Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection

Siddique Latif[1], Muhammad Usman[2], Rajib Rana[3], and Junaid Qadir[1]

[1]Information Technology University (ITU)-Punjab, Pakistan
[2]COMSATS Institute of Information Technology, Islamabad
[3]University of Southern Queensland, Australia

***Abstract*—Deep learning based cardiac auscultation is of significant interest to the healthcare community as it can help reducing the burden of manual auscultation with automated detection of abnormal heartbeats. However, the problem of automatic cardiac auscultation is complicated due to the requirement of reliable and highly accurate systems, which are robust to the background noise in the heartbeat sound. In this work, we propose a Recurrent Neural Networks (RNNs) based automated cardiac auscultation solution. Our choice of RNNs is motivated by their great success of modelling sequential or temporal data even in the presence of noise. We explore the use of various RNN models, and demonstrate that these models significantly outperform the best reported results in the literature. We also present the run-time complexity of various RNNs, which provides insight about their complexity versus performance trade-offs.**

I. Introduction

Cardiovascular diseases (CVDs) are the major health problem and have been the leading cause of death globally. They are causing nearly 48% deaths in Europe [1], 34.3% in America [2], and more than 75% in developing countries [3]. Early diagnosis of CVDs is crucial as it can drastically decrease the potential risk factors of these deaths [4].

Auscultation is a widely used CVD diagnosis method that relies on the use of stethoscope to determine signs of cardiac abnormalities. It is, however, important to note that proper auscultation requires extensive training and experience as it is extremely difficult to perform. Expert cardiologists have been reported to achieve approximately 80% accuracy [5], while primary care physicians and medical students usually achieve around 20-40% accuracy [6].

Another effective solution is echocardiograms that visualise the heart beating and blood pumping. However, this procedure is expensive with an average cost of $1500 per procedure [7].

The rise of phonocardiography (PCG), which is an effective and non-invasive method for early detection of cardiac abnormality, has addressed the above-mentioned challenges to a large extent. In PCG, heart sound is recorded from the chest wall using a digital stethoscope and this sound is analysed to detect whether the heart is functioning normally or the patient should be referred to an expert for further diagnosis. Due to its high potential, automatic detection of cardiac abnormalities through PCG signal is an emerging field of research [8]. However, interestingly, most of these automated cardiac auscultation attempts have utilised either classical machine learning models (e.g., [9]–[12]) or feed-forward neural networks [13], [14] rather than Recurrent Neural Networks (RNNs)—which are intrinsically better suited for this task.

Heart sound is a physiologic time series and it possesses temporal dynamics that change based on the different heart symptoms. Due to their well-known capabilities for modelling and analysing sequential data even in the presence of noise [15], we propose the use of RNNs for automated cardiac auscultation. We use 2016 PhysioNet Computing in Cardiology Challenge dataset [16] that contains phonocardiograms recorded using sensors placed at the four common locations of human body: pulmonic area, aortic area, mitral area, and tricuspid area. The work presented in this study is the *first attempt* that investigates the performance and run-time complexity of various state-of-the-art RNNs for heartbeat classification using PCG signals.

II. Related Work

In the past few years, automatic analysis of phonocardiogram (PCG) has been widely studied especially for automated heartbeat segmentation and classification. According to Liu et al. [16] there was no existing study that applied deep learning for automatic analysis on heartbeat before the 2016 PhysioNet Computing in Cardiology Challenge. Recently, there have been few attempts using deep learning models for classifying normal and abnormal heart sounds, which we describe next.

A deep learning based approach was used in [17] for automatic recognition of abnormal heartbeat using a deep Convolutional Neural Network (CNN). The authors computed a two-dimensional heat map from one-dimensional time series of PCG signal with the overlapping segment length of $T = 3$ seconds and used for training and validation of the model. They achieved the highest specificity score 0.9521 as compared to all entries made in PhysioNet Computing in Cardiology challenge but their sensitivity and accuracy scores were low: 0.7278 and 0.8399, respectively.

A fully connected neural network (NN) consisting of 15 hidden layers was used in [14] for the classification of PCG signals. The authors achieved the accuracy of 0.80 with the specificity of 0.82, however, the sensitivity was low: 0.63. Potes et al. [18] used an ensemble of AdaBoost and the CNN classifiers to classify normal/abnormal heartbeats. This ensemble approach achieved the highest score in the among others and achieved *rank one* in the competition with the specificity, sensitivity, and overall score of 0.7781, 0.9424, and 0.8602, respectively. We have achieved a significantly higher score than this best-reported result in the literature. Other approaches in this challenge were based on the classical machine learning based classifiers. Interestingly, none of the above studies have attempted RNN, which is, however, the most powerful model for time series data.

RNN architectures such as Long Short-term Memory (LSTM) and Gated Recurrent Units (GRU) have not been used for PCG analysis but they have achieved state-of-the-art performances in various other applications with sequential data including speech recognition [19], [20], machine translation [21], [22], and emotion detection [23]. Bidirectional Long Short-Term Memory (BLSTM) [24], which processes the information in both directions with two different LSTM layers show even better performance than LSTM and other conventional RNNs in phoneme classification and recognition [25]. Deep BLSTMs with the objective function of Connectionist Temporal Classification (CTC) are very effective in optimising word error rate in speech recognition system with no prior linguistic or lexicon information [26]. Bidirectional Gated Recurrent Units (BGRUs) have also been used for audio processing. They achieved a significant improvement in the performance of sound events detection from the recording of real-life, by capturing their temporal boundaries [27].

III. Proposed Approach and RNN Models

Our proposed approach for heart sound classification using RNNs is depicted in Figure 1. The heart sound is first prepossessed for first and second heart sounds (S1 and S2, respectively) detection and segmented into smaller chunks. Features extracted from these segments are given to the RNNs for classification, which classifies these into normal or abnormal.

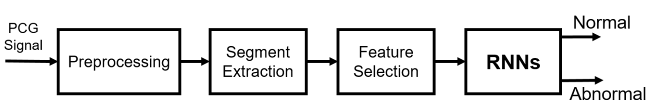


Figure 1: Block diagram of proposed approach

A. Recurrent Neural Networks (RNNs)

Recurrent neural networks are specialized to process sequences, unlike the CNNs, which are specialized for grid-like structures, such as, images. It takes an input sequence $x(t) = (x_1, ..., x_T)$ and at the current time step t calculates the hidden state or memory of the network h_t using previous hidden state h_{t-1} and the input x_t. The standard equations for RNN are given below:

$$h_t = \boldsymbol{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{1}$$

$$y_t = (W_{xh}x_t + b_y) \tag{2}$$

where W terms are the weight matrices (i.e., W_{xh} is weight matrix of an input-hidden layer), b is the bias vector and $\boldsymbol{H}$ denotes the hidden layer function.

For classification, the outputs from RNN are projected to the number of classes. In particular, a softmax function is used to project the output vector into the probability vector having values in [0,1]. While using RNNs for abnormal heartbeat detection, its end layer is projected to a two class vector: normal and abnormal. The class receiving the higher probability gets outputted as the label for the current heartbeat segment.

1) Long Short-Term Memory (LSTM) Units: The LSTM [15] network is a special type of RNN that consists of a recurrent memory block to store representations for extended time intervals. A memory block consists of three gates: input, output and forget gate. These multiplicative gates learn to control the constant error flow within each memory cell. The memory cell decides what to store, and when to enable reads, writes and erasures of information. Graphical representation of LSTM memory cell is shown in Figure 2a.

When features from a sequence of heart signal are given to the network, each LSTM unit holds a memory c_t at a specific time t. The activation function is given by:

$$h_t = o_t \tanh(c_t) \tag{3}$$

The output gate o_t modulates the memory content and calculated by:

$$o_t = \sigma\big(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\big) \tag{4}$$

The forget gate f_t control the memory in the network and update it by forgetting the existing memory c_t with the incoming information.

$$f_t = \sigma\big(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_t + b_f\big) \tag{5}$$

The extent of incoming information is controlled by the input gate i_t.

$$i_t = \sigma\big(W_{if}x_t + W_{hi}h_{t-1} + W_{ci}c_t + b_i\big) \tag{6}$$

The existing memory in the network is finally updated by the following equation under the control of these three gates.

$$c_t = f_t c_{t-1} + i_t \tanh\big(W_{xc}x_t + W_{hc}h_{t-1} + b_c\big) \tag{7}$$

2) Gated Recurrent Units (GRUs): The Gated Recurrent Unit (GRU) [21] is a slightly simplified version of LSTM that combines the input and forget gates into a single gate known as update gate. GRU architecture has an additional reset gate as compared to LSTM (see Figure 2b). GRUs control the information flow from the previous activation while computing candidate activation but unlike LSTM do not control the amount of incoming memory using the forget gate.

B. Bidirectional Recurrent Neural Networks

One shortcoming of standard RNNs is that they can only use previous information for making decisions. Bidirectional RNNs [24] using LSTM units or GRUs can process the information both in the forward and backward direction which

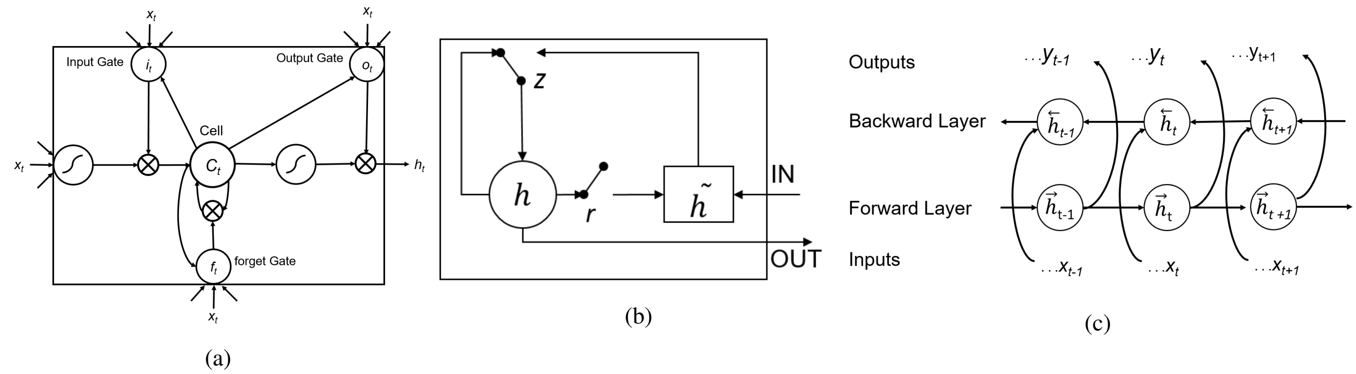


Figure 2: Graphical representation of (2a) LSTM memory cell; (2b) Gated Recurrent Unit (GRU); (2c) bidirectional LSTM.

enable them to exploit future context. This is achieved by computing the hidden sequence both in forward $\overrightarrow{h}$ and backward direction $\overleftarrow{h}$, and updating the output layer using backward layer from time step $t = T$ to 1 and forward layer from $t = 1$ to T (see Figure 2c).

Similarly, a bidirectional GRU (BGRU) exploits the full use of contextual information and produces two sequences $[h_1^f, h_2^f, ..., h_T^f]$ and $[h_1^b, h_2^b, ..., h_T^b]$ by processing the information both in forward and backward directions, respectively. These two sequences are concatenated at the output by the following equation:

$$\overleftrightarrow{GRU}(X) = |h_1^f||h_T^b, ..., h_T^f||h_1^b|. \quad (8)$$

Here, the $\overleftrightarrow{GRU}(X)$ term represents the full output of BGRU produced by concatenating each state in forward direction h_i^f and backward direction $h_{(T-i+1)}^b$ at step i given the input X. In this paper, we use both RNNs and bidirectional RNNs and compare their performances on abnormal heartbeat detection.

IV. Experimental Procedure

The performance of different RNN models is evaluated on the publicly available datasets. The details of the datasets and experimental procedure are presented in this section.

A. Database Description

To evaluate the proposed methodology, largest dataset provided at the Physionet Challenge 2016 [16] has been used. The Physionet dataset consists of six databases (A through F) containing a total of 3240 raw heart sound recordings. These recordings were independently collected by different research teams using heterogeneous sensing equipment from seven countries spanning three continents both in clinical and nonclinical (i.e., home visits) settings. The recordings were collected from both healthy subjects, and patients with a variety of heart conditions, especially coronary artery disease and heart valve disease. The subjects were also from different age groups including children, adults and the elderly. Due to the uncontrolled environment, recordings are corrupted by different noises such as stethoscope motion, intestinal activity sounds, breathing, and talking. The length of heart sound recordings varied from 5 seconds to just over 120 seconds. For our experiments, we use all six databases containing normal and abnormal heart sound recordings.

B. Preprocessing of Heart Sound

The heart sound recorded by electronic stethoscope often has background noise. The preprocessing of heart sound is an essential and crucial step for automatic analysis of heartbeat recordings. It reveals the inherent physiological structure of the heart signal by detecting the abnormalities in the meaningful regions of PCG signal and allows for the automatic recognition of pathological events. The detection of the exact locations of the first and second heart sounds (i.e., S1 and S2) within PCG is known as the segmentation process. The main goal of this process is to ensure that incoming heartbeats are properly aligned before their classification as it significantly improves the recognition scores [28].

In this paper, we used state-of-the-art method Logistic Regression-Hidden Semi-Markov Models (HSMM) for identification of heart states proposed by Springer et al. [29]. This method uses LR-derived emission or observation probability estimates and provides significantly improved results as compared to the previous approaches based on the Gaussian or Gamma distributions [30], [31].

The working of Logistic Regression-HSMM is similar to SVM based emission probabilities [32] and it allows for greater discrimination between different states. Logistic regression is a binary classifier that maps the feature space or predictor variables to the binary response variables by using a logistic function. The logistic function $\sigma(a)$ is defined as:

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (9)$$

The probability of a state or class given the input observations O_t can be defined using the logistic function:

$$P\big[q_t = \xi | O_t\big] = \sigma(w^{'} O_t). \quad (10)$$

The term w represents the weights of the model that are applied to each observation or input features. The model is trained iteratively and re-weighted least squares on the training data. For one-vs-all logistic regression, the probability of each observation given the state $b_j(O_t|\xi_j)$ is found using Bayes'rule:

$$b_j(O_t) = P[O_t|q_t = \xi] = \frac{P[q_t = \xi|O_t] \times P(O_t)}{P(\xi_j)}. \quad (11)$$

The $P(O_t)$ is calculated from a multivariate normal distribution of the entire training data and $P(\xi_j)$ is the initial state probability distribution.

The Logistic Regression-HSMM algorithm use the combination of four type of features: Homomorphic envelope, Hilbert envelope, Wavelet envelope, and Power spectral density envelope. The details of these features can be found in [29]. The overall PCG recordings are given to the model for accurate detection of most probable states (i.e., S1 and S2).

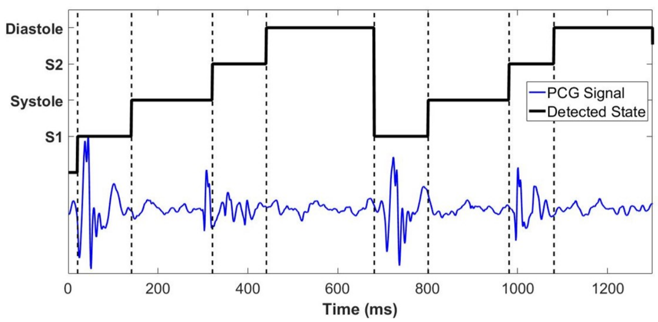


Figure 3: Four states (S1, S2, systole, and diastole) of the heart cycle using Logistic Regression-HSMM

Figure 3 shows the detected four states: S1, S2, systole, and asystole of two heart cycles. Note that, generally, it is called as S1 and S2 detection, although it detects all the four states. The blue line is for the heart signal and the black line shows the detected states using Logistic Regression-HSMM algorithm.

C. Segment Extraction

After detecting the position of heart states, we segmented the overall PCG waveform into shorter instances by locating the beginning of each heartbeat. This is because the number of audio recordings (i.e., 3240) of heart signal is not adequate to evaluate RNNs for designing a robust system. Segment extraction was also used in previous studies to divide the overall heart sound in smaller chunks. For instance, Rubin et al. [17] used a segment of $T = 3$ seconds for training and validation of CNN. In this paper, we segmented the heart signal into three sequences of heart cycles: 2, 5, and 8. Table I shows the number of instances for each segment. These instances contain enough information and at the same time are small enough to generate many samples for training the model.

Table I: Dataset details

Number of Cycles	Abnormal Instances	Normal Instances	Total Instances
Two Cycles	9724	35703	45427
Five Cycles	6436	21579	28015
Eight Cycles	4603	15826	20429
Overall Data	665	2575	3240

Figure 4a & 4b shows the five cycles of normal and abnormal heart sound, respectively. The abnormal heart sound is different from the normal one in temporal context. It has heart cycle states of longer duration in the segment.

D. Feature Selection

In this paper we used Mel-frequency cepstral coefficients (MFCCs) [33] to represent PCG signal in compact representation. MFCCs are used almost in every study on automatic heart sound classification (for example, [18], [34]–[36]) due to their effectiveness in speech analysis. We compute MFCCs from 25ms of the window with a step size of 10ms. We select the first 13 MFCCs for compact representation of PCG signal as a large feature space does not always improve the recognition rate of the model [37].

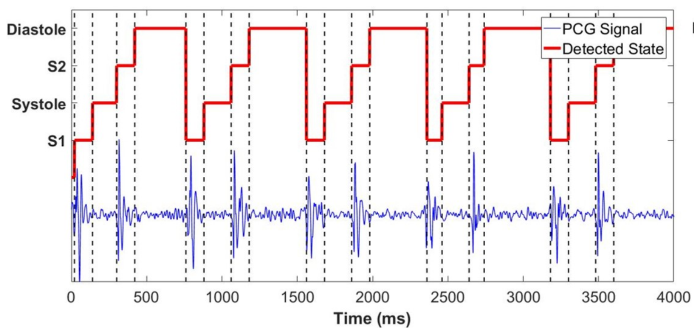
E. Model Parameters

We built our RNN models using Keras [38] with a TensorFlow backend [39]. In order to find the best models' structure, we evaluated a different number of gated layers from one to four. A smaller model with only one LSTM or GRU layer did not perform well on this task. Experiments with a larger number of gated layers and dense layers also failed to give improvements in performance. This is due to overfitting we anticipate. We got the best classification results using 2 gated layers for both LSTM and GRU models. Therefore, our LSTM and BLSTM models consist of two LSTM layers with $\tanh$ [40] as the activation function. For each heartbeat, the outputs of LSTM or BLSTM layers were given to the dense layer and the outputs of the dense layer were given to the softmax layer for classification (refer to Section III).

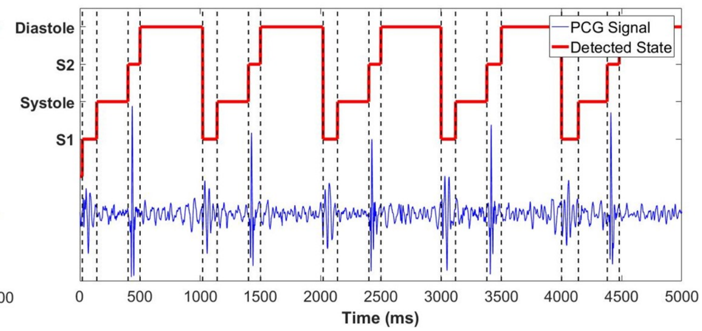
We trained all these models using the training set and development data was used for hyper-parameter selection. We started training the network with a learning rate of 0.002 and learning rate was halved after every 5 epochs if the classification rate of the model did not improve on the validation set. This process continued until the learning rate reached below 0.00001 or until the maximum epochs (100) were reached. We used batch normalisation after the dense layer for normalisation of learned distribution to improve the training efficiency [41]. In order to incorporate the effect of initialisation, we repeated each hyperparameter combination for three times and used the averaged prediction for validation and testing.

V. Results

The overall dataset of PhysioNet Computing in Cardiology Challenge consisted of six heart sound databases gathered from seven different countries. There is a total of 3240 publicly available heartbeat recordings. We detected heart states in each PCG waveform and segmented these signals into smaller chunks containing exact five heart cycles. MFCCs were computed from these chunks and both models were trained on 75% of data, 15% data was used for validation and remaining 10% of data was used for testing. As each model was trained on MFCC computed from the smaller segment, to predict



(a)



(b)

Figure 4: Extracted segments of five heart cycles using (4a) Normal and (4b) Abnormal heartbeats. **Abnormal heart cycles have longer duration.**

the score for full instance, an averaging was performed on posterior probabilities of the respective chunks. In this work, the performance of RNNs has been tested on three different cycles: 2, 5, and 8 in a heart signal (see Figure 5). All RNN models consistently performed well on 5 cycles. We anticipate that heart signals consisting of 5 cycles carry more information compared to a shorter signal with 2 cycles yet have enough number training instances for accurate classification. We also anticipate that although heartbeats of 8 cycles have more information, but incur a performance loss (insignificant) due to the decrease in the number of training instances. Onward we mention results using 5 cycles.

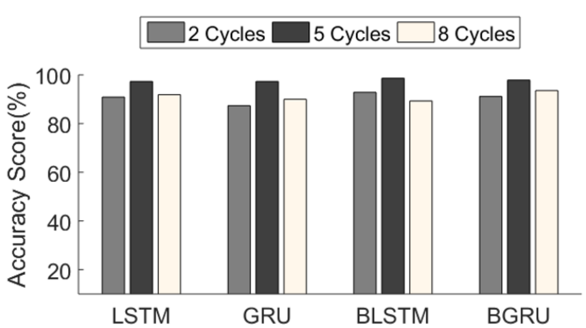


Figure 5: Performance (MAcc) of RNNs using different number of heart cycles, where RNNs give the best results with 5 cycles.

A. Baseline models

In order to compare the performance of RNNs, we selected a number of powerful classifiers including Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Artificial Neural Networks (ANNs) as baseline models. We trained these models on MFCCs computed from 5 heart cycles. Hyperparameters of these models were selected using validation data. In SVM, we tested three kernels: linear, polynomial, and radial basis function (RBF) to obtain the best classification results. Table II shows the best results obtained by these models and presents a comparison with RNNs. We observe RNNs significantly outperform all these models in every performance measures.

Table II: Comparison of different baseline classifiers with RNN

Model	Sensitivity	Specificity	Score (MAcc)
SVM (best)	0.8259	0.8324	0.8291
LR (best)	0.7121	0.6879	0.6991
RF (best)	0.6901	0.6850	0.6875
KNN (best)	0.6231	0.7041	0.6636
ANNs (best)	0.8056	0.8826	0.8441
RNNs (best)	**0.9886**	**0.9836**	**0.9861**

B. Comparison with non-recurrent deep models

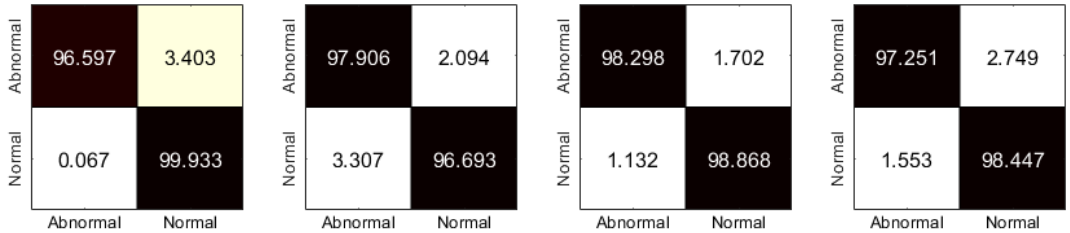
We compare the performance of RNNs with most recent studies on PCG classification using DNNs in Table III. It can be clearly noticed that RNNs outperform all non-recurrent deep learning models used on heart sound classification with a significant improvement. Worth mentioning, heart sound classification using the ensemble of AdaBoost and CNN achieved first rank in PhysioNet Computing in Cardiology Challenge 2016 with the best recognition rate. In our approach, RNNs using LSTM has performed better than that in every performance measures defined in [44].

Table III: Comparison of results with previous attempts.

Author (Year)	Approach	Sensitivity	Specificity	Score (MAcc)
Potes et al. [18] (2016)	AdaBoost and CNN	0.9424	0.7781	0.8602
Tschannen et al. [42] (2016)	Wavelet-based CNN	0.855	0.859	0.828
Rubin et al. [17] (2017)	CNN	0.7278	0.9521	0.8399
Nassralla at al. [14] (2017)	DNNs	0.63	0.82	0.80
Dominguez at al. [43] (2018)	Modified AlexNe	0.9512	0.9320	0.9416
Our Study (2018)	LSTM	**0.9995**	0.9671	0.9833
	BLSTM	0.9886	**0.9836**	**0.9861**
	GRU	0.9669	0.9793	0.9731
	BGRU	0.9846	0.9728	0.9787

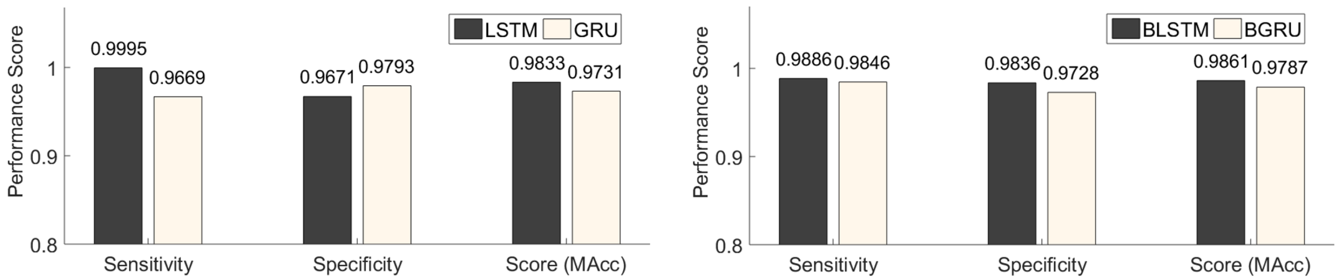
C. Performance Comparison of RNNs

In this study, our evaluation focused on the sequence modeling of heart sound using RNNs. We explored the performance of different state-of-the-art RNNs for this task. We first present



(a) LSTM (b) GRU (c) BLSTM (d) BGRU

Figure 6: Confusion matrices using RNN models.



(a) LSTM vs GRU (b) BLSTM vs BGRU

Figure 7: Performance comparison of RNNs on heartbeat classification

the confusion matrices in Figure 6. Based on these results, different RNNs have achieved comparable performances in general. We then calculate sensitivity, specificity, and mean accuracy (MAcc) score and present these results in Figure 7. Another important finding, despite having a simpler architecture compared to LSTM, the performance of GRU is strikingly promising on detecting both abnormal and normal heartbeats.

VI. Discussion

A. Theoretical Implications

Our proposed approach using RNNs for pathological heart murmurs detection provides almost 100% classification score. The obtained results are presented using three metrics: sensitivity, specificity, and challenge scores (MAcc). RNNs show significant improvements in performance on all three measures compared to leading approaches presented in the PhysioNet Computing in Cardiology Challenge and compared to other approaches recently been published (see Section V-B). These improvements suggest the significance of RNNs for designing a reliable tool for automatic auscultation. This is due to RNNs' ability to learn contextual information from the connective heart cycles that helps the model to classify accurately. We also found GRUs are computationally efficient compared to other RNN models. It takes 35% (2.8 times) less run-time compared to BLSTM while achieving a comparable result. This is an important finding for real-time implementation of automatic cardiac auscultation using mobile devices with limited battery power.

B. Practical Implications

Cardiac auscultation is one of the widely used methods for the diagnosis of different cardiac abnormalities. However, physicians are required to have extensive training and expertise to accurately detect heart abnormalities [45]. It is reported that the accuracy for auscultation performed by expert cardiologists is roughly 80% [43]. There are two type of errors: Type-I and Type-II, usually caused by physicians during the auscultation process. A Type-I error (also known as alpha error) is the detection of the wrong abnormality and a healthy person is suggested for the echocardiogram. The Type-II or beta error occurs when an expert failed to detect the presence of heart abnormalities and pathological patients are sent home without any treatment. This is essentially a false negative and therefore more serious than the Type-I and needs to be avoided. Encouragingly, our proposed approach achieved the best result on sensitivity, which ensures that it correctly identifies patients with abnormality, minimising the false negatives. Also, our proposed approach achieves the best results on specificity, which minimises the Type I error, which is essentially a false positive.

VII. Limitation and Future Work

We leverage various RNNs for abnormal heartbeat detection problem. Through extensive experiments, we show that RNNs outperform other approaches even CNNs that are very powerful deep models. For practical application, computationally efficient models are crucial to build a portable system. We compared the run-time of four models using 3.40 GHz Intel Core i7 with 20 GB memory and 8 GB NVIDIA Quadro M5000 GPU and performed each experiment five times to compare the average training time for each model. We found that GRUs take 28% (1.4 times) less time compared to LSTM and BGRUs incur 25% (1.3 times) less time compared to BLSTM. Encouragingly, GRUs have almost similar performance in all three matrices compared to other models. Therefore, they can be the potentially suitable models in real-time scenarios using mobile devices. The limitation of our work lies in the fact that these results are generated using a desktop computer, not using an embedded device. However, it provides important insights into the time complexity of different RNNs for automatic auscultation. In our future work, we aim to evaluate the performance of GRUs for abnormal heartbeat detection on mobile devices. Moreover, we will also keep searching for larger heartbeat datasets and consider options for collecting our own dataset, for our future studies.

VIII. Conclusion

In this paper, we conduct an empirical study on abnormal heartbeat detection using PCG signal and demonstrate that Recurrent neural networks (RNNs) produce promising results. In particular, the results are significantly better than the conventional deep learning models. For example, the proposed RNN models significantly outperform the AdaBoost-CNN model which was placed Rank 1 in the PhysioNet Computing in Cardiology Challenge 2016. RNNs have the most important architectures to capture the temporal statistics and dynamics in the sequence of heartbeats more efficiently as compared to the other popular DNNs like CNN. Our results also show interesting information on performance and complexity of various RNN architectures. In particular, the comparison of GRU and BLSTM is quite interesting and insightful. For instance, BLSTM performs 1.3% better than GRU while incurring 2.8 times higher computational load. In our future studies, we aim to produce further empirical evidence using RNNs on mobile devices.

References

[1] S. Allender, P. Scarborough, V. Peto, M. Rayner, J. Leal, R. Luengo-Fernandez, and A. Gray, "European cardiovascular disease statistics," 2008.

[2] D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, T. B. Ferguson, E. Ford, K. Furie, C. Gillespie *et al.*, "Heart disease and stroke statistics2010 update," *Circulation*, vol. 121, no. 7, pp. e46–e215, 2010.

[3] S. Latif, J. Qadir, M. Usman, M. Y. Khan, A. Qayyum, S. M. Ali, Q. H. Abbasi, and M. A. Imran, "Mobile technologies for managing non-communicable diseases in developing countries," *Mobile Applications and Solutions for Social Inclusion*, p. 261, 2018.

[4] Z.-J. Yang, J. Liu, J.-P. Ge, L. Chen, Z.-G. Zhao, and W.-Y. Yang, "Prevalence of cardiovascular disease risk factor in the chinese population: the 2007–2008 china national diabetes and metabolic disorders study," *European heart journal*, vol. 33, no. 2, pp. 213–220, 2011.

[5] S. L. Strunic, F. Rios-Gutiérrez, R. Alba-Flores, G. Nordehn, and S. Bums, "Detection and classification of cardiac murmurs using segmentation techniques and artificial neural networks," in *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*. IEEE, 2007, pp. 397–404.

[6] M. Lam, T. Lee, P. Boey, W. Ng, H. Hey, K. Ho, and P. Cheong, "Factors influencing cardiac auscultation proficiency in physician trainees," *Singapore medical journal*, vol. 46, no. 1, p. 11, 2005.

[7] A. A. H. Care, "How Much Does an EKG Cost?, accessed on: 25-May-2018," http://health.costhelper.com/ecg.html#extres2, Mar. 2018.

[8] I. Maglogiannis, E. Loukis, E. Zafiropoulos, and A. Stasis, "Support vectors machine-based identification of heart valve diseases using heart sounds," *Computer methods and programs in biomedicine*, vol. 95, no. 1, pp. 47–61, 2009.

[9] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, "Phonocardiographic signal analysis method using a modified hidden markov model," *Annals of Biomedical Engineering*, vol. 35, no. 3, pp. 367–374, 2007.

[10] R. SaraçOğLu, "Hidden markov model-based classification of heart valve disease with pca for dimension reduction," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1523–1528, 2012.

[11] L. Avendano-Valencia, J. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Feature extraction from parametric time–frequency representations for heart murmur detection," *Annals of Biomedical Engineering*, vol. 38, no. 8, pp. 2716–2732, 2010.

[12] Y. Zheng, X. Guo, and X. Ding, "A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2710–2721, 2015.

[13] H. Uğuz, "A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases," *Journal of medical systems*, vol. 36, no. 1, pp. 61–72, 2012.

[14] M. Nassralla, Z. El Zein, and H. Hajj, "Classification of normal and abnormal heart sounds," in *Advances in Biomedical Engineering (ICABME), 2017 Fourth International Conference on*. IEEE, 2017, pp. 1–4.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] 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.

[17] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing abnormal heart sounds using deep learning," *arXiv preprint arXiv:1707.04642*, 2017.

[18] 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 *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 621–624.

[19] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[20] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.

[21] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[22] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[23] S. Latif, R. Rana, J. Qadir, and J. Epps, "Variational autoencoders for learning latent representations of speech emotion," *arXiv preprint arXiv:1712.08708*, 2017.

[24] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[25] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional lstm networks for improved phoneme classification and recognition," in *International Conference on Artificial Neural Networks*. Springer, 2005, pp. 799–804.

[26] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning*, 2014, pp. 1764–1772.

[27] R. Lu and Z. Duan, "Bidirectional gru for sound event detection," *Detection and Classification of Acoustic Scenes and Events*, 2017.

[28] O. Deperlioglu, "Classification of phonocardiograms with convolutional neural networks," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 9, no. 2, pp. 22–33, 2018.

[29] 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, 2016.

[30] S. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden markov model," *Physiological measurement*, vol. 31, no. 4, p. 513, 2010.

[31] N. P. Hughes and H. Term, "Probabilistic models for automated ecg interval analysis," Ph.D. dissertation, University of Oxford, 2006.

[32] F. Marzbanrad, Y. Kimura, K. Funamoto, R. Sugibayashi, M. Endo, T. Ito, M. Palaniswami, and A. H. Khandoker, "Automated estimation of fetal cardiac timing events from doppler ultrasound signal using hybrid models," *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1169–1177, 2014.

[33] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

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

[35] M. Zabihi, A. B. Rad, S. Kiranyaz, M. Gabbouj, and A. K. Katsaggelos, “Heart sound anomaly and quality detection using ensemble of neural networks without segmentation,” in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 613–616.

[36] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, “Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients,” in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 813–816.

[37] R. Rana, “Gated recurrent unit (gru) for emotion classification from noisy speech,” *arXiv preprint arXiv:1612.07778*, 2016.

[38] F. Chollet *et al.*, “Keras,” 2015.

[39] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, “Tensorflow: A system for large-scale machine learning.” in *OSDI*, vol. 16, 2016, pp. 265–283.

[40] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[41] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.

[42] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, and T. Wiatowski, “Heart sound classification using deep structured features,” in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 565–568.

[43] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, “Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors,” *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 1, pp. 24–34, 2018.

[44] G. D. Clifford, C. Liu, B. Moody, D. Springer, I. Silva, Q. Li, and R. G. Mark, “Classification of normal/abnormal heart sound recordings: The physionet/computing in cardiology challenge 2016,” in *Computing in Cardiology Conference (CinC), 2016*. IEEE, 2016, pp. 609–612.

[45] D. Roy, J. Sargeant, J. Gray, B. Hoyt, M. Allen, and M. Fleming, “Helping family physicians improve their cardiac auscultation skills with an interactive cd-rom,” *Journal of Continuing Education in the Health Professions*, vol. 22, no. 3, pp. 152–159, 2002.