1-1-2021

Early detection of lung cancer - A challenge

Fatma Taher
Zayed University

Neema Prakash
Zayed University

Ashraf AlZaabi
Zayed Military Hospital

Early Detection of Lung Cancer- A Challenge

F. Taher[1], N. Prakash[1] and A. AlZaabi[2]

[1] College of Technological Innovation, Zayed University, Dubai, UAE
[2] Department of Respiratory Division, Zayed Military Hospital, Abu Dhabi, UAE

Abstract: Lung cancer or lung carcinoma, is a common and serious type of cancer caused by rapid cell growth in tissues of the lung. Lung cancer detection at its earlier stage is very difficult because of the structure of the cell alignment which makes it very challenging. Computed tomography (CT) scan is used to detect the presence of cancer and its spread. Visual analysis of CT scan can lead to late treatment of cancer; therefore, different steps of image processing can be used to solve this issue. A comprehensive framework is used for the classification of pulmonary nodules by combining appearance and shape feature descriptors, which helps in the early diagnosis of lung cancer. 3D Histogram of Oriented Gradient (HOG), Resolved Ambiguity Local Binary Pattern (RALBP) and Higher Order Markov Gibbs Random Field (MGRF) are the feature descriptors used to explain the nodule's appearance and compared their performance. Lung cancer screening methods, image processing techniques and nodule classification using radiomic-based framework are discussed in this paper which proves to be very effective in lung cancer prediction. Good performance is shown by using RALBP descriptor.

Keywords: Lung Cancer, MGRF, Image Processing

1. INTRODUCTION

Lung cancer is the root cause for large number of deaths all over the world. Lung cancer classified as Non-small cell lung cancer (NSCLC) and Small cell lung cancer (SCLC). NSCLC have been identified for a large percentage of people when compared with SCLC [1]. Adenocarcinoma or squamous cell carcinoma are the two types of NSCLC. Smoking tobacco is the main cause for all lung cancer cases which affect a large percentage of population. Lung cancer can develop in people who don't have smoking habit, but they may be exposed to passive smoke resulting in early diagnosis. By the use of the CT scans in lung cancer screening, death rate decreases by 20%. Computer aided diagnostic (CAD) systems for lung nodules are developed in order to help radiologists by utilizing machine learning and image processing techniques to assist in cancer detection in a fast and accurate manner. Many researchers used different algorithms such as hierarchical semantic convolutional neural network (HS-CNN) [2], optimal deep neural network (ODNN) [3] etc. to classify nodules into cancerous and non-cancerous. Presently, lung nodules analysis has many shortcomings. Some frameworks utilize the Hounsefield Unit (HU) [4] values as an appearance descriptor, ignoring any spatial correlations. Some

frameworks are imprecise due to their sensitivity to the segmentation because of the use of simple shape descriptor and some frameworks disregard the nodule morphology by only feeding the data as it is. To maintain a strategic distance from these shortcomings, and to provide a competitive accuracy, appearance and shape descriptors are incorporated. To utilize the 3D spatial correlations in the CT intensity scale and the association between voxels, a 3D Histogram of Oriented Gradient, (3D-HOG), RALBP and a Higher-order MGRF are explained in detail [4]. The organization of this paper is as follows: Literature survey is described in Section 2. Test methods are discussed in Section 3. Section 4 is dedicated to procedure overview. Section 5 is presented with image preprocessing and the feature extraction is proposed in Section 6. Section 7 explains the nodule classification and finally, the conclusion is given in Section 8.

2. LITERATURE SURVEY

Major contributions of some of the researchers who aimed at developing a system for early detection of lung cancer are summarized below.

CT scan is used to detect the presence of cancer and its uncontrolled spread. Visual analysis of CT can lead to cancer detection at later stages, which increases the cancer

mortality rate. Therefore, Vas et al. [1] proposed a lung cancer detection algorithm by employing image processing techniques, and then Haralick features are extracted for classification of cancer by using artificial neural networks (ANN). Median filters are used to eliminate impulse noise in the images which proved to be very advantageous. By using the ANN classifier, good accuracy of 92% is also obtained. As a future scope, they explained how to improve the accuracy of the cancer detection system by using a different segmentation technique by employing curve let transformation and morphological features together.

LDCT examinations are very difficult to read; therefore, different methods are available for nodule detection and measurement. Moreover, different nodules require different approaches for their detection, measurement and management, knowledge of the significance of different nodule types is also essential. Radiologists who have undergone specific training will explain about the CT scans. Therefore, Rampinelli et al. [5] proposed a technique called as low-dose computed tomography (LDCT) technique, reading methods and interpretation in lung cancer screening. This is helpful for those radiologists who have little experience in these detection techniques.

Due to the overlapping of cancer cells, early detection of lung cancer is a challenging problem. Classification is a computational procedure used in digital image analysis to sort images into groups according to their similarities. Therefore, Dwivedi et al. [6] presented a perfect classification by multivariate multinomial Bayesian. Kernelised Bayesian techniques are used in this research for classifying CT lung image into normal and abnormal. They aimed to prove that the kennel technique will give good accuracy.

Image processing techniques are used for early detection of lung cancer. Developing new method for lung cancer prevention depends on genetic identification as well as environmental factors. In various lung cancer tumor images, abnormality can be determined by using time factor. Decision tree algorithm is used for the early diagnosis of lung cancer with the help of significant pattern tool. Bhaskar et al. [7] proposed a system that uses histogram equalization. Cost and time can be saved by using their system. Online Prediction process is available to patients from far areas. Therefore, people all over the world can access this system.

Detecting cancer at early stage is very essential because of the high death rate reported annually. If the affected nodule is able to find, then early detection of lung cancer is possible. Therefore, Ravindranath et al. [8] describes a method for nodule extraction. Normal and abnormal nodules are classified based on the changes in the intensity level. Dynamic thresholding technique and

fuzzy classifiers can be used to improve the performance of early cancer detection.

Wever et al. [9] explained diagnosis of lung cancer by employing different imaging techniques. This paper discussed that CT scan is having some limitations and used integrated positron emission tomography (PET) /CT for classifying patients with lung cancer. It gives the combination of morphological and metabolic features.

Chaudhary et al. [10] discussed image processing system to identify the presence of lung cancer in CT-images. In this work, MATLAB is used for processing the images. All image processing steps are explained in detail to get good accuracy in lung cancer detection. This paper concludes by explaining the extraction of lung nodule as the emerging field of research for early diagnosis.

Mokhled et al. [11] discussed about the improvement of image quality for prediction of various diseases. The main factors in this research are image quality and accuracy. Gabor filter is used for processing images and obtained promising results when compared with other techniques. Pixels percentage and mask-labelling proved to be the better features to improve the accuracy of images comparison.

The above-mentioned research works shows less accuracy in feature extraction, therefore this work focuses mainly on the survey of different test methods used for lung cancer detection and used one of the test methods to obtain images for further explanation of different image processing steps and feature extraction, and classification using radiomic based framework. This will thus help to obtain better accuracy and understanding in this area of research.

3. TEST METHODS

A. Chest Radiograph

Lung cancer screening came to existence by the invention of chest radiography. It is a simple and radiation less technique providing large amount of information. It is cheap and available extensively. Chest X-Ray is having three views namely posteroanterior, anteroposterior and lateral. Temporal and dual-energy subtraction are the improved radiographic methods. High- and low-energy X-ray images will be helpful in the pulmonary nodule detection for lung cancer testing. In order to find the use of chest radiography in testing lung cancer, so many trials were performed in the USA and in Europe. Chest radiography can be used for classification of lung lesions as cancerous or non-cancerous. This is possible by checking the cell replication rate. In lung cancer patients, tumor growth is very fast.

B. Computed tomography

Computed tomography (CT) is the next commonly used chest imaging method. Helical CT gives high quality CT images which was introduced several years back [9].

Scanning is done continuously to get single volumetric dataset within one breath-hold. A curve of helical or spiral pattern is traced by the X-ray beam with respect to the patient. New scanners are developed that uses multiple rows of detectors known as multi-detector CT (MDCT). MDCT offers many advantages such as good temporal resolution, fast data acquisition, better contrast material administration, increased spatial resolution and high image quality. Lung cancer staging can be studied by using multiplanar reconstruction (MPR).

C. Magnetic Resonance

Magnetic resonance (MR) is very difficult to take in lungs. MR exhibits many drawbacks such as intrinsic low proton density of lung parenchyma, high susceptibility to motion artifacts, and the decrease in signal intensity. A new method to be introduced in near future is known as whole-body MR (WBMR) imaging that will increase the sensitivity of MR. Advantage of this technique is that it is used for the prediction of diseases that spread easily. Tumor cellularity and the integrity of the cellular membrane are obtained by using an imaging tool called diffusion weighted imaging (DWI). This method can be used for tumor detection and characterization.

D. Positron emission tomography and PET/CT

One of the sensitive and specific imaging techniques used is positron emission tomography (PET). PET uses radioisotopes of natural elements for the imaging. Cancer detection imaging system uses fluoro-deoxyglucose as radioisotope which is very useful in detection and classification of lung cancer. Positron-emitting tracers can be viewed by a PET camera. PET/CT combines a positron emission tomography and X-ray computed tomography to take sequential images from both devices at the same time which are combined to get a single superposed image. Finding the accurate position of the lesions, better separation of lesions from their surrounding structures and better lung lesions classification are some of the main features of PET/CT. Tumor and distal atelectasis can be easily identified by using PET/CT.

E. Bone scans

In order to check if the cancer has spread to the bones, bone scans are used. An intravenous (IV) injection of a radioactive substance called a tracer is given to patients prior to the bone scan. How much radioactive tracer collected in the bones are captured which will helps to identify changes in the bones. Cancer spread to the bones are represented by these changes.

F. Needle Biopsy

Needle biopsy extract a part of tissue from the lungs for examination of cancerous cells. A fine/core needle is used for this purpose. For removing a small sample of tissue from the cancer cell expecting area, fine needle is used which is very thin and this technique is known as fine needle aspiration. For extracting larger pieces of tissue, core needle biopsy is done with a slightly larger needle. Needle biopsy can be done very fast avoiding the stay of patient in the hospital. Patients may be given a mild sedative. In order to numb the path of the needle, a local anesthetic is given. The patient may be upright or lying down depending on the imaging techniques used. To remove samples of tissue, needle is inserted through the skin to the nodule area by using guidance from the imaging.

G. Sputum Cytology

Sputum cytology is a process in which lung secretions or phlegm is taken to determine the cancerous cells. Examination of the mucus sample collected at early morning for some days is by using a microscope to find the cancer cells. Sputum cytology is a very fast and cost-effective test, but it is not always accurate. For diagnosing the lung cancer cell located in the central airways, sputum cytology is more useful. Sputum cytology is not useful in identifying other types of lung cancer located in distant parts of the lungs which is one of the drawbacks of this method. Sputum cytology detects tumors located centrally more than those located in the outer regions of the lungs. Sputum cytology has many limitations therefore, additional testing is recommended.

H. Thoracentesis

A procedure in which a needle or small tube is used to clear additional fluid in the gap between the lungs and the chest wall is called thoracentesis. This fluid buildup called pleural effusion occurs, when a patient is having lung cancer or other medical conditions like heart or liver problem, infections or injuries. In order to remove the fluid and find the cause of the fluid build-up, thoracentesis is used. Thoracentesis can be one of the procedures used to assist a physician in diagnosing lung cancer. The presence and size of pleural effusion can be determined using ultrasound or chest X-ray before thoracentesis. Thoracentesis is generally safe. Bleeding, infection, collapsed lung and respiratory distress are some of the possible risks of thoracentesis. The patient will be able to breathe easier after this procedure.

I. Bronchoscopy

In order to diagnose and treat lung conditions, bronchoscopy is used which is also called as fiberoptic bronchoscopy. To view the airways (bronchi), this test is done by using a thin, lighted tube containing a camera. Either flexible or rigid tube can be used in bronchoscopy. The flexible scope is most commonly used. For viewing the larger airways, rigid scope is used which is a straight tube. Bronchoscopy can be used to view the tumor, assess the airway obstruction, and collect the test samples, if lung cancer is suspected. When the suspicious lesion is located

at the center, bronchoscopy is used in common for diagnosing lung cancer. Bronchoscope is a tube-like instrument placed through the nose or mouth and is passed through the airways of the lungs. A camera is placed at the end of the tube which will displays images back to a video screen.

J. Thoracoscopy

Thoracoscopy is a technique to view the space inside the lungs. Thoracoscope is used for this purpose [9]. In order to evaluate if a suspicious area is cancerous or noncancerous, a small tissue sample is removed during this procedure. A biopsy sample is taken from nearby lymph nodes and the fluid surrounding the lungs to determine whether the cancer is spreading. Thoracoscopic surgery, pleuroscopy, or video assisted thoracic surgeries (VATS) are the other names of thoracoscopy. Thoracoscopy is used in the case of pleural tumors and pleural effusions [9]. Before doing thoracoscopy, blood tests, a pulmonary function test, CT scan, and an electrocardiogram are done. The patient is anesthetized, or given general sedation during this procedure. Thoracoscope is inserted into the small incisions made in the chest. Thoracoscope also have a camera at its end to display images of the lungs and chest cavity on a video screen. Through the other incisions, surgical instruments are inserted. After thoracoscopy, to drain fluid and air, one or more tubes may be placed in the chest and sutures or staples are used to close the incisions.

K. Pulmonary function test

Lung cancer patients are advised to do pulmonary function tests. Different types of pulmonary function test are available [9]. Airflow and lung volume can be measured by spirometry and body plethysmography /helium dilution respectively. Patient needs to sit in a transparent airtight box and allowed to inhale and exhale to measure airflow. Lung volume is determined by measuring the pressure changes inside the box or by allowing the patient to breathe helium gas through a tube for a short time. Some of diseases that affect heart and lung functions can be detected by using pulmonary function tests [9]. Based on the patient's features, pulmonary function test values may vary. Pulmonary function tests are generally safe and it is not good for people with collapsed lungs.

4. PROCEDURE OVERVIEW

Among all the test methods explained in the section 3, we have used computed tomography to explain the various steps involved in lung cancer detection. CT images are collected from hospitals and MATLAB is the software used to process these images. Different steps of image processing procedure are discussed as follows. Fig.1 shows the basic block diagram explaining the lung cancer diagnosis process.
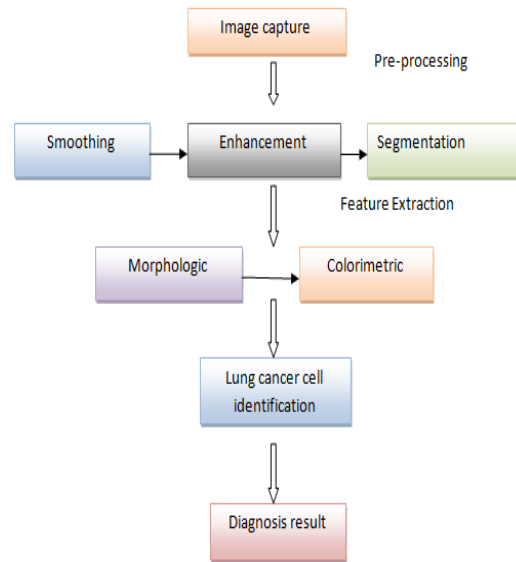

Figure 1. Diagnosis Process

Convolution filters with Gaussian pulse are used for lung cancer detection to smooth the cell images. The image enhancement is done by varying the color and contrast. Then segmentation of the nucleuses is done by thresholding. Fig. 2 shows the non-cancerous and cancerous lung images.

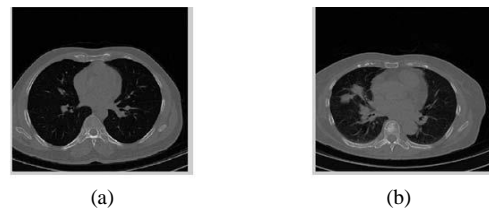(a)   (b)

Figure 2. (a) Non-cancerous and (b) Cancerous Images

These are some of the main steps of digital image processing. Feature extraction is the next step after image processing. Pulmonary nodule is identified from the extracted features to detect the presence of cancerous cells [10].

5. IMAGE PREPROCESSING

Many steps are involved in the preprocessing [12] stage of image processing which are explained as follows. Fig. 3 shows the pre-processed images.
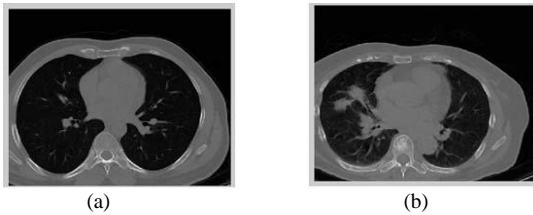
(a)   (b)
Figure 3. Pre-processed Images

A. Image Smoothing

Smoothing is used for reducing the noise or other small variations in the image and also blurring of the edges where the main data about the image is present. These problems can be rectified by using median filter.

B. Image Enhancement

Two broad categories of image enhancement are spatial and frequency domain methods. Three techniques are involved in enhancement namely Gabor filter, auto-enhancement and fast Fourier transform (FFT) [10]. Gabor filter is a linear filter whose impulse response is given by a harmonic function multiplied by a Gaussian function [11] and it is introduced by Dennis Gabor. This technique is shown in Fig. 4. Gabor function is used for texture analysis which makes it applicable in many fields.
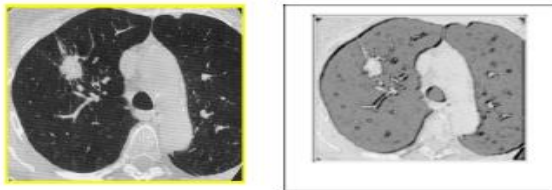

Figure 4. Gabor Filter Enhancement Technique

The second technique is used to automatically enhance the brightness, color and contrast of the image and it depends on calculation of mean and variance. Fig. 5 shows the auto enhancement technique. In the Fast Fourier transform technique, image will be in the Fourier domain at the output and will be in the spatial domain at the input of the transformation. FFT is a faster version of the Discrete Fourier Transform (DFT) [10]. Fig. 6 shows the FFT enhancement technique. Gabor filter enhancement gives a good quality image compared to other techniques.
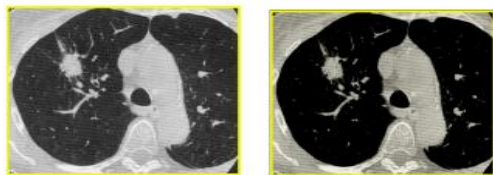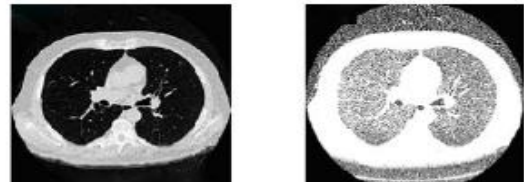

Figure 5. Auto Enhancement Technique

Figure 6. FFT Enhancement Technique

C. Image Segmentation

Image processing starts by dividing an image into multiple segments. For applications like image compression or object recognition, segmentation is commonly used [11]. Fig. 7 depicts the segmented lungs. Thresholding and watershed segmentation are the two commonly used image segmentation methods.

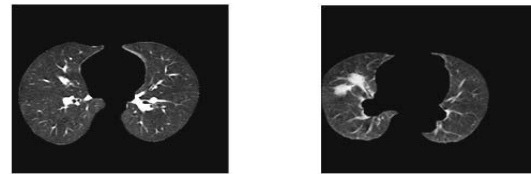

(a)   (b)
Figure 7. Segmented Lungs

Thresholding is a non-linear operation where a gray-scale image is converted into a binary image in which the two levels are given to pixels that are below or above the specified threshold value. One above the threshold is called an object point, otherwise, the point is called a background point. Fig. 8 shows the threshold approach.
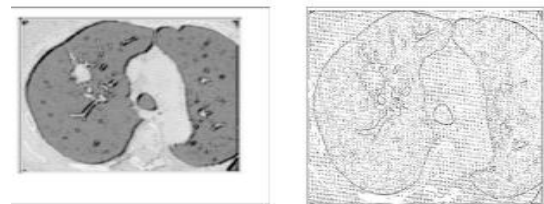

Figure 8. Thresholding Approach

Watershed segmentation uses colors to differentiate the touching objects in an image. This type of segmentation is classified as external associated with the background and internal associated with the objects of interest. Water shed segmentation (Fig. 9) will give good resolution images and is reliable compared to other types.
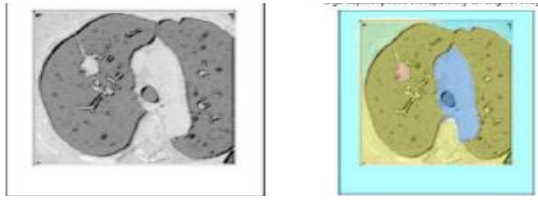
Figure 9. Watershed Segmentation Approach

6. FEATURE EXTRACTION

This method is used to pluck out some of the features of an image such as area, perimeter, eccentricity and average intensity [12]. It is a form of dimensionality reduction. The extracted features are then used for the classification. Binarization and masking are the commonly used feature extraction methods for early detection of lung cancer. These methods are based on structure of the lung and CT imaging. Binarization is a process of taking a grayscale image and converting it to black-and-white. It is also called thresholding [12]. In the masking approach [11], as the white masses increases inside region of interest (lungs), chances of the cancer presence also increase. Both the methods are combined, which helps to identify whether the patient is normal or abnormal.

The classification of pulmonary nodules is possible by extracting the features. Appearance and shape features are explained as follows:

A. Appearance Features: The accelerated outgrowth of the affected nodules causes non-uniformity in their textures. There will be variations in the HU values due to this non-uniformity and will be described using the following descriptors.

1) 3D Histogram of Oriented Gradient, (3D-HOG): Shaffie et al. [13] calculated HOG feature vector for the classification of benign and malignant nodules. Nodule appearance can be studied from this feature vector easily. There are two steps used for the calculation of 3D HOG (Fig. 10), mean gradient and orientation binning.

Mean gradient for 3D volumes can be calculated as follows:

At x, y, z, integral gradient volume for a voxel equals x, y, z. $\overline{intVol} = (intVol_x, intVol_y, intVol_z)$, $intVol_x = \sum_{\acute{x} \le x, \acute{y} \le y, \acute{z} \le z} V_{dx}(\acute{x}, \acute{y}, \acute{z})$, and similarly, for $intV\,ol_y$ and $intV\,ol_z$. Where V_{dx} is the the partial derivatives along x direction. V_{dx} can be calculated using the following recurrences:

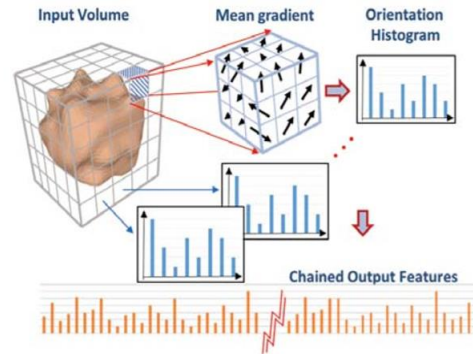$$line(x, y, z) = line(x, y-1, z) + V_{dx}(x, y, z) \qquad (1)$$

Figure 10. 3D HOG illustration.

$$plan(x, y, z) = plan(x-1, y, z) + line(x, y, z) \qquad (2)$$

$$intVol_x(x, y, z) = intVol_x(x, y, z-1) + plan(x, y, z) \qquad (3)$$

Where, *line* (x, y, z) denotes the accumulative row sum, *plan* (x, y, z) denotes the accumulative plan sum, for initialization *line* (x, −1, z) = 0, *plan* (−1, y, z) =0, and $intVol_x$ (x, y, −1) = 0. The mean gradient $\bar{G}$ for any 3D sub-volume Q as shown in Fig. 11 is calculated after getting $\overline{intV\,ol}$. Base point V_1, its width (W), height (H), and length (L) are used for getting the mean gradient. $\bar{G} = \overline{D} - \overline{H}$, $\overline{D} = \overline{V_8} - \overline{V_6} - \overline{V_7} + \overline{V_5}$, $\overline{H} = \overline{V_4} - \overline{V_2} - \overline{V_3} + \overline{V_1}$. The next step is orientation binning which means that, each voxel votes for an orientation according to the closest bin in the range 0 to 360 degrees in 2D. Based on the gradient magnitude, this vote is weighted. Number of sides of polygon is determined by counting the number of bins. A polyhedron is extended from the polygon in order to make binning in 3D. Division of bins is possible by using a regular and convex polyhedron.
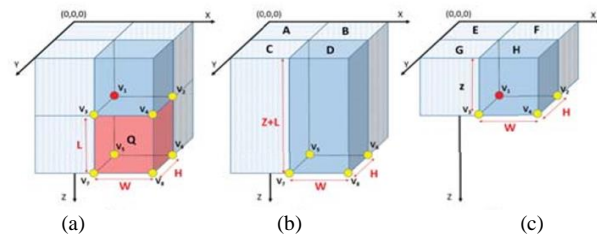

Figure 11. Mean gradient computation

Dodecahedron which is having 20 vertices meet this criterion. Resolution for the mean gradient vector $\bar{G}$ is calculated by dot product. Then it is simplified by matrix multiplication. Let M be the matrix of vertices coordinates m_1, m_2,..., m_n. $M = (m_1, m_2, \ldots, m_n)^T$, $m_i = (x_i, y_i, z_i)$. The dodecahedron could be presented by: (±1, ±1, ±1), (0, ± (α -1), ±α), (±α, 0, ± (α -1)), where α is the golden ratio. The normalized component J of $\bar{G}$ is calculated as follows: $J = (j_1, j_2, \ldots, j_n)^T = \frac{M.G}{\|\bar{G}\|}$. Thus, each j_i holds the normalized

component of $\bar{G}$ on axes through the vertex m_i. Binning threshold was 1:291 for dodecahedron. After normalization, the magnitude of each bin is obtained using this equation: $\tilde{J} = \frac{\|\bar{G}\| \cdot J}{\|J\|}$. To get the HOG features for each block, all of these histograms for the cells in the same block are added and normalized.

2) 3D Resolved Ambiguity Local Binary Pattern(3D-RALBP):

The RALBP feature vector is explained by Soliman et al. [4], and is used for classification. Fig. 12 shows the detailed illustration. The 3D-RALBP can be also used to face some of existing challenges in the CT scans.

The Original Local Binary Pattern (LBP) is a feature descriptor. In LBP, first the threshold is computed and then the features are extracted based on this threshold value. In the 2D LBP, a binary code generated is as follows [4]:

$$LBP_{N,R} = \sum_{i=1}^{N} 2^{i-1} * f(g_p - c) \qquad (4)$$

where

$$f(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & otherwise \end{cases}$$

Due to the low quality of some CT scanners, there will be presence of noise or distortion when using the 3D-LBP descriptor. Therefore, in the nodule classification application, LBP is showing less accuracy.
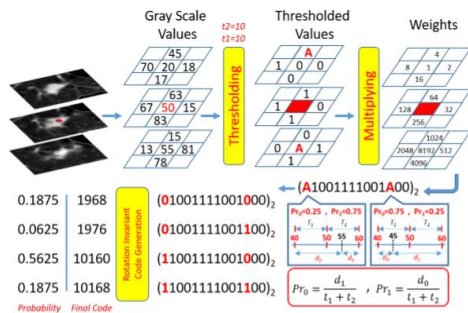

Figure 12. 3D RALBP code calculation

In this situation, 3D Resolved Ambiguity LBP is implemented, in which two thresholds were added, t_1 and t_2 and changes the condition for (4).

$$f(x) = \begin{cases} 1 & \text{if} \quad x \geq t_2 \\ a & \text{if} \quad -t_1 < x < t_2 \\ 0 & \text{if} \quad x \leq -t_1 \end{cases}$$

In order to remove the effect of the noises, these two thresholds, t_1 and t_2 will help. 0 and 1 in the above

condition are solid codes and a is an ambiguous code. Code ambiguity has to be resolved and should be transferred to a solid code. This is possible by using the following equation $P(0) = \frac{c-g+t_2}{t_1+t_2}$, $P(1) = \frac{g-c+t_1}{t_1+t_2}$ where c denotes the value of the central voxel, g denotes the voxel value of the code which is calculating, and t_1, t_2 are the lower bound of the differences of the voxels that will assign 0 and 1 respectively[4]. Now, different combinations for the solid code will be generated for all ambiguous codes. If m bits with an ambiguous code are present, results in 2^m codes. 3D-RALBP is designed to be a rotation invariant descriptor to avoid classification errors in the detection phase due to the generation of different codes for same appearance feature during rotation.

3) MGRF energy:

MGRF model is depicted in Fig. 13. Lung nodule characteristics is the basis of MGRF modeling [14]. Here, with benign nodules, more homogeneous intensity exists when compared to malignant nodules where more heterogeneity can be seen. The Gibbs potential energy is calculated as $E_7(\mathbf{g})$, $v_{7:\rho}(g(r') : r' \in \mathbb{R}(r))$, where $\mathbf{g} = (g(r): r \in \mathbb{R})$, where, ρ is the interaction radius. Maximum likelihood estimates (MLE) helps to understand the 7-voxel neighborhood from a set of nodules.
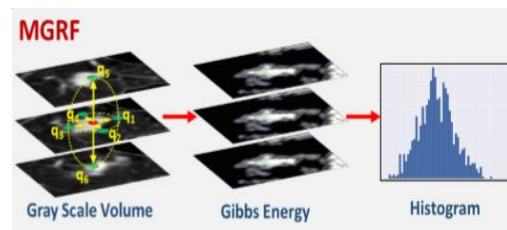

Figure 13. MGRF Illustration

The intensity patterns for the training nodules, $\mathbf{g}$ is considered as a sample of higher order MGRF [4]. In this model, $E_7(\mathbf{g})$, $\mathbf{g}$, $g(\mathbf{r})$, are related by: $P_7(\mathbf{g}) = 1/Z \; \psi(\mathbf{g}) \exp(-E_7(\mathbf{g}))$, where, Z is the normalizing factor.

B. Shape Features

Spherical and Geometric features are described as shape features. Fig. 14 shows the shape feature illustration.

1) Spherical Harmonic: A crucial discriminative radiomic can be seen in the pulmonary nodule's shapes. Based on the malignancy status, lung nodules have different shapes. Nodule shape modeling is a better method to avoid such problems. Spectral SH analysis [15] is used for modeling the surface of the nodule.
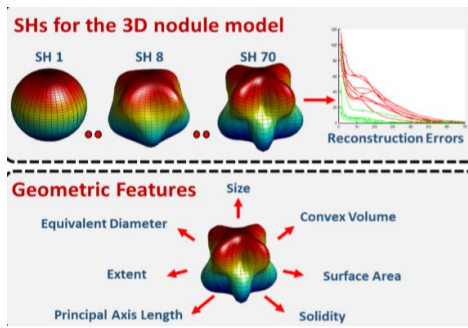
Figure 14. Shape Features Illustration

First step is to build a 3D mesh in triangle shape by which the nodule's surface is approximated. Before, SH decomposition, it is mapped to the unit sphere. By using Attraction-Repulsion Algorithm [14], precise modeling is achieved. Nodules are shown as linear combination of spherical harmonics in the mapping process. Benign nodules which are fewer complex shapes are shown by lower-order harmonics, while malignant nodules are represented by higher-order harmonics as they are called as complex shapes. The iterative residual fitting [14] is used to estimate the coefficients for the linear combination modeling of the nodule surface that mapped to the unit sphere through the solution of isotropic heat equation.

2) Geometric features: A set of geometric features will be calculated for each segmented nodule, (delineated by the radiologist), to provide a modeled discrimination between different types of pulmonary nodules (benign vs. malignant). Quantitative measures for some of the commonly used parameters used for calculation [4] are provided by the extracted features. To tackle the problems stemming from the differences in acquisition, a predefined isotropic volume of size $40\times40\times40$ mm^3 is extracted for each nodule. This cubic volume is centered around the nodule center [16].

7.   NODULE CLASSIFICATION

A. Radiomic- Based Framework

A noninvasive diagnostic system for pulmonary nodule is presented by combining both shape and appearance descriptors, which are derived from one CT scan (Fig. 15).

A two-stage framework is designed for the classification of pulmonary nodules. Different classifiers are used in the first stage. An autoencoder followed by a SoftMax classifier is used in the second stage to give the final result of classification [13]. Spherical harmonics and MGRF features use stacked auto encoders which are having more than one hidden layer.
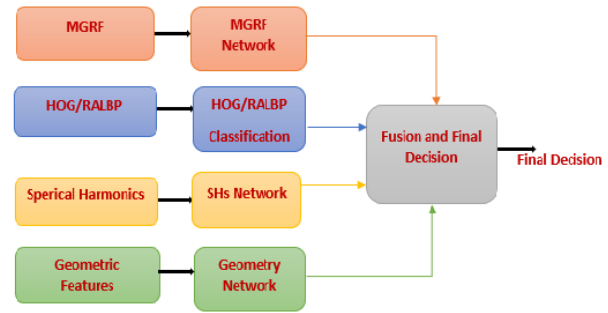

Figure 15. Comprehensive Framework for Lung Nodule Diagnosis

An SVM classifier is used for the HOG/RALBP. SVM showed a greater accuracy than the autoencoder for this feature. SoftMax is used to classify the geometric features which are of small size when compared with other features. An autoencoder is followed by a SoftMax in the second stage for the fusion to obtain better classification accuracy.

B. Experimental Results

LIDC dataset [17] is utilized for the system validation which is a publicly available dataset. Clinical chest CT image of 1080 cases are present in this database which is linked to eXtensible Markup Language files for further explanation. A two-phase scan information collected by a group of radiologists are recorded in these files. The training and testing dataset are generated by selecting the nodules that decided by these radiologists' group. The average score of these radiologists are taken in concern and reached a conclusion that, if it is $\geq 3:5$, it is considered malignant and is benign if $\leq 1:5$. More than 700 nodules are used for testing and training to evaluate the performance which consists of both benign and malignant nodules. Extracted nodule is having volume of $40 \times 40 \times 40$mm^3. Accuracy, sensitivity, specificity and precision are measured for comparison of HOG and RALBP feature descriptors which are used for the classification. The performance metrics is shown in Table I. These measurements of the performance parameters are explained in detail in [18].

TABLE I.   PERFORMANCE METRICS

Combined Features	Performance metrics			
	Accuracy	*Sensitivity*	*Specificity*	*Precision*
Using HOG	93.12	92.47	93.60	91.49
Using RALBP	94.95	94.62	95.20	93.62

8. CONCLUSIONS

Lung cancer is one of the common and serious type of cancer and it affects a large population worldwide in every year. Early diagnosis of lung cancer is very challenging. In this paper, different test methods which helps in the prediction of lung cancer at earlier stages and the steps involved in image processing such as preprocessing, image enhancement, image segmentation which helps to improve the quality of the images obtained after prediction test for extracting features are also explained. Furthermore, we discussed a comprehensive framework for pulmonary nodule classification by combining appearance and shape feature descriptors. The HOG/RALBP, higher order MGRF, and appearance descriptors are modeled in order to describe the non-uniformity of the nodule texture due to the sudden increase in the spread of the affected nodules. Spherical harmonic and basic geometric features are used for modeling shape descriptors to explain the shape intricacy. Finally, a deep neural network is used to combine all these features which helps to identify the pulmonary nodules which are affected or not affected. Among the two feature descriptors, better performance is shown by using RALBP.

REFERENCES

[1] M. Vas, A. Dessai, "Lung cancer detection using lung CT images image processing", *ICCUBEA*, August 2017, proceedings 2018, pp. 1-5.

[2] Shen, S., et al., "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification", *Expert Systems with Application*, pp. 1-23, June 2018.

[3] L. Prabu, S., et al., "Optimal deep learning model for classification of lung cancer on CT images", *Future Generation Computer Systems*, pp.374-382, 2019

[4] A. Soliman, A. Shaffie, F. Taher, "A novel CT- based descriptors for precise diagnosis of pulmonary nodules", *ICIP*, pp.1400-1404, 2019.

[5] C. Rampinelli, D. Origgi, M. Bellomi, "Low-dose CT: technique, reading methods and image interpretation", *International Cancer Imaging Society*, pp.548-556, 2012.

[6] S. A.Dwivedi, R.P. Borse, A. M. Yametkar, "Lung cancer detection and classification by using Machine learning & Multinomial Bayesian", *IOSR Journals,*pp.69-75, Jan. 2014.

[7] N. Bhaskar, B. Chaitanya, "A survey on early detection and prediction of lung cancer", *IJTIMES*, pp.1291-1298, May 2018.

[8] K. Ravindranath, K. Somashekar, "Early detection of lung cancer by nodule extraction- a survey", *ICEECCOT*, pp.927-930, 2017.

[9] W. D. Wever, J. Coolen, J.A. Verschakelen, "Imaging techniques in lung cancer", *Breathe*, pp.339-346, June 2011.

[10] A. Chaudhary, S. S. Singh "Lung cancer detection on CT images using image processing", *International conference on computing sciences*, pp.142-146, 2012.

[11] M. S. Al-Tarawneh, "Lung Cancer detection using Image processing techniques", *Leonardo electronics journal of practices and technologies*, no.20, pp. 147-158, Jan-Jun 2012.

[12] B.V. Ginneken, B.M. Romeny and M.A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", *IEEE, transactions on medical imaging*, vol.20, no.12(2001).

[13] A. Shaffie, A. Soliman, H.A. Khalifeh, F. Taher, A. Elmaghraby, and A. El-Baz, "Radiomic-based framework for early diagnosis of lung cancer", *proceeding at IEEE*, pp. 1293-1297, 2019.

[14] A. Shaffie, A. Soliman, F. Taher, N. Dunlap, B. Wang, A. Elmaghraby, G. Gimel'farb and A. El-Baz, "A new Framework for Incorporating Appearance and Shape Features of Lung Nodules for Precise Diagnosis of Lung Cancer", *proceeding at IEEE International Conference on Image Processing (ICIP) conference*, pp. 1372-1376, China, Sept. 17-20, 2017.

[15] Chung, M.K., et al., "Weighted Fourier series representation and its application to quantifying the amount of gray matter", *IEEE transactions on medical imaging*, pp.566-581, 2007.

[16] Olson, E., et al., "Particle shape factors and their use in image analysis part 1: theory", Journal *of GXP Compliance*, pp.85-96, 2011.

[17] Samuel III, G., et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on ct scans", *Medical physics*, pp. 915-931, 2011.

[18] A. Shaffie, A. Soliman, F. Taher, B. Wang, A. Elmaghraby, G. Gimel'farb and A. El-Baz, "A Generalized Deep Learning-Based Diagnostic System for Early Diagnosis of Various Types of Pulmonary Nodules", *Technology in Cancer Research & Treatment*, pp. 1-9, 2018.

Fatma Taher is an associate professor and assistant dean for research and outreach in the College of Technological Innovation at Zayed University, Dubai, UAE. Her research interests are in the areas of signal and image processing, pattern recognition, deep learning, Machine learning, artificial intelligence, medical image analysis, especially in detecting of the cancerous cells, kidney transplant and autism. In addition to that, watermarking, remote sensing and satellite images researches. Dr. Fatma Taher has published more than 70 papers in international Journals and conferences. She served as a member of the steering, organizing and technical program committees of many international conferences, and has served on many editorial and reviewing boards of international journals and conferences. Dr. Fatma is the chair of IEEE UAE section and the chair of the Education Committee in British Society in UAE. Dr. Fatma has received many distinguished awards such as: Best paper award of the first prize in the PhD Forum of the 20th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), PhD Forum, Dec. 8-11, 2013. And recently, she was awarded the UAE Pioneers award as the first UAE to create a computer aided diagnosis system for early lung cancer detection based on the sputum color image analysis, awarded by H.H Sheik Mohammed Bin Rashed Al Maktoum, 15th Nov. 2015. In addition to that she was awarded an innovation award at the 2016 Emirati Women Awards by H. H. Sheik Ahmed Bin Saeed Al-Maktoum. Chairman of Civil Aviation Authority, and LOreal-UNESCO For Women in Science Middle East Fellowship 2017.

Neema Prakash is currently working as research assistant in the College of Technological Innovation at Zayed University, Dubai, UAE. She received M.tech degree in Optoelectronics and Optical communication from University of Kerala, India (2016), B.tech degree in Electronics and Communication from Mahatma Gandhi University, Kerala, India (2012). Her research interests are in the areas of image processing, designing electronic circuits and telecommunication. She has gained experience in image processing working in Raman Research Institute, India and has good experience in testing electronic circuits and components, working in Indian Space Research Organization, India.

Ashraf AlZaabi is a professor and the Head of Respiratory Division at Zayed Military Hospital in Abu Dhabi. UAE. He graduated from the Royal College of Surgeons in Ireland 1997 and did his training in Internal medicine and Respirology for six years at the University of Toronto, Canada 2000-2006 where he was board certified in both specialties in Canada and USA. Since his arrival to UAE he managed to develop a state-of-the-art patient services in the Respiratory Medicine Division at Zayed Hospital and this included establishing the first sleep laboratory in Abu Dhabi and the first endoscopy center performing Endobronchial ultrasound guided biopsy procedures in UAE. In 2007, he was appointed as a Clinical Assistant Professor in Internal Medicine at the College of Medicine and Health Sciences (CMHS) at UAE University and was promoted to a full professor on October 21, 2018. Ashraf Alzaabi won many distinction awards such as the prestigious Murray Muirhead Award (2001) from the University of Toronto during his residency program. In 2013, He won the Abu Dhabi Medical Distinction Award in Clinical Performance launched by the Health Authority in Abu Dhabi (HAAD). He was elected as a president for the Emirates Thoracic Society February 2019. He engineered a memorandum agreement between the faculty of Medicine at UAE University and Zayed Hospital. The signing of this agreement recognized Zayed Hospital as a training hospital for the medical students and residents during their undergraduate and postgraduate studies. In 2010, He finished a two-year program at Zayed University doing an Executive Master in Health Care Administration. This helped during his posts at many important administrative positions both at Zayed Military Hospital and at other national organizations. He served as the Chair of the Internal Medicine Examination Committee at the Health Authority of Abu Dhabi for the term 2010-2014. Medical research and academic work have always been of great interest to him. He is an editor at the "CHEST" journal, middle east edition. He published the first UAE prevalence study on chronic obstructive pulmonary disease at the Respiratory Medicine Journal in 2011. His main interest in research includes Asthma, COPD and Sleep Medicine.