



OPEN

Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning

Yoonjoo Kim^{1,5}, YunKyong Hyon^{2,5}, Sung Soo Jung¹, Sunju Lee², Geon Yoo³, Chaeuk Chung^{1,4}✉ & Taeyoung Ha²✉

Auscultation has been essential part of the physical examination; this is non-invasive, real-time, and very informative. Detection of abnormal respiratory sounds with a stethoscope is important in diagnosing respiratory diseases and providing first aid. However, accurate interpretation of respiratory sounds requires clinician's considerable expertise, so trainees such as interns and residents sometimes misidentify respiratory sounds. To overcome such limitations, we tried to develop an automated classification of breath sounds. We utilized deep learning convolutional neural network (CNN) to categorize 1918 respiratory sounds (normal, crackles, wheezes, rhonchi) recorded in the clinical setting. We developed the predictive model for respiratory sound classification combining pretrained image feature extractor of series, respiratory sound, and CNN classifier. It detected abnormal sounds with an accuracy of 86.5% and the area under the ROC curve (AUC) of 0.93. It further classified abnormal lung sounds into crackles, wheezes, or rhonchi with an overall accuracy of 85.7% and a mean AUC of 0.92. On the other hand, as a result of respiratory sound classification by different groups showed varying degree in terms of accuracy; the overall accuracies were 60.3% for medical students, 53.4% for interns, 68.8% for residents, and 80.1% for fellows. Our deep learning-based classification would be able to complement the inaccuracies of clinicians' auscultation, and it may aid in the rapid diagnosis and appropriate treatment of respiratory diseases.

The stethoscope has been considered as an invaluable diagnostic tool ever since it was invented in the early 1800s. Auscultation is non-invasive, real-time, inexpensive, and very informative¹⁻³. Recent electronic stethoscopes have rendered lung sounds recordable, and it facilitated the studies of automatically analyzing lung sounds^{4,5}. Abnormal lung sounds include crackles, wheezes, rhonchi, stridor, and pleural friction rubs (Table 1). Crackles, wheezes and rhonchi are the most commonly found among them, and detecting those sounds greatly aids the diagnosis of pulmonary diseases^{6,7}. Crackles, which are short, explosive, and non-musical, are produced by patients with parenchymal lung diseases such as pneumonia, interstitial pulmonary fibrosis (IPF), and pulmonary edema^{1,8,9}. Wheezes are musical high-pitched sounds associated with airway diseases such as asthma and chronic obstructive pulmonary disease (COPD). Rhonchi are musical low-pitched sounds similar to snores, usually indicating secretions in the airway, and are often cleared by coughing¹.

Although auscultation has many advantages, the ability to analyze respiratory sounds among clinicians varies greatly depending on individual clinical experiences^{6,10}. Salvatore et al. found that hospital trainees misidentified about half of all pulmonary sounds, as did medical students¹¹. Melbye et al. reported significant inter-observer differences in terms of discriminating expiratory rhonchi and low-pitched wheezes from other sounds, potentially compromising diagnosis and treatment¹². These limitations of auscultation raised the need to develop a standardized system that can classify accurately respiratory sounds using artificial intelligence (AI). AI-assisted

¹Division of Pulmonology and Critical Care Medicine, Department of Internal Medicine, College of Medicine, Chungnam National University, Daejeon 34134, Republic of Korea. ²Division of Medical Mathematics, National Institute for Mathematical Sciences, Daejeon 34047, Republic of Korea. ³Clinical Research Division, National Institute of Food and Drug Safety Evaluation, Cheongju-si, Chungcheongbuk-do, Republic of Korea. ⁴Infection Control Convergence Research Center, Chungnam National University School of Medicine, Daejeon 35015, Republic of Korea. ⁵These authors contributed equally: Yoonjoo Kim and YunKyong Hyon. ✉email: universe7903@gmail.com; tha@nims.re.kr

	Mechanism of sound production	Location	Characteristics	Acoustics	Related diseases
Fine crackles	Unrelated to secretions	Peripheral lung	Discontinuous High-pitched Inspiratory	Rapidly dampened wave deflection Frequency: about 650 Hz Shorter duration (about 5 ms)	Interstitial lung fibrosis Pneumonia Congestive heart failure
Coarse crackles	Intermittent airway opening, related to secretions	Peripheral lung	Discontinuous Low-pitched Inspiratory	Rapidly dampened wave deflection Frequency about 350 Hz Longer duration (about 15 ms)	Same as fine crackles but usually more advanced disease
Wheezes	Narrowed airway Flow limitation	Bronchi	Continuous High-pitched Expiratory > Inspiratory	Sinusoid Frequency > 100–5000 Hz Duration > 80 ms	Asthma COPD Tumor Foreign body
Rhonchus	Rupture of fluid films of secretions Airway wall vibrations	Bronchi	Continuous Low-pitched Expiratory > Inspiratory	Sinusoid Frequency about 150 Hz Duration > 80 ms	Bronchitis Pneumonia
Stridor	Narrowed airway	Larynx, Trachea	Continuous High-pitched Inspiratory	Sinusoid Frequency > 500 Hz	Epiglottitis After extubation Foreign body
Pleural friction rub	Pleural inflammation	Chest wall	Continuous Low-pitched Inspiratory and expiratory	Rhythmic succession of short sounds Frequency < 350 Hz Duration > 15 ms	Pleurisy Pericarditis Pleural tumor

Table 1. Classification of abnormal lung sounds.

auscultation can help a proper diagnosis of respiratory disease and identify patients in need of emergency treatment. It can be used to screen and monitor patients with various pulmonary diseases including asthma, COPD and pneumonia^{13,14}.

Recently, deep learning is widely applied to some medical fields including a chest x-ray or electroencephalogram analysis^{15,16}. There are several published studies on AI-assisted auscultation of heart and lung sounds^{13,17–20}. AI was used to distinguish different murmurs and diagnose valvular and congenital heart diseases²¹. Auscultation of the lung is different from that of the heart in some aspects. First, the lungs are much larger than the heart; and lung sounds should be recorded at multiple sites of both lungs for an accurate analysis. Second, the quality of lung sound is easily affected by the patient's effort to breathe.

There are several studies that tried to automatically analyze and classify respiratory sounds^{15,22–30}. An interesting study quantified and characterized lung sounds in patients with pneumonia for generating acoustic pneumonia scores²². The sound analyzer was helpful for detecting pneumonia with 78% sensitivity and 88% specificity. In another study, crackles and wheezes in 15 children were applied to feature extraction via time–frequency/scale analysis; the positive percent agreement was 0.82 for crackle and 0.80 for wheezing²³. Tomasz et al. used neural network (NN)-based analysis to differentiate four abnormal sounds (wheezes, rhonchi, and coarse and fine crackles) in 50 children¹⁸. Intriguingly, the results showed that the NN F1-score was much better than that of doctors. Gorkem et al. used a support vector machine (SVM), the k-nearest neighbor approach, and a multilayer perceptron to detect pulmonary crackles²⁴. Gokhan et al. proposed the automatic detection of the respiratory cycle and collected synchronous auscultation sounds from COPD patients^{28,31,32}. Interestingly, they demonstrated that deep learning is useful for diagnosing COPD and classifying the severity of COPD with significantly high-performance rates^{28,29}.

Another study employed a sound database of the international conference on biomedical and health informatics (ICBHI) 2017 for classifying lung sounds using a deep convolutional NN. They converted the lung sound signals to spectrogram images by using the time–frequency method, but the accuracy was relatively low (about 65%)²⁵. There are many feature extractors with CNN classifiers including inception V3, DenseNet201, ResNet50, ResNet101, VGG16, and VGG19^{33–38}. In this study, we tried to combine pre-trained image feature extraction from time-series, respiratory sound, and CNN classification. We also compared the performances of these feature extractors.

Although this field has been being actively studied, it is still in its infancy with significant limitations. Many studies enrolled patients of a limited age group (children only), and some studies analyzed the sounds of a small numbers of patients. The studies that used the respiratory sounds of the ICBHI 2017 or the R.A.L.E. Repository database have a limitation in types of abnormal sounds. The ICBHI database contained crackles and wheezes only, and R.A.L.E. database lacked rhonchi³⁹.

In this study, we aimed to classify normal respiratory sounds, crackles, wheezes, and rhonchi. We made a database of 1,918 respiratory sounds from adult patients with pulmonary diseases and healthy controls. Then we used transfer learning and convolutional neural network (CNN) to classify those respiratory sounds. We tried to combine pre-trained image feature extraction from time-series, respiratory sound, and CNN classification. In addition, we measured how accurately medical students, interns, residents, and fellows categorized breathing sounds to check the accuracy of auscultation classification in real clinical practice.

Results

The general characteristics of the enrolled patients and the collected lung sounds. We recorded 2840 sounds and the respiratory sounds were evaluated by three pulmonologists independently and classified. Then we made a respiratory sound database contained 1222 normal sounds (63.7%) and 696 abnormal sounds (36.3%) including 297 crackles (15.5%), 298 wheezes (15.5%), and 101 rhonchi (5.3%). Our database of classi-

871 cases (recording sounds: n = 1918)	Characteristics
Age, mean (years)	67.7 ± 10.9
Sex	Male: 562 (64.5%), Female: 309 (35.5%)
Diagnosis	Pneumonia: 11.1%, IPF: 4.9%, COPD: 21%, Asthma: 12.3%, Lung cancer/mass: 13.1%, Healthy: 5.9%, Tuberculosis: 5.4%, Bronchiectasis: 4.4% ILD except IPF: 8.0%, ACO: 0.3% ETC: 13.5%
Respiratory sounds	Normal: 1222 (63.7%), Crackles: 297(15.5%), Wheezes: 298 (15.5%), Rhonchi 101 (5.3%)
Auscultation location	RULF: 10.4%, RMLF: 6.9%, RLLF: 34.8% LULF: 16.3%, LMLF: 9.0%, LLLF: 21.7%, Neck 0.9%

Table 2. Characteristics of respiratory sound database. *RULF* right upper lobe field, *RMLF* right middle lobe field, *RLLF* right lower lobe field, *LULF* left upper lobe field, *LMLF* left middle lobe field, *LLL* left lower lobe field, *COPD* chronic obstructive pulmonary disease, *ILD* interstitial lung disease, *IPF* idiopathic pulmonary disease, *ACO* asthma-COPD overlap.

fied sounds was consisted of 1918 sounds from 871 patients in the clinical field. Their demographic and clinical characteristics are presented in Table 2. The mean patient age was 67.7 (\pm 10.9) years and 64.5% of patients were male. Sounds were collected from patients with pneumonia, IPF, COPD, asthma, lung cancer, tuberculosis, and bronchiectasis, as well as healthy controls. The proportions of COPD and asthma patients were 21% and 12.3% respectively, the pneumonia proportion 11.1%, the IPF proportion 8.0%, and the healthy control proportion 5.9%. The location of auscultation was most common in both lower lobe fields. (Table 2).

Performance of AI-assisted lung sound classification. *Discriminating normal sounds from abnormal sounds (crackles, wheezes, and rhonchi).* In clinical settings, distinguishing abnormal breathing sounds and normal sounds is very important in screening emergency situations and deciding whether to perform additional tests. Our sound database included 1222 normal sounds and 696 abnormal sounds. We first checked how accurately our deep-learning based algorithm can classify abnormal respiratory sounds from normal sounds (Fig. 1). The precision, recall, and F1 scores for abnormal lung sounds were 84%, 80%, and 81% respectively (Table 3). The accuracy was 86.5% and the mean AUC was 0.93 (Fig. 2).

Categorization of abnormal sounds into crackles, wheezes, and rhonchi. Next, we categorized abnormal sounds as specific types of sounds: crackles, wheezes, or rhonchi using deep learning. The sound database included 297 crackles, 298 wheezes, and 101 rhonchi that were confirmed by specialists. The precision, recall, and F1 scores for crackles were 90%, 85%, and 87% respectively. In the case of wheezes, the precision, recall, and F1 scores were 89%, 93%, and 91%. Finally, the precision, recall, and F1 scores for rhonchi were 68%, 71%, and 69% respectively (Table 4). The average accuracy was 85.7% and the mean was AUC 0.92 (Fig. 3).

Comparison of performances of different feature extractors with CNN classifier. Respiratory sounds, especially abnormal sounds, have very complicated structures with noise, and positional dependency in time. In the sound analysis, particularly mathematical point of view, its 2-D spectral-domain has more information rather than one dimensional time-series. Moreover, the deep learning structure gives an automatic feature extraction overcoming the difficulties on complicate data, especially image data. For this reason, we adopted CNN, which is a powerful method in image classification. To find out the most optimized strategy for the classification of respiratory sounds, we also compared the accuracy, precision, recall score and F1 score of each analytic method (Table 5). CNN classifier showed the best performance with VGG, especially, VGG16 rather than InceptionV3, DenseNet201, ResNet50, and ResNet101. Since VGG architecture has a better capability, especially in extracting image features for classification using transfer learning^{40,41}, we adopted it for our AI models.

Additionally, we compared the performance between CNN and SVM classifiers in order to investigate classifier dependency of feature extractor. CNN showed better performance than SVM, and VGG16 was the best classifier for both CNN and SVM. Moreover, CNN was more efficient in computation time than SVM (Table 6).

Accuracy of auscultation analysis in real clinical practice. To verify the accuracy of auscultation analysis in real clinical practice and to evaluate the need for deep learning-based classification, we checked how exactly medical students, interns, residents, and fellows categorize breathing sounds (Fig. 4). We made several test sets of normal sounds and three types of abnormal lung sounds: crackle, wheezes, and rhonchi. 25 medical students, 11 interns, 23 residents, and 11 fellows of the internal medicine department of four teaching hospitals were asked to listen to the sounds and identify them. Regarding each breath sounds, the mean correct answer rates of normal sounds, crackles, wheezes, and rhonchi were 73.5%, 72.2%, 56.3%, and 41.7%, respectively. The overall correct answer rates of medical students, interns, residents, and fellows were 59.6%, 56.6%, 68.3%, and 84.0% respectively. The average correct answer rates for normal breathing were 67.1% for medical students, 75.7% for interns, 73.2% for residents, and 87.7% for fellows, while those for crackles were 62.9% for medical students, 72.3% for interns, 76.0% for residents, and 90.3% for fellows. The accuracies for wheezes were 55.6% for medical students, 41.0% for interns, 57.4% for residents, and 69.1% for fellows respectively, while those for rhon-

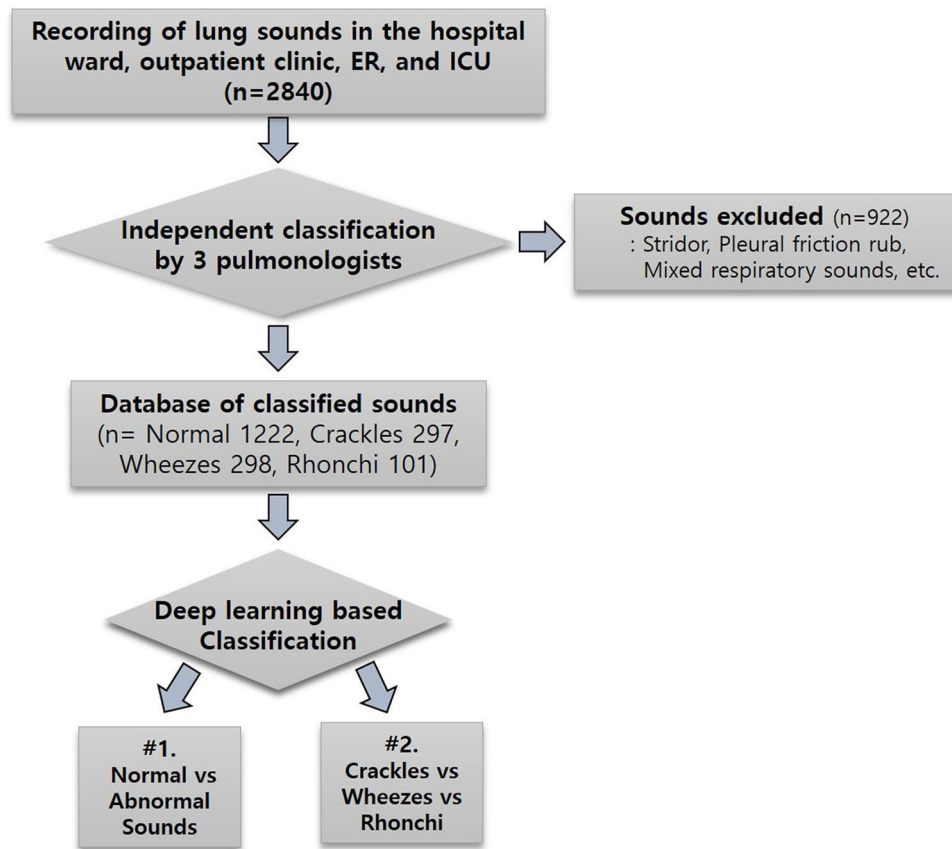


Figure 1. Scheme of the classification of respiratory sounds using deep learning. Lung sounds database contains normal sounds, crackles, wheezes, and rhonchi. Deep learning was used for two types of classification: The first step is the discriminating normal sounds from abnormal sounds. The second is to categorize abnormal sounds into crackles, wheezes, and rhonchi. (ER: Emergency room, ICU: intensive care unit).

	Precision	Recall score	F1 score
Normal	0.88	0.91	0.89
Abnormal	0.84	0.80	0.81

Table 3. The averages of Precision, Recall and F1 score in discriminating normal sounds from abnormal sounds.

chi were 42.5% for medical students, 15.0% for interns, 37.1% for residents, and 82.2% for fellows (Fig. 4). There was no significant difference between each group in analyzing normal breathing sound, but in all three types of abnormal breathing sound, the fellows showed the highest accuracy. Among the abnormal breath sounds, interns and residents classified crackles most accurately. Rhonchi was revealed to be the most difficult sound to discriminate (Fig. 4).

Discussion

The stethoscope has been considered an invaluable diagnostic tool for centuries^{2,3}. Although many diagnostic techniques have been developed, auscultation still plays major roles^{1,7}. For example, a pulmonologist can detect early-stage IPF or pneumonia based on inspiratory crackles even when the chest X-ray appears near-normal⁴². Changes in wheezes sometimes indicate the onset of asthma or COPD exacerbation. Therefore, early detection and accurate classification of abnormal breathing sounds can prevent disease progression and improve a patient's prognosis.

Several studies have tried to automatically classify lung sounds. Chamberlain et al. classified lung sounds with a semi-supervised deep learning algorithm. The AUC were 0.86 for wheezes and 0.74 for crackles, respectively²⁶. Guler et al. used a multilayer perceptron running a backpropagation training algorithm to predict the presence or absence of adventitious sounds²⁷. They enrolled 56 patients and two hidden layers yielded 93.8% rated classification performance²⁷.

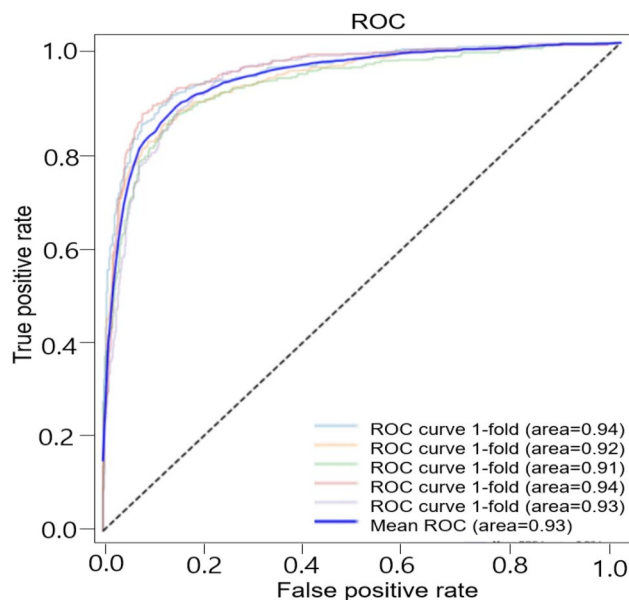


Figure 2. ROC of the model for discrimination of abnormal lung sounds. Each plot illustrates the ROC of the algorithm on the independent testing set for abnormal lung sounds, with AUC of 0.93.

	Precision	Recall score	F1 score
Crackles	0.90	0.85	0.87
Wheezes	0.89	0.93	0.91
Rhonchi	0.68	0.71	0.69

Table 4. The averages of Precision, Recall and F1 score in discriminating crackles, wheezes, and rhonchi.

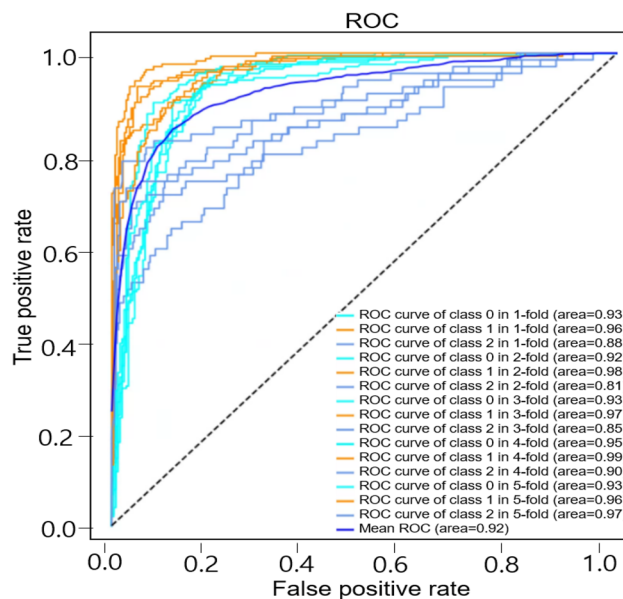


Figure 3. ROC of the model for classifying abnormal lung sounds into crackles, wheezes, and rhonchi. Each plot illustrates the ROC of the algorithm on the independent testing set for crackles, wheezes, and rhonchi with the mean AUC of 0.92.

Feature extractor	Accuracy	Precision	Recall score	F1 score
InceptionV3	0.748	0.722	0.727	0.720
DenseNet201	0.786	0.749	0.735	0.729
ResNet50	0.783	0.747	0.738	0.738
ResNet101	0.809	0.787	0.766	0.769
VGG16	0.857	0.823	0.828	0.824
VGG19	0.848	0.814	0.817	0.814

Table 5. Comparison of Performance among feature extractors with CNN classifier.

Classifier	Feature extractor	Accuracy	Precision	Recall score	F1 score
CNN	InceptionV3	0.748	0.722	0.727	0.720
	DenseNet201	0.786	0.749	0.735	0.729
	ResNet50	0.783	0.747	0.738	0.738
	ResNet101	0.809	0.787	0.766	0.769
	VGG16	0.857	0.823	0.828	0.824
	VGG19	0.848	0.814	0.817	0.814
SVM	InceptionV3	0.746	0.727	0.622	0.620
	DenseNet201	0.754	0.733	0.616	0.604
	ResNet50	0.747	0.768	0.599	0.569
	ResNet101	0.749	0.757	0.607	0.589
	VGG16	0.755	0.762	0.614	0.594
	VGG19	0.750	0.752	0.617	0.609

Table 6. Comparison of performance between CNN and SVM with feature extractors (Inception V3, DenseNet201, VGG16).

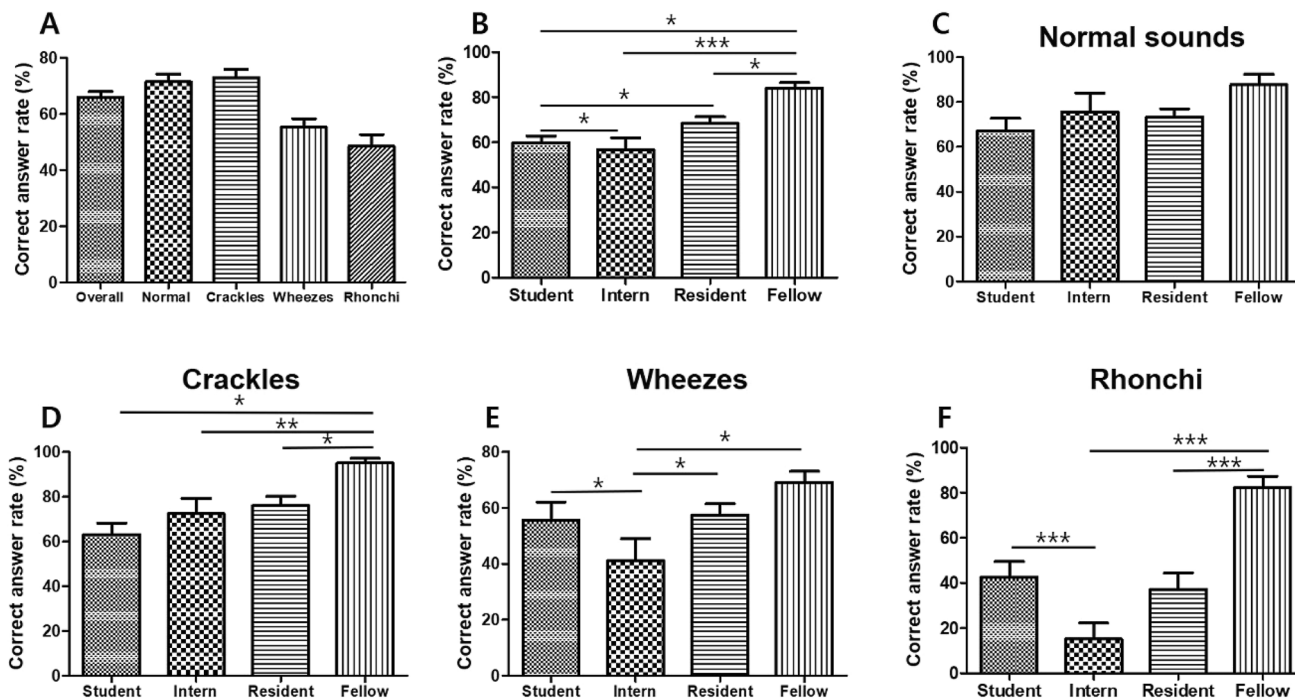


Figure 4. Accuracy of auscultation analysis in real clinical practice. (A) Mean correction answer rates for the overall sounds, normal sounds, crackles, wheezes, and rhonchi. (B) Mean correction answer rates of students, interns, residents, and fellows for overall sounds. (C) Mean correction answer rates of students, interns, residents, and fellows for normal sounds (D) Mean correction answer rates of students, interns, residents, and fellows for crackles (E) Mean correction answer rates of students, interns, residents, and fellows for wheezes. (F) Mean correction answer rates of students, interns, residents, and fellows for rhonchi. * $p < 0.05$, ** $p < 0.05$ *** $p < 0.001$ (Student’s t-test).

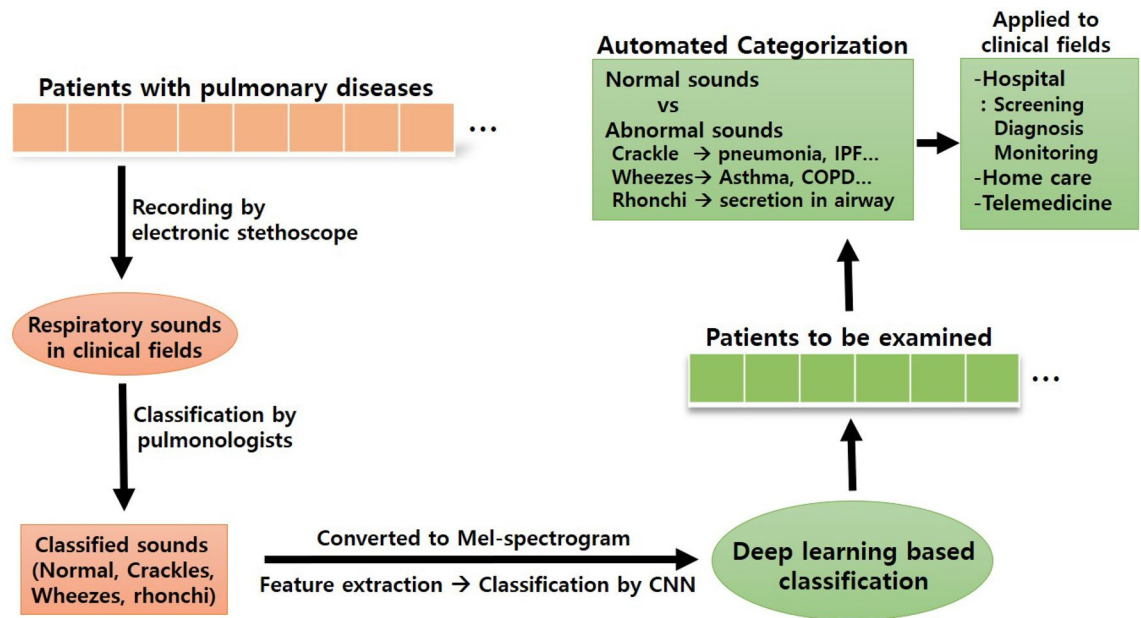


Figure 5. Summary of deep learning assisted classification of respiratory sounds. Respiratory sounds were corrected from the patients with pulmonary diseases. The sounds were validated and classified by pulmonologists. The sounds were converted to Mel-spectrogram and features were extracted by VGG16 (transfer learning). Respiratory sounds were classified by CNN. Deep learning-based classification of respiratory sounds can be helpful for screening, monitoring, and diagnosis of pulmonary diseases.

In this study, we used deep learning for the classification of respiratory sounds. Compared with several lung sound classification studies that applied machine learning or deep learning for lung sounds classification^{27,43–47}, we modified the deep learning algorithm of Bardou's study which applied SVM⁴⁶. In our study, we utilized the transfer learning method, which is easy, fast and able to use various features, but one has to be careful in connecting two deep learning networks, feature extractor and classifier. Moreover, there is a certain dependency between these two. We applied CNN instead of SVM because CNN is more efficient than a SVM for image classification.

Besides, our comparison of performances of different feature extractors demonstrated that CNN classifier showed much better performance with VGG, especially, VGG16 than InceptionV3 and Densenet201. The main contribution of this study is to develop the predictive model for respiratory sound classification combining pretrained image feature extractor of time-series, respiratory sound, and CNN classifier.

Our deep learning-based classification can detect abnormal lung sounds with an AUC of 0.93 and an accuracy of 86.5%. It has similar results in categorizing abnormal sounds into subcategorical sounds: crackles, wheezes, or rhonchi. Considering these are the result of analyzing the sounds recorded in a real clinical field with various noises, these are impressive results. We believe that these accuracies are adequate for primary screening and follow-up testing of patients with respiratory diseases.

Our test results showed that the auscultation accuracy of interns and residents were less than 80% in all four kinds of sounds and rhonchi was the most difficult sound to discriminate. The result of the test is not conclusive since the number of participants is small. However, it looks obvious that there are marked differences in the ability of each clinician to classify breathing sounds. This suggests that AI-assisted classification standardize the identification and categorization of breath sounds and greatly aid the diagnosis of pulmonary diseases.

There are several respiratory sounds in which two or more abnormal breath sounds are mixed. Such sounds are sometimes difficult even for experts and there may be disagreements between them. Few published studies have classified mixed abnormal breathing sounds, so research about these sounds is necessary. Also, since noises such as coughs, voices, heart sounds, and medical alarms are frequently recorded with breath sound, which reduces the accuracy of analysis, the technology for noise filtering is required.

Conclusion

We found that our deep learning-based classification could classify the respiratory sounds accurately. Utilizing the transfer learning method, combining pre-trained image feature extraction from respiratory sound and CNN classification, worked well and was helpful for improving the classification accuracy. Though the analysis of mixed abnormal sounds and filtering noises remain challenging, recent innovations in analytic algorithm and recording technology will accelerate the advance of respiratory sound analysis more rapidly. Soon, deep learning-based automated stethoscope is expected to be used in telemedicine and home care (Fig. 5).

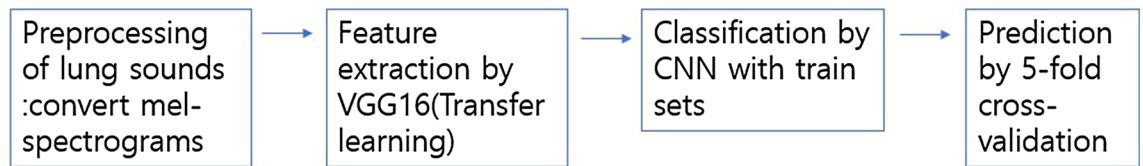


Figure 6. Overview of our AI models.

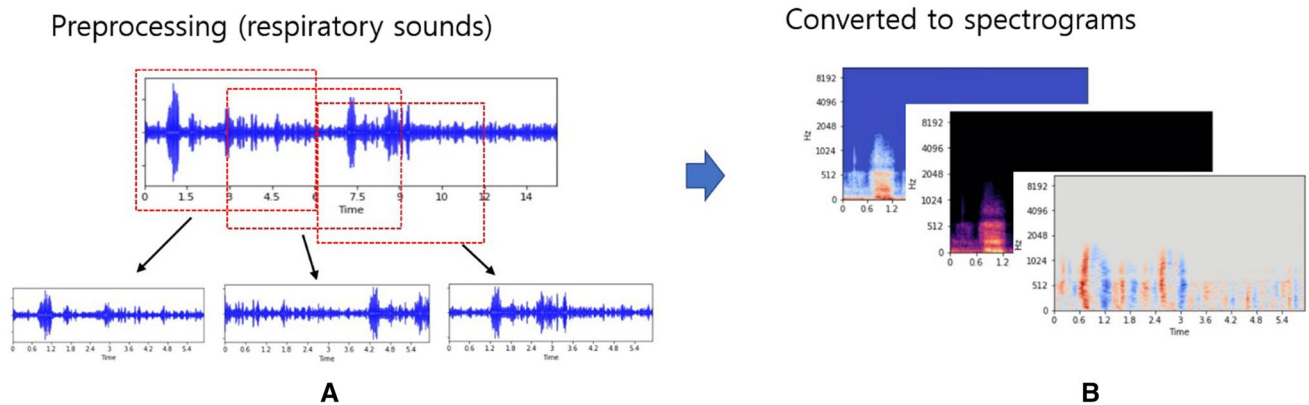


Figure 7. Process to obtain spectrograms. (A) Given lung sound, dividing lung sound files with overlapping during 3 s (B) Obtaining three types of Mel-spectrograms with log-scale.

Methods

Patient selection and data collection. Patients who visited an outpatient clinic or were hospitalized at Chungnam National University Hospital, regardless of the presence or type of respiratory diseases were enrolled from April 2019 to December 2020. The recording was proceeded in the actual clinical field (outpatient clinic, hospitalization room, emergency room, intensive care unit). Lung sounds were obtained from two to six sites of the posterior thorax using a Littman 3200 electronic stethoscope, downloaded to a computer, and converted into “wave” files. All sounds were carefully checked and validated by three pulmonologists. All patients gave written informed consent, and we obtained human research ethics committee approval of Chungnam National University Hospital Institutional Review Board (No. 2020-10-092). All methods were performed in accordance with the relevant guidelines and regulations. We recorded 2840 sounds and made a respiratory sound database containing 1222 normal breath sounds, 297 crackles, 298 wheezes, and 101 rhonchi.

AI models with transfer learning and CNN. *Overview of AI models.* Lung sounds were converted to Mel-spectrograms and features were extracted by VGG16. CNN was applied for the classification and fivefold cross-validation was used for prediction (Fig. 6).

Preprocessing of lung sounds. Recorded sounds were ranged from a few seconds to several tens of seconds. We divided them into 6 s each with 50% overlapping. For example, the audio file is a 14.5-s audio file of wheezing, which is divided into 3 cycles according to the start and end times (Fig. 7). And, to process the feature extraction and use the 3-dimensional input data, we used Mel-spectrogram, average of harmonic and percussive Mel-spectrogram, and the derivative of Mel-spectrogram using the Python library librosa⁴⁷.

Feature extractor and classification. We thought at least two or three cycles of respiratory sounds are needed for accurate analysis of lung sounds. Approximately, normal respiratory rate is 15–20 per one minute (three–four seconds per one respiratory cycle) and it tends to be more rapid at pathologic conditions. So, after testing several options, we finally have decided six seconds as the length of the respiratory sound.

We used pre-trained models VGG16 as feature extractors in transfer learning, which was built by Karen Simonyan^{48,49}. VGG16 is a model with 16 layers trained on fixed-size images and the input is processed through a set of convolution layers that use small-size kernels with a receptive field 3×3 . The default input size of VGG16 is 224×224 , but the input size for our model is 256×256 (Fig. 8). We used weights pre-trained on ImageNet by freezing all the five convolution blocks without fully-connected layer, and predicted the test sets with simple CNN with only one-layer.

Evaluation of our models. To avoid overfitting, we utilized the fivefold cross-validation method³⁴ (Fig. 9). The dataset has been chosen randomly to split into 80% training set and 20% test set, and 20% of training set is used for validation. The main idea of the fivefold cross validation is to split the training set into 5 partitions. Each time one of the 5 partitions are used for validating the model and the other 4 partitions are used for training the

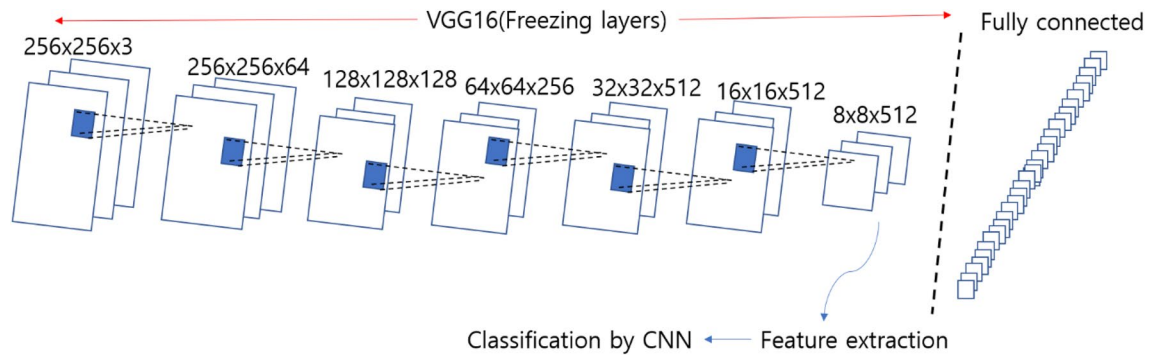


Figure 8. VGG16 architecture for our model. The input size for our model is 256×256 . We freeze all layers in VGG16 without fully-connected layer, and with extracting features, we classified the respiratory sounds.

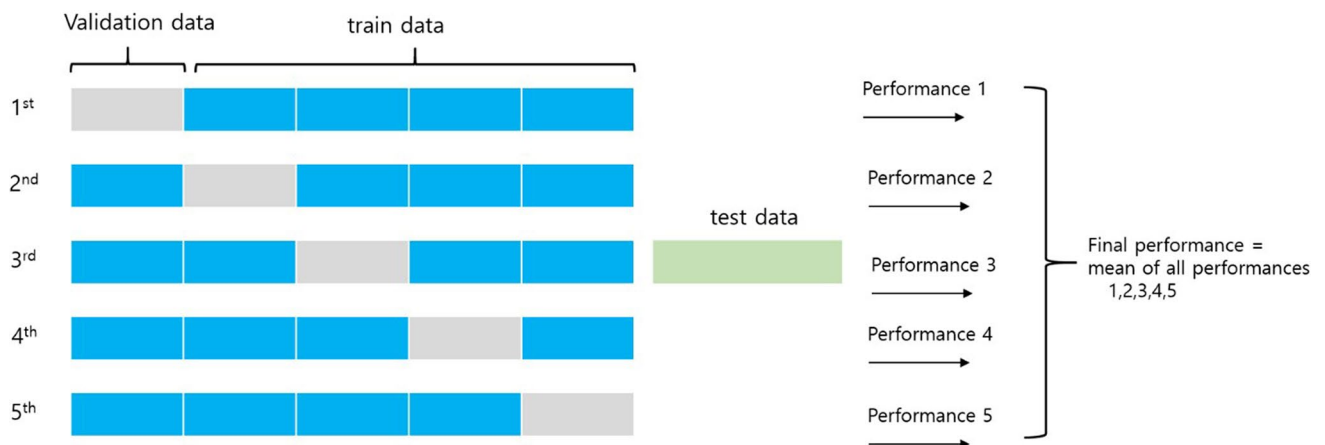


Figure 9. 5-folds cross validation. Trying 5 iterations and getting final results with mean of all performances.

model. So, each instance in the data set is used once in testing and 4 times in training. All results of the different metrics are then averaged to return the result. From results by our models, we obtained accuracy, precision, recall score and ROC curve.

Statistical analysis. All values are presented as means \pm standard deviation (SD). Significant differences were determined using GraphPad 5 software. The Student's t-test was used to determine statistical differences between two groups. The receiver operating characteristic curve was plotted and the area under the curve was calculated with the 95% confidence intervals.

Received: 24 April 2021; Accepted: 12 August 2021

Published online: 25 August 2021

References

- Bohadana, A., Izbicki, G. & Kraman, S. S. Fundamentals of lung auscultation. *N. Engl. J. Med.* **370**, 744–751. <https://doi.org/10.1056/NEJMra1302901> (2014).
- Bloch, H. The inventor of the stethoscope: René Laennec. *J. Fam. Pract.* **37**, 191 (1993).
- Roguin, A. Rene Theophile Hyacinthe Laennec (1781–1826): The man behind the stethoscope. *Clin. Med. Res.* **4**, 230–235. <https://doi.org/10.3121/cm.4.3.230> (2006).
- Swarup, S. & Makaryus, A. N. Digital stethoscope: Technology update. *Med. Dev. (Auckland N. Z.)* **11**, 29–36. <https://doi.org/10.2147/mder.s135882> (2018).
- Leng, S. *et al.* The electronic stethoscope. *Biomed. Eng. Online* **14**, 66. <https://doi.org/10.1186/s12938-015-0056-y> (2015).
- Arts, L., Lim, E. H. T., van de Ven, P. M., Heunks, L. & Tuinman, P. R. The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: A meta-analysis. *Sci. Rep.* **10**, 7347. <https://doi.org/10.1038/s41598-020-64405-6> (2020).
- Sarkar, M., Madabhavi, I., Niranjan, N. & Dogra, M. Auscultation of the respiratory system. *Ann. Thoracic Med.* **10**, 158–168. <https://doi.org/10.4103/1817-1737.160831> (2015).
- Vyshedskiy, A. *et al.* Mechanism of inspiratory and expiratory crackles. *Chest* **135**, 156–164. <https://doi.org/10.1378/chest.07-1562> (2009).
- Fukumitsu, T., Obase, Y. & Ishimatsu, Y. The acoustic characteristics of fine crackles predict honeycombing on high-resolution computed tomography. *BMC Pulm. Med.* **19**, 153. <https://doi.org/10.1186/s12890-019-0916-5> (2019).
- Hafke-Dys, H., Bręborowicz, A. & Kleka, P. The accuracy of lung auscultation in the practice of physicians and medical students. *PLoS ONE* **14**, e0220606. <https://doi.org/10.1371/journal.pone.0220606> (2019).
- Mangione, S. & Nieman, L. Z. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am. J. Respir. Crit. Care Med.* **159**, 1119–1124. <https://doi.org/10.1164/ajrccm.159.4.9806083> (1999).

12. Melbye, H. *et al.* Wheezes, crackles and rhonchi: Simplifying description of lung sounds increases the agreement on their classification: a study of 12 physicians' classification of lung sounds from video recordings. *BMJ Open Respir. Res.* **3**, e000136. <https://doi.org/10.1136/bmjresp-2016-000136> (2016).
13. Andr s, E., Gass, R., Charloux, A., Brandt, C. & Hentzler, A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J. Med. Life* **11**, 89–106 (2018).
14. Ohshimo, S., Sadamori, T. & Tanigawa, K. Innovation in analysis of respiratory sounds. *Ann. Intern. Med.* **164**, 638–639. <https://doi.org/10.7326/L15-0350> (2016).
15. Altan, G., Yayık, A. & Kutlu, Y. Deep learning with ConvNet predicts imagery tasks through EEG. *Neural Process. Lett.* **53**, 2917–2932 (2021).
16. Tang, Y. X., Tang, Y. B. & Peng, Y. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Dig. Med.* **3**, 70. <https://doi.org/10.1038/s41746-020-0273-z> (2020).
17. Coucke, P. A. Laennec versus Forbes : Tied for the score ! How technology helps us interpret auscultation. *Rev. Med. Liege* **74**, 543–551 (2019).
18. Grzywalski, T. *et al.* Practical implementation of artificial intelligence algorithms in pulmonary auscultation examination. *Eur. J. Pediatr.* **178**, 883–890. <https://doi.org/10.1007/s00431-019-03363-2> (2019).
19. Palaniappan, R., Sundaraj, K. & Sundaraj, S. Artificial intelligence techniques used in respiratory sound analysis: A systematic review. *Biomedizinische Technik. Biomed. Eng.* **59**, 7–18. <https://doi.org/10.1515/bmt-2013-0074> (2014).
20. Ono, H. *et al.* Evaluation of the usefulness of spectral analysis of inspiratory lung sounds recorded with phonopneumography in patients with interstitial pneumonia. *J. Nippon Med. School Nippon Ika Daigaku Zasshi* **76**, 67–75. <https://doi.org/10.1272/jnms.76.67> (2009).
21. Thompson, W. R., Reinisch, A. J., Unterberger, M. J. & Schriefel, A. J. Artificial intelligence-assisted auscultation of heart murmurs: Validation by virtual clinical trial. *Pediatr. Cardiol.* **40**, 623–629. <https://doi.org/10.1007/s00246-018-2036-z> (2019).
22. Murphy, R. L. *et al.* Automated lung sound analysis in patients with pneumonia. *Respir. Care* **49**, 1490–1497 (2004).
23. Kevat, A., Kalirajah, A. & Roseby, R. Artificial intelligence accuracy in detecting pathological breath sounds in children using digital stethoscopes. *Respir. Res.* **21**, 253. <https://doi.org/10.1186/s12931-020-01523-9> (2020).
24. Serbes, G., Sakar, C. O., Kahya, Y. P. & Aydin, N. Feature extraction using time-frequency/scale analysis and ensemble of feature sets for crackle detection. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* **2011**, 3314–3317. <https://doi.org/10.1109/IEMBS.2011.6090899> (2011).
25. Demir, F., Sengur, A. & Bajaj, V. Convolutional neural networks based efficient approach for classification of lung diseases. *Health Inform. Sci. Syst.* **8**, 4. <https://doi.org/10.1007/s13755-019-0091-3> (2020).
26. Chamberlain, D., Kodgule, R., Ganelin, D., Miglani, V. & Fletcher, R. R. Application of semi-supervised deep learning to lung sound analysis. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* **2016**, 804–807. <https://doi.org/10.1109/EMBC.2016.7590823> (2016).
27. Guler, I., Polat, H. & Ergun, U. Combining neural network and genetic algorithm for prediction of lung sounds. *J. Med. Syst.* **29**, 217–231. <https://doi.org/10.1007/s10916-005-5182-9> (2005).
28. Altan, G., Kutlu, Y. & Allahverdi, N. Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE J. Biomed. Health Inform.* <https://doi.org/10.1109/JBHI.2019.2931395> (2019).
29. Altan, G., Kutlu, Y., Pekmezci, A.  . & Nural, S. Deep learning with 3D-second order difference plot on respiratory sounds. *Biomed. Signal Process. Control* **45**, 58–69 (2018).
30. Altan, G., Kutlu, Y. & G k en, A. Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds. *Turk. J. Electr. Eng. Comput. Sci.* **28**, 2979–2996 (2020).
31. Altan, G., Kutlu, Y., Garbi, Y., Pekmezci, A.  . & Nural, S. Multimedia respiratory database (RespiratoryDatabase@ TR): Auscultation sounds and chest X-rays. *Nat. Eng. Sci.* **2**, 59–72 (2017).
32. Aras, S.,  zt rk, M. & Gangal, A. Automatic detection of the respiratory cycle from recorded, single-channel sounds from lungs. *Turk. J. Electr. Eng. Comput. Sci.* **26**, 11–22 (2018).
33. Zheng, L. *et al.* Artificial intelligence in digital cariology: A new tool for the diagnosis of deep caries and pulpitis using convolutional neural networks. *Ann. Transl. Med.* **9**, 763. <https://doi.org/10.21037/atm-21-119> (2021).
34. Rezaeijoo, S. M., Ghorvei, M. & Mofid, B. Predicting breast cancer response to neoadjuvant chemotherapy using ensemble deep transfer learning based on CT images. *J. Xray Sci. Technol.* <https://doi.org/10.3233/xst-210910> (2021).
35. Arora, V., Ng, E. Y., Leekha, R. S., Darshan, M. & Singh, A. Transfer learning-based approach for detecting COVID-19 ailment in lung CT scan. *Comput. Biol. Med.* **135**, 104575. <https://doi.org/10.1016/j.compbiomed.2021.104575> (2021).
36. Jin, W., Dong, S., Dong, C. & Ye, X. Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph. *Comput. Biol. Med.* **131**, 104252. <https://doi.org/10.1016/j.compbiomed.2021.104252> (2021).
37. Wang, Q. *et al.* Realistic lung nodule synthesis with multi-target co-guided adversarial mechanism. *IEEE Trans. Med. Imag.* <https://doi.org/10.1109/tmi.2021.3077089> (2021).
38. Pu, J., Sechrist, J., Meng, X., Leader, J. K. & Sciruba, F. C. A pilot study: Quantify lung volume and emphysema extent directly from two-dimensional scout images. *Med. Phys.* <https://doi.org/10.1002/mp.15019> (2021).
39. Gharehbaghi, A. & Linden, M. A deep machine learning method for classifying cyclic time series of biological signals using time-growing neural network. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 4102–4115. <https://doi.org/10.1109/TNNLS.2017.2754294> (2018).
40. Raghu, M., Chiyuan, Z., Jon, K. & Samy, B. Transfusion: Understanding Transfer learning for medical imaging. In *33rd Conference on Neural Information Processing Systems* (2019).
41. Park, S., Kim, J. & Kim, D. A study on classification performance analysis of convolutional neural network using ensemble learning algorithm. *J. Korea Multimedia Soc.* **22**, 665–675 (2019).
42. Epler, G. R., Carrington, C. B. & Gaensler, E. A. Crackles (rales) in the interstitial pulmonary diseases. *Chest* **73**, 333–339. <https://doi.org/10.1378/chest.73.3.333> (1978).
43. Horimasu, Y. *et al.* A machine-learning based approach to quantify fine crackles in the diagnosis of interstitial pneumonia: A proof-of-concept study. *Medicine* **100**, e24738. <https://doi.org/10.1097/md.00000000000024738> (2021).
44. Naves, R., Barbosa, B. H. & Ferreira, D. D. Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. *Comput. Methods Programs Biomed.* **129**, 12–20. <https://doi.org/10.1016/j.cmpb.2016.02.013> (2016).
45. Aykanat, M., Kılıç,  ., Kurt, B. & Saryal, S. Classification of lung sounds using convolutional neural networks. *J. Image Video Proc.* <https://doi.org/10.1186/s13640-017-0213-2> (2017).
46. McFee, B. *librosa*. *librosa* 0.8.0. <https://doi.org/10.5281/zenodo.3955228> (2020).
47. Bardou, D., Zhang, K. & Ahmad, S. M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **88**, 58–59. <https://doi.org/10.1016/j.artmed.2018.04.008> (2018).
48. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* **25**, 1097–1105 (2012).
49. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *CoRR abs* **1409**, 1556 (2014).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. NRF-2017R1A5A2015385). T.H, Y.H. and S.L. was supported by National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government (No. B21910000). Authors appreciate Y.P, I.O, S.Y and I.J for collecting the data related to the accuracy of auscultation analysis in real clinical practice.

Author contributions

Y.K., C.C., T.H., and Y.H made substantial contributions to conception and design. Y.K., C.C., and S.S.J. performed data acquisition and validation for this study. T.H., Y.H. and S.L. implemented deep learning with lung sounds. G.Y. edited the manuscript. All authors contributed to the elaboration and redaction of the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.C. or T.H.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021