

Clinical decision support system using a machine learning model to assist simultaneous cardiopulmonary auscultation: Open-label randomized controlled trial

DIGITAL HEALTH
Volume 10: 1–10
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241233689
journals.sagepub.com/home/dhj



Takanobu Hirosawa¹ , Tetsu Sakamoto¹, Yukinori Harada¹, Kazuki Tokumasu² and Taro Shimizu¹

Abstract

Background: The utility of a clinical decision support system using a machine learning (ML) model for simultaneous cardiac and pulmonary auscultation is unknown.

Objective: This study aimed to develop and evaluate an ML system's utility for cardiopulmonary auscultation.

Methods: First, we developed an ML system for cardiopulmonary auscultation, using cardiopulmonary sound files from our previous study. The technique involved pre-processing, feature extraction, and classification through several neural network layers. After integration, the output class was categorized as “normal,” “abnormal,” or “undetermined.” Second, we evaluated the ML system with 24 junior residents in an open-label randomized controlled trial at a university hospital. Participants were randomly assigned to the ML system group (intervention) or conventional auscultation group (control). During training, participants listened to four cardiac and four pulmonary sounds, all of which were correctly classified. Then, participants classified a series of 16 simultaneous cardiopulmonary sounds. The control group auscultated the sounds using noise-cancelling headphones, while the intervention group did so by watching recommendations from the ML system.

Results: The total scores for correctly identified normal or abnormal cardiopulmonary sounds in the intervention group were significantly higher than those in the control group (366/384 [95.3%] vs. 343/384 [89.3%], $P = 0.003$). The cardiac test score in the intervention group was better (111/192 [57.8%] vs. 90/192 [46.9%], $P = 0.04$); there was no significant difference in pulmonary auscultation.

Conclusions: The ML-based system improved the accuracy of cardiopulmonary auscultation for junior residents. This result suggests that the system can assist early-career physicians in accurate screening.

Keywords

Telemedicine, artificial intelligence, clinical medicine, physical examination

Submission date: 9 August 2023; Acceptance date: 1 February 2024

Introduction

Cardiopulmonary disease is the leading cause of mortality worldwide. Auscultation serves as a powerful and convenient screening tool to detect cardiopulmonary diseases.^{1,2} Following the outbreak of the coronavirus pandemic (COVID-19), the global demand for telemedicine increased to facilitate contactless healthcare services.³ While telemedicine has the potential to enhance health outcomes, it

¹Department of Diagnostic and Generalist Medicine, Dokkyo Medical University, Tochigi, Japan

²Department of General Medicine, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan

Corresponding author:

Takanobu Hirosawa, Department of Diagnostic and Generalist Medicine, Dokkyo Medical University, 880 Kitakobayashi, Mibu-cho, Shimotsuga, Tochigi 321-0293, Japan.

Email: hirosawa@dokkyomed.ac.jp



requires more efficient and effective technologies.⁴ Several strategies, including telemedical palpation,⁵ telemedical percussion,⁶ telemedical patient-assisted comprehensive clinical examination,⁷ and remote auscultation, have been developed to address the lack of physical examinations in telehealth services.

Remote auscultation, particularly through electronic stethoscope systems, plays a critical role in telemedical physical examinations.^{8–14} Our previous study found that conventional auscultation was superior to that of real-time remote cardiac auscultation.¹³ Thus, an appropriate support system is required to improve remote cardiac auscultation.

Remote auscultation systems,^{11,14} including diagnostic clinical decision support systems that use automated classification,¹⁵ have been developed. Recent advances in computing have enabled diagnostic support systems that utilize sound analysis.¹⁶ Sound analysis comprises three major steps, that is pre-processing, feature extraction, and classification.^{17,18} First, the pre-processing step handles noise processing or signal-frame decomposition. Then, to extract the characteristics of the original signals, feature extraction for time and/or frequency domain features is performed. Finally, classification is performed to identify the classification rules. Artificial neural networks, which imitate the human brain, are the most widely used machine learning (ML)-based approach.¹⁸ Recently, such methods have exhibited good reliability in various fields,^{19–21} which is due in part to the sophistication of algorithms and the improvements in calculation.

Despite these advancements, the development of an ML-based clinical decision support system for simultaneous cardiac and pulmonary auscultation remains understudied. Most studies have focused either on cardiac sounds,^{10,12,16,18} including specific valvular disease¹¹ or on pulmonary sounds.^{22,23} The overlapping nature of these sounds on a spectrum chart,²⁴ and the absence of a suitable auscultation site on the human body,²⁵ are significant challenges. Existing research²⁴ primarily targets reducing heart sounds from lung sounds but does not achieve simultaneous analysis. Furthermore, another study²⁵ has focused on the dataset classification but has not evaluated the impact on the diagnostic skills of physicians, including junior resident doctors. In contrast, our previous studies explored the diagnostic skills in remote auscultation of physicians via Bluetooth²⁶ and internet-connections¹³ but did not address simultaneous cardiopulmonary auscultation.

Addressing these challenges and meeting the high demand for ML-based clinical decision support systems capable of cardiopulmonary auscultation,¹⁴ we have undertaken this study. Our goal is to develop an ML-based system that not only enhances diagnostic accuracy in cardiopulmonary auscultation but also evaluates its effectiveness in improving the diagnostic skills of junior resident doctors.

Materials and methods

Overview

This study involved development and evaluation of an ML system for cardiopulmonary auscultation. A unique aspect of our methodology was not only the development of the ML system but also included an evaluation study involving junior resident doctors. The purpose was to assess the system's practical impact on enhancing diagnostic accuracy in cardiopulmonary auscultation. The evaluation provided valuable insights into the real-world application and effectiveness of our ML system in a clinical setting.

Sample collection and preparation

In this study, we used a sample of cardiopulmonary sound files from our previous study,^{13,26} which evaluated the utility of real-time remote cardiopulmonary auscultation through a simulator-based approach. To ensure data validity, we have primarily selected sound files from cardiac patient simulators and pulmonary auscultation simulators. This selection was based on their proven reliability in our previous studies,^{13,26} which demonstrated the utility of these simulators in accurately replicating cardiopulmonary sounds. These files, recorded using a state-of-the-art electronic stethoscope (JPES-01; MEMS CORE CO., Ltd), ensure the clarity and quality of the auscultation sounds. The sound files were captured in Waveform Audio File format (WAVE) at a sampling rate of 44 100 Hz. The cardiac sounds were recorded for 15 seconds at four chest locations, that is, the second intercostal space at the right sternal border, the second intercostal space at the left sternal border, the fourth intercostal space at the left sternal border, and the apex of the heart. In addition, pulmonary sounds were recorded for 15 seconds at four chest locations, including the anterior chest (right front and left front) and the posterior chest (right back and left back). This approach was designed to capture a wide range of respiratory sounds, which can vary significantly in different parts of the chest due to the anatomy of the lungs and the distribution of airways.

Development process for ML model

Figure 1 shows the development of the ML model. The development of the ML model involved four steps, that is, pre-processing, feature extraction, classification, and integration for the user interface. Another unique aspect of our methodology was the parallel processing of overlapping cardiopulmonary sounds. This approach was useful for distinguishing between heart and lung sounds, which often occur simultaneously. Additionally, our ML models independently evaluated each cardiac and pulmonary phase and employed a majority rule with a defined margin to

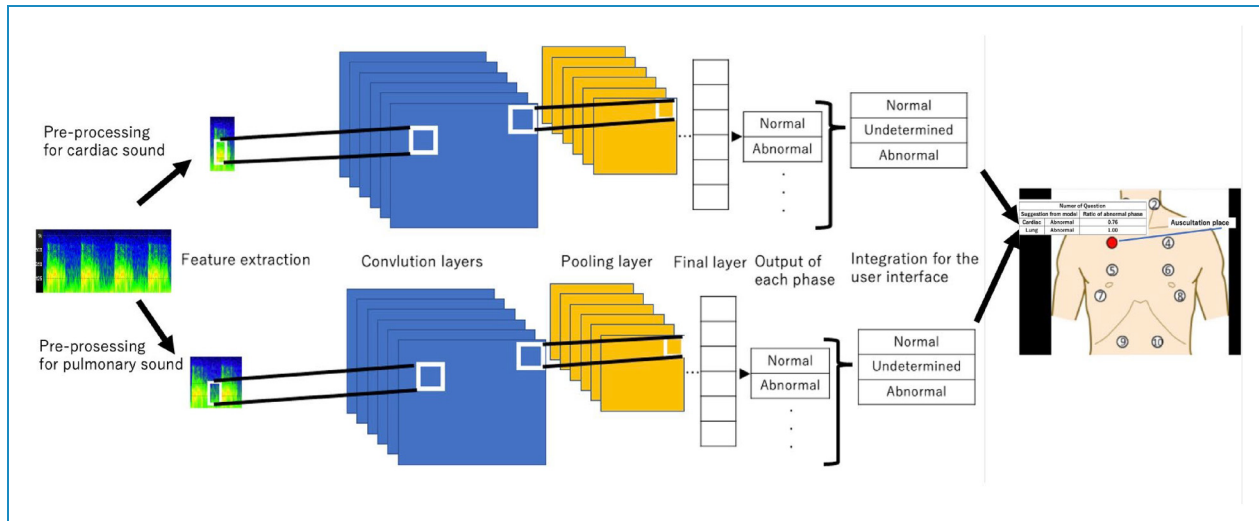


Figure 1. System development flow. Our methodology was the parallel processing of overlapping cardiopulmonary sounds, utilizing separate ML models for cardiac and pulmonary sound classification. Upper and lower sections were used for analyzing cardiac sound and pulmonary sound, respectively. The development of the ML model involved four steps, i.e. pre-processing, feature extraction, classification, and integration for the user interface. The overall predictions in cardiac and pulmonary sounds were shown in the left upper window of the intervention group as the user interface.

integrate the binary classifications into a comprehensive prediction for the user interface.

Pre-processing for cardiac and pulmonary sounds

First, we trimmed and augmented the audio files using the Librosa²⁷ Python library in the pre-processing step. The step included normalization and augmentation with white noise and introduced speed and volume variations.

Feature extraction

Second, we calculated the mel-spectrogram using the Librosa²⁷ Python library in the feature extraction process. To effectively differentiate between overlapping cardiac and pulmonary sounds, we implemented an envelope-based method.^{28,29} This method is particularly useful for distinguishing between two different rhythms, characteristic of both cardiac and lung sounds, by isolating the temporal patterns that differentiate heartbeats from breathing cycles. The distinct rhythmic patterns of heartbeats compared to the more variable patterns of breath sounds allow for the effective segmentation of our audio data, ensuring that our ML models process and classify cardiac and pulmonary sounds accurately.

Classification

Third, the proposed model is a convolutional neural network (CNN) using Keras and a Tensorflow back-end for classification. The sequential model comprised artificial

neural network layers, including an input layer, followed by four two-dimensional (2D) convolution layers, associated dropout layers for regulation, and max pooling layers for feature reduction. The input sizes for the 2D convolution layers are progressively increased to capture features at different scales: 16 for the first layer, 32 for the second, 64 for the third, and 128 for the fourth. The final layer includes the GlobalAveragePooling2D layer, followed by a dense layer with a SoftMax activation function. We employed one-hot encoding for the output labels to categorize the auscultation sounds into either “normal” or “abnormal.” The model was compiled using adaptive moment estimation optimizer and learning rate was 0.001. It is imperative to note that due to the limitations of our dataset, the output layer of the CNN categorizes sounds solely into two classes: “normal” or “abnormal.”

The training process involved multiple epochs with a batch size optimized for our computing resources. For validation, we employed a split of the data into training, testing, and validation sets. Table 1 details the confusion matrix for classification of cardiac and pulmonary sounds. The sensitivity and specificity for classification of cardiac sounds were 99.5% and 99.5%, respectively. The sensitivity and specificity for classification of pulmonary sounds were 92.0% and 99.9%, respectively.

Integration for the user interface

In the final step of system development, we integrated the binary output from our CNN into the user interface. The CNN was designed to classify auscultation sounds into

two primary categories for each cardiac and pulmonary phase: “normal” or “abnormal.” Due to dataset limitations, which offered a limited range of labeled examples, the model was not equipped to differentiate subtypes within the “abnormal” category.

To construct a comprehensive user interface, we devised a system to amalgamate the binary classifications from individual phases into an overall prediction. If the model classified more than two-thirds of the phases as “abnormal,” the final output would be categorized as an “abnormal” prediction. Similarly, an overall “normal” prediction required more than two-thirds of phases to be classified as “normal.” Any ratio other than these is categorized as “undetermined” in the overall prediction.

Table 1. Confusion matrix for model’s classification of cardiac and pulmonary sounds.

Actual classification	Predicted classification	
	Normal	Abnormal
Cardiac sounds		
Normal	2132	10
Abnormal	5	194
Pulmonary sounds		
Normal	1099	8
Abnormal	4	49

This tri-level classification system allowed us to present a nuanced view of the ML model’s predictions, reflecting a consensus from the individual phase predictions. During the study, the participants in the intervention group auscultated and viewed the overall prediction, which were displayed as “normal,” “abnormal,” or “undetermined” in the left upper window of the user interface (Figure 2A). The window provided an overall prediction along with the ratio of phases determined to be abnormal for both cardiac and pulmonary sounds, in addition to the number of questions presented.

Control group interface

On the other hand, the control group interface was designed to show only the number of questions without ML model prediction in the left upper window (Figure 2B). This approach was intended to isolate the effect of ML support in the intervention group and to serve as a comparative baseline for the study’s outcomes.

Evaluation process for ML model

We evaluated the utility of the proposed ML system to improve the correct answer rate of auscultation in an open-label randomized controlled trial. This evaluation process involved four major steps; that is, preparing sound material, a training session, randomization, and a test session (Figure 3).

Type of study for evaluation

A randomized controlled trial is a prospective study that measures the effectiveness of a new intervention by

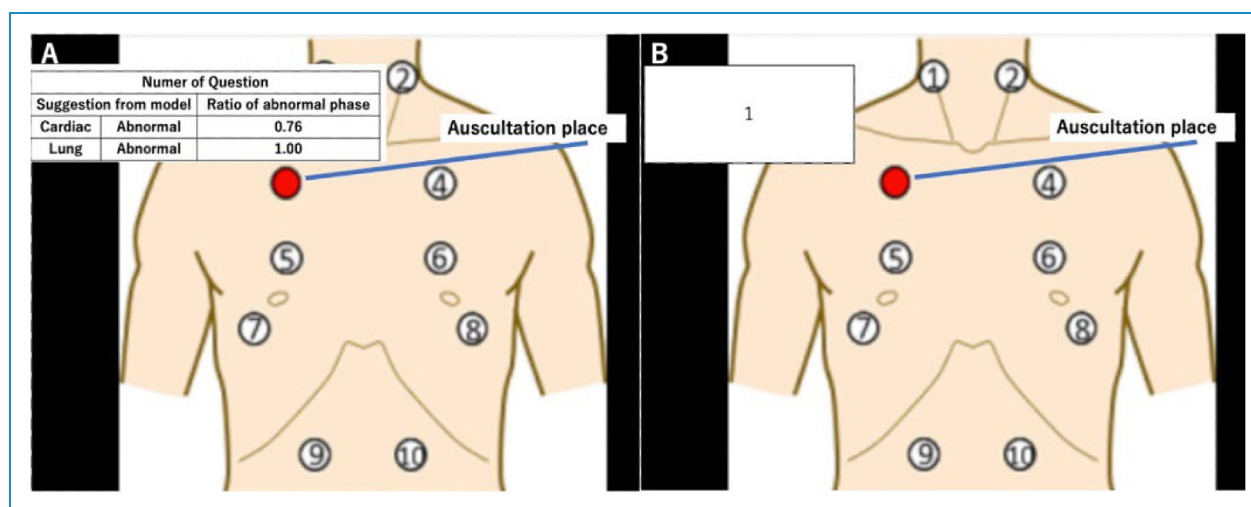


Figure 2. Example images of the test session. In the intervention group, the suggestions from ML-model were shown in the left upper window (A). In the control group, only the numbers of questions were shown in the left upper window (B). Each cardiopulmonary sound consisted of four locations: second intercostal space at the right sternal border (No. 3, red circle indicates the auscultation position in the figure), second intercostal space at the left sternal border (No. 4), apex (No. 8), and ipsilateral apex (No. 7).

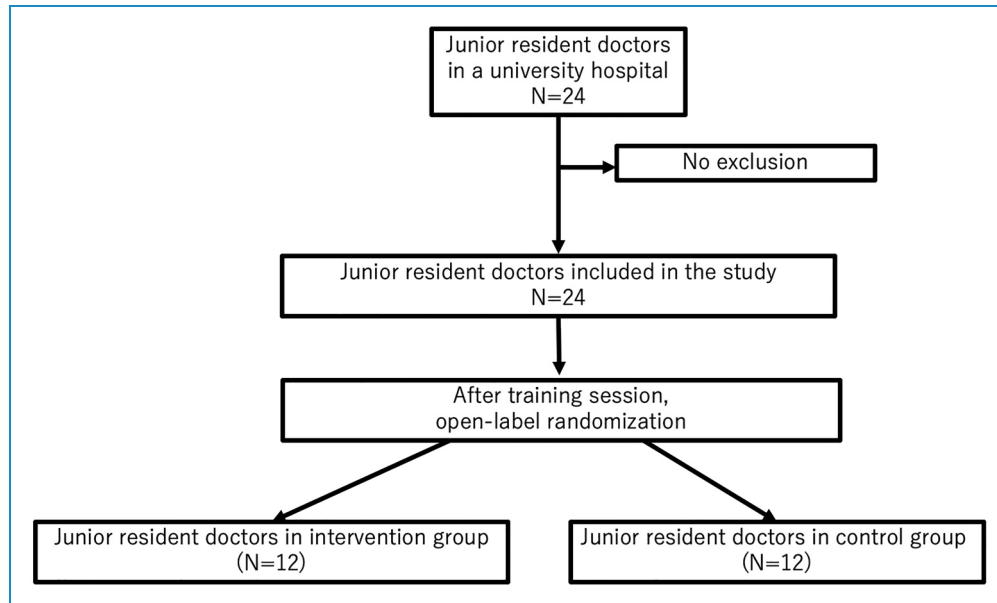


Figure 3. Flowchart of the participants in the study.

randomly assigning participants to an intervention group or a control group. An open-label trial is a type of randomized controlled study. In this approach, both the researcher and the participants are aware of their group assignments and the interventions they receive after randomization. Note that a randomized controlled trial must be approved by a research committee. It must also be registered to be published, according to the International Committee of Medical Journal Editors.

Recruitment

We recruited 24 junior resident doctors from Dokkyo Medical University Hospital as participants between August 2022 and February 2023. A junior resident refers to a doctor who has been practicing clinically for one to two years after graduating from medical school. Note that participation was voluntary. In addition, the exclusion criterion was hearing loss. All participants were included. The Research Committee of Dokkyo Medical University Hospital approved this study's procedures (R-60-8J). In addition, all participants provided written informed consent. The protocol was registered in the University Hospital Medical Information Network Clinical Trials Registry (UMIN ID: UMIN000048499, https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000055270).

Sound materials for evaluation

In preparing the sound materials for evaluation, we ensured a selection process to maintain the integrity of the testing phase. We select eight distinct sound categories, comprising four cardiac and four pulmonary sounds. The cardiac sound

categories included normal, aortic stenosis, aortic regurgitation, and mitral regurgitation. For pulmonary sounds, the categories were normal, fine crackles, coarse crackles, and wheeze.

It is crucial to note that these sounds were meticulously curated from a separate dataset not utilized in the ML model's training or development process. This distinction ensures that the model's performance is evaluated on truly unseen data.

The selection of these test materials was based on their clinical relevance and prevalence in diagnostic settings. By including a variety of pathological sounds along with normal sounds, we aimed to challenge the model across a spectrum of clinically significant scenarios. The sounds were sourced from a private available database.

Training session

Second, in the training session, the participants auscultated the four cardiac sounds and the four pulmonary sounds in a previously determined order. Here, each sound was played for a maximum of 15 seconds.

Randomization

Third, the participants were randomized into one of two groups; the intervention group (i.e. where the proposed ML-based support system was used) or the control group (conventional auscultation). Here, the main researcher was blinded in terms of allocation, by using a block randomized generator in Excel (Microsoft, Redmond, WA, USA) to assign participants. The generator was made by another researcher.

Test session

Finally, the participants in the control group auscultated 16 cardiopulmonary sounds using noise-cancelling headphones (WH-1000XM3; Sony Corp., Tokyo, Japan). The participants in the intervention group auscultated 16 cardiopulmonary sounds like the control group using the proposed support system in a test session (Figure 3 and Supplementary Movie Material including the first 2 questions out of the 16 questions in the test session). Here, the 16 cardiopulmonary sounds were a combination of four cardiac sounds (i.e. normal, aortic stenosis, aortic regurgitation, and mitral regurgitation sounds) and four pulmonary sounds (i.e. normal, fine crackles, coarse crackles, and wheeze sounds). Each cardiopulmonary sound consists of four locations: second intercostal space at the right sternal border (No. 3, red mark of the auscultation place in Figure 2), second intercostal space at the left sternal border (No. 4 in Figure 2), apex (No. 8 in Figure 2), and ipsilateral apex (No. 7 in Figure 2). Each cardiopulmonary sound was played for 60 seconds. During the test session, all participants filled in forms to indicate the types of sounds they recognized.

Outcomes

The primary outcome was the total scores for correctly identified normal or abnormal cardiopulmonary sounds. The total number of cardiopulmonary sounds was used as denominators, consisting of 12 participants in each group who answered 16 questions including cardiac and pulmonary sounds (12 by 16 by 2). The numerator was the number of correctly identified normal or abnormal cardiopulmonary sounds. The secondary outcomes were the total correct rate to identify cardiac sounds, the total correct rate to identify pulmonary sounds, and the correct rate to identify each cardiac sound and each pulmonary sound.

Sample size calculation

In previous studies,^{13,26,30} the diagnostic accuracy of cardiopulmonary auscultation was nearly 70% to 90% depending on the setting and the study design. To determine the required sample size, we set the α error (type I error) to 0.05, the power ($1 - \beta$, type II error) to 0.95, and both groups contained the same number of participants. The expected mean was set to 0.8 for the intervention group, and 0.7 for the control group. Accordingly, the required total sample size was calculated as 325. In this study, each participant answered 16 questions. Thus, a minimum of 22 participants were required to facilitate an effective comparison between the experimental groups. In consideration of participant dropout or missing data, we enrolled 24 participants.

Analysis

The Mann–Whitney U test was used to analyze the continuous variables, and a Chi-squared test was used for the categorical or binary variables. Here, the continuous variables for the participant baseline are presented as the mean and standard deviation. In this evaluation, a P -value of <0.05 was considered statistically significant. Note that all statistical tests were performed using the R (version 4.2.2) for macOS X (The R foundation for Statistical Computing, Vienna, Austria). We used G*power version 3.1.9.6 (Department of Psychology of Heinrich Heine University, Düsseldorf, Germany) for statistical power analysis.

Results

Participant profiles

In total, 24 participants (junior residents from a university hospital) were included in the final analysis. Among the participants, 13 (54.2%) were male, the mean age of 27.9 ± 3.8 years, and the mean time since graduation was 1.3 ± 0.4 years. There were no statistically significant differences between the intervention and the control groups in participant age ($P=0.2$), sex ($P=0.2$), or the time since graduation ($P=0.6$). Twelve participants were assigned to each group.

Diagnostic performance

In terms of diagnostic performance, the scores for correctly identified cardiopulmonary sounds by junior resident doctors are presented in Table 2. The total scores for correctly identified normal or abnormal cardiopulmonary sounds in the intervention group (366/384 [95.3%]) were significantly better than those of the control group (343/384 [89.3%], $P=0.003$).

For cardiac auscultation, the total cardiac test score in the intervention group (111/192, [57.8%] vs. 90/192 [46.9%], $P=0.04$) was significantly better than that of the control group. The intervention group also had a significantly higher percentage of correct answers for the normal cardiac sounds compared to the control group (48/48 [100%] vs. 38/48 [79.2%], $P=0.003$). No statistically significant difference was observed in other cardiac sounds between the two groups.

For pulmonary auscultation, no statistically significant difference was observed in the total pulmonary scores between the two groups (148/192 [77.1%] vs. 153/192 [79.7%], $P=0.62$). The percentages of correct answers of each pulmonary sound were not different between the two groups.

Power analysis indicated that statistical power for comparison of the total score for correctly identified normal

Table 2. Cardiopulmonary sounds correctly identified by junior resident doctors.

Variable	Intervention group (12 participants)	Control group (12 participants)	p-Value*
Total scores for correctly identified normal or abnormal cardiopulmonary sounds, n/N (%)	366/384 (95.3)	343/384 (89.3)	0.003
Total cardiac sounds, n/N (%)	111/192 (57.8)	90/192 (46.9)	0.04
Normal, n/N (%)	48/48 (100)	38/48 (79.2)	0.003
Aortic stenosis, n/N (%)	26/48 (54.2)	20/48 (41.7)	0.31
Aortic regurgitation, n/N (%)	14/48 (29.2)	13/48 (27.1)	>0.99
Mitral regurgitation, n/N (%)	23/48 (47.9)	19/48 (39.6)	0.54
Total pulmonary sound, n/N (%)	148/192 (77.1)	153/192 (79.7)	0.62
Normal, n/N (%)	44/48 (91.7)	47/48 (97.9)	0.13
Wheezes, n/N (%)	48/48 (100)	47/48 (97.9)	>0.99
Coarse crackles, n/N (%)	26/48 (54.2)	25/48 (52.1)	>0.99
Fine crackles, n/N (%)	30/48 (62.5)	34/48 (70.8)	0.52

* Chi-squared test.

or abnormal cardiopulmonary sounds exceeded 0.99. The statistical powers for comparing the total cardiac sounds and the total pulmonary sounds were 0.98 and 0.19, respectively.

Discussion

Model performance in system development

Both in cardiac and pulmonary sound classification for each phase, the models' specificity of over 99% indicates an exceptionally low rate of Type-I error, which refers to the incorrect rejection of a true null hypothesis. This high specificity translates to a low rate of false positive diagnoses in clinical terms. It demonstrates that the models are adept at correctly identifying the majority of normal cardiac conditions. Additionally, the models also showed a sensitivity of 99% for cardiac sounds, indicating high accuracy in identifying abnormal cardiac conditions, thereby minimizing the risk of Type-II errors, commonly known as false negatives.

These performance metrics for cardiac sounds are promising, especially considering that remote cardiac auscultation has previously been identified as less effective than traditional methods.¹³ Enhancing the detection capabilities could thus improve remote diagnostics, making it more viable and reliable.

Regarding pulmonary sound classification, the models maintain high sensitivity at 92%. Although this is slightly lower than that for cardiac sounds, it remains significant. However, this does suggest a margin for error, with a risk of Type-II errors that could have substantial clinical implications. Given the serious nature of missing a pulmonary abnormality, which might lead to delayed treatment, physicians should exercise additional caution and corroborate the ML model's findings with other diagnostic information when interpreting pulmonary sounds.

Principal results

In the following, we discuss our principal findings. First, the proposed ML-based clinical decision support system for cardiopulmonary auscultation improved the total scores for correctly identified normal or abnormal cardiopulmonary sounds by 6%. This was because the system provided feedback and decision support, which enhanced the diagnostic accuracy of junior residents. This demonstrated that the system was able to assist early-career physicians in accurately auscultating cardiopulmonary sounds for screening purposes. Second, the total cardiac correct answer rate was improved by more than 10% with the assistance of the ML system. Even though the system did not suggest extensive details about abnormal cardiac sounds, the proposed system was found to improve auscultation accuracy. Third,

no statistical difference in the total pulmonary test score was observed between the two groups. This was partly due to the fact that the pulmonary auscultation accuracy was already high, and the suggestion offered by the proposed system did not include extensive details about abnormal pulmonary sounds. In the future, there is a possibility to improve the accuracy of pulmonary auscultation by the suggestion regarding the details of abnormal pulmonary sounds. Fourth, we observed no statistical difference in the test score for specific abnormal sounds. This was partly because the system did not suggest extensive details about abnormal cardiopulmonary sounds due to limited dataset.

Strengths

This study had significant strengths. First, we evaluated the proposed ML-based clinical decision support system in an open-label randomized controlled study, which is considered the gold standard for clinical research. Second, we enrolled the doctors who were completing their training as resident physicians. We found that the proposed system effectively assisted participants in their cardiac auscultation evaluations. Third, the participants auscultated both cardiopulmonary sounds simultaneously with only four points on the anterior chest compared to more than eight points in conventional cardiopulmonary auscultation. This suggested that the proposed system can assist in conducting physical examinations both quickly and accurately in clinical and telemedicine environments.

Limitations

Several limitations should be discussed. First, the dataset used to develop the proposed ML model was limited in size and variability from simulators. Consequently, with a limited dataset, the model's predictive power is confined to distinguish only normal or abnormal sounds. It lacks the extensive and varied examples necessary to learn the subtle distinction and complexities of different abnormal sound patterns. Future research should focus on acquiring a more diverse and extensive dataset from real patients, which could enable the model to learn and recognize a broader spectrum of abnormal sound patterns. Second, the optimal timing for presenting the ML system's output (before, after, or simultaneously physicians' auscultation) was unknown. Third, an appropriate interface for integrating the ML model's clinical decision support into the auscultation process was not available. Fourth, in response to valid concerns raised, we acknowledge that the cardiopulmonary sounds cannot be accurately represented by simply combined heart and lung sounds. Future studies should investigate novel methodologies for analyzing these complex sounds. Advanced signal processing techniques or

more sophisticated ML algorithms would enhance diagnostic performance. Finally, the responses may have exhibited dependencies within the same participant. This indicates that the answers from an individual might not have been entirely independent of one another.

Future direction

In future studies, we aim to explore the application of our ML-based clinical decision support system across a spectrum of medical expertise, including medical students, junior physicians, and senior physicians with diverse specialities. Specifically, we plan to conduct comparative studies to assess how different levels of clinical experience might affect the utilization and perceived value of the system. For senior physicians, with their extensive experience, we anticipate the system will primarily serve as a confirmatory tool. This could yield insights into the nuanced differences in how experienced clinicians interact with ML tools compared to their less experienced counterparts.

Additionally, further research will focus on expanding the dataset to include a wider range of abnormal sound patterns, enhancing the ML model's learning capability. Collaborations with experts in cardiopulmonary medicine will be sought to enrich the dataset and refine the model's accuracy in diagnosing specific abnormalities.

Furthermore, we plan to extend our research to evaluate the impact of the ML system in real-patient settings across different clinical environments, such as outpatient clinics, inpatient wards, and emergency departments. This would involve measuring patient outcomes and the system's diagnostic accuracy in real-world scenarios.

Lastly, we intend to establish a feedback loop that allows physicians at different career stages to contribute their insights and experiences to refine the ML system. This feedback will not only help in refining the ML model but also ensure its relevance and applicability in diverse clinical contexts for the use of the system. Collaboration with senior physicians in the research and development phase will be crucial in achieving the sufficient feedback, leveraging their expertise to enhance the system's capabilities.

Conclusions

In an open-label randomized controlled study, the proposed ML-based clinical decision support system to assist cardiopulmonary auscultation improved the auscultation accuracy of junior residents. Specifically, we observed a 6% improvement in identifying normal or abnormal cardiopulmonary sounds and more than a 10% increase in the correct answer rate for cardiac auscultation. These findings underscore the potential of ML tools in augmenting medical training and improving patient care.

Acknowledgements: This study was made possible using the resources from the Department of Diagnostic and Generalist Medicine, Dokkyo Medical University.

Contributorship: TH and YH researched literature and conceived the study. TH was involved in protocol development, gaining ethical approval, participant recruitment and data analysis. TH wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The ethics committee of Dokkyo Medical University Hospital approved this study (REC number: R-60-8J).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 22K10421).

Guarantor: TH

Trial Registration: UMIN ID: UMIN000048499, https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000055270.

ORCID iD: Takanobu Hirosawa  <https://orcid.org/0000-0002-3573-8203>

Supplemental material: Supplemental material for this article is available online.

References

- Bohadana A, Izbicki G and Kraman SS. Fundamentals of lung auscultation. *N Engl J Med* 2014; 370: 744–751.
- Bank I, Vliegen HW and Bruschke AV. The 200th anniversary of the stethoscope: can this low-tech device survive in the high-tech 21st century? *Eur Heart J* 2016; 37: 3536–3543.
- Lee SM and Lee D. Opportunities and challenges for contactless healthcare services in the post-COVID-19 era. *Technol Forecast Soc Change* 2021; 167: 120712.
- Kruse C and Heinemann K. Facilitators and barriers to the adoption of telemedicine during the first year of COVID-19: systematic review. *J Med Internet Res* 2022; 24: e31752.
- Myers DR, Weiss A, Rollins MR, et al. Towards remote assessment and screening of acute abdominal pain using only a smartphone with native accelerometers. *Sci Rep* 2017; 7: 12750.
- Krumholz R, Fuchtmann J, Berlet M, et al. Telemedical percussion: objectifying a fundamental clinical examination technique for telemedicine. *Int J Comput Assist Radiol Surg* 2022; 17: 795–804.
- Benziger CP, Huffman MD, Sweis RN, et al. The telehealth ten: a guide for a patient-assisted virtual physical examination. *Am J Med* 2021; 134: 48–51.
- Fan M, Wang Q, Liu J, et al. Real-world evaluation of the stethoscope electronic tele-auscultation system. *Biomed Eng Online* 2022; 21: 63.
- Behere S, Baffa JM, Penfil S, et al. Real-world evaluation of the eko electronic teleauscultation system. *Pediatr Cardiol* 2019; 40: 154–160.
- Kalinauskienė E, Razvadauskas H, Morse DJ, et al. A comparison of electronic and traditional stethoscopes in the heart auscultation of obese patients. *Medicina (Kaunas)* 2019; 55: 94.
- Ghanayim T, Lupu L, Naveh S, et al. Artificial intelligence-based stethoscope for the diagnosis of aortic stenosis. *Am J Med* 2022; 135: 1124–1133.
- Belmont JM and Mattioli LF. Accuracy of analog telephonic stethoscopy for pediatric telecardiology. *Pediatrics* 2003; 112: 780–786.
- Ito T, Hirosawa T, Harada Y, et al. Evaluation of internet-connected real-time remote auscultation: an open-label randomized controlled pilot trial. *J Pers Med* 2022; 12: 1950.
- Lee SH, Kim YS, Yeo MK, et al. Fully portable continuous real-time auscultation with a soft wearable stethoscope designed for automated disease diagnosis. *Sci Adv* 2022; 8: eabo5867.
- Swarup S and Makaryus AN. Digital stethoscope: technology update. *Med Devices (Auckl)* 2018; 11: 29–36.
- Tiwari S, Jain A, Sharma AK, et al. Phonocardiogram signal based multi-class cardiac diagnostic decision support system. *IEEE Access* 2021; 9: 110710–22.
- Yuenyong S, Nishihara A, Kongprawechnon W, et al. A framework for automatic heart sound analysis without segmentation. *Biomed Eng Online* 2011; 10: 13.
- Li S, Li F, Tang S, et al. A review of computer-aided heart sound detection techniques. *BioMed Res Int* 2020; 2020: 5846191.
- Abiodun OI, Jantan A, Omolara AE, et al. State-of-the-art in artificial neural network applications: a survey. *Heliyon* 2018; 4: e00938.
- Agatonovic-Kustrin S and Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal* 2000; 22: 717–727.
- Deng C, Ji X, Rainey C, et al. Integrating machine learning with human knowledge. *iScience* 2020; 23: 101656.
- Gurung A, Scrafford CG, Tielsch JM, et al. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med* 2011; 105: 1396–1403.
- Huang DM, Huang J, Qiao K, et al. Deep learning-based lung sound analysis for intelligent stethoscope. *Mil Med Res* 2023; 10: 44.
- Hadjileontiadis LJ and Panas SM. Adaptive reduction of heart sounds from lung sounds using fourth-order statistics. *IEEE Trans Biomed Eng* 1997; 44: 642–648.

25. Wu YC, Han CC, Chang CS, et al. Development of an electronic stethoscope and a classification algorithm for cardiopulmonary sounds. *Sensors (Basel)* 2022; 22: 4263.
 26. Hirosawa T, Harada Y, Ikenoya K, et al. The utility of real-time remote auscultation using a bluetooth-connected electronic stethoscope: open-label randomized controlled pilot trial. *JMIR Mhealth Uhealth* 2021; 9: e23109.
 27. McFee B, Raffel C, Liang D, et al. Librosa: audio and music signal analysis in python. In: Proceedings of the 14th python in science conference, vol. 8, pp. 18–25. (2015).
 28. Nivitha Varghees V and Ramachandran KI. Multistage decision-based heart sound delineation method for automated analysis of heart sounds and murmurs. *Healthc Technol Lett* 2015; 2: 156–163.
 29. Liu C, Springer D, Li Q, et al. An open access database for the evaluation of heart sound algorithms. *Physiol Meas* 2016; 37: 2181–2213.
 30. Bernardi S, Giudici F, Leone MF, et al. A prospective study on the efficacy of patient simulation in heart and lung auscultation. *BMC Med Educ* 2019; 19: 275.
-