





Article

How Many Private Data Are Needed for Deep Learning in Lung Nodule Detection on CT Scans? A Retrospective Multicenter Study

Jeong Woo Son ^{1,†}, Ji Young Hong ^{2,†} , Yoon Kim ^{1,3}, Woo Jin Kim ⁴ , Dae-Yong Shin ⁵, Hyun-Soo Choi ^{1,3,*} , So Hyeon Bak ^{6,*} and Kyoung Min Moon ^{7,*} 

¹ ZIOVISION, Chuncheon 24341, Korea; jeongwoo12@ziovision.co.kr (J.W.S.); yooni@kangwon.ac.kr (Y.K.)

² Division of Pulmonary and Critical Care Medicine, Department of Medicine, Chuncheon Sacred Heart Hospital, Hallym University Medical Center, Chuncheon 24253, Korea; mdhong@hallym.or.kr

³ Department of Computer Science and Engineering, College of IT, Kangwon National University, Chuncheon 24341, Korea

⁴ Department of Internal Medicine, Kangwon National University, Chuncheon 24341, Korea; pulmo2@kangwon.ac.kr

⁵ KNU-Industry Cooperation Foundation, Kangwon National University, Chuncheon 24341, Korea; dyshin@kangwon.ac.kr

⁶ Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul 05505, Korea

⁷ Department of Pulmonary, Allergy and Critical Care Medicine, Gangneung Asan Hospital, University of Ulsan College of Medicine, Gangneung 25440, Korea

* Correspondence: choi.hyunsoo@ziovision.co.kr (H.-S.C.); arsgnm17@gmail.com (S.H.B.); pulmogicu@ulsan.ac.kr (K.M.M.); Tel.: +82-33-250-8452 (H.-S.C.); +82-2-3010-3491 (S.H.B.); +82-33-610-3058 (K.M.M.)

† These authors contributed equally to this work.



Citation: Son, J.W.; Hong, J.Y.; Kim, Y.; Kim, W.J.; Shin, D.-Y.; Choi, H.-S.; Bak, S.H.; Moon, K.M. How Many Private Data Are Needed for Deep Learning in Lung Nodule Detection on CT Scans? A Retrospective Multicenter Study. *Cancers* **2022**, *14*, 3174. <https://doi.org/10.3390/cancers14133174>

Academic Editors: Andreas Stadlbauer, Anke Meyer-Baese and Max Zimmermann

Received: 15 May 2022
Accepted: 24 June 2022
Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: The early detection of lung nodules is important for patient treatment and follow-up. Many researchers are investigating deep-learning-based lung nodule detection to ease the burden of lung nodule detection by radiologists. The purpose of this paper is to provide guidelines for collecting lung nodule data to facilitate research. We collected chest computed tomography scans reviewed by radiologists at three hospitals. In addition, several experiments were conducted using the large-scale open dataset, LUNA16. As a result of the experiment, it was possible to prove the value of using the collected data compared to using LUNA16. We also demonstrated the effectiveness of transfer learning from pre-trained learning weights in LUNA16. Finally, our study provides information on the amount of lung nodule data that must be collected to stabilize lung nodule detection performance.

Abstract: Early detection of lung nodules is essential for preventing lung cancer. However, the number of radiologists who can diagnose lung nodules is limited, and considerable effort and time are required. To address this problem, researchers are investigating the automation of deep-learning-based lung nodule detection. However, deep learning requires large amounts of data, which can be difficult to collect. Therefore, data collection should be optimized to facilitate experiments at the beginning of lung nodule detection studies. We collected chest computed tomography scans from 515 patients with lung nodules from three hospitals and high-quality lung nodule annotations reviewed by radiologists. We conducted several experiments using the collected datasets and publicly available data from LUNA16. The object detection model, YOLOX was used in the lung nodule detection experiment. Similar or better performance was obtained when training the model with the collected data rather than LUNA16 with large amounts of data. We also show that weight transfer learning from pre-trained open data is very useful when it is difficult to collect large amounts of data. Good performance can otherwise be expected when reaching more than 100 patients. This study offers valuable insights for guiding data collection in lung nodules studies in the future.

Keywords: lung nodule; radiologist; deep learning; computed tomography; nodule detection; publicly available data; transfer learning

1. Introduction

According to the World Health Organization, lung cancer is a life-threatening disease worldwide [1]. In 2018, lung cancer was the largest new cancer diagnosis (11.6%) and the largest cause of cancer-related deaths (18.4%) globally [2,3]. One of the main causes of high mortality in lung cancer is the absence of early overt clinical symptoms. The extent of lung cancer at the time of diagnosis is closely related to prognosis. Cancer Research UK reported that the survival rates for stages I and IV were 87% and 19% in patients with lung cancer, respectively [4]. Therefore, early detection and diagnosis are important for the curative treatment of lung cancer.

The National Lung Screening Trial showed that low-dose computed tomography (LDCT) reduces mortality by 20% compared with chest X-rays [5]. Posterior studies have shown that the early detection of lung nodules can increase the five-year survival rate of lung cancer patients. The five-year survival rate of lung cancer, which was estimated at 18% of all cases diagnosed between 2004–2010, is the lowest among several types of cancer [4]. However, manual detection of lung nodules in many CT scans is a very difficult, time-consuming, and tedious process, even for experienced radiologists [6–8]. Although chest CT is the latest and most widely used imaging tool to capture lung images [5], the manual detection of lung nodules by radiologists on CT images can easily lead to misdiagnosis [8]. An object detection system for lung nodules is necessary and several benchmark datasets have been used [9–13].

Consequently, artificial intelligence (AI)-assisted lung nodule detection systems may be used to provide a second opinion for radiologists and make final decisions faster and more accurately. Deep-learning models are highly dependent on datasets; thus, they can effectively achieve the performance of advanced learning algorithms when high-quality datasets are used for training [14]. Public databases such as LIDC/IDRI [9], LUNA16 [10], NSCLC [11], ELCAP [12], and ANODE09 [13] have been used for lung nodule diagnosis research. Nevertheless, these databases cannot be easily used to develop nodule detection systems because their format is not standardized. Therefore, it is important to create high-quality, real-world lung nodule datasets in a standardized format for training.

Recently, deep learning has been actively and successfully developed in various fields, such as image analysis, natural language processing, and speech recognition [15–18]. A convolutional neural network (CNN), a type of neural network that uses convolution operators, has been widely applied with excellent performance in image analysis. This success has contributed to the adoption of CNN in medical image analysis [19,20]. CNN performs well in deep learning technologies such as classification, object detection, and segmentation. In this study, we used CNN-based object detection to detect lung nodules by analyzing CT scan images. Object detection is a task that allows the drawing of a bounding box on a lung nodule to determine its exact location. CNN-based object detection tasks are making a lot of progress. Faster R-CNN [21], RetinaNet [22], EfficientDet [23], and YOLOX [24], etc., which were state-of-art models at the time, show the development of object detection. In this study, we experimented with lung nodule detection based on the YOLOX model. The specific details of the model used are described in Section 2.4.

It is necessary to train with a large amount of data to accurately detect lung nodules. However, it is difficult for most researchers to collect large amounts of medical data. To mitigate the difficulty of the lack of medical data, we conducted a study focusing on how much data should be collected at the beginning of a lung nodule study and what to do if it is difficult to collect a large amount of data. This study also considers how to leverage publicly available datasets when difficult to collect large amounts of data. Open datasets related to lung nodules [9–13] contain large amounts of data. However, open and privately

collected data differ in many ways, for example in the information they offer on patient race and the way they annotate nodules. Therefore, open data are used; however, training that is suitable for private data is also required. Transfer learning is used to pre-train models on open data and transfers that knowledge to private data [25]. In this study, we demonstrated the usability of open datasets and the importance of building private datasets through various experiments using both open and private data.

The main contributions of this study are: (1) The construction of a high-quality chest CT scan dataset reviewed by radiologists from three hospitals for lung nodule detection; (2) The demonstration of the importance of collecting private data by comparing open data with collected private data; (3) The demonstration that transfer learning using open data is highly effective when a large amount of data cannot be collected; (4) Through empirical experiments, the provision of insights into the necessary amount of data to collect.

2. Materials and Methods

2.1. Datasets

2.1.1. LUNA16

LUNA16 is a publicly available dataset for lung nodule detection and a subset of the LIDC/IDRI dataset. The LIDC/IDRI dataset contains 1018 CT scans and annotations confirmed by four experienced radiologists. The radiologists reviewed each CT scan and marked the lesions in three categories: nodules ≥ 3 mm, nodules < 3 mm, and non-nodules. The LUNA16 dataset used in this study only considers nodules ≥ 3 mm in diameter. The LUNA16 dataset also excludes CT scans with a slice thickness greater than 3 mm, inconsistent slice spacing, and missing slices from the LIDC/IDRI dataset. The LUNA16 dataset contains 888 chest CT scans and annotations consisted of the centroid coordinates and diameters of lung nodules.

2.1.2. Private Dataset

We retrospectively reviewed the medical records and contrast-enhanced chest CT scans of patients with solitary pulmonary nodules (SPN) that were followed up at Gangneung Asan Hospital (GNAH), Kangwon National University Hospital (KNUH), and Hallym Sacred Heart Hospital (HSHH) between March 2016 and February 2021. The private dataset was reviewed by radiologists in three hospitals, and chest CT scan data were collected from 515 patients with lung nodules (GNAH, 287; KNUH, 181; HSHH, 47). We used the Digital Imaging and Communications in Medicine (DICOM) file format.

There were four inclusion criteria: (1) the patient was aged between 18 and 90 years; (2) the nodule diameter in the chest CT axial image was more than 7 mm and less than 30 mm [26]; (3) the pathological diagnosis was performed using percutaneous needle aspiration (PCNA) biopsy, surgical operation, bronchoscopic biopsy, or endobronchial ultrasonography-guided transbronchial needle aspiration (EBUS-TBNA); and (4) in the absence of a pathological diagnosis, follow-up was performed with contrast-enhanced chest CT scan for more than two years. Tables A1–A3 in Appendix A showed the clinical data of the private dataset that explained inclusion criteria. In the NSCLC category in Tables A2 and A3, the sub-NSCLC category is NSCLC with pathological findings that do not fall into any category of adenocarcinoma, squamous cell carcinoma, or large cell endocrine carcinoma.

In this study, we defined an SPN as benign when (1) the pathological diagnosis was confirmed; and (2) there was no increase in size, or there was disappearance of SPN when followed up by a chest CT scan for more than two years [27]. An SPN was defined as malignant only when the pathological diagnosis was confirmed.

The three datasets were formed as follows, using the chest CT protocols outlined in Table 1:

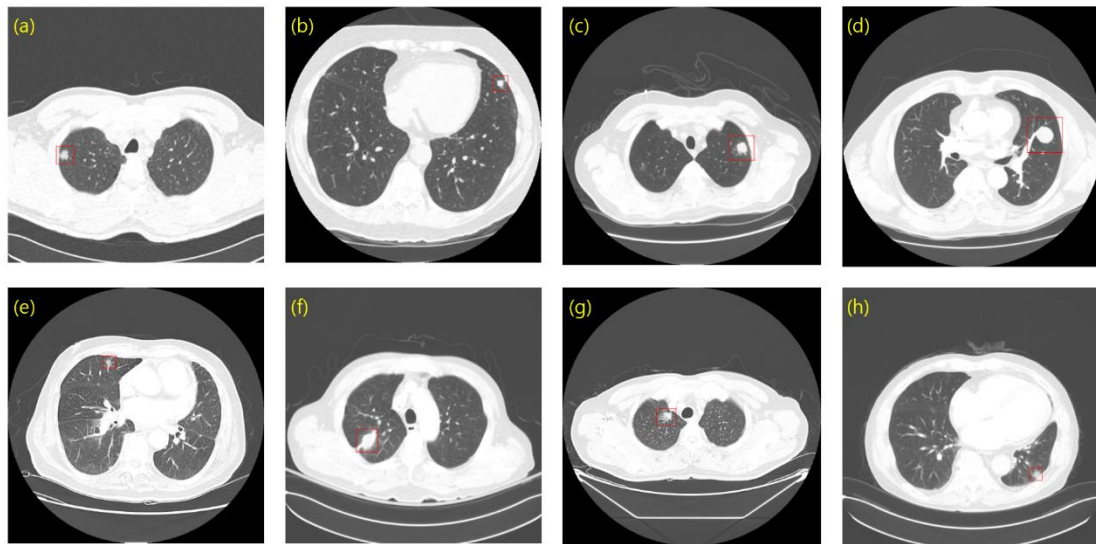
- GNAH dataset: 287 nodules (102 benign SPNs, and 185 malignant SPNs);
- KNUH dataset: 181 nodules (36 benign SPNs, and 145 malignant SPNs);
- HSHH dataset: 47 nodules (17 benign SPNs, and 30 malignant SPNs).

Table 1. Chest CT protocols at Gangneung Asan Hospital (GNAH), Kangwon National University Hospital (KNUH), and Hallym Sacred Heart Hospital (HSHH).

CT Protocols	GNAH Dataset	KNUH Dataset	HSHH Dataset
Name	SIEMENS/SOMATOM Definition Edge 2 (128 ch)	SIEMENS/SOMATOM Definition & Definition Flash	SIEMENS/SOMATOM Flash (128 ch)
kVp/mAs	120/35	120/110	120/35
kernel	B41f medium	B41f	B40f medium
slice/gap (mm)	5	3	3

2.1.3. Annotation

Using LUNA16 and private data, we created bounding boxes of lung nodules, which were used as the ground truth of the object detection task. To determine the coordinates of a bounding box in LUNA16, we used an annotation consisting of the centroid coordinates and diameters of the 3D nodules. The bounding boxes and annotations of private data were created with the help of the experts at each hospital. To this end, we first obtained the nodule centroid coordinates for each slice as identified by an expert. Based on the nodule centroid coordinates, we constructed a lung nodule bounding box using a computer vision annotation tool (CVAT). CVAT is an open-source platform that can build and share private servers. We then shared the lung nodule bounding boxes with experts to review and modify them. Using the bounding boxes in LUNA16 and our private data, we prepared annotations in pascal visual object classes challenge (VOC) format, which is a representative object detection data format [28]. The annotation of the VOC format consists of the filename, size, object name, and coordinates of the bounding box. Figure 1 shows a set of bounding boxes in LUNA16 and the private data. Figure 1c,d corresponding to the LUNA16 dataset show how the bounding box can at times represent the nodule position while having a size larger than that of the nodule. These cases negatively affect the detection of sophisticated boxes, an issue that we will further describe in the Results.

**Figure 1.** Ground truth sample. (a–d) LUNA16; (e–h) Private data.

2.1.4. Dataset Configuration

The LUNA16 dataset consisted of 10 subsets. We used eight subsets (712 patients) for the training dataset and two subsets (176 patients) for the validation dataset. The private dataset included a total of 515 CT scans consisting of 181 data from KNUH, 287 data from GNAH, and 47 data from HSHH. To learn from historical data and test relatively recent data, we sorted each hospital's data by chronological order based on the date of the CT scan.

The GNAH and KNUH data were accordingly divided into 60% of the data for training, 20% of the data for validation (a dataset used to tune the hyper parameters of a detector), and 20% of the data for testing (a dataset used only to assess the performance of a fully specified detector) (Figure 2). The HSHH dataset was used as an external validation set, which was used only for testing, with no involvement in training and validation.

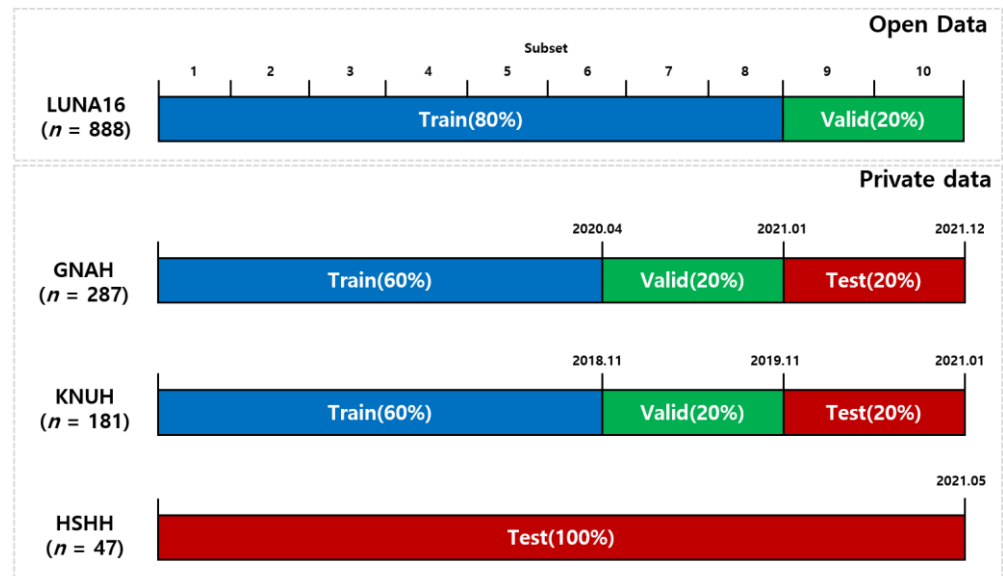


Figure 2. Illustration of dataset configuration.

2.2. Preprocessing

Unlike normal image data with values of 0–255, CT scan data consist of 10–16 bits of Hounsfield units (HU). To use CT data for deep learning, we thus need to set their values within the 0–1 range. Certain tissues and organs, such as the lung, brain, bone, and skin, generally have a constant range of HU values and adjusting the width center (W_C) and the window width (W_W) can emphasize the areas of interest. W_C is the HU value of the specific area to be emphasized and W_W is the HU range to be observed. In this study, we used window-setting values suitable for lung nodule detection ($W_C = -700$; $W_W = 1500$) [29–32]. The range of the CT scan data can be defined as:

$$X_{CT} \in \left[W_C - \frac{W_W}{2}, W_C + \frac{W_W}{2} \right] \quad (1)$$

We clipped the CT scan data into $(-1450, 50)$ using the minimum and maximum values obtained through Equation (1). Then, the range of the CT scan data was linearly normalized to $(0, 1)$, as follows:

$$\tilde{X}_{CT} = \frac{X_{CT} - \min(X_{CT})}{\max(X_{CT}) - \min(X_{CT})} \quad (2)$$

2.3. Data Augmentation

Data augmentation generates artificial data to increase the amount of data [33]. This technology alleviates the overfitting of deep-learning models. Data augmentation is essential in the medical field where large amounts of data are difficult to obtain. We applied the following augmentation techniques to the training model in the following order: (1) flip; (2) mosaic [34]; (3) random affine transformation; and (4) MixUp [35]. After doing image flipping with a probability of 0.5, we applied a mosaic, a strong data augmentation technique that combines up to four images into one. In addition to learning from four images with one input, because the number of small objects increases in the process of combining the images, the detection performance of small objects can be expected to improve. After

mosaic application, random affine transformations such as rotation, translation, and scaling were applied. Finally, MixUp, in which two images are mixed at a certain ratio, was used to produce a single image. MixUp helps generalize deep learning models. Figure 3 shows an example of applying data augmentation in LUNA16.

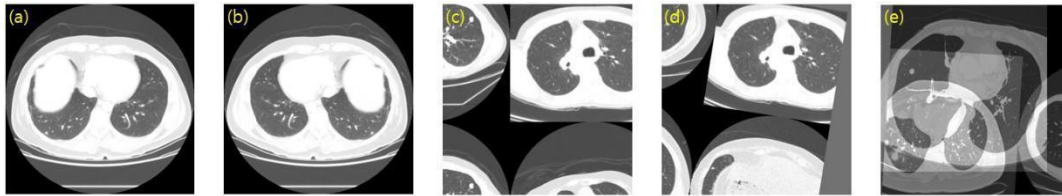


Figure 3. Example of data augmentation applied to lung nodule CT images. (a) Original; (b) Flip; (c) Mosaic; (d) Random Affine; (e) MixUp.

2.4. Object Detection and Deep Learning Model

In this study, an object detection algorithm, which is a deep learning technique, was used to detect lung nodules in lung CT scans. Object detection is the task of classifying and localizing all the objects present in an image. Localization is a regression operation that finds a bounding box, whereas classification is an operation that classifies objects in a bounding box.

We used YOLOX to detect lung nodules [24]. The architecture of YOLOX is depicted in Figure 4. YOLOX is an object detection model with excellent accuracy and speed that simultaneously applies various detection techniques. This model is a one-stage detector and consists of a backbone-neck decoupled head. YOLOX extracts a feature map from the input image using a darknet-53 backbone. The neck connects the backbone and the head and consists of spatial pyramid pooling (SPP) and feature pyramid network (FPN) [36,37]. SPP improves performance by playing a role in maintaining spatial structure information. FPN receives a feature map extracted from the backbone and obtains a multiscale feature map. FPN adds semantic information preserved by a high-level layer to a low-level layer to make it robust against scale variations. The feature map obtained from FPN enters the decoupled head and finds the class and box of the object to derive the detection results. The decoupled head improves the performance by separating the classification and regression heads to solve the problem caused by the conflict between classification and regression work [38]. In addition, YOLOX uses various technologies, such as anchor-free, SimOTA, and multi-positive, to improve performance.

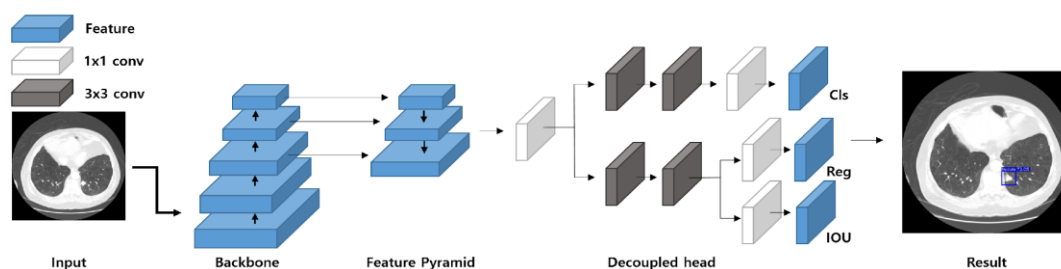


Figure 4. Detailed architecture of YOLOX. Cls, Class; Reg, Regression; IOU, intersection over union; conv, convolution.

2.5. Evaluation Metrics

In general, we computed true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) by comparing the predicted and the ground truth to measure the predictive performance of the trained model. In this study, TP refers to instances where the deep learning model correctly predicted lung nodules. FP refers to cases where the deep learning model incorrectly predicted non-lung nodules as lung nodules. FN refers to instances where the deep learning model incorrectly predicted lung nodules as non-lung

nodules. Finally, TN refers to cases where the deep learning model correctly predicted that non-lung nodules were not lung nodules. In the object detection task, the intersection over union (IOU) of the predicted bounding box and the actual bounding box were calculated. If the IOU was above the threshold value, the bounding box was judged to be true, and if IOU was under the threshold value, the bounding box was judged to be false. Sensitivity and precision can be calculated using this information. The sensitivity is the ratio of the lung nodules that were correctly predicted to the actual lung nodules. The precision is the ratio of the correctly predicted lung nodules to all predicted lung nodules. Finally, based on this information, lung nodule detection performance was evaluated using the average precision (AP) and competition performance metric (CPM) [39] utilizing free-response receiver operating characteristic (FROC). AP and CPM were used as quantitative evaluation metrics. AP is an evaluation metric that considers both precision and sensitivity and is widely used to evaluate the performance of object detection algorithms. Lung nodules are very small objects, and the IOU values are very low, even if the size or position of the box is slightly different from the ground truth. For this reason, even an actual true box can be judged false if the IOU threshold is high. Therefore, the performance measurements were carried out using AP10 (AP at IOU = 0.1) and AP50 (AP at IOU = 0.5). The AP10 metric has the advantage of having a high sensitivity because it is recognized as true, even if the overlap between the prediction and ground truth is low. The AP50 metric may have a lower sensitivity performance because the IOU of the prediction and ground truth must be 0.5 or more to be true, but more sophisticated lung nodule bounding box performance can be measured. FROC analysis is utilized for nodule detection evaluation in the LUNA16 challenge. In the FROC curve, sensitivity is plotted as a function of the average number of false positives per scan (FPs/scan). The CPM score is the average sensitivity at seven predefined false-positive rates (1/8, 1/4, 1/2, 1, 2, 4, and 8FPs per scan).

2.6. Dataset Configuration and Experimental Strategy

The dataset configurations used in the experiments are presented in Table 2. We prepared five configurations using both open and private datasets for training and validation, whereas for testing we used only one configuration using private datasets, GNAH (20%), KNUH (20%), and HSHH (100%). In configuration (A), we used only the open dataset LUNA16 and eight subsets for training and two subsets for validation. In configuration (B), we used LUNA16 for training and two private datasets (GNAH and KNUH) for validation. In the other configurations (C, D, and E), we used only private datasets for both training and validation.

Table 2. Dataset configurations used in the experiments.

Dataset	Training	Validation	Test
(A)	LUNA16 (subsets 1–8)	LUNA16 (subsets 9, 10)	
(B)	LUNA16 pre-train (100%)	GNAH (20%) KNUH (20%)	GNAH (20%) KNUH (20%)
(C)	GNAH (60%)	GNAH (20%)	HSHH (100%)
(D)	KNUH (60%)	KNUH (20%)	
(E)	GNAH (60%), KNUH (60%)	GNAH (20%) KNUH (20%)	

To analyze the effects of open and private datasets, we then conducted three experiments depending on the settings of the LUNA16 and private datasets (Figure 5). Experiment 1 was conducted to compare the usefulness of the open dataset and private datasets when used for training. Experiment 2 was conducted to evaluate the effect of pre-training using LUNA16 before training with private data. In Experiment 3, performance improvement was measured by gradually increasing private data to analyze the amount

of private data required. All experiments were conducted on an NVIDIA GeForce GTX 1080Ti GPU with 11GB of graphics memory. We trained nodule detection models using the Pytorch framework. For the detection model, we used YOLOX-S. In the training, we used a stochastic gradient descent (SGD) optimizer with a momentum of 0.9, a weight decay of 0.0005, and a learning rate of $lr \times \text{batch size}/64$. The initial lr was 0.01 and the batch size was set to 16. The learning rate scheduler adopts a cosine LR schedule. The input size was 640×640 pixels. The model was then trained for 200 epochs. All experiments were conducted using a fixed seed = 1.

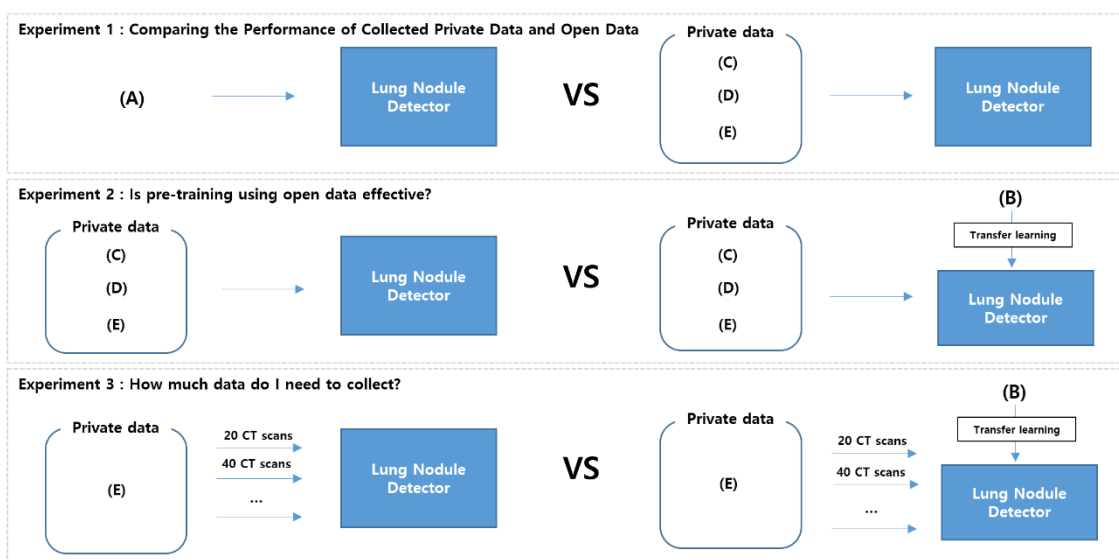


Figure 5. Experimental settings.

3. Results

3.1. Experiment 1: Comparing the Performance of Collected Private Data and Open Data

This experiment compared the performance of a model trained on the LUNA16 dataset with that of a model trained on a private dataset. Table 3 summarizes the performance obtained in Experiment 1, demonstrating that learning from private data led to a competitive performance of the model compared to that obtained when learning from LUNA16. In addition, when the same hospital data were used for both training and testing, the performance was better than that obtained when using LUNA16 for training. Furthermore, the AP50 performance of the model trained on private data was significantly improved compared with that of the model trained on LUNA16. Moreover, the performance using the combined dataset of GNAH and KNUH was higher than that using the individual hospital datasets. This result suggests that collecting a large amount of private data is important for performance improvement.

Figure 6 shows an image superimposing the prediction results of the models trained for the ground truth and LUNA16, as well as the prediction results of the models trained for private data. As shown in Figure 6b,c, the prediction results of the LUNA16 training model tended to draw a box larger than the lung nodule, suggesting a difference in the performance of AP50. Figure 6d shows a case in which the private data training model predicted the lung nodules unforeseen in the LUNA16 training model.

Table 3. Performance Comparison of Collected Private Data and Open Data.

Training Dataset	Test Dataset	AP10	AP50	CPM
LUNA16	GNAH	0.8590 (+0.0)	0.5482 (+0.0)	0.8886 (+0.0)
	KNUH	0.9151 (+0.0)	0.4499 (+0.0)	0.9340 (+0.0)
	HSHH	0.7262 (+0.0)	0.1111 (+0.0)	0.7628 (+0.0)
GNAH	GNAH	0.8674 (+0.0084)	0.8160 (+0.2678)	0.8967 (+0.0081)
	KNUH	0.8548 (−0.0603)	0.8228 (+0.3729)	0.8844 (−0.0496)
	HSHH	0.7878 (+0.0616)	0.6270 (+0.5159)	0.8114 (+0.0486)
KNUH	GNAH	0.8455 (−0.0135)	0.7996 (+0.2514)	0.8717 (−0.0169)
	KNUH	0.9167 (+0.0016)	0.8817 (+0.4318)	0.9395 (+0.0055)
	HSHH	0.8307 (+0.1045)	0.7729 (+0.6618)	0.8561 (+0.0933)
GNAH \cup KNUH	GNAH	0.8960 (+0.0370)	0.8450 (+0.2968)	0.9290 (+0.0404)
	KNUH	0.9525 (+0.0374)	0.9341 (+0.4842)	0.9660 (+0.0320)
	HSHH	0.8391 (+0.1129)	0.7935 (+0.6824)	0.8645 (+0.1017)

AP, average precision; CPM, competition performance metric.

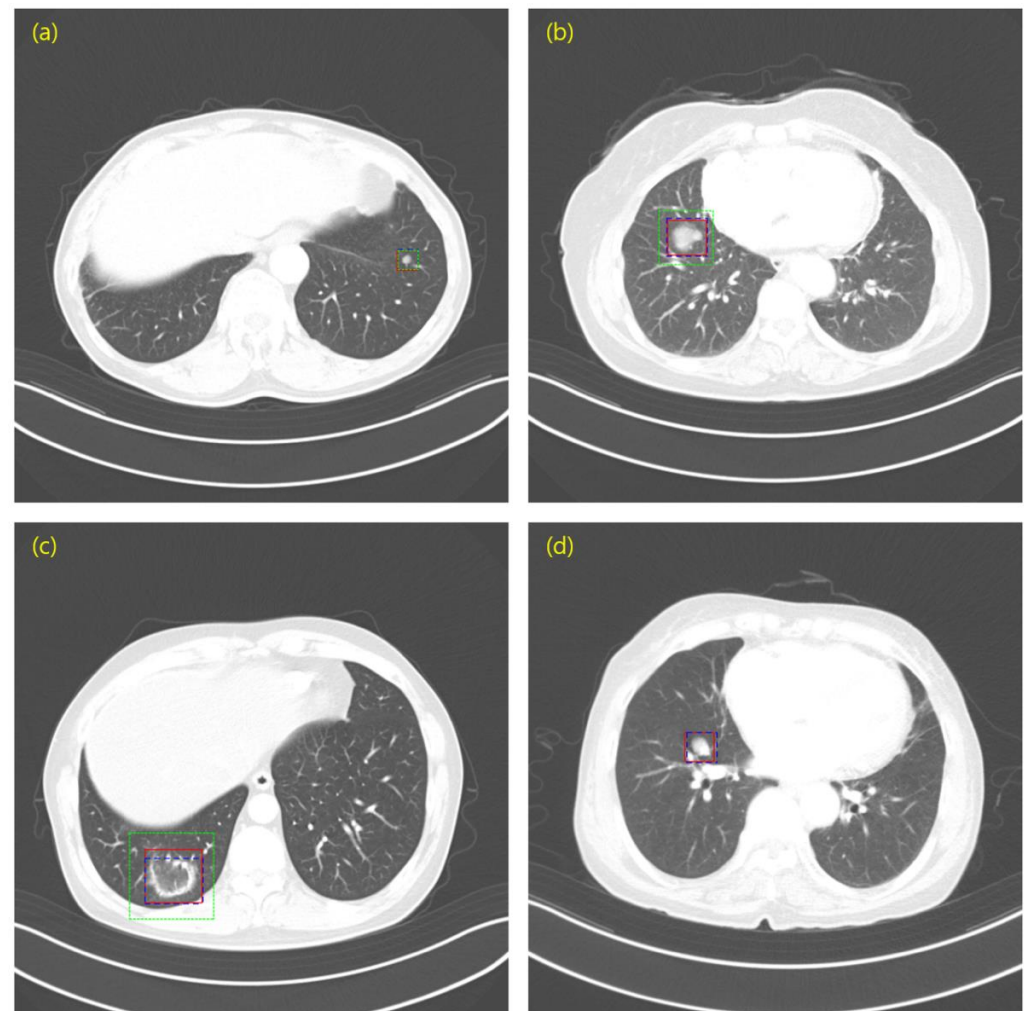


Figure 6. Image sample overlapping the prediction results of the LUNA16 and private data trained models and the ground truth. (a) Case where the predictions of the model trained on LUNA16 and the model trained on private data is very similar to the ground truth; (b,c) Case where the predicted box of the model trained on LUNA16 is significantly larger than the size of the ground truth; (d) Case in which the LUNA16 trained model did not predict lung nodules. Red (Solid line) = Ground Truth; Blue (Dashed line) = Private data; Green (Dotted line) = LUNA16.

3.2. Experiment 2: Is Pre-Training Using Open Data Effective?

We found that although training on LUNA16 only in Experiment 1 was not sophisticated, the model learned the characteristics of lung nodules well enough to predict their locations. Hence, we can expect pre-training on LUNA16 to have a positive effect on training with private data. To confirm this, we experimented with a transfer learning method using pre-training on LUNA16. We compared the cases with and without transfer learning that adopted weights pre-trained on LUNA16. Figure 7 shows the CPM results from Experiment 2. Most of the performances were higher when applying transfer learning. However, when the combined data from GNAH and KNUH were used for training, the performance was higher when transfer learning was not applied. The resulting graphs for Experiment 2 in its entirety are shown in Figure A1 in Appendix B.

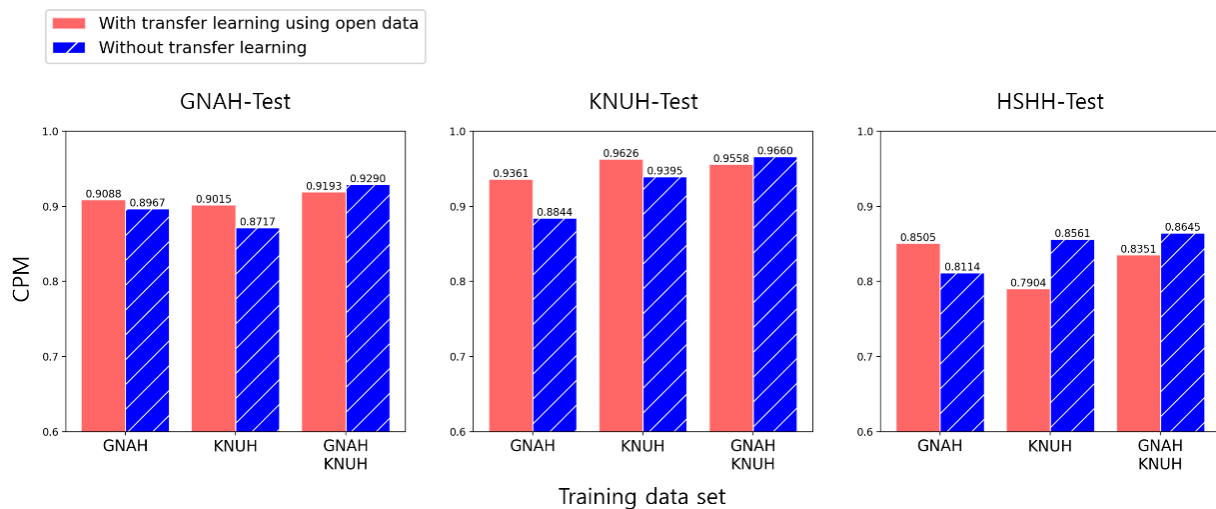


Figure 7. Comparison of the performance of the models with transfer learning from open data and without transfer learning.

3.3. Experiment 3: How Much Data Do I Need to Collect?

Experiment 3 evaluated performance while gradually increasing the amount of private data. The data were sorted based on the date of the CT scan, and old data were extracted first. The combined data of GNAH and KNUH were gradually increased with 10 patients from GNAH and 10 patients from KNUH and the performance was evaluated. We also trained and evaluated the EfficientDet [23] model as well as the YOLOX model for the generalization of the conclusion (Figure 8). When using a small amount of data without transfer learning, the performance was very poor. However, when using approximately 100 patient data points, the performance became similar to the case using transfer learning. When transferring knowledge from open data, high performance could be achieved, even with a small amount of data. Additional graphs for Experiment 3 are shown in Figures A2 and A3 of Appendix B.

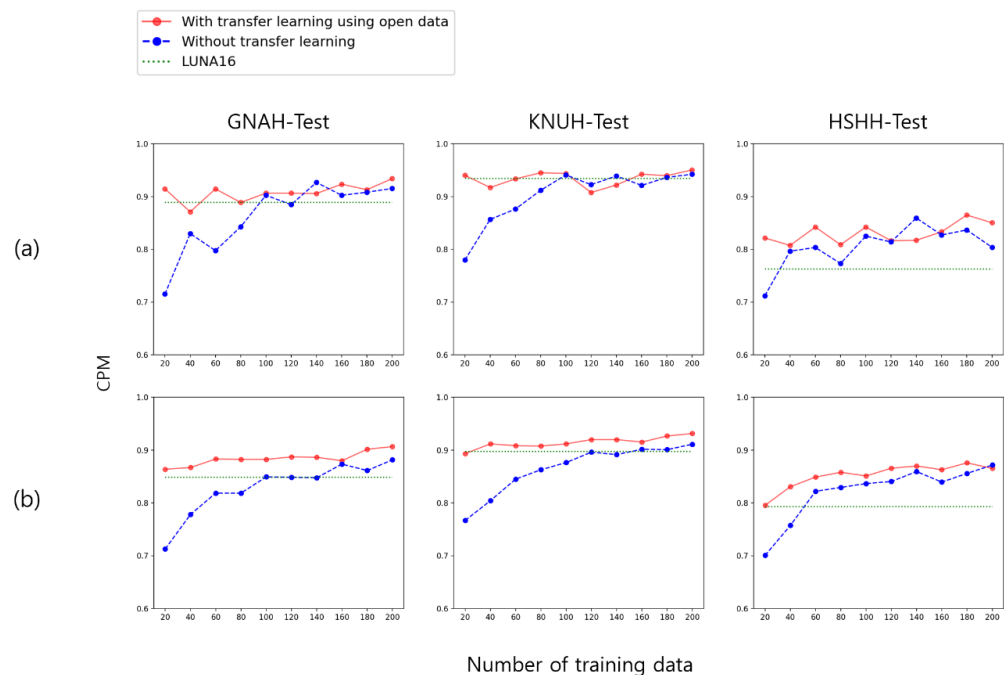


Figure 8. CPM results for models trained with incremental growth of GNAH and KNUH combined data. (a) YOLOX; (b) EfficientDet.

4. Discussion

Public datasets are essential for the better application of deep learning to medical problems in clinical settings [40]. Identifying issues and biases in commonly used benchmark datasets requires the use of private datasets [41,42]. In this study, several experiments were conducted using publicly available LUNA16 data and private data collected from three hospitals.

In Experiment 1, we were able to compare the results when training the model on LUNA16 and on private data. Compared to training the model on LUNA16 with large numbers of CT scans, private data collected in each hospital led to a similar performance in lung nodule detection. Using training and testing data from the same hospital led to higher performance than when training on LUNA16. These results indicate that using training data from the same hospital can be expected to lead to good performance in the detection of lung nodules in private hospitals. In Experiment 1, it was confirmed that even if the model was trained on LUNA16 alone, it was possible to identify and detect lung nodule features in private test data. However, the annotation of LUNA16 consists only of the central coordinates and diameters of the 3D nodules, not the bounding box coordinates for each slice, making it impossible to create a sophisticated box for each slice. Private data included slice-by-slice boxes reviewed by radiologists as annotations, allowing sophisticated boxes to be detected. The AP50 results showed that learning on private data can yield more sophisticated results than learning on LUNA16 only. Finally, training could be done by combining the training datasets of KNUH and GNAH. These results show that a good performance can be expected for lung nodule detection if a dataset is built by collecting a large amount of data.

In Experiment 2, the performance of transfer learning based on the weight pre-trained on LUNA16 and the performance without transfer learning were compared. When training on individual hospital data, most of the cases of transfer learning were better than those that did not. However, the performance was found to be better without transfer learning when the model was trained using KNUH in the HSHH-Test. This phenomenon appears to be due to differences in the CT slice thickness between the datasets [43]. The KNUH and HSHH datasets had the same slice thickness (3 mm). For this reason, even though only KNUH data were used for training, the performance may be higher than that of transfer

learning when HSHH data are tested. In this case, it seems that applying transfer learning in the HSHH-test is a factor reducing the performance. These results show that if the CT protocol is similar and the slice gap is the same, the performance is good when testing the CT data under the same conditions. Moreover, AP10 and CPM performance were higher when applying transfer learning to graphs trained with GNAH in the GNAH-TEST, but the AP50 performance was lower. In this case, the location of the lung nodule was better identified when applying transfer learning, but the sophistication of the nodule box was reduced. In the future, it will be necessary to collect more private data and analyze these cases in depth. When using the training dataset combining KNUH and GNAH, the performance improved without transfer learning. These results demonstrate that high performance can be achieved without applying transfer learning if sufficient data can be collected.

Experiment 3 tested the amount of data that needed to be collected to detect the lung nodules. Performance was measured as the amount of private data gradually increased, starting with the data from 20 patients. When no transfer learning was performed, the performance was very poor when training was performed with a small amount of data. However, when close to 100 patient data points were collected, the performance was similar to or increased more strongly than when the model was trained on LUNA16. When LUNA16 transfer learning was performed under the same conditions, the performance was high, even with a small amount of private data training. Transfer learning from LUNA16 is very effective in situations where it is not possible to collect a large amount of data.

To summarize the results, to evaluate AI algorithms for detecting lung nodules of chest CT scans in private hospitals, it is better to collect and build a private dataset that can produce more sophisticated results than using only open data. However, if it is difficult to collect a large amount of private data, it is effective to utilize pre-trained weights with open data. If a large amount of private data can be collected, good performance can be expected without transfer learning. These results provide guidelines for building the private datasets necessary for the model to learn lung nodules.

To the best of our knowledge, this paper is the first to determine the amount of private data required to perform object detection for an SPN on an enhanced chest CT scan. Because it is a multicenter study of three hospitals, it can better represent real-world datasets and their heterogeneity than public datasets. The size of private datasets was sufficient to perform an object detection model without using public data. Most recently, Xu et al. [44] reported a YOLOv3 network for detecting lung nodules on chest CT using 314 private data points. However, our private dataset of 515 patients was larger than Xu's private dataset.

Although this study obtained meaningful results through lung nodule detection experiments using data from various hospitals, it had several limitations. First, we did not study a technology to improve lung nodule detection performance. For further study, we will focus on enhancing detection performance itself by adapting state-of-the-art techniques such as generative adversarial networks-based augmentation. Second, the classification of benign and malignant nodules was not considered because only the task of detecting the location of the lung nodules was performed.

This study shows that when detecting lung nodules from private data, training the model on private data collected from the same hospital can yield more sophisticated results than training the model on massive open data. Moreover, the performance based on the amount of private data was confirmed, and it was found that transfer learning had a significant influence when the amount of data was small. However, the effect of transfer learning was reduced if a large amount of data was collected and used for training. Therefore, if a large amount of data is collected, good performance can be expected without having to transfer the weights pre-trained on the open data.

The final goal of our study is to present guidelines on how many private data should be collected when conducting a study to detect lung nodules with chest CT images. Most deep learning researchers who want to analyze medical data use public data because it is difficult to obtain private data. However, as the reviewer has already mentioned, public

data does not reflect the characteristics of data based on the real world, such as when a CT machine is changed, a new protocol is released, or the definition of a pulmonary nodule is changed. This is because even if a deep learning model is developed with public data, it cannot be applied to private data for the same reason as above, and training suitable for private data may need to be renewed again. Additionally, the issues of whether artificial intelligence can replace radiologists and who should be held responsible for making a wrong decision are still controversial, and I think that AI researchers' consensus is needed to solve these problems. At least for now, it should be considered as one of the diagnostic aids that reduce the fatigue of medical staff.

5. Conclusions

The experimental results of this study will help researchers study lung nodule detection for the collection of private data and obtaining results. In future research, we plan to build larger datasets with continuous data supply and demand. Unlike ordinary images, medical images must be labeled by specialists. The more data there are, the more labeling is needed, but the number of specialists is limited. Therefore, it is necessary to increase the time spent by specialists and the efficiency of the labor force. We plan to study active learning algorithms to proceed efficiently with data labeling. Active learning allows the selection of only the data with high uncertainty and requires labeling from a specialist. Consequently, data can be smoothly built by reducing the amount of data required for labeling. In addition, we plan to study AI technologies that are suitable for the characteristics of chest CT scans and to conduct research that can contribute to engineering. First, it will be valuable to experiment by applying various window settings for chest CT scanning. Depending on the window setting, the range of HUs can be adjusted to obtain various chest CT images. It is also possible to consider combining images as inputs into deep learning models using multiple sets of lung CT images or combining the results of each image. Second, we are starting to use object detection models using slices, which include box annotation; therefore, there is a lot of data that remains unused. We also plan to conduct a study using CT slices that do not include lung nodules to devise anomaly detection technology. Finally, because there are many slices containing lung tissues among the unused data, research on lung detection or lung segmentation can also be a good approach.

Author Contributions: Conceptualization, J.Y.H., W.J.K., K.M.M. and S.H.B.; methodology, J.W.S., K.M.M., S.H.B. and H.-S.C.; software, J.W.S., Y.K. and H.-S.C.; validation, J.W.S., J.Y.H., W.J.K., K.M.M. and S.H.B.; formal analysis, J.W.S., J.Y.H., W.J.K. and H.-S.C.; investigation, J.W.S., J.Y.H., K.M.M., S.H.B., D.-Y.S. and H.-S.C.; resources, J.W.S. and Y.K.; data curation, J.Y.H., K.M.M. and S.H.B.; writing—original draft preparation, J.W.S. and K.M.M.; writing—review and editing, J.W.S., K.M.M., S.H.B. and H.-S.C.; visualization, J.W.S.; supervision, K.M.M., S.H.B. and H.-S.C.; project administration, Y.K., K.M.M., S.H.B. and H.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant of the Medical data-driven hospital support project through the Korea Health Information Service (KHIS), funded by the Ministry of Health & Welfare, Republic of Korea, by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1F1A1076454), and by 2021 Research Grant from Kangwon National University.

Institutional Review Board Statement: The study was approved by the Ethics Committee of Chuncheon Sacred Heart hospital (approval no. 2021-05-016, date of approval: 11 June 2021).

Informed Consent Statement: Patient consent was waived due to the fact that this was a retrospective study with de-identified data.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. GNAH clinical data.

Clinical Data	Total (n = 287)	Malignancy (n = 185)	Benign (n = 102)
Age, yr (min-max)	67.5 (21–87)	68.3 (32–87)	66.1 (21–87)
Sex, male (%)	181 (63.1)	110 (59.5)	71 (69.6)
Pathological diagnosis, n (%)			
NSCLC	145 (50.5)	145 (78.4)	0
Adenocarcinoma	107 (37.3)	107 (57.8)	
Squamous cell carcinoma	33 (11.5)	33 (17.8)	
Adenosquamous carcinoma	2 (0.7)	2 (1.1)	
NSCLC	3 (1.0)	3 (1.6)	
SCLC	9 (3.1)	9 (4.9)	0
Other type of lung malignancy	7 (2.4)	7 (3.8)	0
Malignancy other than in the lung	24 (8.4)	24 (13.0)	0
Negative for malignancy	58 (20.2)	0	58 (56.9)
NA	44 (15.3)	0	44 (43.1)
Methods for pathological diagnosis			
PCNA	105 (36.6)	57 (30.8)	48 (47.1)
Surgical operation	77 (26.8)	71 (38.4)	6 (5.9)
Bronchoscopy	4 (1.4)	4 (2.2)	0
EBUS-TBNA	2 (0.7)	2 (1.1)	0
Contrast-enhanced Chest CT (≥ 2 years)			
No growth	35 (12.2)	0	35 (34.3)
Disappearance	9 (3.1)	0	9 (8.8)

NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; NA, not applicable; PCNA, percutaneous needle aspiration; EBUS-TBNA, endobronchial ultrasound-guided transbronchial needle aspiration.

Table A2. KNUH clinical data.

Clinical Data	Total (n = 181)	Malignancy (n = 145)	Benign (n = 36)
Age, yr (min-max)	67.2 (31–88)	69.0 (46–88)	60.3 (31–85)
Sex, male (%)	116 (64.1)	90 (62.1)	26 (72.2)
Pathologic diagnosis, n (%)			
NSCLC	118 (65.2)	118 (81.4)	0
Adenocarcinoma	87 (48.1)	87 (73.3)	
Squamous cell carcinoma	25 (13.8)	25 (21.1)	
Large cell endocrine carcinoma	3 (1.7)	3 (2.5)	
NSCLC	3 (1.7)	3 (2.5)	
SCLC	5 (2.8)	5 (3.4)	0
Other type of lung malignancy	6 (3.3)	6 (4.1)	0
Malignancy other than lung	16 (8.8)	16 (11.0)	0
Negative for malignancy	33 (18.2)	0	33 (91.7)
NA	3 (1.7)	0	3 (8.3)
Methods for pathologic diagnosis			
PCNA	92 (50.8)	70 (48.3)	22 (61.1)
Surgical operation	86 (47.5)	75 (51.7)	11 (30.6)
Contrast-enhanced Chest CT (≥ 2 years)			
No growth	3 (1.7)		3 (8.3)
Disappearance	0		0

Table A3. HSHH clinical data.

Clinical Data	Total (n = 47)	Malignancy (n = 30)	Benign (n = 17)
Age, yr (min-max)	68.0 (35–88)	70.4 (50–88)	63.6 (35–87)
Sex, male (%)	28 (63.1)	18 (60.0)	10 (58.8)
Pathologic diagnosis, n (%)			
NSCLC	24 (51.1)	24 (80.0)	0
Adenocarcinoma	17 (36.2)	17 (56.7)	
Squamous cell carcinoma	7 (14.9)	7 (23.3)	
SCLC	2 (4.3)	2 (6.7)	0
Other type of lung malignancy	3 (6.4)	3 (10.0)	0
Malignancy	1 (2.1)	1 (3.3)	0
Negative for malignancy	17 (36.2)	0	17 (100.0)
Methods for pathologic diagnosis			
PCNA	19 (40.4)	13 (43.3)	6 (35.3)
Bronchoscopy	16 (34.0)	9 (30.0)	7 (41.2)
Transbronchial lung biopsy	5 (10.6)	1 (3.3)	4 (23.5)
Surgical operation	3 (6.4)	3 (10.0)	0
EBUS-TBNA	1 (2.1)	1 (3.3)	0
Unknown	3 (6.4)	2 (1.1)	0

Appendix B

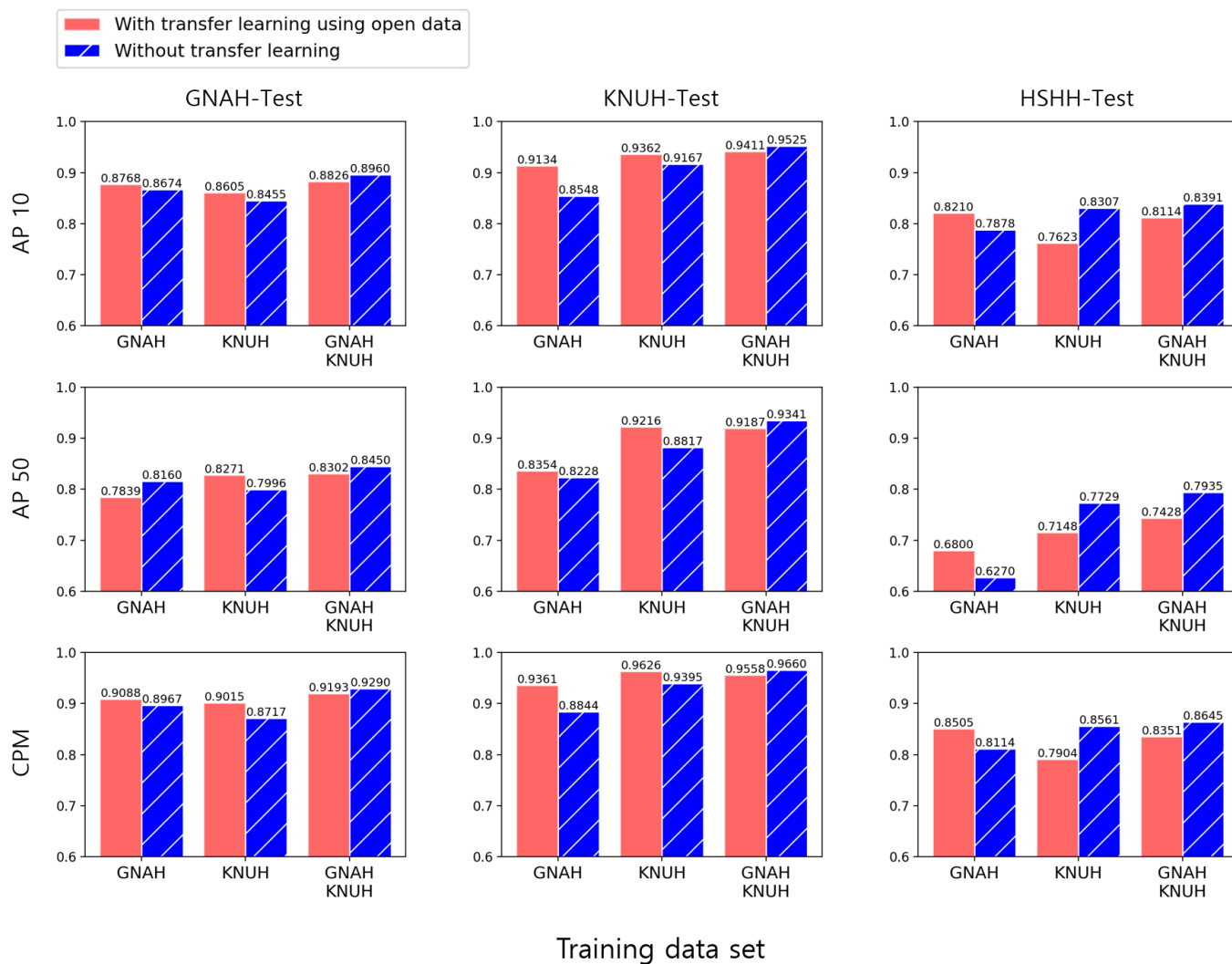


Figure A1. Experiment 2.

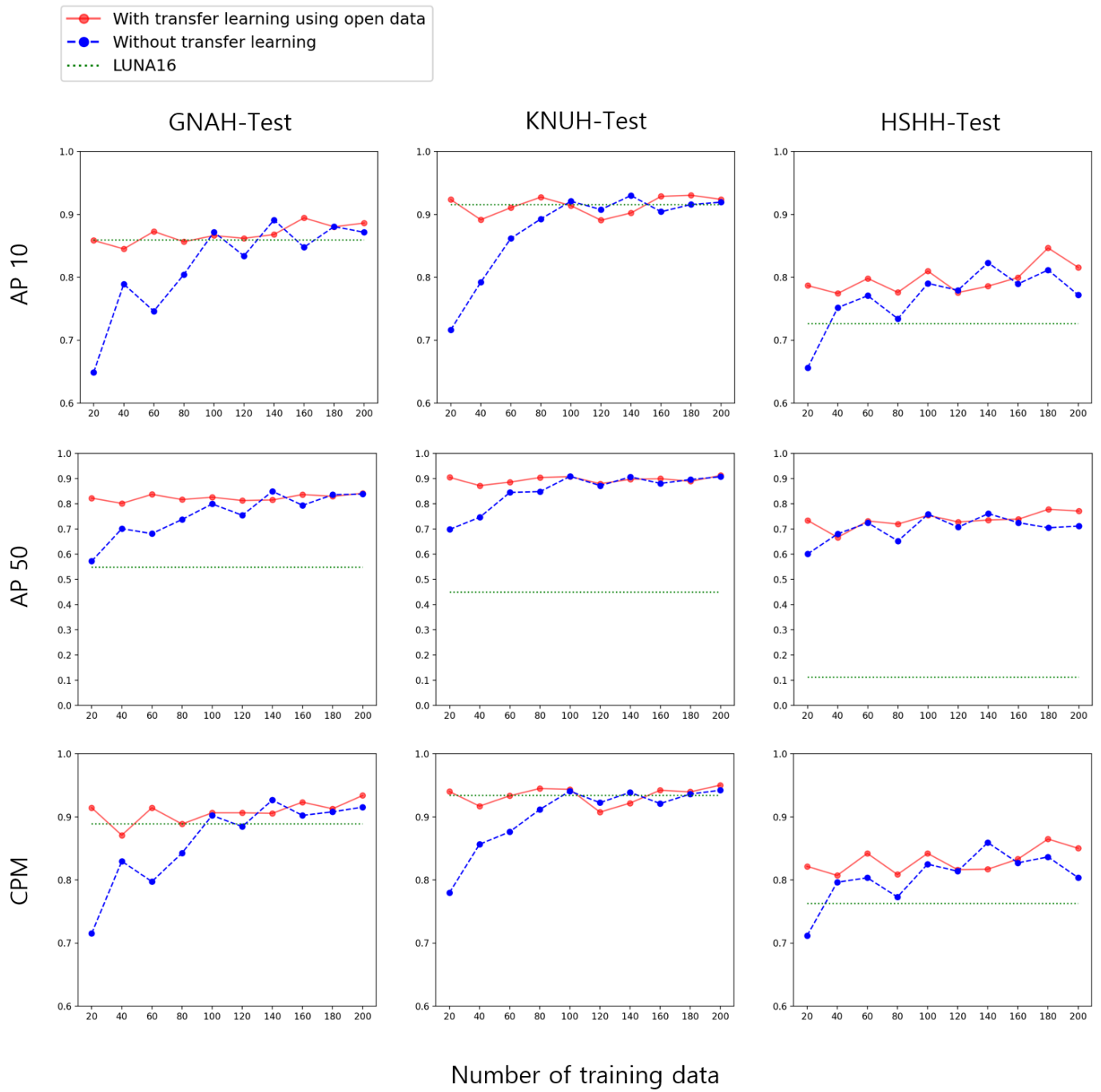


Figure A2. Experiment 3—YOLOX.

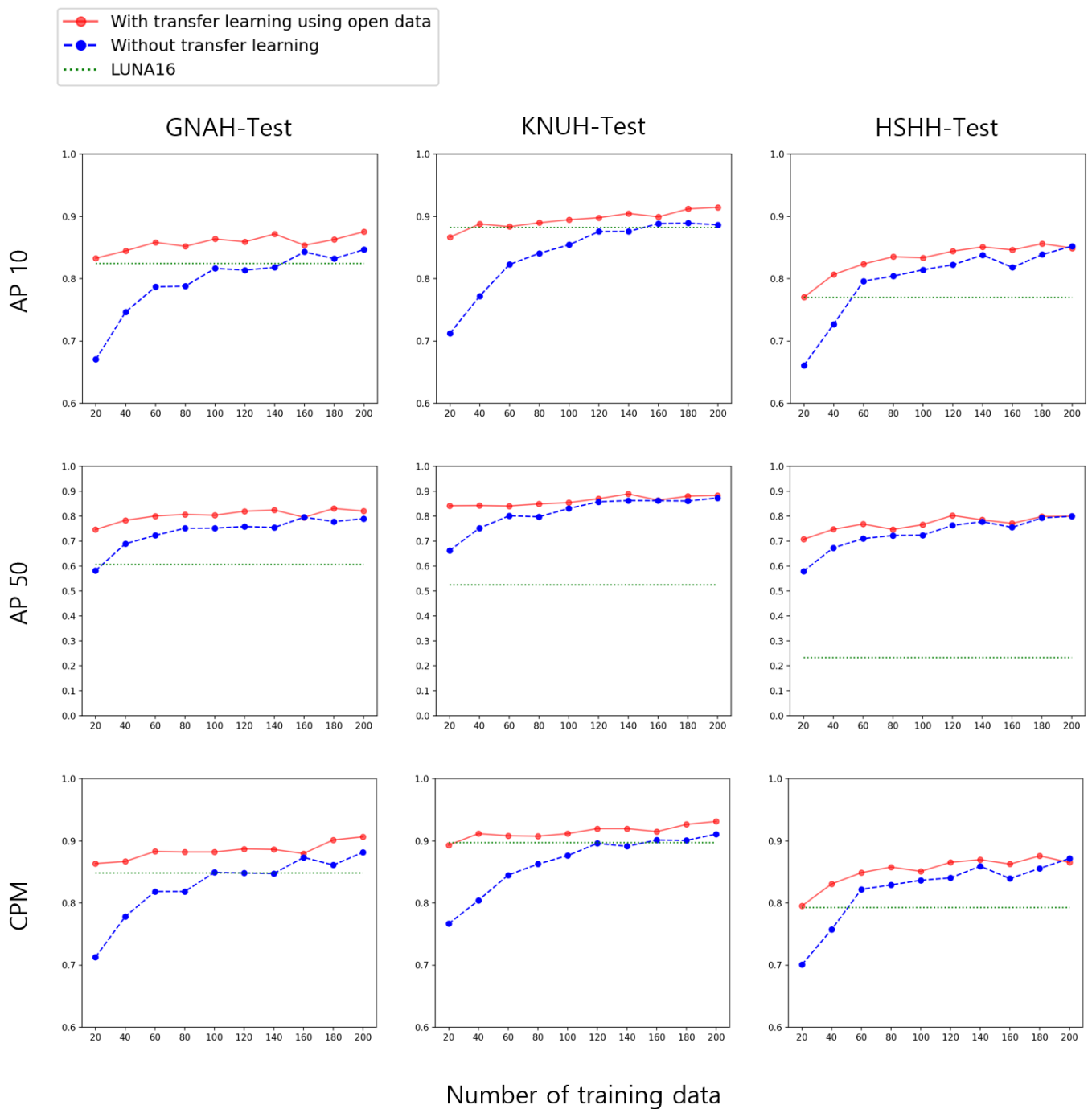


Figure A3. Experiment 3—EfficientDet.

References

1. World Health Organization. Available online: <http://www.who.int/en/> (accessed on 5 September 2017).
2. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
3. Feng, R.M.; Zong, Y.N.; Cao, S.M.; Xu, R.H. Current cancer situation in China: Good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun.* **2019**, *39*, 1–12. [[CrossRef](#)]
4. Cancer Research UK. Available online: <https://www.cancerresearchuk.org/> (accessed on 12 October 2017).
5. The National Lung Screening Trial Research Team; Aberle, D.R.; Adams, A.M.; Berg, C.D.; Black, W.C.; Clapp, J.D.; Fagerstrom, R.M.; Gareen, I.F.; Gatsonis, C.; Marcus, P.M.; et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **2011**, *365*, 395–409. [[CrossRef](#)] [[PubMed](#)]

6. Demir, Ö.; Yılmaz Çamurcu, A. Computer-aided detection of lung nodules using outer surface features. *Bio-Med. Mater. Eng.* **2015**, *26*, S1213–S1222. [[CrossRef](#)] [[PubMed](#)]
7. Bogoni, L.; Ko, J.P.; Alpert, J.; Anand, V.; Fantauzzi, J.; Florin, C.H.; Koo, C.W.; Mason, D.; Rom, W.; Shiau, M.; et al. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. *J. Digit. Imaging* **2012**, *25*, 771–781. [[CrossRef](#)] [[PubMed](#)]
8. Al Mohammad, B.; Brennan, P.C.; Mello-Thoms, C. A review of lung cancer screening and the role of computer-aided detection. *Clin. Radiol.* **2017**, *72*, 433–442. [[CrossRef](#)] [[PubMed](#)]
9. Armato, S.G., III; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* **2011**, *38*, 915–931. [[CrossRef](#)]
10. Setio, A.A.A.; Traverso, A.; De Bel, T.; Berens, M.S.; Van Den Bogaard, C.; Cerello, P.; Chen, H.; Dou, Q.; Fantacci, M.E.; Geurts, B. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med. Image Anal.* **2017**, *42*, 1–13. [[CrossRef](#)]
11. Zarogoulidis, K.; Zarogoulidis, P.; Darwiche, K.; Boutsikou, E.; Machairiotis, N.; Tsakiridis, K.; Katsikogiannis, N.; Kougioumtzi, I.; Karapantzos, I.; Huang, H. Treatment of non-small cell lung cancer (NSCLC). *J. Thorac. Dis.* **2013**, *5*, S389.
12. ELCAP Public Lung Image Database. Available online: <http://www.via/cornell.edu/lungdb.html> (accessed on 21 January 2022).
13. Van Ginneken, B.; Armato, S.G., III; de Hoop, B.; van Amelsvoort-van de Vorst, S.; Duindam, T.; Niemeijer, M.; Murphy, K.; Schilham, A.; Retico, A.; Fantacci, M.E. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study. *Med. Image Anal.* **2010**, *14*, 707–722. [[CrossRef](#)]
14. Li, R.; Xiao, C.; Huang, Y.; Hassan, H.; Huang, B. Deep Learning Applications in Computed Tomography Images for Pulmonary Nodule Detection and Diagnosis: A Review. *Diagnostics* **2022**, *12*, 298. [[CrossRef](#)] [[PubMed](#)]
15. Mu, R.; Zeng, X. A review of deep learning research. *KSII Trans. Internet Inf. Syst. (TIIS)* **2019**, *13*, 1738–1764.
16. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*. [[CrossRef](#)] [[PubMed](#)]
17. Howard, J.; Ruder, S. Universal language model fine-tuning for text classification. *arXiv* **2018**, arXiv:1801.06146.
18. Xiong, W.; Wu, L.; Allewa, F.; Droppo, J.; Huang, X.; Stolcke, A. The Microsoft 2017 conversational speech recognition system. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5934–5938.
19. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **2019**, *29*, 102–127. [[CrossRef](#)]
20. Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. *Classif. BioApps* **2018**, 323–350. [[CrossRef](#)]
21. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
22. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
23. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
24. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
25. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40. [[CrossRef](#)]
26. Fleischner Society. Available online: <https://fleischner.memberclicks.net/> (accessed on 16 October 2017).
27. MacMahon, H.; Naidich, D.P.; Goo, J.M.; Lee, K.S.; Leung, A.N.C.; Mayo, J.R.; Mehta, A.C.; Ohno, Y.; Powell, C.A.; Prokop, M.; et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* **2017**, *284*, 228–243. [[CrossRef](#)]
28. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
29. Cho, H.-H.; Lee, H.Y.; Kim, E.; Lee, G.; Kim, J.; Kwon, J.; Park, H. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. *Commun. Biol.* **2021**, *4*, 1286. [[CrossRef](#)] [[PubMed](#)]
30. Song, Y.; Zheng, S.; Li, L.; Zhang, X.; Zhang, X.; Huang, Z.; Chen, J.; Wang, R.; Zhao, H.; Chong, Y. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 2775–2780. [[CrossRef](#)] [[PubMed](#)]
31. Ahn, Y.; Lee, S.M.; Noh, H.N.; Kim, W.; Choe, J.; Do, K.-H.; Seo, J.B. Use of a commercially available deep learning algorithm to measure the solid portions of lung cancer manifesting as subsolid lesions at CT: Comparisons with radiologists and invasive component size at pathologic examination. *Radiology* **2021**, *299*, 202–210. [[CrossRef](#)] [[PubMed](#)]
32. Kim, H.; Goo, J.M.; Lee, K.H.; Kim, Y.T.; Park, C.M. Preoperative CT-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. *Radiology* **2020**, *296*, 216–224. [[CrossRef](#)]
33. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]

34. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
35. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
37. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
38. Wu, Y.; Chen, Y.; Yuan, L.; Liu, Z.; Wang, L.; Li, H.; Fu, Y. Rethinking classification and localization for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10186–10195.
39. Niemeijer, M.; Loog, M.; Abramoff, M.D.; Viergever, M.A.; Prokop, M.; van Ginneken, B. On combining computer-aided detection systems. *IEEE Trans. Med. Imaging* **2010**, *30*, 215–223. [[CrossRef](#)]
40. Matuszewski, D.J.; Sintorn, I.-M. TEM virus images: Benchmark dataset and deep learning classification. *Comput. Methods Programs Biomed.* **2021**, *209*, 106318. [[CrossRef](#)]
41. Kerner, H. Too many AI researchers think real-world problems are not relevant. *Opin. MIT Technol. Rev.* **2020**, 1–15. Available online: <https://www.technologyreview.com/2020/08/18/1007196/ai-research-machine-learning-applications-problems-opinion/> (accessed on 23 June 2022).
42. Wagstaff, K. Machine learning that matters. *arXiv* **2012**, arXiv:1206.4656.
43. Gong, J.; Liu, J.; Li, H.; Zhu, H.; Wang, T.; Hu, T.; Li, M.; Xia, X.; Hu, X.; Peng, W. Deep Learning-Based Stage-Wise Risk Stratification for Early Lung Adenocarcinoma in CT Images: A Multi-Center Study. *Cancers* **2021**, *13*, 3300. [[CrossRef](#)] [[PubMed](#)]
44. Xu, Y.; Wang, S.; Sun, X.; Yang, Y.; Fan, J.; Jin, W.; Li, Y.; Su, F.; Zhang, W.; Cui, Q.; et al. Identification of Benign and Malignant Lung Nodules in CT Images Based on Ensemble Learning Method. *Interdiscip. Sci.* **2022**, *14*, 130–140. [[CrossRef](#)] [[PubMed](#)]