Scientific
Research

Improves Treatment Programs of Lung Cancer Using Data Mining Techniques

Zakaria Suliman Zubi[1], Rema Asheibani Saad[2]

[1]Computer Science Department, Faculty Of Science, Sirte University, Sirte, Libya; [2]Computer Science Department, Faculty of Science, Tripoli University, Tripoli, Libya.
Email: zszubi@yahoo.com, ranem_reem@yahoo.com

ABSTRACT

Lung cancer is a deadly disease, but there is a big chance for the patient to be cured if he or she is correctly diagnosed in early stage of his or her case. At a first glance, lung X-ray chest films being considered as the most reliable method in early detection of lung cancers, the serious mistake in some diagnosing cases giving bad results and causing the death, the computer aided diagnosis systems are necessary to support the medical staff to achieve high capability and effectiveness. Clinicians could predict patient's behavior future and improve treatment programs by using data mining techniques and they can be better managing the health of patients today, in addition they do not become the problems of tomorrow. The lung cancer biological database which contains the medical images (chest X-ray) classifies the digital X-ray chest films into three categories: normal, benign and malignant. The normal ones are those characterizing a healthy patient (non nodules);, lung nodules can be either benign (non-cancerous) or malignant (cancer). Two steps are major in computer-aided diagnosis systems: pattern recognition approach, which is a combination of a feature extraction process and a classification process using neural network classifier.

1. Introduction

Cancer is a disease of the body's cells. Our bodies are always making new cells: so we can grow, replace worn-out cells, or heal damaged cells after an injury. This process is controlled by certain genes. All cancers are caused by changes to these genes. Changes usually happen during our lifetime, although a small number of people inherit such a change from a parent. Normally, cells grow and multiply in an orderly way. However, changed genes can cause cells to behave abnormally. They may grow into a lump. These lumps can be benign (not cancerous) or malignant (cancerous). Benign lumps do not spread to other parts of the body. A malignant lump (more commonly called a malignant tumor or a cancer) is made up of cancer cells. When it first develops, this malignant tumor is confined to its original site. If it is not treated, cancer cells may spread into surrounding tissue and to other parts of the body. When these cells reach a new site, they may continue to grow and form another tumor at that site. Such tumors are called secondary cancers or metastases. For a cancer to grow bigger than the head of a pin, it must grow on its own blood vessels, and this is so called angiogenesis [1]. The efficient use of information and technology is essential for the health care organizations to stay competitive in today's complex, evolving environment. The challenges faced when trying to make sense of large, diverse, and often complex data source of significant. In an effort to turn information into knowledge, health care organizations are implementing data mining technologies to help control costs and improve the effectiveness of patient

care. Data mining can be used to help predict future patient behavior and to improve treatment programs. By identifying high-risk patients, clinicians can better manage the care of patients today so that they do not become the problems of tomorrow. For instance, early stage lung and oral cancers are very hard to diagnose by conventional means; genomic signatures can be used to provide more timely and perhaps more accurate diagnosis [2]. Data mining refers to extracting or "mining" knowledge from large amounts of data. It is an increasingly popular field that uses statistical, visualization, machine learning, and other data manipulation and knowledge extraction techniques aimed at gaining an insight into the relationships and patterns hidden in the data [3]. In principle, data mining should be applicable to any kind of information repository. This includes relational databases, data warehouses, transactional databases, advanced database systems, and the World-Wide Web. Advanced database systems include object-oriented, object-relational databases, and specific application-oriented databases, such as biological database, spatial databases, time-series databases, text databases, and multimedia databases [4]. Multimedia data are usually *unstructured* by nature. There are no well-defined fields of data with precise and nonambiguous meaning, and the data must be processed to arrive at fields that can provide content information about it. Such processing frequently leads to nonunique results with several possible interpretations. In fact, multimedia data are often subject to varied interpretations even by human beings. For example, it is not uncommon to have different interpretations of an image by different experts, for example radiologists. Another difficulty in mining of multimedia data is its heterogeneous nature. The data are often the result of outputs from various kinds of sensor modalities with each modality needing its own way of processing. Yet another distinguishing feature of multimedia data is its sheer volume. All these characteristics of multimedia data make mining demanding and interesting [5]. There are various kinds of data mining methods; some of the major data mining methods are known as deviation detection, summarization, classification, generalized rule induction and clustering [6]. The lung cancer biological database which contains the medical images (chest X-ray) classifies the digital X-ray chest films into three categories: normal, benign and malignant. The normal ones are those characterizing a healthy patient (non nodules), lung nodules can be either benign (non-cancerous) or malignant (cancer). Two steps are major in computer-aided diagnosis systems: pattern recognition approach, which is a combination of a feature extraction process and a classification process using neural network classifier. The performance of the classifier depends directly on the ability of characterization of candidate regions by the adopted features.

Many kinds of features have been proposed for discriminating between normal tissues and abnormal ones [7].

2. The State of the Problem

Many radiologists adopt a wait-and-see approach that allows nodules to become larger. "*But many oncologists now wants to take a shot at the small nodules, which could lead to new and innovative treatment and outcome improvements*", said Hirakawa in [8]. "*The chest X-ray is quite complex and a difficult area to navigate*", said Riverain Medical Chairman and CEO Diane Hirakawa in [8]. "*Small nodules are only millimeters in size and can be hidden behind ribs and the surrounding chest structures. On Guard uncovers these subtle nodules, leading to earlier diagnosis and improved therapy options*" [8]. The detection of lung cancer at an early stage is very important to cure it; however, it is very difficult for radiologists to detect and to diagnose lung cancer on chest X-ray images because of following reasons (Lee *et al.*, 2007). There are many tissues that overlap each other on chest radiographs. The presence of cancerous tumors is obscured by the overlying ribs, bronchia, blood vessels and other normal anatomic structures. The shadows of cancerous tumors seen on chest radiograph are usually vague and subtle, and they tend to be missed [9]. Chest X-ray can deduction the nodules in picture as a spot, nodule means a spot on the pictures taken of lungs that does not look like normal lung tissue or blood vessels. A nodule may also be called a lesion. The nodule may range in size, from something the size of a grain of rice. These nodules have many attributes (Class, shape, size of nodule…) used for diagnosis. Most very small nodules are not cancer. The nodule may be an infection. It might also be scar tissue from a previous infection. If nodule is very small, for example, the size of a grain of rice, a marble, or even a walnut, then surgery to remove the lobe in which the nodule was found may be the only treatment that is needed. Almost everyone who has lung cancer, that is found when it is still this small, is cured by surgery. The smallest tumor that can be seen on a chest X-ray is about 1/2-inch in diameter, if nodule is larger than the size of a walnut, lemon, or even larger, it needs to have more treatment than just the surgery. The other treatments might include chemotherapy, radiation, or both [10]. Overall, the chances that a lung nodule is cancer are 40%, but the risk of a lung nodule being cancerous varies considerably depending on several things. In people less than 35 years of age, the chance that a lung nodule is cancer is less than 1%, whereas half of lung nodules in people over age 50 are malignant (cancerous). Other factors that raise or lower the risk that a lung nodule is cancer include:

- Size—Larger nodules are more likely to be cancerous than smaller nodules.
- Smoking—Current and former smokers are more like-

ly to have cancerous lung nodules than never smokers.

- Shape—Smooth, round nodules are more likely to be benign, whereas irregular or "spiculated" nodules are more likely to be cancerous [10].

In our work we will focus on the shape and size to identify the nodules to be classified images as normal or benign or malignant. To detect the small nodules, early diagnosing lung cancer system deploys advanced image enhancement, feature extraction, and classification techniques (neural networks, for instance, that differentiate normal chest structures from solitary pulmonary nodules).

Early diagnosing lung cancer system allows physicians to deal with enormous data sets more efficiently. Essentially, it makes their job easier and, in turn, makes them more accurate. But it won't replace what they do in [9]. It is unlikely that the proposed system will ever replaced the physicians and the radiologist but they will likely to become crucial elements of medical image analysis. The essence of developing a system like that needed to focus on detecting nodules in their early stages from images x-ray, which are the very small nodules that are likely to be overlooked by the radiologists. The difficulties for detecting lung nodules in radiographs are below:

- Nodule sizes are varying widely: Commonly a nodule diameter can take any value between a few millimeters up to several centimeters [11].
- Nodules exhibit a large variation in density—and Hence visibility on a radiograph—some nodules are only slightly denser than the surrounding lung tissue, while the densest ones are calcified [11].
- As nodules can appear anywhere in the lung field, they can be obscured by ribs, and structures beneath the diaphragm, resulting in a large variation of contrast to the background [11].

3. Proposed Method

Our work aims to apply some data mining techniques in the form of chest X-ray images for the diagnosis of lung cancer in its early stages. So we need to classify these images to a number of categories: normal and abnormal, which includes lung cancer in its early stages. Data mining includes many of the techniques or methods such as the characterization and classification of discrimination and association and compilation, and analysis of the development trend and deviation analysis, similarity analysis, etc... We chose the classification technique based on the type of knowledge to be mined. By inputting a training data set and building a model of the class attribute based on the rest of the attributes. In this way the classification will produce a function that is sorting out a data item into one of several predefined classes [12].

Our proposed system will classify the lung nodule into benign (non-cancerous) or malignant (cancer). Most of the cancers start with the appearance of small nodules. Nodule pixels are often brighter than the surrounding areas, but in some cases, the difference in grey levels is not significant. Furthermore, ribs and pulmonary arteries, which often have higher grey levels, also contribute to the complexity of lung tissue and make some nodules undetectable.

To detect and classify early nodules we employed data mining approach to extract these nodules from the X-ray images, these steps were shown in the following figure (**Figure 1**).

In the first step we applied pre-processing techniques to the images before the feature extraction phase, data preprocessing is necessary to get better the quality of the actual data for mining. Lung field segmentation procedures to isolate lung from chest X-ray, after enhancing the images with represent to the data cleaning phase, and image segmentation. Features relevant to the classification are extracted from the cleaned images. The extracted features are organized in a database in the form of transactions, which constitute the input for classification algorithm used.

In the classification phase we used artificial neural network (ANN) technique. It classifies the image into a normal or abnormal status.
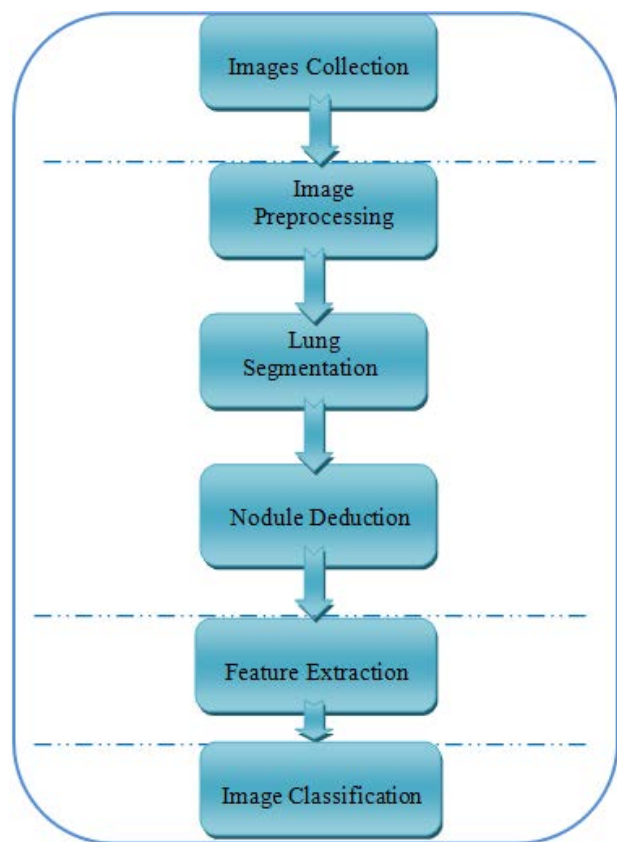

Figure 1. The steps of early detect the nodules by using data mining method.

Once a nodule has been detected as an abnormal status, it may be quantitatively analyzed as follows: Abnormal once classify to lung cancer or no lung cancer (it might also be scar tissue from a previous infection or pulmonary).

3.1. Data Preprocessing

Data preprocessing involves data cleaning, data transformation, data integration, data reduction or data compression for compact representation, etc. Preprocessing techniques varies depending on the type of data, and on the type of knowledge to be mined. In our work we will apply two preprocessing techniques, namely Data Cleaning and Data Transformation, were applied to the images collection. Data Cleaning is the process of cleaning the data by removing noise. It could mislead the actual mining process. In our case, there are some images that are very large and almost 50% of the whole image comprised the background with a lot of noise. In addition, these images were scanned at deferent illumination conditions, and therefore some images appeared too bright and some were too dark, **Figure 2** shows image before apply preprocessing techniques.

The first step toward noise removal was resizing the images with the help of the resizing operation in Image Processing. We used also resizing operation to reduce size image to 1/4. The next step towards preprocessing the images was using image enhancement techniques. Image enhancement is a set of image processing operations applied on images to produce good images. Image enhancement helps in qualitative improvement of the image with respect to a specific application Enhancement can be done either in the spatial domain or in the frequency domain. Here we work with the spatial domain and directly deal with the image plane itself, in other word enhancement techniques are based on direct manipulation of pixels in an image. In order to diminish the effect of over-brightness or over-darkness in images, and

at the same time accentuate the image features, we applied the Histogram Equalization method, which is a widely used technique. But the noise removal step was necessary before this enhancement because, otherwise, it would also result in enhancement of noise.

Order filters are based on a particular treatment of image statistics called order statistics. These filters run in the neighborhood of a certain pixel, known as window, and they replace the value of the central pixel.

Order statistics is a technique that organizes all pixels in a window in sequential order, on the basis of their grey level. The M mean in a set of values is such that half of the values in the set are smaller than M and half of the values are greater than M. In order to filter the mean in the area around the neighborhood, we ranked the intensities in the neighborhood, we determined the mean, and assigned the last to the intensity of the pixel.

The main point of mean filtering is to cause the points with very different intensities to become like to their neighbors, thus eliminating any isolated intensity peaks that become visible in the area of the filter mask. The median filter is a nonlinear filter, used in order to eliminate the high-frequency filter without eliminating the significant characteristics of the image. A 5×5 mask is used, which is centered on each image pixel, replacing each central pixel by the mean of the twenty five pixels covering the mask, **Figure 3** shows image after apply median filter.

The window size allows e characteristics of the image to me preserved while at the same time eliminating high frequencies next the automatic cropping is performed. The idea of this step is to focus the process exclusively on the relevant lung region, which reduces the possibility for erroneous classification by areas which are not of interest. Histogram Equalization increases the contrast range in an image by increasing the dynamic range of grey levels. As the low-contrast image's histogram is narrow and centred towards the middle of the gray scale, by distributing the histogram to a wider range will improve the quality of the image, **Figure 4** shows image (after apply histogram equalization).

Figure 2. Image before apply preprocessing techniques.

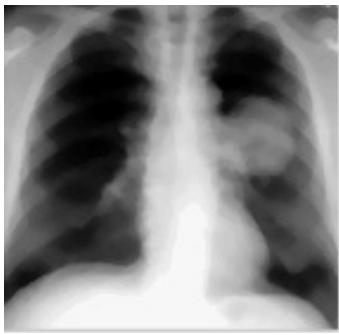**Figure 3. Image after apply median filter.**

Figure 4. Image after apply Histogram Equalization.

3.2. Image Segmentation

In numerous image applications image segmentation is a significant step. A host of techniques and algorithms usually fall into this general category as a starting point for edge detection, region labeling, and transformations. All these techniques, region labeling, and analyses, are relatively simple algorithms that have been used for many years to isolate, measure, and identify potential regions [13].

Morphology is an image processing methodology, a solid knowledge of this methodology enables the quick development of algorithms and products for solving complex detection, segmentation, and pattern recognition problems. Morphology has been successfully applied in many areas, including medicine (MR imagery, CT imagery, X-rays), biology, radar, sonar, infra-red, and remote sensing images, industrial inspection, robotics, etc. It addresses binary and grayscale images, as well as color and 3-D imagery. Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors by choosing the size and shape of the neighborhood [14].

Segmentation procedures have been proposed due to the approximate intensity value between the lung region and the unwanted background region. The most important step for the process is thresholding, followed by morphological operation to remove unwanted noise. Other morphological operation such as erosion and dilation operation will also be implemented. Erosion operator makes a region smaller while dilation operator enlarges a region. Median filtering also have been used to smooth the lung boundaries and finally, masking of image is done obtaining the segmented lung (**Figure 5**).

A stack method is used for region labeling, as it is one of the fastest and simplest ways to implement. After labeling, the regions which are not of interest to the study are eliminated. It is known that the surface covered by
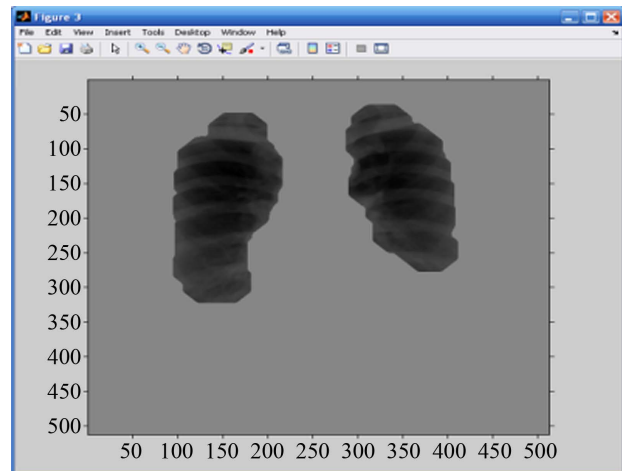

Figure 5. Lung segmentation.

the lung is over 80%; therefore, isolated areas with surfaces smaller than 1% do not belong to the lung and are eliminated through the creation of masks obtained from neighboring pixels.

3.3. Feature Extraction

The feature extraction phase is required in order to build the transactional database to be mined. The features that were extracted were organized in a database, which is the input for the mining phase of the classification system. Many kinds of features have been proposed for discriminating between normal tissues and abnormal ones. But there are optimal feature set from a large number of features which enable a diagnosis system for lung cancer screening to take large step toward a practical application. Our system consists of two processing steps:

1) Extraction of the boundaries of tumor candidates.

By using thresholding, and also the morphological operation of removing unwanted pixel in an image, other morphological operation such as erosion and dilation operation will also be implemented (**Figure 6**).

2) Extraction of feature parameters.

These features act as the basis for classification process. Regarding on the type of the image processed as a binary image presented in black and white colors. Thus, only three features were considered to be extracted; area, shape and perimeter. The features are defined as follows:

1) Area: it is a scalar value that gives the actual number of overall nodule pixel. It is obtained by the summation of areas of pixel in the image that is registered as 1 in the binary image obtained.

2) Perimeter: it is a scalar value that gives the actual number of the outline of the nodule pixel. It is obtained by the summation of the interconnected outline of the registered pixel in the binary image.

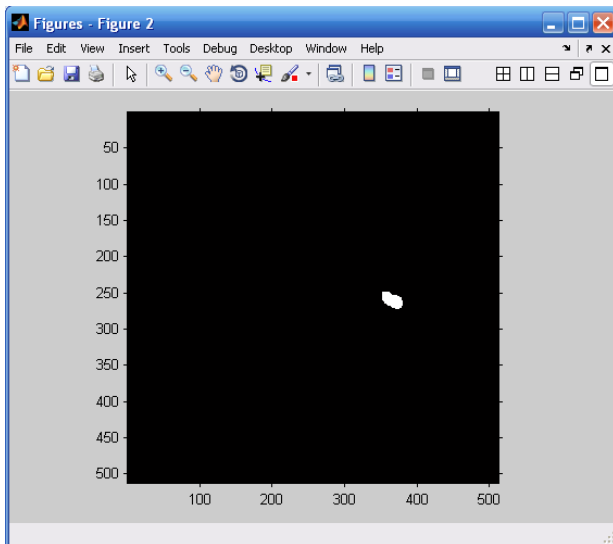3) Shape: a scalar value that can be defined in Equation (1)

Figure 6. Nodule detection.

$$shape = \frac{4*pi*area}{perimeter^2} \qquad (1)$$

This equation is equal to one only for a round (round and regular) and it is less than one for any other shape (irregular) [15].

3.4. Transactional Database Organization

All the extracted features presented above have been computed over smaller windows of the original image. The features extracted were discretized over intervals before organizing the transactional data set. When all the features were extracted the transactional database to be mined was built in the following way. For the normal images, all the features extracted were attached to the corresponding transaction, while those characterizing an abnormal chest X-ray features extracted from abnormal parts are attached as well.

3.5. Artificial Neural Network (ANN)

An **artificial neural network** (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase [16].

The structure of the neural network consists of three layers: the input layer, the hidden layer and the output layer. The number of nodes in the input layer is equal to the number of elements existing in one transaction in the database. In our case, the input layer had 3 nodes. In the hidden layer, we chose 10 nodes, while the output layer

consist of 3 nodes (**Figure 7**).

The nodes represented by output layer will apply the classification procedure on the image. It will classifies the image as normal if the image benign (non-cancerous) or malignant (cancer).

We use a feed-forward network with tan-sigmoid transfer function to generate their output. In this case the net neuron input is mapped into values between +1 and −1 [17]. The neuron transfer function can be shown in **Figure 8**:

In the training phase, back propagation procedures will be used. Back propagation algorithm is based on the error correction rule. Error propagates via forward and backward pass where weight is fixed and adjusted. Finally, a set of outputs is produced as the actual response of the network. The application divides the input vectors and target vectors into three sets:

- 70% are used for training.
- 15% are used to validate that the network is generalizing and to stop training before over fitting.
- The last 15% are used as a completely independent test of network.

The internal weights of the neural network are adjusted according to the transactions used in the learning process. For each training transaction the neural network receives in addition the expected output. This allows the modification of the weights. In the next step, the trained neural network is used to classify new images. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Usually, such an input/target pairs are needed to train a network [17]. The flow chart in **Figure 9** shows the brief description of the Feed Forward Back propagation algorithm to early diagnosing lung cancer system.
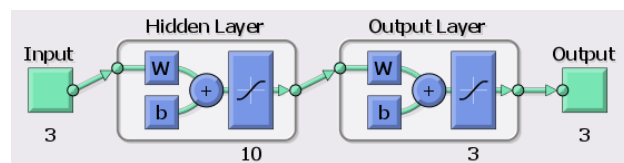

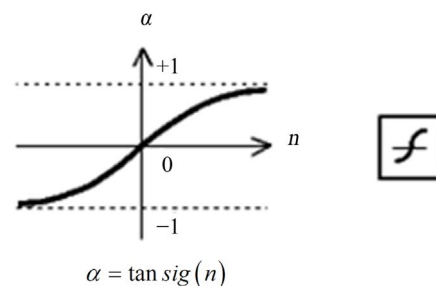**Figure7. The structure of the neural network.**

$$\alpha = \tan sig(n)$$

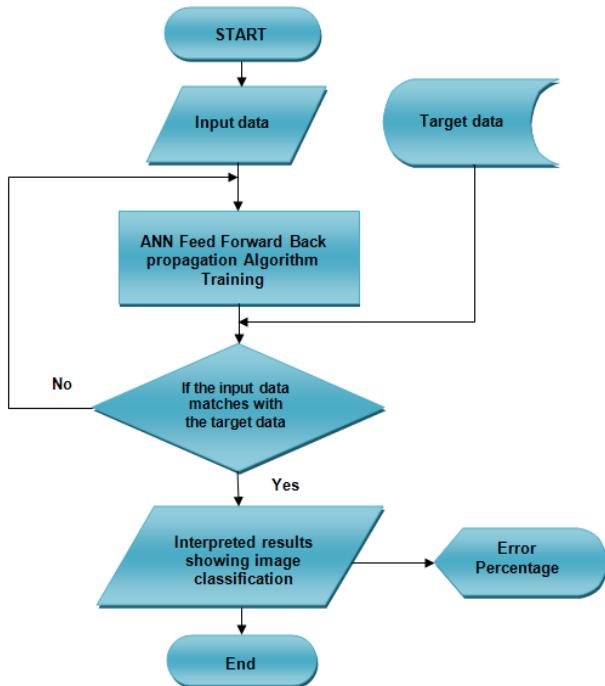**Figure 8. Illustrate the tan-sigmoid transfer function.**

Figure 9. Flow chart showing the description of the ANN-FFBP algorithm.

To define a pattern recognition problem, we arrange a set of Q input. Followed by arrange another set of Q target vectors so that they indicate the classes to which the input vectors are assigned. There are two approaches to create the target vectors. One approach can be used when there are only two classes; it sets each scalar target value to either 1 or 0, indicating which class the corresponding input belongs to. Alternately, target vectors can have N elements, where each target vector is represented as follow:

One element is represented as 1 and the others are represented as 0 respectively. This defines a problem where inputs are classified into N different classes. There are three classes of target vectors (normal, benign, malignant), are shown in **Table 1**. Input vectors are obtained from the feature extracted from the images, are shown in **Table 2**:

This data set consists of 60 records for about 60 patients. In our experiment we randomly pick up 9 elements divided as three-element input vectors and three-element target vectors. We want to classify a chest X-ray to three classes (normal, benign or malignant), based on these features, neural network is adjusted, based on a comparison of the output and the target, until the network output matches the target, it iteratively learns a set of weights for prediction of the class label.

The network is adjusted, based on a comparison of the output and the target data. It will be repeated until the network output matches the target data. The weights are

Table 1. Represents the target vector data.

Probability of malignant	Probability of benign	Probability of normal
0	0	1
0	1	0
1	0	0

Table 2. Results data of input vectors.

Shape	Perimeter	Area
0	0	0
0.8612312	103.84062	739
0.6640143	134.2254	952

modified to minimize the mean squared error between the network's prediction (**output**) and the actual target value. The error values are shown in **Table 3** generated by output neurons and it back propagates as a regulation to the synaptical connections towards the input in order to adjust the output obtained to the expected output as faithfully as possible.

Mainly, the back propagation algorithm is based on error minimization by way of a traditional optimization method called gradient descent which represents the key point. The key point consists of calculating the proper weights of the layers from the errors in the output units. It can be define as a secret lies in evaluating the consequences of an error and dividing the value among the weights of the contributing network connections.

4. Experimentation and Results

The proposed system through all its processing steps will detect whether the supplied lung image had cancer or not. The result obtained from the classification process of the lung cancer image is shown in **Table 4**.

The variation of classification accuracy obtained is due to the features that have been used for each of the image category. It can be observed that there is no misjudgment obtained for normal image, this is due to the high variation of the image features obtained compared to the features obtained for stage two and stage three images. Therefore, from the result it is confirmed that the network is successful in classifying the lung cancer image applied with higher accuracy and lower misjudgment obtained. In varying the number of epochs of the network, the best network with the used of 19 epochs were chosen to be applied for the next variation of parameters. The successful network that has been developed is then ready for the classification process.

5. Conclusion

According to the above concepts we purposed and de-

Table 3. The final results.

Input vectors			Target vector			Output vectors			Error		
0	0	0	0	0	1	0.0098	0.0001	0.9995	−0.0098	−0.0001	0.0004
0.6640143	134.2254	952	1	0	0	0.9978	0.0113	5.0505	0.0021	−0.0113	−0.0005
0.9866212	57.213203	257	0	1	0	4.8257	0.9998	0.0023	0.0001	−0.0002	−0.0023

Table 4. Result obtained for lung cancer image classification.

Image category	Number of samples	Correct recognition	Misclassification	Classification accuracy (%)
normal	24	24	0	100
begign	22	21	1	95
malignant	14	12	2	85

veloped an automatic system for early detection of lung cancer by analyzing chest X-ray images using several steps. In generating the system, MATLAB has been the most important tools in implementing our project. Medical image classification is an important thing in medicine. It allows biological structures to be isolated non-invasively. It is used for diagnostic purposes or practically applied in image guided surgeries; image classification has many forms and uses. Unfortunately, currently there is no classification strategy that can accommodate all its applications.

In image processing procedures, the most challenging part of image processing was the segmentation of the lung region due to the limitation regarding the similarities of the intensity in the X-ray image. The lung cancer nodule detection process does not seem to be the problem since the absent of the similar intensity due to the lung segmentation completed.

Followed by the ANN procedures, a difficulty was obtained for choosing the best architecture for the ANN. Besides, varying several parameters, such as number of epochs, number of hidden nodes, and learning rate value, the best architecture has been successfully chosen thus generating a successful CAD system for the project. Our system has successfully classified several stages of lung cancer. It is unlikely that the proposed CAD system will ever replace the physicians and the radiologist but they will likely become crucial elements of medical image analysis. Therefore, we will summarize the main concern of this work in the following steps:

1) In this work, we used some data mining, techniques such as neural networks for detection and classification Lung Cancer in X-ray chest films.

2) We also considered 60 X-ray chest films multimedia database as a training dataset to be used in our proposed classification system.

3) These sets of images are considered 70 percent as a training value of the systems and 15 percent for testing it. Fifteen splits of the data collection were considered to compute all the results in order to obtain more accurate result of the system potential.

4) A classification of the digital X-ray chest films was obtained into two categories: normal and abnormal. The normal ones are those characterizing a healthy patient. The abnormal ones could include types of lung cancer.

5) Some procedures considered as an essential part to the task of medical image mining theses procedures including Data Preprocessing, Feature Extraction were indicated in this work as well.

Finally, in this work, we used classification, neural network method to classify problems aiming at identifying the characteristics that indicate the group to which each case belongs.

REFERENCES

[1] Cancer Council Victoria, "Lung Cancer: For People with Cancer, Their Family and Friends," Cancer Council Victoria, Melbourne, 2008.

[2] S. Sumathi and S. N. Sivanandam, "Introduction to Data Mining and Its Applications," Springer, Berlin Heidelberg New York, 2006.

[3] Md. Rafiqul Islam, M. U. Chowdhury and S. M. Khan, "Medical Image Classification Using an Efficient Data Mining Technique," School of Information Technology Deakin University, Burwood, 2004.

[4] J. W. Han, "Data Mining: Concepts and Techniques," 2nd Edition, University of Illinois, Urbana-Champaign, 2006.

[5] L. Khan and V. A. Petrushin, "Multimedia Data Mining and Knowledge Discovery," 2007.

[6] A. A. Freitas, "A Genetic Programming Framework for Two Data Mining Tasks: Classification and Generalized Rule Induction," 1997.

[7] Z.-H. Zhou, Y. Jiang, Y.-B. Yang and S.-F. Chen, "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles," National Laboratory for Novel Software Technology, Nanjing University, Nanjing, 2002.

[8] D. Harvey, "CAD Applications in Liver, Prostate, Breast, and Lung Cancer," Radiology Today, 2010.

[9] M. Gomathi and P. Thangaraj, "Computer Aided Medical Diagnosis System for Detection of Lung Cancer Nodules a Survey," The Free Library, 2012, pp. 3-12.

[10] http://lungcancer.about.com

[11] B. Magesh, Pgscholar, P. Vijayalakshmi and M. Abirami, "Computer Aided Diagnosis System for the Identification and Classification of Lessions in Lungs," *International Journal of Computer Trends and Technology*, 2011.

[12] C. Tjortjis and J. Keane, "A Classification Algorithm for Data Mining," Department of Computation, UMIST, Manchester, 2002.

[13] G. Ferrero, P. Britos and R. García-Martínez, "Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks," 2006.

[14] L. Vincent, "Current Topics in Applied Morphological Image Analysis," 2000.

[15] A. A. Abdullah and S. M. Shaharum, "Lung Cancer Cell Classification Method Using Artificial Neural Network," *Information Engineering Letters*, Vol. 2, No. 1, 2012, pp. 49-59.

[16] Y. Singh and A. S. Chauhan, "Neural Networks in Data Mining," Journal of Theoretical and Applied Information Technology, 2005-2009.

[17] M. H. Beale, M. T. Hagan and H. B. Demuth, "Neural Network," The MathWorks, 1992-2012.