precision and recall in 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.

Heart Sound Type	**Precision**	**Recall**
normal	0.949	0.987
early systolic	0.846	0.1.000
mid systolic	0.918	0.918
late systolic	0.957	0.880
holosytolic	0.857	0.842
diastolic	0.933	0.913
systolic & diastolic	0.960	0.923

Table 2: SVM 90–10 Split Precision & Recall

Finally, we examined the accuracy of these same models when trained on our refined feature set of only the top 11 features that emerged in our Gini Gain analysis. With a 90–10 split, we found that our KNN classifier had an average precision of 0.907 and an average recall of 0.872. Our SVM classifier, even better, had an average precision of 0.927 and an average recall of 0.928 (See Table III). We found that Gini Gain was a useful tool for eliminating unimportant or potentially redundant features. In fact, this smaller feature set allowed us to maintain the classification accuracy of our KNN model, and even slightly improved the accuracy of our SVM model (2% for the 80–20 split.) Although we successfully trained our models on 26 features initially (See Table I), we were able to refine our feature set to only 11 features with a marginal difference in classification performance. Interestingly, 10 out of these 11 top features (marked with * in the feature list) were derived from the additional 16 features that we introduced above. This demonstrates the merit of our additional heart sound features for the detection and classification of heart murmurs.

Model	Train/Test Split	Accuracy (%) 10-fold cross-val.	Average Precision	Average Recall
KNN	80–20	89.01	0.890	0.851
	90–10	90.33	0.907	0.872
SVM	80–20	94.18	0.957	0.958
	90–10	95.27	0.927	0.928

Table 3: Classification Results (refined feature set using Gini Gain)

5. Conclusions

We have shown in this study that effective and accurate heart murmur classification is achievable by taking advantage of supervised classification via machine learning. Accurate classification is possible if relevant heart sound and murmur features are extracted and adopted in model training. We developed an easy and effective automatic segmentation method to divide an underlying signal into S1, systole, S2, and diastole segments. Each segment is represented by consecutive slices of time and frequency domain features proposed in our study. The information value embodied in these features were examined using Gini Gain to assess influence on supervised classification. This knowledge, in turn, allowed us to expedite and improve heart sound and murmur classification. Our study has shown that classification accuracy that consistently exceeds 90% in KNN and SVM machine learning classifiers. Encouraged by the promising progress, we are making an effort to expand the satisfactory results from our current study to larger heart sound datasets and to encompass additional types of abnormal heart sounds using machine learning. While powerful machine learning algorithms are emerging, we maintain that suitable features essential to heart sound and murmur behaviors are crucial to achieve satisfactory classification. The results of our study shed light on a promising future of dependable automatic cardiac auscultation via machine learning by introducing information-rich features.

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] 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.

[3] 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.

[4] A. Gharehbaghi, M. Borga, B. Sjöber, & P. Ask. ”A Novel Method for Discrimination between Innocent and Pathological Heart Murmurs,” *Medical Engineering and Physics*, 2015, *37*(7), 674-682.

[5] 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.

[6] M. I. Jordan and T. M. Mitchell. “Machine learning: Trends, perspectives, and prospects.” *Science*, 2015, *349*(6245), 255-260.

[7] 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.

[8] 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.

[9] 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.

[10] W. L. Lim, O. Sourina, and L. P. Wang. “STEW: Simultaneous Task EEG Workload Dataset,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.26, no.11, 2018, 2106-2114.

[11] 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.

[12] V. Arora, R. Leekha, R. Singh, and I. Chana. "Heart sound classification using machine learning and phonocardiogram," *Modern Physics Letters B*, *33*(26), 2019, 1950321.

[13] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller. "Learning Image-based Representations for Heart Sound Classification," *Proceedings of the 2018 International Conference on Digital Health*, 2018.

[14] Yaseen, G. Son, and S. Kwon. "Classification of Heart Sound Signal Using Multiple Features," *Applied Sciences*, *8*(12), 2018, 2344.

[15] 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.

[16] 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.

[17] 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.