

Unsupervised Feature Mapping via Stacked Sparse Autoencoder for Automated Detection of Large Pulmonary Nodules in CT Images

Anindya Gupta¹, Tonis Saar², Olev Martens¹, Yannick Le Moullec¹

¹Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology,
Ehitajate Tee 5, 19086 Tallinn, Estonia

²Elestron OU,
Tallinn, Estonia
anindya.gupta@ttu.ee

Abstract—We present a novel and efficient false positive reduction stage, using stacked sparse autoencoder, for the automatic detection of large nodules in computed tomography (CT) images. The discriminative features are automatically learnt in unsupervised manner. The initial candidates are segmented using candidate detection method specifically designed for the large nodules. For each candidate, 3D grayscale clusters are computed and, are later resized into a uniform size of $10 \times 10 \times 5$ for feature mapping. Finally, a softmax layer is used for the binary classification. Data augmentation, sparsity regularization, and L2 weight regularization are applied to overcome the generalization issue. On 899 CT scans taken from LIDC-IDRI, our method yields a high detection sensitivity of 90 % with only 4 false positives per scan and an area under receiver operating curve of 0.983. An external validation on a completely independent dataset from PCF is also performed to evaluate the potency of the proposed method. We showed that the proposed stacked sparse autoencoder is efficient enough to be accommodated as a false positive reduction phase in a computer-aided-detection system.

Index Terms—Computer-aided detection; pulmonary nodule; stacked sparse autoencoder; unsupervised learning.

I. INTRODUCTION

Worldwide, lung cancer is one of the main cause of cancer-cause death. Detection in the very late stages results in an ineffective treatment planning and high mortality rate. It is thus essential and crucial to detect the cancerous lesions in the early stages. The results of the national lung screening trial (NLST) reported a significant reduction of 20 % in lung cancer mortality rate as a result of using low-dose computed tomography (CT) imaging modality. The findings of NLST also encouraged other countries to organize such large scale lung cancer screening trial [1].

However, data processing during such screening trials remains a major challenges, specifically in terms of radiologists' (human readers) efficiency. In general, interpreting the acquired CT scans is a monotonous, error-

prone and time consuming tasks, which can certainly affect the efficiency of the readers. Hence, automation of CT scan analysis is becoming a crucial field of research [2].

For minimal human interaction with adequate efficiency, computer-aided detection (CAD) systems are utilized as a second reader to automatically detect the lesions during the screening trials. In such scenarios, to improve the detection accuracy, radiologists initially mark the lesions in an unaided manner and later cross-validate their markings with the CAD findings. Eventually, this process yields a high accuracy but also increases the reading time. An efficient way to reduce the reading time could be to accommodate the CAD as first reader followed by a visual inspection of the CAD findings by radiologists to make a quick inspection and to maintain the integrity of workflow [3].

A fundamental requirement for a CAD system to be accommodated as a first reader is that it yields a high sensitivity for all relevant lesions/nodules. The general pipeline of a CAD system consists of an initial candidate detection step followed by a false positive reduction step [4]. The initial candidate detection step aims to achieve a high sensitivity and, unfortunately, typically results in a bigger set of false positives (FP). FPs are then successively reduced in the next step, which also enhances the overall performance of the CAD systems. During the FP reduction step, a large set of low-level features such as blobness, cluster, and intensity are computed for a supervised classification scheme [4].

In last two decades, several CAD systems have been proposed that yielded a high detection accuracy against multiple phenotypes of nodules but that were unable to achieve an adequate FP rate. The outcome of ANODE09 [5] showed that these systems are unable to detect the less prevalent class of nodules (i.e. large nodules). One reason for degradation in the detection performance is the number of large nodules (malignant lesions); which is quantitatively smaller in comparison to the number of other smaller nodules. The FP reduction step is often trained with randomly selected candidates; hence, large nodules are often undersampled in the training set and eventually considered as lower priority candidates. Another reason is the extraction of underlying features of the lesions, which can exhibit the

Manuscript received 23 March, 2017; accepted 8 October, 2017.

This research was partly funded by a European Union's Horizon 2020 research and innovation programme under grant No 668995. This work is also supported by Estonian IT Academy Stipend Program, EU Institutional grant IUT19-11 of Estonian Research Council, and Doctoral School at Tallinn University of Technology.

same characteristics as non-nodules, resulting in a difficult process of low-level feature extraction steps. The motivation behind the proposed work is to overcome these limitations by means of a so-called autoencoder.

Recently, convolutional neural networks (CNN) in supervised learning, together with large amounts of augmented data, are effectively used in several medical imaging applications [6]. Conversely, an autoencoder [7] with multiple hidden layers can be trained in an unsupervised fashion wherein each layer can efficiently learn to map the input features. However, training the autoencoder with multiple hidden layers can be complicated in practice. It is thus effective to train individual layers at a time. Although, being so efficient, there is still no work that use an autoencoders to analyze the CT volumetric data. To the best of our knowledge, this is the first work, which reports an unsupervised feature mapping technique using autoencoders for lung image database consortium/image database resource initiative (LIDC/IDRI) data which is also validated on the PCF dataset. Some examples of the detected lesions from the PCF dataset are shown in Fig. 1.

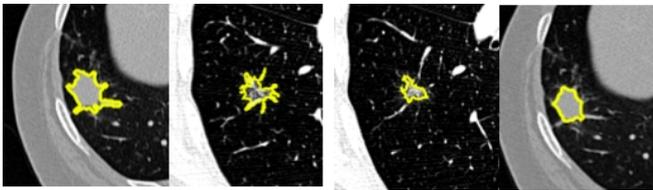


Fig. 1. Some example of detected lesions from the PCF dataset. The detected nodules are marked in the yellow bounding box.

A. Objective and Contribution

In this work, a fully automated computer aided detection (CAD) system specifically designed to detect large pulmonary nodules is proposed. The ultimate objective of this work is to increase the reading efficiency of the CAD system. The contribution of the work are as follows: (1) We formulate a novel FP reduction stage using stacked sparse autoencoder (SSAE) to classify the malignant nodule. The candidates are extracted through our existing algorithm, which also boost the sensitivity of the initial candidate detection stage. (2) a performance benchmark is presented and compared with the previously reported CAD systems; and (3) the proposed method is also validated on a completely independent dataset.

II. IMAGING DATA

In this work, the CT cases (DICOM formatted) from two datasets are utilized for training and independent validation of proposed CAD system. The CT scans from LIDC/IDRI are used for the training and the supervised validation of the proposed CAD system. Additionally, CT scans from the PCF dataset are used for independent validation of proposed CAD system.

A. Lung Image Database Consortium (LIDC)/Image Database Resource Initiative (IDRI)

We utilized the largest publically available dataset to train and validate the proposed CAD system. LIDC-IDRI consists of a heterogeneous set of 1018 CT scans captured at seven different institutes. Each scan is annotated by up to 4 radiologists in a two-phase annotation process (blind and

unblind). Additionally, boundary outlines and subjective ratings of each lesion are also provided. [8], [9]

In this work, the CT cases of slice thickness up to 3 mm are utilized, resulting in a set of 899 CT scans. Other CT cases are rejected due to unacceptable slice thickness [10]. To overcome the variance in the boundary outlines marked by multiple radiologists, the individual outlines of each reader are overlapped to compute a centre-of-mass.

The diametric size of each nodule is calculated according to the criteria reported in [3]. Then, a diameter of 10 mm (diametric size of sphere) is considered as the size threshold criterion to develop a reference set, which consists of 289 nodule candidates. Note that, only those nodule candidates, which were marked by at least 3 radiologists are selected.

B. External Validation data (PCF)

VIA Cornell University released a public dataset of thoracic CT scans for the early detection and diagnosis of lung cancer. The data can be accessed on VIA website [11]. PCF consists of heterogeneous cancer cases. We selected 33 CT scans consisting 40 large nodule candidates for independent validation of the proposed CAD system. For every case, nodule annotations (spatial coordinates of nodule center, diameter, and volume) are also provided.

III. METHODOLOGY

The outline of our proposed CAD system for the detection of the large nodule candidates is shown in Fig. 2. The CAD pipeline is divided into two stages: i) initial candidate detection and ii) FP reduction. Current work emphasizes on the FP reduction stage. The initial candidate detection stage is reported elsewhere [12], and briefly discussed here.

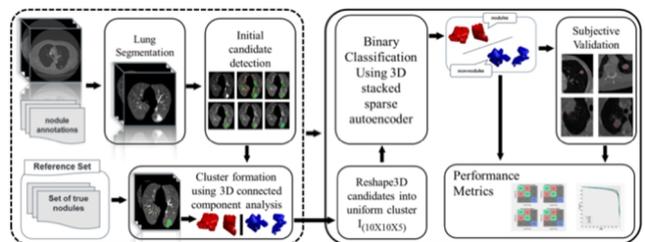


Fig. 2. Overview of the developed CAD pipeline. The dotted box is the initial stage and the box with bold line is the FP reduction stage.

A. Initial Candidate Detection Stage

The initial candidate detection stage consists of multiple steps. Initially, a thresholding based method is implemented to segment the lung regions in each section of the CT volume. Then, a morphological closing operation is used to further refine the segmented lung region. For the initial candidate detection, a multistage rule based sub-algorithm module consisting of six stages is developed.

The grayscale masks of all labelled candidates are initially extracted from the lungs. Secondly, a normalized intensity value of 0.007 is calculated to segment the grayscale masks. This is followed by a morphological branchpoint operation is used to refine the juxta-vascular candidates. Next, a morphological erosion operation using disk kernel with radius of 2 pixels is used for further segmentation. Subsequently, circularity is considered as a rule to detect the candidates wherein a candidate that ranges between the threshold values of 9–380 is considered for next

stage. Lastly, an additional morphological dilation operation using disk kernel with a radius of 2 pixels is applied on the final candidates.

B. Connected Component Analysis

Connected voxels of each candidate are clustered via 3D connected component analysis using a 26-points connectivity scheme. The centroids of the candidates in the consecutive slices (I_s) of a 3D cluster is computed using (1). The clusters of size between 268 mm^3 – $34,000 \text{ mm}^3$ (diametric size of sphere: 8 mm–40 mm) are considered as the potential candidates for the next stage of the CAD system

$$[C_x, C_y, C_z] = \left[\frac{\sum_{i=1}^N x_i}{N}, \frac{\sum_{i=1}^N y_i}{N}, \frac{\sum_{i=1}^N z_i}{N} \right], \quad (1)$$

where C_x, C_y, C_z and are the centroids of the cluster and N is the total number of the coordinates for the cluster.

C. 3D Cluster Resizing

The initially detected candidates are resized into a cluster of $10 \times 10 \times 5$ to compute a uniform size for the next stage of the CAD system. We adopted the 3D affine transformation using cubic interpolation to resize the extracted cluster.

D. Candidate Augmentation

Prior to the FP reduction stage, the initially computed nodule candidates are augmented using a 3D affine transformation based rotation method. Each candidate was rotated, along its center-of-mass, at different degrees in the $x, y,$ and z directions, resulting in 150 angular rotated candidates for each original candidate. The reason for doing so is to overcome the generalization issue of the classifier and also to enhance the performance of the classifier.

IV. AUTOENCODER

An autoencoder is an unsupervised neural network that tries to learn high-level features and maps the output exactly as the input (feature mapping), resulting in the same size as the input [13]. The hidden layers (see Fig. 4) are individually trained in an unsupervised fashion to map the high-level features. A softmax layer can then be stacked with the autoencoders to perform the classification. Finally, all the layers are integrated to form a deep network and to be trained for the final time in a supervised fashion to improve the overall performance. The architecture of the proposed SSAE is illustrated in Fig. 3. During this study, several parameters (such as number of neurons in the hidden layers, learning rate, L2 weight regularization, numbers of encoders) were optimized. The scaled conjugate gradient descent algorithm is used for training.

A. Feature Mapping: Stacked Sparse Autoencoder (SSAE)

SSAE tries to find the optimal training parameter $\theta = \{W, b_h, b_x\}$ by minimizing the error between the input and the reconstructed output. Here, W is the weight parameter; b_h and b_x are the bias of each layer. After obtaining the optimal θ , the SSAE computes a function

$f: R^{d_x} \rightarrow R^{d_{h^{(2)}}}$ to map the intensities of the input cluster in a new feature representation $h^{(2)} = f(x) \in R^{d_{h^{(2)}}}$.

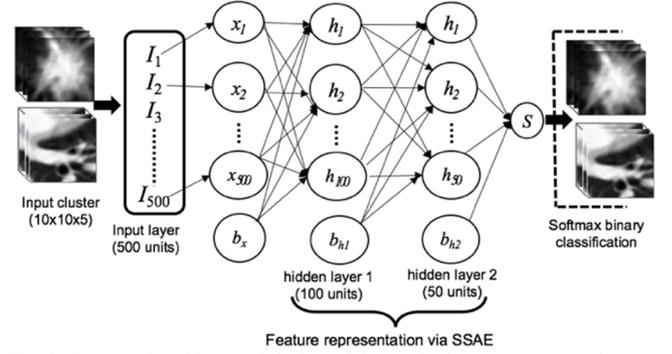


Fig. 3. Proposed architectural design of the stacked sparse autoencoder.

The voxel intensities of each training cluster $x(k)$ are represented as a low-level structured representation of input candidates in the first hidden layer. The second hidden layer ($h^{(2)}$) represents the high-level features. After the 2nd hidden layer, all the training clusters can be represented as $\{h^{(2)}(k), y(k)\}_{k=1}^N$; where k is the total number of cluster, and $\{h^{(2)}(k), y(k)\}$ represents the high level features and their label. Informatively, the label information y is not used during the SSAE training process. Next, the high level features and the label of each candidate are fed into the softmax classification stage (SCS).

B. FP Reduction: SCS

A softmax classifier is a supervised model used to solve the binary classification problems. Mainly, it aims at minimizing the cross entropy error loss using (2)

$$f_{w^{(3)}}(z) = \frac{1}{1 + \exp\left(-w^{(3)T}z\right)}, \quad (2)$$

where $f_{w^{(3)}}(z)$ is a cross-entropy function and z is the output of $h^{(2)}$ and T is the transposed matrix.

TABLE I. DISTRIBUTION OF CANDIDATES IN THE DIFFERENT FOLDS. FINAL SELECTED CANDIDATES ARE MARKED IN BOLD.

Candidates	Training	Validation	Test
nodule	190	41	42
aug (nod)	28,500	6,150	6,300
Non-nodule	35,803	7,850	8,020

C. FP Reduction: Training

During the training process, the parameters θ and the high level features (computed from $h^{(2)}$) are determined. The general procedure of feature mapping is shown in Fig. 4. The training of the classifier is done according to a 3-fold cross validation scheme. For each fold, 70 % of the data is used for training the model, 15 % for validating the parameters, and 15 % for testing the model. The statistics on the distribution of the detected candidates in the different folds are shown in Table I.

D. FP Reduction: Evaluation

The proposed CAD system is evaluated using two performance metrics: 1) area under the ROC curve (AUC) and 2) Competition performance metric (CPM) [14]. CPM computes the average sensitivity at seven false positive/scan points $\{1/8, 1/4, 1/2, 1, 2, 4, \text{ and } 8\}$ of the free-response receiver operating characteristic (FROC) curve. FROC is an extension of ROC. It is more sensitive at detecting small difference between performances and has higher statistical power [15]. Additionally, the 95 % confidence interval using bootstrapping with 1000 bootstrap is also computed.

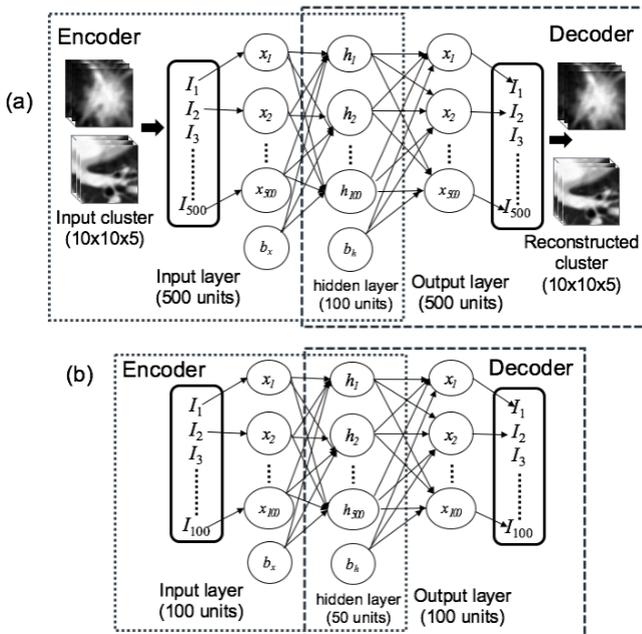


Fig. 4. Feature mapping process in hidden layers of the autoencoder.

TABLE II. CPM AT DIFFERENT FP/SCAN AND AREA UNDER CURVE (A_z) FOR BOTH DATASET.

Dataset	1/8	1/4	1/2	1	2	4	8	CPM	A_z
LIDC/IDRI	0.40	0.56	0.80	0.84	0.88	0.90	0.92	0.77	0.98
PCF	0.02	0.17	0.32	0.39	0.48	0.74	0.79	0.42	0.82

V. RESULTS AND DISCUSSION

On the 899 CT scans from the LIDC/IDRI dataset, the initial candidate detection stage yields a sensitivity of 94.4 % (273/289). As on the given data (see, Table I), the performance of the proposed CAD scheme is satisfactory in terms of sensitivity across different operating points of FROC. The FROC curve for both datasets (i.e. LIDC/IDRI and PCF) is shown in Fig. 5. The figure also shows the average CPM for both datasets. The 95 % bootstrap confidence interval for both datasets is also shown.

For quantitative evaluation, the CPM at different FP/scan points for each dataset are also shown in Table II. The CAD system achieved considerable overall sensitivity of 90 % at 4 FP/scan and AUC (A_z) of 0.98. Notably, the adequate performance of the classification stage is boosted by the sensitivity of the initial candidate detection stage. It reflects that the classification stage correctly classifies the 96.3 % (263/273) nodules from the initially detected candidates at 8 FP/scan. For the 33 CT scans of the independent dataset (PCF), the classification stage yielded a sensitivity of 74 % at 4 FPs/scan and AUC (A_z) of 0.82.

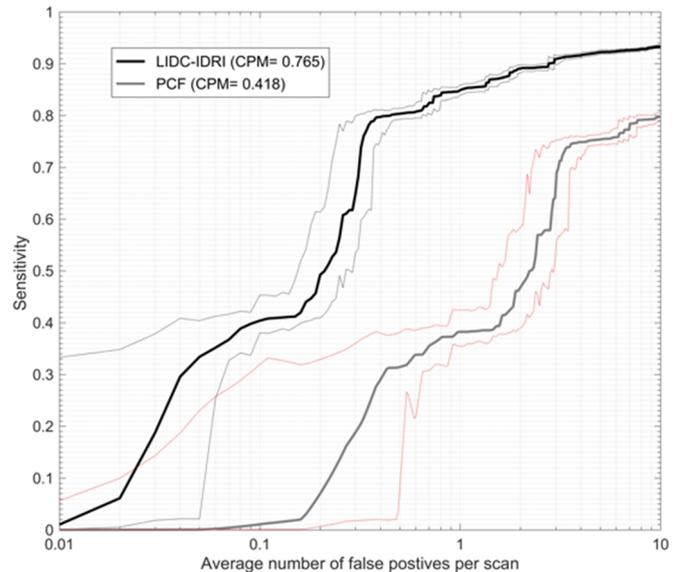


Fig. 5. FROC curve for the LIDC-IDRI dataset and the independent datasets (PCF). The two light gray curves and the two red curves show the 95 % bootstrap confidence intervals for the LIDC/IDRI and PCF datasets, respectively. The number of false positives are shown on a logarithmic scale.

It is seemingly challenging to directly compare the proposed CAD system with different existing CAD systems due to the high variance in the selection criteria of lesions in the training, validation and testing datasets. Additionally, performance evaluation on secluded datasets and insufficient information may also influence the direct comparison. However, we still compared our results with the previously reported CAD system in Table III.

TABLE III. STATISTICAL COMPARISON OF PREVIOUSLY REPORTED CAD SYSTEMS ON LIDC/IDRI DATABASE.

Authors	No of scans	Nodule size (mm)	No of nodules	Sensitivity (%)	FPS/scan
Setio <i>et al.</i> [4]	888	≥ 3	1186	90.1	4
Torres <i>et al.</i> [16]	949	≥ 3	1749	80	8
Brown <i>et al.</i> [17]	108	≥ 8	68	79.3	1
Tan <i>et al.</i> [18]	360	≥ 3	NA	83	4
Lu <i>et al.</i> [19]	294	≥ 5	631	85.2	4
Proposed method	849	≥ 10	289	90	4

In the future, it will be interesting to investigate the same methodology for the smaller size nodule candidates (benign candidates), which is a more complex problem in terms of features mapping. It will also be interesting to investigate the performance of the ensemble of the unsupervised classifiers for robust and reliable predictions.

VI. CONCLUSIONS

In this work, an unsupervised FP reduction step for automatic detection of large nodules in CT images is presented. For LIDC/IDRI, the proposed CAD system trained using an SSAE achieved a competitive sensitivity of 90 % at 4 FPs per scan. The proposed false positive reduction step could be integrated with the previous methods to further supplement the performance of the CAD

framework. The proposed CAD system can thus be considered as a decision aider in the lung nodule detection scenario. However, evaluation of the proposed CAD system on more datasets is still a foremost requirement for effective and reliable usability.

ACKNOWLEDGMENT

The authors acknowledge the National Cancer Institute for the creation of the free publicly available LIDC/IDRI database used in this study. Thanks are due to Cornell University for maintaining the public databases.

REFERENCES

- [1] R. Siegel, K. Miller, A. Jemal, "Cancer statistics, 2016", *CA: A Cancer Journal for Clinicians*, vol. 66, no. 1, pp. 7–30, 2016. [Online]. Available: <https://doi.org/10.3322/caac.21332>
- [2] D. Gierada, P. Pinsky, F. Duan, K. Garg, E. Hart, E. Kazerooni, H. Nath, J. Watts, D. Aberle, "Interval lung cancer after a negative CT screening examination: CT findings and outcomes in National Lung Screening Trial participants", *Eur. Rad.*, pp. 1–8, 2017.
- [3] A. A. A. Setio, C. Jacobs, J. Gelderblom, B. van Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic CT images", *Med. Phys.*, vol. 42, pp. 5642–5642, 2015. [Online]. Available: <https://doi.org/10.1118/1.4929562>
- [4] A. A. A. Setio, F. Ciompi, *et al.*, "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks", *IEEE Trans. Med. Imaging*, vol. 42, pp. 1–10, 2016. [Online]. Available: <https://doi.org/10.1109/TMI.2016.2536809>
- [5] B. van Ginneken, S. G. Armato, *et al.*, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study", *Med. Image Anal.*, vol. 14, pp. 7–22, 2010. [Online]. Available: <https://doi.org/10.1016/j.media.2010.05.005>
- [6] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, vol. 521, pp. 436–444, 2015. [Online]. Available: <https://doi.org/10.1038/nature14539>
- [7] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85–117, 2015. [Online]. Available: <https://doi.org/10.1016/j.neunet.2014.09.003>
- [8] S. G. Armato, G. McLennan, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans", *Med. Phys.*, vol. 38, pp. 915–931, 2011. [Online]. Available: <https://doi.org/10.1118/1.3528204>
- [9] A. Gupta, O. Martens, Y. Le Moullec, T. Saar, "A tool for lung nodules analysis based on segmentation and morphological operation", in *IEEE 9th Intelligent Symposium on Signal Processing (WISP)*, 2015, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/WISP.2015.7139186>
- [10] S. van Riel, C. Sanchez, *et al.*, "Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management", *Radiology*, vol. 277, no. 3, pp. 863–871, 2015. [Online]. Available: <https://doi.org/10.1148/radiol.2015142700>
- [11] PCF database. Online. [Available]: <http://www.via.cornell.edu/databases/lungdb.html>
- [12] A. Gupta, O. Martens, Y. Le Moullec, T. Saar, "Methods for increased sensitivity and scope in automatic segmentation and detection of lung nodules in CT images", in *IEEE International Symposium on Signal Processing and Information Technology*, 2015, pp. 375–380. [Online]. Available: <https://doi.org/10.1109/ISSPIT.2015.7394363>
- [13] J. Xu, L. Xiang, R. Hang, Q. Liu, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images", *IEEE Trans. Med. Imaging*, pp. 119–130, 2016. [Online]. Available: <https://doi.org/10.1109/TMI.2015.2458702>
- [14] M. Niemeijer, M. Loog, *et al.*, "On combining computer-aided detection system," *IEEE Trans. Med. Imag.*, vol. 30, pp. 215–223, 2011. [Online]. Available: <https://doi.org/10.1109/TMI.2010.2072789>
- [15] D. Chakraborty, "A status report on free-response analysis", *Radiation protection dosimetry*, vol. 139, pp. 20–25, 2010. [Online]. Available: <https://doi.org/10.1093/rpd/ncp305>
- [16] E. L. Torres, E. Fiorina, *et al.*, "Large scale validation of the MSL lung CAD on heterogeneous CT datasets", *Med. Phys.*, vol. 42, pp. 1477–1489, 2015. [Online]. Available: <https://doi.org/10.1118/1.4907970>
- [17] M. S. Brown, P. Lo, *et al.*, "Toward clinically usable CAD for lung cancer screening with CT", *European Radiology*, 2014. [Online]. Available: <https://doi.org/10.1007/s00330-014-3329-0>
- [18] M. Tan, R. Deklerck, J. Cornelis, B. Jansen, "Phased searching with neat in a time-scaled framework: Experiments on a computer-aided detection system for lung nodules", *Artificial Intelligence in Medicine*, 2013. [Online]. Available: <https://doi.org/10.1016/j.artmed.2013.07.002>
- [19] L. Lu, Y. Tan, L. Schwartz, "Hybrid detection of lung nodules on CT scan images", *Med. Phys.*, vol. 42, no. 9, pp. 42–54, 2015. [Online]. Available: <https://doi.org/10.1118/1.4927573>