

Lung Nodule Texture Detection and Classification Using 3D CNN

Ivan William Harsono¹, Suryadiputra Liawatimena², and Tjeng Wawan Cenggoro³

¹Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University

²Computer Science Department, BINUS Graduate Program - Doctor of Computer Science, Bina Nusantara University

²Computer Engineering Department, Faculty of Engineering, Bina Nusantara University

³Computer Science Department, School of Computer Science, Bina Nusantara University

³Bioinformatics and Data Science Research Center, Bina Nusantara University
Jakarta 11480, Indonesia

Email: ¹ivan.harsono@binus.ac.id, ²suryadi@binus.edu, ³wcenggoro@binus.edu

Abstract—Following artificial intelligence implementation in computer vision field, especially deep learning, many Computer-Aided Diagnosis (CAD) tools are proposed to help to detect lung cancer by the scoring system or by identifying the characteristics of nodules. However, lung cancer is a clinical diagnosis which does not provide detailed information needed by radiologists and clinician to prevent unnecessary invasive diagnostic procedures compared to lung nodule texture detection and classification. Hence, to answer this problem, this research explores the steps needed to implement 3D CNN on raw thorax CT scan datasets and usage of RetinaNet 3D + Inception 3D with transfer learning. The 3D CNN CAD tools can improve the speed, performance, and ability to detect lung nodule texture instead of malignancy status done by previous studies. This research implements 3D CNN on Moscow private datasets acquired from NVIDIA Asia Pacific. The proposed method of data conversion can minimize information loss from raw data to 3D CNN input data. On training phase, after 100 epochs, the researchers conclude that the best-proposed model (3D CNN with transfer learning of pretrained LIDC public datasets weight) provides 22.86% of mean average precision (mAP) detection capability and 70.36% of Area Under the Curve (AUC) in Moscow private dataset lung texture detection tasks. It outperforms non-transfer learning 3D CNN model (trained from scratch) and 3D CNN with transfer learning of pre-trained ImageNet weight.

Index Terms—Image Processing, Lung Nodule, Pattern Recognition, Convolutional Neural Network, Transfer Learning

I. INTRODUCTION

LUNG cancer or malignant lung nodules is one of the leading causes of death in the world. However, lung cancer is difficult to diagnose clinically, so it requires imaging tests such as Computed Tomography scan (CT-scan), Positron Emission Tomography (PET), and lung biopsy with a gold standard examination. Lung nodules or pulmonary masses are abnormal growths of lung parenchymal cells. Lung nodules can generally be found in every patient during a medical check-up. However, only a few of them are at risk of becoming malignant and causing death. Epidemiologically, lung nodules are generally found in male population with the age group of 60 years and history of heavy smoking (20 – 40 packs per year for more than 20 years) [1, 2]. The size of the lung nodule varies greatly. The appearance on the CT scan resembles a small dot, fine needle, up to the size of a golf ball (golf-ball appearance). Usually, the doctor ignores the nodules less than 3 mm (generally not visible on CT scans) because of the small potential of the nodules to become a malignancy [2, 3]. One of the important characteristics to determine malignancy is the texture or solidity of lung nodules. Generally, lung nodules can be categorized into solid, sub-solid, and ground-glass based on their texture. Nodules that are at risk of being malignant generally have a sand-like appearance or are referred to as a ground-glass appearance. Those are larger than 4 mm. The location of lung nodules can be in all parts of the lung, including the lung wall, airway, lung fissure, or blood vessels. They are usually accompanied by enlargement of the local lymph nodes making it difficult to diagnose because the location and shape resemble the anatomical structure of the

lung [1, 3–9].

Indonesia is one of the countries with the highest population growth and the low ratio of healthcare professionals [10, 11]. Additionally, Indonesia is also one of the largest cigarette producers in Indonesia. People can easily buy cigarettes at kiosks and minimarkets, causing lung cancer as the most common cancer in Indonesia. Difficult access to adequate health concentrating on large islands such as Java and Sumatra is also one of the causes of the increase in undiagnosed cancer at an early stage and cancer mortality [10, 12, 13].

Generally, when radiologists interpret Computed Tomography scan (CT-scan) images, whether patients have lung cancer or a suspicion of malignancy, especially in the early stages, expertise and consensus of three or more radiologists are needed to establish the diagnosis of lung cancer [7, 9, 14]. While radiologist needs to reach consensus in making a diagnosis of lung cancer, there are still outpatients and inpatients who do radiological examinations and require readings by a radiologist. Thus, this issue causes an increase in radiologists' workload, which affects the detection and diagnosis of lung nodules [1, 15–17]. Surveys have shown that the detection of lung nodules performed by radiologists manually has a high False-Positive Rate (FPR) of 51–83.2% with a sensitivity of 94.4–96.4%. To address workload imbalance suffered by radiologists and high FPR of lung nodule detection, radiologists need reliable Computer-Aided Diagnosis (CAD) tools to help them to interpret accurately and reach consensus faster [7–9, 18–20].

Studies in the last five years have proven that CAD using deep learning techniques such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and Deep Belief Networks have successfully surpassed conventional CAD techniques. For example, there are Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), and fractal analysis [7, 21]. However, to the best of researchers' knowledge, there is the only implementation of 3D CNN to classify and detect lung nodule malignancy. There is no research about implementing 3D CNN to detect and classify lung nodule texture [1, 7–9, 14, 18, 19, 22–24]. Detection of a lung nodule, especially lung nodule texture detection and classification, is more challenging compared to lung nodule malignancy because many lung nodules have ambiguous opacity. It can mimic another adjacent structure, so it causes different interpretations of radiologists, which usually solved by comparing the interpretation of multiple radiologists, which is time-consuming. Nodule texture information can be used to predict lung biopsy results. It can reduce the risk of unneeded invasive actions such as

lung biopsy or bronchoalveolar lavage, especially in annual screening settings. The diagnosis of malignancy is clinical and not useful to reduce the number of invasive actions needed to ascertain the nature of malignant nodules [1, 16]. Therefore, in this research, the researchers explore the use of 3D CNN for lung nodule texture detection and classification instead of malignancy detection. It can be used as CAD tools to conduct initial screening, help consensus, prevent unneeded invasive diagnostic procedures, and provide a rough picture of nodule location. It adapts the Inflated 3D Inception (I3D) architecture combined together with RetinaNet using pre-trained ImageNet weight and LIDC weight transfer learning.

II. RELATED WORK

CAD tool is a device used to support various aspects of diagnosis in the health environment. Generally, various hospitals, indirectly use CAD without them knowing, such as automatic interpretation of ECG machines by using signal analysis with machine learning [3, 18]. However, CAD can function more than just interpreting binary signals/data. CAD can help to stratify risks and detect certain objects from medical images [25]. The purpose of using CAD in interpreting medical images is to ease the workload of medical image interpreters (pathologists or radiologists). Thus, they can provide fast, accurate, complete, and precise results of the object of interest class and location, such as solid and non-solid nodule locations [9, 26].

The CAD design process can be carried out in a supervised or semi-supervised manner. Nowadays, it is closely related to the implementation of artificial intelligence in healthcare. Generally, artificial intelligence can be classified into two based on the algorithm used, namely machine learning and deep learning. Machine learning engages in learning, exploring, and constructing algorithms that can learn adaptively and make predictions from existing data or algorithms following static program instructions based on prediction or decisions predetermined from previous samples. This method is applied to complete computational tasks in which the design and programming of explicit algorithms with good performance are difficult. The examples are the email filtering applications, network intruder detection or data breaches, optical character recognition, and computer vision learning [27]. Deep learning is a branch of machine learning that uses the abstraction capabilities of the neural network (the smallest unit of deep learning design) to solve problems in the Machine Learning (ML) domain. Deep learning design is inspired by the workings of the human brain that can do extraction, abstraction, and

decision making by the interconnection between neurons (the smallest unit of the human nerve) [26, 27]. Because of its ability to adapt and extract complex features by conducting hierarchical and abstract learning, deep learning, especially CNN, is the first choice for completing various computer vision tasks. It includes medical image analysis [3, 26].

A typical CNN framework generally consists of several convolutional layers and subsampling or pooling layers. Those are fully connected to a multilayer perceptron. To analyze 2D images, the dimensions of the convolutional layer are generally set to two to take the local spatial pattern of the object under research. The size of the convolutional layer is smaller than the input layer and can be stretched with multiple parallel map features. Feature maps can be interpreted as input or map images that are convolved with linear filters (sigmoid functions). They are parameterized with synaptic weights and biases [28–30]. Neighboring hidden units on the feature map are the result of replication of the unit through the same weight vector and bias parameters to reduce the number of free parameters that can be examined. Each subsampling layer performs non-linear subsampling of the feature map. Non-linear subsampling chooses the maximum value of each overlapping subregion on the input map. The main function of the subsampling layer is to reduce the learning complexity of the upper layer and is invariant to the translation effect of image input. The CNN model can be built from training data with the gradient back-propagation method of multiple layer layers connecting the top convolutional layer, down to the next convolutional layer, until reaching the lowest convolutional screen. It is by adjusting the weight and bias between layers [9, 30, 31].

Generally, the close layer to the input image has fewer filters. In the farther layer from the input image, the number of filters will increase. Although CNN's performance is tremendously good, it has the following shortcomings. First, CNN requires much training data that has been labeled, which is difficult to fulfill, especially reading expertise from experts like radiologists. It requires expensive costs, and the number of specific diseases is small in number. Second, training deep CNN requires computational ability, large memory allocation, and time-consuming. Third, there is the possibility of overfitting and convergence problems, so that it requires repeated adjustments to the architecture and parameters of the neural network to ensure all layers of learning with good speed [3, 22, 29, 30, 32].

The rapid growth of image data on the Internet has provided opportunities for models and algorithms to index, organize, and interact with the existing image and multimedia data. However, there is still no compre-

hensive annotated image database that can be used as standard in object detection and classification. Thus, to solve this problem, ImageNet is created to provide a hierarchical solution to image recognition by annotating images on the internet [14, 28, 31, 33]. AlexNet [33], Overfeat [34], and Region-Based Convolutional Neural Network (R-CNN) family [35–38] are some of the popular CNN using ImageNet database to train its model. After successfully proving that the CNN method as the state-of-the-art in computer vision tasks, some developments made in object detection and classification are transfer learning methods [14, 28, 33, 39, 40], action recognition using stream CNN on kinetics datasets (such as Two-Stream Inflated 3D ConvNet) [41], Region Proposal Network (RPN) and Region of Interest (RoI) used by Faster R-CNN and Mask R-CNN (two-stage object detectors) [37, 38]. The presence of RetinaNet as single-stage object detectors is faster and outperform two-stage object detector (Faster R-CNN) and its predecessor (YOLO and DSSD) performance by balancing foreground and background classes with the function of focal loss cross-entropy and feature pyramid network (FPN) for multiscale extraction on image feature maps [42–47].

Previous researchers propose several techniques for lung nodule detection. First, it is multi-scaling CNN layers combined with machine learning methods (Support Vector Machine (SVM) or Random Forest (RF)) as classifiers after CNN [3, 7, 26]. Second, it transfers learning from 2D CNN [3, 7, 9, 15, 21, 48]. Third, they use multiple parallel layers 2D CNN (2.5D CNN). Fourth, they use 3D CNN trained from scratch [3, 8, 15, 20–23, 26, 48]. Fifth, it is transferring learning from pre-trained 3D CNN with previously trained natural images or lung nodule database weights [15, 20, 23, 24, 39, 48]. Sixth, it is using Dense Neural Network (DenseNet) or dual path Fully Convolutional Network (Dual FCN) to classify, detect, or segment lung nodules [15, 20, 24, 49, 50].

Most medical imaging studies implement 2D and 2.5D CNN for medical image analysis. It is because the image size is large compared to small objects of interest, such as lung nodules. 3D CNN is the most accurate approach, but it uses more computing resources (memory and GPU). It is also difficult to apply because of the depth dynamics presenting in medical images [12, 15]. All methods mentioned before are implemented to classify lung nodule malignancy on public lung nodule datasets. However, in healthcare professional points of view, it is presumptuous for medical detection tools to diagnose patients with cancer. The diagnosis is not made by CT-scan interpretation, but by requiring higher abstraction,

negotiation, decision making, and emotions (empathy) to deal with patients clinically condition. The CAD is unable to fulfill psychological needs and anxiety resulting from diagnosis. Thus, instead of directly verdict patients with the final diagnosis, this research sees CAD as a tool to help radiologists and clinicians (pulmonologists) to achieve a clinical diagnosis. It is by providing lung nodule location and texture information. To the best of researchers' knowledge, there is still no research regarding implementing 3D CNN to detect and classify lung nodule texture.

III. RESEARCH METHOD

The method proposed in this research consists of three main phases. Those are data profiling, data cleaning, filtering, and data compatibility, 3D CNN architecture design and transfer learning, and training and evaluation.

A. Data Profiling

This research aims to implement 3D CNN on Moscow private dataset. Moscow private dataset is obtained from the Medical Radiology Center of Moscow Healthcare Department in collaboration with NVIDIA Asia-Pacific via a private link provided by the NVIDIA-Binus AI R&D Center. Moscow private dataset has 546 patient data annotated by at least three radiologists for each patient. Moscow private dataset also provides annotation information on nodule coordinates (x, y , and z), nodule size in mm, and nodule texture (solid/subsolid/ground-glass). All annotation information for 546 patients is presented in one Microsoft Excel (.xlsx) file in Russian and one Microsoft Excel (.xlsx) file in English. From 546 patients, 472 patients have lung lesions (pulmonary nodules), and 74 have no pulmonary nodules. About 472 patients may contain one or a combination of two or three types of pulmonary nodules. From patients who have pulmonary nodules, the number of solid: subsolid: ground-glass annotations are obtained as many as $2136 : 778 : 384 = 5.56 : 2.02 : 1$. To reduce the imbalance classes between solid, subsolid, and ground-glass classes, non-solid classes are formed. Those are the sum of the subsolid classes and ground-glass classes. The final ratio obtained after averaging the doctor's expertise for each nodule observation gives a ratio of solid nodules: non-solid ($2136 : 1420 = 1.83 : 1$). The detailed data class distribution can be seen in Table I. However, due to the annotation structure, annotation agreement, and raw image quality in Moscow private dataset resulting in poor quality, a data cleaning process is needed. It consists of several

processes. First, it is matching the annotation coordinates of each patient with a CT scan of the chest. Second, it is matching the z coordinates (depth) with the DICOM metadata information that each CT image slice has a thorax scan (unique alphanumeric). Third, it is generating pseudo mask coordinates (pixel-wise annotation) and converting CSV annotations to XML format.

B. Data Cleaning, Filtering, and Data Compatibility

The processing of the Moscow private dataset consists of 546 patients with a total of 674 thorax (body/tissue) CT or lung CT scans with slice thickness (distance/thickness between slices) of 0.5 mm or 1 mm. The annotation contains information on the patients' code (or the name of the main folder), doctors' code (000, 001, 002, 003, 004, 006, 007, 008, 008, 009, 010, 011, 012, 013, 014), 3D coordinates of picture dataset based on x, y , and z coordinates of lesions and texture (solid, subsolid, ground-glass) of lung nodule lesions. Moscow private dataset requires data cleaning because of several reasons.

First, there is no agreement on coordinates naming by radiologists. The origin coordinates system is supposedly located in the upper left/upper right side on the corner of the patient. As for the $z = 0$ coordinates, the majority of doctors still use the benchmark in which the closest location to the head will be designated as $z = 0$ (contrary to the World Coordinate System standard). Meanwhile, other doctors use World Coordinate System standards in which the closest location to the foot is designated as $z = 0$.

Second, data cleaning is 10% of coordinating the annotation data (x, y , and z) by pointing to areas outside the lungs (bone, liver, or neck). It may be caused by the coordinating process carried out before the CT scan slice thickness. It is standardized to ensure 0.5 mm and 1.0 mm slice thickness. There are also image data providing two types of slice thickness (0.5 mm and 1.0 mm) with single annotation not specifying slice thickness. Thus, it increases the difficulty of determining whether the coordinates given are suitable for images with 0.5 mm or 1.0 mm of slice thickness. The 0.5 mm slice thickness is not necessarily two times the number of 1.0 mm slice thickness.

Third, the data cleaning is needed as the patients who do not have nodules are mixed together with positive nodule data. Thus, it results in 96 non-nodule datasets to be excluded from the training process. Fourth, the quality of the resolution of the CT scan in Moscow private dataset has a bad quality of image resolution compared to other publicized CT scan datasets such as LIDC public dataset [3, 7–9, 15, 20–23, 26, 39, 48]. It may be caused by the suboptimal

TABLE I
THE DETAILED CLASS DISTRIBUTION OF MOSCOW PRIVATE DATASET.

		Solid	Subsolid	Ground-Glass
Number of Data	Number of patients	413	246	144
	Number of Annotation	2136	778	384
Size of Nodule (mm)	Average	8.27	7.27	8.89
	Standard Deviation	7.25	5.08	5.77
	Max	113.00	53.00	48.00
	Min	1.50	2.00	2.00
Number of Slice	Average	493.14	504.15	443.30
	Standard Deviation	214.86	226.40	178.66
	Max	1630.00	1284.00	1484.00
	Min	60.00	301.00	173.00
Pixel Spacing (mm)	Average	0.71	0.69	0.70
	Standard Deviation	0.09	0.09	0.09
	Max	1.04	1.04	1.04
	Min	0.53	0.52	0.54
Distribution of Data	0.5 mm Slice Thickness	319	131	30
	1.0 mm Slice Thickness	1816	647	354
	5.0 mm Slice Thickness	1	-	-

regulation of power source (kilovolt peak, tube current, and tube speed rotation), noise filtering, or non-optimal iterative reconstruction. The Poisson noise by sub-optimal iterative reconstruction causes bright and dark lines, so lowers the image quality [16, 25, 51].

Fifth, Moscow private dataset also includes patients who have pulmonary radiological lesions besides the pulmonary nodules into datasets (interstitial lung and pleura inflammation or infection such as pneumonia, bronchiolitis, interstitial lung disease, pleuritis, hydrothorax, and post-pneumectomy) in addition to nodule datasets. However, it is understandable since most terminally ill lung cancer patients have a high risk of lung infection and inflammation. Sixth, manual data cleaning in the existing datasets cannot be viewed using DICOM viewer programs such as Aliza and MITK. It can only be viewed and reconstructed correctly using DICOM reconstruction tool applications such as 3D Slicer [52] due to error in orientation. It is suspected due to the corrupted metadata information caused by machine malfunctions.

The steps taken to overcome the first to third problems are the researchers review the raw CT scan dataset manually and match them carefully with their respective annotations one by one. It is to ensure that none of the annotations is out of bounds of the lung area. The reviews of the pulmonary CT scan dataset show that there is only one consistent doctor (code number: 011) out of a total of 15 radiologists participating in reading pulmonary CT scan data.

To overcome the z coordinate standardization problem mentioned in the first problem, the researchers match the depth coordinates by using spatial information on DICOM metadata (code: 0020, 0032). It is to accurately describe z coordinates, rather than using

inconsistent references by Moscow radiologists. After matching the coordinates with the DICOM metadata information, the data cleaning process is continued by dynamically generating pseudo mask pixel-wise annotation coordinate and is saved in XML format. After cleaning the data for the Moscow private dataset, each DICOM file (standardized file of medical images) of the Moscow private dataset is converted into 3D volumetric Nearly Raw Raster Data (NRRD) format to ensure data volatility and intensity (Hounsfield unit (HU)) normalization [2, 6, 53]. As for the Moscow private dataset, due to a large number of DICOM files in which the metadata corruption arises, the DICOM datasets are reconstructed and converted one by one manually using the third-party apps (3D Slicer) [52]. However, the outputs of NRRD files are still lower in terms of quality compared to LIDC public dataset because of Poisson noise.

After the conversion process of DICOM to NRRD format, the next step is to convert annotation information from XML to NifTi and planar (.pf) format. Those serve as a pseudo mask. The NRRD datasets are converted to 4D NumPy array with dimensions of c (picture channels), x (length of picture datasets, default = 512), y (width of picture datasets, default = 512), and z (depth of picture datasets, variable depending on patient age, and body size) = 2, 512, 512 for the images and Region of Interest (RoI) mask. NumPy array image size adjustment methods are as follows. Each NumPy array with dimensions (channel: z, x, y) = (number of slices: 512, 512, 2) will be standardized in size by readjusting the fixed pixel spacing (distance between pixels on the x and y axis is 0.7 mm) and fixed slice thickness (the distance between the images on the z -axis is 1.25 mm). The reason for using 0.7

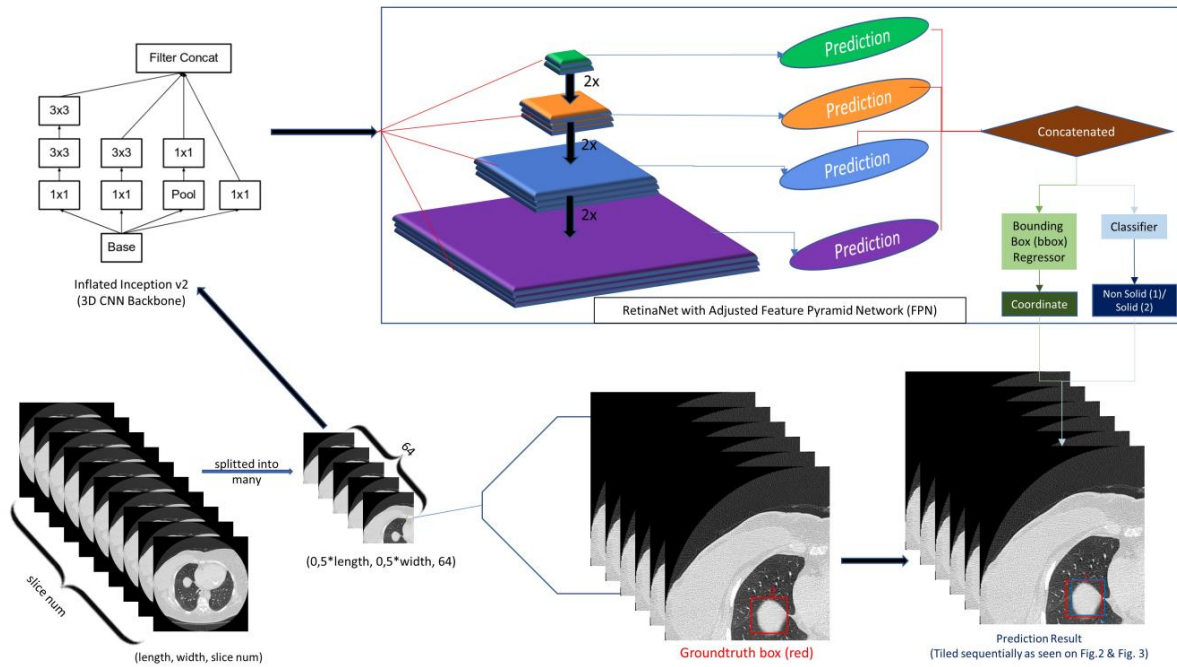


Fig. 1. The 3D CNN architecture.

mm pixel spacing is to accommodate the smallest pulmonary nodule size of 3 mm (4.3 pixels). Thus, it can be read with the smallest detection anchors size, which is 4×4 pixels. However, since Moscow private dataset is high in-depth variable, slice thickness, and pixel spacing, it is impossible to be fitted into permanently fixed input of $96 \times 96 \times 96$ as proposed by Ref. [50]. First, Moscow dataset uses 0.5 mm and 1 mm slice thickness with high pixels spacing variable. It is widely ranging from 0.54–1.04 pixels spacing, which encompasses adults, geriatric, and pediatric population. Meanwhile, LIDC in DeepLesion uses 1.25 mm and 2.5 mm slice thickness with 0.6–0.9 pixels spacing. Because most of LIDC population are adults, Moscow private datasets need the standardization of using 0.7 mm pixels. Second, there are a proportion of Moscow datasets contain a large nodule (around 113 mm). It is larger than 96 pixels in width and length [50]. Thus, the approach using dynamically larger pixel length and width with a smaller depth size ($0.5 \times \text{length}$, $0.5 \times \text{width}$, 64) reduces the size of the input data to minimize RAM and GPU memory. It also preserves volumetricity or 3D spatial information of a lung CT scan.

C. 3D CNN Architecture Design and Transfer Learning

The 3D CNN architecture adopts the I3D backbone and single-stage of state-of-the-art object detector and RetinaNet 3D, which each the source code can be downloaded from Github public repository. The researchers also include Feature Pyramid Network usage as original RetinaNet architecture by adjusting the feature size with anchors resizing to one-eighth of original RetinaNet paper. It results in the smallest anchors size equalling to 4×4 pixels to enable nodule (small object) detection. Instead of using generic ResNet-50 and ResNet-101 backbone for feature extraction, the researchers use I3D backbone. It is acquired by inflating Inception V2 (2D CNN), which has been already proven to be state-of-the-art architecture for kinetic video datasets feature extraction. It can also transfer learning using pre-trained weight from the natural image database (ImageNet) to 3D CNN. Because 3D medical image datasets consist of many slices, it can also be viewed as multiple frames of kinetics dataset such as the kinetics dataset database used by two-stream I3D research for action recognition classification by Ref. [41]. The researchers can empirically justify the use of the pre-trained I3D backbone to replace the backbone ResNet used by the original RetinaNet architecture to enable transfer learning [41].

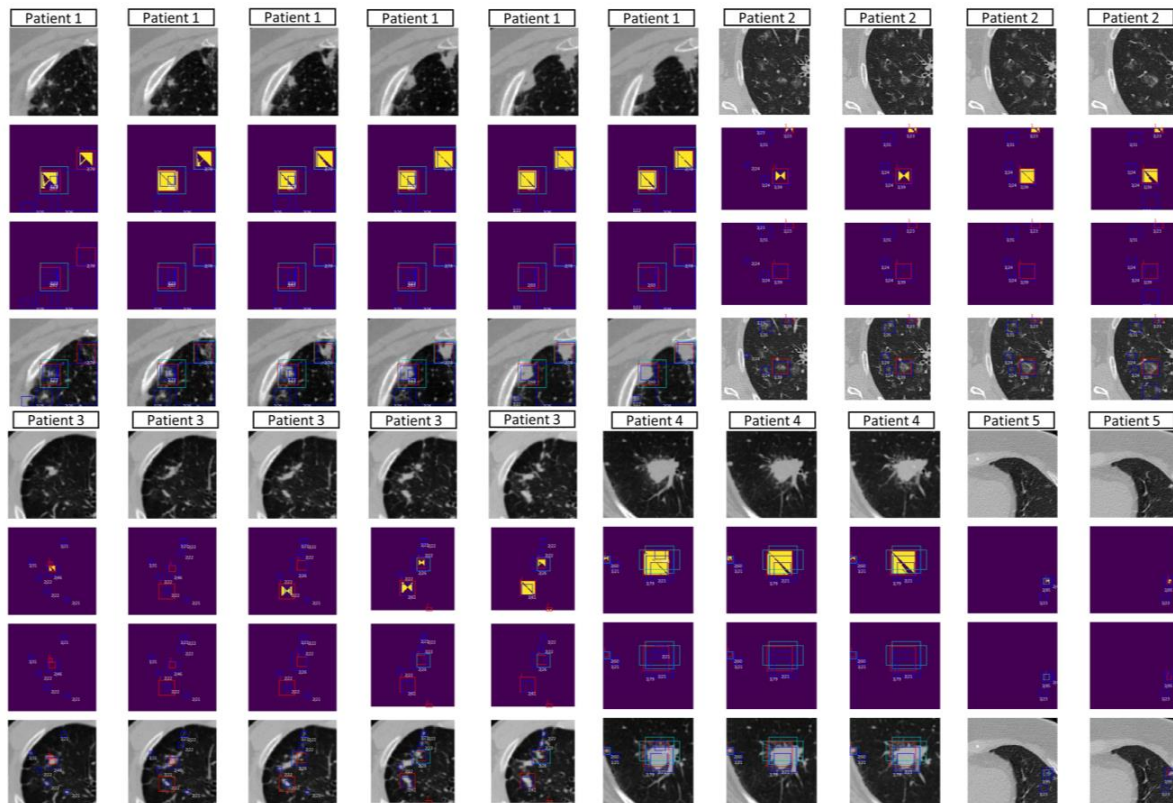


Fig. 2. The example of 20 patch sequential patches (10×2) taken from five different patients. It is correctly detected and predicted by 3D CNN architecture with pre-trained LIDC weight for Detection and Classification of lung nodule texture in Moscow private dataset.

The illustration of the 3D CNN architecture can be seen on Fig. 1.

D. Training and Evaluation Setup

Moscow private dataset is separated with the ratio of 3:1:1 (346:116:116) for training, validation, and testing. Experimental design for 3D CNN evaluation in detecting and classifying texture for Moscow private dataset uses 100 hyperparameter epochs, learning rate of 1×10^{-4} , 15 batch size, 115 batch numbers, 10 validation samplings, Non-Maximum Suppression (NMS) of 1×10^{-5} , and temporal ensemble of 5 top epochs which are maximized by GPU memory consumption on 16GB Tesla P100 server.

IV. RESULTS AND DISCUSSION

This research evaluates 3D CNN performance with and without transfer learning and separates it into three sub-experiments. Those are 3D CNN with pre-trained ImageNet I3D weight (3D CNN T-ImageNet) versus 3D CNN trained from zero (3D CNN T-Zero) versus 3D CNN with pre-trained LIDC weight (3D CNN T-LIDC) for 578 Moscow private dataset.

Figure 2 shows the output of the successful true positive prediction of lung nodule texture detection and classification generated by 3D CNN. Then, Fig. 3 presents an example of false-positive non-nodule artifact detected or wrongly classified class. False-positive prediction in detection may be caused by a very small nodule or very lucent nodule (Fig. 2: Patient 6 and Patient 9), ambiguous nodule that retains both solid and non-solid characteristics (semi-opaque) predicted as other classes (Fig. 2: Patient 7 and Patient 9), nodule-like structure (Fig. 2: Patient 8 and Patient 9), or possibly missed or misdiagnosed nodule by annotating radiologists.

Figure 2 is a compilation of the results of CT scans of five patients. They have nodules of different sizes and texture (solid, subsolid, and ground glass). They are presented sequentially from left to right. The top picture of each subsample (below the patient’s number) is a picture of raw data patches. The second row shows a yellow pixel mask ground-truth with a red bounding box and a class description (1 = non-solid, 2 = solid) below the box. The second row also shows a cyan prediction anchor box and a dark blue

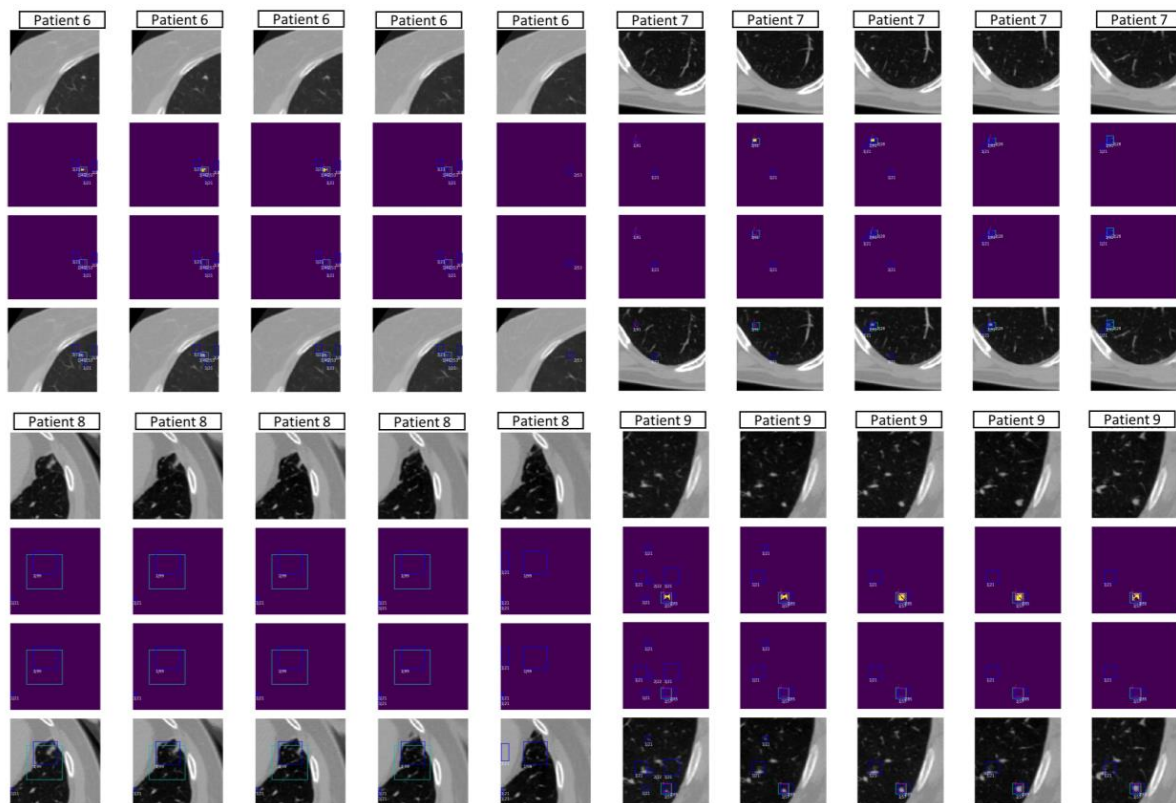


Fig. 3. The example of 20 patch sequential patches (10×2) which are taken from four different patients. The False Positive Rate is predicted by 3D CNN architecture with pre-trained LIDC weight for detection and classification of lung nodule texture in Moscow private dataset.

prediction box with a prediction class and confidence score. For example, dark blue box with description $2 | 100$ signifies a prediction of a solid class with a 100% confidence score. The third row is the filter-out mask result, so it has similarities to the second row except there is no pixel mask ground-truth. The last row is the result of combining the third row with the first row (original patches raw data).

Moreover, Fig. 3 is a compilation of the results of CT scans of five patients who have nodules of different sizes and texture (solid, subsolid, and ground-glass). It is presented sequentially from left to right. Patient 6 shows a very small detected nodule but poorly classified (solid nodule is predicted as non-solid with a score of 53% confidence and solid with 46% confidence). Then, Patient 7 shows a detected semi-opaque solid nodule. It is falsely classified as a non-solid nodule with confidence as high as 91%. This type of misprediction is caused by the semi-opacity of the nodule in the CT scan. It is generally interpreted as a non-solid (subsolid) nodule. However, there is a probability that this type of ambiguous nodule is

misdiagnosed by annotating radiologists. Next, Patient 8 shows a nodule-like structure that has a high non-solid confidence score (91%) caused by underlying disease besides lung nodules such as tuberculosis and lung infection or possibly missed out by annotating radiologists.

Meanwhile, Patient 9 shows an ambiguous nodule that is falsely classified as a solid nodule rather than a non-solid nodule (85% vs 57%) near the falsely detected nodule. There is some hazy structure in the lung detected as a non-solid nodule with a low confidence score of 21%. It can be caused by other lung diseases or missed by radiologists.

Figures 4 and 5 show the Receiver Operating Characteristics (ROC) curve for solid and nonsolid nodules. The complete performance summary is shown in Fig. 6. The 3D CNN T-LIDC and 3D CNN T-ImageNet performance are not different, but these two models implementing transfer learning consistently show superior results compared to 3D CNN T-Zero (mean mAP 22.86% vs 22.16% vs 9.70%, AUC 70.36% vs 69.00% vs 56.10%, Sen 65.46% vs 65.90% vs 47, 54%, FPR

253.51 vs 222.91 vs 278.02) with shorter training times (85.44 vs 85.44 vs 113.88).

Based on the results, it can be determined that the approach can integrate lung nodule CT-scan datasets into the 3D CNN model. 3D CNN is a state-of-the-art of 3D CNN novel model that can provide good detection (mAP) and classification (AUC, sensitivity) performance on a small number of datasets (Moscow public datasets). 3D CNN performance with transfer learning techniques will provide superior results, especially if the transferred weight is trained from similar datasets for the detection of the same object. LIDC transfer learning for Moscow private dataset provides better performance than ImageNet transfer learning. The 3D CNN model is unable to completely predict Moscow private dataset due to the lack of high-quality datasets and low level of trust in annotations. The average mAP for 3D CNN cannot reach 50%. It is because of the strict and demanding recall criteria in assessing the accuracy of 3D bounding boxes (all boxes in 64 adjacent slices must overlap with the Intersection of Union (IoU) 0.1).

Several factors are affecting relatively lower yields in the Moscow private dataset. First, Moscow private dataset contains high noise compared to other cancer or lung nodule datasets (LIDC public dataset). Second, some private datasets cannot be read and converted to 3D volumetric format. Those need advanced DICOM reconstruction software such as 3D Slicer. The conversion results are still far from adequate in terms of quality. Third, there is less amount of patient training data on private datasets compared to the thousand natural image database (ImageNet). Fourth, there is a class imbalance between solid and nonsolid on the Moscow private dataset. Fifth, the Moscow private dataset also includes patients who have pulmonary radiological lesions other than pulmonary nodules, such as patients with lung and pleural inflammation (pneumonia, bronchiolitis, interstitial lung disease, pleurisy, hydrothorax, and post-pneumonectomy). It increases the difficulty of datasets to be abstractly interpreted by CNN.

V. CONCLUSION

The techniques propose a unique way for lung CT scan feature extraction. This model can standardize various age groups and body size (reflected in highly variable slice thickness and pixel spacing) CT scan data into a dynamic dimension of (0.5*length, 0.5*width, 64). It can also give solutions to the unforeseen issue of data. In this case, it is an error in orientation caused by machine malfunctions or sudden movement inside CT scan machines. The data corruption can be partially recovered by using a high-end reconstruction library like 3D Slicer.

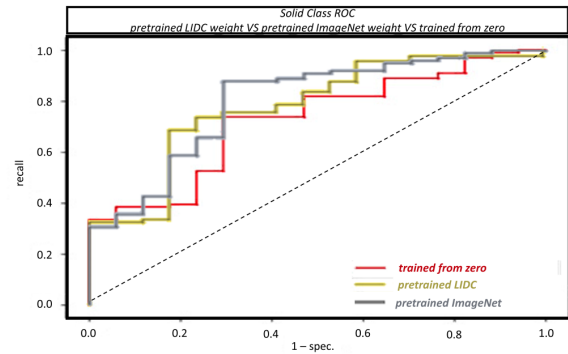


Fig. 4. Receiver Operating Characteristics (ROC) curve for 3D CNN T-ImageNet vs 3D CNN T-Zero vs 3D CNN T-LIDC for Moscow private dataset for solid nodule.

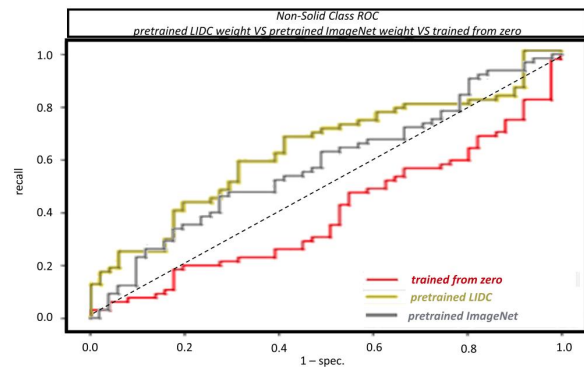


Fig. 5. Receiver Operating Characteristics (ROC) curve for 3D CNN T-ImageNet vs 3D CNN T-Zero vs 3D CNN T-LIDC for Moscow private dataset for nonsolid nodule.

Meanwhile, the noise can be suppressed by adequate iterative reconstruction on raw uncompressed CT scan data (*.raw). Although the architecture is not satisfactory enough due to the limited amount and low quality of the private dataset, the researchers successfully implement and prove the benefit of transfer learning of pre-trained weight on 3D CNN. It has higher performance metrics (higher mAP, AUC, and training time) to non-transferred ones.

The performance metric of 3D CNN pre-trained LIDC weight (similar datasets) is the architecture having the best performance metrics. It is followed by slightly lower 3D CNN pre-trained ImageNet and poor performance metrics of 3D CNN trained from zero, as shown in Fig. 6. Pre-trained LIDC architecture shows a slightly higher FPR compared to pre-trained ImageNet architecture. It is probably caused by high variable characteristics of data on Moscow private datasets compared to LIDC public datasets. The patients with mixed lesions (non-nodule lesions) are also included in Moscow private datasets. Thus, it can be concluded

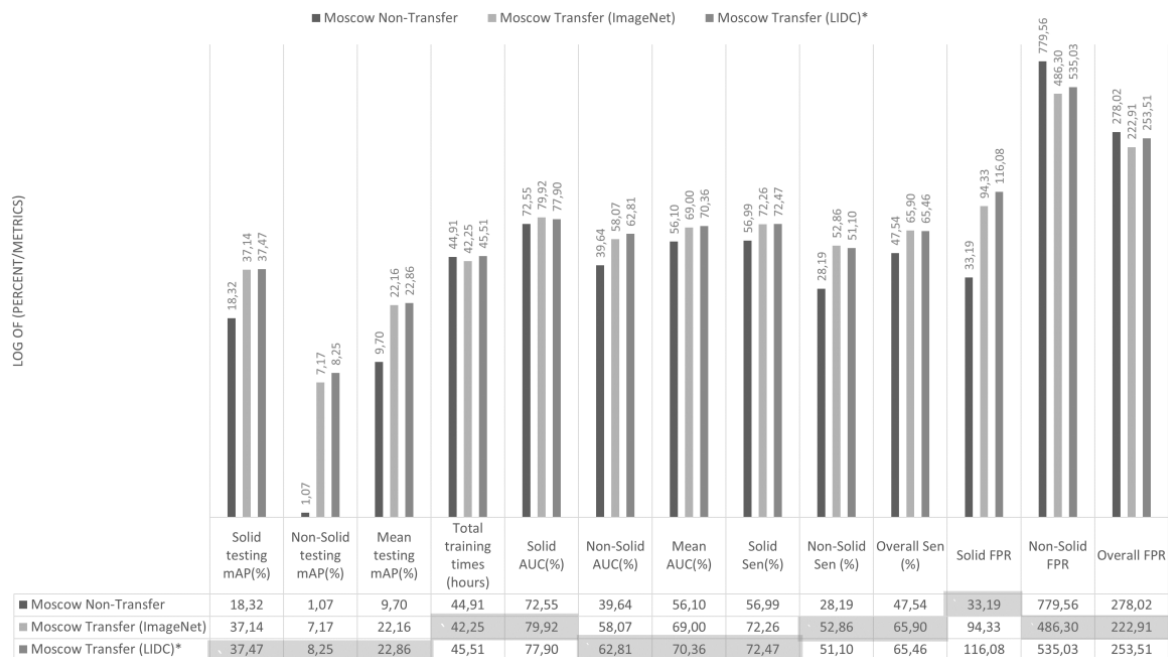


Fig. 6. Prediction performance summary of 3D CNN T-ImageNet vs 3D CNN T-Zero vs 3D CNN T-LIDC for Moscow private dataset.

that this research has successfully proven that private raw datasets such as Moscow private dataset can be integrated into 3D CNN. It also shows the benefit of previously trained transfer learning in a small amount of weak labeled private dataset data. Additionally, this research also uses a unique approach compared to the previous research. All previous studies only classify malignancy status using the LIDC public dataset. This research detects and classifies the texture information contained in Moscow private dataset. The information about textures can help the initial screening, reduce the time needed for consensus (comparing opinion between radiologist), and indirectly score the process to determine lung malignancies. In research involving specific medical images and requiring greater accuracy, the researchers suggest the future researchers to collaborate with medical institutes. Thus, they can obtain highly accurate annotations with high-quality images. Annotation processes should be assisted using special programs that can directly record metadata information from each image in detail (coordinates, pixel spacing, slice thickness, the origin of the image, and pixel-wise perimeter) with the friendly user interface. Then, it can be used by specialist health workers (radiologists) without the needs for special training.

REFERENCES

- [1] G. D. Rubin, “Lung nodule and cancer detection in CT screening,” *Journal of Thoracic Imaging*, vol. 30, no. 2, p. 130, 2015.
- [2] S. J. Swensen, R. W. Viggiano, D. E. Midthun, N. L. Muller, A. Sherrick, K. Yamashita, D. P. Naidich, E. F. Patz, T. E. Hartman, J. R. Muhm, and A. L. Weaver, “Lung nodule enhancement at CT: Multicenter study,” *Radiology*, vol. 214, no. 1, pp. 73–80, 2000.
- [3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [4] S. A. Bolliger, L. Oesterhelweg, D. Spendlove, S. Ross, and M. J. Thali, “Is differentiation of frequently encountered foreign bodies in corpses possible by hounsfield density measurement?” *Journal of Forensic Sciences*, vol. 54, no. 5, pp. 1119–1122, 2009.
- [5] M. J. Budoff and J. S. Shinbane, *Handbook of cardiovascular CT: Essentials for clinical practice*. Springer Science & Business Media, 2008.
- [6] W. Herring, *Learning radiology: Recognizing the*

- basics*. Saunders, 2011.
- [7] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. van Ginneken, "Automatic classification of pulmonary peri-fissural nodules in Computed Tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Medical Image Analysis*, vol. 26, no. 1, pp. 195–202, 2015.
- [8] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2017.
- [9] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, and Y. J. Chen, "Computer-aided classification of lung nodules on Computed Tomography images via deep learning technique," *OncoTargets and Therapy*, vol. 8, pp. 1–8, 2015.
- [10] Kementerian Kesehatan Republik Indonesia. (2016) Informasi SDM kesehatan Indonesia. [Online]. Available: <http://www.depkes.go.id/development/site/tenaga-kesehatan/>
- [11] W. Novitaria, P. A. Wigati, and A. Sriatmi, "Analisis kesiapan pelaksanaan sosialisasi Program Ambulance Hebat dalam rangka dukungan terhadap sistem penanggulangan gawat darurat terpadu di Kota Semarang," *Jurnal Kesehatan Masyarakat (E-Journal)*, vol. 5, no. 4, pp. 164–171, 2017.
- [12] C. R. Nasution. (2016) Kebijakan dalam implementasi SPGDT di Indonesia. [Online]. Available: <https://bit.ly/2NyYqUN>
- [13] Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia. (2015) Situasi global penyakit kanker. [Online]. Available: <http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin-kanker.pdf>
- [14] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep Convolutional Neural Networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [15] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [16] C. Rampinelli, D. Origgi, and M. Bellomi, "Low-dose CT: Technique, reading methods and image interpretation," *Cancer Imaging*, vol. 12, no. 3, p. 548, 2012.
- [17] R. Kakinuma, N. Moriyama, Y. Muramatsu, S. Gomi, M. Suzuki, H. Nagasawa, M. Kusumoto, T. Aso, Y. Muramatsu, T. Tsuchida *et al.*, "Ultra-high-resolution Computed Tomography of the lung: Image quality of a prototype scanner," *PloS One*, vol. 10, no. 9, p. e0137165, 2015.
- [18] B. Van Ginneken, A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf Convolutional Neural Network features for pulmonary nodule detection in Computed Tomography scans," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. New York, NY, USA: IEEE, April 16–19, 2015, pp. 286–289.
- [19] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [20] K. Yan, L. Lu, and R. M. Summers, "Unsupervised body part regression via spatially self-ordering convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC, USA: IEEE, April 4–7, 2018, pp. 1022–1025.
- [21] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale Convolutional Neural Networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*. Sabhal Mor Ostaig, Isle of Skye, UK: Springer, June 28–July 3, 2015, pp. 588–599.
- [22] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network," *IEEE Transactions on Neural Networks and Learning Systems*, 2019.
- [23] S. Hussein, K. Cao, Q. Song, and U. Bagci, "Risk stratification of lung nodules using 3D CNN-based multi-task learning," in *International conference on information processing in medical imaging*. Boone, NC, USA: Springer, June 25–30, 2017, pp. 249–260.
- [24] R. Dey, Z. Lu, and Y. Hong, "Diagnostic classification of lung nodules using 3D neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Washington, DC, USA: IEEE, April 4–7, 2018, pp. 774–778.
- [25] S. P. Power, F. Moloney, M. Twomey, K. James, O. J. O'Connor, and M. M. Maher, "Computed

- Tomography and patient risk: Facts, perceptions and uncertainties," *World Journal of Radiology*, vol. 8, no. 12, pp. 902–915, 2016.
- [26] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [27] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*. Bangkok, Thailand: IEEE, Nov. 22–24, 2017, pp. 1–6.
- [28] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE, June 20–25, 2009, pp. 248–255.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, United States, June 26–July 1, 2016, pp. 770–778.
- [30] Y. Le Cun, "A theoretical framework for back-propagation," *Proceedings of the 1988 Connectionist Models Summer School, CMU, Pittsburg, PA*, vol. 1, pp. 21–28, 1988.
- [31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, Dec. 3–6, 2012, pp. 1097–1105.
- [34] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.
- [35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, June 23–28, 2014, pp. 580–587.
- [36] R. Girshick, "Fast R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea, Oct. 27–Nov. 2, 2015, pp. 1440–1448.
- [37] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Oct. 22–29, 2017, pp. 2961–2969.
- [38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, Montreal, Canada, Dec. 7–12, 2015, pp. 91–99.
- [39] W. Lindsay, J. Wang, N. Sachs, E. Barbosa, and J. Gee, "Transfer learning approach to predict biopsy-confirmed malignancy of lung nodules from imaging data: A pilot study," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Granada, Spain: Springer, Sep. 16 and 20, 2018, pp. 295–301.
- [40] E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Actor-mimic: Deep multitask and transfer reinforcement learning," *arXiv preprint arXiv:1511.06342*, 2015.
- [41] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, United States, July 21–26, 2017, pp. 6299–6308.
- [42] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Oct. 22–29, 2017, pp. 2980–2988.
- [43] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, Oct. 11–14, 2016, pp. 21–37.
- [44] P. Soviany and R. T. Ionescu, "Optimizing the trade-off between single-stage and two-stage object detectors using image difficulty prediction," *arXiv preprint arXiv:1803.08707*, 2018.
- [45] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, United States, June 26–July 1, 2016, pp. 779–788.
- [46] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid net-

- works for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, Hawaii, United States, July 21–26, 2017, pp. 2117–2125.
- [47] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.
- [48] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," *ArXiv Preprint ArXiv:1811.08661*, 2018.
- [49] R. Anirudh, J. J. Thiagarajan, T. Bremer, and H. Kim, "Lung nodule detection using 3D Convolutional Neural Networks trained on weakly labeled data," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785. San Diego, California, United States: International Society for Optics and Photonics, Feb. 27–March 3, 2016, p. 978532.
- [50] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, NV, USA: IEEE, Mar. 12–15, 2018, pp. 673–681.
- [51] F. E. Boas and D. Fleischmann, "CT artifacts: Causes and reduction techniques," *Imaging in Medicine*, vol. 4, no. 2, pp. 229–240, 2012.
- [52] A. Fedorov, R. Beichel, J. Kalpathy Cramer, J. Finet, J. C. Fillion Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3D slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [53] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Muller, and J. Remy, "Fleischner society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.