Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230525

Recurrent vs Non-Recurrent Convolutional Neural Networks for Heart Sound Classification

Arash GHAREHBAGHI[a,1], Elaheh PARTOVI[b] and Ankica BABIC[a,c]
[a] *Department of Biomedical Engineering, Linköping University, Linköping, Sweden*
[b] *Department of Electrical Engineering, Amir Kabir University, Tehran, Iran*
[c] *Department of Information Science and Media Studies, University of Bergen, Norway*
ORCiD ID: Arash Gharehbaghi https://orcid.org/0000-0002-3413-2859
Ankica Babic https://orcid.org/0000-0002-7532-6828

Abstract. Convolutional Neural Network (CNN) has been widely proposed for different tasks of heart sound analysis. This paper presents the results of a novel study on the performance of a conventional CNN in comparison to the different architectures of recurrent neural networks combined with CNN for the classification task of abnormal-normal heart sounds. The study considers various combinations of parallel and cascaded integration of CNN with Gated Recurrent Network (GRN) as well as Long- Short Term Memory (LSTM) and explores the accuracy and sensitivity of each integration independently, using the Physionet dataset of heart sound recordings. The accuracy of the parallel architecture of LSTM-CNN reached 98.0% outperforming all the combined architectures, with a sensitivity of 87.2%. The conventional CNN offered sensitivity/accuracy of 95.9%/97.3% with far less complexity. Results show that a conventional CNN can appropriately perform and solely employed for the classification of heart sound signals.

Keywords. Heart sound, deep learning, intelligent phonocardiography, convolutional neural network

1. Introduction

The application of Convolutional Neural Network (CNN) has received interest from researchers of the biomedical signal processing domain, especially in the heart sound analysis field in which a reliable classification method for detecting abnormal heart sound from the phonocardiogram is still challengeable [1]. The development of a reliable machine learning method capable of responding to this research challenge can provide a breakthrough in cardiovascular disease detection, thanks even to the Internet of Things (IoT) technological platform. [2,3]. Although sophisticated machine learning methods have been introduced for classification of heart sound signals [4–6], the CNN-based methods which have been broadly employed for this classification task, showed very promising results to be incorporated into the appropriate apparatus to serve as an inexpensive and noninvasive approach for detecting heart abnormalities from the sounds

[1] Corresponding Author: Arash Gharehbaghi, E-mail: arash.gharehbaghi@liu.se

[7–8]. In those CNN-based methods, the dynamics of the signal is preserved at the feature level, where a time-frequency representation of the signal was employed as a mathematical tool for feature extraction. A number of the recent studies on CNN for heart sound classification suggested the use of recurrent architecture of neural network such as Long and Short Term Memory (LSTM), in combination with CNN, in which the signal dynamics are preserved not only at the feature level, but also in the architecture of the LSTM for a better classification [7, 9–10]. However, inconsistent validation manner in terms of the input data makes the comparison flawed. This can negatively affect development of a reliable system for screening heart disease.

This paper presents results of innovative experimentation on various combined architectures of the recurrent neural network and CNN. In this paper, Gated Recurrent Unit (GRU) and LSTM are selected as the recurrent architectures and the accuracy and sensitivity of the various combinations of the recurrent architectures and CNN is evaluated using a public dataset of the heart sound. Such an innovative study can help the researchers of biomedical informatics to select suitable machine-learning method for heart sound classifications.

2. Materials and Methods

2.1. Datasets

We employed a well/known dataset for validation: the Physionet heart sound dataset. This dataset is publicly available and can be found on the website of the Physionet (https://physionet.org/content/challenge-2016/1.0.0/).

2.2. Classification Methods

The heart sound signals are divided into nonoverlapping segments of 5 seconds. Considering the group sizes of the normal and abnormal signals, one can easily see the class imbalance. We employed the SMOTE augmentation method, applied on the signal segments of 5 seconds to provide consistent group size over the two classes: abnormal and normal. Details can be found in [7]. Next, the time-frequency representation of the 5-second signals is obtained using the mel-frequency method [7]. To explore the effect of integration recurrent architectures with CNN, two well-known architectures, LSTM and GRU, are combined with a CNN in two different manners: parallel and cascaded manner. The CNN is independently optimized using Adam optimizer, and the set of the optimized hyperparameters is selected equivalently for all the architectures.

2.3. Validation Methods

The conventional CNN, along with the different combinations of the recurrent architecture and CNN, are independently validated. Each combination composes a classifier whose performance is validated and compared to the CNN's. The classifiers are named and denoted by: Serial GRU (SGRU), parallel GRU (PGRU), serial LSTM (SLSTM), and Parallel LSTM (PLSTM). The accuracy and sensitivity of the classifiers are invoked as the performance measures. We employed the Physionet dataset for the classifier validation using 75%, 10%, and 15% of the data as the training, validation and test data, respectively. Each classifier is examined in 5 complete runs of train/validation/test, and the statistics of the accuracy and the sensitivity are obtained.

3. Results

Table 1 demonstrates the two descriptive statistics: the average value, and the standard deviation (STD) of the performance measures, resulted from 5 complete runs of the train/validation/test split, as applied to the abovementioned classifiers.

Table 1. Average value (Mean) and standard deviation (STD) of the accuracy and sensitivity of the CNN, the Serial GRU (SGRU), parallel GRU (PGRU), serial LSTM (SLSTM), and Parallel LSTM (PLSTM)

Classifier	Accuracy (%)		Sensitive (%)	
	Mean	STD	Mean	STD
CNN	95.9	4.9	97.3	3.7
PGRU	97.9	0.9	93.8	4.9
PLSTM	98.0	0.5	87.2	10.1
SGRU	95.6	4.8	95.0	5.0
SLSTM	95.2	3.5	93.8	3.4

The parallel CNN-LSTM shows superior accuracy with excellent reproducibility, as implied by the very low standard deviation of 0.5%. However, this is not true when it comes to sensitivity, which is the worst for this architecture. The serial architecture of CNN-LSTM exhibits a better performance in terms of the balance between accuracy and sensitivity. The serial architecture of CNN-GRU performs better in this sense, but with higher instability as reflected by the standard deviations. The serial CNN-LSTM delivers further stability in its results. It is obvious that the performance of a conventional CNN is degraded while either of the recurrent architectures is incorporated serially. The parallel integration of both recurrent architectures enhances accuracy at the expense of sensitivity. Serial integration degraded the performance of a CNN in terms of both performance measures. A crucial requirement of any screening approach is a high value of sensitivity which is impaired by the integration of any of the recurrent architecture.

4. Discussion

This paper innovatively explored the effect of the integration of different recurrent architectures with a CNN on detecting an abnormal heart from the heart sound. This approach, which was previously named intelligent phonocardiography, showed its potential to be employed as a screening tool in clinical settings [11–13]. The paper investigated different combinations of serial and parallel integration of two well-known recurrent methods: LSTM and GRU. Integration of recurrent architecture, in any of the described forms, effectively degraded the sensitivity, which is considered a key feature of any screening tool. Even though the accuracy was slightly improved in some of the architectures, the heavy load of the complexity inflicted by the recurrent architecture cannot justify the use of these architectures compared to the described CNN. In this study, the mel-frequency representation of the segmented signal with a 5-second length was employed to be used as the mathematical tool whose energy contents constitute the input features to the CNN. The dynamics of the signal is, therefore, preserved by this representation. Results show that such a representation sufficiently considers the signal dynamic and avoids the necessity of using other dynamic architecture. Nevertheless, the high standard deviation of architecture for some of the performance measures implies the high structural risk for the architecture that requires sophisticated methods for the evaluation [14]. Noise and artifact can randomly affect the classification performance which is regarded as the study limitation. This can be addressed in continuation to this study as the future work.

5. Conclusions

The paper innovatively explored the performance of a CNN in the presence of a recurrent neural network architecture for the classification of abnormal heart sound. Serial and parallel integration of an LSTM and a GRU with the CNN was explored. It is concluded by this study that increasing complexity of a classification method is not necessarily accompanied by an improvement in the performance.

References

[1] Gharehbaghi A, Partovi E. Accuracy of a Deep Learning method for heart sound analysis is unrealistic. Neural Networks: the Official Journal of the International Neural Network Society. 2022 Dec 13;159:107-8.

[2] Watrous RL, Thompson WR, Ackerman SJ. The impact of computer-assisted auscultation on physician referrals of asymptomatic patients with heart murmurs. Clinical Cardiology. 2008 Feb;31(2):79-83.

[3] Gharehbaghi A, Lindén M. An Internet-Based Tool for Pediatric Cardiac Disease Diagnosis Using Intelligent Phonocardiography. IoT 360°; 2015 October 27-29; Rome, Italy; Revised Selected Papers, Part I. Cham: Springer International Publishing; 2016. p. 443-7.

[4] Gharehbaghi A, Sepehri AA, Lindén M, Babic A. Intelligent Phonocardiography for Screening Ventricular Septal Defect Using Time Growing Neural Network. Informatics Empowers Healthcare Transformation: IOS Press; 2017. 238:108-11.

[5] Gharehbaghi A, Lindén M, Babic A. An Artificial Intelligent-Based Model for Detecting Systolic Pathological Patterns of Phonocardiogram based on Time-Growing Neural Network. Applied Soft Computing, 2019. 83:105615.

[6] Gharehbaghi A, Lindén M. A Deep Machine Learning Method for Classifying Cyclic Time Series of Biological Signals Using Time-Growing Neural Network. IEEE Transactions on Neural Networks and Learning Systems. 2018; 29(9) 4102-4115.

[7] Deng M, Meng T, Cao J, Wang S, Zhang J, Fan H. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Netw. 2020 Oct;130:22-32.

[8] Gharehbaghi A, Babic A. Deep Time Growing Neural Network vs Convolutional Neural Network for Intelligent Phonocardiography. Stud Health Technol Inform. 2022 Jun 29;295:491-494. doi: 10.3233/SHTI220772. PMID: 35773918.

[9] Al-Issa Y, Alqudah AM. A lightweight hybrid deep learning system for cardiac valvular disease classification. Scientific Reports. 2022 Aug 22;12(1):14297.

[10] Chen Y, Sun Y, Lv J, Jia B, Huang X. End-to-end heart sound segmentation using deep convolutional recurrent network. Complex & Intelligent Systems. 2021 Aug;7:2103-17.

[11] Gharehbaghi A, Sepehri AA, Lindén M, Babic A, A Hybrid Machine Learning Method for Detecting Cardiac Ejection Murmurs. EMBEC & NBC 2017; 787-790.

[12] DeGroff CG, Bhatikar S, Hertzberg J, Shandas R, Valdes-Cruz L, Mahajan RL. Artificial neural network–based method of screening heart murmurs in children. Circulation. 2001 Jun 5;103(22):2711-6.

[13] Singh SA, Meitei TG, Majumder S. Short pcg classification based on deep learning, in: Deep Learning Techniques for Biomedical and Health Informatics. Elsevier, pp. 141–164

[14] Gharehbaghi A, Babic A. Structural Risk Evaluation of a Deep Neural Network and a Markov Model in Extracting Medical Information from Phonocardiography. Studies in health technology and informatics, 2018, 251:157-160.