applied sciences

MDPI

Article

Active Semi-Supervised Learning via Bayesian Experimental Design for Lung Cancer Classification Using Low Dose Computed Tomography Scans

Phuong Nguyen [1,2,*], Ankita Rathod [3], David Chapman [1,2], Smriti Prathapan [1,*], Sumeet Menon [1], Michael Morris [1,4,5] and Yelena Yesha [1,2,6]

[1] Institute for Data Science and Computing, University of Miami, Coral Gables, FL 33146, USA
[2] Department of Computer Science, University of Miami, Coral Gables, FL 33124, USA
[3] Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA
[4] National Institutes of Health Clinical Center, 10 Center Dr, Bethesda, MD 20892, USA
[5] Networking Health, Glen Burnie, MD 21061, USA
[6] Department of Radiology, University of Miami, Coral Gables, FL 33124, USA
* Correspondence: phuong3@umbc.edu (P.N.); sxp3486@miami.edu (S.P.)

Abstract: We introduce an active, semisupervised algorithm that utilizes Bayesian experimental design to address the shortage of annotated images required to train and validate Artificial Intelligence (AI) models for lung cancer screening with computed tomography (CT) scans. Our approach incorporates active learning with semisupervised expectation maximization to emulate the human in the loop for additional ground truth labels to train, evaluate, and update the neural network models. Bayesian experimental design is used to intelligently identify which unlabeled samples need ground truth labels to enhance the model's performance. We evaluate the proposed Active Semi-supervised Expectation Maximization for Computer aided diagnosis (CAD) tasks (ASEM-CAD) using three public CT scans datasets: the National Lung Screening Trial (NLST), the Lung Image Database Consortium (LIDC), and Kaggle Data Science Bowl 2017 for lung cancer classification using CT scans. ASEM-CAD can accurately classify suspicious lung nodules and lung cancer cases with an area under the curve (AUC) of 0.94 (Kaggle), 0.95 (NLST), and 0.88 (LIDC) with significantly fewer labeled images compared to a fully supervised model. This study addresses one of the significant challenges in early lung cancer screenings using low-dose computed tomography (LDCT) scans and is a valuable contribution towards the development and validation of deep learning algorithms for lung cancer screening and other diagnostic radiology examinations.

Keywords: active learning; artificial intelligence; computer-aided diagnosis; deep learning; expectation maximization; lung cancer screening; semisupervised learning

check for updates

Citation: Nguyen, P.; Rathod, A.; Chapman, D.; Prathapan, S.; Menon, S.; Morris, M.; Yesha, Y. Active Semi-Supervised Learning via Bayesian Experimental Design for Lung Cancer Classification Using Low Dose Computed Tomography Scans. *Appl. Sci.* **2023**, *13*, 3752. https://doi.org/10.3390/app13063752

Academic Editor: Maria Stefania Sinicropi

Received: 5 February 2023
Revised: 2 March 2023
Accepted: 10 March 2023
Published: 15 March 2023

Copyright: © 2023 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/).

1. Introduction

Lung cancer is a highly prevalent cancer type and a major cause of cancer-related deaths, accounting for 18.4% of such deaths [1]. It is also among the most commonly diagnosed cancers, accounting for 11.4% of new cancer cases [2,3]. Low-dose computed tomography (LDCT) has been shown to be an effective tool for early lung cancer screening, resulting in a 20% reduction in mortality rates [4]. One of the major challenges of LDCT is the high false-positive rate (23%) [4], leading to unnecessary tests, patient anxiety, and invasive procedures. Currently, all low-dose CT scans (LDCTs) for screening are adjudicated for risk by radiologists using the Lung-screening Reporting and Data System (Lung-RADS) [5], which assigns a score (1, 2, 3, 4A, 4B, or 4X) based on the most worrisome nodule in an LDCT. An algorithm then suggests the next step in the workup of these findings for the clinician. Balancing the high probability of malignancy with the significant difficulty

associated with investigating the nodule's etiology leaves the clinicians with a management conundrum of intervening versus watchful waiting. Biopsy strategies for these nodules are either remarkably invasive (e.g., surgical resection) or fraught with increased risk of complications and poor yields. Waiting for the nodule to grow to a size that is more amenable to biopsy also allows the potential to metastasize whilst waiting. AI strategies demonstrate excellent performance in predicting the malignancy potential of a nodule within a year [6,7]. Computer-aided diagnostic (CAD) tools based on AI have demonstrated high efficacy in low-dose CT lung cancer screening, with the performance levels approaching those of double readings by radiologists. These tools have been recommended for use as second reader [7]. A review of AI-based CAD research revealed that deep learning networks had a pooled sensitivity and specificity of 93% and 68%, respectively, [8].

Advancements in deep learning have led to significant improvements in medical image analysis [9], the accuracy of CAD-based cancer screening in particular [10–16], and the use of CAD tools to support clinical findings [9]. Most fully supervised deep learning algorithms are often limited by scarcity of annotations for medical image analysis. The cost of annotating 3D medical images at the voxel level is high, since it is labor-intensive, requires proficient clinical experts, and could be inaccurate [17–21]. Therefore, algorithms that can learn and generalize from limited, sparse annotations with limited supervision are essential for deep learning models in medical image analysis [22]. Given the rapid expansion of AI applications in healthcare, it is crucial to establish standardized AI software specifications and implement a safe process for continuous evaluation, updating, and evolution that can be used in clinical practice [7,23,24]. In real-world clinical operation, obtaining ground truth data for evaluation and monitoring of AI models require clinicians and experts to provide feedback and annotation, which is impractical for real-time or near-real-time practice. In many cases, the ground truth labels are not available until further procedural tests are performed and confirmed. Transfer learning [9,25,26], self-supervised learning [19,20,27,28], semi-supervised learning [29,30], uncertainty estimations [31], and active learning [32,33] are some approaches to overcome the need for human annotations.

Our study employs both active learning and semisupervised learning techniques to train and evaluate AI models to classify lung cancer malignancy from CT scans using a smaller subset of labeled images. We demonstrate an active, semisupervised expectation maximization for CAD tasks (ASEM-CAD) using pulmonary CT scans. Semisupervised learning is performed using an expectation maximization (EM) algorithm to train CNN-based CAD algorithms with observed and latent image labels (pseudolabels). Previously, we demonstrated semi-supervised and active, semisupervised algorithms for lung cancer screening with limited experiments [34,35]. Our key contributions are:

- We present a novel algorithm that combines active learning with semisupervised expectation maximization to simulate human-in-the-loop by adding additional ground truth (i.e., labeled data) to train and update the model. This algorithm employs a Bayesian experimental design to estimate the uncertainty of prediction outcomes without access to ground truth and to identify unlabeled samples that require expert labeling to enhance the model performance during training or updating.
- The ASEM-CAD algorithm was evaluated using three public CT scan datasets and two deep learning architectures (3D CNN and ResNet34) for lung cancer classification. The experimental results demonstrate high true positive rates and lower false positive rates with significantly fewer labeled images. This accomplishment represents a valuable step towards developing accurate and efficient deep learning algorithms for computer-aided diagnosis (CAD).

To the best of our knowledge, this is the first study to combine semisupervised algorithm with an active learning CAD system that involves human-in-the-loop and demonstrates its effectiveness in classifying malignancy in lung cancer cases (whole CT scans) and malignancy in lung nodules using three low-dose CT scan datasets. The training algorithm and the Bayesian experimental design for estimating uncertainty of the prediction outcome

without accessing to the ground truth labels are generic in that they can be applied to other AI applications.

The paper is organized as follows: In Section 2, we review the current semisupervised approaches. Section 3 outlines our proposed architecture, methods, and algorithm. The experimental design is presented in Section 4, while Section 5 provides a detailed evaluation. Finally, we conclude this work with the main findings and suggest potential future directions in this field of research.

2. Related Work

Applying artificial intelligence (AI) algorithms and deep learning (DL) techniques to automate medical image analysis is an active area of research. Medical image analysis using deep convolutional neural networks (DCNNs) [36–41] relies heavily on annotated 3D medical images, which are difficult to create, owing to the labor-intensive process and necessary prowess that is often unavailable [19–21]. Recent studies have used LDCT-based radiomic signatures for lung cancer screening [42] to predict the survival in patients based on the risk score generated by radiomic models [43].

Semi-Supervised learning (SSL) is a learning approach that can utilize a limited amount of labeled data along with a significant amount of unlabeled data to improve model performance [44]. Typical SSL methods include entropy minimization [45–47], self learning or pseudo labeling [48–51], and consistency regularization [52–55]. SSL is extensively employed in medical image processing tasks such as image detection, segmentation, and classification [30,56–61].

Generative models are another category of techniques commonly utilized in SSL and have been proven to generate additional realistic samples [62]. In [63,64], generative adversarial networks (GANs) utilized labeled and unlabeled data with a localization classifier to extract information from unlabeled data, which was otherwise insufficient. One of the most influential among the generative techniques is expectation maximization (EM), which was introduced by Dempster et al. [65]. Although EM assumes an underlying generative model, it has been shown to be compatible with CNNs [66], in which EM can be applied to improve the semantic segmentation of natural images.

Semi-Supervised Active Learning: Several recent studies have explored the combination of SSL and AL, resulting in semisupervised active Learning (SSAL). For example, in [67,68], SSL was employed to learn the latent representation of both labeled and unlabeled samples, with AL performed on the learned semantic distribution. One popular AL method is uncertainty sampling [69], which selects the least certain data point for acquisition of its ground truth label. Additionally, an AL-based approach [70] was used to overcome data scarcity issues by iteratively selecting the most informative unlabeled samples for labeling and include them in the labeled dataset.

Various techniques have been utilized in prior research to assess the degree of uncertainty in active learning, including Monte Carlo dropout [71], ensemble models [72], and data augmentation for classification tasks [73]. In a recent study, McKinly et al. [74] proposed a loss function, a generalized version of binary cross entropy, that accounts for label uncertainty.

Reversed active learning (RAL) [75] is an approach that remove samples found to be uninformative based on computed confidence intervals. While active learning adds ground truth labels to the dataset, the reverse active learning approach removes samples from the dataset, thereby reducing the number of samples in the overall training set.

3. Methodology

In this section, we discuss the basic components of ASEM-CAD.

3.1. ASEM-CAD Design

ASEM-CAD uses a combination of active and semisupervised learning to improve the prediction accuracy using minimal data labels. The active learning approach is based

on two key principles: (i) a large amount of unlabeled data is available frequently, and (ii) the learning model queries the oracle (e.g., human annotator) for labels during training iterations. Figure 1 shows an overview of the ASEM-CAD learning model. The algorithm initiates the training of a deep learning model by utilizing a subset of training data that has been fully labeled (observed) by experts. Once this process is complete, the algorithm generates an initial model. Then, the initial model is utilized to assign labels to unlabeled (unobserved) images and to retrain the model.

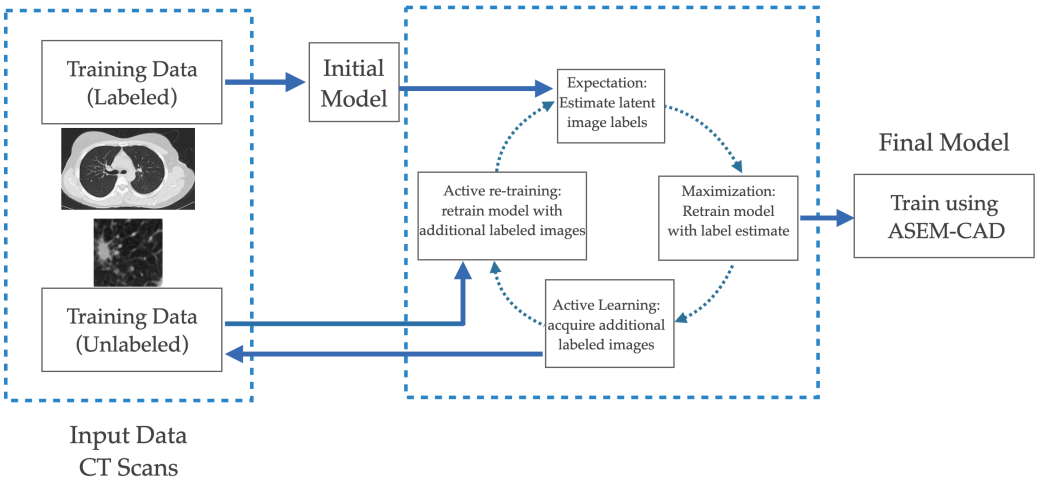


Figure 1. An active, semisupervised expectation maximization (ASEM) framework for CAD tasks.

The expectation maximization (EM) algorithm is employed to perform semisupervised learning. This algorithm estimates the maximum likelihood of unobserved (latent) image labels, given the current model. Each ASEM iteration retrains the model in maximization and active retraining phases, with improved latent variable estimates in each iteration. This approach reduces the computational burden of retraining the model in each iteration by reusing the weights from the previous iterations. While EM attempts to maximize the likelihood, active Learning minimizes the cross entropy. In practical machine learning applications, these have equivalent global optima, assuming statistical independence and approximately equivalent optima, as shown in Equation (1).

$$-\sum_{i} p(X_i) log(p(X_i|\theta)) \approx -log(p(X|\theta)) \quad (1)$$

3.2. Active Learning with Expectation Maximization

Expectation maximization [65] is a widely used clustering algorithm that enables parameter estimation in probabilistic models with incomplete data. The EM algorithm involves two steps per iteration: (i) expectation, which computes the conditional expectation of the log likelihood (i.e., probability distribution) over completions of missing labeled data, given the current model; and (ii) maximization, which involves re-estimating the model parameters to maximize the expected log likelihood of the model, given the current expected value of the latent variables.

In an unsupervised model, EM is initiated in cases in which the classes are predefined and parameters are random. However, in a semisupervised context, the initial model is trained using a subset of training data with ground truth labels. In the subsequent phase, which is unsupervised, the model is retrained on all the unlabeled data to infer the latent variables (i.e., labels). Here, the classification entropy provided by the neural network serves as the clustering metric, which adds on to the loss function.

EM initially generates a classifier (θ). The algorithm uses θ to classify the data and generate a hypothesis based on the labels inferred in the previous step. EM attempts to

determine latent variable Z (i.e., unknown labels), thereby maximizing the likelihood of observing the image X in the neural network model.

The likelihood of the latent variable Z is given by the integral of the joint probability density of all the possible values of the latent variable Z, as in Equation (2):

$$L(\theta; X) = P(X|\theta) = \int p(X, Z|\theta) dZ \tag{2}$$

EM alternates between the expectation and maximization phases to solve the integral in Equation (2). The expected value of the latent variable Z, given the t^{th} iteration of the model θ^t, is calculated in expectation. $E_{Z|X,\theta^t}$, the expected value of the latent variable (Z), can be computed by classifying the probabilities of the unlabeled image labels using the model coefficients (θ^t) in the t^{th} iteration,as shown below in Equation (3):

$$Q(\theta|\theta^t) = E_{Z|X,\theta^t}[logL(\theta; X, Z)] \tag{3}$$

The maximum likelihood model (θ^{t+1}) is computed in the maximization phase, given the expected value of the latent variable (Z). The deep learning model is retrained using the expected values of the image labels in the t^{th} iteration to compute the maximum likelihood model as in Equation (4):

$$\theta^{t+1} = argmax_\theta Q(\theta|\theta^t) \tag{4}$$

Furthermore, active learning optimizes the expected posterior cross entropy of the model, given an alternate experimental design ξ, along with the labeled sample (y_i). The expected posterior cross entropy is expressed as:

$$U(\xi) = -\int log(p(X|\theta, y_i, \xi)) dy_i \tag{5}$$

Applying the Bayes rule to Equation (5):

$$U(\xi) = -\int log(p(y_i|\theta, X, \xi) \frac{p(y_i|\theta, \xi)}{p(X|\theta, \xi)} dy_i \tag{6}$$

The integral in Equation (6) is computationally expensive since it requires retraining of the algorithm for each sample choice and sample label prior to choosing an appropriate sample. To simplify this process, we make an approximation since a single sample has a negligible impact on the model prediction for most samples, and the predicted sample y_i makes the greatest local contribution to posterior cross entropy. Therefore, the change in posterior cross entropy is approximately equal to the normalized classification entropy across all possible K labels, as expressed in Equation (7):

$$\Delta U(\xi) \approx I_{norm}(y_i) = \frac{-1}{log(K)} \sum_{k=1}^{K} p(y_{ik}) log(p(y_{ik})) \tag{7}$$

When active learning is performed in small batches rather than in a single step, the average classification entropy of all of the selected samples in the batch is given as:

$$avg(I_{norm}(Y)) = \frac{1}{|Y|} \sum_{y \in Y} I_{norm}(y) \tag{8}$$

The deep learning model is retrained after the additional labels are annotated by Oracle. The ASEM-CAD algorithm alternates between steps (3), (4) and (8) to enhance the maximum likelihood estimate and reduce the classification cross entropy when latent variables are present, while optimizing Bayesian experimental design. It is essential to note that the algorithm does not ensure convergence to a global optimum. Instead, it attains a local optimal experimental design and a local optimal estimate of latent variables for semisupervised learning.

3.3. Bayesian Approximation

The main factors impacting the performance of the EM meta-algorithm (Algorithm 1) are (i) accuracy of the trained initial model; (ii) the label quality, which depends on Equation (2); and (iii) the acquisition policy for additional ground truth [76]. Monte Carlo (MC) dropout has been used as a Bayesian approximation method to generate labels in a semisupervised approach [77] and to evaluate the prediction uncertainty with active learning [76].

The posterior cross entropy in Equation (5) can be optimized to decrease the model loss, given the model weights θ, experimental design ξ, and the labeled sample (y_i). MC dropout is applied to infer the model output ($\hat{y}$) for a fixed number of iterations to calculate the average probability of outputs as:

$$\hat{y} = \hat{f}_T(X, \theta) = \frac{1}{T} \sum_{t=1}^{T} f^{(t)}(X, \theta) \forall X \tag{9}$$

where $f^{(t)}$ indicates the neural network function with dropout. Classification entropy, as defined in (7), is the metric used to determine how close the output of neural network ($\hat{y}$) is from the predicted labels (pseudolabels) of the samples (as defined in line 8 of Algorithm 1). Other approaches for label acquisition are discussed in [76].

Algorithm 1: Active Learning with Bayesian Approximation

Data: Input Data X, labels G, unlabeled data X', ground truth g'
Result: weights w

1 Initialize the data *(X,G)* with minimal samples and the weights w
2 train w on *(X,G)* to minimize the loss $\mathcal{L}$
3 **repeat**
4 Expectation Maximization step
5 **while** $\forall x' \in X'$ **do**
6 infer the model output y' using f, w and x' with Bayesian Approximation
7 estimate the classification uncertainty via entropy
8 **if** *classification entropy is below threshold* θ **then**
9 $(X, G) \leftarrow (X,G) \cup (x', \psi(y'))$
10 $X' \leftarrow X' \setminus x'$
11 **end**
12 **if** *classification entropy higher than threshold* θ **then**
13 Active Learning: ask the oracle for ground truth g'
14 $(X, G) \leftarrow (X, G) \cup (x', g')$
15 $X' \leftarrow X' \setminus x'$
16 **end**
17 **end**
18 train w for N epochs with the updated (X, G) to minimize the loss $\mathcal{L}$
19 **until** *convergence*;

MC dropout inference is applied to all data points for which the ground truth is acquired from the oracle during training for threshold estimation. The average classification entropy is calculated as in Equation (8). In every iteration of the EM algorithm, all the data samples with entropy values below the chosen threshold are added. The entropy of the average classification results under MC dropout is used for for both semi-supervised and active learning (lines 8 and 12 in Algorithm 1) for all $x' \in X'$, where X' represents unlabeled data points. For the samples with highest entropy, the oracle provides the ground truth (g'). The newly labeled data point (x', g') is added to the set of original data points (X) and labels (G), thereby removing x' from X'.

4. Experimental Design

4.1. Datasets

We demonstrate ASEM-CAD using three publicly available computed tomography (CT) datasets: the National Lung Screening Trial (NLST), the Lung Image Database Consortium (LIDC-IDRI), and Kaggle Data Science Bowl 2017 (Kaggle2017). NLST was a randomized trial conducted at multiple centers to screen lung cancer with low-dose CT for individuals aged 55–74 years with a significant smoking history [78]. Electronic calipers were used by radiologists specialized in standardized image interpretation to measure nodule size. Of the 26,722 patients in the CT screening arm of the NLST, 16,684 were excluded, as no abnormality was recorded in the NLST database. In our study, CT studies from 4075 scans from the NLST dataset were used. Each CT scans was annotated with 1 if the patient was diagnosed with cancer, and the non-cancerous CT scans were labeled as 0. Of the 4075 scans, 639 patients (15.7%) were diagnosed with lung cancer.

Kaggle Data Science Bowl provided labeled low-dose CT scans from 1375 patients to facilitate the development of novel machine learning algorithms for automated CT diagnosis [79]. This includes images with associated binary labels for 356 patients (25.8%) diagnosed with lung cancer within one year of the scan.

The LIDC-IDRI dataset [80] is a publicly available and web-accessible international resource initiated by the National Cancer Institute (NCI), further developed by the Foundation for the National Institutes of Health (FNIH), and supported by the U.S. Food and Drug Administration (FDA). The LIDC dataset consists of thoracic computed tomography (CT) scans for diagnostic and lung cancer screening purposes. The scans contain marked-up annotated lesions provided by four experienced thoracic radiologists. This dataset is the result of a collaboration between seven academic centers and eight medical imaging companies, consisting of 1018 CT cases with thoracic CT scan images in the DICOM format, along with an additional XML file per patient. Based on the nodule size, the lesions were categorized into three categories: nodules > or =3 mm, nodules <3 mm, and non-nodules > or =3 mm. Furthermore, it consists of malignancy ratings from 1 to 5 based on the nodule size, type, and characteristics. In the LIDC-IDRI dataset, nodules that are equal to or larger than 3 mm were annotated by four board-certified radiologists to include subtlety, internal structure, calcification, sphericity, margin, lobulation, spiculation, and texture image characteristics. Subtlety refers to the degree of contrast between the lung and its surrounding tissues [80]. Internal structure relates to the components that are present within the nodule. Calcification is the appearance of calcium in the nodule, with smaller nodules more likely to exhibit visible calcium. A central, non-central, laminated, or popcorn calcification rating is highly indicative of a benign nodule. Sphericity refers to the roundness of the nodule. Margin relates to how well-defined the edges of the nodule appear. Lobulation is an assessment of whether the nodule has a visible lobular shape, which is a sign that the nodule is benign. Spiculation refers to the presence of spicules or spike-like formations along the border of the nodule, with a spiculated margin indicating malignancy. Finally, texture refers to the internal density of the nodule and is an important characteristic when differentiating between partly solid and non-solid textures, which can complicate the process of establishing the nodule boundary. A summary of the three datasets is shown in Table 1.

Table 1. Summary of datasets.

Dataset	Samples	Positive Samples	Type
NLST	4075	639 (15.7%)	Whole CT scans
Kaggle17	1375	356 (25.8%)	Whole CT scans
LIDC-IDRI	4253	1653 (38.9%)	3D Nodules from 1018 CT scans

Data Preprocessing

The Kaggle and NLST data contain chest whole-slice CT scans with varied slice numbers and a slice thickness of less than 3 mm. The CT scans are a 3D volume, with a single intensity value expressed in Hounsfield scale standardized units for each voxel. An axial slice of the dataset is 512 × 512, and the number of axial slices per CT scan varies between 150 and 225 in each volume. Chunks of 20 slices per patient were created by averaging adjacent slices, and each 512 × 512 image was resized to 50 × 50 due to the burden of computing resources such as RAM and CPU processing time. Thus, the dimensions of the input 3D volume for each patient is 50 × 50 × 20, with a label of 1 or 0 representing cancer and non-cancerous cases, respectively. Some benign and malignant nodules from an example CT scan are shown in Figures 2 and 3, in which the difference in appearance between a benign and a malignant nodule can be observed.

The Kaggle dataset includes the data for 1375 patients with 356 cancer cases, accounting for 25.8% of the entire dataset, with the non-cancer cases representing 74% of the data. NLST data include 2538 cases, with 397 and 2141 cancer and non-cancer outcomes, respectively. The LIDC-IDRI dataset consists of 1018 CT scans, and each slice has a thickness of 512 × 512 pixels. Radiologists outlined nodules larger than 3 mm in all sections where they appeared by marking the pixels that comprise the outline at the first pixel outside the nodule. These annotations are provided in the form of nodule regions of interest, along with their z-positions. These spatial co-ordinates where used to create a 3D box and a 3D mask centered on the annotations of the lung nodule. Thus, we cropped the nodules using a box size of 32 pixels × 32 pixels ×, with 16 slices centered on the annotated location of the CT scan. To assign a label for each nodule, the malignancy scores provided by board-certified radiologists were used. The malignancy scores range from 1 to 5 (1 indicates highly unlikely to be malignant, 2 is moderately unlikely, 3 is intermediate, 4 is moderately suspicious to be malignant, and 5 is highly suspicious to be malignant). The nodules of intermediate malignancy (malignancy score of 3) were not considered for binary classification in this work.

In summary, a total of 4253 nodules with dimensions 32 × 32 × 16 were analyzed, each associated with a label of 1 (cancer) or 0 (non-cancer). In total, 1653 cancerous and 2600 non-cancerous nodules were included in the analysis.

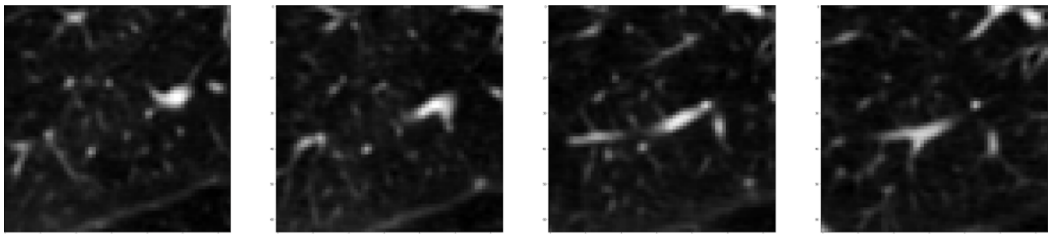


Figure 2. An example of a benign lung nodule from the CT scans.

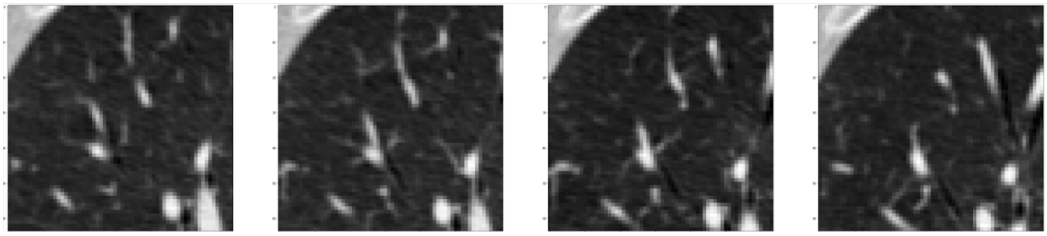


Figure 3. An example of a malignant lung nodule from the CT scans.

4.2. Network Architectures

ASEM-CAD was evaluated using a 3D convolutional neural network (3D CNN) and 3D residual networks.

4.2.1. 3D Convolutional Neural Networks (3D CNNs)

The 3D-CNN model is a variant of the traditional CNN and is composed of a feature extractor and an ANN (artificial neural network) classifier. Figure 4 depicts the model architecture for this evaluation, which consists of five CNN layers and two fully connected dense layers. The input 3D lung volume is passed through the first layer of the CNN using a sliding window of $3 \times 3 \times 3$. The output of the first layer has dimensions of (32,32,16,1) and is passed to the next convolution layer, resulting in an output of with dimensions of (32,32,16,8). Each convolution layer uses a $3 \times 3 \times 3$ kernel and regularizers with a regularization factor of 0.001. The activation function after each convolution layer is ReLU. The max pooling layer downsamples the feature maps before passing them on to the next convolution layer, using a pool size of (3,3,3), resulting in an output of shape of (11,11,6,8). Two more convolution layers are applied to the image, producing an output of shape (11,11,6,16). The final max pooling layer downsamples the output to (4,4,2,16). A dropout layer and a flattening layer are then applied to the image, with the latter reshaping the features into a 1D array that can be fed to the dense layer. The ANN component consists of two fully connected dense layers with ReLU as the activation function. The flattened image passes through the dense layer with 64 neurons, followed by batch normalization and dropout layers with 64 features each. The final layer is an output dense layer with one neuron.

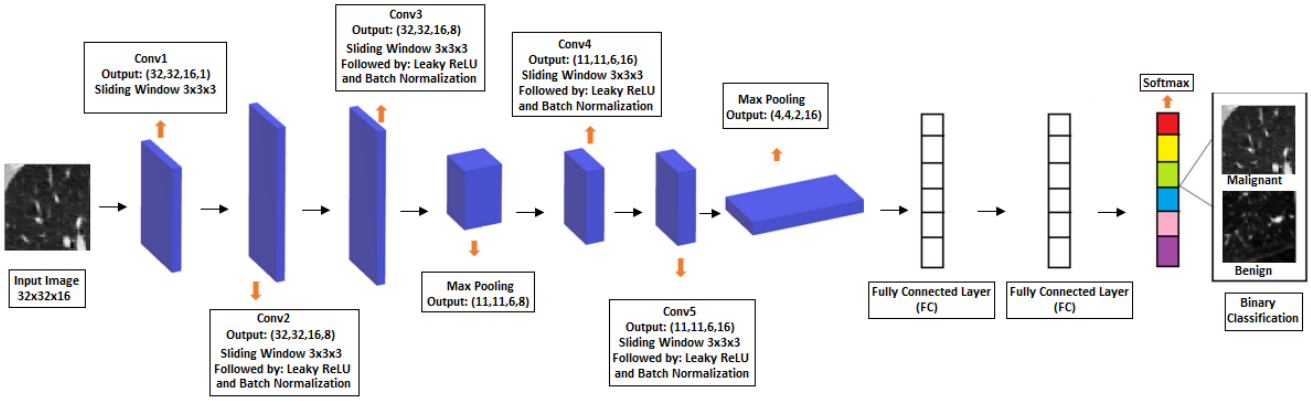


Figure 4. Three-dimensional CNN architecture.

4.2.2. Residual Networks (ResNet)

Residual networks (ResNets) were introduced to tackle the performance degradation in the training of deep CNNs [81] by adding skip connections to allow information to flow through identity mappings, thereby eliminating extra parameters and computational complexity.

We evaluated ASEM-CAD on a 34-layer residual network (ResNet-34). The ResNet-34 model has 34 layers of convolutional and identity blocks [81], with 3 convolution layers in each of the blocks. The ResNet34 architecture has an input layer. The 3D volume data are passed through the input layer. Our architecture uses the following five stages: Stage one includes the convolution layer, batch normalization, and activation, followed by the max pooling layer. Stage two includes the convolution block and two identity blocks (Res block $\times$ 3). Stage 3 includes the convolution block and three identity blocks (Res block $\times$ 4). Stage 4 includes the convolution block and five identity blocks (Res block $\times$ 6). Stage 5 includes the convolution block and two identity blocks (Res block $\times$ 3). Stage 5 is followed by the average pooling layer and dense 512 and two classes.

For supervised learning experiments, the ASEM-CAD model was trained on 80% of the input data, and 20% of the data was used for testing. The initial model was trained with 50% of labeled data using category cross-entropy loss until convergence. The training phase employed an RMSprop optimizer with a learning rate of 0.0001. The initial model was fully trained over 500 epochs using a batch size of 32 samples, and each metaiteration

(EM iteration) was trained over 10 epochs. Once fully trained, the initial model was saved. Subsequent EM iterations loaded the initial model parameters and started the active EM training phase. The active learning component selected 10 samples and requested the oracle to label them during an ASEM-CAD iteration. Up to 10 active EM iterations were performed.

4.2.3. Evaluation Metrics

The AUC ROC curve is a commonly used performance measurement for classification problems. It is based on the receiver operating characteristic (ROC) probability curve and is calculated as the integral of sensitivity with respect to the specificity over the ROC curve domain, with a resulting value between 0 and 1. This score reflects the ability of the model to distinguish between classes. A higher AUC value indicates that the model is better at predicting 0 s as 0 s and 1 s as 1 s. A higher AUC also indicates that the model can differentiate between cancerous and non-cancerous patients more accurately. The ROC curve is a graphical representation of true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis. Other entities, such as sensitivity, specificity, and predictive values, are used to express test performance in different ways. Sensitivity is also known as the true positive rate. The binary classifier confusion matrix is shown in Table 2.

Table 2. Binary classifier confusion matrix.

Labels	Predicted (True)	Predicted (False)
Actual (True)	True Positive	False Negative
Actual (False)	False Positive	True Negative

Sensitivity is the ability of a screening test to determine whether a test correctly generates a positive result for people who have the test condition (true positive rate). Sensitivity is defined in Equation (9).

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \tag{10}$$

Specificity is the ability of a screening test to determine whether a test correctly generates a negative result for people who do not have the test condition (true negative rate). Specificity is expressed as in Equation (10).

We calculated an inflection point along this ROC curve and present sensitivity, specificity, and AUC as the performance metrics to evaluate ASEM-CAD in comparison with other approaches.

$$Specificity = \frac{TrueNegative}{TrueNegative + FalseNegative} \tag{11}$$

5. Evaluation

This section presents the experimental results for ASEM-CAD. We evaluated the performance of a fully supervised model, a semi-supervised model with EM, and the ASEM-CAD model using high-entropy and average classification entropy-based label acquisition policies. We evaluated the results based on metrics such as AUC/ROC, accuracy, sensitivity, and specificity.

The initial model was fully trained until convergence, and for each additional EM iteration, the model was retrained for 10 epochs. We conducted experiments on an AMD processor with a clock rate of 1885 MHz, 32 cores, 658 GB of storage space, and three NVIDIA GeForce RTX GPUs, each with 11 GB of memory. Table 3 displays the training time (wall clock time) for the ASEM-CAD model. The overall increase in the training time is around 30–50%, rather than a tenfold increase in runtime, as the model reuses the parameters (weights) after each iteration instead of retraining. During the maximization phase of

EM, each ASEM-CAD iteration strives to enhance the maximum likelihood estimates of the model parameters. Hence, the weights from the previous ASEM-CAD iteration can be utilized to assess the weights of the next iteration, which significantly reduces the number of epochs, ultimately decreasing the training time for the ASEM-CAD iterations.

Table 3. Training time for the ASEM-CAD algorithm estimated based on 10 EM iterations.

Dataset	Number of Images	Initial Model (Minutes)	ASEM-CAD Iterations (Minutes)	Total Time (Minutes)	% Increase
Kaggle	1357	26	8	34	31
NLST	2538	47	15	62	32
LIDC	4253	8	4	12	50

In addition to active, semisupervised learning with Bayesian approximation, we applied upsampling, label smoothing, batch normalization, hyperparameter tuning, and early termination techniques to improve the model performance.

ASEM-CAD was trained using batches of 32 samples, after which the initial model was saved. In the following phase, EM iterations load the parameters that were saved from the initial model and begin the active EM training process. The active component selects 10 samples and determines which labels to acquire ground truth labels for based on an entropy-based label acquisition policy. For each dataset, we evaluated ASEM-CAD with increasing labeled samples in steps of 10 iterations (step10) and analysis of active learning phases (active10) on two different neural network architectures.

5.1. Initial Accuracy

The model was initially trained with a small percentage of the labeled dataset until convergence. The initial achieved model accuracy is shown in Figure 5 for Kaggle and NLST datasets. Early stopping was implemented, which stops the algorithm execution once the model starts converging. The model started converging around 350 epochs for Kaggle dataset and 230 epochs using the NLST dataset, as observed in the initial accuracy plots presented in Figure 5. The model stabilized after 400 epochs for the Kaggle dataset and 300 epochs for the NLST dataset. The model showed very good stability thereafter up to 500 epochs.

We conducted experiments with 25%, 30%, 40%, and 50% of total training labels. The initial models were trained until convergence. Subsequently, there were 10 EM iterations in which each iteration in the active learning step selects 10 unlabeled samples for which ground truth labels are provided by oracles (experts). Thus, the total number of labeled samples add 9.1%, 3%, and 2% to the initial labels for the Kaggle, NLST, and LIDC datasets, respectively. For example, when the initial model was trained in the active learning phase with 50% of training labels from the Kaggle dataset, by the final iteration, the model would have used 59.1% of the labels (with additional labels being added at each iteration). The fully supervised learning models use 100% of the training labels. The model performances were tested using unseen testing data (20% of dataset). For each dataset, all models were tested using the same unseen test data.

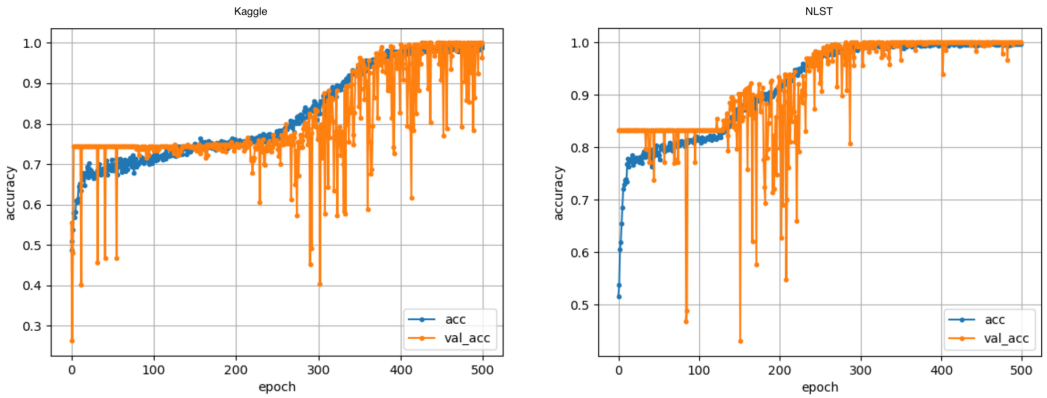


Figure 5. The initial training and validation accuracy for ASEM-CAD with (**left**) Kaggle and (**right**) NLST datasets. The accuracy curves illustrate the ASEM-CAD performance, which is not underfit or overfit with the improvement in validation accuracy over the number of epochs.

5.2. Experiments on the Kaggle17 Dataset

For active phase iterations, the initial model was trained with a fixed number of labeled samples. Subsequently, the model was trained using 10 EM iterations, and 10 ground truth labels were added in each EM iteration.

As shown in Figure 6 (left), the model was trained with 25% of initially labeled samples, and 10 ground truth labels were added per EM iteration. The ROC curve using 3D-CNN architecture for active phase iterations (active 10) is shown in Figure 6. ASEM-CAD with high classification entropy shows a high true positive rate and low false positive rate, with an AUC of 0.89, as shown in Figure 6 (left). The AUC for high classification entropy has a better AUC value in comparison with average classification entropy and semi-supervised classification. A fully supervised model has an AUC of 0.92. However, with 25% of labeled data, ASEM-CAD with high classification entropy provides good results.

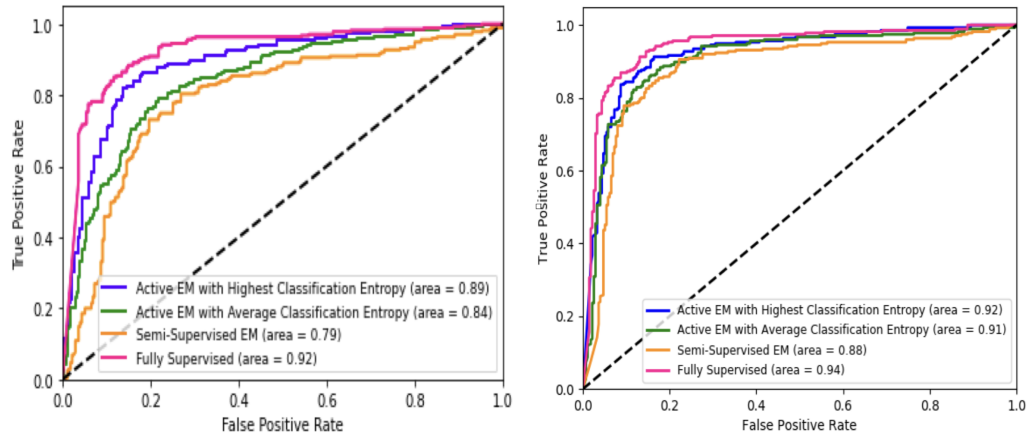


Figure 6. Active phase iterations for the Kaggle dataset using a 3D CNN with (**left**) 34% and (**right**) 59% of training labels. Fully supervised learning using 100% of training labels.

Figure 6 (right) shows the ROC curve with 50% of initially labeled samples for active phase iterations. ASEM-CAD with high classification entropy achieves an AUC of 0.92. This is comparable to a fully supervised model, which uses 100% of the labels, with an AUC of 0.94. The ROC curve shows similarity in TPR and FPR between the performance of our ASEM-CAD model and that of a fully supervised learning model. Our model achieved sensitivity of 0.9 and specificity of 0.81.

The ASEM-CAD performance increases when the number of samples used by ASEM-CAD is increased (see Figure 7). The ASEM-CAD achieved an average and maximum AUC of 0.93 and 0.94, respectively compared with 0.925 and 0.96 for the fully supervised models.

With an increasing number of samples for training experiments, the ASEM-CAD performed as well as the supervised learning model (using 100% of labeled samples) but only used 59% of labeled samples.

Avg Our ASEM-CAD	Avg fully supervised	Max Our ASEM-CAD	Max fully supervised
0.93	0.925	**0.94**	0.96
Sensitivity		Specificity	
0.9		**0.81**	

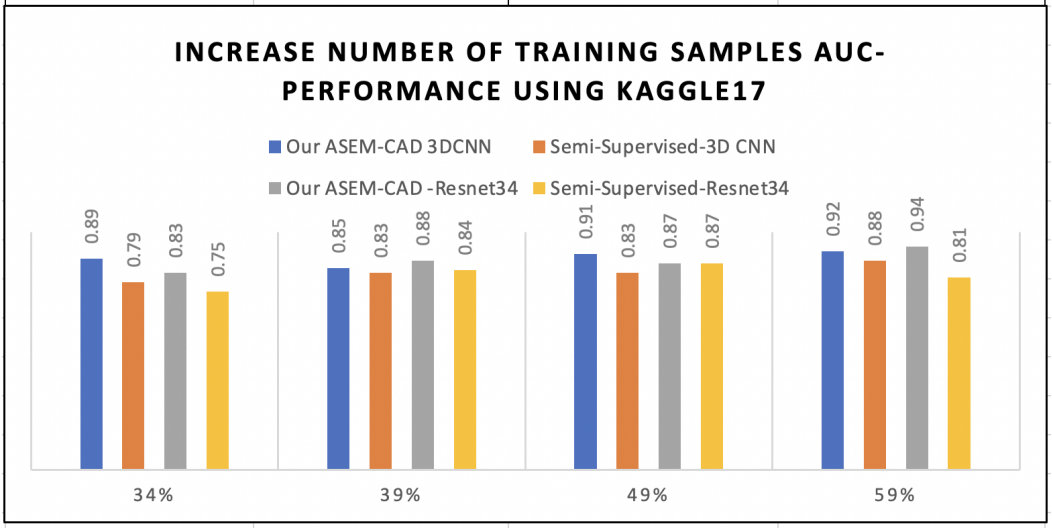


Figure 7. Comparison of algorithms with an increasing number of samples for training.

5.3. Experiments on the NLST Dataset

5.3.1. Active Phase Iterations Using 3D CNN

Figure 8 shows the ROC curve for active phase iterations using the NLST dataset with 3D CNNs. The AUC values with 25 % initial labels for ASEM-CAD with high classification entropy, average classification, and the semisupervised model are nearly the same as those indicated in Figure 8a. Using all samplesand 53% of labels (50% initially) for training, we observe that AUC values for our ASEM-CAD model are very close (0.95) to those of a fully supervised model using 100% of the labels (0.97).

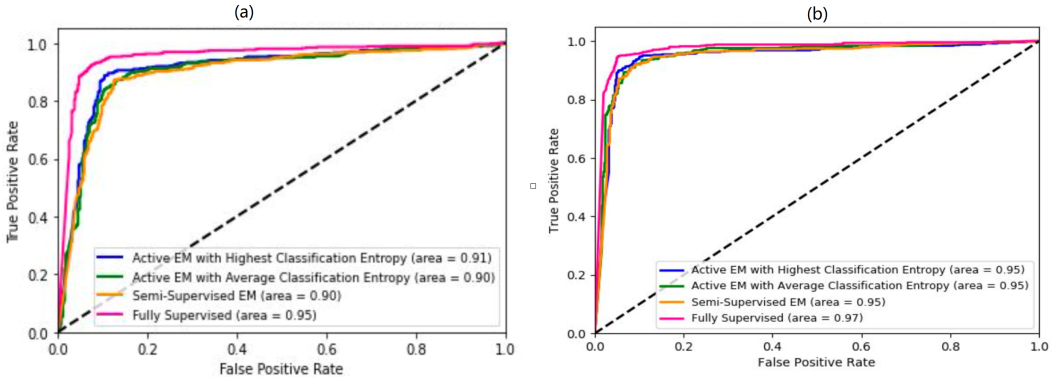


Figure 8. Active phase: 10 labels per iteration for the NLST dataset using a CNN with (**a**) 25% and (**b**) 50% of initial labels.

5.3.2. Active Phase Iterations Using Resnet34

The performance of active phase iterations with Resnet34 architecture is shown in Figure 9. With 25% of initial labels for ASEM-CAD, the fully supervised model outperformed all the configurations of ASEM-CAD, as shown in Figure 9a. However, with 50% of

initial labels, the AUC values for ASEM-CAD with high and average classification entropies (0.95) are as good as the fully supervised model (0.96), which uses 100% of the labels, as shown in Figure 9b.

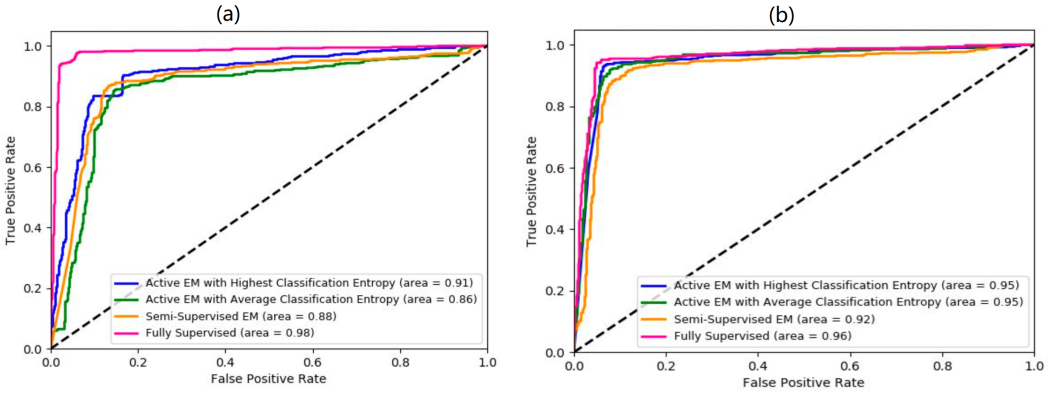


Figure 9. Active phase: 10 labels per iteration for the NLST dataset using ResNet34 with (**a**) 25% and (**b**) 50% of initial labels.

The ROC curves in Figures 8 and 9 show that the curve of our ASEM-CAD model with high classification is hugs the top left-hand side which, indicating a higher true positive rate and lower false positive rate compared to other algorithms, achieving almost equal performance to that of the fully supervised learning model.

5.4. Experiments on the LIDC-IDRI Dataset

ASEM-CAD was evaluated using supervised and semisupervised models for nodule malignancy classification using the LIDC-IDRI dataset. ASEM-CAD achieved an AUC of 0.81, which is comparable with the performance of a fully supervised model (with an AUC 0.82). The ROC curves for a fully supervised model and ASEM-CAD are shown in Figure 10. The ROC curves for both models indicate similar accuracy and performance characteristics. It is important to note that ASEM-CAD uses 52% of the data labels, as opposed to 100% of the labels in the fully supervised model. For the LIDC-IDRI dataset, ASEM-CAD achieves comparable AUC performance, and these algorithms outperformed those reported in our previous work on a semisupervised EM (SEM) algorithm [35]. We did not observe an improvement in training LIDC using 3D images of the nodule alone (image only) with a Bayesian experimental design via the Monte Carlo dropout method. To make an improvement, we incorporated additional input variables, known as image biomarkers or image characteristics, assigned by radiologists to the LIDC 3D images and trained the model to predict whether the nodule was malignant or benign. These additional variables include subtlety, internal structure, calcification, sphericity, margin, lobulation, speculation, and texture [80]. The values for these variables are scalar numbers. Subtlety is rated based on the contrast between the lung nodule and its surroundings, indicating the level of difficulty in nodule detection. Its possible ratings are: 1, extremely subtle; 2, moderately subtle; 3, fairly subtle; 4, moderately obvious; and 5, obvious. For more information on the methodology, please refer to our previous publication [34]. As a result, the model was able to achieve an AUC of 0.8822 and a sensitivity of 0.87, which represents an improvement over the use of 3D images only as input, with which an AUC of 0.81 and a sensitivity of 0.79 were obtained. Figure 10 shows the ROC curve.

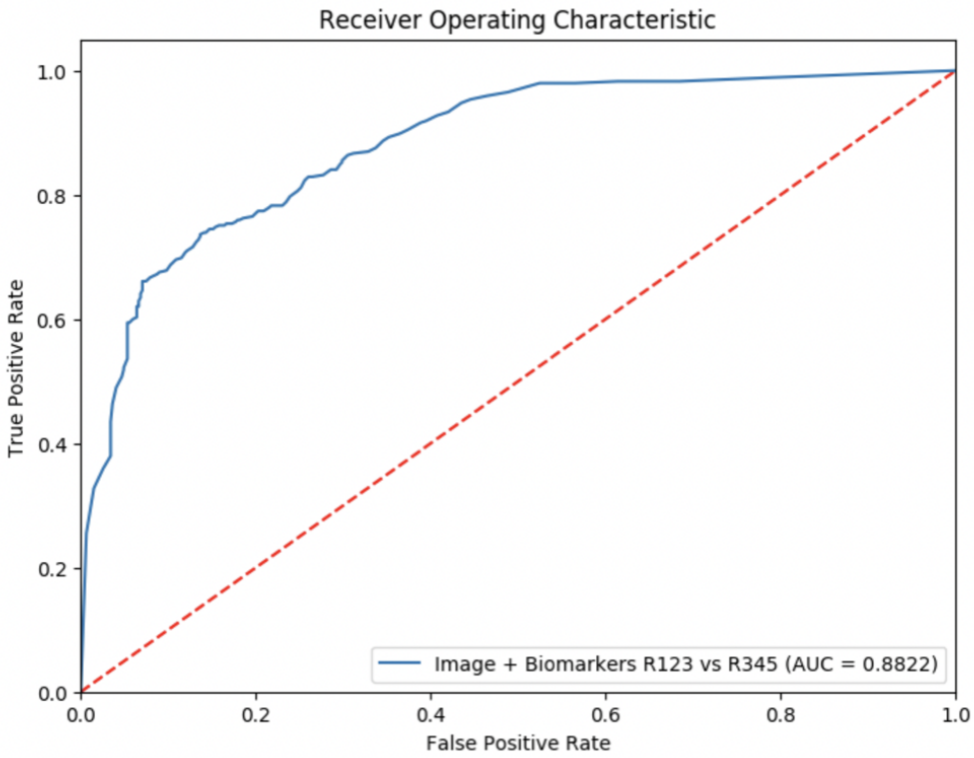


Figure 10. ROC curve using LIDC 3D images and image characteristics as input. The red dotted diagonal represents ROC curve for a random classifier.

6. Conclusions

Our study presents an active, semi-supervised algorithm called ASEM-CAD, which utilizes expectation maximization to classify lung cancer. We evaluated the performance of the algorithm on three publicly available datasets using two distinct neural network architectures. Using active learning strategies and label acquisition policies, we showed that ASEM-CAD using 52% to 59% of labels achieved a performance comparable to that of a fully supervised model (using 100% of labels). The ASEM-CAD model achieved high positive rates of 0.9, 0.88, and 0.87 (sensitivity) for the Kaggle, NLST, and LIDC datasets, respectively.

The active learning component of ASEM-CAD acquires additional ground truth labels during the EM training phase based on two label acquisition policies: high classification entropy and average classification entropy. The overall outcome of these two approaches proved to be better in terms of performance when compared to the semi-supervised model.

ASEM-CAD will be a useful tool for the enhancement of medical imaging research. This training algorithm can be integrated with annotation and evaluation tools. ASEM-CAD intelligently asks experts (human in the loop) for ground truth labels during training, which results in better prediction outcomes. It can also be used to continuously retrain the model over time to learn features from more data and identify changes in datasets. The changes may be the outcome of new CT scanners, adjustments to radiation doses, or changes in scanning procedures.

Moreover, commercial vendors can use the proposed methods to train and evaluate AI-based virtual radiology assistants that can accurately predict oncology imaging outcomes within the context of lung cancer screening and other diagnostic radiology examinations. Semi-supervised and active learning approaches have the potential to significantly enhance CAD algorithms by facilitating knowledge acquisition from large clinical PACS (picture archiving and communication system) datasets, thereby reducing the need for manual annotation by radiologists.

7. Future Directions

Although deep learning algorithms have tremendously enhanced medical image analysis, learning from few labeled samples while exploiting the vast majority of unlabeled data is a long-standing problem in the machine learning domain. It is worth exploring more powerful semi-supervised and self-supervised learning methods so that the novel methods may prove to be as promising as supervised models in the future. Self-supervised methods have gained popularity in the medical imaging field due to their ability to outperform supervised approaches on specific tasks, as reported in previous studies [82]. However, when implementing self-supervised frameworks in the medical domain, it is essential to address the challenge of data imbalance [83]. Another area of future research is to use self-supervised learning methods to extract data features by constructing negative examples by enhancing the contrastive learning frameworks to facilitate downstream tasks in medical imaging.

Author Contributions: Conceptualization, P.N. and A.R.; Methodology, P.N., A.R., D.C., S.M. and S.P.; Software, P.N. and A.R.; Validation, P.N. and A.R.; formal analysis, P.N. and A.R.; investigation, P.N. and D.C.; resources, D.C. and Y.Y.; data curation, S.M.; writing—original draft preparation, P.N., S.P., D.C. and A.R.; writing—review editing, M.M., Y.Y. and S.M.; Visualization, P.N., A.R. and S.P.; Supervision, P.N. and D.C.; Funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was sponsored by the NSF IUCRC Center for Accelerated Real Time Analytics (CARTA), (https://carta.umbc.edu/, https://carta.miami.edu/, accessed on 9 March 2023) (NSF grant award #1747724).

Data Availability Statement: National Lung Screening Trial (NLST): https://cdas.cancer.gov/nlst/, accessed on 9 March 2023; Kaggle 2017 lung cancer data: https://www.kaggle.com/competitions/data-science-bowl-2017/overview, accessed on 9 March 2023; the Lung Image Database Consortium image collection (LIDC): https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254, accessed on 9 March 2023.

Conflicts of Interest: This manuscript is not under consideration by another journal, nor has it been published. None of the authors have any competing financial interest.

References

1. Oudkerk, M.; Liu, S.; Heuvelmans, M.A.; Walter, J.E.; Field, J.K. Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives. *Nat. Rev. Clin. Oncol.* **2021**, *28*, 135–151. [CrossRef] [PubMed]
2. Chang, H.T.; Wang, P.H.; Chen, W.F.; Lin, C.J. Risk Assessment of Early Lung Cancer with LDCT and Health Examinations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4633. [CrossRef] [PubMed]
3. Ruan, J.; Meng, Y.; Zhao, F.; Gu, H.; He, L.; Gong, X. Development of deep learning-based automatic scan range setting model for lung cancer screening low-dose CT imaging. *Acad. Radiol.* **2022**, *29*, 1541–1551. [CrossRef] [PubMed]
4. Lee, H.Y. Time to Scrutinize and Revise the Fine Print of Lung Cancer Screening Using Low-Dose CT: Seeking Greater Confidence in Cancer Detectability. *Radiology* **2022**, *303*, 213084. [CrossRef] [PubMed]
5. American College of Radiology. Lung-Screening Reporting and Data System (Lung-RADS)®. 2022. Available online: https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/Lung-RADS-2022.pdf (accessed on 9 March 2023).
6. Yeh, M.C.H.; Wang, Y.H.; Yang, H.C.; Bai, K.J.; Wang, H.H.; Li, Y.C.J. Artificial Intelligence-Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach. *J. Med. Internet Res.* **2021**, *23*, e26256. [CrossRef]
7. Grenier, P.A.; Brun, A.L.; Mellot, F. The potential role of artificial intelligence in lung cancer screening using low-dose computed tomography. *Diagnostics* **2022**, *12*, 2435. [CrossRef]
8. Forte, G.C.; Altmayer, S.; Silva, R.F.; Stefani, M.T.; Libermann, L.L.; Cavion, C.C.; Youssef, A.; Forghani, R.; King, J.; Mohamed, T.-L.; et al. Deep Learning Algorithms for Diagnosis of Lung Cancer: A Systematic Review and Meta-Analysis. *Cancers* **2022**, *14*, 3856. [CrossRef]
9. Serena Low, W.C.; Chuah, J.H.; Tee, C.A.T.; Anis, S.; Shoaib, M.A.; Faisal, A.; Khalil, A.; Lai, K.W. An overview of deep learning techniques on chest X-ray and CT scan identification of COVID-19. *Comput. Math. Methods Med.* **2021**, 1–17. [CrossRef]
10. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
11. Bejnordi, B.E.; Veta, M.; Van Diest, P.J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **2017**, *318*, 2199–2210. [CrossRef]

12. Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; Beck, A.H. Deep learning for identifying metastatic breast cancer. *arXiv* **2016**, arXiv:1606.05718.
13. Jin, T.; Cui, H.; Zeng, S.; Wang, X. Learning deep spatial lung features by 3D convolutional neural network for early cancer detection. In Proceedings of the 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, Australia, 29 November–1 December 2017; IEEE: Piscataway, NJ, USA, 2017.
14. Xu, J.; Luo, X.; Wang, G.; Gilmore, H.; Madabhushi, A. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing* **2016**, *191*, 214–223. [CrossRef]
15. Hua, K.L.; Hsu, C.H.; Hidayati, S.C.; Cheng, W.H.; Chen, Y.J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Ther.* **2015**, *8*, 2015–2022.
16. Lakhani, P.; Sundaram, B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **2017**, *284*, 574–582. [CrossRef]
17. Setio, A.A.A.; Ciompi, F.; Litjens, G.; Gerke, P.; Jacobs, C.; Van Riel, S.J.; Wille, M.M.W.; Naqibullah, M.; Sánchez, C.I.; Van Ginneken, B. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 1160–1169. [CrossRef]
18. Valente, I.R.S.; Cortez, P.C.; Neto, E.C.; Soares, J.M.; de Albuquerque, V.H.C.; Tavares, J.M.R. Automatic 3D pulmonary nodule detection in CT images: A survey. *Comput. Methods Programs Biomed.* **2016**, *124*, 91–107. [CrossRef]
19. Zhou, Z.; Sodha, V.; Rahman Siddiquee, M.M.; Feng, R.; Tajbakhsh, N.; Gotway, M.B.; Liang, J. Models genesis: Generic autodidactic models for 3d medical image analysis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2019; pp. 384–393.
20. Wang, S.; Cao, S.; Wei, D.; Wang, R.; Ma, K.; Wang, L.; Meng, D.; Zheng, Y. LT-Net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9162–9171.
21. Zhao, A.; Balakrishnan, G.; Durand, F.; Guttag, J.V.; Dalca, A.V. Data augmentation using learned transformations for one-shot medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8543–8553.
22. Asgari Taghanaki, S.; Abhishek, K.; Cohen, J.P.; Cohen-Adad, J.; Hamarneh, G. Deep semantic segmentation of natural and medical images: A review. *Artif. Intell. Rev.* **2021**, *54*, 137–178. [CrossRef]
23. Aquila, I.; Sicilia, F.; Ricci, P.; Antonio Sacco, M.; Manno, M.; Gratteri, S. Role of post-mortem multi-slice computed tomography in the evaluation of single gunshot injuries. *Med. Leg. J.* **2019**, *87*, 204–210. [CrossRef]
24. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)—Based Software as a Medical Device (SaMD). Available online: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf (accessed on 9 March 2023).
25. Olivas, E.S.; Guerrero, J.D.M.; Martinez-Sober, M.; Magdalena-Benedito, J.R.; Serrano, L. (Eds.) *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2009.
26. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.lytera.de/Terahertz_THz_Spectroscopy.php?id=home (accessed on 5 June 2014).
27. Jing, L.; Tian, Y. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4037–4058. [CrossRef]
28. Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; Tang, J. Self-supervised learning: Generative or contrastive. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 857–876. [CrossRef]
29. Cheplygina, V.; de Bruijne, M.; Pluim, J.P. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* **2019**, *54*, 280–296. [CrossRef] [PubMed]
30. Wang, D.; Zhang, Y.; Zhang, K.; Wang, L. Focalmix: Semi-supervised learning for 3d medical image detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3951–3960.
31. Baier, L.; Schlör, T.; Schöffer, J.; Kühl, N. Detecting concept drift with neural network model uncertainty. *arXiv* **2021**, arXiv:2107.01873.
32. Santosh, K.C. AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. *J. Med. Syst.* **2020**, *44*, 93. [CrossRef] [PubMed]
33. Zhang, W.; Zhu, L.; Hallinan, J.; Zhang, S.; Makmur, A.; Cai, Q.; Ooi, B.C. Boostmis: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 20666–20676.
34. Nguyen, P.; Chapman, D.; Menon, S.; Morris, M.; Yesha, Y. Active semi-supervised expectation maximization learning for lung cancer detection from Computerized Tomography (CT) images with minimally label training data. In *Medical Imaging 2020: Computer-Aided Diagnosis*; International Society for Optics and Photonics: Bellingham, WA, USA, 2020; Volume 11314, p. 113142E.
35. Menon, S.; Chapman, D.; Nguyen, P.; Yesha, Y.; Morris, M.; Saboury, B. Deep expectation-maximization for semi-supervised lung cancer screening. *arXiv* **2020**, arXiv:2010.01173.
36. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Ontervention, Athens, Greece, 17–21 October 2016; Springer: Cham, Switzerland, 2016; pp. 424–432.

37. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
38. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2019**, *39*, 1856–1867. [CrossRef]
39. Zhang, J.; Xie, Y.; Wang, Y.; Xia, Y. Inter-slice context residual learning for 3D medical image segmentation. *IEEE Trans. Med. Imaging* **2020**, *40*, 661–672. [CrossRef]
40. Tajbakhsh, N.; Jeyaseelan, L.; Li, Q.; Chiang, J.N.; Wu, Z.; Ding, X. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Med. Image Anal.* **2020**, *63*, 101693. [CrossRef]
41. Azour, L.; Hu, Y.; Ko, J.P.; Chen, B.; Knoll, F.; Alpert, J.B.; Brusca-Augello, G.; Mason, D.M.; Wickstrom, M.L.; Kwon, Y.J.F.; et al. Deep Learning Denoising of Low-Dose Computed Tomography Chest Images: A Quantitative and Qualitative Image Analysis. *J. Comput. Assist. Tomogr.* **2023**, *10*, 1097. [CrossRef]
42. Li, Y.; Liu, J.; Yang, X.; Wang, A.; Zang, C.; Wang, L.; He, C.; Lin, L.; Qing, H.; Ren, J.; et al. An ordinal radiomic model to predict the differentiation grade of invasive non-mucinous pulmonary adenocarcinoma based on low-dose computed tomography in lung cancer screening. *Eur. Radiol.* **2023**, 1–11. [CrossRef]
43. Le, V.H.; Kha, Q.H.; Hung, T.N.K.; Le, N.Q.K. Risk score generated from CT-based radiomics signatures for overall survival prediction in non-small cell lung cancer. *Cancers* **2021**, *13*, 3616. [CrossRef]
44. van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [CrossRef]
45. Grandvalet, Y.; Bengio, Y. Semi-supervised learning by entropy minimization. *Adv. Neural Inf. Process. Syst.* **2004**, *17*.
46. Huang, J.-T.; Hasegawa-Johnson, M. Semi-supervised training of gaussian mixture models by conditional entropy minimization. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Chiba, Japan, 26–30 September 2010.
47. Vu, T.H.; Jain, H.; Bucher, M.; Cord, M.; Pérez, P. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 2517–2526.
48. Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N.E.; McGuinness, K. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 17–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8.
49. Ding, G.; Zhang, S.; Khan, S.; Tang, Z.; Zhang, J.; Porikli, F. Feature affinity-based pseudo labeling for semi-supervised person re-identification. *IEEE Trans. Multimed.* **2019**, *21*, 2891–2902. [CrossRef]
50. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on Challenges in Representation Learning*; ICML: Atlanta, GA, USA, 2013; Volume 3, p. 896.
51. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10687–10698.
52. Bachman, P.; Alsharif, O.; Precup, D. Learning with pseudo-ensembles. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, QC, Canada, 8–13 December 2014.
53. Dai, Z.; Yang, Z.; Yang, F.; Cohen, W.W.; Salakhutdinov, R.R. Good semi-supervised learning that requires a bad gan. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
54. Verma, V.; Kawaguchi, K.; Lamb, A.; Kannala, J.; Bengio, Y.; Lopez-Paz, D. Interpolation consistency training for semi-supervised learning. *arXiv* **2019**, arXiv:1903.03825.
55. Xie, Q.; Dai, Z.; Hovy, E.; Luong, T.; Le, Q. Unsupervised data augmentation for consistency training. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6256–6268.
56. Zhou, H.Y.; Wang, C.; Li, H.; Wang, G.; Zhang, S.; Li, W.; Yu, Y. SSMD: Semi-supervised medical image detection with adaptive consistency and heterogeneous perturbation. *Med. Image Anal.* **2021**, *72*, 102117. [CrossRef]
57. Gyawali, P.K.; Ghimire, S.; Bajracharya, P.; Li, Z.; Wang, L. Semi-supervised medical image classification with global latent mixing. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–9 October 2020; Springer: Cham, Switzerland, 2020; pp. 604–613.
58. Mahapatra, D.; Bozorgtabar, B.; Ge, Z. Medical image classification using generalized zero shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3344–3353.
59. Shang, H.; Sun, Z.; Yang, W.; Fu, X.; Zheng, H.; Chang, J.; Huang, J. Leveraging other datasets for medical imaging classification: evaluation of transfer, multi-task and semi-supervised learning. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: Cham, Switzerland, 2019; pp. 431–439.
60. Nie, D.; Gao, Y.; Wang, L.; Shen, D. ASDNet: attention based semi-supervised deep networks for medical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Cham, Switzerland, 2018; pp. 370–378.
61. Zhou, Y.; He, X.; Huang, L.; Liu, L.; Zhu, F.; Cui, S.; Shao, L. Collaborative learning of semi-supervised segmentation and classification for medical images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 2079–2088.

62. Njima, W.; Bazzi, A.; Chafii, M. DNN-based Indoor Localization Under Limited Dataset using GANs and Semi-Supervised Learning. *IEEE Access* **2022**, *10*, 69896–69909. [CrossRef]
63. Chen, K.M.; Chang, R.Y. Semi-supervised learning with GANs for device-free fingerprinting indoor localization. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
64. Mangalagiri, J.; Chapman, D.; Gangopadhyay, A.; Yesha, Y.; Galita, J.; Menon, S.; Yesha, Y.; Saboury, B.; Morris, M.; Nguyen, P. Toward Generating Synthetic CT Volumes using a 3D-Conditional Generative Adversarial Network. In Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 16–18 December 2020; IEEE: Piscataway, NJ, USA; pp. 858–862.
65. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22.
66. Papandreou, G.; Chen, L.C.; Murphy, K.P.; Yuille, A.L. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1742–1750.
67. Ebrahimi, S.; Elhoseiny, M.; Darrell, T.; Rohrbach, M. Uncertainty-guided continual learning with bayesian neural networks. *arXiv* **2019**, arXiv:1906.02425.
68. Sinha, S.; Ebrahimi, S.; Darrell, T. Variational adversarial active learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 5972–5981.
69. Lindenbaum, M.; Shaul, M.; Dmitry, R. Selective sampling for nearest neighbor classifiers. *Mach. Learn.* **2004**, *54*, 125–152. [CrossRef]
70. Mahapatra, D.; Bozorgtabar, B.; Thiran, J.P.; Reyes, M. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; Springer: Cham, Switzerland, 2018; pp. 580–588.
71. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York City, NY, USA, 19–24 June 2016; pp. 1050–1059.
72. Beluch, W.H.; Genewein, T.; Nürnberger, A.; Köhler, J.M. The power of ensembles for active learning in image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9368–9377.
73. Tran, T.; Do, T.T.; Reid, I.; Carneiro, G. Bayesian generative active deep learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6295–6304.
74. McKinley, R.; Meier, R.; Wiest, R. Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation. In *Proceedings of the International MICCAI Brainlesion Workshop*; Springer: Cham, Swizterland, 2018; pp. 456–465.
75. Xie, X.; Li, Y.; Shen, L. Active learning for breast cancer identification. *arXiv* **2018**, arXiv:1804.06670.
76. Gal, Y.; Islam, R.; Ghahramani, Z. Deep bayesian active learning with image data. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2018; pp. 1183–1192.
77. Hyams, G.; Greenfeld, D.; Bank, D. Improved training for self-training. 2017. Corr abs/1710.00209.
78. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **2011**, *365*, 395–409. [CrossRef] [PubMed]
79. Kareem, H.F.; Al-Huseiny, M.S.; Mohsen, F.Y.; Al-Yasriy, K. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. *Indones. J. Electr. Eng. Comput. Sci.* **2021**, *21*, 1731. [CrossRef]
80. 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.; et al. 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. [PubMed]
81. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
82. Nikroorezaei, F.; Esmaili, S.S. Application of Models based on Human Vision in Medical Image Processing: A Review Article. *Int. J. Image Graph. Signal Process. (IJIGSP)* **2019**, *11*, 23–28. [CrossRef]
83. Chen, S.; Ma, K.; Zheng, Y. Transfer learning for 3d medical image analysis. *arXiv* **1904**, arXiv:1904.00625.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.