Migration Letters
Volume: 21, No: S4 (2024), pp. 1167-1174
ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

A Comparative Analysis of Lung Image Classification using different Classification Techniques

K. Vasanthi[1], Dr. K. Karthikeyan[2]

Abstract

The urgent need for precise and prompt detection is further underscored by the fact that lung cancer is still the largest killer in the cancer category. In order to identify and categorize lung scans as normal or abnormal, this study suggests three separate image categorization techniques. Before anything else, lung images are preprocessed with Histogram Equalization to make them clearer and noise-free. After that, we use feature extraction methods, and then we use RF with Generalized Discriminant Analysis (GDA) to pick the best subset of features. During the classification stage, four classifiers are used: K-Nearest Neighbour (KNN), Naïve Bayes (NB), Neural Network (NN), and Random Forest (RF). Based on the comparison, the RF-GDA approach achieves better classification accuracy than the state-of-the-art methods. Second, computed tomography (CT) scans can be used to find early stages of lung cancer. Before feature extraction, the CT images are preprocessed using isotropic diffusion. The HOG approach is then used. One way to train a Neuro Fuzzy Classifier with Binary Cuckoo Search (NFCBCS) to better detect cancerous growth is to employ adjusted weights. The results of this study provide credence to the idea that image processing methods could be useful in the fight against lung cancer. The third tactic is the DF-PTDNN CAD model, which stands for Deep Features with Parameter-Tuned Deep Neural Networks. Prior to feature extraction using DenseNet121 and Local Binary Patterns (LBP), the model undergoes contrast augmentation and pre-processing based on Gaussian filtering. Various forms of lung cancer are detected and categorized by a deep neural network that employs a soft max classifier. Hyperparameters are fine-tuned using the quasi-oppositional moth swarm optimization (QOMSO) method, which yields remarkable results: sensitivity of 98.81%, specificity of 97.41%, and accuracy of 99.85%.

Keywords: *Computed Tomography, Deep Features with Parameter-Tuned Deep Neural Network, Generalized Discriminant Analysis, Lung cancer.*

1. INTRODUCTION

The leading cause of mortality worldwide is lung cancer. The majority of individuals are diagnosed with incurable lung cancer in the postponed phase. In order to increase survival rates for individuals with lung cancer, early identification is crucial [1]. In order to retrieve data from picture databases, image mining makes use of a number of methods, including

[1] Research Scholar(Ph.D)., PG and Research Department, Pioneer College of Arts and Science, Coimbatore, Tamilnadu, India.
[2] Assistant Professor of Computer Science, Puratchi Thalaivi Amma Government Arts and Science College, Palladam, Coimbatore, Tamilnadu. India

picture categorization, picture grouping, and picture representation based outline and association rule mining. Analysis of the dataset's lung cancer images using image classification is crucial to the study (2) to (3).

Uncontrolled cell development in one or both lungs is known as malignant pulmonary growth. A person's risk of developing a serious illness increases dramatically when they smoke [4]. The two most common types of lung tumours produced by cancer are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [5]. There are anywhere from twenty thousand to twenty-five thousand therapies in a single human cell. All living things contain these components. As ribonucleic acid (RNA) models the arrangement of deoxyribonucleic acid (DNA) and subsequently changes it into a protein group, the central rule of life explains the expression of quality [6-7]. The development of DNA microarrays has made it possible to assess the visual quality of individual cells and tissues. Results from tumour type classifications and evaluations of malignant growth have been greatly improved since the discovery of DNA chips [8-9]. The complete treatment of the condition is still an open question, despite the availability of numerous remediation procedures such as chemotherapy, radiotherapy, and clinical practice. Researchers and physicians are increasingly concerned about patients' inconsistent infections, susceptibility to medications, and adverse drug reactions [10–11]. One contributing element could be the delayed identification of the appropriate drug. Factors such as heredity, qualitative communication, epigenetics, metagenomics, and natural elements are among those that should be considered by a pharmaceutical system when considering genetic markers and the drug's intended usage [12].

As a straightforward and efficient dimensionality reduction model in image processing (IP), feature extraction is the job at hand. The non-intrusive nature of CT imaging is one of its most remarkable features. You might think the improved is strange compared to parallel imaging modules [13]. Input data connected to the reduction operation is used to filter the decided features set. Then, Support Vector Machines (SVMs) are given degraded characteristics so that they can perform testing and training [14]. Neural Network (NN) methods utilizing binarization IP are employed for the purpose of lung cancer classification [15]. Prior research on lung cancer categorization has used NN in some cases, and it has shown to be very accurate. The first definition of SVM is a globally applicable learning scheme that uses statistical learning hypothesis [16]. Although the aforementioned methods can extend life expectancy by detecting lung cancer in its early stages, they are more expensive. Therefore, more effective infection cures can be achieved with the help of early cancer prediction. This led to the development—and ongoing maintenance—of a requirement for lung cancer prediction models [17–18].

2. BACKGROUND STUDY

Demir, F. et al. [2] the automatic identification of lung disorders was the main emphasis of this work. This was a crucial issue in public health. Despite a dearth of research on the subject, lung sound classification has failed to take advantage of difficult data sets that include noises, background sounds, and different sampling frequencies. For the most part, traditional methods were also employed. In order to overcome the recognition issue of lung sounds, cutting-edge deep learning was employed to enhance classification performance.

Kaznowska, E. et al. [4] Finding out whether SCC and ACC were different was the driving force for this research. The authors also aimed to determine whether FTIR spectroscopy could distinguish between the two subtypes of lung cancer and measure the severity of each. The author discovered that infrared spectroscopy did provide the biochemical data necessary to distinguish between cancer types and grades of malignancy.

Lakshmanaprabu, S. et al. [6] the research looked at the use of a big data map-reduce system for SIoT categorization. In this instance, the author has eliminated superfluous database records by employing the Gabor filter. To further improve the work's efficiency, Hadoop Map Reduce has been investigated for data mapping. From the database, the most desirable attributes were selected using the data classification method. In this study, the EHO has been used to choose the most relevant subset of attributes from each dataset. Finally, the LK-SVM classifier was proposed using the expression data; it would sort the data into two classes, 1 and 2.

Lu, Z. et al. [8] a novel DICFM method was proposed for the purpose of investigating the link between claudin-7 expression and the advancement of lung cancer. The author pioneered the use of the fragmentation index, which differs from earlier research. Using the degree of dispersion of the spatial structure of claudin-7 expression as a discriminant characteristic, these authors results show that healthy lung tissue can be distinguished from malignant lung tissue.

Ren, Y. et al. [10] Findings from this study support the accuracy with which these authors novel manifold learning system categorises lung nodules. These authors intention in developing this MCR-DNN was to enhance classification accuracy while reducing the impact of over-fitting caused by insufficient data. This will allow us to classify images directly on a manifold while still keeping their actual structure. The proposed framework outperforms a conventional deep learning-based classification method that relies on a network built by just removing the manifold constraints in MCR-DNN, according to experimental results.

Ü. Özsandıkcıoğlu et al. [12] two different breath samples were taken and analysed throughout the experiment. The study's breath samples came from both healthy people and those who had lung cancer. The results show that the classification process was enhanced when the feature matrix was reduced in dimension using the principal component analysis (PCA) method. In addition, the results clearly show how the classification system was used. This exploratory study aims to determine the feasibility of using breath analysis in the early detection of lung cancer.

3. MATERIALS AND METHODS

In this study, we used a comprehensive methodology to develop and evaluate three distinct image classification systems for the aim of early identification and classification of lung cancer. For preprocessing, feature extraction, and classification, these methods employ cutting-edge machine learning and image processing techniques.

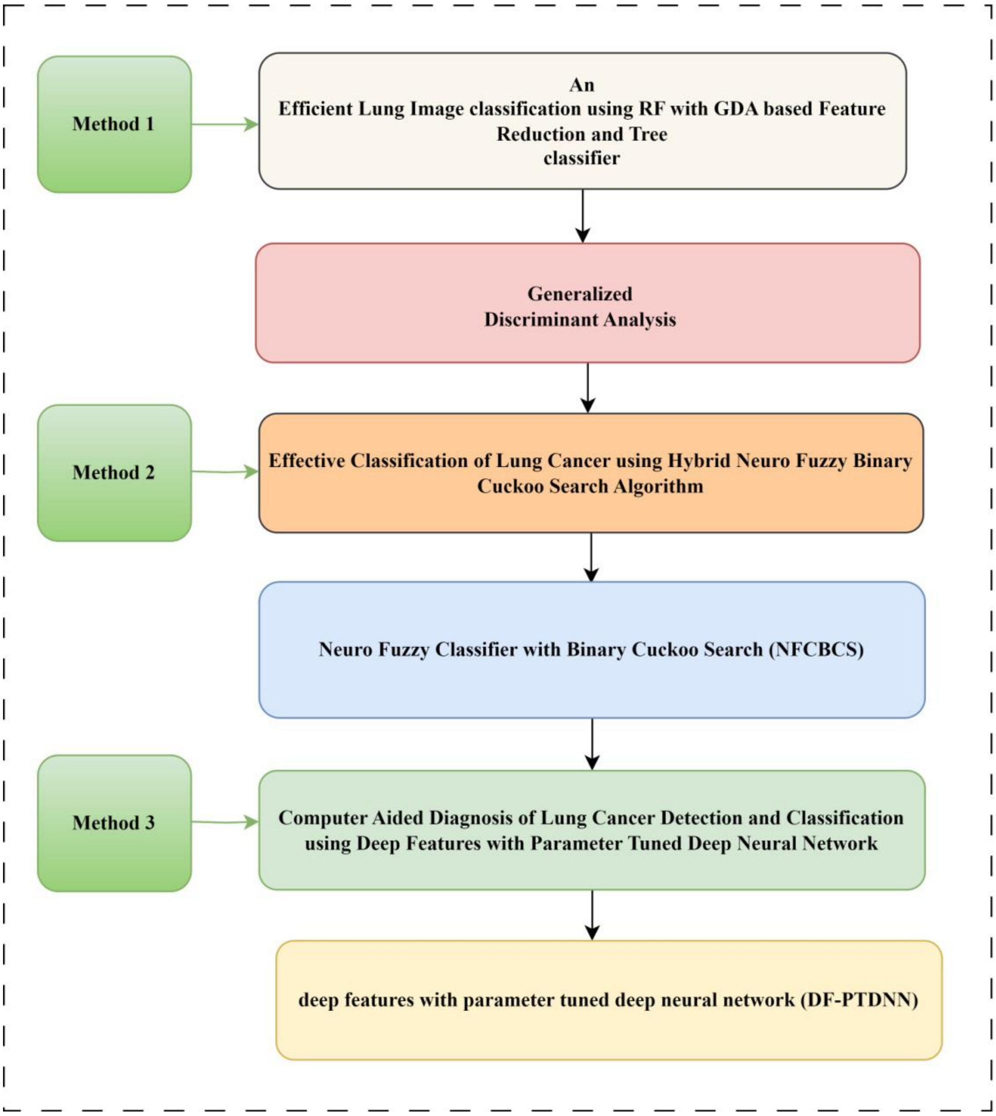

Figure 1: Proposed work flow architecture

3.1 RF with Generalized Discriminant Analysis (GDA)

Generalised Discriminant Analysis (GDA) is a statistical technique for dimensionality reduction and classification that is utilized in pattern recognition and machine learning. An improvement on the original Linear Discriminate Analysis (LDA), it gives a structure for determining which linear feature combinations may distinguish between dataset classes the most effectively. Maximizing the separation between classes and minimizing the dispersion inside each class is the fundamental purpose of GDA. This is accomplished by solving a generalized eigenvalue problem, where the directions of maximum separation are described by the eigenvalues and eigenvectors. By applying GDA to the input data, discriminative features can be extracted, allowing for the categorization of images into predetermined categories in the context of image classification. For medical image analysis of diseases like lung cancer, among others, it provides a potent instrument for improving the precision and efficacy of classification algorithms, which is especially helpful when working with high-dimensional datasets.

If you're facing a classification problem with several classes or feature space overlap, Generalised Discriminant Analysis is the way to go. GDA is generally designed for non-linear classification based on the kernel function$\emptyset$. The step by step procedure in the GDA feature reduction is deliberated as follows:

Assume the original space is S and then transform the original space to new high dimensional feature space.

$T: \emptyset: S \rightarrow T$ ------------- (1)

From the database, the non-linearly mapped data is calculated as in two ways: namely within the class data and between the class and scatter matrix using equation (4) and (5),

$G^{\varphi} = \sum_{k=1}^{k} \quad N_k n_k^{\varphi} \left(n_k^{\varphi}\right)^T$ ----------------- (2)

$H^{\varphi} = \sum_{k=1}^{k} \quad \sum_{s \epsilon s_k}^{k} \quad \varphi(S)\varphi(S)^T$ ------------------ (3)

Where, N_k represents the number of samples in the $S_k$in T and $n_k^{\varphi}$is the mean of S_k and K is a vector of weights. Projection Matrix Ratio: The main objective of GDA is to maximize the projection matrix ratio by finding

3.2 Neuro Fuzzy Classifier with the aid of Binary Cuckoo Search

A neuro-fuzzy classifier is employed for the purpose of organising lung pictures. By highlighting characteristics such as Ag, Ap, Hm, Cn, and Cr, the neuro-fuzzy classifier is able to classify all provided lung pictures into two categories: regular and irregular. Using building blocks borrowed from neural networks, the neuro-fuzzy classifier constructs a fluff-based framework. Computational learning produces data from close by and displays changes to the equilibrium system from close by. Instead of employing the parts of the arrangement independently, a neuro-fuzzy structure often builds all the more significant arrangements.

Membership functions (MF) reorganise individuals from each component into distinct classes based on the values in the information, which is then interpreted by grouping the ratings into structure Ag, Ap, Hm, Cn, and Cr. By excluding closed and interconnected data from the characteristics of the MF classes, the neuro-fuzzy classifier ensures the utmost accuracy during the order phase. Five lines and two sections make up the partial grid; the lines' sizes are proportional to the highlights' and the sections' to the classes'.

$X = [x_1, x_2, x_3, x_4, x_5]$ ------------- (4)

A kind of X –type MF is used as a partial capability for group images. This is a range of work with a structure such as Gaussian / exponential efficiency. - X –type fuzzifier (m) has a parameter that can be balanced by the needs of the problem. This limits the projecting ability by selecting the correct estimation of fuzzifier m, and it provides significant adaptive ability to classify images.

3.3 Deep features with parameter tuned deep neural network

An enhanced CAD model for early detection of lung cancer is presented by Deep Features with Parameter-Tuned Deep Neural Network (DF-PTDNN). The preprocessing improves picture quality and begins with a Gaussian filter and contrast enhancement. To extract comprehensive features that capture both local patterns and high-level representations, the model's novel technique mixes DenseNet121 features with Local Binary Patterns (LBP). A deep neural network (DNN) classifier uses a softmax layer as its compass to decipher complex data patterns. Crucially, the model is optimized for high sensitivity, specificity, and accuracy by means of hyperparameter fine-tuning using the quasi-oppositional moth swarm optimization (QOMSO) technique. The DF-PTDNN performed exceptionally well in the evaluation, with sensitivity at 98.81%, specificity at 97.41%, and accuracy at 99.85%, demonstrating its usefulness as a precise tool for the early detection of lung cancer.

$C_{adaptive}(w,b) = C(w,b) + \beta \sum_{n=1}^{s2} \quad KL(\rho // \rho_n)$ ---------------- (5)

Since neurons aren't actively processing any data during the first stage of a DNN's training, the activation function measure is close to zero. A greater activation function is being penalized. The procedure is carried out using different forms of the average activation function measure. The function of penalty is expressed in Eq. (6)

$$P_{penalty} = \sum_{n=1}^{s_2} \quad \rho//\rho_n \text{ ----------------- } (6)$$

Where, $s2$ implies overall count of secret layer neurons, KL (.) denotes a Kullback–Leibler divergence (KL divergence) and implied as:

$$KL(\rho//\rho_n) = \rho_{log}\frac{\rho}{\rho_n} + (1-\rho)\log\frac{1-\rho}{1-\rho_n} \text{ ---------------- } (7)$$

The enhanced divergence, which is actually an adaptive constant, is denoted by $(\rho//\rho n)$=0 when ρn=p. Applying a cost function achieves the desired result. "H" denotes the penalty weight that is implied by KL divergence. Given the elevated cost function, the examination of weight $'w\ '$ and biases $'b'$ is deemed crucial; thus, the aforementioned constraints are utilized to establish a connection between them in this regard. Excess optimizing concerns are addressed and the end conclusion of a similar problem is assumed in order to reduce the expression $aadapbia$ (w,). It implies the optimized model for iterative inform bias and weight which is referred as:

4. RESULTS AND DISCUSSION

This section presents the study's findings and their consequences, explaining the results of the suggested picture categorization systems and their relevance to the early identification of lung cancer. Performance measures, comparative analyses, and the possible impact of the study's advanced techniques are the main points of discussion when it comes to the results.

Table 1: Classification performance metrics comparison

Models	Sensitivity	Specificity	Accuracy
MLP	77.00	72.00	82.00
ODNN	96.20	94.20	94.56
RF-GDA	96.34	95.03	97.99
NF-CBCS	97.57	91.44	99.73
DF-PTDNN	98.81	97.41	99.85

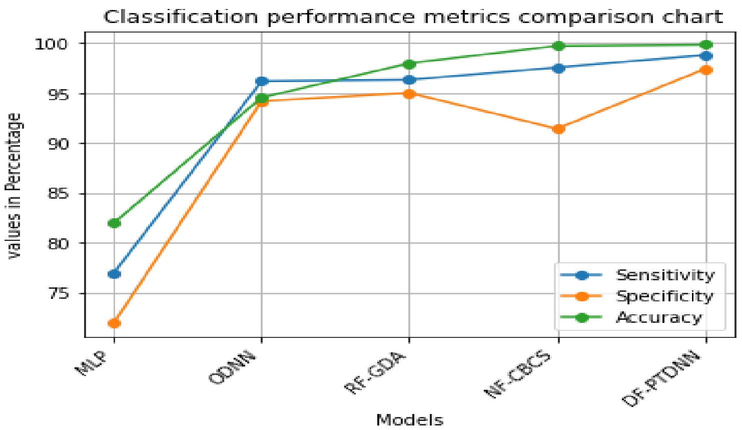

Figure 2: Classification performance metrics comparison chart

The table 1 and figure 2 shows the performance metrics—sensitivity, specificity, and accuracy—across different models employed for a particular task. Among the models, "DF-PTDNN" emerges as the top-performing model, showcasing exceptional results with the highest sensitivity (98.81%), specificity (97.41%), and accuracy (99.85%). "NF-CBCS" also demonstrates robust performance, particularly excelling in accuracy (99.73%) and boasting a high sensitivity (97.57%). "RF-GDA" follows closely, achieving notable sensitivity (96.34%), specificity (95.03%), and accuracy (97.99%). "ODNN" exhibits strong performance, notably in sensitivity (96.20%) and specificity (94.20%). On the other hand, "MLP" shows relatively lower performance across all metrics, suggesting potential areas for improvement. These findings underscore the varying efficacy of different models in the context of the specific task, emphasizing the importance of selecting models based on the desired balance of sensitivity, specificity, and accuracy for a given application.

5. CONCLUSION

Better early identification of lung cancer is critically needed, as it is a primary cause of cancer-related deaths. All three of these picture classification algorithms have the ability to accurately identify and categories lung anomalies, which is a testament to the power of modern image processing and machine learning methods. The first method outperforms the state-of-the-art by combining Histogram Equalization, Generalized Discriminant Analysis (GDA), and Random Forest (RF) classification. This proves that robust classification methods, feature extraction, and preprocessing all work together effectively. The second approach involves utilizing computed tomography (CT) scans to detect early-stage lung cancer. Applying anisotropic diffusion preprocessing, the Histogram of Oriented Gradient (HOG) algorithm, and the Neuro Fuzzy Classifier with Binary Cuckoo Search (NFCBCS) algorithm demonstrates the potential of these technologies to boost the sensitivity of lung cancer diagnosis. The third and most recent method presents the DF-PTDNN model, which stands for Deep Features with Parameter-Tuned Deep Neural Network. This model accomplishes outstanding sensitivity, specificity, and accuracy by means of contrast augmentation, Gaussian filtering, and the use of Local Binary Patterns (LBP) and DenseNet121. A key component of the suggested CAD model's overall performance is the hyperparameter tuning procedure known as quasi-oppositional moth swarm optimization (QOMSO).

Bibliography

Ahmad P. Tafti*1, Fereshteh S. Bashiri, "Diagnostic Classification of Lung CT Images using Deep 3D Multi-Scale Convolutional Neural Network", In Proceedings of IEEE International Conference on Healthcare Informatics, pp. 412 414, 2018.

Demir, F., Sengur, A., & Bajaj, V. (2019). Convolutional neural networks based efficient approach for classification of lung diseases. Health information science and systems, 8(1), 4.

K. Vasanthi, N. Balakumar, Effective Classification of Lung Cancer Using Hybrid Neuro Fuzzy Binary Cuckoo Search Algorithm, Jour of Adv Research in Dynamical & Control Systems, Vol. 12, 04-Special Issue, 2020

Kaznowska, E., Depciuch, J., Łach, K., Kołodziej, M., Koziorowska, A., Vongsvivut, J., Cebulski, J. (2018). The classification of lung cancers and their degree of malignancy by FTIR, PCA-LDA analysis, and a physics-based computational model. Talanta, 186, 337–345.

L. Anifah, Haryanto, R. Harimurti, Z. Permatasari, P. W. Rusimamto and A. R. Muhamad, "Cancer lungs detection on CT scan image using artificial neural network backpropagation based gray level coocurrence matrices feature," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, 2017, pp. 327-332.

Lakshmanaprabu, S. K., Shankar, K., Khanna, A., Gupta, D., Rodrigues, J. J., Pinheiro, P. R., & De Albuquerque, V. H. C. (2018). Effective Features to Classify Big Data Using Social Internet of Things. IEEE Access, 6, 24196-24204.

Li, J., Wang, Y., Song, X., & Xiao, H. (2018). Adaptive multinomial regression with overlapping groups for multi-class classification of lung cancer. Computers in Biology and Medicine, 100, 1–9.

Lu, Z., Liu, Y., Xu, J., Yin, H., Yuan, H., Gu, J., …Xie, B. (2018). Immunohistochemical quantification of expression of a tight junction protein, claudin-7, in human lung cancer samples using digital image analysis method. Computer Methods and Programs in Biomedicine, 155, 179–187.

N. S. Nadkarni and S. Borkar, "Detection of Lung Cancer in CT Images using Image Processing," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 863-866.

Ren, Y., Tsai, M.-Y., Chen, L., Wang, J., Li, S., Liu, Y., … Shen, C. (2019). A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification. International Journal of Computer Assisted Radiology and Surgery. doi:10.1007/s11548-019-02097-8

S. Rattan, S. Kaur, N. Kansal, J. Kaur, An optimized lung cancer classification system for computed tomography images, in: Image Information Processing (ICIIP), 2017 Fourth International Conference on, IEEE, 2017, pp. 1–6.

Ü. Özsandıkcıoğlu, A. Atasoy and Ş. Yapıcı, "Hybrid Sensor Based E-Nose For Lung Cancer Diagnosis," 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, 2018, pp. 1-5.

Vasanthi, K. and Kumar, N.B., 2019. An Efficient Lung Image Classification Using GDA Based Feature Reduction and Tree Classifier. In Handbook of Multimedia Information Security: Techniques and Applications (pp. 645-666). Springer, Cham.

W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-5.

Wei, G., Ma, H., Qian, W., Han, F., Jiang, H., Qi, S., &Qiu, M. (2018). Lung nodule classification using local kernel regression models with out-of-sample extension. Biomedical Signal Processing and Control, 40, 1–9.

Y. Xie, J. Zhang, Y. Xia, M. Fulham, Y. Zhang, Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct, Information Fusion 42 (2018) 102–110.

Zia urRehman, M., Javaid, M., Shah, S. I. A., Gilani, S. O., Jamil, M., & Butt, S. I. (2018). An appraisal of nodules detection techniques for lung cancer in CT images. Biomedical Signal Processing and Control, 41, 140–151.

Zuo, W., Zhou, F., Li, Z., & Wang, L. (2019). Multi-resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection. IEEE Access, 1–1. doi:10.1109/access.2019.2903587