

Deep Wavelets for Heart Sound Classification

Kun Qian*, Zhao Ren[†], Fengquan Dong[‡], Wen-Hsing Lai[§], Björn W. Schuller[¶], and Yoshiharu Yamamoto*

*Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Japan

[†]ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany

[‡]Department of Cardiology, Shenzhen University General Hospital, P.R. China

[§]Department of Computer and Communication Engineering,

National Kaohsiung University of Science and Technology, Taiwan

[¶]GLAM – Group on Language, Audio & Music, Department of Computing, Imperial College London, UK

Email: qian@p.u-tokyo.ac.jp, zhao.ren@informatik.uni-augsburg.de

Abstract—Cardiovascular diseases have a high morbidity, and remain the leading cause of mortality. In the past two decades, developing an intelligent auscultation system has attracted tremendous efforts from the field of signal processing and machine learning. We propose a novel framework based on wavelet representations and deep recurrent neural networks for recognising three heart sounds, i.e., *normal*, *mild*, and *severe*. The Heart Sounds Shenzhen corpus ($n = 170$) is used to validate the proposed method. The experimental results demonstrate the efficacy of the proposed method in a rigorous subject independent scenario, which can reach an unweighted average recall at 43.0 % (chance level: 33.3 %).

Index Terms—Heart Sound, Healthcare, Cardiology, Wavelets, Deep Learning

I. INTRODUCTION

Cardiovascular diseases (CVD), cause 45 % of all deaths in Europe annually [1]. On one hand, as a simple, convenient, and less-expensive method, auscultation is widely used in clinical and medical practice. On the other hand, physicians need tremendous and extensive training for gaining experiences and skills in auscultation [2]. In addition, it was reported that, only approximately 20 % of the medical interns on average can make an efficient use of the stethoscope to measure a subject’s heart status [3]. In the past two decades, plenty of endeavours were made in the field of developing an intelligent auscultation system, which can facilitate an automatic analysis of the heart sound for monitoring the health status of the subject [4]. From a recent literature survey [4] on heart sound classification we can see that, the previous work has achieved encouraging and promising results, which demonstrated the feasibility of using state-of-the-art techniques from signal processing and machine learning to automatically monitor a subject’s health status from the heart sound, or the respective Phonocardiogram (PCG). However, there are still some limitations among the existing

This work was partially supported by the Zhejiang Lab’s International Talent Fund for Young Professionals (Project HANAMI), P.R.China, the JSPS Postdoctoral Fellowship for Research in Japan (ID No. P19081) from the Japan Society for the Promotion of Science (JSPS), Japan, the Grants-in-Aid for Scientific Research (17H00878) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, the Horizon H2020 Marie Skłodowska-Curie Actions Initial Training Network European Training Network (MSCA-ITN-ETN) project under grant agreement No. 766287 (TAPAS), Germany, and the Natural Science Foundation of Shenzhen University General Hospital (No. SUGH2018QD013), P.R. China. The first author would like to thank Ms. Xiao (Lexie) Li for her love and support.

work: Firstly, publicly accessible heart sound databases are extremely limited. The PhysioNet Cinc Challenge database [5] is the biggest one currently. Nevertheless, this database was combined by multiple medical centres, which share inconsistent data acquisition, annotation, and pre-processing methods. Secondly, most of the previous work ignored subject independence, which renders results overoptimistic. Thirdly, advanced signal processing methods, e. g., wavelets combined with deep learning methods were not comprehensively studied. Therefore, we propose a novel framework based on wavelet representations and deep recurrent neural networks (DRNNs) for the task of heart sound classification. To the best of our knowledge, it is the first time to investigate the capacity of combining multi-resolution analysis and deep sequential learning for heart sound classification. In addition, the experiments were implemented in a publicly accessible database, i. e., the heart sounds Shenzhen (HSS) corpus, which makes this study reproducible and comparable. The remainder of this paper will be organised as follows: First, a relation to prior work will be given in Section II. Then, we will introduce the database and methods in Section III. Subsequently, the experimental results are given in Section IV. Finally, we conclude the study in Section V.

II. RELATION TO PRIOR WORK

In our previous studies, wavelets had been found to be efficient for extracting robust representations from body acoustical signals, e. g., snore sounds [6]. In this study, we first introduce wavelet energy features (WEF) into the field of automatically recognising heart sounds. In addition, we investigate the performance of sequential learning by DRNNs, which has been ignored in the aforementioned studies.

III. MATERIALS AND METHODS

A. HSS Database

The HSS corpus was first released in the INTERSPEECH 2018 COMPARE challenge heart beats sub-challenge [7]. This study was approved by the ethics committee of the Shenzhen University General Hospital. There were 170 participants (female: 55, male: 115, 65.4 ± 13.2 years, 21 to 88 years) involved in the data collection (the heart sounds were annotated as *normal*, *mild*, and *severe*). To make a subject independent

TABLE I
THE DATA SET PARTITIONING OF THE HSS CORPUS.

#	Train	Dev	Test	Σ
Normal	84	32	28	144
Mild	276	98	91	465
Severe	142	50	44	236
Σ	502	180	163	845

evaluation, and taking the gender, age, and class distributions into account, we split the whole database into train, development (dev), and test sets (see Table I). All the heart sound recordings were collected with an electronic stethoscope (Eko CORE, USA) set up via a Bluetooth 4.0 at a 4 kHz sampling rate. The average length of the recordings is 30 s (from 29.8 s to 30.2 s), which results in a whole length of approximately 423 min with 845 recordings.

B. Features and Classifier

We use WEF [6] as the low-level descriptors (LLDs) for representing the characteristics of the heart sounds. Compared to the traditional Fourier transformation, wavelet transformation can provide a multi-resolution analysis of the signal [8]. We further chose gated recurrent unit (GRU) cells [9] to overcome the *vanishing gradient* issue when training a DRNN [10]. The WEF based LLDs are extracted from 1 s length frames (with 50% overlap) segmented from one recording, and stacked as a sequence when feeding into the DRNN model.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

We selected ‘coif3’ as the wavelet type for extracting WEF (dimension: 287) from 7 decomposition levels (refer to [11]) by experiments on the development set. The DRNN model was optimised to have three hidden layers (512-256-128) with the ‘Adam’ optimiser. The learning rate, batch size, and iteration number was set to .001, 64, and 500, respectively. To improve the reproducibility of this work, we set the random seed as 12 in the experiments. To evaluate the performance of the proposed model, the unweighted average recall (UAR), i. e., the averaged recall for each class of heart sounds, was used due to the imbalanced distribution of HSS.

B. Results

The confusion matrix of the proposed model on the test set is shown in Fig. 1. Generally speaking, the wavelet based DRNN works for the three-class heart sound recognition task (UAR: 43.0%, Chance Level UAR: 33.3%). We can see that, the model has an excellent performance for recognising the *mild* class. However, its capacity in classifying *normal*, and *severe* needs to be improved. In particular, it is difficult for the current model to distinguish *normal* and *mild*, or *severe* and *mild*.

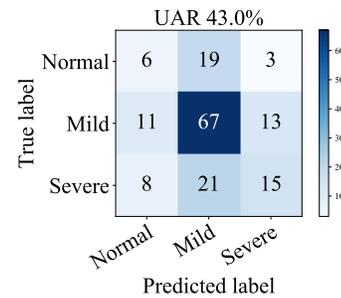


Fig. 1. Confusion Matrix of the Proposed Model on the Test Set.

V. CONCLUSION

In this study, we proposed a novel framework based on wavelet representations and deep recurrent neural networks for classifying heart sounds into *normal*, *mild*, and *severe*. The experiments showed the encouraging and promising performance of the model. In future work, we will involve more sophisticated methods like attention based models [12].

REFERENCES

- [1] E. Wilkins, L. Wilson, K. Wickramasinghe, P. Bhatnagar, J. Leal, R. Luengo-Fernandez, R. Burns, M. Rayner, and N. Townsend, *European Cardiovascular Disease Statistics 2017*. Brussels, Belgium: European Heart Network, 2017.
- [2] 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.
- [3] 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.
- [4] S. Ismail, I. Siddiqi, and U. Akram, “Localization and classification of heart beats in phonocardiography signals: a comprehensive review,” *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, p. 26, 2018.
- [5] C. Liu, D. Springer, Q. Li, B. Moody *et al.*, “An open access database for the evaluation of heart sound algorithms,” *Physiological Measurement*, vol. 37, no. 12, p. 2181, 2016.
- [6] K. Qian, M. Schmitt, C. Janott, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, “A bag of wavelet features for snore sound classification,” *Annals of Biomedical Engineering*, vol. 47, no. 4, pp. 1000–1011, 2019.
- [7] I. Daubechies, A. Steidl, A. Batliner, P. B. Marschik *et al.*, “The INTERSPEECH 2018 Computational Paralinguistics Challenge: Atypical & self-assessed affect, crying & heart beats,” in *Proc INTERSPEECH*, Hyderabad, India, 2018, pp. 122–126.
- [8] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.
- [9] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” in *Proc. NIPS Deep Learning and Representation Learning Workshop*, Montreal, Canada, 2014, pp. 1–9.
- [10] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber *et al.*, “Gradient flow in recurrent nets: The difficulty of learning long-term dependencies,” in *A Field Guide to Dynamical Recurrent Neural Networks*. Piscataway, NJ, USA: IEEE Press, 2001, pp. 237–244, J. F. Kolen, and S. C. Kremer, Ed.
- [11] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, “Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, 2011.
- [12] Z. Ren, Q. Kong, J. Han, M. D. Plumbley, and B. Schuller, “Attention-based atrous convolutional neural networks: Visualisation and understanding perspectives of acoustic scenes,” in *Proc. ICASSP*, Brighton, UK, 2019, pp. 56–60.