PLOS ONE

Diagnostic performance of artificial intelligence model for pneumonia from chest radiography

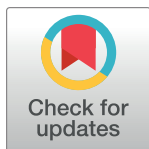**TaeWoo Kwon[1]ⓨ, Sang Pyo Lee[2]ⓨ, Dongmin Kim[1]ⓨ, Jinseong Jang[1], Myungjae Lee[1], Shin Uk Kang[1], Heejin Kim[3], Keunyoung Oh[3], Jinhee On[3], Young Jae Kim[4], So Jeong Yun[4], Kwang Nam Jin[5], Eun Young Kim[6‡]*, Kwang Gi Kim[4‡]***

1 JLK, Incorporated, Eonju-ro, Gangnam-gu, Seoul, South Korea, **2** Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, South Korea, **3** Korea National Tuberculosis Association (KNTA), Seoul, South Korea, **4** Department of Biomedical Engineering, Gachon University College of Medicine, Incheon, South Korea, **5** Department of Radiology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, South Korea, **6** Department of Radiology, Gil Medical Center, Gachon University College of Medicine, Incheon, South Korea

ⓨ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* oneshot0229@gmail.com (EYK); kimkg@gachon.ac.kr (KGK)

Abstract

Objective

The chest X-ray (CXR) is the most readily available and common imaging modality for the assessment of pneumonia. However, detecting pneumonia from chest radiography is a challenging task, even for experienced radiologists. An artificial intelligence (AI) model might help to diagnose pneumonia from CXR more quickly and accurately. We aim to develop an AI model for pneumonia from CXR images and to evaluate diagnostic performance with external dataset.

Methods

To train the pneumonia model, a total of 157,016 CXR images from the National Institutes of Health (NIH) and the Korean National Tuberculosis Association (KNTA) were used (normal vs. pneumonia = 120,722 vs.36,294). An ensemble model of two neural networks with DenseNet classifies each CXR image into pneumonia or not. To test the accuracy of the models, a separate external dataset of pneumonia CXR images (n = 212) from a tertiary university hospital (Gachon University Gil Medical Center GUGMC, Incheon, South Korea) was used; the diagnosis of pneumonia was based on both the chest CT findings and clinical information, and the performance evaluated using the area under the receiver operating characteristic curve (AUC). Moreover, we tested the change of the AI probability score for pneumonia using the follow-up CXR images (7 days after the diagnosis of pneumonia, n = 100).

Results

When the probability scores of the models that have a threshold of 0.5 for pneumonia, two models (models 1 and 4) having different pre-processing parameters on the histogram

Competing interests: The authors have read the journal's policy and the authors of this manuscript have the following competing interests: TK, DK, ML, and SUK are paid employees of JLK Inc. and JJ was a paid employee of JLK Inc. at the time of study. KGK received a research grant from JLK Inc. for activities not related to the present article. This does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

equalization distribution showed best AUC performances of 0.973 and 0.960, respectively. As expected, the ensemble model of these two models performed better than each of the classification models with 0.983 AUC. Furthermore, the AI probability score change for pneumonia showed a significant difference between improved cases and aggravated cases (Δ = -0.06 ± 0.14 vs. 0.06 ± 0.09, for 85 improved cases and 15 aggravated cases, respectively, P = 0.001) for CXR taken as a 7-day follow-up.

Conclusions

The ensemble model combined two different classification models for pneumonia that performed at 0.983 AUC for an external test dataset from a completely different data source. Furthermore, AI probability scores showed significant changes between cases of different clinical prognosis, which suggest the possibility of increased efficiency and performance of the CXR reading at the diagnosis and follow-up evaluation for pneumonia.

Introduction

Pneumonia is a contagious disease characterized by acute lower respiratory infection, usually caused by viruses or bacteria and less commonly, other microorganisms [1]. Pneumonia affects approximately 450 million people (7% of the population) globally, results in approximately 4 million deaths each year [2]. Typical symptoms include a combination of productive or dry cough, chest pain, fever, and breathing difficulties [3, 4]. The presentation might be atypical since it is influenced by patient age, comorbid conditions, lifestyle factors, and causative pathogens. Furthermore, because pneumonia shares many symptoms with other conditions such as the common cold, bronchitis, and asthma, the diagnosis is difficult with only symptoms and physical examination. CXR remains the most readily available and common imaging modality for the assessment of pneumonia [5, 6]. Radiologically, CXR presentations of pneumonia can be classified as lobar pneumonia, bronchopneumonia, and interstitial pneumonia. Bacterial, community-acquired pneumonia classically show airspace consolidation of one lung segmental lobe, which is known as lobar pneumonia [7]. However, findings may vary, and other patterns are common in different types of pneumonia. Detecting pneumonia from chest radiography is a challenging task since radiologic findings may be absent or subtle in the early stages of the disease and a high prevalence of pulmonary conditions can mimic pneumonia [8]. Furthermore, some studies showed that errors are common in the interpretation of chest radiographs, due to inter-observer variation [9]. The limitation of human expert-based diagnosis has provided a strong motivation for the use of computer technology to improve the speed and accuracy of the detection process. Now, with the advent of digital CXR, there is renewed interest in using CXR interpreted by computer-aided diagnosis (CAD) software programs for the diagnosis of pneumonia [10–12]. Recent advance of deep learning algorithm for CAD on CXR, diagnosis of pneumonia is expected to improve physician's performance for diagnosing pneumonia [13, 14].

The purpose of this study was to develop an artificial intelligence (AI) model for the diagnosis of pneumonia using the public open database of CXR and to evaluate the diagnostic accuracy of the model using an external dataset from a tertiary hospital in Korea.

Materials and methods

Data source

In this study, data collected from the National Institutes of Health (NIH), the Korean National Tuberculosis Association (KNTA), and a tertiary university hospital (Gachon University Gil Medical Center GUGMC, Incheon, South Korea) were used. The summary of the demographic information on the data are provided in Table 1 and S1 Data.

The institutional review board of GUGMC approved this retrospective study and waived the requirement for informed patient consent (approval number: GBIRB-2019-337).

Dataset for model development

For model development, dataset was extracted from large-scale public open dataset of the National Institutes of Health (NIH), containing approximately 112,000 CXR images (portable network graphics, png format) of multiple labels using natural language processing (NLP) [15]; The CXR labels of abnormal data are as follows; atelectasis, cardiomegaly, consolidation, edema, effusion, emphysema, fibrosis, infiltration, mass, nodule, pleural thickening pneumonia, pneumothorax, and hernia. Along with normal CXR (n = 60,361), pneumonia (n = 36,294) was chosen based on having one or more of the following CXR labels; consolidation, effusion, infiltration, nodule, and pneumonia.

The Korean National Tuberculosis Association (KNTA) dataset consists of CXRs (DICOM format) for tuberculosis screening by KNTA and has been annotated by multiple radiologists for the presence of tuberculosis abnormalities. To allow the pneumonia model to learn various types of data, an equal number of normal data (n = 60,361) from KNTA were randomly sampled.

A total of 70% was used for model training and the remaining 30% dataset was used as a validation dataset.

Dataset for evaluation

The test dataset was extracted from a pneumonia cohort (between January and June 2018) of a tertiary university hospital (GUGMC) as DICOM files (n = 212). The diagnosis based on the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes and the initial CXR was collected at the time of admission. Normal data were randomly extracted from NIH and KNTA (212 pieces of normal data, n = 106 for each) to set the ratio of test dataset between pneumonia and normal as 1:1. In addition, the

Table 1. Dataset for model development and evaluation.

	Dataset for model development		Test dataset
CXR images according to data source (normal vs. pneumonia)	n = 157,016 (120,722 vs. 36,294) from NIH (60,361 vs. 36,294) from KNTA (60,361 vs. 0)		normal: NIH(n = 106), KNTA (n = 106) pneumonia from GUGMC (n = 212)
	Training dataset (70%)	**Validation dataset (30%)**	
	n = 109,912 (84,506 vs. 25,406)	n = 47,104 (36,216 vs. 10,888)	n = 424 (212 vs. 212)
Number of patients (male vs. female)	62,703 (32,065 vs. 30,638)	28,463 (14,746 vs. 13,717)	424 (288 vs. 136)
Age (mean ± SD)	47 ± 16	47 ± 16	54 ± 13

Abbreviations: CXR, chest X-ray; NIH, National Institutes of Health; KNTA, Korean National Tuberculosis Association; GUGMC, Gachon University Gil Medical Center.

https://doi.org/10.1371/journal.pone.0249399.t001

longitudinal response was evaluated when the follow-up CXR became available on the 7-day after initial CXR (FU CXR) from a subset of pneumonia patients (n = 100). On FU CXR, the cases were determined as improved cases and aggravated cases using medical record based on the clinical presentation (deterioration of clinical symptom and required ventilator care and ICU transfer) and on the radiologic changes.

Pre-processing module

The pre-processing module consists of Invert Checker, Lung Finder, and Data Rescaler. Since the file format and data type of CXR are different, the image input is put through the pre-processing module.

Invert Checker checks whether the input image is inverted in black-white format. The invert check model is a simple convolutional neural network (CNN) model designed to classify whether the image is inverted to black-white format. For the training data, 1,000 randomly extracted data from NIH and KNTA were used and the images were composed of black-white inverted images and non-inverted images on a 1:1 ratio. The Invert Check Model consists of one input layer, three hidden layers, and one output layer. The size of the input shape at the input layer is 64×64, the hidden layer has 128, 64, and 32 nodes respectively, and the Rectified Linear Unit (ReLU) [16] is used as an activation function. The output layer has a value between 0 and 1, and softmax [17] is used as the activation function.

Lung Finder sets the lung area as the region of interest (ROI) in the CXR data using the lung segmented image. It can minimize the learning areas other than lungs in pneumonia models. For lung segmentation, one of the artificial neural networks, U-net was used [18, 19]. The training data used were the Montgomery and Japanese Society of Radiological Technology (JSRT) datasets [20]. The lung segmentation model consists of three down-sampling layers and an up-sampling layer. The size of both the input and output shapes is 256×256. In the results of the lung segmentation model, parts smaller than the stated size are removed by post-processing and the lungs are cropped to the ROI.

The Data Rescaler resizes, normalizes, and standardizes the data input size to fit the input layer in the pneumonia model. The data processed by the Data Rescaler is used as training or test data in the pneumonia model.
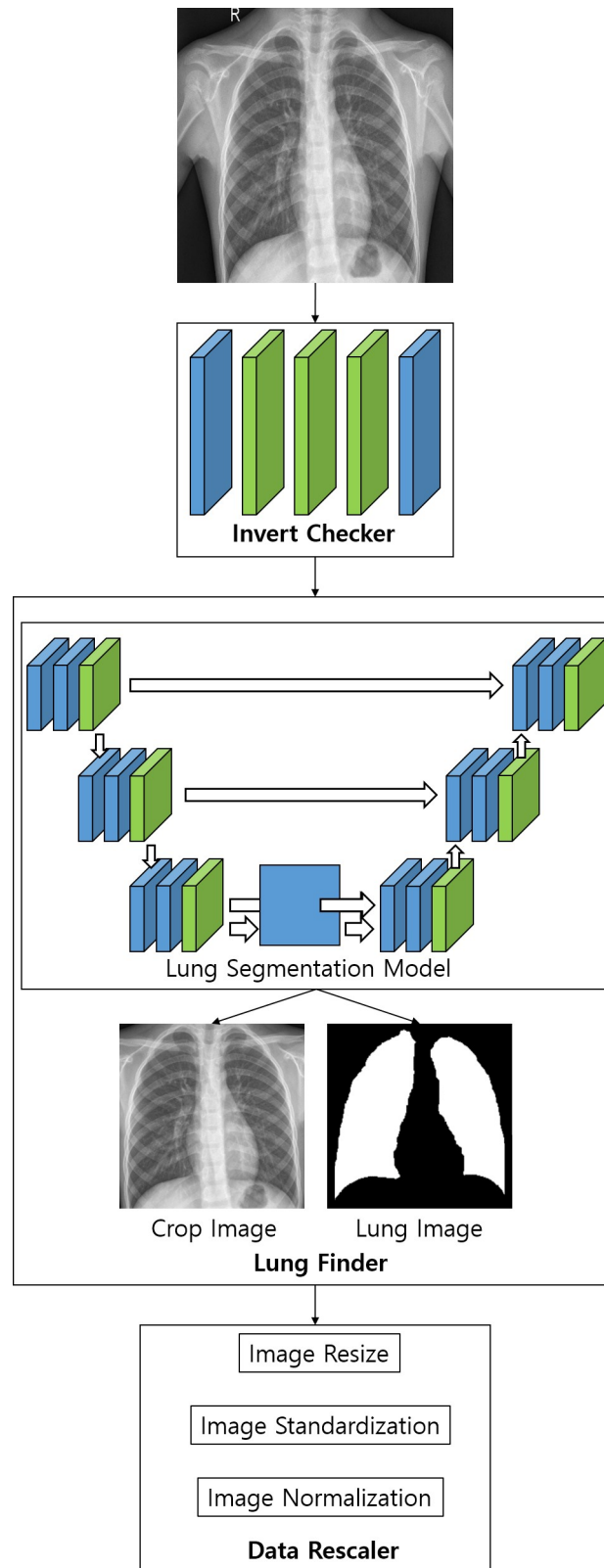
Fig 1 shows the process flow of the pre-processing module used in this study. The data processed in the pre-processing module is used as input data for the pneumonia model.

Classification module

The pneumonia model is based on DenseNet [21]. DenseNet has a structure in which some layers are bundled into dense blocks and then the dense blocks are repeated. It allows the information of the feature map at the input layer to be maintained even when the AI model is deep. The basic architecture of DenseNet is shown in Fig 2.

In this study, DenseNet was used to combine datasets and ten models were trained; a summary of each model and the data augmentation used are shown in Table 2.

The performance of the AI models was quantitatively compared using sensitivity, specificity, and accuracy. In addition, to find the optimal model, each model was combined with others to create and to be used an ensemble model. Since the ensemble model uses a combination of several models, it has the advantage of improving performance as well as reducing the effect of overfitting. The number of gadgets of the ensemble model can be obtained by combining the models. For this study, the number equals to $\sum_{m=1}^{10} {}_{10}C_m$, a total of 1,023 combinations. The probability score of the ensemble model was obtained by using the RMS (Root Mean Square) method on the probability score of the models composing the ensemble model.

Fig 1. Flow diagram of the pre-processing module process.
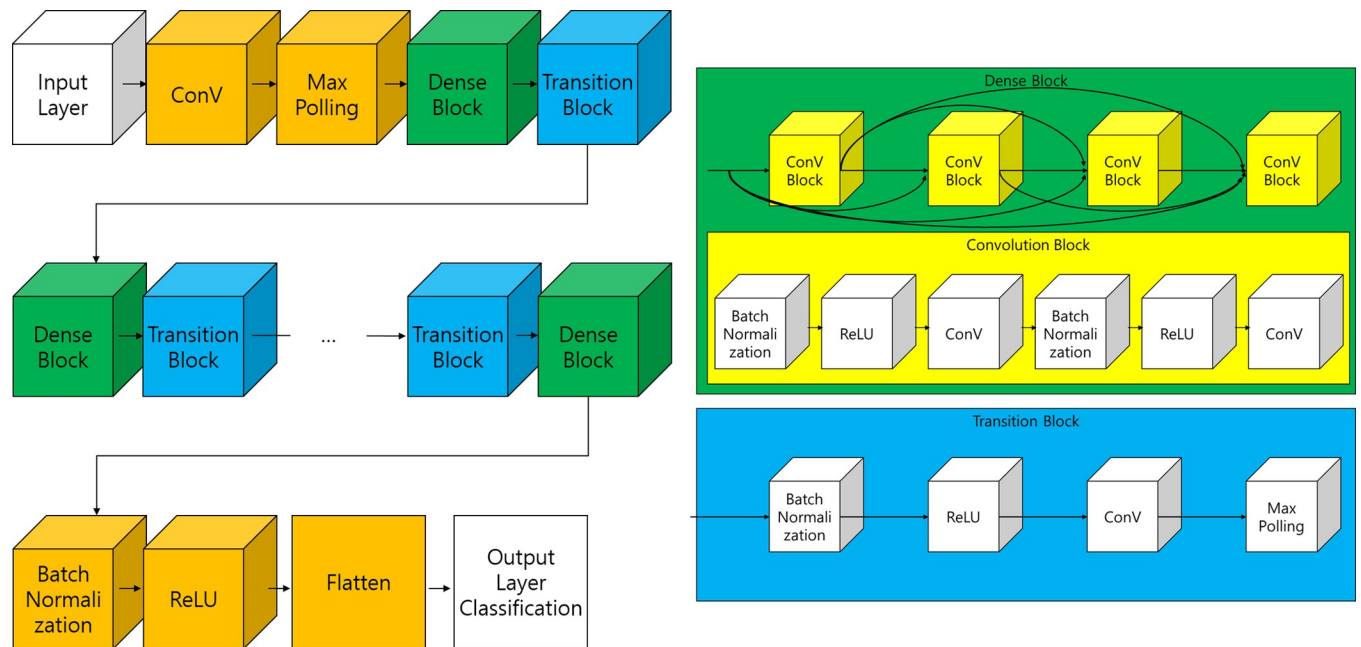
Fig 2. Architecture of DenseNet.

The optimal ensemble model is obtained by sorting the area under the receiver operating characteristic curve (AUC) [22] for the entire combination in descending order, where the top 20 combinations are selected, based on sensitivity and specificity. Since the CXR is regarded as an initial simple assessment method for pneumonia diagnosis, we set the sensitivity to outperform the specificity for the pneumonia model. Therefore, the sensitivity is multiplied by a weight of 0.7 and specificity is multiplied by 0.3. The optimal ensemble model is established by sorting it in descending order based on this score.

Statistical analysis

Results are presented as a percentage for categorical variables and as a mean (± standard deviation) for continuous variables. The diagnostic performance of AI models for pneumonia was compared using test dataset, in terms of sensitivity, specificity, and accuracy, to enable classification as pneumonia or not. Moreover, the AUC for each model was also calculated to evaluate general classification performance.

To evaluate the probability score of the AI pneumonia model that could represent the longitudinal change of pneumonia, the change of probability score (Δ = probability score on FU CXR − probability score on baseline CXR) was compared using the paired sample T-test between improved cases and aggravated cases on CXR taken at the follow-up, 7 days after the initial diagnosis.

The open-source statistical software Python version 3.6.5 was used to analyze and evaluate the AI model and the diagnostic performance. *P*-values of less than 0.05 (two-sided) were considered significant.

Results

Table 3 shows the results of the performance of variable AI models for the diagnosis of pneumonia with external test dataset. AUC, sensitivity, and specificity are used as

Table 2. Summary of AI models for the diagnosis of pneumonia.

Model No.	Description	Data Augmentation
Model 1	Model trained with NIH dataset	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
Model 2	Model trained with NIH dataset	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Histogram equalization
Model 3	Model trained with NIH and KNTA dataset	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
Model 4	Model trained with NIH and KNTA dataset	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Histogram equalization
Model 5	Model trained with NIH dataset first and transfer learning KNTA dataset (using Model 1)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Invert
Model 6	Model trained with NIH dataset first and transfer learning KNTA dataset (using Model 1)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Histogram equalization
		Invert
Model 7	Model trained with NIH dataset (abnormal: pneumonia, effusion, consolidation)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Invert
Model 8	Model trained with NIH dataset (abnormal: pneumonia, effusion, consolidation)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270)
		Histogram equalization
		Invert
Model 9	Model trained with NIH dataset first and transfer learning KNTA dataset (using Model 1)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270, -15 to 15)
		Invert
Model 10	Model trained with NIH dataset first and transfer learning KNTA dataset (using Model 1)	Flip(Vertical, Horizontal)
		Rotation(90, 180, 270, -15 to 15)
		Histogram equalization
		Invert

performance evaluation indicators. Table 3 shows that the AUC of Model 1 was 0.973 and the sensitivity was 0.972, which was also the highest. However, the specificity shows a low value of 0.764, which is difficult to use as a pneumonia model. Therefore, in this study, the optimal ensemble model combination was found by combining each model. The ensemble model compared and analyzed the performance of the values obtained by calculating the probability, which is the result of each model, with the probability of another model that does not allow duplication. The ensemble model's result is calculated by the RMS method. AUC, which is not affected by the threshold, was used as a performance indicator, and through this, a combination of ensemble models for optimal pneumonia diagnosis was found. The results are shown in Table 4, which shows the top three combination ensemble models with high AUC for the diagnosis of pneumonia.

Table 3. Performance of AI models for pneumonia.

Model Number	AUC	Sensitivity	Specificity
Model 1	0.973	0.972	0.764
Model 2	0.311	0.873	0.024
Model 3	0.630	0.802	0.425
Model 4	0.960	0.646	0.995
Model 5	0.726	0.490	0.962
Model 6	0.711	0.448	0.975
Model 7	0.665	0.412	0.914
Model 8	0.623	0.310	0.935
Model 9	0.723	0.624	0.822
Model 10	0.641	0.297	0.987

Abbreviations: AUC, area under the receiver operating characteristic curve.

https://doi.org/10.1371/journal.pone.0249399.t003

It was confirmed that the optimal ensemble model occurs when Models 1 and 4 are combined. The receiver operating characteristic (ROC) curves on the graphs of Models 1 and 4 and the ensemble models (Model 1 and Model 4) are shown in Fig 3.

In addition, the paired sample T-test was performed to evaluate whether the probability score of the AI pneumonia model represents the longitudinal change of pneumonia. Using a follow-up CXR (CXR Day 8) taken seven days after the diagnosis of pneumonia (n = 100) with the original test dataset (CXR Day 1), the patients showed either improvement (n = 85) or aggravation (n = 15) based on the clinical progression and radiologic change (Figs 4 and 5).

The probability score change showed a significant difference between improvement cases and aggravation cases (Δ = -0.06 ± 0.14 vs. 0.06 ± 0.09, for improved cases vs. aggravated cases. P = 0.001) (Fig 6).

Discussion

The technical advance in artificial intelligence (AI) offers rapid and accurate automated diagnosis in medicine. Deep learning algorithm showed excellent performance diagnosis of CXR in lung cancer, pneumonia, tuberculosis, and multiple abnormal findings [13, 14, 23–25]. In a study of deep-learning based neural network algorithm for major thoracic diseases including pneumonia on CXR, the algorithm demonstrated significantly higher performance than that of the physician groups (AUC, 0.983 vs 0.814–0.932; all P < .005) and significant improvements (0.814–0.932 to 0.904–0.958; all P < .005) in performance were observed in physician groups with the assistance of the algorithm [13].

The CXR is one of the most frequently performed procedures in medicine and the first imaging modality for diagnosis of pneumonia. The main role is in the detection or exclusion of the presence of pneumonia. Other roles include the narrowing down of differential

Table 4. Performance of the ensemble models.

Used Model Number	AUC	Sensitivity	Specificity
Model 1 and Model 4	0.983	0.962	0.901
Model 1, Model 4, and Model 5	0.982	0.939	0.929
Model 1, Model 4, Model 5, and Model 10	0.981	0.962	0.885

Abbreviations: AUC, area under the receiver operating characteristic curve.

https://doi.org/10.1371/journal.pone.0249399.t004
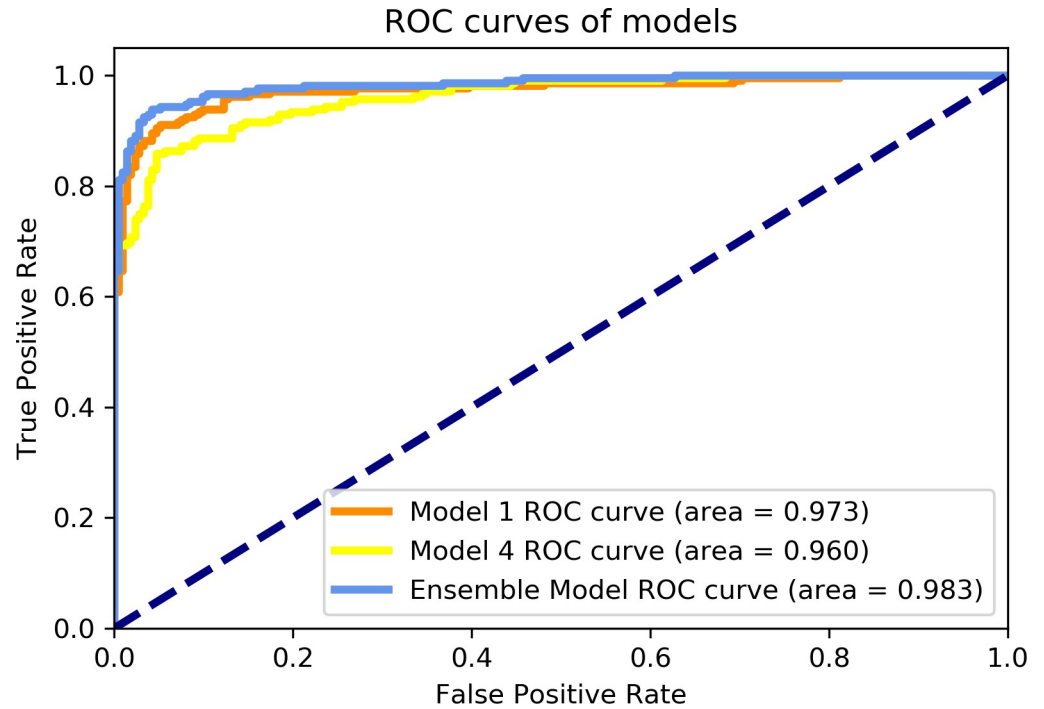
Fig 3. ROC curves of AI models for pneumonia.

https://doi.org/10.1371/journal.pone.0249399.g003

diagnosis of pneumonia with regards to the causative organisms involved. Furthermore, the response to antibiotic therapy could be assessed by evaluating follow-up imaging studies. However, the reading of CXR is also one of the most complex radiology tasks; it represents a huge workload in all institutions and some countries have very few radiologists to interpret CXR. Furthermore, detection errors are common due to the low contrast between lesions and
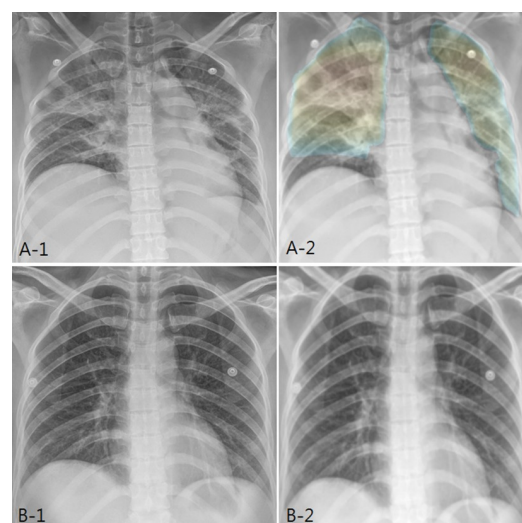
Fig 4. AI probability score changes and heatmap in improved patient. (a) In a 42-year-old female patient, initial chest X-ray shows bilateral patchy increased opacity, compatible with pneumonia. AI pneumonia model (A-2) shows the probability score of 0.787 and color map for pneumonia. (b) After 7 days of treatment, the chest X-ray shows marked improvement. AI pneumonia model (B-2) shows the decrease of probability score (0.318) for pneumonia.

https://doi.org/10.1371/journal.pone.0249399.g004
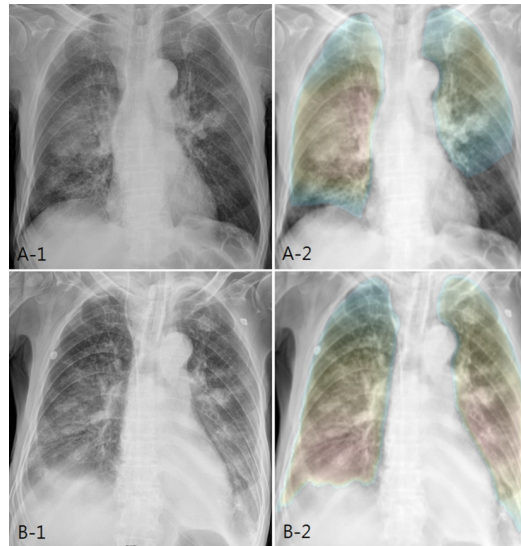
Fig 5. AI probability score changes and heatmap in aggravated patient. (a) In an 81-year-old male patient, initial chest X-ray shows bilateral patchy increased opacity, compatible with pneumonia. AI pneumonia model (A-2) shows the probability score of 0.833 and color map for pneumonia. (b) After 7 days of treatment, the chest X-ray shows slight aggravation. AI pneumonia model (B-2) shows slight increase of probability score (0.876) and increased extent of color map for pneumonia.

https://doi.org/10.1371/journal.pone.0249399.g005
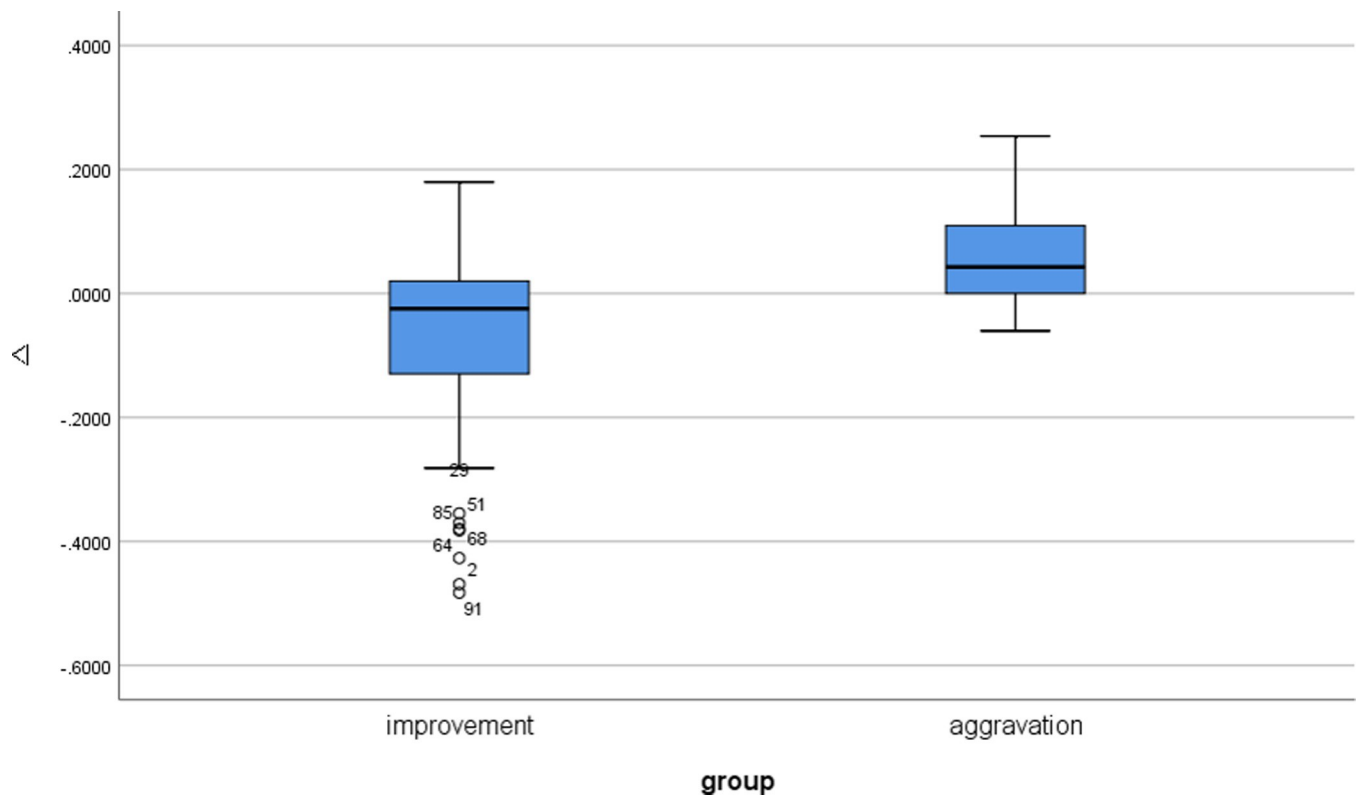
Fig 6. Box plot of probability score changes on follow-up images on pneumonia test dataset (n = 100).

https://doi.org/10.1371/journal.pone.0249399.g006

non-lesion regions such as the surrounding lung and the superposition of bone/mediastinal structures. Computer-assisted tools can be helpful in the reading of CXR. The CAD system has been reported to improve reader accuracy for the detection of lung lesions previously missed on CXR [26]. AI algorithms derived from CXR would also be of great benefit to undertake massive screening programs that could take place in any country with access to X-ray equipment and to support the diagnosis of pneumonia. Since CXR is used across the world due to its affordability, including those areas with few or no radiologists, AI can be applied to facilitate automated screening for pneumonia even in a low-resource setting.

In this study, AI model for pneumonia on CXR was designed and developed. For the developed AI models, not only the performance of single models but also the performance of the ensemble model was tested. As a result, the AUC of 0.973, the highest performance among the single models, was increased to AUC of 0.983 through the ensemble modelling for the diagnosis of pneumonia with the test dataset from a completely different data source to the training data.

Furthermore, the probability score change on FU CXR can represent the clinical course of the pneumonia. One of main role of imaging is to evaluate the response to antibiotic therapy by observing follow-up imaging studies. Thus, we evaluated the AI probability score on a follow-up CXR taken 7 days after the diagnosis of pneumonia. The conventional 10- to 14-day treatment duration anecdotal patterns of behavior is not supported by current evidence. Recent guidelines recommend 5 to 7 days of treatment based on studies showing that a period shorter than 7 days was as effective as longer duration of therapy for mild to moderate community-acquired pneumonia [27]. These suggest the possible application of screening and quantitative analysis to identify patients at risk of poor clinical course.

To evaluate the severity of infection and monitoring of the disease, imaging studies play an important role. The AI-based pneumonia model on CXR images can be one of the fastest and the most quantitative image-based triage tools for diagnosis and treatment monitoring with the fully automated image analysis function to effectively control a large number of suspected or infected patients. A more accurate analysis can be provided by combining the patient's information such as fever, respiratory condition, and underlying conditions.

The present study has several limitations. First, we trained the AI model for pneumonia from NIH and KNTA dataset; the annotation methods were heterogeneous and were not intentionally for pneumonia. The NIH dataset was loosely annotated using NLP and text mining method with several abnormal findings under various clinical settings. The KNTA dataset was annotated by radiologists for the existence of tuberculosis lesion under tuberculous screening setting. Second, we evaluated the diagnostic performance using external dataset of pneumonia cohort from a tertiary university hospital from Korea. Moreover, the normal data of the test dataset, extracted from KNTA and NIH, were not used for training. Finally, we utilized and validated with data solely from Korea and it does not guarantee the same diagnostic performance with other populations of different nationalities, races, and clinical settings.

In conclusion, our study developed an AI pneumonia model on CXR and demonstrates good performance for the diagnosis and follow-up monitoring of pneumonia. There is likely a promising role of the AI pneumonia model as an effective screening method for pneumonia at the primary care level in settings where access to radiologists is limited. Further study is warranted to evaluate the usefulness of the AI model for pneumonia in the diagnosis and monitoring of treatment.

Supporting information

S1 Data.
(XLSX)

Author Contributions

Conceptualization: Sang Pyo Lee, Dongmin Kim, Jinseong Jang, Myungjae Lee, Shin Uk Kang, Heejin Kim, Keunyoung Oh, Jinhee On, Young Jae Kim, Kwang Nam Jin, Eun Young Kim, Kwang Gi Kim.

Data curation: TaeWoo Kwon, Sang Pyo Lee, Heejin Kim, Keunyoung Oh, Jinhee On, So Jeong Yun, Eun Young Kim.

Formal analysis: TaeWoo Kwon, Eun Young Kim.

Funding acquisition: Eun Young Kim.

Investigation: TaeWoo Kwon, Dongmin Kim, Jinseong Jang.

Methodology: TaeWoo Kwon, Dongmin Kim, Eun Young Kim.

Project administration: Dongmin Kim.

Resources: Dongmin Kim, Eun Young Kim.

Software: Dongmin Kim.

Supervision: TaeWoo Kwon, Dongmin Kim, Shin Uk Kang, Young Jae Kim, Kwang Nam Jin, Eun Young Kim, Kwang Gi Kim.

Validation: TaeWoo Kwon, Dongmin Kim.

Visualization: TaeWoo Kwon, Dongmin Kim, Eun Young Kim.

Writing – original draft: TaeWoo Kwon, Dongmin Kim, Eun Young Kim.

Writing – review & editing: TaeWoo Kwon, Sang Pyo Lee, Dongmin Kim, Jinseong Jang, Myungjae Lee, Shin Uk Kang, Heejin Kim, Keunyoung Oh, Jinhee On, Young Jae Kim, So Jeong Yun, Kwang Nam Jin, Eun Young Kim, Kwang Gi Kim.

References

1. Honkinen M, Lahti E, Osterback R, Ruuskanen O, Waris M. Viruses and bacteria in sputum samples of children with community-acquired pneumonia. Clin Microbiol Infect. 2012; 18(3):300–7. Epub 2011/08/20. https://doi.org/10.1111/j.1469-0691.2011.03603.x PMID: 21851481; PubMed Central PMCID: PMC7128628.

2. Ruuskanen O, Lahti E, Jennings LC, Murdoch DR. Viral pneumonia. Lancet. 2011; 377(9773):1264–75. https://doi.org/10.1016/S0140-6736(10)61459-6 PMID: 21435708.

3. Diehr P, Wood RW, Bushyhead J, Krueger L, Wolcott B, Tompkins RK. Prediction of pneumonia in outpatients with acute cough—a statistical approach. J Chronic Dis. 1984; 37(3):215–25. Epub 1984/01/01. https://doi.org/10.1016/0021-9681(84)90149-8 PMID: 6699126.

4. Riquelme R, Torres A, el-Ebiary M, Mensa J, Estruch R, Ruiz M, et al. Community-acquired pneumonia in the elderly. Clinical and nutritional aspects. Am J Respir Crit Care Med. 1997; 156(6):1908–14. Epub 1997/12/31. https://doi.org/10.1164/ajrccm.156.6.9702005 PMID: 9412574.

5. Franquet T. Imaging of pneumonia: trends and algorithms. Eur Respir J. 2001; 18(1):196–208. Epub 2001/08/21. https://doi.org/10.1183/09031936.01.00213501 PMID: 11510793.

6. Mahomed N, Fancourt N, de Campo J, de Campo M, Akano A, Cherian T, et al. Preliminary report from the World Health Organisation Chest Radiography in Epidemiological Studies project. Pediatr Radiol. 2017; 47(11):1399–404. Epub 2017/10/19. https://doi.org/10.1007/s00247-017-3834-9 PMID: 29043423; PubMed Central PMCID: PMC5608771.

7. Sharma S, Maycher B, Eschun G. Radiological imaging in pneumonia: recent innovations. Curr Opin Pulm Med. 2007; 13(3):159–69. https://doi.org/10.1097/MCP.0b013e3280f3bff4 PMID: 17414122.

8. Qin C, Yao D, Shi Y, Song Z. Computer-aided detection in chest radiography based on artificial intelligence: a survey. Biomed Eng Online. 2018; 17(1):113. Epub 2018/08/24. https://doi.org/10.1186/s12938-018-0544-y PMID: 30134902; PubMed Central PMCID: PMC6103992.

9. Young M, Marrie TJ. Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. Arch Intern Med. 1994; 154(23):2729–32. https://doi.org/10.1001/archinte.1994.00420230122014 PMID: 7993157.

10. Malhotra P, Gupta S, Koundal D. Computer Aided Diagnosis of Pneumonia from Chest Radiographs. Journal of Computational and Theoretical Nanoscience. 2019; 16(10):4202–13.

11. Oliveira LL, Silva SA, Ribeiro LH, de Oliveira RM, Coelho CJ, AL SA. Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. Int J Med Inform. 2008; 77(8):555–64. Epub 2007/12/11. https://doi.org/10.1016/j.ijmedinf.2007.10.010 PMID: 18068427.

12. Omar H, Babalık A. Detection of Pneumonia from X-Ray Images using Convolutional Neural Network. Proceedings Book. 2019:183.

13. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. JAMA Netw Open. 2019; 2(3):e191095. https://doi.org/10.1001/jamanetworkopen.2019.1095 PMID: 30901052; PubMed Central PMCID: PMC6583308.

14. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018; 15(11):e1002686. https://doi.org/10.1371/journal.pmed.1002686 PMID: 30457988; PubMed Central PMCID: PMC6245676 following competing interests: CPL holds shares in whiterabbit.ai and Nines.ai, is on the Advisory Board of Nuance Communications and on the Board of Directors for the Radiological Society of North America, and has other research support from Philips, GE Healthcare, and Philips Healthcare. MPL holds shares in and serves on the Advisory Board for Nines.ai. None of these organizations have a financial interest in the results of this study.

15. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM, editors. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

16. Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint arXiv:180308375. 2018.

17. Goodfellow I, Bengio Y, Courville A. Deep learning: MIT press; 2016.

18. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention; 2016: Springer.

19. Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention; 2015: Springer.

20. Shi Y, Qi F, Xue Z, Chen L, Ito K, Matsuo H, et al. Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. IEEE Transactions on medical Imaging. 2008; 27(4):481–94. https://doi.org/10.1109/TMI.2007.908130 PMID: 18390345

21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

22. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. Crit Care. 2004; 8(6):508–12. Epub 2004/11/30. https://doi.org/10.1186/cc3000 PMID: 15566624; PubMed Central PMCID: PMC1065080.

23. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. Clin Infect Dis. 2019; 69(5):739–47. https://doi.org/10.1093/cid/ciy967 PMID: 30418527; PubMed Central PMCID: PMC6695514.

24. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology. 2019; 290(1):218–28. https://doi.org/10.1148/radiol.2018180237 PMID: 30251934.

25. Hwang EJ, Nam JG, Lim WH, Park SJ, Jeong YS, Kang JH, et al. Deep Learning for Chest Radiograph Diagnosis in the Emergency Department. Radiology. 2019; 293(3):573–80. https://doi.org/10.1148/radiol.2019191225 PMID: 31638490.

26. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. Eur J Radiol. 2020; 123:108774. https://doi.org/10.1016/j.ejrad.2019.108774 PMID: 31841881.

27. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007; 44 Suppl 2:S27–72. https://doi.org/10.1086/511159 PMID: 17278083; PubMed Central PMCID: PMC7107997.