A Study on the Clinical Effectiveness of Deep Learning CAD Technology

Han Ju-Hyuck
Department of Medical Engineering
Konyang University
Daejeon, South Korea
dnfwlq203@gmail.com

Oh Hyun-Woo
Bio Convergence Technology Training
Project Group, Konyang University
Daejeon, South Korea
osj0805@naver.com

Kim Woong-Sik*
Department of Medical A.I.
Konyang University
Daejeon, South Korea
wskim@konyang.ac.kr

***Abstract*—Chest radiography is the most common method of examining chest disease. However, interpretation of chest X-rays is difficult, and the diagnosis may vary depending on the doctor's proficiency. In order to solve this problem, additional diagnosis using a computer is attracting attention in the medical imaging field. In addition, the recently developed artificial intelligence technology has been applied to the analysis of chest X-rays, and commercialization has entered the stage as a computer-aided diagnostic tool. However, the reading model based on artificial intelligence has different performance depending on the type of data. In addition, current medical data is a weak standardization stage and the data form varies from institution to institution. Therefore, the performance of the model may not be guaranteed if the data for training artificial intelligence and the data from the real institution are different. The purpose of this study is to verify the clinical effectiveness of a computer-aided diagnostic tool based on chest X-rays. To this end, data from a different source than the training data were applied to the reading model. In addition, for validation, we prepared a doctor's lung lesion labeling findings for clinical validation. In this study, OPT (Observer Performance Test) was conducted by clinical experience level to evaluate the reading model.**

***Keywords*—Chest Radiography, Deep learning Algorithm, Observer Performance Test, CAD**

I. INTRODUCTION

Chest radiography is the most common method for examining chest diseases and monitoring chest abnormalities such as lung cancer[1]. However, interpretation of chest X-Ray (CXR) is difficult and misreadable, requiring a lot of image analysis experience[2]. In other words, the reading of chest radiographs depends on the physician's clinical experience. Recently, in order to solve this problem, the development of computer aided diagnosis (CAD) is growing. In addition, advanced artificial intelligence technology is being applied to the medical imaging field. The purpose of this paper is to verify the clinical effectiveness of artificial intelligence models that support the interpretation of chest radiographs. In addition, three cohorts were organized and studied as methods for verification.

II. METHODE

A. CAD and Data Configuration

In this paper, Lunit INSIGHT for Chest Radiography certified by the Ministry of Food and Drug Safety in Korea, was used to clinically evaluate CAD performance based on deep learning. The performance evaluation was applied by dividing a cohort of patients who visited respiratory outpatient hospitals in three institutions in 2018 and underwent chest radiography. All data used in this study are retrospective data approved by the instrumental review boards, and the requirements for patient consent have been omitted. Data screening targeted 26,988 patients.

There are two criteria for not selecting data. First, the unexamined case of chest CT (n=17,871), secondly, the interval between chest CT and chest X-Ray is more than 1 month (n=3,165). Therefore, as shown in Table 1, the final data for performance comparison is 6,006 people's data. Data collection is based on CXR, and meaningful patient information (age, gender, chest CT, smoking history, past history) was collected from electronic medical records. In addition, in the case of images, they were collected by the picture archiving and communication system (PACS) and all of them were de-identified for research.

TABLE I. DEMOGRAPHIC DATA OF RESPIRATORY PATIENTS

Category	Institutions			Total	Dataset for OPT	P value
	B	*G*	*K*			
No. of Patients	2536	1470	2000	6006	230	-
Female	1166	643	798	2607	107	0.53
Male	1370	827	1202	3398	123	0.50
Age	61 ±16	61 ±14	61 ±16	61 ±16	60 ±16	0.21
Interval CXR & CT	3 ±9	3 ±11	1 ±7	2 ±9	2 ±9	0.42
No. of PA image	2536	1421	1952	5908	229	0.15

Table 1 compares the notations of radiologists and CAD at three institutions. CAD marks Activation Map at the location of the abnormal lesion. This was considered positive when there was a coincidence rate of more than 15% at the center of the lesion based on pixels. Previous studies showed that the AUC of Test1 was 0.814 and the AUC of Test2 was 0.904 [3]. It has a test set number of 230, and constitutes a dataset of positive 0.4, negative 0.6. In this study, 230 random sampling data were composed of OPT Data Set to apply the same criteria as in previous studies. In addition, CT images are used to compare with CXR. The CT image was selected based on the closest shooting date from the CXR shooting date. The criteria for chest abnormalities were selected by radiologists. They analyzed the dataset to define reference settings for chest abnormalities. In this study, when disagreement occurs in the diagnosis, an investigation was conducted to agree on it. The CXR lesion in this study consisted of nodules/species, integration, and pneumothorax as target lesions. In addition, lethargy or fibrosis, bronchiectasis, heart lesions, diffuse lung genes, longitudinal lesions, and pleural effusion were composed of comparative lesions for patient comparison. This classification referred to the labeling standard of the ChestX-ray14 dataset or MIMIC-CXR database[4][5][6]. This study was conducted with the approval of the Institutional Review Committee[7].

TABLE II. LESION TYPES OF CHEST RADIOGRAPHY IN PATIENTS

Lesion	Institutions			Total	Random sample for OBT	P value
	B	*G*	*K*			
Target Lesions						
Nodule/ mass	446 (18)	259 (18)	468 (23)	1173 (20)	41 (18)	0.79
Consolidation	341 (13)	212 (14)	366 (18)	919 (15)	35 (15)	0.99
Pneumothorax	5 (0.3)	2 (0.1)	8 (0.4)	15 (0.2)	2 (0.9)	0.87
Non-target Lesions						
Atelectasis or fibrosis	93 (4)	62 (4)	185 (9)	340 (6)	15 (7)	0.90
Bronchiectasis	217 (9)	286 (20)	107 (5)	610 (10)	27 (12)	0.80
Cardiomegaly	21 (0.8)	48 (3)	67 (3)	136 (2)	4 (2)	0.94
Diffuse interstitial lung opacities	115 (5)	73 (5)	65 (3)	253 (4)	10 (4)	0.99
Mediastinal lesion	11 (0.4)	27 (2)	36 (2)	74 (1)	4 (2)	0.93
Pleural effusion	81 (3)	29 (2)	76 (4)	186 (3)	7 (3)	0.99
Other	188 (7)	172 (12)	198 (10)	558 (9)	28 (12)	0.61
Total						
Sum of target or non-target lesions	1518	1170	1576	4264	173	N/A
Participants with any types of lesions	1317 (52)	889 (61)	113 (57)	3337 (56)	137 (60)	0.36
No. of lesion type per patient	1.2 (1-3)	1.3 (1-4)	1.4 (1-5)	1.3 (1-5)	1.3 (1-4)	0.83

Table 2 shows the configuration of the data set used in this study. According to this, out of 4,274 reference chest abnormal lesions of 6,006 CXR, pulmonary nodules/tumor, aggregation, pneumothorax, and other reference chest abnormalities were found in 1,173 (20%), 919 (15%), 15 (0.2%), and 2,157 (51%) CXRs, respectively. Among the 26 classified final diagnoses, pneumonia was the most common diagnosis (n = 696, 12%). Pulmonary tuberculosis and malignant neoplasm (neoplasm) of the bronchial tubes or lungs were found in CXRs of 550 (9%) and 355 (6%), respectively.

B. Verification Method

In this study, OPT was constructed to evaluate CAD. The OPT of this study was conducted by dividing the number of data collection days to prevent data bias. In OPT Test 1, observers conducted CXR evaluation alone without CAD help. It provided observers with CXR and CT, and patient information (gender, age, etc.). The observers consisted of 12 doctors, including 3 chest radiologists, 3 board-certified radiologists, 3 radiologists, and 3 lung specialists. They constructed the same form as the result of CAD by marking chest anomalies on the image. In OPT test 2, observers were assisted by CAD. This is based on the patient's CXR image and patient information, indicating chest abnormalities in the image. In addition, if CAD 's Activation Map matches the observer's lesion indication, it was treated as true positive. On the other hand, if the CAD's Activation Map and the observer's lesion indication did not match, they were marked as false positive and false negative. In the case of false positive, CAD was marked positive, but there was no indication from the observer. In the case of false negative, the observer's notation exists, but there is no CAD notation. We confirmed the effect of CAD by doctors' clinical experience through OPT.

C. CAD Model

In this study, Lunit Insight-CXR was used as CAD[8]. According to this, the model was used by extracting data sets for model training from six multinational multi-centers. In addition, a technique to mark chest abnormalities in CXR was used. The model consists of 27 layers and 12 residual connections based on convolutional neural networks (CNNs). This model uses a semisupervised localization approach with partial data annotation.

III. Result

In this study, receiver operating characteristics (ROC) curves and jackknife free-response receiver operating characteristic (JAFROC) curves are used for outcome indicators. The receiver operating characteristic curve is calculated as a true-positive rate and a false-positive rate. In addition, the jackknife replacement free response ROC curve is calculated with the local fraction of the lesion to the probability of false positive (FP) per normal CXR. The number of false positive marks per image is defined as the value obtained by dividing the number of false positive marks by the total number of radiographs. Statistical analysis was performed using Medcalc version 19.5.1 or R version 3.5.3. All statistical analyses were performed using R software version 3.6.1.

TABLE III. Performance of observer group in the randomly sampled dataset (n=230)

Observer Group		***Test1***	***Test2***	****P value***	***†P value***
AUC	Thoracic radiologist(n=3)	0.89 (0.84,0.93)	0.89 (0.84,0.95)	0.21	0.58
	Board-certified radiologist(n=3)	0.87 (0.83,0.91)	0.89 (0.83,0.95)	0.14	0.12
	Radiology residents(n=3)	0.85 (0.80,0.89)	0.88 (0.85,0.91)	0.07	0.03
	Pulmonologist (n=3)	0.84 (0.80,0.85)	0.88 (0.85,0.92)	0.03	0.01
JAFROC	Thoracic radiologist(n=3)	0.82 (0.75,0.89)	0.84 (0.76,0.91)	0.03	0.60
	Board-certified radiologist(n=3)	0.80 (0.76,0.85)	0.82 (0.76,0.88)	0.29	0.12
	Radiology residents(n=3)	0.79 (0.72,0.85)	0.83 (0.79,0.87)	0.05	0.10
	Pulmonologist (n=3)	0.78 (0.73,0.83)	0.81 (0.77,0.75)	0.07	0.04

Table 3 shows the OPT results. This was constructed to confirm the effect of CAD according to the clinical experience of doctors. According to Table 3, Test 1 and Test 2 to view the influence of CAD, the average AUC was 0.86 (95% CI: 0.82, 0.90) to 0.89 (95% CI: 0.85, 0.92). the average JAFROC was 0.92 at 0.80 (95% CI: 0.76, 0.84). First, for Test 1 without CAD, a thoracic radiologist showed AUC: 0.87 and JAFROC: 0.80 for radiation resistors, AUC: 0.85 and JAFROC: 0.79, pulmonary tuberculosis specialist showed AUC: 0.89 and JAFROC: 0.82 for radiation residents. On the other hand, Test 2 assisted by CAD showed AUC: 0.89, JAFROC: 0.82 for chest radiologists, AUC: 0.88, JAFROC: 0.83 for radiation resistors, and AUC: 0.89 for pulmonary tuberculosis specialists showed AUC: 0.89 and JAFROC: 0.84. In OPT to verify the clinical effectiveness of CAD, Test 2 showed better performance than Test 1. In particular, the less experienced doctors, the more significantly their diagnostic ability in the assisted state of CAD. In addition, it can be seen that JAFROC, which quantifies the local features of the lesion, increased by about 2%–3% in all observer groups.

IV. Conclusion

This study was conducted to verify the clinical validity of CAD. This confirmed the results by directly applying data from the learned data set and data from a different real-world medical environment to CAD. The effectiveness of CDA was discovered by using CAD in OPT, which consisted of groups by clinical experience. The clinical data application performance of CAD used in this study showed an AUC of 0.87 at 0.86 and an JAFROC of 0.87 at 0.86 as a sole test. Also, the performance of OPT is from AUC 0.814 to 0.932, 0.904 to 0.958 and JAFROC from 0.781 to 0.907, 0.873 to 0.938. The results of the observer performance test show that CAD improved the ability of the observer (chest radiologist, radiologist, lung cancer specialist) to detect chest abnormalities. This means that CAD has a clinical effect on locating lesions. Therefore, it can be said that the accuracy of CAD is guaranteed even if it is clinical data from a source (hospital) other than the learned data. In addition, in the case of doctors with long clinical experience, it can be effective as CAD.

Acknowledgment

This thesis was conducted with the support of the Korea Industrial Technology Promotion Agency's "Bio Convergence Technology Professional Training Project" with the funding of the government (Ministry of Trade, Industry and Energy) in 2021 (No. P0017805).

References

[1] B.P. Little, M.D. Gilman, K.L. Humphrey, T.K. Alkasab, F.K. Gibbons, J.O. Shepard, and C.C. Wu, "Outcome of recommendations for radiographic follow-up of pneumonia on outpatient chest radiography," American Journal of Roentgenology, vol. 202, pp. 54-59, 2014.

[2] H.B. Harvey, M.D. Gilman, C.C. Wu, M.S. Cushing, E.F. Halpern, J. Zhao, P.V. Pandharipande, J.O. Shepard, and T.K. Alkasab, "Diagnostic yield of recommendations for chest CT examination prompted by outpatient chest radiographic findings," Radiology, vol. 275, pp. 262-271, 2015.

[3] E.J. Hwang, S. Park, K-N. Jin et al, "Development and validation of a deep learning–based automated detection algorithm for major thoracic diseases on chest radiographs," JAMA network open 2.3, e191095-e191095, 2019.

[4] P. Rajpurkar, J. Irvin, K. Zhu, et al, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint, arXiv:1711.05225, 2017.

[5] D.M. Hansell, A.A. Bankier, H. MacMahon, T.C. McLoud, N.L. Müller, and J. Remy, et al, "Fleischner Society: glossary of terms for thoracic imaging," Radiology, vol. 246, pp. 697-722, 2008.

[6] A.E.W. Johnson, T.J. Pollard, N.R. Greenbaum, et al, "MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs," arXiv preprint, arXiv:1901.07042, 2019.

[7] World Health Organization, ICD-10 Version: 2016, apps.who.int/classifications/icd10/browse/2016/en. F00-F09, February 2016.

[8] E.J. Hwang, S. Park, K.N. Jin, J.I. Kim, S.Y. Choi, J.H. Lee, J.M. Goo, J. Aum, J.J. Yim, C.M. Park, and Deep Learning-Based Automatic Detection Algorithm Development and Evaluation Group, "Development and validation of a deep learning–based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs," Clinical Infectious Diseases, vol. 69, pp. 739-747, 2019.