

RESEARCH

Open Access



# Gated recurrent unit-based heart sound analysis for heart failure screening

Shan Gao<sup>1</sup>, Yineng Zheng<sup>2</sup> and Xingming Guo<sup>1\*</sup>

\*Correspondence:

guoxm@cqu.edu.cn

<sup>1</sup> Key Laboratory of Biorheology Science and Technology, Ministry of Education, College of Bioengineering, Chongqing University, Chongqing 400044, China  
Full list of author information is available at the end of the article

## Abstract

**Background:** Heart failure (HF) is a type of cardiovascular disease caused by abnormal cardiac structure and function. Early screening of HF has important implication for treatment in a timely manner. Heart sound (HS) conveys relevant information related to HF; this study is therefore based on the analysis of HS signals. The objective is to develop an efficient tool to identify subjects of normal, HF with preserved ejection fraction and HF with reduced ejection fraction automatically.

**Methods:** We proposed a novel HF screening framework based on gated recurrent unit (GRU) model in this study. The logistic regression-based hidden semi-Markov model was adopted to segment HS frames. Normalized frames were taken as the input of the proposed model which can automatically learn the deep features and complete the HF screening without de-noising and hand-crafted feature extraction.

**Results:** To evaluate the performance of proposed model, three methods are used for comparison. The results show that the GRU model gives a satisfactory performance with average accuracy of 98.82%, which is better than other comparison models.

**Conclusion:** The proposed GRU model can learn features from HS directly, which means it can be independent of expert knowledge. In addition, the good performance demonstrates the effectiveness of HS analysis for HF early screening.

**Keywords:** Heart sound, Heart failure screening, Deep learning, Gated recurrent unit

## Background

Heart failure (HF) has attracted widespread attentions due to the high morbidity and mortality, especially with the aging of population. The risk indicators of HF are numerous and complicated. Beside the well-known factors, like obesity, smoking and alcohol abuse, some cardiovascular diseases such as hypertension, earlier heart attack and myocardial infarction have also been verified as the precursors for HF developing in clinical practice [1, 2]. Therefore, keeping a healthy lifestyle and paying attention to the early screening of HF play an important role in the preventive and timely treatment.

HF can be divided into two categories—HF with reduced ejection fraction (HFrEF) and HF preserved ejection fraction (HFpEF), and the following conditions are often used to diagnose of HFrEF and HFpEF in clinical [3]: (1) typical symptoms and/or signs of HF; (2) the indicator of left ventricular ejection fraction; (3) the levels of natriuretic peptides;



(4) relevant structural heart disease or diastolic dysfunction. However, these common ways have their own limitations. For instance, the symptoms or signs may be non-specific in the early stages of HF [3], and the invasive measurement [4, 5] is not suitable for promotion among people. The insufficiency in the existing methods prompted us to explore new measures for HF screening.

Nowadays, the non-invasive methods are widely explored for the detection of cardiovascular diseases. For instance, Gao et al. [6, 7] utilized the elasticity-based and a nonlinear state-space approaches to track the motion of carotid artery wall which can be used in the status evaluation of atherosclerotic disease. Many studies used the electrocardiograph signals for cardiac arrhythmia detection [8, 9]; however, the cardiac contractility may not be reflected by electrocardiograph, whose variation is an important sign of HF [10]. Heart sound (HS) can reflect the mechanical dysfunctions of myocardial activity directly, which is a non-stationary physiological signal produced by the beat of muscles [11]. In addition, HS analysis is another non-invasive method. Zheng et al. [12] built a HS-based computer-assisted model in distinguishing HF patients and normal by analyzing the cardiac reserve.

In traditional HS analysis, the feature extraction and/or selection is a crucial step, and various features have been used in HS field, such as wavelet transform [13], wavelet packet transform [14], energy entropy [15] and Mel-frequency cepstral coefficients [16]. These features may be more intuitive to reflect the physical meaning of HS in different states. However, three main limitations also exist: (1) feature extraction and/or selection depends largely on professional knowledge in the fields of medicine and signal processing; (2) extraction of hand-crafted features may miss valuable deep features which contain the latent information of HS; (3) some hand-crafted features are ineffective when the sample quality varies greatly [17]. Deep learning methods, as the new field in machine learning, can learn the features automatically from the inputs without the process of hand-crafted feature extraction and have become popular in the field of biomedical. A convolutional neural network-based transfer learning approach is proposed by Zhang et al. [18] for automatic colorectal cancer diagnosis. Gao et al. [19] proposed a novel deep neural network to learn the implicit strain reconstruction from 2D-radio frequency images and assess the conditions of disease. However, these models have limited ability to mine the features from time-series signals. The improved recurrent neural networks (RNN), including long short-term memory (LSTM) and gated recurrent unit (GRU), can keep the relation of input sequences; therefore, they have been successfully used in sequential data prediction or classification. Yu et al. [20] have adopted the LSTM with attention mechanisms to predict the patient mortality in hospital. Vetek et al. [21] applied LSTM to classify temporal sleep stage using several physiological signals. Similar studies based on EEG were tested by Michielli [22]. Xu et al. [23] reported a LSTM-based architecture for motion-feature extraction from the region of interest sequences. Although RNN-based networks have been extensive used and gained resounding success in biomedical sequence processing, they are barely applied in HS classification.

To address the above issues, we proposed a novel GRU-based method for HF screening using HS. The contributions of this paper lie in: (1) to our best knowledge, this is the first study to distinguish the normal, HFpEF and HFrEF subjects using HS; (2) without heavy reliance on expert knowledge and any hand-crafted features, the proposed

method screens HF utilizing HS signals; (3) the performances show that our method is substantially better than two other deep learning models and one traditional features extraction method. The main framework of this paper is depicted in Fig. 1.

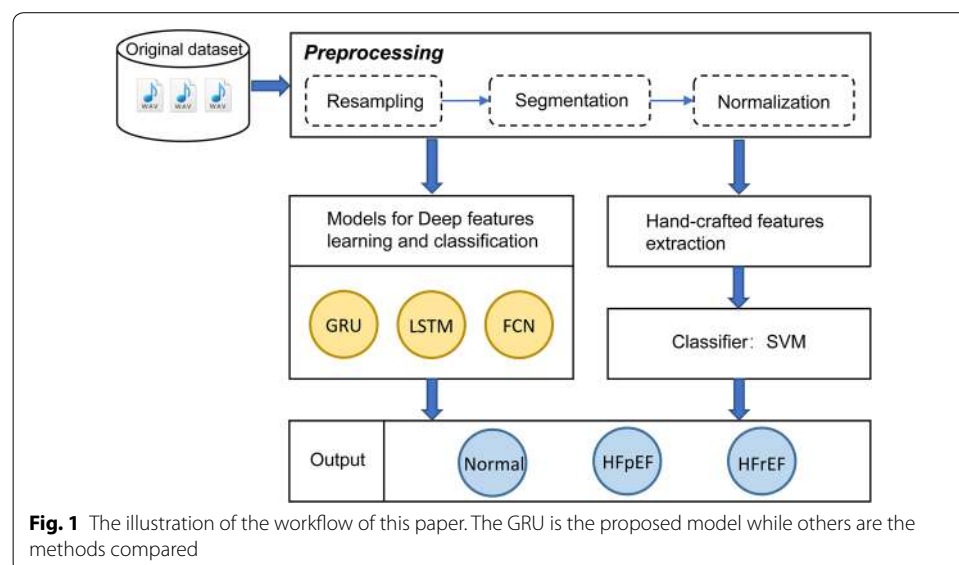
## Results

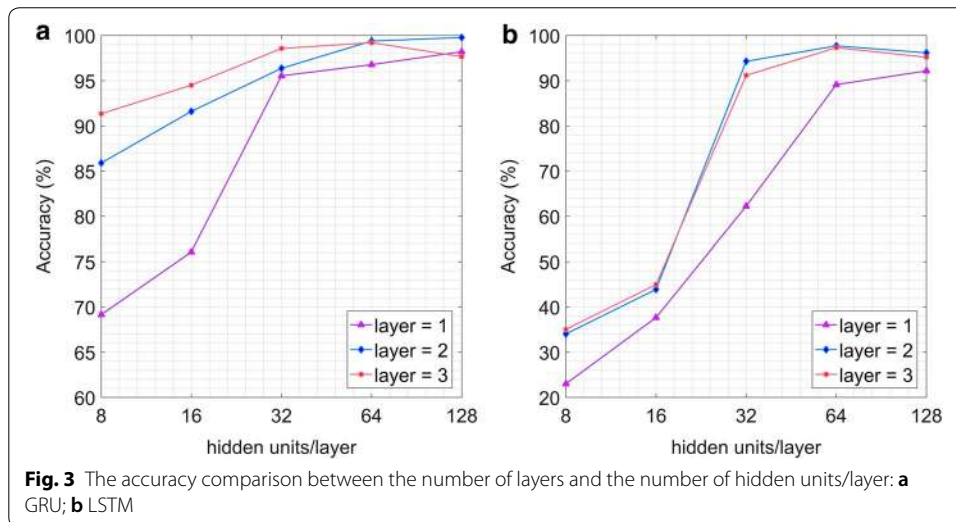
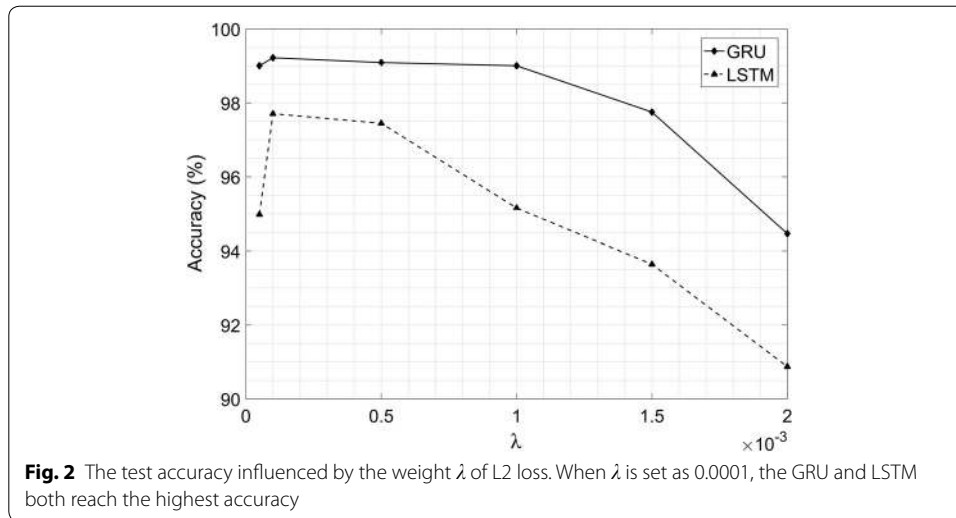
The algorithms of signal preprocessing (resampling, segmentation and normalization), hand-crafted feature extraction and classification with support vector machine (SVM) were all implemented on Matlab (version R2016b) programming. The deep learning models in this work were implemented using python (version 3.5.4) on Tensorflow library (version 1.12.0). The computer used with a 3.7-GHz Intel Core i7-8700 K CPU, GTX 2080Ti GPU with 11 GB video memory and 64 GB RAM to train the networks.

### Model setting experiments

The basic settings of GRU model are determined as follows: Adam is selected as the optimizer and the learning rate is set as 0.001. Softmax cross entropy with logits v2 is chosen as the main loss function. Besides, L2 norm is added in the loss function to prevent model overfitting [24]. The L2 norm of the weight  $\lambda$  for weight decay is calculated by some experiments carefully, and finally set as 0.0001 according to Fig. 2. All the parameters in this paper are trained with the batch size of 64, and the models are trained for 50 epochs in total.

Considering the experimental results about the number of layers and hidden units/layer, the structures of GRU are finally determined. The number of layers varies in {1,2,3}, and the number of units for per layer ranges in {8,16,32,64,128}. As the experimental results show in Fig. 3a, the overall effect of two layers is better than one layer. When the number of units exceeds 64, the performance of three layers is even worse than that of two layers. Considering the complexity of model and the recognition accuracy comprehensively, the GRU structure finally is chosen as two layers with 64 hidden units/layer. Figure 4 shows the final architecture of the GRU network. Moreover, the structure of



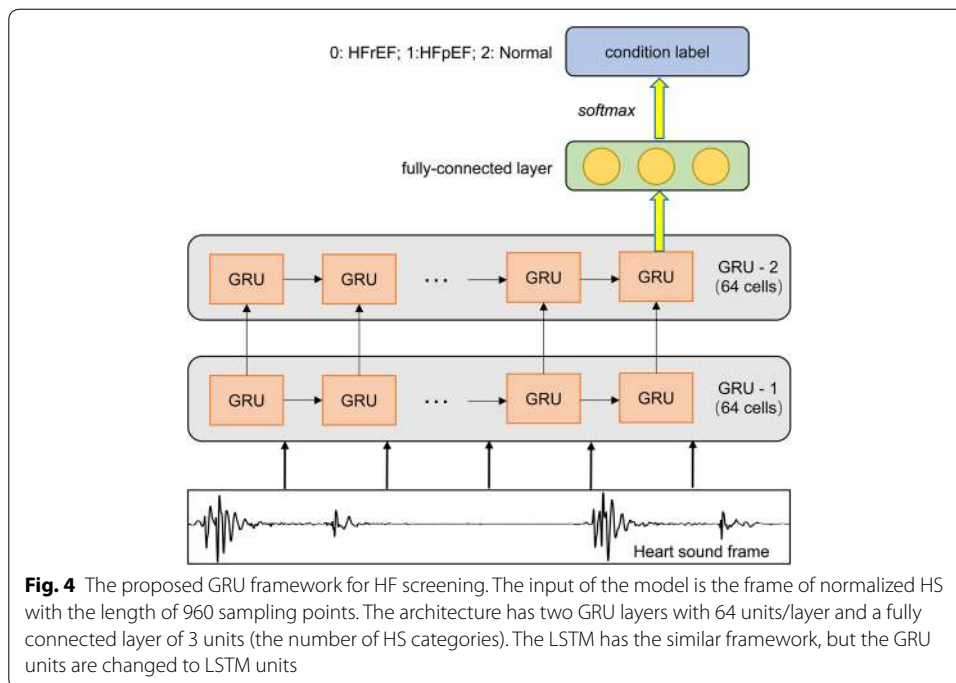


LSTM is defined the same with that of GRU. Figure 3b exemplifies the relevant experimental results of LSTM.

### Screening performance

To evaluate the robustness and to ensure the repeatability of proposed models, the ten-fold cross-validation was used in this work. For each fold, 90% of the HS frames are used for training and the remaining 10% is used to test the performance of our models. To monitor and tune the parameters of training process, 20% frames of the training set are sampled to be used as validation set.

The performance of tenfold cross-validation for all methods is summarized in Table 1. It can be seen that GRU achieves the best average accuracy of 98.82%, which is 2.53%, 4.17% and 11.2% higher than LSTM, fully convolutional network (FCN) and SVM, respectively. SVM is the lowest performing model compared with the other three deep



**Table 1** The tenfold cross-validation results of different models and their average accuracy

Models	Value of accuracy in each fold (%)										Average
	1	2	3	4	5	6	7	8	9	10	
SVM	87.32	<i>89.80</i>	86.05	84.62	89.29	87.27	88.52	89.36	88.46	85.48	87.62
FCN	94.38	<i>97.92</i>	91.61	89.27	95.29	97.36	96.06	96.76	97.10	90.70	94.65
LSTM	96.97	97.02	94.42	95.68	97.15	96.54	96.54	95.16	<i>97.71</i>	95.68	96.29
GRU	99.22	98.92	98.14	97.97	99.09	98.83	99.05	98.53	<i>99.31</i>	99.14	98.82

The best result is highlighted in italics

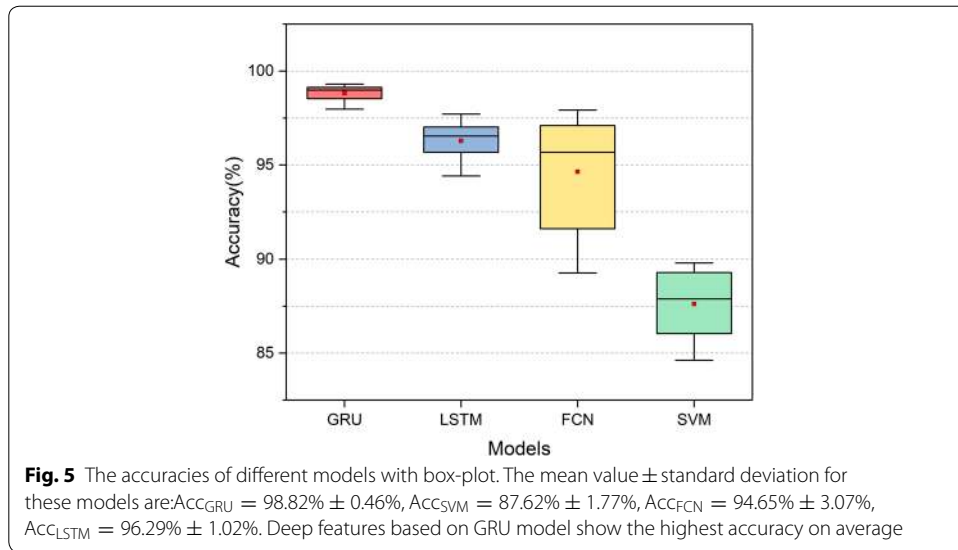
learning models. In addition, the performance of the GRU is more stable as the accuracy deviation is the minimum compared with that of the other three models, which is depicted in the box-plot in Fig. 5.

Table 2 shows the confusion matrix of GRU with all tenfold testing data. The values of precision in three categories are in the range of 98.7–98.93%, and the values of recall are in the range of 98.31–99.46%. It shows that the proposed GRU model can recognize three classes of HS precisely, in which the accuracy of normal class is recognized best. Figure 6 shows an intuitive normalized confusion matrix.

## Discussion

### The impact of the length of frames on classification results

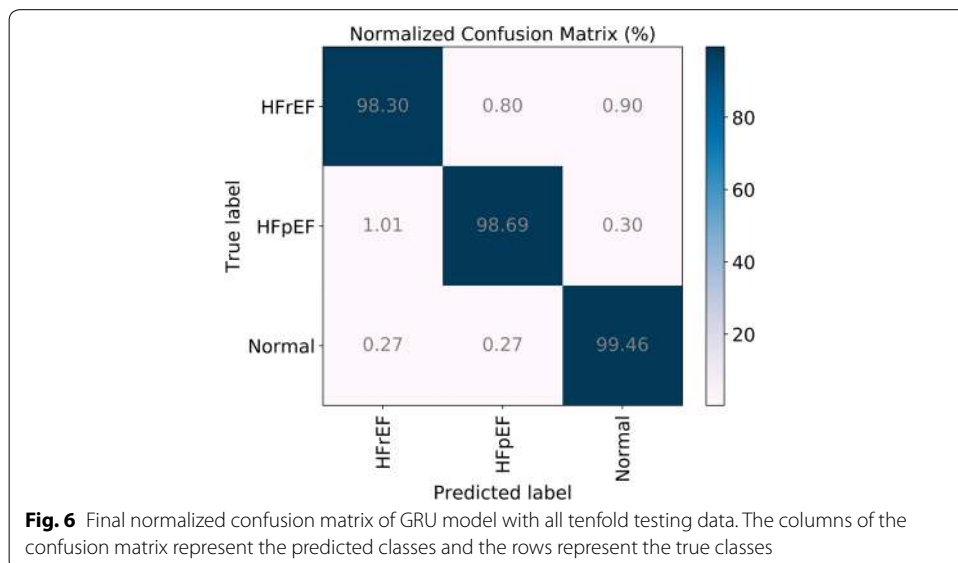
In this paper, the HS signals were segmented to fixed length (1.6 s) frames, and the length of frames might affect the classification stage. To evaluate the possible effect of frame length on final performance, the experiments with fixed length of 0.8 s (approximately one cycle) frames were explored. The corresponding tenfold cross-validation



**Table 2** A confusion matrix of HF for GRU across all tenfold testing data

	HFrEF	HFpEF	Normal	Recall
HFrEF	7540	61	69	98.30%
HFpEF	78	7609	23	98.69%
Normal	21	21	7698	99.46%
Precision	98.70%	98.93%	98.82%	98.82%

The columns represent the predicted categories and the rows represent the true categories



results using the proposed GRU model are listed in Table 3. The results show that the dataset with 1.6 s frames could obtain the average accuracy about 2% higher than 0.8 s frames. The deviation may be caused by the missing of interval features in one cycle frame, which contributes a lot on the classification stage.

**The comparison of the methods used in this study**

In this paper, four models were used to compare the performance for HF screening. GRU and LSTM models are modified kind of RNN architectures. Generally, RNN models can achieve better results than others used in this study. It is because the RNN models can keep the relation of the input time series while others cannot [24]. The results of tenfold cross-validation show that GRU model can achieve higher performance than LSTM model in every attempt of HF screening. Moreover, our comparative experiments have proven that deep learning models outperform the SVM in HF screening. As a representative of traditional knowledge-driven methods, the unsatisfactory results of SVM may be related to the selection of features. Additionally, taking HS signals directly as the input, deep learning models can realize automatic classification without any hand-crafted feature extraction or selection; therefore, our model with fine-tuned parameters can also be applied into other signal processing areas. In sum, the deep learning models can get the higher precision and better performance than traditional SVM, especially the proposed GRU model.

**The comparison of the relevant studies**

Over the years, many studies on screening of HFrEF and HFpEF have been conducted. However, most of the studies were based upon biochemical indicators, phenotype and statistical analysis of medical records information. For instance, Savarese et al. [25] used N-terminal pro-B-type natriuretic peptide to distinguish different HF category. These biochemical indicators are useful to diagnose HF and predict prognosis in HF, but they play a very limited role in the early screening of HF. In addition, such invasive diagnostic methods are not suitable for pervasive application. Xanthopoulos et al. [26] proposed a method to classify the HFpEF based on the phenotype of hypertension, which requires researchers to have a wealth of medical knowledge.

HS signals are closely related to cardiovascular diseases and have been widely studied, while objects of these researches were different. For example, the identification and classification of HS components [27, 28], classification of normal and other abnormal HS [29–31], differentiating the murmurs between physiological and pathological [32, 33]. However, the previously published papers about classification of HFrEF, HFpEF and normal were few and incomplete. Liu et al. [34] explored the difference between HFpEF and

**Table 3 Tenfold cross-validation results of GRU model with two types of frame length**

Frame length	Value of accuracy in each fold (%)										Average
	1	2	3	4	5	6	7	8	9	10	
0.8 s	96.37	97.71	95.37	95.85	97.63	96.19	95.42	96.93	97.92	96.71	96.61
1.6 s	99.22	98.92	98.14	97.97	99.09	98.83	99.05	98.53	99.31	99.14	98.82

The best result is highlighted in italics



normal, but they omitted the study about HFrEF. Zheng et al. [35] reported a HF identification method using HS; however, the HFrEF and HFpEF were not explored separately. It can be seen that the study on HF screening, which included normal, HFpEF and HFrEF, has not been studied sufficiently. Hence, this study could be an efficient complement for HF screening.

#### **The limitations and future work of this study**

This study has three limitations. Firstly, for the lack of HS databases about HFrEF and HFpEF, the experimental tests for generalization ability on other public databases using our method could not be made. Secondly, experimental method was used for the hyper-parameters setting of GRU and LSTM in this study. This method needs to run many experiments to involve approximating optimal value. In the future work, other methods of tuning parameters like grid search may be used in our model to improve the efficiency. In addition, the normal HS may be quite different from that of HF patients, in order to better verify the performance of the proposed method, the abnormal HS with normal systolic and diastolic function can be considered as the control group in the feature.

#### **Conclusions**

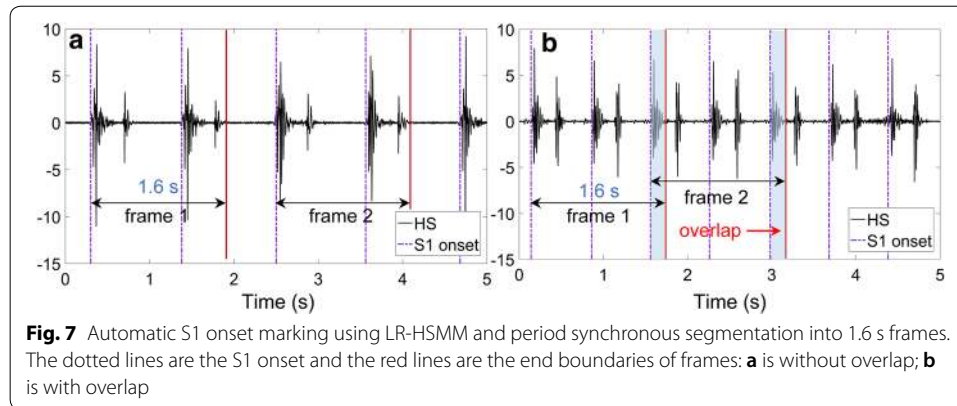
Early screening of HF can provide a timely guide for treatment. In this paper, GRU-based HS analysis method was proposed to screen HF automatically. Taking HS signals as input, the method eliminates the dependence on hand-crafted feature extraction. To verify the screening accuracy, LSTM, FCN and SVM models were carried out as the comparative experiments. The results show that the performance of GRU model is competitive with the methods compared, especially the traditional method of SVM, and it is promising as an effective method for the non-invasive HF screening. In future, the applicability of the method mentioned in this paper will be validated in other cardiovascular diseases, like cardiac murmurs, valvular disease.

#### **Methods**

##### **Experimental data description**

The HS data used in this paper contain three categories—HFrEF, HFpEF and normal. The HS signals of HF patients were acquired from University-Town Hospital of Chongqing Medical University using the HS acquisition system (Patent No.: CN2013093000306700) with the sampling frequency at 11,025 Hz. HF samples were collected from 42 HFrEF and 66 HFpEF patients, respectively. Moreover, all the patients of HFrEF and HFpEF were diagnosed and confirmed by the cardiologists. All patients signed informed consent forms before participating this study, and this study has been ratified by Ethical Commission Chongqing University. The normal HS was obtained from the PhysioNet/Computing in Cardiology Challenge 2016. It contains nine databases from different research groups, and all recordings in the dataset were resampled to 2000 Hz. The dataset includes 2435 normal HS recordings collected from 1297 healthy subjects. Details of the dataset can be referenced in [36, 37]. In this paper, 1286 recordings were randomly selected as the normal group.





### Signal preprocessing

HS preprocessing is an essential part to achieve a good identification performance. In this study, the preprocessing includes three steps introduced as follows.

#### Resampling

In general, HS mainly comprises two components: the first HS (S1) and the second HS (S2). S1 is the transient low-frequency acoustic signals, which is mainly between among 10 and 200 Hz, produced by the vibrations of heart chambers, heart valves and blood in systolic. S2 is produced at the end of systole, following the closure of semilunar valves about aortic and pulmonary [27, 38]. S2 has a higher-pitch than S1, with its frequency range between 20 and 250 Hz [39]. Since the original sampling frequency may cause high computational cost, all recordings are down-sampled at 600 Hz in accordance with Nyquist Sampling Theorem.

#### S1 marking and segmentation

In order to standardize the input length for the model, one strategy was used in this paper to obtain HS frames. Two main steps are involved in this process: marking S1 onset and segmentation HS with fixed frame length.

*Marking S1 onset* Positioning the boundaries of HS components is the critical operation of segmentation. A cardiac period contains four states, namely S1, systole, S2 and diastole. Since S1 is the start of a cardiac cycle, the S1 onset is considered as the boundary of frames.

In this paper, logistic regression-based hidden semi-Markov model (LR-HSMM) is selected to localize the onset of S1. The method of LR-HSMM, developed by Springer et al. [40] and verified by Liu et al. [36], is usually treated as the state-of-the-art method for HS segmentation or marking the onset of cycles, which has great robustness in processing noisy recordings. To preserve more details of HS, the step of signal denoising was skipped in this study. Thanks to the advantages of LR-HSMM, the onset of S1 can be located accurately as shown with the dotted line in Fig. 2.

*Segmentation HS with fixed frame length* The mechanical activity of heart is captured in one cardiac period [41]. Moreover, the interval features may vary between each cycle. In view of these two factors, period synchronous segmentation with the fixed frame length

was applied in this study. The duration of a cardiac cycle is about 0.6–0.8 s, thus the frame length is fixed as 1.6 s, which includes approximately two cardiac cycles. Depicted in Fig. 7a, we segmented the frames with an interval of one cardiac cycle. Whenever the frame length exceeds two periods, overlap is inherent, which is exemplified in Fig. 7b. A total of 23,120 HS frames have been segmented, which, respectively, include the frames of HFrEF, HFpEF and normal are 7670, 7710 and 7740.

### **Normalization**

Normalization is necessary to eliminate the difference of HS amplitude caused by the differences of acquisition locations and individual variation of subjects [15, 16]. All frames used in this paper were normalized by the following formula:

$$X = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (1)$$

### **RNN-based structures**

RNN models, including LSTM and GRU, were used in this work to learn deep features from HS. In this part, some detailed information about the RNN, LSTM and GRU are described as follows.

#### **RNN**

Generally, neural networks assume that inputs and outputs are independent from each other, while many relatedness exist between outputs and previous inputs in reality. Different from other deep learning models, RNN is a network with memory capabilities that can be used to process time sequence data. Hidden layers inputs  $h^{(t)}$  include both the previous hidden output  $h^{(t-1)}$  and the current input  $x^{(t)}$ . It can be expressed as:

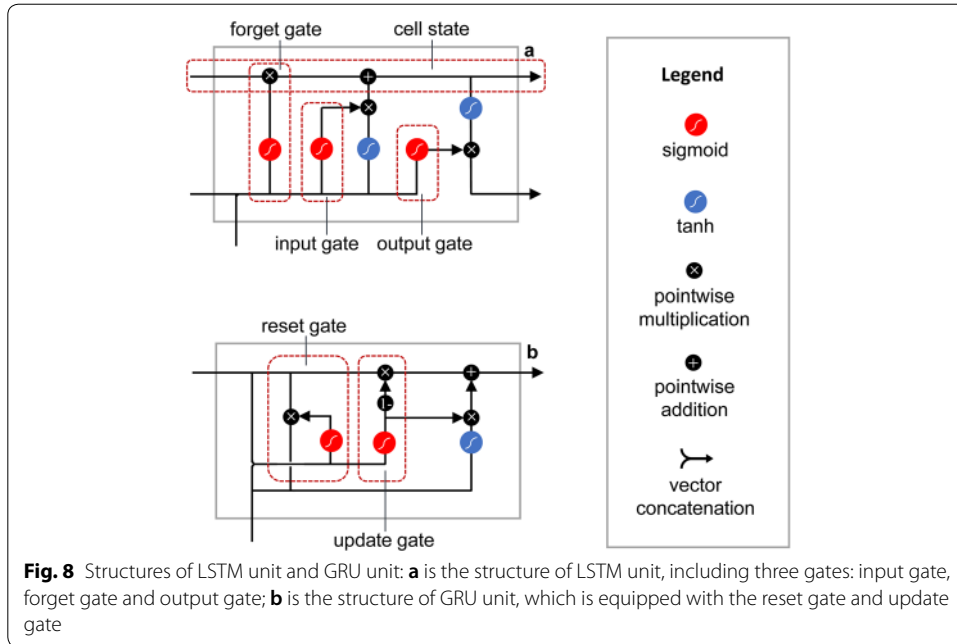
$$h^{(t)} = f(Ux^{(t)} + Wh^{(t-1)} + b), \quad (2)$$

where  $U$ ,  $W$  and  $b$  represent the input weight, hidden unit weight and bias, severally. RNN networks can mine information from arbitrarily long sequences theoretically, but they are limited to just a few steps in practice. For engineering application, LSTM and GRU, the improved RNN networks, are used widely.

#### **LSTM**

As an advanced version of general RNN, LSTM was proposed by Hochreiter and Schmidhuber [42] firstly and improved by Graves [43]. It solved the problem of weight explosion or gradient disappearing due to recursion under long-term time correlation conditions.

The architecture of LSTM contains a cluster of cyclically connected memory cells, and each LSTM unit is equipped with input gate, forget gate and output gate. These gates control the manner of which internal states are retained or discarded. The structure of LSTM unit is shown in Fig. 8a. The algorithm equations of LSTM cell from inputs to outputs are specified as follows:



$$g^{(t)} = \sigma(b_g + U_g x^{(t)} + W_g h^{(t-1)}), \tag{3}$$

$$f^{(t)} = \sigma(b_f + U_f x^{(t)} + W_f h^{(t-1)}), \tag{4}$$

$$o^{(t)} = \sigma(b_o + U_o x^{(t)} + W_o h^{(t-1)}), \tag{5}$$

$$s^{(t)} = f^{(t)} s^{(t-1)} + g^{(t)} \sigma(b + U x^{(t)} + W h^{(t-1)}), \tag{6}$$

$$h^{(t)} = \tanh(s^{(t)}) o^{(t)}, \tag{7}$$

where the  $\sigma$  represents the sigmoid function keeping the weights at 0–1, and  $g^{(t)}$ ,  $f^{(t)}$ ,  $o^{(t)}$ ,  $s^{(t)}$  indicate the external input gate, forget gate, output gate and cell state unit, respectively. The  $b$ ,  $U$  and  $W$  mean the biases, input weights and circular weights, respectively.

Behind the LSTM layers, a fully connected layer with a softmax function is applied for classification. The softmax function is as follows:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_i \exp(x_i)}, \tag{8}$$

where  $x_i$  is the output of former layer.

**GRU**

GRU, a special variant of the LSTM network, was proposed by Cho et al. [44] in 2014. The structure of the GRU is simplified from the LSTM, with two gates, but not separate

memory cell. A single update gate  $z^{(t)}$ , which replaced the input gate and the forget gate in LSTM, is used to estimate the current state of output. Furthermore, the reset gate  $r^{(t)}$  is introduced to control the influence of the previous hidden state on the  $x^{(t)}$  directly. The update gate and reset gate are described as below:

$$z^{(t)} = \sigma(b_z + U_z x^{(t)} + W_z h^{(t-1)}), \quad (9)$$

$$r^{(t)} = \sigma(b_r + U_r x^{(t)} + W_r h^{(t-1)}), \quad (10)$$

and the state of the hidden layer  $h^{(t)}$  is computed as below:

$$h^{(t)} = z^{(t)} h^{(t-1)} + (1 - z^{(t)}) \tilde{h}^{(t)}, \quad (11)$$

where  $\tilde{h}^{(t)} = \tanh(b_h + U_h x^{(t)} + W_h r^{(t)} h^{(t-1)})$ ,  $U, W$  are the weight matrices of different gate referring to the subscripts, and  $b$  represents the bias. Figure 8b gives the structure of GRU unit.

Output states of GRU are calculated using a softmax function (Eq. (8)), which is the same with LSTM.

### Methods compared

**FCN:** FCN with a softmax output layer has been used for time series classification [45]. The model comprised three convolutional blocks with the filter size of 128, 256, 128 and kernel sizes 8, 5, 3, respectively. A batch normalization layer and a ReLU layer are followed by every block. Then the global average pooling layer is added before the softmax layer to reduce the number of weights. The model is trained for 50 epochs with the batch size and learning rate of 64 and 0.001, respectively.

**SVM:** A one-versus-one SVM classifier with radial basis function kernel is adopted. Grid search method is used for parameters tuning. Following Ref. [46], we extracted multiple-type features from HS of HFrEF, HFpEF and normal. Three features with  $P$ -value less than 0.001 in Tamhane's T2 one-way ANOVA are chosen as the feature vector for SVM. To ensure the compactness of this paper, the hand-crafted feature selection and analysis are presented in the "Appendix" at the end of the paper.

**LSTM:** A structure with two layers and 64 hidden units/layer is adopted. The details are explained in the results.

**GRU:** Proposed method.

### Abbreviations

HF: heart failure; HS: heart sound; GRU: gated recurrent unit; LVEF: left ventricular ejection fraction; HFrEF: heart failure with reduced ejection fraction; HFpEF: heart failure with preserved ejection fraction; SVM: support vector machine; RNN: recurrent neural networks; LSTM: long short-term memory; FCN: fully convolutional network; LR-HSMM: logistic regression-based hidden semi-Markov model.

### Acknowledgements

The authors would like to thank National Natural Science Foundation of China for financial support, and the physicians of University-Town Hospital of Chongqing Medical University for professional instructions.

### Authors' contributions

SG, YZ and XG collected the experimental data, reviewed literatures and discussed the method for this study. SG performed the experiments and drafted the manuscript. YZ and XG reviewed and edited the writing. All authors SG, YZ and XG finalized the manuscript for submission. All authors read and approved the final manuscript.

**Funding**

This research was funded by National Natural Science Foundation of China, Grant numbers 31570003, 31870980 and 31800823.

**Availability of data and materials**

The normal HS database is available on PhysioNet. (<https://www.physionet.org/physiobank/database/challenge/2016/>). The HFrEF and HFpEF databases are not publicly available due to the interest of National Natural Science Foundation of China.

**Ethics approval and consent to participate**

All patients signed informed consent forms before participating in this study, and this study has been ratified by Ethical Commission Chongqing University.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup> Key Laboratory of Biorheology Science and Technology, Ministry of Education, College of Bioengineering, Chongqing University, Chongqing 400044, China. <sup>2</sup> Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.

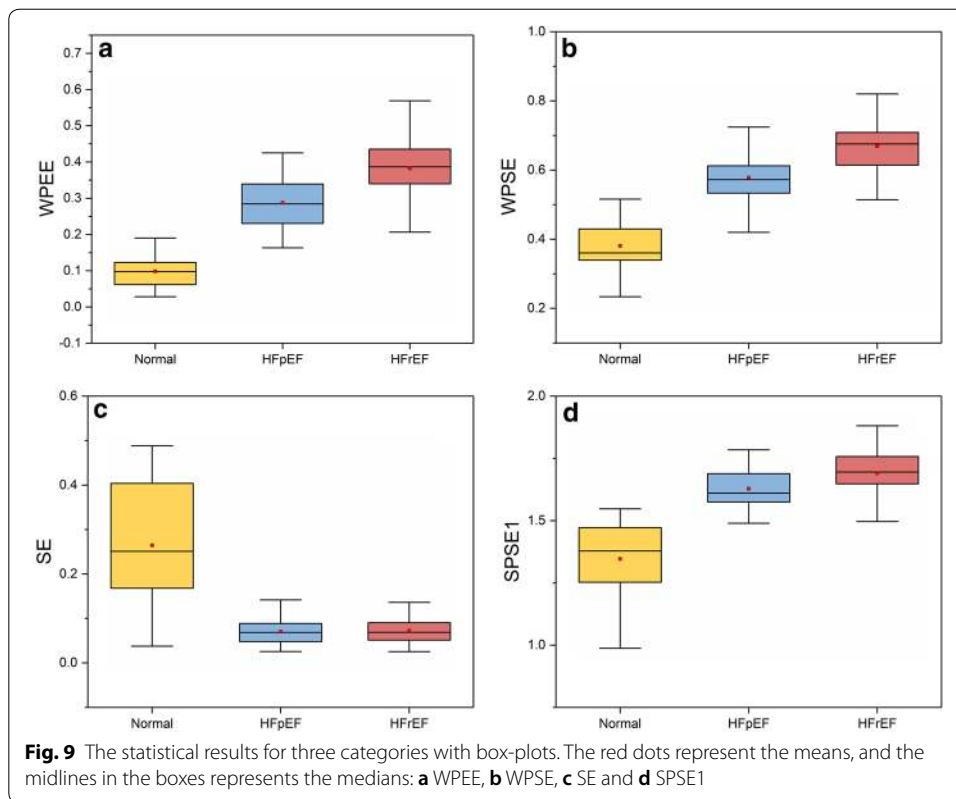
**Appendix**

The hand-crafted features we extracted include wavelet packet energy entropy (WPEE), wavelet packet singular entropy (WPSE), sample entropy (SE) and eight components of sub-band power spectral entropy (SPSE), respectively. For the detailed description of these features, refer to [46]. Tamhane's T2 one-way ANOVA is adopted for multiple comparisons, which is a reliable pairwise comparison based on independent sample *T*-test. The *P* values of extracted features are presented in Table 4, and the *P*-values of WPEE, WPSE and SPSE1 are less than 0.001, indicating that these three features are significantly different among three categories. The SE has the difference between normal and HF groups, but no difference in HF groups. The rest of the features almost have no differences. Therefore, WPEE, WPSE and SPSE1 are finally chosen as the feature vector for SVM.

Figure 9 shows the qualitative results of WPEE, WPSE, SPSE1 and SE using box-plots. The values of WPEE, WPSE, SPSE1 keep the same trends among the three groups, i.e., the normal group is the lowest, while HFrEF group is the highest. These trends indicate the myocardial contractility changes in cardiac energy and information complexity during the development of HF.

**Table 4 The P-values of Tamhane's T2 one-way ANOVA**

Multiple comparison	WPEE	WPSE	SE	SPSE1	SPSE2	SPSE3	SPSE4	SPSE5	SPSE6	SPSE7	SPSE8
HFREF	< 0.001	< 0.001	0.989	< 0.001	0.191	0.169	0.612	0.067	0.294	0.024	0.066
HFREF	< 0.001	< 0.001	< 0.001	< 0.001	0.039	0.387	0.495	0.999	0.006	< 0.001	< 0.001
HFpEF	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.993	0.992	0.029	< 0.001	0.001	0.001



Received: 7 November 2019 Accepted: 6 January 2020  
 Published online: 13 January 2020

## References

- Xu L, Huang X, Ma J, Huang J, Fan Y, Li H, et al. Value of three-dimensional strain parameters for predicting left ventricular remodeling after ST-elevation myocardial infarction. *Int J Cardiovasc Imaging*. 2017;33:663–73.
- Ford I, Robertson M, Komajda M, Böhm M, Borer JS, Tavazzi L, et al. Top ten risk factors for morbidity and mortality in patients with chronic systolic heart failure and elevated heart rate: the SHIFT Risk Model. *Int J Cardiol*. 2015;184:163–9.
- McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. *Eur Heart J*. 2012;33:1787–847.
- Nair N, Gupta S, Collier IX, Gongora E, Vijayaraghavan K. Can microRNAs emerge as biomarkers in distinguishing HFpEF versus HFrEF? *Int J Cardiol*. 2014;175:395–9.
- Faxén UL, Hage C, Benson L, Zabarovskaja S, Andreasson A, Donal E. HFpEF and HFrEF display different phenotypes as assessed by IGF-1 and IGFBP-1. *J Card Fail*. 2017;23:293–303.
- Gao Z, Li Y, Sun Y, Yang J, Xiong H, Zhang H, et al. Motion tracking of the carotid artery wall from ultrasound image sequences: a nonlinear state-space approach. *IEEE Trans Med Imaging*. 2018;37:273–83.
- Gao Z, Xiong H, Liu X, Zhang H, Ghista D, Wu W, et al. Robust estimation of carotid artery wall motion using the elasticity-based state-space approach. *Med Image Anal*. 2017;37:1–21.
- Yıldırım Ö, Plawiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med*. 2018;102:411–20.
- Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inf Sci*. 2017;405:81–90.
- Mabote T, Wong K, Cleland JG. The utility of novel non-invasive technologies for remote hemodynamic monitoring in chronic heart failure. *Expert Rev Cardiovasc Ther*. 2014;12:923–8.
- Hofmann S, Groß V, Dominik A. Recognition of abnormalities in phonocardiograms for computer-assisted diagnosis of heart failures. In: 2016 computing in cardiology conference (CinC), Vancouver, BC, Canada, 11–14 September 2016, vol. 43, p. 561–4. <https://doi.org/10.22489/CinC.2016.161-187>.



12. Zheng Y, Guo X, Qin J, Xiao S. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Comput Methods Programs Biomed.* 2015;122:372–83.
13. Eslamizadeh G, Barati R. Heart murmur detection based on wavelet transformation and a synergy between artificial neural network and modified neighbor annealing methods. *Artif Intell Med.* 2017;78:23–40.
14. Safara F, Doraisamy S, Azman A, Jantan A, Ramaiah ARA. Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Comput Biol Med.* 2013;43:1407–14.
15. Zheng Y, Guo X, Ding X. A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification. *Expert Syst Appl.* 2015;42:2710–21.
16. Chauhan S, Wang P, Lim CS, Anantharaman V. A computer-aided MFCC-based HMM system for automatic auscultation. *Comput Biol Med.* 2008;38:221–33.
17. Gao Z, Chung J, Abdelrazek M, Leung S, Hau WK. Privileged Modality Distillation for Vessel Border Detection in Intracoronary Imaging. *IEEE Trans Med Imaging.* 2019. <https://doi.org/10.1109/TMI.2019.2952939>.
18. Zhang R, Zheng Y, Mak TWC, Yu R, Wong SH, Lau JYW, et al. Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform.* 2017;21:41–7.
19. Gao Z, Wu S, Liu Z, Luo J, Zhang H, Gong M, et al. Learning the implicit strain reconstruction in ultrasound elastography using privileged information. *Med Image Anal.* 2019;58:101534.
20. Yu R, Zheng Y, Zhang R, Jiang Y, Poon CCY. Using a multi-task recurrent neural network with attention mechanisms to predict hospital mortality of patients. *IEEE J Biomed Health Inform.* 2019. <https://doi.org/10.1109/JBHI.2019.2916667>.
21. Vetek A, Muller K, Lindholm H. A compact deep learning network for temporal sleep stage classification. In: 2018 IEEE life sciences conference (LSC). 2018. p. 114–7. <https://doi.org/10.1109/lsc.2018.8572286>.
22. Michielli N, Acharya UR, Molinari F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput Biol Med.* 2019;106:71–81.
23. Xu C, Xu L, Gao Z, Zhao S, Zhang H, Zhang Y, et al. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. *Med Image Anal.* 2018;50:82–94.
24. Zhao Y, Yang R, Chevalier G, Xu X, Zhang Z. Deep residual Bidir-LSTM for human activity recognition using wearable sensors. *Math Probl Eng.* 2018;2018:7316954.
25. Savarese G, Orsini N, Hage C, Vedin O, Cosentino F, Rosano GMC, et al. Utilizing NT-proBNP for eligibility and enrichment in trials in HFpEF, HFmrEF, and HFrEF. *JACC Heart Fail.* 2018;6:246–56.
26. Xanthopoulos A, Triposkiadis F, Starling RC. Heart failure with preserved ejection fraction: classification based upon phenotype is essential for diagnosis and treatment. *Trends Cardiovasc Med.* 2018;28:392–400.
27. Amit G, Gavriely N, Intrator N. Cluster analysis and classification of heart sounds. *Biomed Signal Process Control.* 2009;4:26–36.
28. Giordano N, Knaflitz M. A novel method for measuring the timing of heart sound components through digital phonocardiography. *Sensors.* 2019;19:1868.
29. Ren Z, Cummins N, Pandit V, Han J, Qian K, Schuller B. Learning image-based representations for heart sound classification. In: The 2018 international conference on digital health. 2018. p. 143–7. <https://doi.org/10.1145/3194658.3194671>.
30. Boutana D, Djeddi M, Benidir M. Identification of aortic stenosis and mitral regurgitation by heart sound segmentation on time-frequency domain. In: 5th international symposium on image and signal processing and analysis. 2007. p. 1–6. <https://doi.org/10.1109/ispa.2007.4383654>.
31. Beritelli F, Capizzi G, Lo Sciuto G, Napoli C, Scaglione F. Automatic heart activity diagnosis based on Gram polynomials and probabilistic neural networks. *Biomed Eng Lett.* 2018;8:77–85.
32. Jiang Z, Choi S, Wang H. A new approach on heart murmurs classification with SVM technique. In: 2007 international symposium on information technology convergence. 2007. p. 240–4. <https://doi.org/10.1109/itsic.2007.40>.
33. Sanei S, Ghodsi M, Hassani H. An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *Med Eng Phys.* 2011;33:362–7.
34. Liu Y, Guo X, Zheng Y. An automatic approach using ELM classifier for HFpEF identification based on heart sound characteristics. *J Med Syst.* 2019;43:285.
35. Zheng Y, Guo X. Identification of chronic heart failure using linear and nonlinear analysis of heart sound. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). 2017. p. 4586–9. <https://doi.org/10.1109/embc.2017.8037877>.
36. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. *Physiol Meas.* 2016;37:2181–213.
37. Clifford GD, Liu C, Moody B, Springer D, Silva I, Li Q, et al. Classification of normal/abnormal heart sound recordings: the PhysioNet/computing in cardiology challenge 2016. In: 2016 computing in cardiology conference (CinC), Vancouver, BC, Canada, 11–14 September 2016, vol. 43, p. 609–12. <https://doi.org/10.22489/CinC.2016.179-154>.
38. Tang H, Chen H, Li T. Discrimination of aortic and pulmonary components from the second heart sound using respiratory modulation and measurement of respiratory split. *Appl Sci.* 2017;7:690.
39. Dwivedi AK, Imtiaz SA, Rodriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds—a systematic review. *IEEE Access.* 2019;7:8316–45.
40. Springer DB, Tarassenko L, Clifford GD. Support vector machine hidden semi-Markov model-based heart sound segmentation. In: 2014 computing in cardiology conference, Cambridge, MA, USA, 7–10 September 2014, vol. 41, p. 625–8. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7043120&isnumber=7042955>.
41. Deng SW, Han JQ. Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Gener Comput Syst.* 2016;60:13–21.
42. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9:1735–80.
43. Graves A. Generating sequences with recurrent neural networks. *Comput Sci.* 2013. <http://arxiv.org/abs/1308.0850>.

44. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Eprint Arxiv. 2014. <http://arxiv.org/abs/1412.3555v1>.
45. Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: a strong baseline. In: Proc Int Jt Conf Neural Networks. 2017. p. 1578–85.
46. Li H, Guo X, Zheng Y. An automatic approach of heart failure staging based on heart sound wavelet packet entropy. *J Mech Med Biol* (accepted).

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

