Basic Research for High-speed Heart Sound Determination using AI

Riku Nakashima
Faculty of Engineering, University of Miyazaki,
1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Praveen Nuwantha Gunaratne
Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Hiroki Tamura
Faculty of Engineering, University of Miyazaki,
1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Email: hi20031@student.miyazaki-u.ac.jp, ti20060@student.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

In recent years, many researchers have focused on enhancing the current situation of home healthcare systems. In this paper, we developed a system that uses AI to quickly determine whether auscultation sounds are normal or not. In our analysis, heart sounds were imaged, and their abnormalities were identified using machine learning tools. The proposed approach uses a Yolo_v7 model-based anomaly detection and the Wavelet Transform was employed to analyze the acquired heart sound data. Proposed system resulted with an 90% accuracy rate or over on the basis of 5 seconds of heartbeat data, in various recording environments.

Keywords: Auscultation sounds, abnormality identification, YOLO_v7, Wavelet Transform

1. Introduction

In today's society, the demand for home medical care is rapidly increasing due to the aging population. However, home medical care requires a physician to visit the patient in person to make a diagnosis, which poses problems such as securing time for physicians. A possible solution to this problem would be to make medical equipment readily accessible to ordinary households, enabling individuals to personally use and evaluate the effectiveness of the equipment. However, as proficiency in medical knowledge is crucial for operating medical equipment, the current situation hinders its widespread utilization. To identify an approach suitable for the general public, eliminating the necessity for a visit to a physician or medical expertise, our team has attempted to create a system capable of swiftly assessing the normality of heart sound data obtained through auscultation.

In this paper, machine learning tools were employed to analyze the heart sound data. Heart sound data from diverse recording environments was utilized in this paper. The acquired heart sound data was transformed into images through wavelet transforms, and the yolov7 model was employed for abnormality detection. Additionally, a *k*-fold approach was applied to a limited dataset to enhance accuracy. The findings revealed a 90% accuracy rate based on 5 seconds of heartbeat data. This paper compares Yolo model with different durations of phonocardiography data and elucidates strategies for achieving higher confidence scores.

2. Methodology

Initially the Speech data, illustrating frequency features of heart sounds, went through wavelet transformation at various intervals to generate images. Afterwards, the transformed images were utilized to train multiple Yolo models, facilitating a comparative analysis to identify the superior model. Here, the assessment was conducted based on confidence scores assigned to the normal and abnormal classes. For the Yolo training data set, the images were divided into three parts: training data, valid data, and test data. Table 1 illustrates the learning environment used in this analysis, and the Fig. 1 depicts the experimental procedure in brief.

3. Experiment

In the section 3.1, the ratio of training data to validation data and test data was set at 7:2:1 in the Yolo model. The subsequent phases, starting from section 3.2, the ratio was adjusted to 6:2:2. The offline datasets used were heart sound data auscultated by Sono-Support clinical technologists (dataA – 10 healthy subjects and 40 sick subjects), heart sound data from former laboratory members (dataC - 23 healthy subjects), and heart sound data of few laboratory members (jibun - 2 healthy subjects). The data from physioNet (dataB~dataB_F), comprising 2558 healthy subjects, 665 sick subjects was used as online data [1].

Table 1. Learning environment of the YOLO model

Learning environment	
Memory	32GB
CPU	Intel(R) Core(TM) i7 – 10700 CPU@2.90GHz
GPU	NVIDIA GeForce RTX 3070
CUDA	11.2

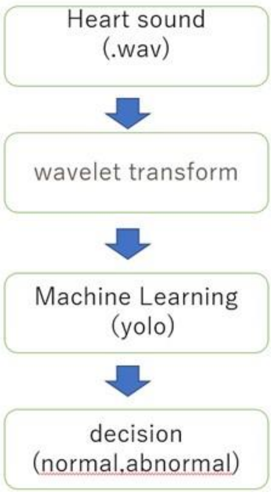


Fig. 1. Experimental Procedure

Fig. 2, 3, and 4 illustrate the images generated by the wavelet transform used in this paper. The vertical axis represents frequency, the horizontal axis signifies a specific time width of the experimental data, and the color bar indicates amplitude. The wavelet transform is output as an image after being processed using the Morlet function with MATLAB. A one second shift was used to increase the number of data by shifting the image by few data. Due to the variation in the number of seconds in heartbeat, the number of images produced varies as well. The *k*-fold method involves creating *k* splits of the training data images and training them *k* times (*k*=5 in this case). The splits were randomly divided into 8 for the training data and 2 for the validation data. Those with a confidence score of less than 0.1 between normal and abnormal were excluded due to their classification falling within gray zones as lacking in reliability. The formulas for calculating the percentage of correct answers Eq. (1) and the percentage of target hits Eq. (2) are presented below.

$$Percentage\ of\ correct\ answers = \frac{Number\ of\ correct\ images}{Total\ number\ of\ images} \times 100 \quad (1)$$

$$Percentage\ of\ target\ hits = \frac{true\ positive}{true\ positive + false\ negative} \times 100 \quad (2)$$

3.1. Experiment 1

In experiment 1, multiple Yolo models (Yolo_v2, Yolo_v5, and Yolo_v7) were employed to jointly train normal and abnormal heart sound images and assess the superiority of each model.[2], [3].

3.2. Experiment 2

Multiple Yolo models with favorable confidence scores were employed, and machine learning was conducted on each of them for varying durations (3, 5, and 10 seconds), comparing their performance in terms of the percentage of correct responses. However, a 10-second wavelet transform could not be generated due to insufficient seconds in imaging dataB_B. Fig. 3, 4 was generated by performing a wavelet transform on a single speech sample, cut out for a fixed time width of 1 second intervals and converting it to an image.

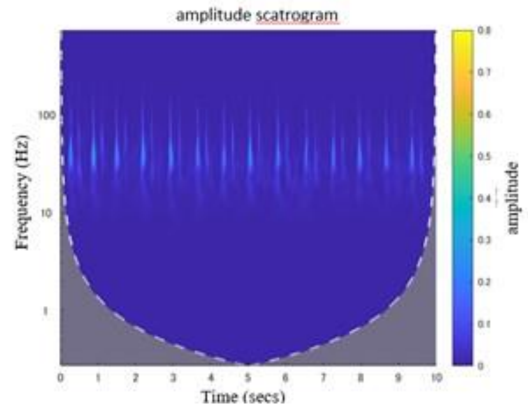


Fig. 2. Wavelet transform 3 sec.

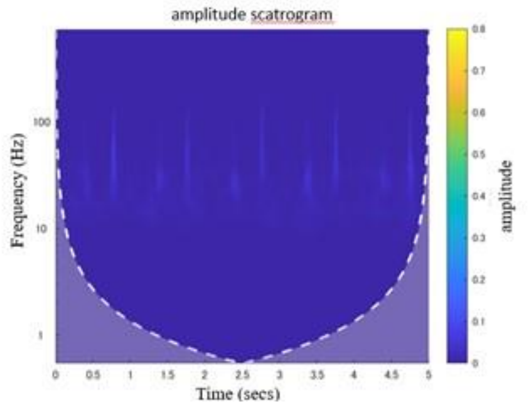


Fig. 3. Wavelet transform 5 sec.

3.3. *Experiment 3*

In Experiment 3, the accuracy of the 5-second wavelet transforms, identified as the best on real data in Section 3.2. was enhanced by employing the *k*-fold method. This method involved covering a small data set and increasing the number of training patterns.

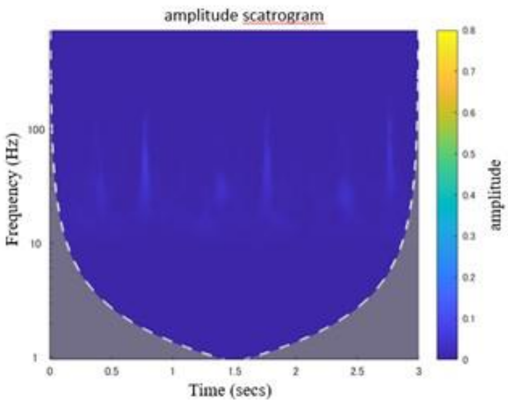


Fig. 4. Wavelet transform 10 sec.

4. Results

4.1. Results of Experiment 1

The correctness of machine learning responses for each Yolo was determined based on the agreement between the label and the test judgment. A comparison of the mean confidence scores of the correct answers revealed the superiority of Yolo_v7, as illustrated in Table 2.

Table 2. Confidence score results of each YOLO

	Yolov2	Yolov5	Yolov7
Confidence Score	0.957	0.973	0.997

4.2. Results of Experiment 2

Machine learning with Yolo_v7 was carried forward, as it came out as the best among the other Yolo models based on the results of Section 4.1. Consequently, the correct response rates were as follows: 3 seconds, 90.6% for normal and 89.2% for abnormal (as shown in Table 3); 5 seconds, 86.8% for normal and 90.6% for abnormal (Table 4); and in 10 seconds, 68.7% for normal and 94.4% for abnormal (Table 5). These three results demonstrate that the 3-second cutout yields the highest average correct response rate. However, upon comparing with actual data (dataA), the 5-second wavelet transform exhibits the best results. While deciding which result to adopt presents a doubt, move forwarded with *k*-fold using the 5-second wavelet transform, as it yields the best results on real data.

4.3. Results of Experiment 3

The outcomes obtained through *k*-fold training on a 5-second wavelet transform revealed accuracy rates of 90.3% on normal and 93.6% on abnormal, which are over 90%.

5. Conclusion

Yolo_v7 enhanced the accuracy to 90% in diverse recording environments, achieved by optimizing a limited dataset and enhancing learning patterns through *k*-fold techniques. Future research will focus on strategies to enhance accuracy by expanding the dataset. Further, considering the satisfactory average accuracy of the 3-second wavelet transform, experiment with *k*-fold on the 3-second version will be continued.

Table. 3. Wavelet transform 3-second correct response rate

3s	data A	data B	data B_B	data B_C	data B_D	data B_E	data B_F	data C	jibun	Aver age
Nor mal	84.1	67.6	80.6	89.8	100	100	100	100	100	92.4
Abno rmal	80	89.4	95.2	100	100	99.3	100			94.8

Table. 4. Wavelet transform 5-second correct response rate

5s	data A	data B	data B_B	data B_C	data B_D	data B_E	data B_F	data C	jibun	aver age
Nor mal	84.2	93.5	91.9	100	35.9	100	75.4	100	100	86.8
abno rmal	85.6	80	100	100	95	99.8	73.9			90.6

Table. 5. Wavelet transform 10-second correct response rate

10s	data A	data B	data B_C	data B_D	data B_E	data B_F	data C	jibun	Aver age
Nor mal	81.8	22	19	33	94	100	100	100	68.7
abno rmal	76	100	98.5	100	100	91.9			94.4

References

1. “ Classification of Heart Sound Recordings: The PhysioNet/Computing in Cardiology Challenge 2016 v1.0.0,” physionet.org. https://physionet.org/content/challenge-2016/1.0.0/
2. “Yolov7”(Japanese) Qiita, Oct. 10, 2022. https://qiita.com/johnrocky/items/a4c24e55acaf7dcccdd9 (accessed Dec. 01, 2023).
3. “YOLOv5”(Japanese) Qiita, Oct. 27, 2021. https://qiita.com/gotty21/items/37b52cec9542fd87bd5f (accessed Dec. 01, 2023)

Authors Introduction

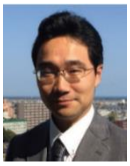
Mr. Riku Nakashima

He was born in 2001. He is currently studying indepartment of environmental robotics, and will receive the B.Eng from University of Miyazaki in 2024. His current research is Basic Research for High-speed Heart Sound Determination using AI.

Mr. Praveen Nuwantha Gunaratne

He received his Bachelor's degree in Engineering in 2018 from the Faculty of Engineering, University of Moratuwa, Sri Lanka. He is currently a Doctoral student in University of Miyazaki, Japan

Prof. Hiroki Tamura



He received the B.E. and M.E. degree from Miyazaki University in 1998 and 2000, respectively. From 2000 to 2001, he was an Engineer in Asahi Kasei Corporation, Japan. In 2001, he joined Toyama University, Toyama, Japan, where he was a Technical Official in the Department of Intellectual Information Systems. In 2006, he joined Miyazaki University, Miyazaki, Japan, where he was an Assistant Professor in the Department of Electrical and Electronic Engineering. Since 2015, he is currently a Professor in the Department of Environmental Robotics. His main research interests are Neural Networks and Optimization Problems. In recent years, he has had interest in Biomedical Signal Processing using Soft Computing.