IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 October 2023, accepted 7 December 2023, date of publication 15 December 2023,
date of current version 21 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343451

RESEARCH ARTICLE

Full-Resolution Lung Nodule Localization From Chest X-Ray Images Using Residual Encoder-Decoder Networks

MICHAEL J. HORRY[1], (Member, IEEE), SUBRATA CHAKRABORTY[1,2,3], (Senior Member, IEEE), BISWAJEET PRADHAN[1,4], (Senior Member, IEEE), MANORANJAN PAUL[5], (Senior Member, IEEE), JING ZHU[6], PRABAL DATTA BARUA[2,7,8], (Member, IEEE), HASAN SAEED MIR[9], (Senior Member, IEEE), FANG CHEN[10], (Senior Member, IEEE), JIANLONG ZHOU[10], (Senior Member, IEEE), AND U. RAJENDRA ACHARYA[11], (Senior Member, IEEE)

[1]Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2351, Australia
[2]School of Science and Technology, Faculty of Science, Agriculture, Business and Law, University of New England, Armidale, NSW 2007, Australia
[3]Griffith Business School, Griffith University, Nathan, Brisbane, QLD 4111, Australia
[4]Earth Observation Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor 43600, Malaysia
[5]Machine Vision and Digital Health (MaViDH), School of Computing, Mathematics, and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia
[6]Department of Radiology, Westmead Hospital, Westmead, NSW 2145, Australia
[7]Cogninet Brain Team, Cogninet Australia, Surry Hills, NSW 2010, Australia
[8]School of Business (Information Systems), Faculty of Business, Education, Law & Arts, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[9]Department of Electrical Engineering, American University of Sharjah, Sharjah, United Arab Emirates
[10]Data Science Institute, University of Technology Sydney, Ultimo, NSW 2351, Australia
[11]School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia

Corresponding author: Subrata Chakraborty (Subrata.chakraborty@une.edu.au)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Technology Sydney under Approval No. ETH21-6193.

ABSTRACT Lung cancer is the leading cause of cancer death, and early diagnosis is associated with a positive prognosis. Chest X-ray (CXR) provides an inexpensive imaging mode for lung cancer diagnosis. Computer vision algorithms have previously been proposed to assist human radiologists in this task; however, leading studies use down-sampled images and computationally expensive methods with unproven generalization. In contrast, this study localizes lung nodules from CXR images using efficient encoder-decoder neural networks that have been crafted to process full resolution input images, thereby avoiding signal loss resulting from down-sampling. Encoder-decoder networks are trained and tested using the Japanese Society of Radiological Technology dataset. The networks are used to localize lung nodules from an independent CXR dataset. These experiments allow for the determination of the optimal network depth, image resolution, and pre-processing pipeline for generalized lung nodule localization. We find that more subtle nodules are detected in earlier training epochs. Therefore, we propose a novel self-ensemble model from three consecutive epochs centered on the validation optimum. This ensemble achieved a sensitivity of 85% in 10-fold internal testing with false positives of 8 per image. A sensitivity of 81% is achieved at a false positive rate of 6 following morphological false positive reduction. This result is comparable to more computationally complex systems, but with a sub-second inference time that is faster than other methods presented in the literature. The proposed algorithm achieved excellent generalization results against a challenging external dataset with a sensitivity of 77% at a false positive rate of 7.6.

INDEX TERMS Chest X-ray, lung nodule, deep learning, segmentation, generalization.

The associate editor coordinating the review of this manuscript and approving it for publication was Manuel Rosa-Zurera.

I. INTRODUCTION

Lung cancer is the third most common cancer globally, and it is the most deadly, accounting for around 1 in 5 cancer deaths [1]. In the United States, lung cancer deaths have declined by around 50% since 1990 for males and by 30% from 2002 for females; this reduction is attributable to reductions in smoking and improvements in early detection and treatment [2]. Unfortunately, the United States' pattern of lung cancer decline is not global. Whilst lung cancer has declined for males in 19 countries, it has increased for women in 26 countries over the same period. Concerningly, although lung cancer is associated with aging populations [3], the risk among younger generations shows increasing trends in some geographies [4].

This paper focusses on the automatic localization of lung nodules, which are small roughly spherical growths on the lung, from Chest X-ray (CXR) images. The detection of lung nodules is important because they often represent early-stage lung cancer [5], and there is a directly proportional relationship between the size and growth rate of a nodule, and nodule malignancy [6]. On a typical workday, a clinical radiologist may be required to analyze a CXR image every 3-4 seconds [7]. Over a work shift of 8-10 hours, it can be expected that radiologists experience fatigue with a negative effect on the quality of diagnosis [8]. Automated tools that facilitate the tracking of lung nodules have been shown to help reduce lung cancer mortality, and associated public health costs [9], estimated to average around $51,000 per case in Australia [10] in 2020, and $45,897 per case in the US in the most recent comparable assessment in 2005 [11].

Segmentation of lung nodules from CXR images is a challenging task, as the complex background of pulmonary cavity anatomy makes it difficult to distinguish a true positive lung nodule from a false positive feature, such as a vascular bundle or rib crossing. The lung nodule itself appears as a relatively small feature against a large and noisy background. This differs from typical computer vision/classification tasks where the image feature of interest, say cat or dog, appears as a large foreground shape that dominates the overall image, as is the case for the frequently cited ImageNet archive [12]. This is also a more challenging computer vision problem than binary classification of interstitial lung diseases from CXR images as recently reviewed by [13] and frequently applied to Covid-19 detection studies [14], [15].

Studies into automated lung nodule detection from CXR images usually involve feature extraction tasks using parameterized linear and spatial filtering methods such as wavelet, convergence index, sliding-band, and Laplacian of Gaussian (LoG) [16], [17], [18], [19], [20]. Deep learning algorithms are then used for the classification of these features as nodule/non-nodule for false positive reduction. The filter parameters for these feature extraction methods are determined empirically and require retuning where datasets exhibit systematic quality differences [21]. Hence, the generalization properties of nodule detection systems based on LoG and other filtering techniques tend to be poor, even when applied to relatively high-resolution CT-based imaging [22]. Filtering-based feature extraction can also result in long inference times due to resource-intensive pixel-wise calculations. For example, [23] reported image processing times of up to 70 seconds per image. In a clinical context, where radiologists are required to interpret an image every few seconds [7], long inferencing times will adversely impact radiologists' workflow and hinder clinical adoption of automation tools.

Previous studies into nodule detection from CXR images using an end-to-end deep learning pipeline employ a sliding-window segmentation methodology. This involves creating a small patch around every pixel in a CXR image, using the nodule/non-nodule label for that image patch as the training data for an ensemble of CNNs. Such methods result in millions of image patches from a typical 2048×2048-pixel image and are resource intensive to train.

This study localizes lung nodules from CXR images using highly efficient encoder-decoder (E-D) neural networks. The E-D networks have been crafted to process full resolution input images to avoid signal loss from down-sampling. The resulting algorithm can localize lung nodules from CXR images in sub-second inference time with high sensitivity. To achieve this, we train a series of E-D neural networks based on U-Net [24], using a training set of 140 lung nodule examples with nodule masks generated from metadata provided by [25]. The E-D network is modified in several novel ways to achieve acceptable performance in this difficult task. The first innovation is a modification of the E-D filter sizes to allow network training using full-resolution CXR images. This eliminates any need to down-sample the CXR image, with the highly undesirable effect of weakening the nodule signal against the noisy CXR background. Since the E-D network is symmetrical, the decoded output is also full-resolution and provides a good indication of the location, size, and shape of predicted nodules. Secondly, following the work of [26], we improved the network recovery from difficult CXR images by using instance normalization [27] rather than batch normalization between layers in the E-D network. Thirdly, we observed that E-D network sensitivity varied by nodule subtlety and degree of network training. This problem is solved by combining the outputs of the E-D models trained to the validation optimum, along with one epoch lower and one epoch higher, thereby reinforcing nodule signal and diffusing noise.

The study also considers the influence of commonly employed pre-processing methods, such as global histogram equalization and lung field segmentation, showing that these operations are needed to achieve good generalization results. The effect of E-D network depth and image resolution are also explored. The resulting E-D algorithm yields excellent results that generalize well to an external dataset.

The novelty of this work can be attributed to three main aspects. Firstly, this is the first study in the literature to effectively utilize an efficient end-to-end E-D architecture for the generalized localization of lung nodules from full-resolution CXR images. This approach sets the study apart from previous works. Secondly, the proposed algorithm is externally tested using a publicly available dataset, which showcases the importance of lung field segmentation and histogram equalization as crucial pre-processing steps to achieve generalization. This aspect adds value to the study as it allows for replication by other researchers using code shared at https://github.com/mikejhorry/Lung-Nodule-Localization-from-CXR-using-Residual-E-D. Thirdly, this study conducts a comprehensive analysis of the proposed algorithm's performance in relation to nodule location, diagnosis, malignancy, and patient sex. This analysis provides insights into the strengths and weaknesses of the algorithm and contributes to a more complete understanding of its abilities.

II. RELATED WORK

The personal and public health costs of lung cancer have driven a comprehensive body of research into computer-based detection of lung cancer [9], [28], [29]. It is known that radiologists fail to detect lung cancer from CXR images in around 20% of cases [30], [31]. This is concerning because CXR is frequently used as a first-presentation diagnostic for lung cancer [32], and it is known that early detection of lung cancer dramatically improves patient outcomes [33]. The idea of using computer vision methods to assist radiologists in finding lung nodules on CXR images, thereby reducing this miss rate, is appealing and research into the use of computer vision methods to automate the detection of lung nodules has been a popular field of research from the first availability of economically priced computer systems in the 1970s and 1980s [34], [35], [36], [37].

Early studies made use of gray-level thresholding [38] and wavelet filtering [39], [40] techniques to distinguish lung nodules from other lung field tissues and structures. These methods were frequently used in combination with difference image techniques such as dual-energy subtraction [41], [42] to reduce the intensity of bony and vascular structures in the source image [38], [43], thereby enhancing the lung nodule signal. Although these early studies delivered promising lab-based results using small test datasets of homogenous CXR images, the thresholding and filtering methods employed are inherently sensitive to the systematic image contrast and pixel intensity differences that result from CXR acquisition parameters including radiation dose, acquisition geometry, detector performance, and post-processing [44]. Thresholding techniques do not penalize false positive, or false negative pixels which makes achieving a stable threshold extremely difficult [45]. Consequently, these methods, require fine-tuning of threshold values to account for image quality differences [46], which limits their generalization to real clinical settings.

More recent studies into lung nodule detection from CXR images make use of machine learning techniques (ML) to automatically find the best-fitting algorithm for nodule classification [47]. ML algorithms are trained on example images for the classes of interest, learning to differentiate diseased tissue from normal tissue using hundreds or thousands of sample images. Over the past decade, the introduction of widely available GPU computing has allowed for cost-effective training of multi-layer neutral networks, a process known as deep learning. Deep neural networks, especially the Convolutional Neural Network (CNN) architecture proposed by [48] and first adapted to computer vision applications by [49], have excelled at many computer vision tasks including automated medical image analysis [50]. Deep learning CNNs automatically combine feature extraction and feature classification into a single trainable algorithm [51], thereby eliminating hand-crafted feature extraction steps in the pipeline.

The work by [21], presents the first study to use an artificial neural network (ANN) for both nodule feature extraction and classification. The authors noted that ANNs have improved generalization properties, stemming from their ability to learn from example data, eliminating the need for highly tuned *a priori* rules. Their study used a source dataset of 60 Posterior-Anterior (PA) CXR images, with each displaying single or multiple lung nodules of size 8 – 20mm, along with 60 images containing simulated small nodule images. The results achieved 94% sensitivity at 5 false positives per image for real image examples, and sensitivity of 89% at 7 false positives per image for smaller simulated test cases thereby proving the potential of using a fully ANN-based pipeline for computer vision-based lung nodule detection, albeit against a small non-standard dataset with simulated disease positive cases.

The use of a standardized lung nodule database in the automated detection of lung nodules using an ANN was first described by [18]. This study used a multiscale ANN that identified candidate nodules through multiscale blob detection, leveraging Gabor wavelet and LoG filters. This candidate nodule feature space was fed into a simple ANN to classify blobs as either nodule or non-nodule. This system was trained and tested against the standard JSRT dataset [25] using a subset of 199 images with nodules in the range of 5-20mm. Under internal testing, this system achieved a sensitivity of 75% with 10.2 false positives per image. The authors also conducted external validation against a private dataset comprising 40 images with at least one nodule, achieving a sensitivity of 75% at a false positive rate of 8.9. This proved that ANN-based feature classification generalized well to an external dataset.

The problem of detecting very subtle lung nodules was addressed by [52] using thresholding based on edge-gradient values, followed by average radial gradient filtering to create a 35-dimensional feature space. This feature space was then classified using an ANN. Experiments were performed against a private dataset consisting of 1000 CXR images,

each containing at least one lung nodule, of which 73 were subtle. The subtle lung nodules of interest were in opaque areas of the CXR image, such as the mediastinum, cardiac and lung area below the diaphragm. The system achieved a sensitivity of 52.1% with 1.89 false positives per image. However, no external generalization study was performed.

Although ANN or CNN algorithms are typically used as a means of false positive reduction, [5] showed that support vector machine (SVM) was also a suitable machine learning classifier for this task. Their best SVM models achieved a sensitivity of 71% at 1.5 false positives per image, 78% at 2.5 false positives per image, 85% at 4 false positives per image. At the highest achieved sensitivity of 92% and 100% false positives per image were 7 and 8, respectively. These results were achieved using a pre-processing pipeline lung field segmentation, histogram equalization, and difference imaging. SVM alone resulted in 70% sensitivity at 16 false positives per image, which was unsatisfactory. Therefore, the authors calculated a set of 160 features using techniques such as shape/size analysis, grey level distribution, and multi-scale LoG, and applied a univariate Golub statistic to-tune the scheme and achieve the reported results.

A common theme in the above studies is the utilization of a very large feature vector, achieved through methods such as shape, intensity, texture, and gradient analysis for feature extraction. However, in cases where datasets are small, such as lung nodule datasets like JSRT with only 154 nodule positive examples, the number of features can be close to, or even larger than, the number of discrete nodules. This results in a lack of robustness, as the developed algorithms are easily overfitted to the small data corpus. This issue was addressed by [16], who proposed augmenting nineteen texture/gradient-based features with a feature describing the degree of nodule isolation, as early-stage lung cancer may be characterized by the presence of a solitary pulmonary nodule. AdaBoost [53] was used to extract a white-nodule likeness feature map. This scheme resulted in an algorithm with a sensitivity of 80% at two false positives per image and a sensitivity of 93% at 5 false positives per image against the JSRT dataset.

The leading study on the use of deep learning for the lung nodule segmentation task is [54]. This study employed sliding-window multi-resolution training of three different resolution CNNs with fully fused outputs. This method achieved state-of-the-art results against the JSRT-A dataset, with a sensitivity of 90% at a false positive rate of 0.2 per image, and a sensitivity of 100% achieved with a false positive rate of 0.4 per image. The results reported in this study are much better than those in other studies in the literature and, furthermore, significantly outperformed the JSRT human expert radiologist metrics reported in the source paper [25]. However, when tested against two external datasets, the sensitivity dropped by 55 - 60%. Despite the impressive internal testing results, the disappointing generalization capability was addressed in the study by fine-tuning the CNN to the external datasets.

A summary of works on automated lung nodule segmentation is shown in Table 1, along with the technique used for feature extraction and classification, preprocessing methods, and the results of external testing where available.

A clinically useful system must be both efficient and well-generalized. Thus, the key objective of this study is to find an efficient computer vision method/architecture that can locate lung nodules in CXR images, with algorithm generalization considered a higher priority than internal fitting. The main contributions of this work are as follows:

1) Establishing for the first time that an end-to-end deep learning-based E-D network can generally segment/localize lung nodules when trained on a small data corpus of full-resolution CXR images.
2) We show that the best results are achieved using full-resolution images as input images to the E-D network. This result calls into question the very common practice of down-sampling images to fit out-of-the-box CNN input dimensions. We show that addition of convolution layers to the E-D network is a a better strategy for coping with large images than down-sampling.
3) Systematically showing that the intuitive, but unproven, image pre-processing steps of lung field segmentation and histogram equalization drastically improve algorithm generalization.
4) In our assessment, we take the novel step of investigating the limits of the proposed algorithm in relation to:

- Nodule's subtlety rating,
- Nodule's malignancy,
- Nodule's location,
- Patient's diagnosis, and
- Patient's sex.

III. METHOD

A. STUDY OVERVIEW

The study makes use of the frequently cited JSRT-A dataset detailed in section 3.3.1, as a training dataset for three implementations of a residual E-D-based segmentation algorithm [56], based on U-Net, initially described by [24] and commonly employed for biomedical segmentation tasks. Three different depths of E-D algorithms are tested at 5, 6, and 7 layers for each encoder/decoder branch. The training dataset is divided by preprocessing options, using histogram equalization, lung field segmentation, and a combination of these.

Each algorithm and pre-processing option are tested using three CXR image resolutions: 512 × 512 pixels, 1024 × 1024 pixels, and 2048 × 2048 pixels, with the latter being the full resolution of the JSRT images and typical of the resolution of DICOM images captured in a clinical context. Each training dataset is split into 10 equally sized groups to facilitate 10-fold cross-validation. Splits across training datasets for different pre-processing options and CXR resolutions are made up of identical images to ensure that observed performance differences are the result of the

TABLE 1. Summary of works relating to automated lung nodule segmentation using the JSRT dataset.

Reference	Training Dataset	External Dataset	Image Size (pixels)	Pre-processing	Feature Extraction	Classifier	Internal Sensitivity (%)	Internal FP Rate	External Sensitivity (%)	External FP Rate	Notes
[18]	JSRT Subset (N=119)	University of Florence private dataset. (N=40)	128 x 128	Gabor Filter Lung Field Segmentation	Multiscale LoG	ANN	75.0	10.2	75	8.9	Excluded nodules < 5mm and > 20 mm.
[5]	JSRT (all) (N=247)	240 Niguarda Ca'Granda Hospital in Milan (N=240)	256 x 256	Gaussian filtered image differencing	Multiscale circle finding and LoG	SVM	100.0	8.5	95	N/A	Total of 160 calculated features External testing false positive rate was not specified.
[55]	JSRT (N=154)	None	256 x 256	Gaussian filtered image differencing	Multiscale LoG Ray casting	kNN	67.0	4	N/A	N/A	Noted that lung field segmentation had no positive effect.
[17]	JSRT-A (N=140)	None	512 x 512	Local contrast enhancement Lung field segmentation	Convergence Index Filter	Fisher's Linear Discriminant (FLD)	80.0	5	N/A	N/A	Used an adjacency rule to eliminate false positives resulting in approx. 10% improvement to sensitivity at FP rate.
[23]	JSRT-A (N=140)	None	Not Specified	Lung field segmentation Morphological operations and	Clustering watershed.	SVM	78.6	5	N/A	N/A	Used an adjacency rule to eliminate false positives. Inference times of 70 seconds.
[16]	JSRT-A (N=140)	Beijing Capital Intl Airport Hospital (N=15)	1024 x 1024	Lung field segmentation Rib suppression	Wavelet Transform + Convergence Index Filter	SVM	90.0	4	100	4	Used an adjacency rule to eliminate false positives.
[54]	JSRT-A (N=140)	Guangzhou Hospital (GDH) (N=168) Shenzen Hospital (SZH)l (N=240)	Patches: 224 x 224 168 x 168 112 x 112	Lung field segmentation Rib Suppression Histogram Equalization	Multi-Resolution patch-based CNN.	Late-fusion multi-resolution CNNs	99.0	< 0.2	45 (GDH) 50 (SZH)	1 1	Pixel-level classification is resource intensive. SOTA internal testing results, but 55 – 60% decrease in AUC for external testing.

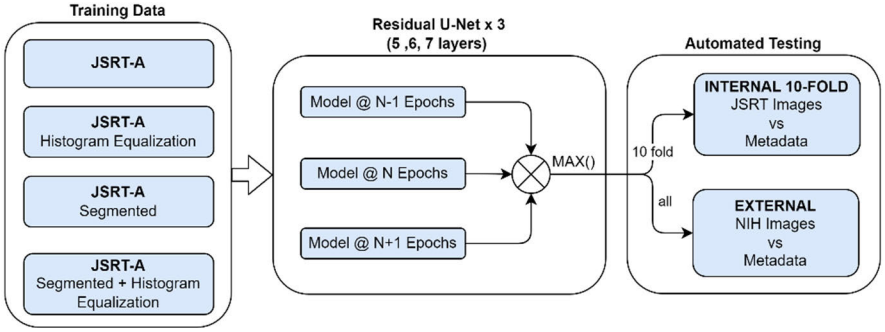


FIGURE 1. Summary of experiments performed in this study.

controlled testing variable rather than an artefact of dataset splitting. External testing is performed against a curated version of the NIH ChestX-ray14 dataset [57] since this dataset provides location and size metadata for a subset of nodule images. This sequence of experiments is illustrated in Fig. 1.

B. THEORY AND CALCULATION

1) MODEL COMPOSITION AND TRAINING

Each E-D variation is trained for each resolution and pre-processing option, with model weights captured every epoch. It was noticed during validation that lung nodules at different locations, and with different subtlety ratings, were located by the algorithm at different training epochs, offset from the point at which validation metrics stabilized before overfitting. Instead of there being a single optimal training epoch, the validation optimum is a saddle-point spread over three epochs. Each of the tested algorithms was therefore composed as an ensemble of three trained E-D models combined using simple element-wise maxima functions over the output arrays of each component model. The highest pixel value from each model output was taken as the final pixel value in a composite lung nodule prediction mask.

Fig. 2 provides an example of an E-D training curve showing no consistent reduction in validation loss following epoch 12. This would tend to indicate that a model trained to epoch 12 would be optimal.

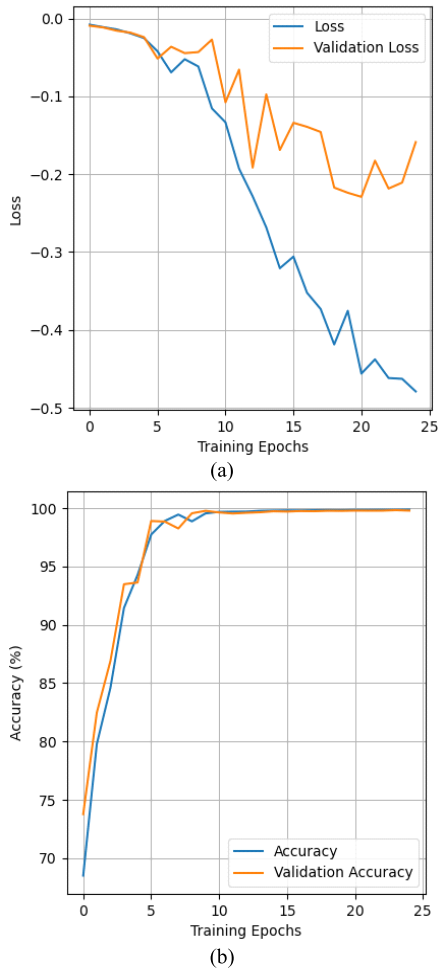


FIGURE 2. Training Curve example for E-D Model. Learning rate was set as 1e-5 and E-D models were trained to 25 epochs. Training Epochs vs Loss (a) and Training Epochs vs Accuracy (b).

The obvious nodule in file JPCLN005 (Fig. 3) is cleanly located with good segmentation at epoch 12, with epoch 11 being noisy, and epoch 13 still good, but starting to erode the nodule detection boundary.

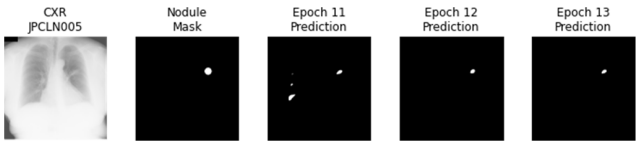


FIGURE 3. Example of obvious lung nodule correctly located at optimal training epoch.

However, more subtle nodules are located earlier in the model training. JPCLN146 is described as extremely subtle. As illustrated in Fig. 4, this nodule is small, and completely missed by the algorithm at epoch 12 onwards. The nodule is, however, visibly segmented in the prediction at epoch 11, along with several false positives. Since false positives are preferable to false negatives, epoch 11 should be included in the overall predicted nodule mask for this image.

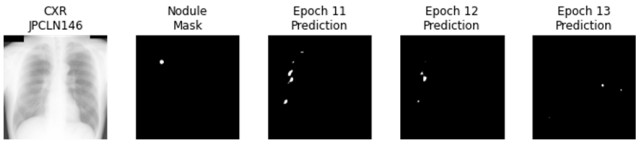


FIGURE 4. Extremely subtle lung nodules are located by earlier epoch in the model training.

Therefore, the proposed segmentation algorithm uses a simple combination of three trained models as described in Equation 1:

$$M' = \max(M(n-1), M(n), M(n+1)) \quad (1)$$

where M^′ is the composition of model output arrays at optimal training epoch n, determined by training validation statistics. The output nodule mask prediction from M^′ is therefore the element-wise maxima of predicted output arrays from the optimally trained model and the model trained to one epoch on either side of that trained model. Combining the algorithm output in this manner reinforces the true positives, with a relative de-emphasis of noise and false positives, resulting in cleaner nodule location prediction masks.

2) AUTOMATED NODULE RATING

Algorithmic performance rating by manual visual comparison is time-consuming and inherently subjective. Instead, this study uses a fully automatic framework for nodule localization rating, using a blob detection algorithm derived from the OpenCV [58] contour finding function, followed by ellipse fitting to detect blob-like activations on predicted nodule masks. Fitted ellipses were checked for eccentricity being less than 0.95 and the area of the ellipse is greater than the total number of pixels in the mask (square of the dimension) divided by the square of tolerance of 128 pixels. For a full-sized image of 2048 pixels, this translates to a minimum area of 256 pixels, or approximately 3mm diameter anatomically, since the JSRT pixel size is 0.175 mm. Detected blobs outside of these parameters were considered noise artefacts from the decoding process and excluded from automated rating since the smallest nodule detectable on CXR images is around 10mm [59], and nodules found/confirmed using CT of size smaller than 6mm are considered to be a micro-nodules that do not require further investigation [60].

A region of interest (ROI) was then calculated using the ellipse center co-ordinates and the minor axis length. If the average pixel intensity within this ROI was less than 0.5, then the detection was filtered out since the ensemble output max function would reinforce a true positive lung nodule signal. These checks filtered out noise in the predicted nodule mask image, leaving suitable nodule candidates for checking against the JSRT metadata for centroid correlation. If the centroid co-ordinates of a fitted ellipse fell within a tolerance of $1/16^{th}$ image dimension tolerance to the JSRT metadata nodule center location, then that detection was logged as a true positive. Other detections were logged as false positives. For full-resolution 2048 × 2048-pixel images, this translates to

a 128-pixel tolerance, or approximately 22.4 mm anatomical centroid difference threshold for a detection to be considered a true positive, which is consistent with other studies using 20 – 25 mm tolerance [17], [54], but computationally neater for the purposes of a fair comparison of different image resolutions. This process is illustrated in the flowchart in Fig. 5.

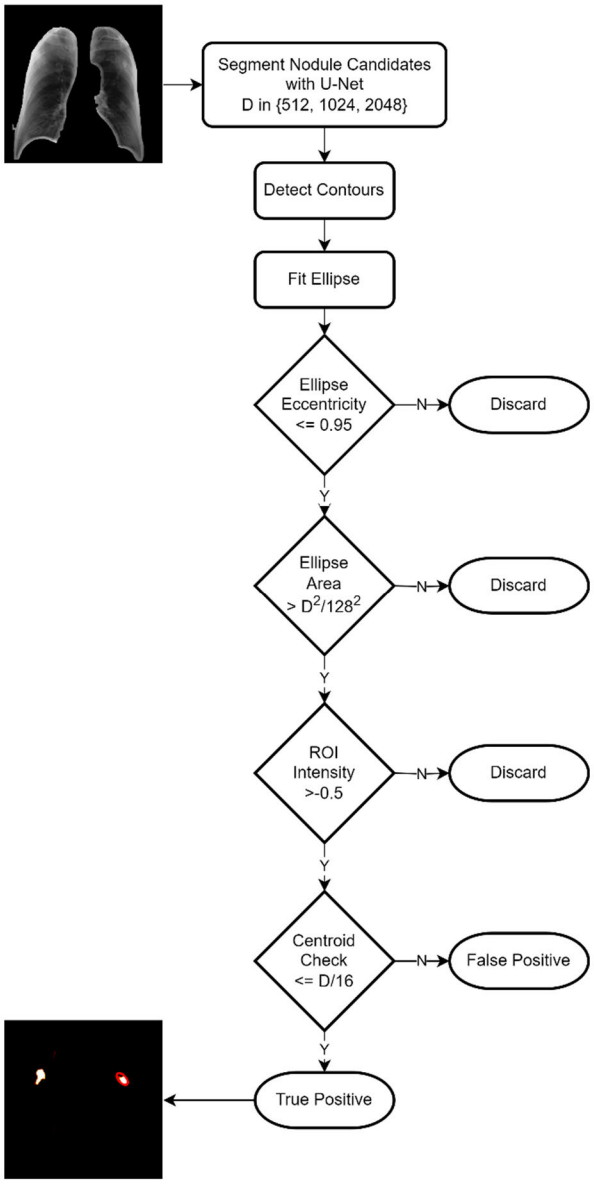


FIGURE 5. Automated nodule localization flowchart.

3) NETWORK DESIGN

Each experiment was performed using a residual U-Net, which is an E-D network with a ResNet backbone [61]. We have previously used this architecture to match state-of-the-art results in the lung field segmentation task [62], as part of a pre-processing pipeline for COVID-19 severity assessment.

The largest input convolution filter in a standard U-Net E-D network is 256 × 256 pixels in size. While this filter size is suitable for many applications, it may miss important features when dealing with full-resolution images of 2048 × 2048 pixels, where nodule features occupy only a small part of the image. To address this issue, we modified the familiar residual U-Net architecture by implementing additional higher dimensional input and output convolution filters.

Three networks were coded using TensorFlow [63] at different depths, consisting of 5, 6, and 7 layers, designated E-D5, E-D6, and E-D7, respectively (Table 2). E-D5 follows a standard residual U-Net design. In E-D6 and E-D7, we increased the depth of the E-D network by further convolving the input image, doubling the number of image channels, and halving the representational image dimensions at each additional layer. This resulted in quadrupling the number of trainable parameters for each added layer in the network.

TABLE 2. Encoder-decoder network summary.

Model	Filter Dimensions	Image Dim at Full Encoding	Trainable Parameters
E-D5	[16, 32, 64, 128, 256]	256 x 32 x 32	4,715,441
E-D6	[16, 32, 64, 128, 256, 512]	512 x 64 x 64	18,882,481
E-D7	[16, 32, 64, 128, 256, 512, 1024]	1024 x 32 x 32	75,528,113

Fig. 6 illustrates the structure of the deepest network used in this study E-D7.

4) INSTANCE NORMALIZATION VS BATCH NORMALIZATION

Instance normalization normalizes features over a single training image sample using the mean and variance calculated over the pixel array for that image [27]. In the context of CXR image, lung nodules appear as very weak signals, which can be easily confused with rib crossings, vascular structures, and other thoracic anatomical features. Since the nodule feature occupies only a tiny proportion of the overall image, this segmentation task is quite different from the tasks for which U-Net is normally used. Table 3 shows a comparison of internal test results at full image resolution for each E-D network with batch normalization and instance normalization for the experiment using histogram equalization and lung field segmentation, with the normalization method as a controlled variable. The results from batch normalization were poor in comparison to instance normalization. Therefore, instance normalization was used for the remainder of the experiments in this study.

TABLE 3. Comparison of instance normalization vs. batch normalization.

Model	Instance Normalization		Batch Normalization	
	Sen (%)	FP	Sen (%)	FP
E-D7	81.4	6.6	2. 9	0.5
E-D6	85.0	7.9	37. 9	6.4
E-D5	61.4	8.4	45.0	9.5

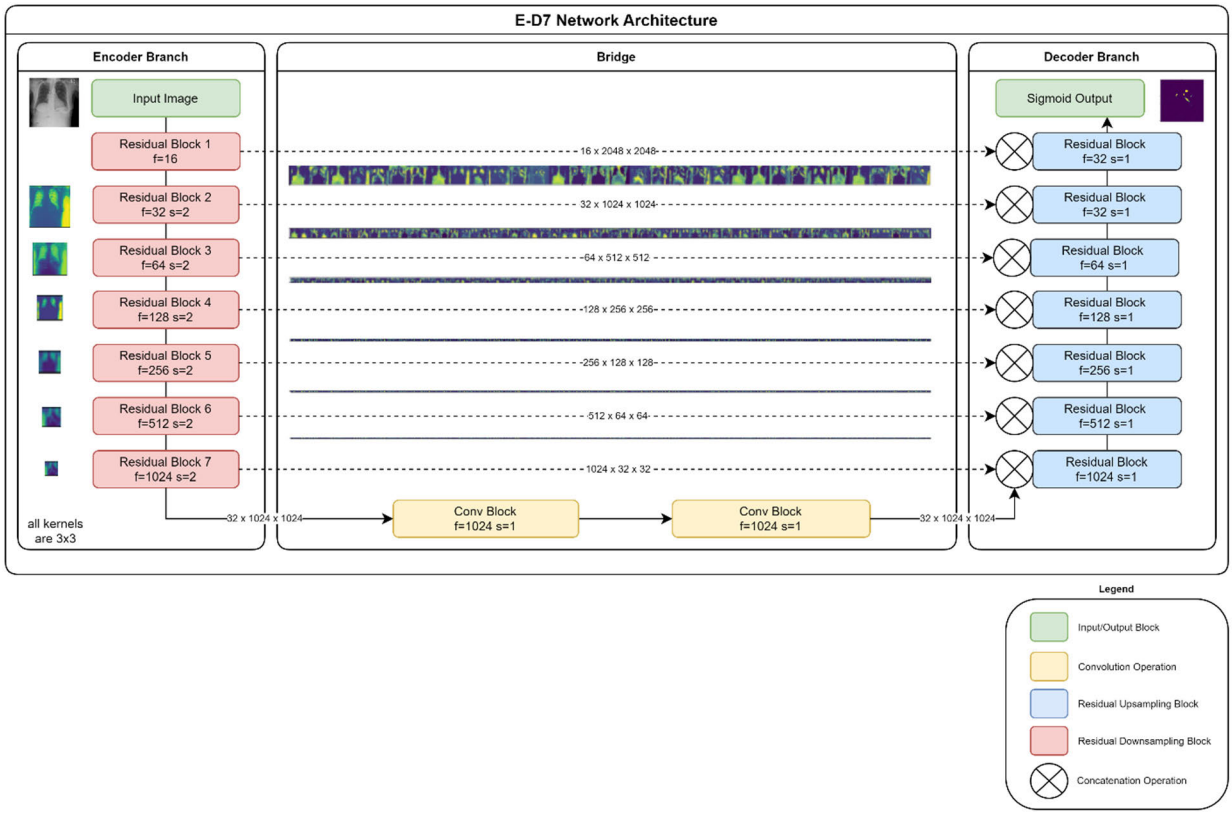


FIGURE 6. Schematic overview of the E-D7 Network.

IV. DATA SETS

Two independent datasets are used in this study. The JSRT dataset [25], consisting of a small, yet high-quality corpus of CXR images featuring solitary pulmonary nodules, serves as the training and internal testing dataset. For external testing, the NIH Chest X-ray14 dataset [57] is employed, providing a challenging and clinically realistic subset of CXR images featuring located nodules. Testing against this external dataset provides valuable insights into the real-world applicability and performance of the proposed algorithm.

A. TRAINING AND INTERNAL TESTING DATASET

Development studies involving the automated identification of lung nodules from CXR images consistently utilize the publicly available JSRT dataset [25]. This dataset comprises 93 control CXR images without lung nodules and 154 CXR images featuring solitary lung nodules, ranging from obvious to extremely subtle. Each of the 154 CXR images featuring solitary lung nodules is accompanied by metadata that describes the patient's age and gender demographics, the malignancy of the lung nodule, the final diagnosis, and, crucially for this study, the spatial co-ordinates and size of the lung nodule.

The JSRT image corpus is relatively small for deep learning segmentation applications, which typically require training on hundreds to thousands of images [64]. The difficulty in lung nodule segmentation arises from the fact that solitary lung nodules occupy only a tiny portion of the overall CXR image, unlike organ segmentation tasks where the organs of interest cover a significant area of the medical image, or biomedical cell segmentation where numerous cells are present for U-Net training. Limited availability of disease-positive training data can be problematic for lung nodule localization studies since ML and deep learning systems tend to overfit on small datasets. This means that the deep neural network may memorize the training dataset features without generalizing well to external datasets.

To maintain consistency with the literature, we excluded 14 images from the JSRT nodule corpus where the identified nodules were located outside the lung-field, resulting in a total of 140 disease-positive images. This created the JSRT-A dataset, as used in previous studies [16], [17], [54], [65], [66]. This exclusion was necessary for our experiment, as the generated lung nodule masks for these 14 images would be completely black after segmentation, invalidating comparisons between segmented experiments and unsegmented experiments. The set of JSRT-A images used in this study is presented in Fig. 7, showing (a) the original CXR images, (b) the effect of global histogram equalization, (c) lung-field segmented images, and (d) nodule mask for each image generated from JSRT metadata.

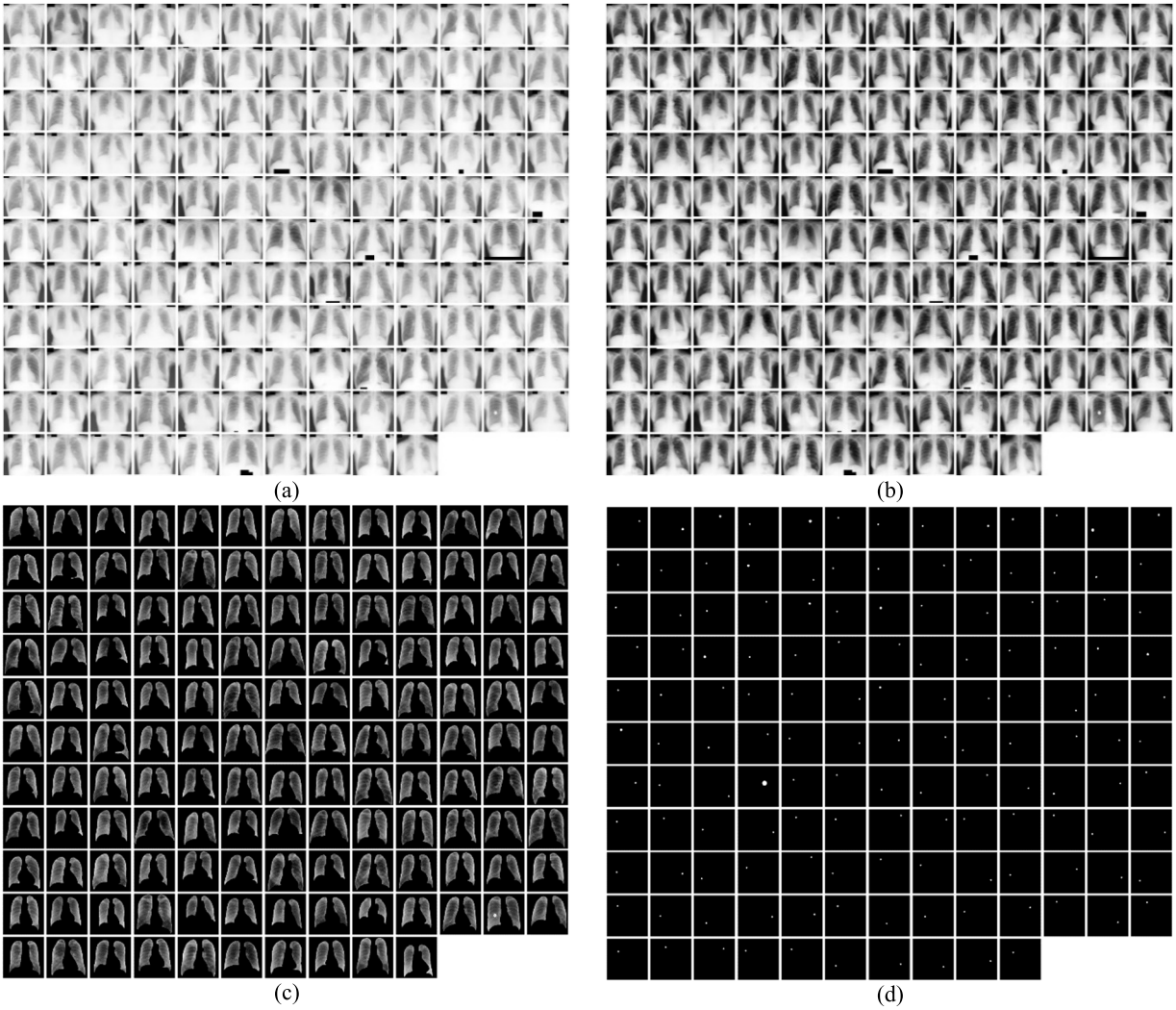


FIGURE 7. JSRT-A dataset summary montage showing: (a) base images; (b) the effect of histogram equalization; (c) the effect of lung-field segmentation using gold-standard masks; and (d) lung nodule locations and size using a circle approximation.

B. EXTERNAL TESTING DATASET

A secondary CXR dataset was required for external generalization testing, with the condition that nodules were identified on the CXR images and location metadata available. The publicly available NIH Chest X-ray14 dataset (NIH) [57] dataset consists of 112,120 frontal views CXR images of 30,805 unique patients, labelled with fourteen thoracic diseases extracted from accompanying radiology reports. The NIH dataset provides bounding box metadata for 984 of these images, of which 79 are classified as 'Nodule', making them suitable candidate images for this study. However, we observed that some images contained multiple nodules and other diseases, making segmentation difficult due to heavy disease obfuscation of the lung-field boundary. Therefore, these images were discarded, leaving 60 usable images for automated external generalization testing, as shown in Fig. 8(a). For each of these images, we generated nodule masks using the NIH boundary box metadata. The center of the circular nodule mask was set to match the center of the boundary box, and the diameter of the nodule mask was calculated from the maxima of width/height of the boundary box, resulting in a set of NIH lung nodule masks as shown in Fig. 8(b).

C. PREPROCESSING – HISTOGRAM EQUALIZATION

Global histogram equalization (HE) is a widely used contrast enhancement technique applied to medical images prior to computer vision analysis. HE is a simple transformer that redistributes the most frequent pixel intensity values across the entire image. This process enhances contrast, making it easier for the human eye to distinguish medical image features [67], [68], [69], [70]. In our previous work [71], we found that combining HE with other techniques such as lung field segmentation and rib suppression improved the generalization of deep learning models. This improvement is attributed to HE's ability to eliminate systematic differences in brightness and contrast between image acquisition sites. Without this correction, the model could become biased

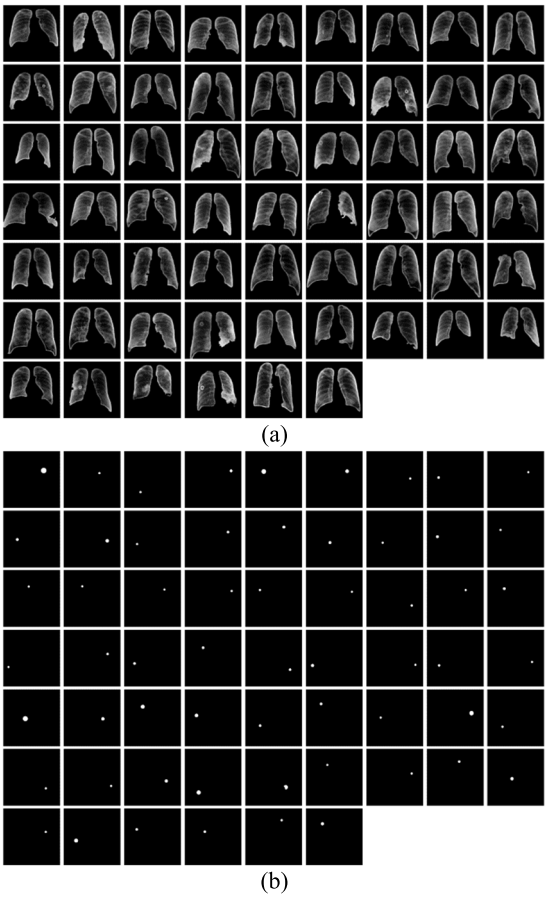


FIGURE 8. NIH image dataset summary montage showing: (a) segmented; and (b) lung nodule mask locations and size using a circle approximation.

towards specific image acquisition parameters used during training, resulting in overly optimistic model performance metrics from internal testing.

D. PREPROCESSING—LUNG-FIELD SEGMENTATION

Since lung nodules appear within the lung-field, it is logical to remove any pixels outside of the lung-field to reduce noise in the CXR image and improve the computer vision algorithm's ability to learn lung nodule features. In the JSRT dataset, there is a set of gold-standard lung field masks prepared by board-certified radiologists. These masks were used to create lung-field segmented versions of the JSRT-A dataset.

V. RESULTS

A. INTERNAL TESTING AT 512 X 512 PIXELS

Table 4 summarizes the internal testing results at a resolution of 512 × 512 pixels. Only one test at 512 × 512 yielded sensitivity results above the 80% threshold, with the highest sensitivity of 81.4% at 7.3 false positive detections per image. The best results at this resolution were obtained using the shallowest 5-layer E-D, indicating possible undertraining of the deeper networks on the smaller image size. For this model, lung field segmentation and HE were effective in improving sensitivity and reducing false positive detections, with the combination of these methods yielding the best results. This combination resulted in a 10% improvement in sensitivity with 2.5 fewer false positives compared to training on the raw image set.

TABLE 4. Results from 10-fold internal testing of each ensemble at 512 × 512 pixels.

Model	Raw Image		HE		Segmentation		HE + Segmentation	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	64.3	6.3	64.3	7.0	60.0	6.8	65.7	6.5
E-D6	75.0	8.7	67.1	7.6	63.6	4.6	72.9	6.8
E-D5	71.4	9.8	77.1	9.3	75.7	7.1	81.4	7.3

B. INTERNAL TESTING AT 1024 X 1024 PIXELS

Table 5 summarizes the internal testing results at a resolution of 1024 × 1024 pixels. At this resolution, two E-D network ensembles achieved sensitivity over the threshold of 80%. The best result was obtained from the 6-layer E-D6 network, with a sensitivity of 83.8% at a false positive rate of 8.3 detections per image. The combination of lung field segmentation and HE improved sensitivity by over 10%. However, this improvement was at the expense of a small increase in FP detections contributed by the HE operator.

TABLE 5. Results from 10-fold internal testing of each ensemble at 1024 × 1024 pixels.

Model	Raw Image		HE		Segmentation		HE + Segmentation	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	70.0	7.2	68.6	7.5	69.3	6.7	69.3	6.7
E-D6	72.9	8.1	77.9	8.6	77.1	6.3	83.6	8.3
E-D5	70.0	9.3	70.0	9.7	81.4	8.3	82.9	8.5

C. INTERNAL TESTING AT 2048 X 2048 PIXELS (FULL RESOLUTION)

Table 6 provides a summary of internal testing results at the full resolution of 2048 × 2048 pixels. At this resolution, two of the E-D network ensembles achieved sensitivity over the threshold of 80%. Using HE and lung field segmentation, E-D6 achieved a sensitivity of 85% at 7.9 false positives per image. The deeper network, E-D7, also met the sensitivity criteria, with 81.4 at a false positive rate of 6.6. This result

TABLE 6. Results from 10-fold internal testing of each ensemble at 2048 × 2048 pixels.

Model	Raw Image		HE		Segmentation		HE + Segmentation	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	65.0	7.9	78.6	8.4	75.0	6.2	81.4	6.6
E-D6	69.3	9.6	72.9	8.7	82.1	7.4	85.0	7.9
E-D5	57.9	9.1	61.4	9.7	78.6	10.2	61.4	8.4

represents the lowest false positive rate for any of the experiments exceeding 80% sensitivity.

D. OVERALL COMPARISON OF THE ENCODER-DECODER APPROACH TO JSRT RADIOLOGISTS

The results of the comparison between the best experiment at each input resolution and the JSRT Radiologist's performance are presented in Table 7. The best E-D model sensitivity was comparable to, or better than, the JSRT radiologists. However, apart from the obvious nodule case, the standard deviation across the 10-folds was higher than the human radiologists. This indicates that the sensitivity of the 10-fold models was influenced by training/testing data splits, with some splits being more challenging than others.

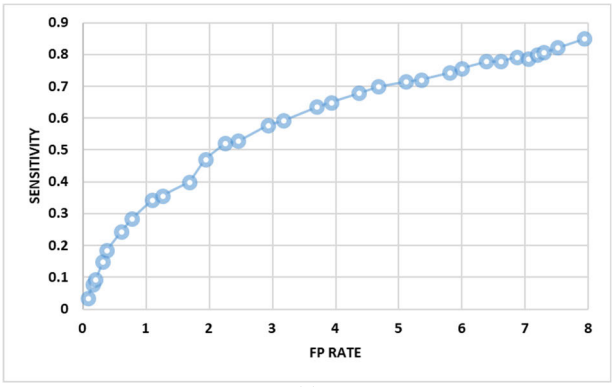
At full resolution, the best-performing E-D model is more sensitive to subtle nodules with comparable standard deviation to the JSRT radiologists. For obvious and relatively obvious categories of nodules, lower-resolution E-D models outperformed the full-resolution model. This may be a result of the down-sampling ''smoothing'' effect noted by [17], which assists with the localization of obvious categories of nodules. However, the same smoothing effect negatively impacts sensitivity to the subtle categories of nodules, resulting in overall better classification metrics using full-resolution images. This is an important result since the clinical utility of nodule detection algorithms would be to assist radiologists in the localization of subtle rather than obvious nodules.

TABLE 7. Breakdown of results from 10-fold internal testing with HE and segmentation compared with JSRT radiologists.

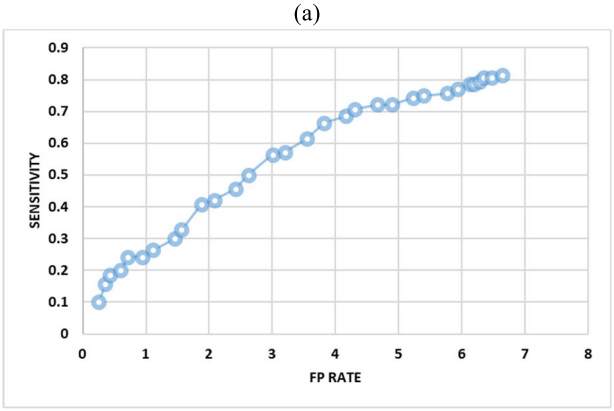
Subtlety	Radiologist Sen (%)		512 x 512 Sen (%)		1024 x 1024 Sen (%)		2048 x 2048 Sen (%)	
	M	SD	M	SD	M	SD	M	SD
Obvious	99.6	1.9	100.0	0	100.0	0	91.7	28.9
Rel. Obvious	92.6	13.1	92.1	27.3	92.1	27.3	81.6	39.3
Subtle	75.7	22.4	83.3	37.7	87.5	33.4	91.7	27.9
V. subtle	54.7	24.0	78.3	42.2	65.2	48.7	73.9	44.9
E. subtle	29.6	15.9	47.4	51.3	68.4	47.8	57.9	50.7

E. RECEIVER OPERATING CHARACTERISTIC FOR FULL-RESOLUTION INPUT IMAGES

In comparison to more traditional feature extraction techniques such as LoG, the E-D architectures used in this study have few parameters that can be modified to trade-off sensitivity against false positive rates. Therefore, to better understand the limitations of the proposed algorithm this study employed a simple method of morphological opening followed by morphological closing using an elliptical kernel



(a)



(b)

FIGURE 9. Comparison of ROC curves for best results at full resolution (2048 × 2048-pixel) input images for E-D6 (a) and E-D7 (b).

starting at 3 pixels and increasing to 90 pixels with increments of 3 pixels before rating output images from E-D6 and E-D7 at full resolution. An elliptical kernel was chosen over a rectangular kernel since detections are rounded rather than angular, matching the actual geometry of lung nodules. This resulted in the ROC curves in Fig. 9(a) and Fig. 9(b), respectively. E-D6 shows a sensitivity of around 80% at a false positive rate of 7. E-D7 achieves a sensitivity of 80% at a false positive rate of around 6. Sensitivity for E-D7 falls off more rapidly than E-D6 so at a false positive rate of 3, E-D6 has a superior sensitivity of around 60% compared to E-D7 at 56%. This is a result of the deeper E-D7 providing cleaner and more accurate initial nodule localization, but with weaker detections that are more rapidly eroded by morphological operations. Nevertheless, E-D7 achieves an overall best result of 81% sensitivity at a false positive rate of 6.4 using a morphological operation kernel of size 6 pixels.

F. PREDICTED NODULE MASK EXAMPLES

Predicted nodule masks from 10-fold testing are presented in Fig. 10, with a representative example from each subtlety rating.

The E-D model sensitivity to nodules is inversely related to the subtlety of the nodules, and FP detections increase as the nodule subtlety increases. We observed rib and collar-bone crossings often caused false positive detections when inspecting the generated nodule masks. The use of rib and bone

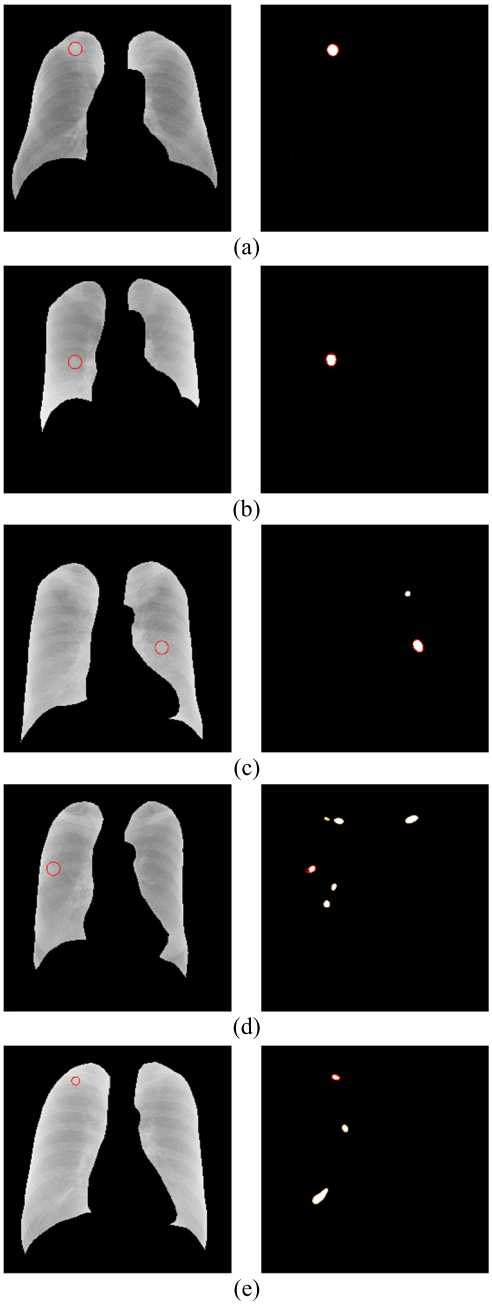


FIGURE 10. Examples of nodule localization predictions from 10-fold testing with various degrees of subtlety. TP is circled in red. (a) obvious, (b) rel. obvious, (c) subtle, (d) v. subtle, (e) e. subtle.

suppression techniques, such as dual energy CXR image acquisition, could be expected to improve the sensitivity of this algorithm and reduce false positive rates.

G. EFFECT OF VARIABLES ON MODEL SENSITIVITY AND FP RATE

1) SENSITIVITY AND FP RATE BY NODULE SUBTLETY AND MALIGNANCY

We were interested in whether the E-D architectures would exhibit differential sensitivity to malignant versus benign nodules, which could potentially enable this technique to assess the severity of detected nodules. Table 8 presents the results of E-D7 using HE and segmented full-resolution images, separated by malignancy. The algorithm demonstrated higher sensitivity to obvious malignant nodules compared to obvious benign nodules. Conversely, it showed higher sensitivity to extremely subtle benign nodules than to malignant nodules. This is a result of benign nodules tending to have a higher degree of calcification resulting in higher visibility on CXR images than subtle malignant nodules [72]. When averaged across 10 folds, sensitivity to benign nodules was 79.6% at 6.7 false positives per image, and sensitivity to malignant Nodules was 82.4% at 6.6 false positives per image.

TABLE 8. Results by the degree of subtlety and malignancy for E-D7 using HE and segmented images at full resolution.

Degree of Subtlety	Malignant			Benign		
	Sen (%)	FP	N	Sen (%)	FP	(N)
Obvious	100.0	5.9	7	80.0	4.6	5
Rel. Obvious	85.7	6.8	28	70.0	7.0	10
Subtle	90.9	6.8	33	93.3	5.7	15
V. subtle	76.9	5.7	13	70.0	8.8	10
E. subtle	40.0	7.5	10	77.8	6.7	9
Mean	82.4	6.6	91	79.6	6.7	49

2) SENSITIVITY AND FP RATE BY NODULE LOCATION

The JSRT dataset metadata describes the anatomical location of each nodule detection by side (left or right) and lung/lobe. The sensitivity of the E-D7 model using full-resolution images is mapped against these locations in Fig. 11. Sensitivity is relatively symmetrical across the left and right lobes, with the result for the right middle lobe being around 10% higher than for the left middle lobe. The most likely explanation for this result is that the heart and associated cardiological shadow is located slightly to the left-hand side of the body, thereby obscuring the visibility of nodules on this side. Overall, the algorithm is much better at localizing nodules in the middle and lower lobes than in the upper lobes, regardless of whether the nodule is located on the left- or right-hand side.

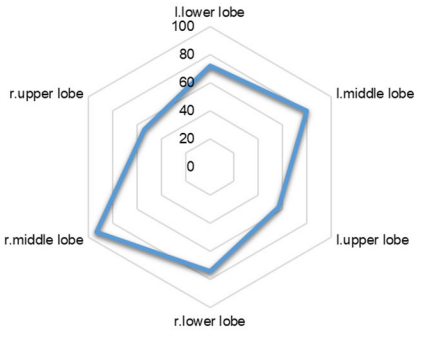


FIGURE 11. Sensitivity by nodule location for the E-D7 algorithm.

3) SENSITIVITY AND FP RATE BY SEX

We were interested in evaluating the performance of the nodule localization algorithm based on the sex of the patients, both before and after applying the image enhancements of HE and lung field segmentation. The sensitivity and FP rate results for each full-resolution test using E-D7 are presented in Table 9. Initial results using raw images showed a significantly higher sensitivity for female over male CXR images, the application of HE and lung field segmentation significantly reduces this difference by debiasing the image samples.

TABLE 9. Sensitivity and FP rate by sex at full resolution using E-D7.

	Raw Image		HE		Segmentation		HE + Segmentation	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
Female	67.9	7.9	78.2	8.3	75.6	5.9	82.1	7.1
Male	61.3	8.1	79.0	8.4	74.2	6.4	80.6	6.0

4) SENSITIVITY AND FP RATE BY DIAGNOSIS

Fig. 12 shows the sensitivity of E-D7 using HE and segmented full-resolution images for each diagnosis in the JSRT Metadata. The algorithm is sensitive to most diagnoses at or above the 80% objective. There are nine diagnoses, excluding the "unknown" category, for which the algorithm provided lower than 80% sensitivity. Table 10 provides examples of these difficult images to highlight the limitations of the proposed algorithm. Most of the missed detections relate to subtle categories of nodules that are obfuscated by ribs or are examples of conditions which are dissimilar to typical solitary pulmonary nodules associated with lung cancer, such as pneumonia and cryptococcosis.

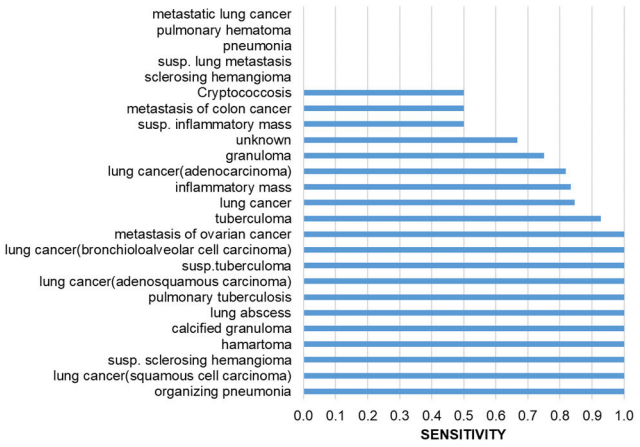
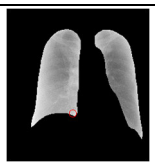
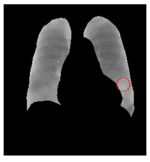
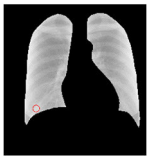
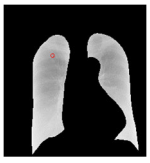
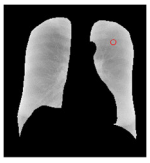
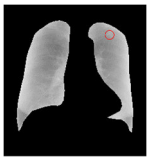
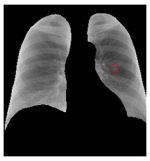
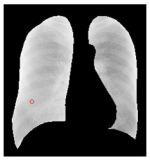
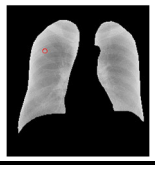


FIGURE 12. Sensitivity by diagnosis for the E-D7 algorithm.

TABLE 10. Diagnoses/examples for which the E-D7 algorithm showed poor performance.

Diagnosis	Subtlety	Example Image	Notes
Metastatic lung cancer	E. Subtle		Overlaps the inner edge of the right lower lobe.
Pulmonary hematoma	Obvious		Large mass occupying lower right lobe. Not a typical nodule.
Pneumonia	V. Subtle		Obfuscated by rib shadow at the lower extremity of the right lung.
Susp. Lung metastasis	E. Subtle		An extremely subtle nodule obfuscated by a rib.
Sclerosing hemangioma	E. Subtle		An extremely subtle nodule obfuscated by a rib crossing.
Cryptococcosis	Rel. Obvious		Appears as an outline rather than a filled circle.
Metastasis of colon cancer	Subtle		Nodule appearance is very poorly defined.
Susp. inflammatory mass	V. Subtle		Overexposed part of the image. Very small dimensions for a suspected mass.
Granuloma	Subtle		Obfuscated by rib shadow.

5) EXTERNAL TESTING

Other studies have also achieved excellent results against the JSRT dataset, sometimes with results surpassing label confidence by a large margin [54]. However, these studies rarely subject JSRT-trained models to external testing. This study externally tested models trained on the JSRT-A dataset against the NIH dataset.

a: EXTERNAL TESTING FOR EACH MODEL AT VARIOUS RESOLUTIONS

The first external test was conducted using models trained at each resolution using JSRT-A as a training corpus. The results of this test are presented in Table 11. Although the NIH metadata describes a single nodule per CXR image, many of the NIH images contain more than one nodule. The presence of multiple lung nodules on the NIH images has the effect of increasing the false positive count since some of the nominal false positives are true positives that have not been logged in the NIH metadata. Medical devices were another source of false positive detections, however, since such devices are pervasive in the clinical context, it is important to understand if the proposed algorithm was resilient to these. Illustrative examples are provided in Fig. 13.

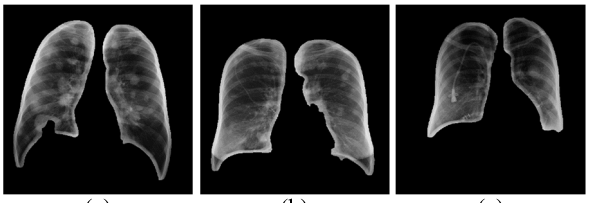


FIGURE 13. NIH Examples with (a, b) multiple pulmonary nodules; and (c) intrusive medical devices.

The best result was obtained at 1024 × 1024 pixels, achieving a sensitivity of 78.3% at 13.2 false positives per image. Recalling that at 1024 × 1024 pixels internal 10-fold testing for JSRT-A using E-D5 achieved a sensitivity of 82.9% at 8.5 false positives per image, this represents a sensitivity penalty of only 4.6% with an increase in false positive rate of 4.6 per image. Bearing in mind that the false positive detections for NIH are artificially high due to the presence of multiple pulmonary nodules, we feel confident to conclude that the E-D5 variant of the proposed system has successfully generalized to the difficult and clinically realistic NIH dataset.

TABLE 11. External testing results.

Model	512 x 512		1024 x 1024		2048 x 2048	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	50.0	16.4	60.0	9.9	70.0	13.4
E-D6	58.3	14.1	68.3	9.0	76.7	13.3
E-D5	78.3	16.1	78.3	13.2	78.3	15.0

Across these experiments, the best-performing E-D architecture is the shallowest 5-layer version. This is most likely a result of the deeper E-D6 and E-D7 networks having learned features of the JSRT training dataset that are specific to the JSRT dataset, resulting in network memorization and overfitting. The E-D6 and E-D7 networks would likely perform better if a larger and more diverse training corpus were available, such as could be obtained using a federated implementation. Upscaling the 1024 × 1024-pixel NIH images to 2048 × 2048 pixels did not improve sensitivity and increased false positives (due to noise introduced by interpolation) by 1.8 per image when using the E-D5 model.

Lung nodule localization examples resulting from external testing against the NIH dataset are presented in Fig. 14. Predictions made by E-D7 (c) provide a very clean segmentation of the metadata located in nodule (b), although missing some of the less obvious nodules that are apparent in the source image (a). These are better detected by the shallower networks E-D6 (d), and E-D5 (e). Fig. 14(f) shows an example of an intrusive medical device on the right lung. This medical device proved to be a source of false positive detections, as all E-D models have incorrectly segmented this device as a nodule candidate (h-j). In practice, a human observer would easily dismiss these detections as obvious false positives, highlighting the importance of an expert overview of medical vision AI algorithms.

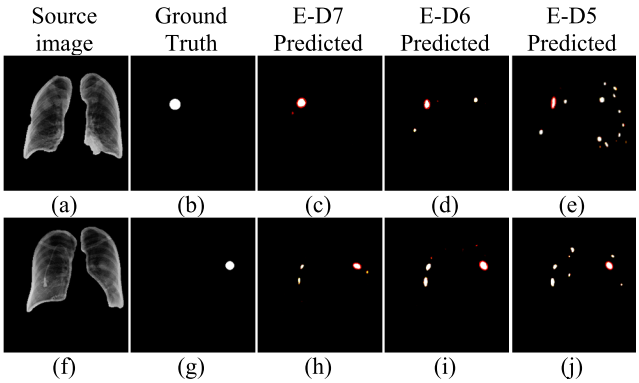


FIGURE 14. External testing using the NIH dataset examples showing: (c-e) sensitivity to nodules and false positive detections due to medical devices (h-j).

b: EXTERNAL TESTING WITH INCLUSIONS/EXCLUSIONS OF NODULE SUBTLETY CATEGORIES

Noting that internal testing performance for the "extremely-subtle" and "very-subtle" categories of nodules were relatively weak, we sought to understand which subtlety categories should be included in model training to facilitate the best external testing results. We hypothesized that nodule masks for the "very-subtle" and "extremely-subtle" categories resulted in nodule masks that were little more than pinpoints. Along with the documented high inter-observer variability/disagreement in the data source metadata, this let us to suspect that these categories would have little bearing on the model performance and may even be detrimental using dilution of model learning. The JSRT-A training dataset was therefore split into four categories A through D by subtlety as described in table 12. E-D models were then trained on each of these splits, resulting in four models at each resolution trained on all nodule images through to only obvious nodule images. These models were then tested against the NIH dataset to establish the contribution of subtle categories of nodules to generalized model sensitivity.

TABLE 12. External testing categories definition.

Category	Obvious	Rel. Obvious	Subtle	V. Subtle	E. Subtle	(N)
A	✓	✓	✓	✓	✓	140
B	✓	✓	✓	✓	×	121
C	✓	✓	✓	×	×	98
D	✓	✓	×	×	×	50

TABLE 13. Nodule categories external testing at 512 × 512 pixels.

Model	A		B		C		D	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	50.0	16.4	51.7	12.6	46.7	11.9	45.0	8.5
E-D6	58.3	14.1	58.3	10.9	78.3	14.4	63.3	9.0
E-D5	78.3	16.1	71.7	10.7	75.0	9.7	70.0	12.5

Table 13 provides the results of this test at a resolution of 512 × 512 pixels. The best result obtained at this resolution used the E-D6 model with very subtle and extremely subtle categories of nodules excluded. Best results were obtained for category C, excluding subtle and extremely subtle nodules, and using E-D6.

Table 14 shows the results of this testing at the NIH native resolution of 1024 × 1024 pixels. Best results were achieved using E-D5 with all categories of nodules included in the training data (A). Notably, excluding extremely subtle nodules per category B had little bearing on these results indicating that these images did not provide a significant training signal. Moreover, when the other E-D architectures are considered, the exclusion of very-subtle and extremely subtle nodules per category C led to a good result for E-D6 with a sensitivity of 76.7 with FP of 7.6, with this being the lowest FP value achieved for any external test. This result corresponds to the good result at 512 × 512 pixels for E-D6 with these categories of nodules excluded.

TABLE 14. Nodule categories external testing at 1024 × 1024 pixels.

Model	A		B		C		D	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	60.0	9.9	65.0	11.7	61.7	14.7	60.0	13.7
E-D6	68.3	9.0	71.7	6. 7	76.7	7.6	73.3	11.0
E-D5	78.3	13.2	78.3	13.4	71.7	10.1	71.7	11.2

Finally, Table 15 presents the result of this round of testing at 2048 × 2048 pixels. Once again, the best result was obtained using category C data, which excludes the very subtle and extremely subtle categories of nodules confirming that the inclusion of these categories of nodules in model training has a negative impact on external generalization.

VI. DISCUSSION

A. VISUALIZATION

Many development studies present computer vision algorithms capable of diagnosing thoracic diseases from CXR images. Most of these studies solve the problem in either binary or multiclass form – giving a computer vision-based probability of a certain disease being present in the lung including the original NIH paper [57] and follow-up studies, with state-of-the-art methods achieving multi-class accuracy for nodule detection of 77.7% [73]. These studies often use saliency techniques such as Grad-CAM [74] to visualize network attention as a proxy for disease location. Saliency techniques tend to have a relatively poor definition, whilst an E-D architecture can generate an image mask that cleanly segments the lung nodule, as shown in Fig. 15.

TABLE 15. Nodule categories external testing at 2048 × 2048 pixels.

Model	A		B		C		D	
	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP	Sen (%)	FP
E-D7	70.0	13.4	71. 7	10.6	71. 7	8.8	66. 7	8.3
E-D6	76.7	13.3	75.0	14.9	70.0	16.1	78.3	12.6
E-D5	78.3	15.0	76. 7	16.5	85.0	16.4	78.3	14.3

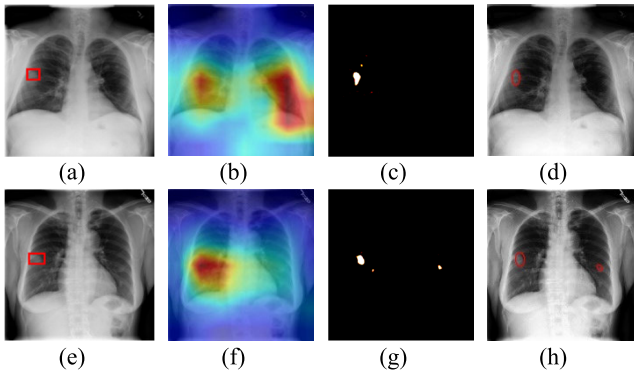


FIGURE 15. Example of a nodule located using saliency technique compared to the E-D method. (a, e) ground truth using NIH provided boundary box, (b, f) Saliency method using DenseNet-121, (c, g) Predicted Nodule Mask using E-D6, (d, h) Nodule location after FP reduction and ellipse fitting.

B. REPLICABILITY

Other studies have successfully located lung nodules in the JSRT dataset, achieving good results. However, these studies have often performed relatively poorly during external testing when the model is applied to datasets not used during training, especially without fine-tuning to the external dataset [54]. Where external testing is performed, the external datasets used are typically a private set of images without indication of clinical realism and/or curation processes, making replication impossible [5], [16], [18], [54]. The tendency for AIs to worsen when applied to patient groups with different characteristics to model development groups is known as ''data set shift'' and has recently been called out as a major problem by [75], who propose evaluating an AI's effectiveness based on real-world clinical evaluation rather than stand-alone model performance, noting that this narrow focus has led to limited penetration of radiologic AI applications in practice.

C. EFFICIENCY

Although our internal testing results may not match those achieved by [54], which exceed label confidence by a considerable margin, the proposed method offers several advantages. One significant benefit is the relatively fast training time, taking approximately 20 minutes per epoch at full resolution and resulting in an overall training time of around 8 hours on the modest hardware used. Additionally, the inference time for a full-resolution CXR measuring 2048 × 2048 pixels is only 768ms, and 198ms at the NIH resolution of 1024 × 1024 pixels, which would easily fit into an augmented clinical workflow. This is much faster than inference times achieved by studies using complex linear filtering methods based around LoG, where inference times of up to 70s per image have been reported [23]. The output of the proposed algorithm is a nodule mask that could be overlaid on the source CXR image to produce an instantaneous visual cue to human radiologists to pay close attention to parts of the CXR image containing nodules.

D. POSITIVE EFFECTS OF INSTANCE NORMALIZATION, IMAGE RESOLUTION, AND PRE-PROCESSING OPERATIONS

This study has demonstrated that an E-D deep learning-based architecture can effectively localize lung nodules when trained on a small data corpus of full-resolution CXR images. Accurate lung nodule segmentation is greatly promoted by substituting instance normalization for batch normalization at the E-D network layer boundaries. Instance normalization is not typically used in CXR deep learning applications, our use is novel, and this may explain why the use of E-D architectures such as U-Net has not been previously reported for lung nodule segmentation tasks for this imaging mode. In our testing, good internal testing results translated well to external testing with only a minimal drop in performance.

Higher resolution images yielded better results in our internal testing, with 2048 × 2048-pixel images achieving a top sensitivity of 85% with 7.9 false positive detections per image. This result is improved to 81% sensitivity at 6 false positive detections per image using morphological false positive reduction. These results are comparable to more computationally complex systems based on linear/spatial filtering using nodule adjacency rules for false positive reduction as summarized in Table 1. The use of HE and lung-field segmentation had a significant positive effect on sensitivity and false positive rate under 10-fold testing. This result was found to be due to an improvement in sensitivity to the subtle categories of lung nodules. Use of full-resolution CXR images avoids the common practice of down-sampling to CNN default image size 299 × 299 pixels, or less, for training. It is apparent from our testing that important pathological signals are weakened by the down-sampling process. This conclusion is supported by external testing achieving the best results at a resolution of 1024 × 1024 pixels, which is the native resolution of the external NIH dataset. Upscaling NIH images to 2048 × 2048 pixels did not improve sensitivity, and the interpolation noise resulting from upscaling caused an increase in the false positive rate for this test.

E. EXTERNAL TESTING AND MODEL GENERALIZATION

In external testing, E-D6 provided an excellent generalization result when very subtle and extremely subtle nodules were excluded from the JSRT training data. Internally, at 1024 × 1024 pixels, E-D6 achieved a sensitivity of 83.6% with 8.3 false positives per image. Externally E-D6 achieved a sensitivity of 76.7% at 7.6 false positives per image. This result shows that the E-D6 algorithm can generalize well across datasets from different clinical origins. This is a promising result to create a federated implementation whereby separate clinics can contribute to a centralized model using local data, whilst preserving patient confidentiality – since only network weights are shared beyond the wall of the clinic. Ideally, several independent clinics could contribute to a centralized model that would learn sensitivity to more subtle nodules via access to a globally diverse and large training data corpus.

F. LIMITATIONS

It is important to understand the limitations of any tool, especially when it comes to applications in human medical diagnosis. AI tools used for CXR computer vision are intended to be used under the supervision of expert radiologists, who will easily note Obvious and Relatively Obvious categories of nodules without automated assistance. The study found that the JSRT expert radiologists outperformed the proposed algorithm for these categories. The situation is different for subtle categories of nodules, where the algorithm outperformed the JSRT radiologists on the sensitivity measure. These results indicate that the clinical usefulness of the algorithm in a clinical setting would be to draw radiologist attention to potentially missed subtle to extremely subtle nodules. One very important limitation to this use case is relatively poor sensitivity to malignant extremely subtle nodules, which was found to be 40%, compared to 78% for benign extremely subtle nodules. Unfortunately, these are the most desirable nodules to find early – since they are likely to grow into life-threatening cancers. We also noted that the algorithm was more sensitive to nodules in the middle and lower lung lobes than the upper lung lobes, most likely because of the presence of scapular and rib crossings mimicking nodules in these geographies. The team is currently working on a full-resolution rib and bone suppression system which is expected to mitigate this limitation in the future.

G. COMPARISON TO EXPERT RADIOLOGISTS

In comparing the performance of the full-resolution deep learning approach with the performance of human radiologists using the provided metadata, we found that the presented method is approximately equal in sensitivity to human

radiologists for obvious nodules. However, it outperforms human radiologists for subtle, and very subtle nodules. Although the algorithm shows promising results, there is a statistical deviation that is currently too wide to immediately support a clinical system. With the availability of more data, the deviation gap could potentially be reduced, leading to the development of a fully generalized tool that could significantly improve the speed and accuracy of lung cancer diagnosis from CXR images. Such a tool would enable earlier intervention and a better prognosis for patients with this deadly cancer.

VII. CONCLUSION

This study proves the feasibility of developing a generalized lung nodule segmentation/localization algorithm using well-known E-D algorithms, with some novel refinements. The achieved results are highly promising, particularly considering the limited size of the training dataset for a deep learning algorithm. One notable advantage of the proposed algorithm is that it does not rely on any parameterized linear filtering methods like LoG. Instead, it utilizes end-to-end deep learning, enabling simultaneous feature extraction and semantic segmentation. This approach allows the algorithm to effectively generalize external datasets without the need for parameter adjustments or fine-tuning against the external dataset. Such generalization is crucial for real-world applications and enhances the potential clinical utility of the algorithm.

The motivation behind this study was to lay the groundwork for a clinically beneficial algorithm that could augment a radiologist's workflow. E-D networks, like the ones evaluated in this research, appear to be well-suited for this objective due to their efficiency, standard library implementation, and ease of deployment as federated algorithms. The primary constraint of the proposed method lies in the availability of labelled and segmented training data for the E-D network. The algorithm's segmentation accuracy could be improved with more data, leading to higher sensitivity and lower false positive scores. Acquiring a larger and more diverse dataset could significantly enhance the algorithm's performance and practical applicability in real-world medical settings.

In the future, we plan to explore two strategies to access additional data. Firstly, we will use the algorithm described in this paper to ''bootstrap'' training data from the NIH dataset, for external testing against a third independent dataset, with human radiologists providing sensitivity and FP scores. This approach will help validate the algorithm's generalization capabilities on new and diverse datasets. Secondly, we will implement the E-D algorithm as a swarm/federated learning [76], [77] model. By conducting the implementation as a federated clinical trial, we can tap into a larger volume of more diverse CXR images, thus creating a tool that can be effectively integrated into a commercial clinical system while still preserving patient's privacy. Additionally, the low resource requirements, especially during inference, make E-D-based designs suitable for integration into mobile and augmented reality applications, supporting telehealth and enhancing radiologist' access to remote and disadvantaged communities. It is hoped that AI-based assistance can facilitate broader lung cancer screening through fast and cost-effective CXR images, leading to earlier intervention and a reduction in mortality rates. By leveraging advanced technology, we aim to improve the overall efficiency and effectiveness of lung cancer diagnosis and management.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization, M.J. Horry; methodology, M.J. Horry; software, M.J. Horry; validation, M.J. Horry; formal analysis, M.J. Horry; investigation, M.J. Horry, S. Chakraborty; resources, M.J. Horry, S. Chakraborty; data curation, M.J. Horry; writing-original draft preparation, M.J. Horry; writing-review and editing, S. Chakraborty, B. Pradhan, M. Paul, J. Zhu, P. Barua, U.R. Acharya, F. Chen, J. Zhou; visualization, M.J. Horry; supervision, S. Chakraborty, B. Pradhan, M. Paul; project administration, M.J. Horry. All authors have read and agreed to the published version of the manuscript.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGMENT

Michael J. Horry would like to thank his employer Rohling International for providing flexibility allowing for the preparation of this manuscript.

REFERENCES

[1] AICR. *Lung Cancer*. World Cancer Reseach Fund-American Institute of Cancer Research. Accessed: Apr. 11, 2022. [Online]. Available: https://www.wcrf.org/dietandcancer/lung-cancer/

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, ''Cancer statistics, 2021,'' *CA, Cancer J. Clinicians*, vol. 71, no. 1, pp. 7–33, Jan. 2021, doi: 10.3322/caac.21654.

[3] Q. Luo, X. Q. Yu, S. Wade, M. Caruana, F. Pesola, K. Canfell, and D. L. O'Connell, ''Lung cancer mortality in australia: Projected outcomes to 2040,'' *Lung Cancer*, vol. 125, pp. 68–76, Nov. 2018, doi: 10.1016/j.lungcan.2018.09.001.

[4] R. Pirker, ''JICC01.07 oral VEGFR inhibitor monotherapy for lung cancer: Should we or should we not?'' *J. Thoracic Oncol.*, vol. 16, no. 3, pp. S99–S100, Mar. 2021, doi: 10.1016/j.jtho.2021.01.162.

[5] P. Campadelli, E. Casiraghi, and D. Artioli, ''A fully automated method for lung nodule detection from postero-anterior chest radiographs,'' *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1588–1603, Dec. 2006, doi: 10.1109/TMI.2006.884198.

[6] A. R. Larici, A. Farchione, P. Franchi, M. Ciliberto, G. Cicchetti, L. Calandriello, A. del Ciello, and L. Bonomo, ''Lung nodules: Size still matters,'' *Eur. Respiratory Rev.*, vol. 26, no. 146, Dec. 2017, Art. no. 170025, doi: 10.1183/16000617.0025-2017.

[7] R. J. McDonald, K. M. Schwartz, L. J. Eckel, F. E. Diehn, C. H. Hunt, B. J. Bartholmai, B. J. Erickson, and D. F. Kallmes, "The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload," *Academic Radiol.*, vol. 22, no. 9, pp. 1191–1198, Sep. 2015, doi: 10.1016/j.acra.2015.05.007.

[8] R. J. M. Bruls and R. M. Kwee, "Workload for radiologists during on-call hours: Dramatic increase in the past 15 years," *Insights Into Imag.*, vol. 11, no. 1, p. 121, Dec. 2020, doi: 10.1186/s13244-020-00925-z.

[9] S. Ziegelmayer, M. Graf, M. Makowski, J. Gawlitza, and F. Gassert, "Cost-effectiveness of artificial intelligence support in computed tomography-based lung cancer screening," *Cancers*, vol. 14, no. 7, p. 1729, Mar. 2022, doi: 10.3390/cancers14071729.

[10] D. E. Goldsbury, M. F. Weber, S. Yap, N. M. Rankin, P. Ngo, L. Veerman, E. Banks, K. Canfell, and D. L. O'Connell, "Health services costs for lung cancer care in australia: Estimates from the 45 and up study," *PLoS ONE*, vol. 15, no. 8, Aug. 2020, Art. no. e0238018, doi: 10.1371/journal.pone.0238018.

[11] L. Kutikova, L. Bowman, S. Chang, S. R. Long, C. Obasaju, and W. H. Crown, "The economic burden of lung cancer and the associated costs of treatment failure in the United States," *Lung Cancer*, vol. 50, no. 2, pp. 143–154, Nov. 2005, doi: 10.1016/j.lungcan.2005.06.005.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[13] K. P. Exarchos, G. Gkrepi, K. Kostikas, and A. Gogali, "Recent advances of artificial intelligence applications in interstitial lung diseases," *Diagnostics*, vol. 13, no. 13, p. 2303, Jul. 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/13/2303

[14] P. Bosowski, J. Bosowska, and J. Nalepa, "Evolving deep ensembles for detecting COVID-19 in chest X-rays," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 3772–3776, doi: 10.1109/ICIP42928.2021.9506119.

[15] S. S. Alahmari, B. Altazi, J. Hwang, S. Hawkins, and T. Salem, "A comprehensive review of deep learning-based methods for COVID-19 detection using chest X-ray images," *IEEE Access*, vol. 10, pp. 100763–100785, 2022, doi: 10.1109/ACCESS.2022.3208138.

[16] X. Li, L. Shen, and S. Luo, "A solitary feature-based lung nodule detection approach for chest X-ray radiographs," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 2, pp. 516–524, Mar. 2018, doi: 10.1109/JBHI.2017.2661805.

[17] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Med. Image Anal.*, vol. 12, no. 3, pp. 240–258, Jun. 2008, doi: 10.1016/j.media.2007.10.004.

[18] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli, "Neural networks for computer-aided diagnosis: Detection of lung nodules in chest radiograms," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 344–357, Dec. 2003, doi: 10.1109/TITB.2003.821313.

[19] A. Anand, S. S. Tripathy, and R. S. Kumar, "An improved edge detection using morphological Laplacian of Gaussian operator," in *Proc. 2nd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Noida, India, Feb. 2015, pp. 532–536, doi: 10.1109/SPIN.2015.7095391.

[20] C. S. Pereira, H. Fernandes, A. M. Mendonça, and A. Campilho, "Detection of lung nodule candidates in chest radiographs," presented at the 3rd Iberian Conf. Pattern Recognit. Image Anal., (IbPRIA), Girona, Spain, 2007.

[21] M. G. Penedo, M. J. Carreira, A. Mosquera, and D. Cabello, "Computer-aided diagnosis: A neural-network-based approach to lung nodule detection," *IEEE Trans. Med. Imag.*, vol. 17, no. 6, pp. 872–880, 1998, doi: 10.1109/42.746620.

[22] H. Yu, J. Li, L. Zhang, Y. Cao, X. Yu, and J. Sun, "Design of lung nodules segmentation and recognition algorithm based on deep learning," *BMC Bioinf.*, vol. 22, no. S5, p. 314, Nov. 2021, doi: 10.1186/s12859-021-04234-0.

[23] S. Chen, K. Suzuki, and H. Macmahon, "Development and evaluation of a computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule enhancement with support vector classification," *Med. Phys.*, vol. 38, no. 4, pp. 1844–1858, Apr. 2011, doi: 10.1118/1.3561504.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, Munich, Germany, 2015, pp. 170–177.

[25] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, Jan. 2000, doi: 10.2214/ajr.174.1.1740071.

[26] S. Park, W. Jeong, and Y. S. Moon, "X-ray image segmentation using multi-task learning," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 3, pp. 1104–1120, 2020, doi: 10.3837/tiis.2020.03.011.

[27] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," 2016, *arXiv:1607.08022*.

[28] T. Agrawal and P. Choudhary, "Segmentation and classification on chest radiography: A systematic survey," *Vis. Comput.*, vol. 39, no. 3, pp. 875–913, Mar. 2023, doi: 10.1007/s00371-021-02352-7.

[29] B. Liu, W. Chi, X. Li, P. Li, W. Liang, H. Liu, W. Wang, and J. He, "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: Three decades' development course and future prospect," *J. Cancer Res. Clin. Oncol.*, vol. 146, no. 1, pp. 153–185, Jan. 2020, doi: 10.1007/s00432-019-03098-5.

[30] S. H. Bradley, S. Abraham, M. E. Callister, A. Grice, W. T. Hamilton, R. R. Lopez, B. Shinkins, and R. D. Neal, "Sensitivity of chest X-ray for detecting lung cancer in people presenting with symptoms: A systematic review," *Brit. J. Gen. Pract.*, vol. 69, no. 689, pp. e827–e835, Dec. 2019, doi: 10.3399/bjgp19X706853.

[31] S. H. Bradley, "Chest X-ray sensitivity and lung cancer outcomes: A retrospective observational study," *Brit. J. Gen. Pract.*, vol. 71, no. 712, pp. e862–e868, Nov. 2021, doi: 10.3399/bjgp.2020.1099.

[32] A. Panunzio and P. Sartori, "Lung cancer and radiological imaging," *Current Radiopharmaceuticals*, vol. 13, no. 3, pp. 238–242, Nov. 2020, doi: 10.2174/1874471013666200523161849.

[33] M. Usman Ali, J. Miller, L. Peirson, D. Fitzpatrick-Lewis, M. Kenny, D. Sherifali, and P. Raina, "Screening for lung cancer: A systematic review and meta-analysis," *Preventive Med.*, vol. 89, pp. 301–314, Aug. 2016, doi: 10.1016/j.ypmed.2016.04.015.

[34] D. H. Ballard and J. Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," *IEEE Trans. Comput.*, vol. C-25, no. 5, pp. 503–513, May 1976, doi: 10.1109/TC.1976.1674638.

[35] K. H. Rotte, "Use of computer-aided X-ray diagnosis in oncology," *Arch. Geschwulstforsch*, vol. 49, no. 4, pp. 347–356, 1979.

[36] W. A. Lampeter, "ANDS-V1 computer detection of lung nodules," *Proc. SPIE*, vol. 555, pp. 253–261, Sep. 1985, doi: 10.1117/12.949501.

[37] W. A. Lampeter and J. C. Wandtke, "Computerized search of chest radiographs for nodules," *Investigative Radiol.*, vol. 21, no. 5, pp. 384–390, May 1986, doi: 10.1097/00004424-198605000-00003.

[38] M. L. Giger, K. Doi, and H. M. MacMahon, "Computerized detection of lung nodules in digital chest radiographs," in *Proc. SPIE*, vol. 767, pp. 384–386, Jan. 1987, doi: 10.1117/12.967022.

[39] H. Yoshida and B. Keserci, "Bayesian wavelet snake for computer-aided diagnosis of lung nodules," *Integr. Comput.-Aided Eng.*, vol. 7, no. 3, pp. 253–269, Jul. 2000, doi: 10.3233/ica-2000-7306.

[40] B. Keserci and H. Yoshida, "Computerized detection of pulmonary nodules in chest radiographys based on morphological features and wavelet snake model," *Med. Image Anal.*, vol. 6, no. 4, pp. 431–447, 2002, doi: 10.1016/S1361-8415(02)00064-6.

[41] H. Macmahon, "Improvement in detection of pulmonary nodules: Digital image processing and computer-aided diagnosis," *RadioGraphics*, vol. 20, no. 4, pp. 1169–1177, Jul. 2000, doi: 10.1148/radiographics.20.4.g00jl211169.

[42] H. MacMahon, F. Li, R. Engelmann, R. Roberts, and S. Armato, "Dual energy subtraction and temporal subtraction chest radiography," *J. Thoracic Imag.*, vol. 23, no. 2, pp. 77–85, 2008, doi: 10.1097/RTI.0b013e318173dd38.

[43] H. Yoshimura, M. L. Giger, K. Doi, H. MacMahon, and S. M. Montner, "Computerized scheme for the detection of pulmonary nodules: A nonlinear filtering technique," *Investigative Radiol.*, vol. 27, no. 2, pp. 124–129, Feb. 1992, doi: 10.1097/00004424-199202000-00005.

[44] A. Tompe and K. Sargar, *X-Ray Image Quality Assurance*. St. Petersburg, FL, USA: StatPearls Publishing, 2022.

[45] P. Ranefall and C. Wählby, "Global gray-level thresholding based on object size," *Cytometry A*, vol. 89, no. 4, pp. 385–390, Apr. 2016, doi: 10.1002/cyto.a.22806.

[46] W. Jasim and R. Mohammed, "A survey on segmentation techniques for image processing," *Iraqi J. Electr. Electron. Eng.*, vol. 17, no. 2, pp. 73–93, Dec. 2021, doi: 10.37917/ijeee.17.2.10.
[47] S. Cui, S. Ming, Y. Lin, F. Chen, Q. Shen, H. Li, G. Chen, X. Gong, and H. Wang, "Development and clinical application of deep learning model for lung nodules screening on CT images," *Sci. Rep.*, vol. 10, no. 1, Aug. 2020, Art. no. 13657, doi: 10.1038/s41598-020-70629-3.
[48] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[50] 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 Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016, doi: 10.1109/TMI.2016.2553401.
[51] S. Amraee, M. Chinipardaz, and M. Charoosaei, "Analytical study of two feature extraction methods in comparison with deep learning methods for classification of small metal objects," *Vis. Comput. Ind., Biomed., Art*, vol. 5, no. 1, p. 13, Dec. 2022, doi: 10.1186/s42492-022-00111-6.
[52] J. Shiraishi, Q. Li, and K. Doi, "Development of a computerized scheme for detection of very subtle lung nodules located in opaque areas on chest radiographs," *Proc. SPIE*, vol. 6144, Mar. 2006, Art. no. 61441U.
[53] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.
[54] X. Li, L. Shen, X. Xie, S. Huang, Z. Xie, X. Hong, and J. Yu, "Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101744, doi: 10.1016/j.artmed.2019.101744.
[55] A. M. R. Schilham, B. van Ginneken, and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database," *Med. Image Anal.*, vol. 10, no. 2, pp. 247–258, Apr. 2006, doi: 10.1016/j.media.2005.09.003.
[56] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, Mar. 2019, Art. no. 014006, doi: 10.1117/1.JMI.6.1.014006.
[57] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.
[58] Open Source Computer Vision Library. (2015). *OpenCV*. [Online]. Available: http://opencv.org/
[59] D. M. Hansell, A. A. Bankier, H. Macmahon, T. C. McLoud, N. L. Müller, and J. Remy, "Fleischner society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, Mar. 2008, doi: 10.1148/radiol.2462070712.
[60] H. MacMahon, D. P. Naidich, J. M. Goo, K. S. Lee, A. N. C. Leung, J. R. Mayo, A. C. Mehta, Y. Ohno, C. A. Powell, M. Prokop, G. D. Rubin, C. M. Schaefer-Prokop, W. D. Travis, P. E. Van Schil, and A. A. Bankier, "Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner society 2017," *Radiology*, vol. 284, no. 1, pp. 228–243, Jul. 2017, doi: 10.1148/radiol.2017161659.
[61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[62] D. P. S. Gomes, A. Ulhaq, M. Paul, M. J. Horry, S. Chakraborty, M. Saha, T. Debnath, and D. M. M. Rahaman, "Features of ICU admission in X-ray images of COVID-19 patients," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 200–204.
[63] M. Abadi, "TensorFlow: A system for large-scale machine learning," presented at the 12th USENIX Conf. Operating Syst. Design Implement. (OSDI'16), Savannah, GA, USA, 2016.
[64] T. Nemoto, N. Futakami, E. Kunieda, M. Yagi, A. Takeda, T. Akiba, E. Mutu, and N. Shigematsu, "Effects of sample size and data augmentation on U-Net-based automatic segmentation of various organs," *Radiolog. Phys. Technol.*, vol. 14, no. 3, pp. 318–327, Sep. 2021, doi: 10.1007/s12194-021-00630-6.
[65] S. Chen and K. Suzuki, "Computerized detection of lung nodules by means of 'virtual dual-energ' radiography," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 369–378, Feb. 2013, doi: 10.1109/TBME.2012.2226583.
[66] S. Chen, K. Suzuki, and H. MacMahon, "Improved computerized detection of lung nodules in chest radiographs by means of'virtual dual-energy'radiography," *Proc. SPIE*, vol. 7963, Mar. 2011, Art. no. 79630C, doi: 10.1117/12.877579.
[67] Z. Al-Ameen, G. Sulong, A. Rehman, A. Al-Dhelaan, T. Saba, and M. Al-Rodhaan, "An innovative technique for contrast enhancement of computed tomography images using normalized gamma-corrected contrast-limited adaptive histogram equalization," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 32, Dec. 2015, doi: 10.1186/s13634-015-0214-1.
[68] J. L. Lehr and P. Capek, "Histogram equalization of CT images," *Radiology*, vol. 154, no. 1, pp. 163–169, Jan. 1985, doi: 10.1148/radiology.154.1.3964935.
[69] R. H. Sherrier, C. Chiles, W. E. Wilkinson, G. A. Johnson, and C. E. Ravin, "Effects of image processing on nodule detection rates in digitized chest radiographs: ROC study of observer performance," *Radiology*, vol. 166, no. 2, pp. 447–450, Feb. 1988, doi: 10.1148/radiology.166.2.3336719.
[70] R. P. M. Sundhari, "RETRACTED ARTICLE: Enhanced histogram equalization based nodule enhancement and neural network based detection for chest X-ray radiographs," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 3, pp. 3831–3839, Mar. 2021, doi: 10.1007/s12652-020-01701-z.
[71] M. J. Horry, S. Chakraborty, B. Pradhan, M. Paul, J. Zhu, H. W. Loh, P. D. Barua, and U. R. Acharya, "Development of debiasing technique for lung nodule chest X-ray datasets to generalize deep learning models," *Sensors*, vol. 23, no. 14, p. 6585, Jul. 2023, doi: 10.3390/s23146585.
[72] L. Ketai, M. Malby, K. Jordan, A. Meholic, and J. Locken, "Small nodules detected on chest radiography," *Chest*, vol. 118, no. 3, pp. 610–614, Sep. 2000, doi: 10.1378/chest.118.3.610.
[73] H. Wang, S. Wang, Z. Qin, Y. Zhang, R. Li, and Y. Xia, "Triple attention learning for classification of 14 thoracic diseases using chest radiography," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101846, doi: 10.1016/j.media.2020.101846.
[74] R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?" 2017, *arXiv:1611.07450*.
[75] P. Rajpurkar and M. P. Lungren, "The current and future state of AI interpretation of medical images," *New England J. Med.*, vol. 388, no. 21, pp. 1981–1990, May 2023, doi: 10.1056/NEJMra2301725.
[76] S. Warnat-Herresthal, "Swarm learning for decentralized and confidential clinical machine learning," *Nature*, vol. 594, no. 7862, pp. 265–270, Jun. 2021, doi: 10.1038/s41586-021-03583-3.
[77] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020, doi: 10.1109/MSP.2020.2975749.

MICHAEL J. HORRY (Member, IEEE) received the double degree in engineering and laws from the University of New South Wales, in 1998. He is currently pursuing the Ph.D. degree with the University of Technology Sydney (UTS). He research advanced computer vision-based diagnosis methods from medical images. Since that time, he has held various industry roles, including a Principal Architect with IBM responsible for solution design across the business partner and SI community. He is currently a Senior Architect with Rohling International, a boutique consultancy progressing technology for good initiatives. He is an Open Group Master Certified IT Architect and TOGAF 9 certified and a member of the Enterprise Architects Association of Australia. More recently, he has developed an interest in machine learning and particularly computer vision applications with over 500 citations, placing him in the top 1% of ResearchGate members who first published, in 2020.

SUBRATA CHAKRABORTY (Senior Member, IEEE) received the Ph.D. degree in decision support systems from Monash University, Australia. He is currently a Senior Lecturer with the School of Science and Technology, Faculty of Science, Agriculture, Business, and Law, University of New England, Australia. Previously, he was an Academician with the University of Technology Sydney, (UTS), Australia, University of Southern Queensland, Charles Sturt University, Queensland University of Technology, and Monash University. He is also a Visiting Fellow with the Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, UTS, and a Visiting Fellow with the Griffith Business School, Queensland. His current research interests include data analytics, machine learning, image processing, and optimization with decision support applications in diverse domains, including health, agriculture, business, and education.

BISWAJEET PRADHAN (Senior Member, IEEE) received the B.Sc. degree (Hons.) from Berhampur University, India, the M.Sc. degree from the Indian Institute of Technology (IIT) Bombay, India, the joint M.Tech. degree in civil engineering from IIT, Kanpur, India, and the Dresden University of Technology, Dresden, Germany, and the Ph.D. degree in GIS and geomatics engineering from University Putra Malaysia, Serdang, Malaysia. He is currently a Distinguished Professor with the School of Information, Systems and Modeling and the Director of the Centre for Advanced Modeling and Geospatial Information Systems, University of Technology Sydney. Before joining the School of Information, Systems and Modeling, Faculty of Engineering and IT, Biswajeet, he was a Full Professor in GIS and remote sensing with the Faculty of Engineering, Universiti Putra Malaysia. He has more than 19 years of teaching, research, and industrial experience in the field of geospatial information systems. He is listed as the World's most Highly Cited Researcher by Clarivate Analytics Report, in 2017 and 2016, as one of the world's most influential mind. He is a recipient of Alexander von Humboldt Research Fellowship from Germany. In 2011, he received his habituation in remote sensing from the Dresden University of Technology. He was also a recipient of the German Academic Exchange Research (DAAD) Fellowship Award, the Saxony State Fellowship, from 1999 to 2002, the Keith Aurtherton Research Award, and the Georg Forster Research Award from the German Government. Since February 2015, he has been serving as the Ambassador Scientist for Alexander Humboldt Foundation, Germany. He has widely travelled abroad visiting more than 55 countries to present his research findings.

MANORANJAN PAUL (Senior Member, IEEE) received the Ph.D. degree from Monash University, Australia, in 2005. He was a Postdoctoral Research Fellow with the University of New South Wales, Monash University, and Nanyang Technological University. Currently, he is a Full Professor, the Director of the Computer Vision Laboratory, and the Leader of the Machine Vision and Digital Health (MaViDH) Research Group, Charles Sturt University, Australia. His major research interests include the field of video coding, image processing, digital health, wine technology, machine learning, EEG signal processing, eye tracking, and computer vision. He has published around 200 peer-reviewed publications, including 72 journals. He was an invited Keynote Speaker in IEEE DICTA-17 & 13, CWCN-17, WoWMoM-14, and ICCIT-10. He was awarded the ICT Researcher of the Year 2017 by Australian Computer Society. He obtained more than $3.6 million competitive external grant, including Australian Research Council (ARC) Discovery grants, Australia-China Grant. He has supervised 15 Ph.D. students to completion. He was the General Chair of PSIVT-19 and the Program Chair of PSIVT-17 and DICTA-18. He is also an Associate Editor of three top ranked journals, such as IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *EURASIP Journal in Advances on Signal Processing*.

JING ZHU received the Bachelor of Medicine/Bachelor of Surgery degree (Hons.) from Western Sydney University, in 2016, and the master's degree in public health from the University of New South Wales, in 2022. Following her residency with Concord Hospital, she was accepted onto the Radiology Training Program. During her master's degree, she concurrently worked with the Sydney Local Health District Public Health Unit. She currently practices as a Radiology Registrar with Westmead and Auburn Hospital.

PRABAL DATTA BARUA (Member, IEEE) received the master's degree in information systems from the University of Ballarat, Australia, and the Ph.D. degree in information systems from the University of Southern Queensland (UniSQ), Australia. He also received a Graduate Certificate in data analytics from the Melbourne Institute of Technology; multiple diplomas and advanced diploma courses in software development from TAFE NSW. He is the author with over 100 publications in top-tier academic journals and has contributed as a guest editor in many international journals. He has secured over A$3M in government and industry research grants for conducting cutting-edge research in applying artificial intelligence in health informatics, education analytics, and ICT for business transformation. He is supervising more than 15 research students. Currently, he is the Academic Dean of the Australian International Institute of Higher Education; an Adjunct Professor with UniSQ; a member of the ICT Industry Advisory Board with UniSQ; a member of the ICT Programs Leadership and Governance Committee with UniSQ; a member of Business in AI Steering Committee with UniSQ; an accredited Higher Degree by Research (HDR) Supervisor with UniSQ. He is also an Adjunct Professor with Taylor's University Malaysia; an Adjunct Professor with the University of New England (UNE); an accredited HDR Supervisor with UNE, and a Visiting Professor with the SRM University of Science and Technology, India. He is also the CEO of Cogninet Australia, which conducts research in medical and educational informatics using artificial intelligence (AI). He is also an Honorary Professional/Industry Fellow with the University of Technology Sydney (UTS).

HASAN SAEED MIR (Senior Member, IEEE) received the B.S. (cum laude), M.S., and Ph.D. degrees in electrical engineering from the University of Washington, Seattle, USA, in 2000, 2001, and 2005, respectively. From 2005 to 2009, he was a member of the Technical Staff with the Air Defense Technology Group, MIT Lincoln Laboratory, Lexington, MA, USA, where he was involved in several projects related to airborne and maritime radar systems. Since 2009, he has been with the Department of Electrical Engineering, American University of Sharjah, where he is currently a Professor. His research interests include radar signal processing and adaptive beamforming. He is a Senior Editor of IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, and an Associate Editor of IEEE ACCESS, *IET Microwaves, Antennas and Propagation*, and *IET Signal Processing*.

FANG CHEN (Senior Member, IEEE) is currently a Distinguished Professor in AI and data science with the University of Technology Sydney (UTS), Australia, where she is also the Executive Director of Data Science and a Distinguished Professor. Her research interests include machine learning, AI ethics, behavior analytics, and their applications. She has created many world-class AI innovations while working in Intel, Motorola, NICTA, CSIRO, and now UTS, and helped governments and industries utilizing data and significantly increasing productivity, safety, and customer satisfaction.

JIANLONG ZHOU (Senior Member, IEEE) is currently an Associate Professor with the Data Science Institute, University of Technology Sydney (UTS), Australia. His research interests include AI ethics, AI for social good, AI explainability, data analytics, visual analytics, behavior analytics, human–computer interaction, VR, and related applications. He is a leading Senior Researcher in trustworthy and transparent machine learning and has done pioneering research in the area of linking human and machine learning.

U. RAJENDRA ACHARYA (Senior Member, IEEE) received the Ph.D., D.Eng., and D.Sc. degrees. He is currently a Professor (Artificial Intelligence in Healthcare) with the University of Southern Queensland, Australia; a Distinguished Professor with the International Research Organization for Advanced Science and Technology, Kumamoto University, Japan; an Adjunct Professor with the University of Malaya, Malaysia; and an Adjunct Professor with Asia University, Taiwan. His research interests include biomedical imaging and signal processing, data mining, and visualization, and the applications of biophysics for better healthcare design and delivery. His funded research has accrued cumulative grants exceeding six million Singapore dollars. He has authored more than 600 publications, including 550 in refereed international journals, 42 in international conference proceedings, and 17 books. He has received more than 73,000 citations on Google Scholar (with an H-index of 134). He has been ranked in the Top 1% of the Highly Cited Researchers for the last seven consecutive years (2016–2022) in computer science, according to the Essential Science Indicators of Thomson. He is on the editorial boards of many journals and he has served as a guest editor on several AI-related issues.