Screening of Heart Sounds using Hidden Markov and Gammatone Filterbank Models

Ben Alexander, Gabriel Nallathambi, Nandakumar Selvaraj
VitalConnect Inc.,
San Jose, CA 95110, USA
email: balexander@vitalconnect.com, gnallathambi@vitalconnect.com, nselvaraj@vitalconnect.com

***Abstract* — Objective classification of heart sound signals can provide significant advancements in the screening of structural heart abnormalities. An algorithm, based on auditory filter models and probabilistic segmentation of heart cycle sequences, is proposed for classifying phonocardiogram (PCG) signals as normal or abnormal. It first involves segmentation of cardiac recordings into a sequence of four heart stages, namely S1, systole, S2 and diastole, using a hidden Markov model approach. Secondly, gammatone frequency cepstral coefficient (GFCC) features are extracted by applying the gammatone filterbank model to two independent binary classification problems: one with PCG segmentation, and one without segmentation. Weighted support vector machine models are trained to classify the PCG signals as normal vs. abnormal records using the GFCC features. The algorithm is trained and cross-validated using the 2016 PhysioNet Computing in Cardiology Challenge database of 3,240 PCG recordings. Based on 10-fold stratified cross-validation, the performance of the proposed "with segmentation" approach is determined to have a sensitivity of 90.3% and a specificity of 89.9%, while the "without segmentation" approach shows a comparable performance of 87.1% sensitivity and 88.5% specificity. Thus, the proposed algorithm demonstrates strong, clinically acceptable performance for automated screening of heart sound signals, and it can be a useful diagnostic tool in clinical practice to screen patients for structural and functional heart diseases.**

Keywords — Auditory model, gammatone filter bank, hidden Markov model, heart sound classification, phonocardiography.

I. INTRODUCTION

Heart auscultation, the technique of listening to the internal sounds of the heart using a stethoscope, is one of the fundamental diagnostic tools of the physician. The heart sounds, produced predominantly by the closure of the atrioventricular and semilunar heart valves, provide doctors with a glimpse into the structural and functional health of the heart and allow for an initial cardiac diagnosis. However, the traditional stethoscope as a spot check diagnostic tool is reliant on the physician's skill and experience, and is inherently limited by the human threshold of audibility in perceiving sounds of only certain magnitude and bandwidth [1]. Also, with traditional stethoscopes, recording sound data for future reference is not possible.

Advances in objective heart sound analysis using electronic recordings are crucial, as more than 5 million people in the U.S. are diagnosed with heart valve disease every year [2]. With an aging population, acquired structural heart disease, which includes diseases such as calcific aortic valve stenosis and mitral regurgitation, is now the third most common cardiovascular disease in the country [3]. In fact, there are now about 1.5 million adult patients with aortic stenosis in the U.S. [4]. Also, congenital heart disease, the most common condition involving structural heart defects at birth, affects more than 71,000 children in the U.S. and European Union every year [5]–[7]. Effective medical management and planning of surgical treatment for structural or congenital heart disease may necessitate objective and reliable monitoring of various functions of the heart, including heart sounds.

With the advent of digital stethoscopes [8], [9], clinicians are now able to record heart sounds as digital waveforms and examine them visually, beyond relying solely on the human perception of sound. The heart sound signals, known as phonocardiograms (PCG), are recorded by microphones typically placed at one of the common heart auscultation sites including the aortic, pulmonic, tricuspid, and mitral areas on the chest [10]. The PCG recordings are mainly comprised of the fundamental heart sounds, S1 and S2, commonly referred to as "lub" and "dub" sounds, respectively. In addition to these normal heart sounds, relatively weak S3 and S4 heart sounds in individuals at risk of heart failure, and other abnormal sounds such as opening snaps, ejection sounds, murmurs, and mid-systolic clicks, may also be present [10]–[13]. The application of digital signal processing and machine learning techniques to these PCG signals can overcome the limitations of manual auscultation for effective clinical diagnosis using heart sounds. However, automated classification of heart sounds is not trivial as PCGs can be obscured by background noise, breathing sounds, intestinal sounds, and sensor motion against the body [10]. Continuous monitoring of heart sounds using novel wearable medical device sensors with sophisticated algorithms may emerge for screening or diagnosing abnormal heart conditions/defects that would allow for early and appropriate medical intervention.

Automated classification of heart sound pathologies has been evolving over the last decade using various feature extractions involving signal amplitudes, frequencies, wavelets, time-frequency spectral measures, and statistical methods in conjunction with classifier methods such as support vector machines (SVM), artificial neural networks, hidden Markov models (HMM), and clustering approaches. Furthermore, segmentation methods based on envelopes, features, machine learning and HMM have also been used to locate S1, systolic, S2 and diastolic states of each cardiac cycle and subsequently classify the heart sounds using localized features from these segments. However, accurate heart sound segmentation and classification into normal versus abnormal classes using PCG signals recorded in hospital or in-home environments remains a challenge [10].

Auditory filter models are designed to imitate the way the human ear filters sound by using an array of independent bandpass filters to effectively recognize sound signals, and thereby discriminates between sounds more closely at low frequencies than at high frequencies. This ability to discern changes more closely at low frequencies is a useful feature for PCG analysis, since the spectral power of most heart sounds primarily occur at low frequencies, particularly under 200 Hz [14]. While auditory filter models are successfully applied in speech analysis [15], they are relatively unexplored in heart

sound classification. Previous studies involving auditory filter models for PCG analysis include noise detection [16] and assessment of the patent ductus arteriosis [17].

The feasibility of applying auditory filter models for classification of heart sounds is explored in this study. More specifically, the heart sounds are segmented using an HMM-based approach into the four heart stages (S1, systole, S2, diastole) for each heartbeat contained in a PCG recording of arbitrary duration. For each heart stage, gammatone frequency cepstral coefficients (GFCC) are extracted as features using the gammatone filterbank, mimicking the impulse response of the auditory nerve [18]. PCG recordings are then classified as normal or abnormal using an SVM. Finally, the classification performances of the proposed GFCC method are studied, and compared with conventional features. The rest of the paper is organized as follows. Section II describes details about the PCG database and the proposed method, which involves heart sound segmentation, GFCC feature extraction, classification, and performance analysis. Section III provides the performance results of the proposed approaches, and in Section IV, the implications of the present work are discussed.

II. Materials and Methods

A. Heart sound database

Automated classification of PCG recordings has been extensively studied to aid the clinical assessment of pathological cardiac conditions. The literature on this topic has wide coverage relying on classical feature extractions including time domain, frequency domain, and wavelet methods [10]. Recently, to facilitate development of clinical grade algorithms for classifying heart sounds, the 2016 PhysioNet Computing in Cardiology Challenge provided a substantially large standardized database by aggregating recordings from multiple smaller databases collected independently in both clinical and non-clinical settings [10]. The training data contains 3,240 PCG recordings generally recorded at four common locations on the chest: the aortic area, the pulmonic area, the tricuspid area, and the mitral area. Each recording in the database is annotated as "normal" or "abnormal" by experts. The database contains 665 abnormal records, which are mostly from patients with heart valve defects such as mitral valve prolapse, mitral regurgitation, aortic regurgitation, aortic stenosis, and valvular surgery or coronary artery disease [10]. This PCG database from the 2016 PhysioNet Challenge was utilized in this paper to validate the proposed method.

Briefly, the proposed algorithm for classifying heart sounds of arbitrary duration into normal and abnormal classes involves preprocessing the PCG signal to extract the spectral contents of interest, segmenting heart sounds into the different phases of the cardiac cycle, extracting features using the auditory filter model, and training prediction models using a weighted SVM with radial basis function (RBF) kernel. Subsequent sections provide a step-by-step detailed description of the algorithmic processes.

B. PCG preprocessing and segmentation

The PCG signals, which originally have a 2000 Hz sampling rate, are filtered using a 2^{nd} order Butterworth bandpass filter between 25 and 400 Hz, and resampled uniformly at 1000 Hz. The preprocessed signals are then segmented using an HMM-based algorithm into the four phases of the heart cycle, namely S1 (first heart sound), systole, S2 (the second heart sound) and diastole, as a sequence. Accurate delineation of heart sounds into the four phases of the heart cycle is highly useful for extracting physiologically relevant features from PCG recordings [9].

An HMM-based method proposed by Schmidt et al. [19] that explicitly incorporates the expected duration of heart phases in the model is employed for the segmentation of each cardiac cycle. Schmidt et al. [19] generate a Gaussian distribution for the expected duration of each of the four cardiac cycle stages using the average duration of the heart sounds as well as autocorrelation analysis of systolic and diastolic durations. Then, using annotated data labeled by experts, the model is trained to derive the emission probabilities of the HMM. Specifically, the true states $Q = \{q_1, q_2, \ldots, q_t\}$ are unknown, but the observed sequence O is known. Therefore, the HMM-based segmentation method is tasked with finding the state sequence that is most likely to produce the observed state [19]:

$$Q^* = argmax_Q\, P(Q|O, \lambda) \qquad (1)$$

where Q^* is the state sequence that is most likely to produce O, the observed state sequence, and λ denotes the model parameters such as the HMM's transition matrix and the cardiac cycle duration distribution. Inclusion of expected durations as prior information resulted in a sensitivity of 99.3% and a positive predictive value of 99.1%, when tested on a population of 60 patients for a total of 744 S1 and S2 occurrences [19].

C. Auditory filter model

Advances in the understanding of cochlear nonlinearities and the physiology of hearing have led to the development of auditory filters [18]. With the goal of imitating the human ear to help with this binary classification problem, a gammatone filter is implemented using an impulse response, $h(t)$, given by the product of a gamma distribution and sinusoidal tone as

$$h(t) = at^{n-1}e^{-2\pi bt}cos(2\pi f_c t + \phi) \qquad (2)$$

where f_c is the center frequency in Hz, ϕ is the phase of the carrier in radians, a is the amplitude gain, n is the order of the filter, b is the filter bandwidth in Hz, and t is the time in seconds. An array of overlapping gammatone filters are organized into a filterbank to mimic the human auditory system.

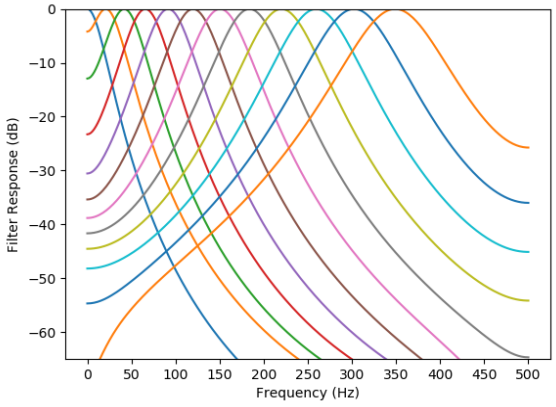

Figure 1. Frequency response of a 12–channel gammatone filterbank.

We developed an $M-$ channel filterbank based on Slaney's implementation of the gammatone filterbank [20]. A 4^{th} order gammatone filter is used currently to simulate the

auditory response. The filters are spaced on the equivalent rectangular bandwidth (ERB) scale, which places them closer together at low frequencies and farther apart at high frequencies, with each filter representing a single frequency sub-band of the original signal. The bandwidth, b, of the gammatone filter is given by $b = 1.019 * ERB(f_c)$, where $ERB(f) = 24.7 + 0.108f$. The center frequency, f_c for a channel m is given by

$$f_c(m) = (f_h + K)e^{\frac{m}{M} log\left(\frac{f_l+K}{f_h+K}\right)} - K \quad (3)$$

where $1 \leq m \leq M$, M is the total number of channels, $K = 228.83$, and f_l and f_h are the lowest and highest cutoff frequencies of the filterbank, respectively.

Figure 1 shows the frequency response of a 12-channel gammatone filterbank based on Slaney's implementation [20]. Next, gammatone filterbanks and cepstral coefficients are combined to compute GFCCs that are related to the spectral envelope on a perceptual auditory scale.

D. Feature extraction

The steps involved in the extraction of GFCCs are as follows. First, the PCG signal, $x_i(n)$, is divided into i frames of 25ms duration with 40 percent overlap, assuming the audio frequencies are relatively stationary on these time frames. A Hanning window, $w(n)$, of length N is also applied to reduce the spectral leakage, as given below. Secondly, the power spectrum in each frame, $P_i(k)$, is calculated to estimate the amount of energy at different frequencies within the signal, similar to the functioning of the cochlea.

$$P_i(k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x_i(n)w(n)e^{-j2\pi n\frac{k}{N}}\right|^2 \quad (4)$$

where $w(n) = 0.5 - 0.5\cos\left(\frac{2\pi n}{N-1}\right)$ and $k \epsilon [0, N-1]$. Thirdly, an $M-$channel gammatone filterbank $h(m,k)$ is applied to the power spectral estimates. Taking the logarithm of the calculated sum of the energy within each gammatone filter in the filterbank, as given below, further distinguishes closely spaced frequencies, similarly to the way that the human auditory system perceives sounds.

$$g_{i,m} = log\left(\sum_{k=0}^{N-1} h(m,k)P_i(k)\right) \quad (5)$$

Finally, the discrete cosine transform (DCT) is applied to decorrelate the logarithm of the filterbank energies, and to obtain the coefficients $c_{i,j}$ corresponding to the j^{th} GFCC for the i^{th} frame.

$$c_{i,j} = \sum_{m=1}^{M} g_{i,m} cos\left(\frac{\pi j(2m-1)}{2M}\right) \quad (6)$$

Currently, features are extracted using the above GFCC method for two independent classification approaches: (i) with PCG segmentation and (ii) without PCG segmentation. In the case with PCG segmentation, the feature vector is calculated by applying statistical measures on each GFCC coefficient across the frames of each of the four heart cycle phases, as segmented by the HMM-based algorithm. The total number of features computed with heart sound segmentation is equal to $n_c * n_p * n_s$ where n_c is the number of extracted coefficients, n_p is the number of phases of the segmented heart cycle (currently, $n_p = 4$), and n_s is the number of statistical measures used. On the other hand, the feature vector without PCG segmentation is calculated by applying statistical measures on each GFCC coefficient across all frames, and the total number of features is equal to $n_c * n_s$. The statistical measures used to obtain the feature vector include the mean, standard deviation, or a combination of both, so n_s is typically 1 or 2.

E. Classifier

Training is performed independently using the feature vectors extracted for with and without segmentation approaches using various classification methods including logistic regression, SVM and gradient boosting. Based on 10-fold stratified cross-validation, the best performance is obtained using a weighted SVM classifier with a RBF kernel, defined by $e^{(-\gamma|X_i - X_j|^2)}$, for both cases of feature extraction. The hyperparameters of the classifier including the misclassification penalty, C, and the kernel coefficient for the RBF kernel, γ, are optimized as described in the next section.

The weighted SVM [21] addresses the imbalance between normal and abnormal heart sound recordings in the dataset with the cost function for a set of weights, w defined as

$$L(w) = \frac{\|w\|^2}{2} + C^+\left(\sum_{i:\, y_i=+1}^{N} \xi_i\right) + C^-\left(\sum_{i:\, y_i=-1}^{N} \xi_i\right) \quad (7)$$

where C^+ and C^- are the regularization or penalty terms for the abnormal and normal heart sound classes, respectively. The weighted classes C^j are obtained by $C^j = C * \left(\frac{n}{kn_j}\right)$, where n is the total number of heart sound observations, k is the number of classes, and n_j is the number of observations in class j. Using different penalty terms for normal and abnormal classes addresses the class imbalance by applying a higher penalty for misclassifying an abnormal recording.

F. Hyperparameter optimization and data analysis

The hyperparameters for the feature extraction and classifier such as C, γ, M (the number of channels), and the subset of the coefficients $c_{i,j}$ are optimized by maximizing the average of sensitivity (the probability of correctly identifying an abnormal case) and specificity (the probability of correctly identifying a normal case). While it is common in speech signal processing to use more than 20 channels and the first 10 coefficients of the gammatone filterbank [22], we determined no substantial improvement in the classification performance beyond 10 channels during optimization. Hence, M and the subset of $c_{i,j}$ are optimized on a 2D grid of 10 channels x 10 coefficients independently for different combinations of statistical measures. The performances of the proposed method are assessed using a feature vector size of 80 ($n_c = 10, n_p = 4, n_s = 2$) and 20 ($n_c = 10, n_s = 2$) for "with" and "without" segmentation, respectively. The classifier hyperparameters are optimized initially using a coarse grid search as recommended in [23] and then fine-tuned on $C \epsilon [2^0, 2^4]$ and $\gamma \epsilon [2^{-4}, 2^0]$.

The classification performance is independently evaluated for the "with" and "without" segmentation approaches using specificity and sensitivity, based on 10-fold stratified cross-validation. The trade-off between specificity and sensitivity for "with" and "without" segmentation methods are also

assessed using ROC curves, computed using Platt's posterior class probabilities [24].

G. Performance assessments from comparative studies

Performance assessments using various conventional features are further carried out per the previous studies in the literature for comparisons of the current GFCC-based method. Specifically, four categories of features are selected for comparative assessments: time-domain features, frequency-domain features, a combination of wavelet and entropy-based features, and mel-frequency cepstral coefficients (MFCC). A total of 36 time-domain features as proposed by Potes et al. [25] were used in the analysis. These include the lengths of intervals such as S1, systole, S2 and diastole intervals. They also include amplitude-based measurements such as the ratio of the absolute amplitude during systole to that of the S1 period in each heartbeat, and the skewness and kurtosis of amplitudes during each of the four heart phases. A total of 60 frequency-domain features were extracted including 36 features that correspond to the power spectrum across 9 frequency bands during the 4 phases of heart cycle as proposed by Potes et al. [25], 9 features based on linear predictive coefficients and power spectral density as proposed by Zabihi et al. [26], and 15 features based on standard deviations, skewness, and kurtosis of periodograms in five equally spaced frequency frames as proposed by Kay et al. [27]. A total of 246 wavelet and entropy-based features were used, including 220 features based on the continuous wavelet transform (CWT) [27], 4 features based on the coefficients of the discrete wavelet transform (DWT) [26], 20 features based on spectral entropy [27], and 2 features based on Natural and Tsallis entropy [26]. Furthermore, MFCCs are calculated using the mel-frequency filterbank with the same number of channels and coefficients as the gammatone filterbank. To facilitate comparisons, the weighted SVM is trained and optimized independently for different categories of features outlined above.

III. Results

Figure 2 shows an example of a PCG signal from the database segmented into four cardiac cycle stages: S1, systole (Sys), S2, and diastole (Dia). The gammatone spectrogram shows the distribution of the signal energy across different frequency sub-bands that are transformed into GFCC features for further classification analysis.

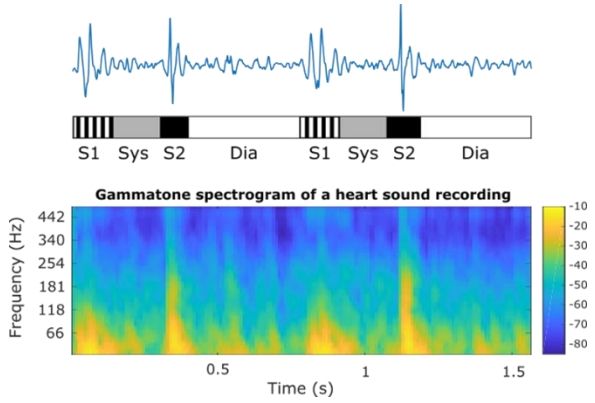

Figure 2. Example of heart sound segmentation and gammatone spectral representation.

A. Classification performance with PCG segmentation

Figure 3 demonstrates the binary classifier performance as a function of the number of filters in the gammatone filterbank, and the number of coefficients (GFCCs) calculated from the PCG recordings. As the number of coefficients increases, performance also increases drastically up to the first 4 coefficients. Thereafter, the performance plateaus and does not result in substantial improvement with further inclusion of additional coefficients or filters. A feature set of 16, corresponding to the first 4 coefficients from each heart phase averaged across frames using an 8-channel filterbank, and setting $C = 2$ and $\gamma = 0.1$ in the SVM, offered the best performance of 90.3% sensitivity and 89.9% specificity.

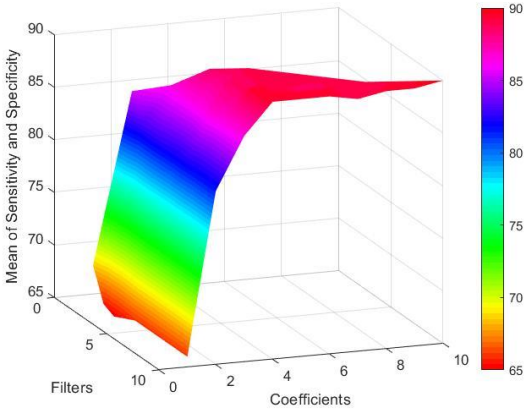

Figure 3. The binary PCG classifier performance versus the number of channels (filters) and coefficients of the gammatone filterbank.

B. Classification performance without PCG segmentation

Similarly, the performance of the proposed method without segmentation offered a performance of 87.1% sensitivity and 88.5% specificity with 16 features corresponding to the mean and standard deviation of the eight coefficients across the frames in an 8-channel gammatone filterbank, using $C = 2$ and $\gamma = 0.2$ in the SVM. Thus, the performance with segmentation is slightly better than the results without segmentation, as shown in the ROC curve below (Figure 4), which suggests that the features extracted with segmentation may outperform with relatively better predictive power. The area under the ROC curve (AUC) of the "with" segmentation approach is 0.96 as compared to 0.94 without segmentation, suggesting that segmentation of cardiac cycles is helpful for accurate classification of normal vs. abnormal heart sounds.

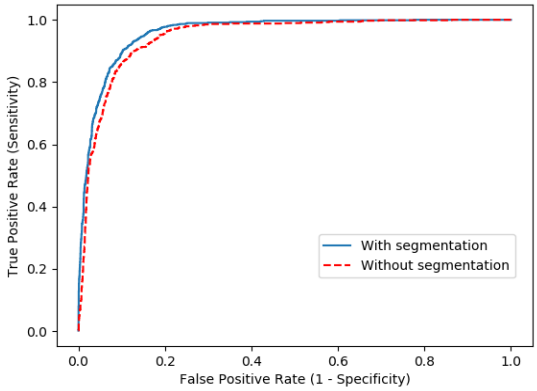

Figure 4. ROC curves for the "with" and "without" segmentation methods.

C. Comparison with conventional features

The 36 time-domain features resulted in a specificity and sensitivity of 79.3% and 77.4%, respectively. While time-domain features are relatively sensitive to noise, frequency-domain and wavelet/entropy features provide comparatively better results. The 60 frequency-domain features yielded a specificity and sensitivity of 82.7% and 85.9%, respectively. The 243 wavelet and entropy features yielded 83.3% specificity and 83.9% sensitivity. On the other hand, 16 MFCC features resulted in comparable performance of 87.9% specificity and 90.7% sensitivity, and 88.8% specificity and 87.4% sensitivity, respectively for "with" and "without" segmentation. The results reveal that the GFCC-based features relatively outperform the classical features for the classification of normal vs. abnormal heart sounds as summarized in Table I.

TABLE I. CLASSIFICATION PERFORMANCES BY FEATURE CATEGORY

Feature category	Number of features	Performance results	
		Specificity	***Sensitivity***
Proposed	16	89.9%	90.3%
Time domain	36	79.3%	77.4%
Frequency domain	60	82.7%	85.9%
Wavelet and Entropy	243	83.3%	83.9%
MFCC	16	87.9%	90.7%

IV. DISCUSSION

Screening of patients may allow for detection of structural heart disease at early stages. This provides more time for planning treatments and analyzing management options, and enables clinicians to perform timely medical intervention, thereby preventing adverse cardiac events and ultimately leading to better health outcomes. The proposed algorithm demonstrates clinically acceptable performance for identifying normal and abnormal heart sounds, indicating that an automated algorithmic solution can be useful for clinical heart sound analysis.

Traditional diagnostic screening primarily involves manual spot check examinations of heart sounds using a stethoscope, followed by other standard clinical diagnostic tests such as chest X-rays and echocardiograms. Manual auscultation relies on subjective assessments through a "process of elimination" that the doctor carries out by sequentially listening to sounds from different heart valves [1]. Most of the fundamental heart sounds begin at 3 cycles/sec and peak at 20 cycles/sec, with the weaker abnormal heart sounds such as S3 and S4 falling below the audible range of the ear, which starts at about 20 cycles/sec [1]. The inherent limitations in manual assessments reveal the necessity for an automated system such as the current algorithm that can augment clinical decision-making by objectively analyzing heart sounds.

The proposed algorithmic solution achieves good performance per its validation using the 2016 PhysioNet heart sound database. It is worthwhile to note that the sound recordings of this database are collected in uncontrolled environments and are corrupted by various noise sources such as speech, barking of dogs, stethoscope motion, breathing and intestinal activity [10]. In this context, the present performance results are encouraging and support successful deployment in real-world settings.

Although a number of methods based on envelopes, machine learning, amplitude, and frequency features [10] have been proposed in the literature for segmentation of cardiac cycles, the HMM is most effective and suitable due to the assumption of a double stochastic process, in which the heart cycle is the hidden Markov process and the heart sounds are the observable stochastic outputs [19]. However, it is to be noted that the study group for Schmidt et al. [19] did not include patients with heart valve disease. Therefore, further refinements to the segmentation method that can ensure clear separation between fundamental and abnormal heart sounds in patients with heart valve abnormalities might improve the overall classification performance.

The cross-validation results indicate that the specificity and sensitivity of the binary classifier with heart sound segmentation are relatively higher than the performance without segmentation. The best performance is obtained using just four GFCC coefficients per heart phase, a smaller feature set that can reduce computational cost, time and data storage requirements. It is worth mentioning that eliminating the segmentation from the algorithmic process further reduces the computation time substantially. Given the relatively close performance between the two approaches, further investigations may be warranted to study the trade-off between performance and computational cost for real-time implementation of the algorithm.

Performance analyses from the comparative literature studies further show that GFCC-based features possess more predictive power for classification of normal vs. abnormal heart sounds than the conventional time domain, frequency domain, wavelet and entropy features. Substituting gammatone filterbanks for mel-frequency filterbanks seems to yield relatively comparable performance. However, GFCCs are intrinsically robust to noise as compared to MFCCs, which has been demonstrated with speaker identification [22]. Moreover, tuning the noise robustness of GFCCs for heart sound classification is an open topic, and needs to be explored further. While the present algorithm exclusively uses GFCC-based features to determine its feasibility for classifying heart sounds signals, future work may aggregate the other feature categories listed in Table I, and investigate any potential improvements in performance.

Another potential path for further research could be to incorporate GFCCs into deep learning algorithms. For example, GFCC heat maps could be created and converted into images, and fed to a neural network for classification [28]. As healthcare applications require high levels of transparency and rationales in their decision making, it could be challenging for deep learning approaches, mostly regarded as black-box models, to get the essential clinical acceptance, which is an issue that may require some consideration [29].

Forty-eight research teams participated in the PhysioNet Computing in Cardiology Challenge and the top entry had a specificity and sensitivity of 77.81% and 94.24%, respectively, on a blind test set. While the public does not have access to the blind test set, the cross-validation results of the proposed algorithm using the available training data are comparable to the performances reported from the challenge.

In this paper, optimization of the algorithm is performed by using the average of sensitivity and specificity as the

cost/scoring function, in order to maintain a balance between those two measures. However, balancing the tradeoff between sensitivity and specificity for practical use cases needs to take the desired outcomes, associated costs, and prevalence of the disease condition into account. Sometimes, it might be desirable to have high sensitivity at the expense of relatively lower specificity, which would increase the probability of identifying true positives at the cost of increased false positives. In other situations, however, this might not be preferable, because the false positives may lead to false alarms, unnecessary expensive tests, and even psychological burden in certain patients. Selecting an appropriate cost function and decision making in such scenarios might also depend on the availability of medical infrastructure, funding for additional tests, and the well-being of the patient. The predictive model could be optimized appropriately to meet the specific goals and patient needs. Furthermore, the proposed method can be easily integrated into clinical decision support systems for assisting technicians and physicians at the point of care testing, which can be beneficial for patients as well as the healthcare system.

V. Conclusion

Automated heart sound analysis allows for non-invasive screening of structural and functional heart abnormalities. The present study proposes a supervised machine learning algorithm for effective and automated identification of normal vs. abnormal heart sound signals, using unique GFCC features that mimic the human auditory system. The performance of the proposed algorithm demonstrates clinically acceptable performance of ~90% for both sensitivity and specificity using 10-fold cross-validation on the 2016 PhysioNet heart sound database, while outperforming a number of other predictive models with traditional feature sets. Thus, the proposed algorithm can be a useful diagnostic tool in clinical settings to help screen for patients with structural and functional abnormalities in the heart.

References

[1] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*. 2006.

[2] V. T. Nkomo, J. M. Gardin, T. N. Skelton, J. S. Gottdiener, C. G. Scott, and M. Enriquez-Sarano, "Burden of valvular heart diseases: a population-based study.," *Lancet (London, England)*, vol. 368, no. 9540, pp. 1005–11, Sep. 2006.

[3] K. Maganti, V. H. Rigolin, M. E. Sarano, and R. O. Bonow, "Valvular Heart Disease: Diagnosis and Management," *Mayo Clin. Proc.*, vol. 85, no. 5, pp. 483–500, May 2010.

[4] B. Iung, "A prospective survey of patients with valvular heart disease in Europe: The Euro Heart Survey on Valvular Heart Disease," *Eur. Heart J.*, vol. 24, no. 13, pp. 1231–1243, Jul. 2003.

[5] J. I. E. Hoffman and S. Kaplan, "The incidence of congenital heart disease.," *J. Am. Coll. Cardiol.*, vol. 39, no. 12, pp. 1890–900, Jun. 2002.

[6] M. D. Reller, M. J. Strickland, T. Riehle-Colarusso, W. T. Mahle, and A. Correa, "Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005.," *J. Pediatr.*, vol. 153, no. 6, pp. 807–13, Dec. 2008.

[7] H. Dolk, M. Loane, and E. Garne, "Congenital Heart Defects in Europe: Prevalence and Perinatal Mortality, 2000 to 2005," *Circulation*, vol. 123, no. 8, pp. 841–849, Mar. 2011.

[8] S. Swarup and A. N. Makaryus, "Digital stethoscope: Technology update," *Med. Devices Evid. Res.*, 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," *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–37, 2015.

[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, 2016.

[11] M. Johnston, S. P. Collins, and A. B. Storrow, "The third heart sound for diagnosis of acute heart failure.," *Curr. Heart Fail. Rep.*, vol. 4, no. 3, pp. 164–8, Sep. 2007.

[12] Y. Minami *et al.*, "Third heart sound in hospitalised patients with acute heart failure: insights from the ATTEND study.," *Int. J. Clin. Pract.*, vol. 69, no. 8, pp. 820–8, Aug. 2015.

[13] S. J. Shah *et al.*, "Association of the fourth heart sound with increased left ventricular end-diastolic stiffness.," *J. Card. Fail.*, vol. 14, no. 5, pp. 431–6, Jun. 2008.

[14] V. N. Varghees and K. I. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomed. Signal Process. Control*, vol. 13, no. 1, pp. 174–188, 2014.

[15] O. Ghitza, "Auditory models and human performance in tasks related to speech coding and speech recognition," *Speech Audio Process. IEEE Trans.*, vol. 2, no. 1, pp. 115–132, 1994.

[16] N. M. T. C. E. Silva, "Phonocardiogram Noise Detection in Realistic Environments," Universidade do Porto, 2012.

[17] P. H. Sung, J. N. Wang, B. W. Chen, L. S. Jang, and J. F. Wang, "Auditory-inspired heart sound temporal analysis for patent ductus arteriosus," *2013 1st Int. Conf. Orange Technol.*, pp. 231–234, 2013.

[18] R. F. Lyon, A. G. Katsiamis, and E. M. Drakakis, "History and future of auditory filter models," *ISCAS 2010 - 2010 IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst.*, pp. 3809–3812, 2010.

[19] S. E. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model," *Comput. Cardiol.*, vol. 35, pp. 345–348, 2008.

[20] M. Slaney, "An Efficient Implementation of the Auditory Filter Bank," *Apple Comput. Tech. Rep.*, vol. 1, no. 35, p. 40, 1993.

[21] E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines : Training and Applications," 1997.

[22] X. Zhao and D. Wang, "Analyzing noise robustness of MFCC and GFCC features in speaker identification," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 7204–7208, 2013.

[23] C. C. Hsu, C. C. Chang, and C. C. Lin, "A practical guide to support vector classification," vol. 1, no. 1, pp. 1–16, 2003.

[24] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. large margin Classif.*, vol. 10, no. 3, pp. 61–74, 1999.

[25] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of Feature:based and Deep learning:based Classifiers for Detection of Abnormal Heart Sounds," pp. 621–624, 2016.

[26] M. Zabihi, A. Bahrami Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, "Heart Sound Anomaly and Quality Detection using Ensemble of Neural Networks without Segmentation," pp. 2–5, 2016.

[27] E. Kay and A. Agarwal, "DropConnected Neural Network Trained with Diverse Features for Classifying Heart Sounds," 2016, pp. 617–620.

[28] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing Abnormal Heart Sounds Using Deep Learning," 2017.

[29] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?," no. Ml, pp. 1–28, 2017.