

Performance Analysis of Various Classification Algorithm for Lung Disease Detection Using Genetic Algorithm

A.Tamilmozhi

Research Scholar, PG & Research Department of Computer Science, Raja Doraisingam Govt. Arts College,
Sivagangai, TamilNadu, India.

Dr.N.Sujatha

Assistant Professor, PG & Research Department of Computer Science, Raja Doraisingam Govt. Arts College,
Sivagangai, TamilNadu, India.

Abstract – In recent days, detecting and classifying the lung disease in a Computed Tomography (CT) image is an important and demanding task. For this purpose, many classification techniques are proposed in the existing works. But, it has some drawbacks such as, inaccurate, not efficient and misclassified results. In order to overcome all these issues, a new classification technique is proposed in this work for lung disease detection and classification. At first, the given lung image is preprocessed by using the median filtering technique to remove the unwanted noise in the image. Then, the features of the filtered image will be segmented by applying the morphological, Fusion Mean Absolute Deviation (MAD) and Walsh-Hadamard Transform (WHT) techniques. The features of the segmented image are extracted by applying the Gabor transformation technique. Then, the optimal features of the image are selected with the help of the Genetic Algorithm (GA). Finally, the normal and abnormal images are classified by applying the classification technique. In this work, there are three different classification techniques includes, Support Vector Machine (SVM), Naïve Bayes (NB) and Back Propagation (BP) are applied. The main intention of this paper is to select the suitable and accurate classification technique for accurate lung disease classification. The experimental results evaluate the performance of the proposed system in terms of Precision, Recall, F-Measure and condition probability.

Index Terms – Lung Disease, Median Filtering, Morphological, Fusion Mean Absolute Deviation (MAD), Walsh-Hadamard Transform (WHT), Gabor Transformation, Support Vector Machine (SVM), Naïve Bayes (NB) and Back Propagation (BP).

1. INTRODUCTION

Lung disease detection and classification is the challenging problem in medical image processing. Due to the structure of disease cells, it is very crucial to accurately classify the location of disease in lung. Generally, the lung can be caused by various disorders, infection, medications and an exposure at the workplace. Early detection of lung disease will significantly reduce the mortality rate. Fig 1 shows the anatomy of lung that has five distinct partitions known as lobes. The right oblique and

horizontal fissures separate the right lung into superior, middle and inferior lobes. Similarly, the left oblique fissures separate the left lung into superior and inferior lobes. For lung disease detection, many technique such as, Chest Radiography (x-ray), Magnetic Resonance Imaging (MRI) and Sputum Cytology are developed in the existing works. But, these techniques are more expensive and time consuming. In order to overcome these issues, a Computed Tomography (CT) based lung disease classification system is introduced in this work. Lung disease classification is defined as the process of identifying the diseased cells by extracting the features of the given image. The lung disease detection and classification is performed in the following stages: preprocessing, segmentation, feature extraction, feature selection and classification.

In this work, the lung includes Bronchitis, Pleural effusion, Emphysema and normal lung are taken for classification. The main objectives of this paper are as follows:

- At first, the median filtering technique is applied to preprocess the given CT image by eliminating the noise in the image.
- Then, the morphological operation, fusion MAD and WHT techniques are employed to segment the filtered image.
- After that, the Gabor transformation technique is used to extract the features of the segmented image.
- Then, the optimal features of the image are selected by using the GA optimization technique.
- Finally, the normal and abnormal lungs are classified by applying the classification.
- Here, there are three different classification techniques such as, SVM, NB and BP are employed

to analyze the suitable technique for efficient classification.

The remaining sections of this paper are organized as follows: Section II reviews some of the existing works related to lung disease detection and classification. Section III presents the detailed description of the proposed lung disease classification system. Section IV evaluates the performance of the proposed classification techniques. Finally, the paper is concluded in Section V.

2. RELATED WORK

This section presents some of the existing works related to lung disease detection and classification. *Tan, et al* [1] suggested a Computer Aided Detection (CAD) system for detecting the lung nodules in a computed tomography images. In this paper, a Lung Image Database Consortium (LIDC) was utilized to train and test the proposed CAD system. This work includes the following stages:

- Preprocessing
- Nodule candidate detection
- Feature selection
- Classification

The main intention of this paper was to detect and classify the lung disease. *Xing, et al* [2] designed a panel of micro RNA markers for the early detection of tumors cell lung disease. This work includes the following stages:

- Marker discovery
- Marker optimization
- Marker validation

In this paper, the diagnostic performance of the panel markers was evaluated by collecting the samples from 67 patients with various stages of tumors cell carcinoma.

Gomathi and Thangaraj[3] recommended a CAD based extreme learning machine for lung disease detection. The processes such as bit-plane slicing, erosion, median filtering, dilation, outlining, lung border extraction and flood fill algorithms were applied to preprocess the lung image. Moreover, the Fuzzy Possibilistic C-Mean (FPCM) technique was developed to segment the lung image. Finally, an Extreme Learning Machine (ELM) technique was used to classify the disease.

Gomathi and Thangaraj[4] proposed a Fuzzy Probabilistic C-Means (FPCM) clustering technique for lung disease segmentation. This technique incorporated both the local spatial context and the non-local information to decrease the effect of noise during the process of segmentation. *Chen and Wang* [5] recommended a 3D CAD system to segment and

reconstruct the lung disease. The main aim of this paper was to offer the more reliable vision for lung disease diagnosis and analysis. Here, the disease regions were extracted and reconstructed in a 3D volume module. *Haider, et al* [6] applied image enhancement and segmentation techniques to detect the lung disease in an early stage. The main aim of this paper was to accurately detect the lung disease for early diagnosis.

Thonnes, et al[7] utilized an integral geometry measures and functional data analysis techniques to classify the lung disease in a High Resolution Computed Tomography (HRCT) scan images. In this paper, the lung images were binarized for a range of various thresholds. *Sharma and Jindal* [8] designed a new detection system using CAD for detecting the lung disease in CT scan images. *Agrabwell, et al* [9] developed an adaptive fuzzy based filtering technique for eliminating the noise in the given image. In this work, the Fuzzy Interference System (FIS) was utilized to enhance the image contrast. The processing time was reduced and the contrast of input image was improved by converting the color image into gray scale image. *Kaur, et al* [10] recommended a Biogeography based optimization technique to detect the abnormal tissue growth in the MRIs. From this work, the requirements of segmentation were discussed for accurate disease segmentation. *Zhang, et al* [11] developed a new classification technique to classify the different types of lung disease, includes, well-circumscribed, juxta-pleural and pleural-tail. This technique integrated both the lung nodule and surrounding anatomical structures based on the contextual analysis. Moreover, the probabilistic estimations were calculated for the relevant images by using a contextual latent semantic analysis based classifier. *Keshani, et al* [12] suggested a Support Vector Machine (SVM) classification and Active Contour Modeling (ACM) techniques for lung disease segmentation and recognition. Here, the lung nodules were classified into four types, such as, lung wall, parenchyma, bronchioles and nodules. *El-Baz, et al* [13] surveyed various CAD systems for lung disease detection and classification. This work includes the following stages: segmentation of lung fields, detection of nodules inside the lung fields, segmentation of detected nodules and diagnosis of nodules.

Bird, et al [14] designed a new tool, namely, Infrared Spectral Histopathology (SHP) for accurate lung disease classification. Moreover, the diagnostic value of SHP was demonstrated to classify the diseaseous and non-diseaseous states of lung histopathology. *Kuruvilla, et al* [15] suggested a Computer Aided Classification (CAC) method for lung disease classification. In this work, the feed forward and feed forward back propagation neural networks were employed to perform the classification process. Moreover, different parameters such as mean, standard deviation, skewness, kurtosis, fifth central moment and sixth central moment were utilized in this work.

3. PROPOSED METHOD

This section presents the detailed description of the proposed lung detection and classification system. The main intention of this paper is to accurately classify the lung image into normal and abnormal. Fig 1 shows the overall flow of the proposed system, which includes the following stages:

- Preprocessing
- Segmentation
- Feature Extraction
- Optimization
- Classification

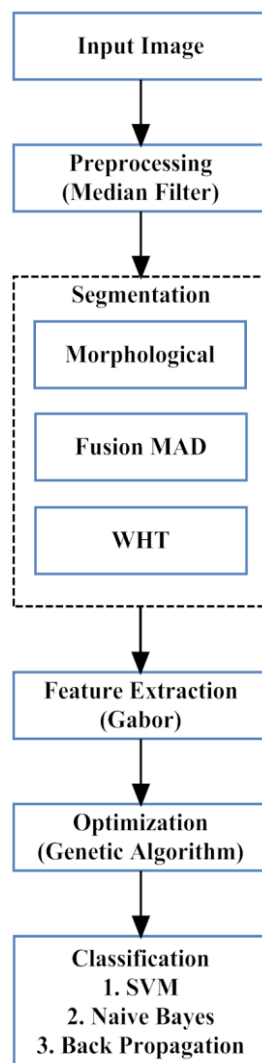


Fig 1. Flow of the proposed lung detection and classification system

A. Preprocessing

Preprocessing is an important and essential application in many medical processing applications. At first, the given lung image is preprocessed by eliminating the noise present in the image. For this purpose, a median filtering technique is utilized in this work. The median filtering is a non-linear method that is mainly used to remove the noise from the given image. It is very effective and extensively used technique in noise removing while preserving edges. It is one of the main building block in image processing applications. It adequately removes the salt and pepper noises by moving through the image pixel by pixel. The pattern of neighbors is referred as window, which slides pixel by pixel over the entire lung image. Initially, the median is calculated by arranging all pixel values from the window into a numerical order. After that, it replaces each value as the median value of the neighboring pixels. From this analysis, it is evaluated that the median filtering technique is the best one that provides the filtered image without noise. The original and preprocessed lung images are shown in Fig 2.

Algorithm 1 – Median Filtering Technique

Input: Image I of size $a \times b$, kernel radius k_r

Output: Image J of the same size as I .

Initialize kernel histogram H .

for $i = 1$ to a **do**

for $j = 1$ to b **do**

for $k = -r$ to r **do**

Remove $I_{i+k, j-r-1}$ to H

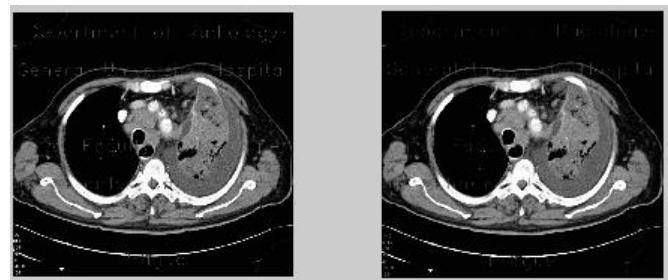
Add $I_{i+k, j+r}$ to H

end for

$J_{i,j} \leftarrow \text{median}(H)$

end for;

end for;



(a)

(b)

Fig 2 (a). Original Lung image and (b). Preprocessed image

B. Segmentation

After preprocessing, the filtered image is further segmented by applying the morphological, fusion Mean Absolute Deviation (MAD) and Walsh-Hadamard Transform (WHT) techniques. Segmentation is defined as the process that splits the given image into regions with some properties such as, gray level, color, texture, brightness and contrast. Automatic segmentation of medical images is a complex task in medical image processing. The main objectives of image segmentation are as follows:

- Analyze the anatomical structure
- Identify the Region of Interest (ROI) for locating the disease and other abnormalities
- Evaluating the volume of disease for measuring the growth of disease

The multisampling is defined as the process of extracting the fusion of multiple samples from the lung CT scan image. The main advantage of image fusion is to derive the most discriminatory information from the original feature set and to eliminate the redundant information from the feature set. Moreover, the MAD uses the median for the deviation scores that is more robust than the other techniques. The WHT is an orthogonal and non-sinusoidal transformation technique that decompose a given signal into a basic function. These functions are known as Walsh functions with the values of +1 or -1. The output image after applying WHT is shown in Fig 3.

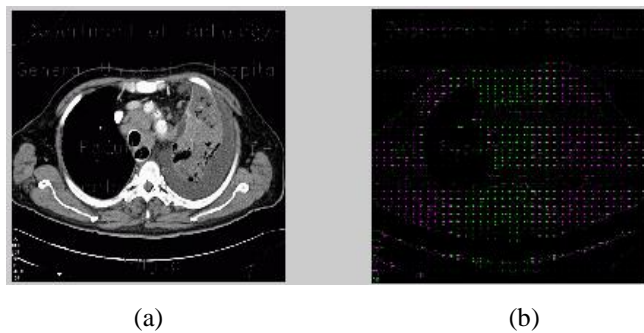


Fig 3 (a). Original image and (b). WHT image

C. Feature Extraction

Feature extraction is defined as the process of capturing the visual content of images for indexing and retrieval. It involves the process of facilitating the amount of resources need to represent a large set of data exactly. The main intention of feature extraction is to represent the raw image in its reduced form to facilitate the decision making process. Moreover, it is an important step to get high classification rate. This stage allows a classifier to identify both the normal and abnormal pattern by extracting a set of features. There are various types of features are used image classification, which includes color, statistical shape features and transform coefficient features.

Selection of feature extraction approach is single but more crucial task in accomplishing high recognition performance. In this work, the Gabor filtering technique is used for feature extraction. Normally, the Gabor filters are band pass filters that has the optimal joint resolution in both spatial and frequency domains. It is more appropriate for texture representation and discrimination, which get the orientation as input and gives a strong response for locations of the target images. The features such as, sum, mean, min, max and median are extracted by using the Gabor filtering technique. The output of Gabor filtering is shown in Fig 4.

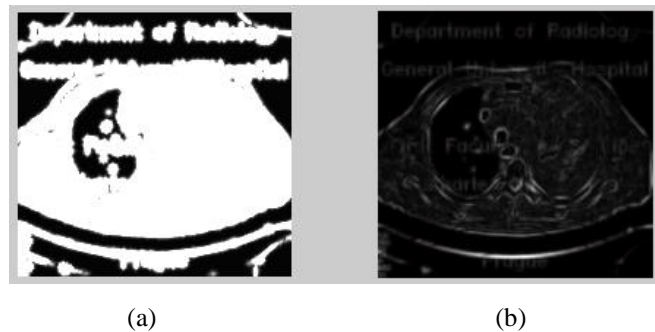
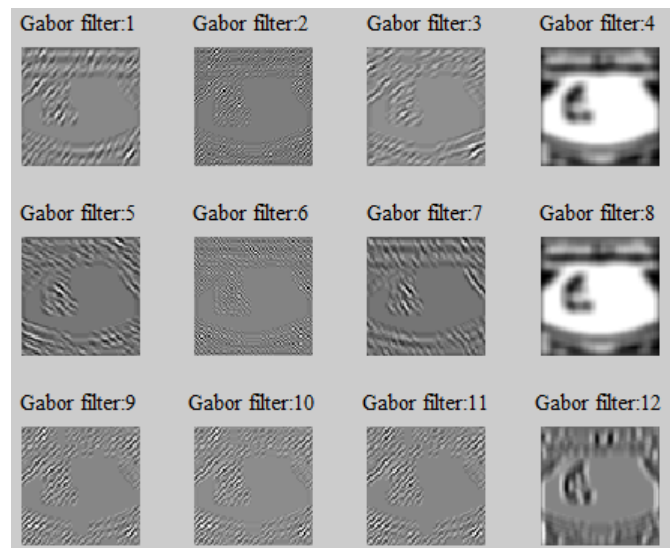


Fig 4 (a). Segmented image and (b). Gabor output



(C). Gabor filtering

D. Feature Selection

After feature extraction, the optimal features are selected by using Genetic Algorithm (GA). Feature selection is one of the key topics in machine learning, which removes the irrelevant and features. Also, it improves the quality of the dataset and the performance of the learning system. It is defined as the process of finding a subset of features from the original set of features. Applying GA to the problem of feature selection is straightforward, where the chromosomes of the individuals

contain one bit for each feature and the value of the bit determines whether the feature will be used in classification.

Algorithm II – Genetic Algorithm (GA)

Step 1: Choose the initial population of individuals;

Step 2: Evaluate the fitness of each individual in the population;

Step 3: Repeat on this generation until reach the time limit and sufficient fitness;

Step 4: Select the best-fit individual by applying crossover and mutation operations;

Step 5: Evaluate the individual fitness of new individuals;

Step 6: Replace the least fit population with new individuals;

E. Classification

After feature selection, the Support Vector Machine (SVM), Naïve Bayes (NB) and Back Propagation (BP) techniques are applied to classify the given image as normal or abnormal. In this work, the performance of three different classification techniques are compared. The SVM is a theoretically superior machine learning technology that provides great results in pattern recognition. The Naïve Bayesian (NB) is a probabilistic learning classification technique. It provides an increased classification accuracy and has the ability to scale high object classes. Back Propagation (BP) is a method of training multilayer neural network that is widely used in many image processing applications. This model contains three layers, includes, input layer, output layer and intermediate layer. The function can be classification function or can be a general regression function. From the evaluation, it is identified that the BP provides the better classification results, when compared to the other techniques.

4. PERFORMANCE ANALYSIS

This section presents the results of the proposed lung disease detection and classification system. The results are analyzed and evaluated in terms of Precision, Recall, F-Measure and probability with true and false values. In this analysis, the performance of three different classification techniques are analyzed and compared.

F. Precision, Recall and F-Measure

Precision, recall and f-measure are the basic measures that are mainly used to evaluate the performance of the classification technique. Precision is defined as a measure of accuracy provided by a specific class has been predicted. It is calculated as follows,

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (10)$$

Recall measures the prediction model's ability that is mainly used to select the instance of a certain class from a dataset. It is also termed as a sensitivity, which is calculated as follows,

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (11)$$

When same weight is provided, the F-measure combines the precision and recall. The computation of the F-measure is based on the following equation,

$$\text{F-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

Table 1 shows the precision, recall and f-measure values of the classification techniques and it is graphically represented in Fig 5.

Table 1. Comparative analysis between the classification techniques

Techniques	Precision (%)	Recall (%)	F-Measure(%)
SVM	98	86.8	97.02
NB	98	88.7	95.60
BP	98	85.2	98

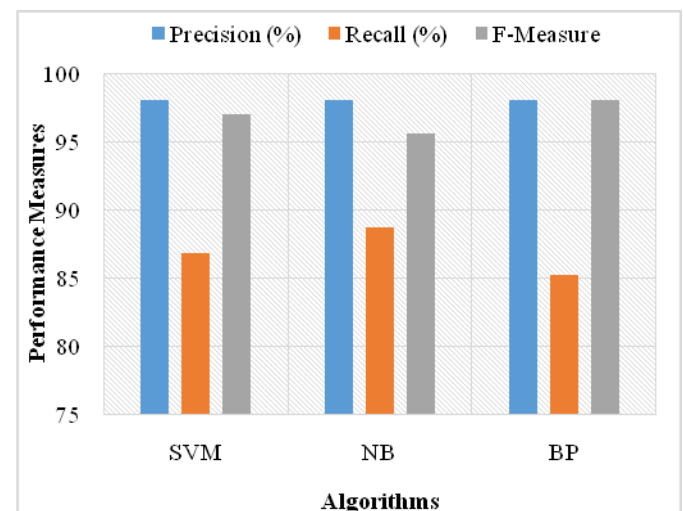


Fig 5. Precision, recall and f-measure

Fig 6 shows the conditional probabilities with evidence of abnormal x-ray results. In this analysis, the probability values of with Bronchitis, without Bronchitis, smoking and lung disease are illustrated. Fig 7 depicts the graphical representation of probability values with true and false values for both bronchitis diagnosis and abnormal x-rays. Here, the red colored bar indicates the false value and the blue colored bar indicates the true value.

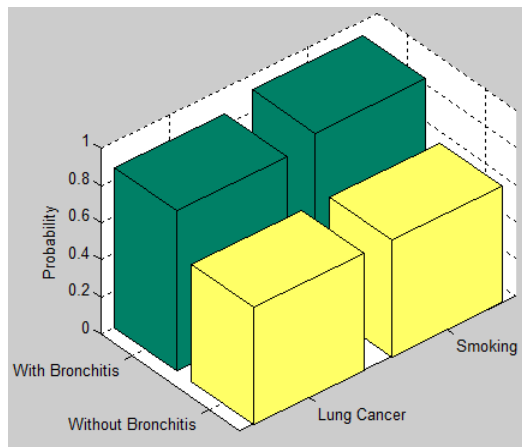


Fig 6. Conditional probability

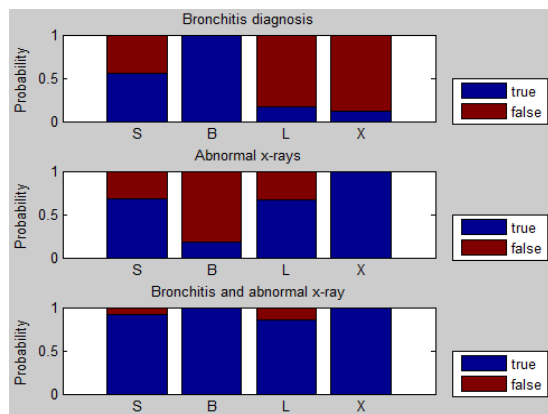


Fig 7. Probability of Bronchitis diagnosis and abnormal x-rays

5. CONCLUSION AND FUTURE WORK

This paper proposed a new classification system for lung disease detection and classification. The main intention of this paper is to accurately detect the disease by improving the performance of classification. For this purpose, different image processing techniques are applied in this work. Initially, the input lung image is preprocessed by using the median filtering technique. It efficiently eliminates the irrelevant and unwanted noise in the image. Then, it will be segmented by applying the morphological operations, fusion MAD and WHT techniques. After that, the features of the segmented image are extracted with the help of Gabor transformation technique. The GA

optimization technique is used to select the optimal features of the image. Finally, the normal and abnormal lung images are identified by implementing the classification techniques. In this work, there are three different classification techniques such as, SVM, NB and BP are implemented to identify the more suitable and efficient technique for lung disease classification. In experiments, the performance of these classification techniques are analyzed and evaluated in terms of Precision, Recall and F-Measure.

REFERENCES

- [1] M. Tan, *et al.*, "A novel computer-aided lung nodule detection system for CT images," *Medical physics*, vol. 38, pp. 5630-5645, 2011.
- [2] L. Xing, *et al.*, "Early detection of squamous cell lung disease in sputum by a panel of microRNA markers," *Modern Pathology*, vol. 23, pp. 1157-1164, 2010.
- [3] M. Gomathi and P. Thangaraj, "A computer aided diagnosis system for detection of lung disease nodules using extreme learning machine," *International Journal of Engineering Science and Technology*, vol. 2, pp. 5770-5779, 2010.
- [4] M. Gomathi and P. Thangaraj, "A new approach to lung image segmentation using fuzzy possibilistic c-means algorithm," *arXiv preprint arXiv:1004.1768*, 2010.
- [5] C.-J. Chen and Y.-W. Wang, "A Preoperative 3D Computer-Aided Segmentation and Reconstruction System for Lung Disease," *Journal of Communication and Computer*, vol. 9, pp. 422-425, 2012.
- [6] W. Haider, *et al.*, "Achieving accuracy in early stage disease identification systems based on image segmentation and 3D structure analysis," *Computer Engineering and Intelligent Systems*, vol. 2, pp. 96-102, 2011.
- [7] E. Thönnies, *et al.*, "Classification of lung in HRCT scans using integral geometry measures and functional data analysis," in *Proceedings of Medical Image Understanding and Analysis*, 2010, pp. 25-30.
- [8] D. Sharma and G. Jindal, "Computer Aided Diagnosis System for Detection of LungDisease in CT Scan Images," *International Journal of Computer and Electrical Engineering*, vol. 3, p. 714, 2011.
- [9] A. Agrawal, *et al.*, "Development of adaptive fuzzy based Image Filtering techniques for efficient Noise Reduction in Medical Images," *IJCSIT International Journal of Computer Science and Information Technologies*, vol. 2, pp. 1457-1461, 2011.
- [10] G. Kaur, *et al.*, "Detection of Abnormal Tissue Growth in MR Imaging using Biogeography Based Optimization," *imaging*, vol. 2, 2013.
- [11] F. Zhang, *et al.*, "Lung nodule classification with multilevel patch-based context analysis," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1155-1166, 2014.
- [12] M. Keshani, *et al.*, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: a complete intelligent system," *Computers in biology and medicine*, vol. 43, pp. 287-300, 2013.
- [13] A. El-Baz, *et al.*, "Computer-aided diagnosis systems for lung disease: challenges and methodologies," *International journal of biomedical imaging*, vol. 2013, 2013.
- [14] B. Bird, *et al.*, "Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung disease," *Laboratory investigation*, vol. 92, pp. 1358-1373, 2012.
- [15] J. Kuruvilla and K. Gunavathi, "Lung disease classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol. 113, pp. 202-209, 2014.