Reproducibility Analysis on Multiple Abnormality Detection with Convolutional Neural Net in Chest PA X-ray Images within Short-term Period

Yongwon Cho  Young-Gon Kim  Geunhwi Ahn  Eunsol Lee  Younghoon Cho
Sangmin B Lee  Joonbeom Seo  Namkug Kim

Asan Medical Center
University of Ulsan College of Medicine
{dragon1won, younggon2.kim, dksrmsgn192, eslee0214, yhcho84}@gmail.com
{sangmin.lee.md, namkugkim, joonbeom.Seo}@gmail.com

Sejin Park
VUNO Seoul Republic of Korea
gnoses@vuno.co

Abstract

Numerous chest PA X-ray images could be used for computer-aided diagnosis (CAD) to improve the radiologist’s reading quality. So far, reproducibility of CAD, however, has not been studied intensively yet. Therefore, we evaluated reproducibility of computer aided detection on abnormality with convolutional neural net (CNN) in chest PA X-ray images of same patient acquired within short-term. We compared reproducibility of four algorithms in pulmonary nodular lesions and evaluated the reproducibility of five abnormalities. The percent positive agreement (PPA) was 91.26%, 91.83%, 90.45%, and 91.50%, with YOLO v2, MASK RCNN, Faster RCNN, and VUNO-Net, respectively. In addition, the PPA with YOLO v2 was 89.10%, 88.37%, 100%, 98.41%, and 89.60% for nodule, consolidation, interstitial opacity, pleural effusion, and pneumothorax, respectively.

1 Introduction

Chest radiography is still the first examination in patients with nonspecific symptoms for evaluation of the thorax in routine clinical practice. As chest radiography is performed fast and easily with low cost as well as readily available in most institutes, a large number of studies are conducted and it becomes a practical burden for radiologists to read all exams maintaining the quality of diagnosis. Missing lesions detected retrospectively are frequently found even by experienced radiologists [1]. More recently, computer-aided diagnosis (CAD) systems have been developed to use in clinical filed that help radiologists to detect potential and hidden lesions on diagnostic radiography. Currently, convolutional neural net (CNN) approach is considered as state of the art of the art of machining learning technique for lesion detection and classification. However, reproducibility of CAD with CNN has not been evaluated intensively yet. Therefore, we compared the reproducibility of not only the detection of small nodular lesions (<24mm), but also middle-sized nodular lesions (>24mm) with four different CNN methods. In addition, we evaluated the reproducibility of computer aided detection on five abnormalities in X-ray images with the YOLO v2 network [4].

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.
2 Materials and Method

2.1 Datasets

We obtained a variety of chest PA X-ray dataset with 3722 patients extracted from the Asan Medical Center (AMC). The datasets of chest PA X-ray images were cleansed and anonymized to analyze reproducibility of CAD with CNN. Among them, 4678 abnormality ROIs were manually drawn by two expert thoracic radiologists in consensus. For training and testing, all data were randomly split into 80 percent and 20 percent for training and test, respectively. To evaluate reproducibility of CAD on five abnormalities including nodule (N), consolidation (C), interstitial opacity (I), pleural effusion (PE), and pneumothorax (PT), chest PA X-ray images (N=121 pairs, C=28 pairs, IO=12 pairs, PE=67 pairs, and PT=20 pairs) of same patient within one week were independently acquired. In addition, to evaluate the reproducibility of the nodule with CAD with four different CNN methods, we have independently trained nodule datasets with 756 images (1093 ROIs) and evaluated reproducibility with chest PA X-ray (N=121 pairs).

2.2 Evaluation metrics of reproducibility

To evaluate reproducibility of five abnormalities including N, C, IO, PE, and PN, we proposed a CAD based on CNN such as YOLO v2 [3] that can detect five lesions. In addition, because it is important to detect small nodular lesions (<24mm) as well as middle-sized nodular lesions (>24mm) in chest PA X-ray images, YOLO v2 [3] was compared with those of three CNNs such as Mask RCNN[4], Faster RCNN[5], and VUNO-Net to analyze reproducibility of detection on pulmonary nodular lesions with percent positive agreement (PPA) (1.1) [2][3]. Percent Positive Agreement (PPA)= 100 x 2a/(2a+b+c) (1.1) where a is top right and b is top left and c is bottom right in the confusion matrix in figure 1.2

3 Results

We have calculated the reproducibility of CAD on five abnormalities in chest PA X-ray images within the short-term period. The PPA in test set was 89.10%, 88.37%, 100%, 98.41%, and 89.60% for N, C, IO, PE, and PN, respectively. The consolidation showed the worst reproducibility and the interstitial opacity is the best reproducibility in Figure 1.

![Figure 1: Examples of computer aided abnormality detection including nodule, consolidation, interstitial opacity, pleural effusion, and pneumothorax with YOLO v2. For reproducibility evaluation, confusion matrixes were evaluated](image)

The reproducibility of CAD with four different algorithms on pulmonary nodules in chest PA X-ray images was evaluated. The PPA (1.1) in test set of nodule was 91.26%, 91.83%, 90.45%, and 91.50% for YOLO v2 [4], MASK RCNN [5], Faster RCNN [6], and VUNO-Net [7], respectively in Figure 2.
Figure 2: Examples of computer aided detection on five abnormalities including nodule, with YOLO v2, MASK RCNN, Faster RCNN, and VUNO-Net. Confusion matrixes of reproducibility were calculated.

4 Discussion and Conclusion

For reproducibility of CAD on five abnormalities, we collected paired data set acquired within one week. The CAD on lung nodules in chest X-ray images with four different CNN models shows comparable reproducibilities. However, range of reproducibility was 8%∼12% depending on evaluation metrics. In addition, CAD on multiple abnormalities in chest X-ray images with CNN showed variable reproducibility. Therefore, for evaluation of CAD performance in later, reproducibility testing will be necessary.

Acknowledgement

This work is partially supported by the Industrial Strategic technology development program (10072064) funded by the Ministry of Trade Industry and Energy (MI, KOREA), and by Kakao and Kakao Brain corporations.

References


