Joint Heart Sounds Segmentation and Murmur Detection with Masked Loss Function

Tomasz Grzywalski*, Adam Maciaszek*, Riccardo Belluzzo*, Krzysztof Szarzyński*, Mateusz Piecuch*, Honorata Hafke-Dys†
*StethoMe, Poznan, Poland
{grzywalski,maciaszek,szarzynski}@stethome.com
†Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University, Poznan, Poland
h.hafke@amu.edu.pl

***Abstract*—In recent years, many approaches have been proposed for automated heart sound analysis. However, most of these algorithms separate the two steps required for an accurate detection of abnormal cardiac sounds: segmentation of stethoscope recording into individual heart cycles and detection of heart murmurs. In this work we propose a method to train a neural network to perform both of these tasks simultaneously achieving a synergy effect. Despite the fact that the method uses both types of labels for training, they don't have to be specified for all training examples. Moreover it also supports negative examples, i.e recordings with no heart sounds. The result is a single neural network model that can detect individual heartbeat cycles, segment these into the 4 heartbeat phases and predict heart murmur presence. This is achieved by using a training loss function that incorporates relations between the different output types and uses masking in case of missing labels. We evaluated our results on the popular 2016 PhysioNet/CinC Challenge dataset for heart sounds and benchmarked it with respect to three state-of-the-art heart murmur detection algorithms. Our method significantly outperforms the latter algorithms achieving, in particular, an F1-score of 83.9% – an enhancement of 7.6 percentage points over the best of the considered alternatives.**

I. Introduction

Cardiac auscultation is the first and most common examination that facilitates the identification of many heart diseases with high predictive power [1]. Despite the advent of new technological innovations like handheld ultrasound, this procedure is still very important in healthcare because it is both quick and cost-effective. Heart auscultation allows for detection of abnormal heart sounds called heart murmurs, which are symptomatic of heart problems. Heart murmurs are caused by turbulent blood flow through the heart. resulting from abnormal opening or closing of heart valves. This may lead to backward blood flow or reduced forward blood flow resulting in murmurs registered by a stethoscope. In children, abnormal murmurs are usually caused by congenital heart disease. In adults, abnormal murmurs are most often due to acquired heart valve problems. Early detection of pathological changes in the heart is clinically beneficial for patients and reduces the risk of serious complications including, in some cases, the need of undergoing risky heart operations [2].

Heart auscultation normally involves the usage of a stethoscope by a trained physician whose diagnosis is highly subjective and relies on his education and experience [3], [4] and [5]. The ability to properly identify heart sounds varies significantly among different groups of medical doctors and trainees. As reported in [6], medical students, residents, and even academic internists in many countries recognize less than 40% of abnormal heart sounds. According to [7], the main reason for such lack of proficiency is inadequate training. In fact, cardiologists, who represent only 5% of physicians in the US, are the only group that has been shown to recognize a majority of abnormal heart sounds and murmurs. However, even experienced examiners often disagree about heart sounds and are additionally unable to detect certain sounds due to human auditory limitations, which include insensitivity to low frequencies, slow responses to rapidly occurring, brief sonic events, and the masking of soft sounds by loud sounds in close proximity [8].

Thankfully, with the advent of digital stethoscopes, computerized and automated heart sound classification systems can play a significant role in solving the inherent subjectivity and overcome limitations of human hearing ability. Additionally, these tools can play an important educational role and help to upskill medical staff in ausculation. All this should contribute to earlier diagnosis of heart murmurs in patients and thus lead to quicker treatment and better health care.

II. Related Work

Historically many attempts have been made to develop algorithms for automatic heart sound analysis. Recently, a comprehensive study of heart sound detection and classification algorithms was published [9] clearly pointing towards great interest in this topic. Indeed, the study lists a total of 117 relevant peer-reviewed articles, of which 53 focus on segmentation and 88 deal with sound classification. Among the many proposed approaches to solve the problem, the most often explored models are: support-vector machines (SVMs), artificial neural networks, hidden Markov models (HMM), k-nearest neighbor classifiers (k-NN) and their hybrids. Moreover the aspect of feature engineering and extraction is another important facet to this problem that has also attracted much focus (72 articles mentioned in the study). Remarkably, none of the articles mentioned in the study or later, to the best of our knowledge, investigate training a single machine learning model for joint

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

heart sound segmentation and murmur detection, which is the core idea that we present in this work.

In 2016, the Research Resource for Complex Physiologic Signals (PhysioNet) organized a challenge to develop algorithm for detection of abnormal heart sounds [10]. The challenge data still remains, to date, the most comprehensive publicly available dataset on the subject which is also confirmed in [9]. The winning submission [11] featured an ensemble of two models: an AdaBoost classifier trained on hand-crafted features extracted from both time and frequency domains and a convolutional neural network trained on raw audio samples. The solution relied on heart cycle segmentation provided in the challenge data which was obtained via a HMM-based algorithm [12]. More solutions have been proposed since the culmination of the challenge whose performance is unfortunately not always possible to fairly gauge. This is because the official test dataset was never released to the public, and moreover there is no commonly accepted performance evaluation procedure. According to [9], the best performance on the PhysioNet 2016 data has thus far been reported in [13] using Gram polynomials and probabilistic neural networks. However, the test was performed on a set of 300 recordings that are not representative of the whole dataset as they are derived from just one of the six constituent databases forming the PhysioNet 2016 training dataset.

To deal with the mentioned types of problems, we have chosen three reference high performing algorithms in heart murmur detection for which the full source code is publically available, thus allowing for a fair comparison. We summarize these below:

A. SVM classifier

The solution proposed by Ortiz et al. [14] utilizes an SVM classifier that is trained on a combination of four types of features. The first type of features are the time interval features that contain information about the length of each heart cycle phase (S1, systole, S2, diastole). The second and third set of features includes basic statistics such as mean and standard deviation computed from a) Mel-Frequency Cepstral Coefficients (MFCCs) and b) the Discrete Wavelet Transform (DWT) of heart cycle signals.

Finally, Discrete Time Warping (DTW) on the DWT is used as a measure of how individual heart cycles differ within one patient (Intra-DTW) and how they compare with class templates (Inter-DTW). Heart cycle segmentation was however not conducted, with the authors relying on this data provided by the challenge organizers.

B. Convolutional neural network trained on MFSC

The algorithm introduced by Maknickas et al. [15] is a simple convolutional neural network trained on MFSC (MFCC with no DCT) patches of size 128 by 128. Interestingly, this solution doesn't make any use of heart cycle segmentation - the MFSC coefficients are calculated for the entire recording and uniformly cut into frames which in general bear no relation to the heart cycles. Partial frames are filled with zeros and all frames are normalized. Data imbalance was addressed by sub-sampling more numerous class (healthy heartbeats) during training. The final prediction regarding the presence of heart murmur is obtained by averaging predictions from all frames extracted from a recording.

This solution was submitted for evaluation during the PhysioNet 2016 challenge. On the final leaderboard it is ranked sixth with a final score of 84.2%, which places it only 1.8 percentage points behind the winner.

C. Convolutional neural network trained on time series

The solution proposed by Humayun et al. [16] was developed as a part of The INTERSPEECH 2018 ComParE Challenge [17], where the task was to classify heart phonocardiograms into one the three classes: *Normal*, *Mild* and *Moderate/Severe*. Authors developed a two-stage algorithm, where the first stage included training the initial model on PhysioNet 2016 train data and second stage, where the knowledge was transferred to new model and fine-tuned for target task. The model is very similar to the CNN classifier used in [11]. Processing starts with resampling of the time series into a 1kHz signal and separation into four frequency bands: 25-45, 45-80, 80-200 and 200-500 Hz. Each of the four time series is cut into individual heart cycles according to externally provided heart segmentation ground truth, zero-padded to a total length of 2.5s (2500 samples) and fed into dedicated branches of the network where 1D convolutions are performed. Features from each branch are then concatenated and fed into a multilayer perceptron network that generates final predictions. Binary predictions from all heart cycles are averaged to obtain the final prediction for the entire recording.

III. Problem formulation

The problem that we solve in this article is simple to state: predict the presence/absence of heart murmur in an input stethoscope recording containing heartbeat signals. This is a simple binary classification task.

We are given a dataset of training stethoscope recordings with ground truth binary information of murmur presence. Additionally, some recordings have ground truth segmentation into the four phases of the heart cycle or information indicating absence of a heartbeat. In accordance with the definition of heart cycle phases provided in [12], the five segmentation classes are: (1) *no heart*; (2) *S1*; (3) *systole*; (4) *S2* and (5) *diastole*.

Our solution, described below, can either generate the predictions directly from the input recording, or it can rely on an external algorithm, e.g. [12], to provide automatic heart cycle segmentation.

IV. Proposed solution

A. Model

We propose to solve the task by using a deep neural network model that jointly performs heart cycle segmentation and predicts murmur presence directly from the input recording. Our

proposed model uses the modified Convolutional Recurrent Neural Network (CRNN) architecture described in [18].

As input, the network accepts spectrograms generated from recordings that were first resampled to 500Hz. The hop length used during generation of the spectrogram is 10ms and it is equal to the time resolution of the output prediction raster.

The final layer of the network consists of two groups of neurons that represent the two different output types. The first group consists of five neurons that encode the segmentation of the input signal into the five possible classes described earlier. The second group consists of two neurons and describes the presence or absence of heart murmurs. The occurance of heart murmur, as well as heart cycle segmentation, are output for each frame. Softmax is applied at the output of the network, independently for each group of neurons.

B. Masked Loss Function

1) Segmentation Loss: The first output group of our proposed network can be represented as a matrix $\hat{\mathbf{S}} \in \langle 0,1 \rangle^{C_s \times N}$, where $C_s = 5$ is the number of segmentation classes and N is the number of time frames. We represent the ground truth training set for this output as $\mathbf{S} \in \{0,1\}^{C_s \times N}$. Given these two matrixes, the training loss function for segmentation is defined as the conditional categorical crossentropy:

$$cross(S, \hat{S}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{5} S_{c,n} \log(\hat{S}_{c,n}), \tag{1}$$

$$loss_{segmentation} = \begin{cases} cross(S, \hat{S}), & \text{if } \mathbf{S} \text{ is known} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

2) Heart Murmur Loss: The second output of the network representing heart murmur presence is handled in similar fashion. $\mathbf{M} \in \{0,1\}^{C_m \times N}$ represents the ground truth raster for heart murmur with $C_m = 2$ referring to (1) *murmur* and (2) *no murmur* classes while $\hat{\mathbf{M}} \in \langle 0,1 \rangle^{C_m \times N}$ represents network predictions for *murmur* and *no murmur* classes.

The ground truth labelling for murmur presence is provided globally for each full recording. As a result, the ground truth raster $\mathbf{M}$ is prepared such that this single label is repeated for all frames of a given recording.

This definition poses a significant obstacle to model training based on the direct comparison of $\mathbf{M}$ and $\hat{\mathbf{M}}$. This is because the particular frames with heart murmur are not known. We note that out of the four heart cycle phases heart murmur can only be heard in two, *viz. systole* and *diastole*. Moreover some recordings might contain a part with no audible heartbeat (that can occur because e.g. the stethoscope did not touch the skin properly during part of the recording). For efficient identification of murmurs, it is hence pertinent that heart murmurs are not inferred during the S1 or S2 phases, as well during intervals where no heartbeat is present.

To implement this understanding into our model, we modify the crossentropy with an additional weighting scheme. The goal of this is to evaluate network predictions regarding murmurs only on valid heart cycle phases for calculation of the loss function. Technically, we introduce a weight mask $\hat{\mathbf{W}} \in \langle 0,1 \rangle^{N}$ defined as:

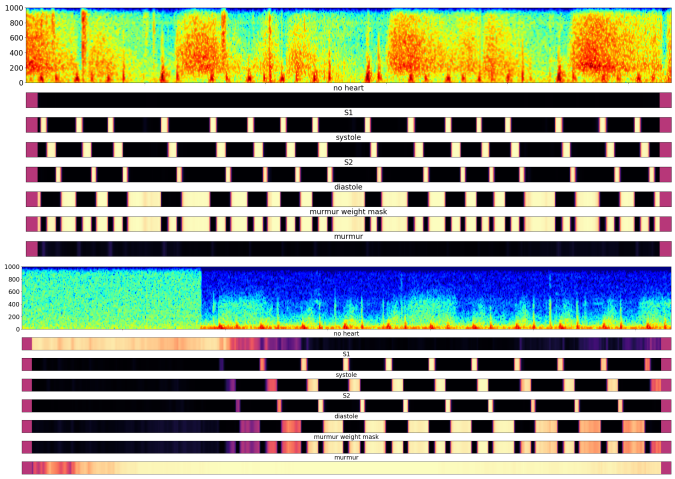

Fig. 1. Visualization of input spectrograms and model predictions for two example recordings. Output rasters are presented with a color map where values close to 0 are dark and values closer to 1 are light. The violet areas at the beginning and end of each raster depict network margins - frames for which the network did not provide predictions due to the lack of padding in convolutional filters. (Top) Network successfully segmented heart cycles despite the fact that heart cycle changes periodically with breathing cycle, which is a known physiological phenomena. The final output shows that network did not detect heart murmur in the recording. (Bottom) In the first part of the recording no heartbeat can be heard, the network was able to correctly detect heart in the second part of the recording. Network correctly predicted presence of heart murmur (indeed the patient was diagnosed with Aortic Stenosis).

$$\hat{W}(\hat{S}) = \max(\hat{S}_3, \hat{S}_5) \tag{3}$$

where $\hat{S}_3$ and $\hat{S}_5$ represent the network prediction rasters for the *systole* class and *diastole* class respectively. We note here that the weight mask is created from current network predictions, since the ground truth segmentation is assumed to be unknown in general during inference. Now the weighted heart murmur crossentropy becomes:

$$w\_cross(M, \hat{M}, \hat{W}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{2} M_{c,n} \log(\hat{M}_{c,n}) \hat{W}_n \tag{4}$$

which leads to the heart murmur loss:

$$loss_{murmur} = \begin{cases} w\_cross(M, \hat{M}, \hat{W}), & \text{if } \mathbf{M} \text{ is known} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

It is important to note here that the weight mask is determined independently of the gradient calculation used in the minimization of the loss function for the model. Specifically, network predictions for segmentation are treated as constants. Without this, the optimal way to minimize murmur loss would be to never detect a heart cycle.

3) Training Loss: Our final optimization criterium is chosen as a weighted sum of segmentation loss and heart murmur

loss. A larger weight is associated with the heart murmur prediction, which is the central task of this paper:

$$loss = loss_{segmentation} + \alpha \times loss_{murmur} \quad (6)$$

In all conducted experiments, α was set to 2.

With this definition, the model is trained for simultaneous heart cycle segmentation and heart murmur detection. This allows for sharing learned features between these two tasks. Such interchange of information is not possible with solutions proposed so far where the two tasks are solved separately.

C. Recording-level Murmur Prediction

The network, trained to minimize training loss (Equation 6), generates predictions for each audio frame of the input signal. Given the two prediction sets: $\hat{S}$ for segmentation and $\hat{M}$ for murmur, we aim to extract a single prediction about murmur presence for the whole recording. To do this we propose to use the following formula:

$$p_{murmur}(\hat{S}, \hat{M}) = \frac{\sum_{n=1}^{N} \hat{M}_{1,n} \hat{W}_n(\hat{S})}{\sum_{n=1}^{N} \hat{W}_n(\hat{S})} \quad (7)$$

where $\hat{M}_{1,n}$ represents the *murmur* class network prediction for the audio frame n.

According to this definition, the recording-level prediction for presence of heart murmur is a weighted sum of network predictions for the *murmur* class over all audio frames. The weighting is consistent with the training loss function and implies that the network's prediction for *murmur* are only taken into consideration in audio frames where the same network also predicts *systole* or *diastole*.

V. Evaluation

A. Challenge Data

We evaluated the performance of our and baseline solutions on the popular PhysioNet Computing in Cardiology Challenge 2016 [10]. In particular, we used the updated official ground truth data posted by the organizers on the challenge forum on July 25th, 2016. This data contained a total of 3153 recordings, of which 665 were labelled as containing murmur and 2488 were labelled as not containing murmur. For all recordings, two versions of heart cycle segmentation were provided: automatic segmentation generated using [12], and the same segmentation that was further manually verified and corrected by an expert cardiologist. For the purpose of our studies we used the latter as heart segmentation ground truth.

According to this segmentation some recordings did not contain any valid heart cycles (although they had ground truth information about murmur). Because some of the reference algorithms we were comparing our solution with relied on the properly detected heart cycle, they were unable to generate prediction for these recordings. Therefore in order to be able to compare all solutions on the same dataset, we decided to exclude these recordings from the final evaluation. In total, the final number or recordings used for evaluation was 2850 (557 with murmur and 2293 without murmur). We emphasize again that our algorithm does not suffer from the above mentioned limitations and can generate predictions for all recordings.

B. Evaluation procedure

We designed a 10-fold cross-validation experiment to get most reliable results. Folds were generated randomly, but for each tested algorithm exactly the same split was used. The trainval dataset was split into training (80%) and validation (20%). Results obtained from the experiments are summarized in two ways. First, we aggregate the test set results from each fold and report the resulting cumulative confusion matrices as well as the following metrics: precision, sensitivity (recall), specificity, F1-score and balanced accuracy (BACC). Secondly, we calculate these statistics separately for each fold and report their mean values and standard deviations. Finally, we evaluate the statistical significance of differences observed in performance between our proposed algorithm and the three alternatives by means of a Wilcoxon signed-rank test (two-sided) and report the obtained p-values.

VI. Implementation

A. Proposed Solution

Our solution was implemented in Python using the Tensorflow backend. The model was trained for 100 epochs using a batch size of 32. After each epoch, the performance of the model was evaluated on the validation set and upon obtaining the new best result, a snapshot of output weights was taken. The final snapshot was then used to generate predictions on the test set of the fold.

The simplicity of the pipeline makes the proposed algorithm computationally efficient. Each single training epoch (*i.e.* single swipe through all training recordings) takes 15 seconds on a modern home-use GPU (GeForce GTX 1070) or 135 seconds on a CPU (Intel Core i7-6700). The inference time for a 20-second recording, including preparation of input data (resampling and spectrogram generation) takes 340ms on CPU and under 50ms on GPU.

Two examples of network predictions are shown in Figure 1.

B. Baseline models

We used the publicly available author implementations of the three reference algorithms considered in this paper. For tunable parameters lacking available default values we tested all available options and chose the best performing ones in our evaluation.

In particular, this was the case for SVM parameters used in the solution proposed by Ortiz et al. In case of Humayun et al., we used only the first stage of their two-stage solution as their solution was designed to solve a different problem than defined here. For the solution of Maknickas et al. we modified the original author code to match the pipeline described in this publication. This applies in particular to implementation of 2-stage pipeline with pre-training model using balanced classes and finetuning the model using original classes ratio.

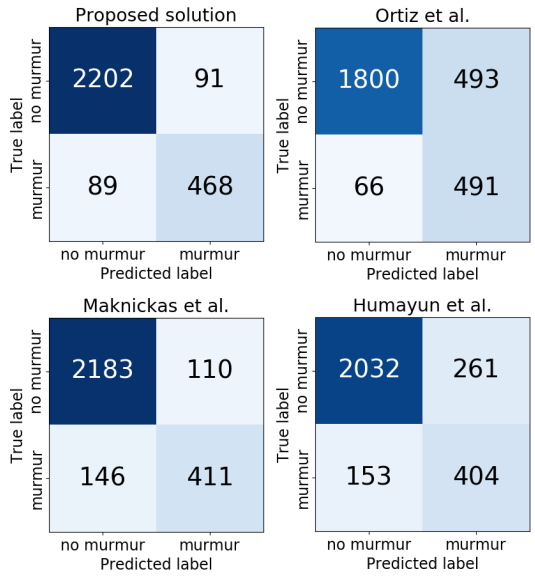

Fig. 2. Cumulative confusion matrices obtained from evaluation experiments for all four considered algorithms.

TABLE I
HEART MURMUR DETECTION PERFORMANCE OF EVALUATED ALGORITHMS (CUMULATIVE)

	Proposed solution	**Ortiz et al.**	**Maknickas et al.**	**Humayun et al.**
Precision	**83.7%**	49.9%	78.9%	60.8%
Sensitivity (Recall)	84.0%	**88.2%**	73.8%	72.5%
Specificity	**96.0%**	78.5%	95.2%	88.6%
F1-score	**83.9%**	63.7%	76.3%	66.1%
BACC	**90.0%**	83.3%	84.5%	80.6%

VII. RESULTS

We now summarize results of the evaluation procedure performed on the proposed solution and three baseline models. Cumulative confusion matrices obtained from all four experiments are depicted in Figure 2.

Based on these results we calculated evaluation metrics, which are presented in Table I.

As can be seen, our solution performed the best with regards to precision, specificity, F1-score and balanced accuracy. Our main evaluation metric, the F1-score, is witness to an improvement of 7.6 percentage points of our method with respect to the best reference algorithm of Maknickas et al. [15].

In terms of sensitivity, the solution of Ortiz et al. [14] performed best. Note, however, that it was characterized by very low precision (49.9%) and specificity (78.5%).

In order to evaluate the significance of the differences in performance of tested models, we summarize the variability of metrics between folds in Table II, where mean values and standard deviations for each metric are given. Additionally the table includes p-values obtained by performing the Wilcoxon signed-rank test for each considered metric (numbers in round brackets). The Wilcoxon test was performed between the proposed solution and each of the three alternative algorithms.

Amongst all tested combinations, assuming a significance level of 5%, we fail to reject the null hypothesis that the distributions are the same in only two cases. In particular, according to the test, precision and specificity of our proposed algorithm is not significantly higher than that of Maknickas et al. However, our proposed model shows significantly higher sensitivity, F1-score and balanced accuracy. Compared do Humayun et al., our solution shows significantly higher performance in terms of all presented metrics. Also with respect to Ortiz et al. all differences turned out to be significant, which means that the baseline solution shows higher sensitivity compared to our algorithm (although the p-value is close to the thresholds), but the performance of our model is significantly higher in all remaining metrics. Moreover our solution shows significant improvement over all considered benchmarks in terms of the F1-Score and balanced accuracy.

Going beyond the performance in detection of hear murmur, we also studied analogous performance with respect to the task of segmentation. This evaluation was performed with 10ms time resolution at the frame level. We accumulated information from all frames of all evaluated recordings, yielding almost 7 million individual ground truth - prediction pairs. For each of the five classes we calculated the same set of metrics as in the discussion of heart murmur performance: precision, sensitivity (recall), specificity, F1-score and balanced accuracy. Results are presented in Table III.

These findings show that our solution performs heart cycle segmentation well. The only class for which the performance is not satisfactory is the class that represents lack of audible heartbeat ('no heart'). The reason for this is that this class is heavily under-represented in the dataset used in this study. In fact, only 8.7% of audio frames in the PhysioNet 2016 [10] hand-corrected augmentations are labelled as 'no heart'. This problem can be easily solved by extending the training dataset with more examples of recordings that have no heartbeat.

VIII. CONCLUSIONS

In this paper a new solution to the heart murmur detection problem was presented. We propose to train the model for simultaneous detection and segmentation of heart cycles and heart murmur prediction. This is in contrast to the standard 2 stage approach consisting first of heart cycle detection followed by classification of heart murmur presence. Our solution displays a significant advantage over the considered benchmarks and has the following additional advantages that make it suitable for implementation:

- Simple pipeline consisting of single model that accepts raw audio signal
- Can be trained on recordings provided only with heart murmur label, only with heart segmentation, or provided with both
- Provides output on the validity of the recording (*i.e.* if a heartbeat was detected) and heart rate.

TABLE II
HEART MURMUR DETECTION PERFORMANCE OF EVALUATED ALGORITHMS (VARIABILITY BETWEEN FOLDS AND WILCOXON TEST P-VALUES)

	Proposed solution	**Ortiz et al.**	**Maknickas et al.**	**Humayun et al.**
Precision	$\mathbf{83.9} \pm 6.2\%$	$50.5 \pm 8.9\%$ (0.005)	$78.9 \pm 6.9\%$ (0.139)	$62.1 \pm 8.3\%$ (0.005)
Sensitivity (Recall)	$84.1 \pm 7.6\%$	$\mathbf{88.2} \pm 5.9\%$ (0.047)	$73.8 \pm 7.4\%$ (0.005)	$72.3 \pm 7.8\%$ (0.037)
Specificity	$\mathbf{96.0} \pm 1.8\%$	$78.5 \pm 5.7\%$ (0.005)	$95.2 \pm 1.6\%$ (0.406)	$88.5 \pm 5.6\%$ (0.005)
F1-score	$\mathbf{83.6} \pm 3.5\%$	$63.7 \pm 6.2\%$ (0.005)	$75.9 \pm 5.1\%$ (0.013)	$66.2 \pm 3.9\%$ (0.005)
BACC	$\mathbf{90.1} \pm 3.3\%$	$83.4 \pm 2.0\%$ (0.005)	$84.5 \pm 3.5\%$ (0.007)	$80.4 \pm 2.5\%$ (0.005)

TABLE III
PERFORMANCE OF PROPOSED SOLUTION IN HEART CYCLE SEGMENTATION TASK (CUMULATIVE)

	no heart	**S1**	**systole**	**S2**	**diastole**
Precision	61.9%	88.9%	88.7%	85.2%	91.4%
Sensitivity (Recall)	61.5%	86.3%	88.9%	80.1%	93.9%
Specificity	96.4%	98.1%	97.1%	98.1%	93.1%
F1-score	61.7%	87.6%	88.8%	82.6%	92.6%
BACC	93.3%	96.3%	95.5%	95.9%	93.5%

We believe that proposed solution is an important step towards developing AI algorithms that can be commonly used by both medical staff and individual home users. In the first case, they can play both an educational as well as supportive role providing aid whenever a physician is uncertain of diagnosis. In principle, this could help enhance the audible heart sound identification skills of physicians and also allow the detection of sounds that are undetectable to the human ear. Home users could benefit from access to automatized basic screening of household members and allow for fast detection of potential problems without the need to visit a doctor in person. This, in turn, should also reduce waiting times for patients who are in more immediate need of consultancy. Moreover, we expect that the general framework we used for solving the heart murmur detection problem will be useful in other signal processing domains as well, especially in those that combine tasks of localized detection and overall signal classification.

REFERENCES

[1] A. J. Taylor, *Learning Cardiac Auscultation*. Springer, 2015.

[2] M. Thoenes, P. Bramlage, P. Zamorano, D. Messika-Zeitoun, D. Wendt, M. Kasel, J. Kurucova, and R. P. Steeds, "Patient screening for early detection of aortic stenosis (as)—review of current practice and future perspectives," *Journal of thoracic disease*, vol. 10, no. 9, p. 5584, 2018.

[3] U. Alam, O. Asghar, S. Q. Khan, S. Hayat, and R. A. Malik, "Cardiac auscultation: an essential clinical skill in decline," *British Journal of Cardiology*, vol. 17, no. 1, p. 8, 2010.

[4] M. R. Montinari and S. Minelli, "The first 200 years of cardiac auscultation and future perspectives," *Journal of multidisciplinary healthcare*, vol. 12, p. 183, 2019.

[5] S. Mangione, "Cardiac auscultatory skills of physicians-in-training: a comparison of three english-speaking countries," *The American journal of medicine*, vol. 110, no. 3, pp. 210–216, 2001.

[6] M. J. Barrett, C. S. Lacey, A. E. Sekara, E. A. Linden, and E. J. Gracely, "Mastering cardiac murmurs: the power of repetition," *Chest*, vol. 126, no. 2, pp. 470–475, 2004.

[7] S. Mangione, L. Z. Nieman, E. Gracely, and D. Kaye, "The teaching and practice of cardiac auscultation during internal medicine and cardiology training: a nationwide survey," *Annals of internal medicine*, vol. 119, no. 1, pp. 47–54, 1993.

[8] M. E. Tavel, "Cardiac auscultation: a glorious past—but does it have a future?" *Circulation*, vol. 93, no. 6, pp. 1250–1253, 1996.

[9] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds–a systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2018.

[10] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, no. 12, p. 2181, 2016.

[11] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 621–624.

[12] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-hsmm-based heart sound segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 822–832, 2015.

[13] F. Beritelli, G. Capizzi, G. L. Sciuto, C. Napoli, and F. Scaglione, "Automatic heart activity diagnosis based on gram polynomials and probabilistic neural networks," *Biomedical engineering letters*, vol. 8, no. 1, pp. 77–85, 2018.

[14] J. J. G. Ortiz, C. P. Phoo, and J. Wiens, "Heart sound classification based on temporal alignment techniques," in *2016 Computing in Cardiology Conference (CinC)*. IEEE, 2016, pp. 589–592.

[15] V. Maknickas and A. Maknickas, "Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients," *Physiological measurement*, vol. 38, no. 8, p. 1671, 2017.

[16] A. I. Humayun, M. T. Khan, S. Ghaffarzadegan, Z. Feng, and T. Hasan, "An ensemble of transfer, semi-supervised and supervised learning methods for pathological heart sound classification," in *Proc. Interspeech 2018*, 2018, pp. 127–131. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2018-2413

[17] B. W. Schuller, S. Steidl, A. Batliner, P. B. Marschik, H. Baumeister, F. Dong, S. Hantke, F. B. Pokorny, E.-M. Rathner, K. D. Bartl-Pokorny *et al.*, "The interspeech 2018 computational paralinguistics challenge: Atypical & self-assessed affect, crying & heart beats." in *Interspeech*, 2018, pp. 122–126.

[18] E. Cakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, "Convolutional recurrent neural networks for polyphonic sound event detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1291–1303, 2017.