

A Review on Automatic Lung Lesion Detection from Imaging Techniques

Raj Khatri

Dept. of Computer Science &
Engineering
SAM College of Engineering &
Technology, Bhopal, Madhya Pradesh,
India
rajkhatri786@gmail.com

Neelesh Jain

Dept. of Computer Science &
Engineering
SAM College of Engineering &
Technology, Bhopal, Madhya Pradesh,
India
neeshcme@gmail.com

Prateek Singhal

Dept. of Computer Science &
Engineering
SAM College of Engineering &
Technology, Bhopal, Madhya Pradesh,
India
prateeksinghal2031@gmail.com

Abstract— Lung cancer is a terrible condition that can be fatal to people. It is the type of cancer that is most frequently diagnosed worldwide and is potentially fatal. Treatment is a potential remedy that might save a person's life if it can be identified sooner. This condition may be identified using a variety of imaging modalities, and it can then be appropriately treated. But compared to other imaging modalities like X-ray, ultrasound, and many more, computed tomography, sometimes known as a CT scan, is a superior alternative for accurately identifying illness. It saves time for medical professionals and spares lives if a condition can be automatically diagnosed. A basic checkup may become quicker and easier to complete. This paper's goal is to examine several already-in-use technologies for the automated detection of lung cancer. Numerous studies have been conducted using machine learning or image processing methods including CNN, DNN, edge detection, and others. This essay's goal is to outline the shortcomings and shortcomings of numerous systems that exist but are not fully utilised.

Keywords— Lung Cancer Detection, Lesion Classification, Imaging Techniques, CNN, DNN, X-Ray, CT Scan.

I. INTRODUCTION

One of the causes of human fatalities is lung cancer. On the ground, it is difficult to tell when it first appears and begins to exhibit symptoms. Nevertheless, early infection discovery and treatment can reduce the fatality rate. There are several imaging techniques, but CT imaging is the best since it can distinguish between malignant and benign lung cancer. In the past, lung CT scan pictures could only be used by medical specialists to identify the condition. It took a long time, and the patient had to wait a day before getting his report. But now that it has been improved, it can be automatically detected using a variety of methodologies and certain pre-processing procedures [1]. Currently, automated diagnosis has emerged as a complement and a potential tool

to aid radiologists and physicians in accurately detecting cancer [2]. Numerous frameworks have been developed, and research into the early diagnosis of lung cancer continues. In some situations, some frameworks don't have adequate detecting precision, and some frameworks really need to be enhanced in order to achieve the highest level of accuracy, which tends to be 97 to 98%. To find lung cancer, image processing techniques and AI processes have been used. It has been thoroughly examined; the approach that is best for creating automated lung cancer diagnosis is discussed in the next section. On the odd occasion that lung cancer is identified early, the 5-year survival percentage for patients increases from 14 to 49%. X-rays can be more effective than CT scans, despite this. Due to the limited number of diagnostic strategies, a problem seemed to have converged because of the length of time needed to discover the existence of lung cancer.

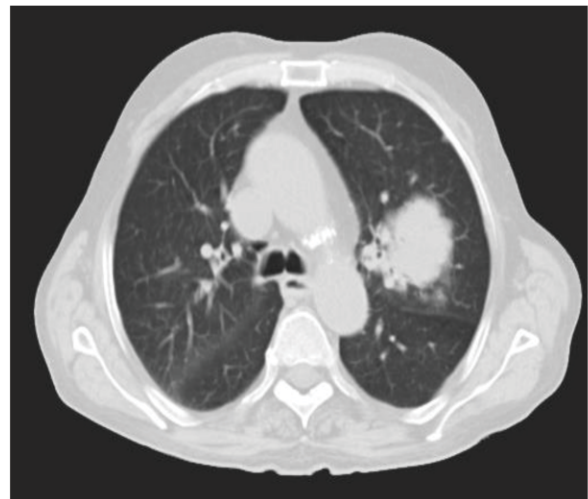


Fig. 1. Lung Cancer Lesion Detection in CT Scan Image [3]

The illness of abnormal cells multiplying and growing into a growth is known as lung cancer. The lymph fluid that surrounds the lung tissue and the circulatory system both allow cells to spread from the lungs. Because lymph moves normally, cancer cells often travel towards the chest, where they are most concentrated. When cancer cells move to various organs, this is called metastasis. Early cancer identification has a crucial role in preventing cancer cells from multiplying and spreading. Investigations into lung cancer have already been undertaken, for instance using microarray data and the bunching approach.

II. RELATED WORK

A. Related Works

EK-Mean Clustering is the foundation of the approach presented by P.B. Sangamithaa et al. [4]. This methodology divides the ROI first, followed by exams for impairment detection that look at the illness. Given that lung growth CT images have some features, such as variability in cancer appearance and ambiguous cancer limits, even with a few newly developed lung growth divisions and improving cancer division techniques, which are still fascinating. The growth division approach for CT images, which was developed to address this problem, separates non-upgrading lung tumours from lung tissues that had previously been clustered. The split of the CT images was initially accomplished using the K-Means clustering approach. The result is grouped using EK-Mean grouping. Furthermore, from the tumorous section of the Fuzzy Ek-Means portioned Image, the highlights like entropy, Contrast, Correlation, Homogeneity, and the area are retrieved. Grey Level Coherence Matrix (GLCM) is a measuring method used for highlight extraction.

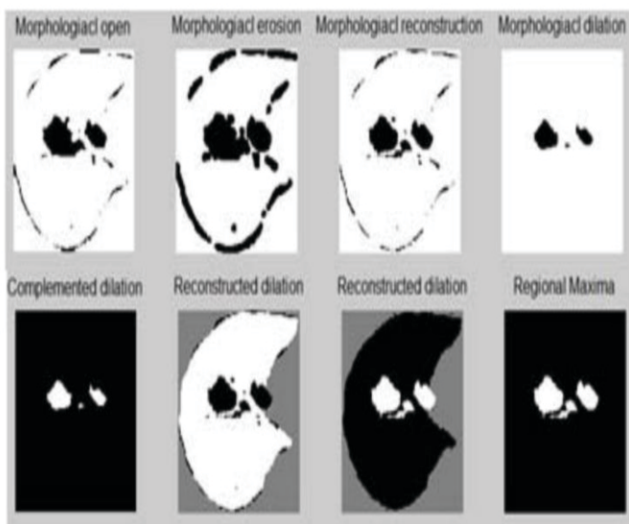


Fig. 2. EK-Means Algorithm [4]

Using a median filter and morphological operation as its foundation, Selin Uzelaltinbulat et al. [5] created a system. The work comprises unique image processing tools that, when gradually integrated into the system, successfully achieved the required goals. The division framework goes through several steps before achieving its ultimate goal, which is to segment the lung development. The first step is picture pre-processing, where several enhancement

techniques are used to improve and lower noise in photos. The subsequent stage is when the various components in the photos are divided into sections to define the cancer region. Since each image's development has unique dim levels, the stage edge was chosen in this case in a way that ensures the correct selection of all pictures. Here, erosion—an additional technique—was also used to remove noise from the thresholded image. Finally, the thresholded picture and the other image are subtracted to accurately portion the lung growth. In the area of automated lung cancer diagnosis, G. Niranjana et al. [6] studied a number of methods. According to a study of several studies on the approaches for dividing lung lesions, putting an edge to various CT images can separate the lung region from non-ROI. The least complicated and most effective strategy for lung division is staggered thresholding. It produces better results than other division approaches for lesion division. The lesion's form and surface components are used to characterise it. In light of the discussion above, it was decided that the component-based technique was preferable for categorization.

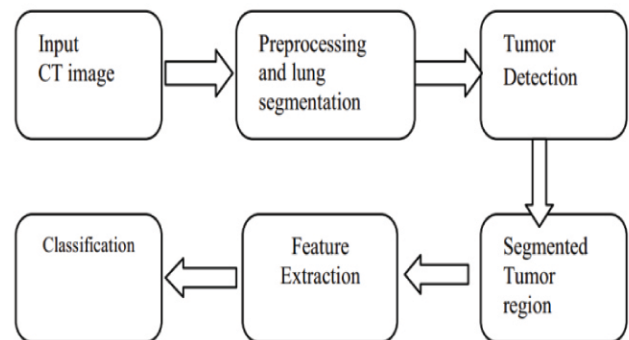


Fig. 3. Structure of Conventional System [6]

An Entropy Degrading-based system was proposed by Qing Wu et al. [7]. In this review, the author put out a method for detecting SCLC for the early diagnosis of malignant cancer using EDM AI calculations and vectorized histogram elements. Although it demonstrates that EDM has a logically excellent forecast, there is still much need for improvement before the computation can be applied in a clinical setting. The primary goal of this study is to develop a clinical dynamic framework that will enable radiologists to predict lung cancer detection from SCLC with processed tomography (CT) imaging with ease. A system based on the Gaussian Filter, Median Filter, and Watershed Segmentation was proposed by Suren Makaju et al. [8]. The suggested model is inaccurate and fails to identify the cells that are impacted by cancer. The suggested framework uses watershed division to identify the cancerous lesion from the lung CT scan picture and classifies the cancer cells as malignant or benign. The proposed model has a precision of 92%, which is greater than that of previous models, while the classifier has a precision of 86.6%. Overall, the suggested framework is better than the best model now in use, although it doesn't classify diseases into stages like cancer's stages I, II, III, and IV. Therefore, carrying out the arrangement in phases should be achievable as future degree development. Likewise, proper pre-processing and excellent noise removal may increase accuracy even more.

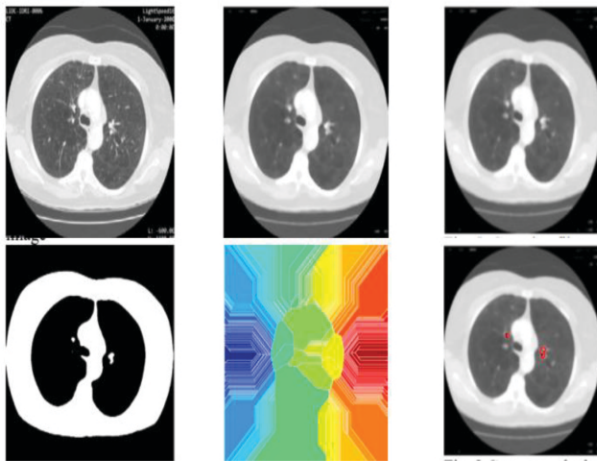


Fig. 4. System Diagnosis Phases [8]

A method based on DenseNet and Adaptive Boosting was proposed by SHANCHEN PANG et al. [9] with the use of computations from DenseNet and AdaBoost, the author of this study created a model of lung cancer CT scans. In order to improve the model's capacity for conjecture and prevent clustering findings from being biased, interpretation and transformation procedures are added to the initial dataset. DenseNet is then developed to handle the lung cancer datasets and characterise the gathered data at that point. Finally, adaboost computation is used to categorise findings in order to improve execution yet further. Results from the test indicate that the model performs a little bit better, and the test's precision was 89.85%. This method of lung cancer picture categorization can aid radiologists in identifying lung cancer later on in clinical practise. To manage the order of lung cancer, it may also use more CT scans of the disease. Here, data has been transformed to improve system performance. However, the densely linked network of DenseNet may make it more difficult and negatively impact the effectiveness of computing. Aman Agarwal et al. [10] proposed a system which is based on Alexnet CNN. In this study, the author sought to identify the cancer's malignant component using an image from a lung CT scan as input. Here, the network has been trained using diverse lung malignant pattern shapes and sizes from CT scanned pictures.

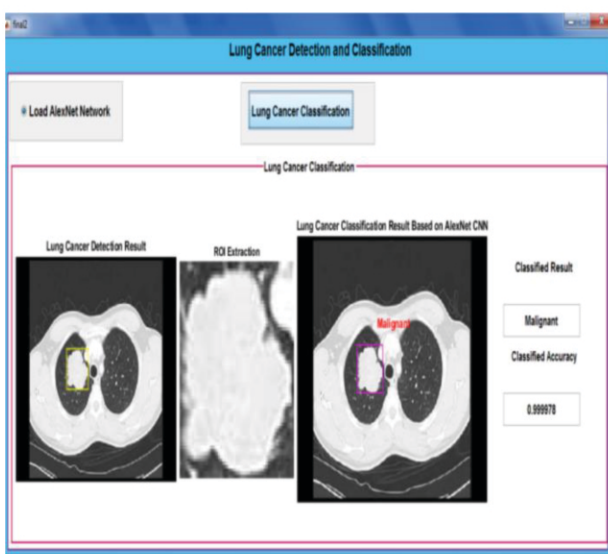


Fig. 5. System GUI [10]

It has been put into place based on the categorization procedure, which determines whether it is malignant or benign. Additionally, the cancer region is being extracted from the picture utilising different thresholding techniques, specific morphological methods, and Alexnet for the final classification. Here, the system's classification of malignant and non-cancerous pictures resulted in an accuracy rate of 96%. The performance of the system or network is weak since Alexnet is not a particularly deep model or network and may have trouble extracting the image's features. Zirong Li et al. [11] proposed a system which is based on ResNet. The author of this work described a method for utilising ResNet to identify density in a chest picture. The RCNN architecture has been shown to be a little quicker and to be able to provide results more efficiently. Component extraction, which involves taking a section of the network and using RESNET to demonstrate improved execution, has also been found to be more effective. Whether the picture depicts a malignant or healthy lung has been determined. It has been demonstrated that it performs better than Alexnet and contains slightly more convolutional layers than Alexnet. The issue with this article is that it uses chest X-rays rather than lung CT scan pictures. Lung cancer cannot be accurately diagnosed with an X-ray picture. The lung cancer's malignant region cannot be seen on an X-ray. As opposed to X-ray pictures, it is preferable to utilise CT scan images since they may generate a high degree of precision. Hui Shi et al. [12] proposed a system which is based on VGG16. In this study, 2 dimensional CNN was employed to identify lung cancer lesions and to determine the lung's overall health. Here, the system has been trained with VGG16, which uses 16 layers with unique lesion shapes and sizes. They use anchors, and anchors are in charge of predicting the location of the lesion and the system's diagnostic score. The system is said to be highly precise in detecting lesions of all shapes and sizes. However, the system employs subpar datasets, including Luna16, for its training and testing modules. It has averagely sensitive lesions, and this dataset cannot identify minor lesions. Here, the system had an accuracy rate of 82.62%, which is not very good compared to the other models. A system built on the Feed Forward ANN was proposed by Moffy Vas et al. [13]. This paper's modified methodology aims to provide a framework for lung cancer diagnosis. Here, the author's use of a median filter to reduce the clamour in the photographs worked out splendidly. Additionally, the system makes use of morphological techniques to further contribute to successful classification and segmentation results. ANN proved to be a respectable classifier with reasonable accuracy. The project's methodology produced an accuracy of 92% for the emergency clinic database. This framework aims to improve the lung cancer detection system's accuracy and speed. Additionally, it aids in early cancer detection. Here, the system reduced the size of the picture by converting all of the lung images into binary images, and it then used an erosion approach to remove noise from the image that was encountered during morphological operations. As blood arteries and veins are extremely similar to lesions, it removes background and foreground noise from the picture but may also destroy critical information.

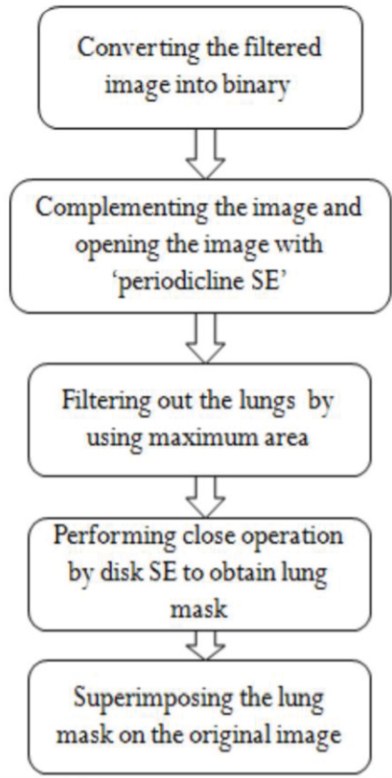


Fig. 6. Segmentation Stage [13]

Wadood Abdul et al. [14] proposed a system which is based on CNN model. System pertains the lesion from the structure and texture and train the network accordingly. Here, the system aims to improve the network's back propagation so that it can extract the malignant region from CT lung pictures more accurately. For testing, the author employed the LIDC-IDRI dataset, and the results were as expected. System evaluated its performance in comparison to the most recent algorithm from a number of angles. Author made the network more complex in order to obtain successful results, although doing so decreased computing speed and efficiency. D. Sharma et al. [15] proposed a system which is based on texture based analysis of cancerous cells. System uses various image pre-processing techniques for extracting the lesion and the location of the lesion. The system also employs local binary patterns for segmentation, although this is a fairly traditional technique that removes sensitive information from the picture while segmenting the image or region of interest. Local binary patterns erode and dilate the pixel in accordance with their comparison of the picture information with the threshold value. Support Vector Machine was employed to categorise the lesion at the last stage of diagnosis, however due to the usual method, SVM is ineffective with local binary patterns. The system obtained specific pictures of both normal and diseased lungs for the testing phase and extracted the histogram characteristics from each image. 80 photos altogether were utilised for categorization. True positive, True Negative, False Positive, and False Negative were the four parameters they employed. However, they make a 100% accuracy promise, which is completely unrealistic. Since no one can guarantee accuracy at 100%, it might eventually deteriorate as testing sample numbers rise. Not all types of photos or samples can be used effectively by a

system. When diverse patterns are received, it may become distracted and return a high false alert rate.

III. PROBLEM IDENTIFICATION

Many studies have attempted to extract the lesion from lung imaging and have had fair success with a limited amount of false alarms. An enhanced deep neural network and ensemble classifier based lung cancer detection system was proposed by P. Mohamed Shakeel et al. [16]. System attempted to address the shortcomings of deep neural networks, which demand large amounts of memory to hold the system's weight model. However, deep learning is still necessary for deep neural networks to provide effective results. The system also employs ensemble classifiers like GAWA, ACO, PSOMS, and HSOGR, which integrate the results from four classifiers. But compared to other classifiers, SVM is far more advanced and effective. But compared to other classifiers, SVM is far more advanced and effective. SVM has a good prediction rate and can handle both linear and nonlinear data. The best classifier for any illness linked to image processing is considered to be SVM. In the sphere of medical science, accuracy is crucial since a false alarm rate with a serious patient might result in the loss of a human life. Table I shows the comparison among various methods which have been used to implement the automatic lung cancer detection. Table shows the problem which has been identified at each model of the system and where it suffers and what kind of perspective should be retain in mind to develop an ideal system that can diagnose lung cancer whether it is malignant or benign from lung CT scan images. System require a better approach that can train the system with less filters or light weight filter with proper training through which percision can be calculated high with less false alarm rate.

TABLE I. RESULT COMPARISON

	Method	Modality	Precision
Zirong Li [11]	ResNet	X-Ray	0.5238
Hui Shi [12]	VGG-16	CT	0.826
Shanchen Pang [9]	DenseNet	CT	0.8985
P.B.Sangamithraa [4]	EK-Means	CT	0.9087
Moffy Vas [13]	FF-ANN	CT	0.92
Aman Agarwal [10]	AlexNet	CT	0.96
P. Mohamed Shakeel [16]	DNN & Ensemble Classifier	CT	0.962

IV. CONCLUSION & FUTURE SCOPE

Automatic lung cancer identification is a procedure that identifies lesions from lung scans and extracts the region that indicates the size of the lesion, assisting medical personnel in

assessing the severity of the condition. This essay examined a number of studies before coming to a conclusion on an issue with the existing situation. Both a poor network and a heavy network cause issues while training and assessing the data. If machine learning is an option, the network should be light weighted with fewer filters or light weighted filters employed. Each layer should provide effective results, and the system should adhere to a high degree of accuracy at the conclusion of the levels. The system should be extremely sensitive to the data if it employs edge detection or any other image processing method since clinical data is extremely delicate and shouldn't be compromised in any way.

REFERENCES

- [1] A.M. Gindi, T.A. Al Attiatalla, M.M. Sami "A Comparative Study for Comparing Two Feature Extraction Methods and Two Classifiers in Classification of Earlystage Lung Cancer Diagnosis of chest x-ray images." *Journal of American Science*, 10 (6) (2014), pp. 13-22
- [2] K. Suzuki, M. Kusumoto, S.I. Watanabe, R. Tsuchiya, H. Asamura "Radiologic classification of small adenocarcinoma of the lung: radiologic-pathologic correlation and its prognostic impact" *The Annals of Thoracic Surgery*, 81 (2) (2006), pp. 413-419
- [3] Su CC, et al "Impact of low-dose computed tomography screening for primary lung cancer on subsequent risk of brain metastasis" *J Thorac Oncol* 2021; DOI: 10.1016/j.jtho.2021.05.010.
- [4] P. B. Sangamithraa and S. Govindaraju, "Lung tumour detection and classification using EK-Mean clustering," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 2201-2206, doi: 10.1109/WiSPNET.2016.7566533.
- [5] Selin Uzelaltinbulat, Buse Ugur, Lung tumor segmentation algorithm, *Procedia Computer Science*, Volume 120, 2017, Pages 140-147.
- [6] G. Niranjana and M. Ponnaivaikko, "A Review on Image Processing Methods in Detecting Lung Cancer Using CT Images," 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), 2017, pp. 18-25, doi: 10.1109/ICTACC.2017.16.
- [7] Q. Wu and W. Zhao, "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm," 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC), 2017, pp. 88-91, doi: 10.1109/ISCSIC.2017.22.
- [8] Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A. Elchouemi, Lung Cancer Detection using CT Scan Images, *Procedia Computer Science*, Volume 125, 2018, Pages 107-114.
- [9] Pang, Shanchen; Zhang, Yaqin; Ding, Mao; Wang, Xun; Xie, Xianjin (2020). A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting. *IEEE Access*, 8(), 4799-4805.
- [10] Agarwal, Aman & Patni, Kritik & Devarajan, Rajeswari. (2021). Lung Cancer Detection and Classification Based on Alexnet CNN. 10.1109/ICCES51350.2021.9489033.
- [11] Li, Zirong; Li, Lian (2017). [IEEE 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) - Kansas City, MO (2017.11.13-2017.11.16)] 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) - A novel method for lung masses detection and location based on deep learning. , (), 963-969.
- [12] Shi, Hui; Peng, Zhenwei; Wan, Honglin (2019). [IEEE 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT) - Jinan, China (2019.10.18-2019.10.20)] 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT) - Pulmonary Nodules Detection Based on CNN Multi-scale Feature Fusion. , (), 86-90.
- [13] M. Vas and A. Dessai, "Lung cancer detection system using lung CT image processing," 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017, pp. 1-5, doi: 10.1109/ICCUBEA.2017.8463851.
- [14] W. Abdul, "An Automatic Lung Cancer Detection and Classification (ALCDC) System Using Convolutional Neural Network," 2020 13th International Conference on Developments in eSystems Engineering (DeSE), 2020, pp. 443-446, doi: 10.1109/DeSE51703.2020.9450778.
- [15] D. Sharma and G. Jindal, "Computer aided diagnosis system for detection of lung cancer in CT scan images," *Int. J. Comput. Elect. Eng.*, vol. 3, no. 5, pp. 714-718, Sep. 2011.
- [16] [4] Shakeel, P.M., Burhanuddin, M.A. & Desa, M.I. Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. *Neural Comput & Applic* (2020). <https://doi.org/10.1007/s00521-020-04842-6>.
- [17] R. K. Samala, H.-P. Chan, C. Richter, L. Hadjiiski, C. Zhou, and J. Wei, "Analysis of deep convolutional features for detection of lung nodules in computed tomography," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, p. 109500Q, International Society for Optics and Photonics, 2019.
- [18] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Conference on Computer and Robot Vision*, pp. 133-138, 2015.
- [19] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Medical Physics*, vol. 38, no. 10, pp. 5630-5645, 2011.
- [20] A. A. A. Setio, A. Traverso, T. de Bel, M. S. N. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, R. van der Gugen, P. A. Heng, B. Jansen, M. M. J. de Kaste, V. Kotov, J. Y.-H. Lin, J. T. M. C. Manders, A. S'ona-Mengana, J. C. Garcia-Naranjo, E. Papavasileiou, M. Prokop, M. Saletta, C. M. Schaefer-Prokop, E. T. Scholten, L. Scholten, M. M. Snoeren, E. L. Torres J. Vandemeulebroucke, N. Walasek, G. C. A. Zuidhof, B. van Ginneken, and C. Jacobs, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1-13, 2017.