IDENTIFY AND CLASSIFICATION OF LUNG CANCER: A DEEP LEARNING

[1] Mrs.M.Ganthimathi,[2] D.Dhasarathan,[3] K.Murali Dharan,[4] D.Parameshwaran.

[1] Associate Professor, Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram, Namakkal.

[2][3][4] Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram, Namakkal.

Abstract: Deep Learning techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the accuracy factor is very important to discover the abnormality issues in target images, especially in various cancer tumors such as lung cancer etc. proposed to segment lung parenchyma using a convolutional neural network (CNN) model. To reduce the workload of manually preparing the dataset for training the CNN, one clustering algorithm based method is proposed firstly. Specifically, after splitting CT slices into image patches, the k-means clustering algorithm with two categories is performed twice using the mean and minimum intensity of image patch, respectively. A cross-shaped verification, a volume intersection, a connected component analysis and a patch expansion are followed to generate final dataset. Secondly, we design a CNN architecture consisting of only one convolutional layer with six kernels, followed by one maximum pooling layer and two fully connected layers. Using the generated dataset, a variety of CNN models are trained and optimized, and their performances are evaluated by eightfold crossvalidation. Following the segmentation principles, an enhanced region of the object of interest that is used as a basic foundation of feature extraction is obtained.

Index Terms— Lung disease Detection, Convolutional Neural Network, Deep Learning, Medical Images.

1. INTRODUCTION

Lung cancer is one of the commonest cancers in the industrialized world, and persons with this grave disease must deal not only with the physical effects but also with the psychosocial aspects. Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease
is detected in time. Among different types of cancer, lung cancer is the most aggressive, and the best practice for its accurate prognosis is the determination of the current stage of the disease.

Three main factors in cancer staging are primary tumor, regional lymph nodes, and metastasis. One of the most important and difficult tasks a doctor has to carry out is the detection and diagnosis of cancerous lung nodules from x-ray image results. Given that lung cancer is one of the common cancers worldwide, the implications of focusing on the quality of life as well as survival require to be understood. Early detection is the most important for reducing death due to lung cancer. The early detection of lung cancer is a challenging problem, due to both the structure of the cancer cells and the stained methods which are used in the preparation of the sputum cells. Now there are three main methods to use for the diagnosis of lung cancer: biochemical diagnosis (serology and immunology), imaging diagnosis, and cytology histology diagnosis [9. Previous researches have shown that the major types of lung cancer are associated with cigarette smoking. A lot of research has been done on the detection of areas with high cancer incidence by domestic and foreign researchers, but the research did not relieve the increased lung cancer incidence and mortality rates.

Most of the techniques are detecting lung cancer in its advanced stages, where the patients' chance of survival is very low. Therefore, there is a great need for new technology to diagnose lung cancer in its early stages. Deep Learning and data mining techniques provide a good quality tool for improving the manual analysis. Medical Data Mining is a promising area of computational intelligence applied to automatically analyse patient's records aiming at the discovery of new knowledge useful for medical decision-making. Applying data mining techniques to cancer data is useful to rank and link cancer attributes to the survival outcome. Further, accurate outcome prediction can be extremely useful for doctors and patients to not only estimate survivability, but also aid in decision making to determine the best course of treatment for a patient, based on patient-specific attributes, rather than relying on personal experiences, anecdotes, or populationwide risk assessments. Lung cancer is the most dangerous
and widespread cancer in the world according to the stage of discovery of the cancer cells in the lungs, so the process of early detection of the disease plays a very important and essential role to avoid seriously advanced stages to reduce its percentage of distribution.

METHOD TYPES

The two types of methods used for Deep Learning are Analog and Digital Image Processing. Analog or visual techniques of Deep Learning can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in Deep Learning through visual techniques. So, analysts apply a combination of personal knowledge and collateral data to image processing.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

In recent years, advances in information technology and telecommunications have acted as catalysts for significant developments in the sector of health care. These technological advances have had a particularly strong impact in the field of medical imaging, where film radiographic techniques are gradually being replaced by digital imaging techniques, and this has provided an impetus to the development of integrated hospital information systems and integrated teleradiology services networks which support the digital transmission, storage, retrieval, analysis, and interpretation of distributed multimedia patient records. One of the many added-value services that can be provided over an integrated teleradiology services network is access to high-performance computing facilities in order to execute computationally intensive image analysis and visualization tasks. In general, currently available products in the field of Deep Learning and image processing (IP) meet only specific needs of different end user groups. They either aim to provide a comprehensive pool of ready to use software within a user-friendly and application specific interface for those users that use IP software, or aim for the specialised IP researcher and developer, offering programmer's libraries and visual language tools. However, we currently lack the common framework that will integrate all prior efforts and developments in the field and at the same time provide added-value features that support and in essence realise what we call a 'service'. In the case of Deep Learning, these features include: computational resource management and intelligent execution scheduling; intelligent and customizable mechanisms for the description, management, and retrieval of figure 1.1 Digital Image Processing software modules; mechanisms for the "plug-and-play" integration of already existing heterogeneous software modules; easy access and user transparency in terms of software, hardware, and network technologies; sophisticated charging mechanisms based on quality of service; and, methods for the integration with other services available within an integrated health telematics network.

MEDICAL IMAGING

Medical imaging is the technique and procedure of creating a visual demonstration of the internal body of a body for experimental analysis and health intervention. Medical imaging seeks out to disclose internal structures hidden by the skin and bones, as well as to diagnose and treat disease. Medical imaging also establishes a database of normal anatomy and physiology to make it possible to identify an abnormality. Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are usually considered part of pathology instead of medical imaging.

As a discipline and in its widest sense, it is part of biological imaging and incorporates radiology which uses the imaging technologies of X-ray radiography, magnetic resonance imaging, medical ultrasonography or ultrasound, endoscopy, elastography, tactile imaging, thermography, medical photography, and medicine functional techniques as positron emission tomography.

In the clinical context, "imperceptible light" medical imaging is generally associated with radiology or "medical imaging" and the medical practitioner responsible for understanding (and sometimes acquiring) the images is a radiologist. "Visible light" medical imaging involves digital video or still pictures that can be seen without special equipment. Dermatology and wound care are two modalities that use visible light imagery. Diagnostic radiography designates the technical aspects of medical imaging and in particular the acquisition of medical images. The radiographer or radiologic technologist is usually responsible for acquiring medical images of diagnostic quality, although some radiological interventions are performed by radiologists.

As a field of scientific investigation, medical imaging constitutes a sub-discipline of engineering, medical, or medicine depending on the context: Research and development in the area of instrumentation, image acquisition (e.g. radiography), modelling and quantification are usually the preserve of biomedical engineering, medical physics, and computer science; Research into the application and interpretation of medical images is usually the preserve of radiology and the medical sub-discipline relevant to medical condition or area of medical science (neuroscience, cardiology, psychiatry, psychology, etc.) under investigation. Many of the techniques developed for medical imaging also have scientific and industrial applications.[3]

Medical imaging is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. In this restricted sense, medical imaging can be seen as the solution of the mathematical inverse. This means that cause (the

properties of living tissue) is inferred from effect (the observed signal). In the case of medical ultrasonography, the probe consists of ultrasonic pressure waves and echoes that go inside the tissue to show the internal structure. In the case of projection radiography, the probe uses X-ray radiation, which is absorbed at different rates by different tissue types such as bone, muscle, and fat. The term non-invasive is used to denote a procedure where no instrument is introduced into a patient's body which is the case for most imaging techniques used. The MIPAV (Medical Image Processing, Analysis, and Visualization) application enables quantitative analysis and visualization of medical images of numerous modalities such as PET, MRI, CT, or microscopy. Using MIPAV's standard user-interface and analysis tools, researchers at remote sites (via the internet) can easily share research data and analyses, thereby enhancing their ability to research, diagnose, monitor, and treat medical disorders. MIPAV is a Java application and can be run on any Java-enabled platform such as Windows, UNIX, or Macintosh OS X.

GOALS FOR MIPAV
MIPAV is to meet the following goals:
▪   To develop computational methods and algorithms to analyse and quantify
biomedical data;
▪   To collaborate with NIH researchers and colleagues at other research canter's in applying information analysis and visualization to biomedical research problems;
To develop tools (in both hardware and software) to give our collaborators the ability to analyse biomedical data to support the discovery and advancement of biomedical knowledge.

NEED FOR MIPAV
Imaging has become an essential component in many fields of biomedical research and clinical practice. Biologists study cells and generate 3D confocal microscopy data sets, virologists generate 3D reconstructions of viruses from micrographs, radiologists identify and quantify tumors from MRI and CT scans, and neuroscientists detect regional metabolic brain activity from PET and functional MRI scans. Analysis of these diverse types of images requires sophisticated computerized quantification and visualization tools. To support scientific research in the NIH intramural program, CIT has made major progress in the development of a platformindependent, n-dimensional, general-purpose, extensible Deep Learning and visualization program.

II.IMPLEMENTATION DETAILS

Upload datasets
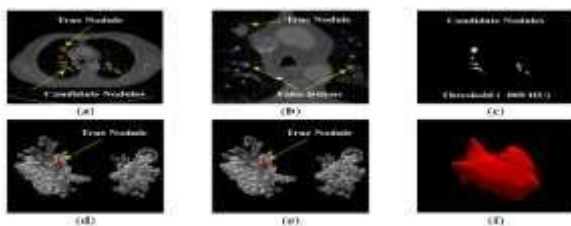
Lung cancer is one of the most common and deadly diseases in the world. Detection of lung cancer in its early stage is the key of its cure. In general, measures for early stage lung cancer diagnosis mainly includes those utilizing X-ray chest films, CT, MRI, isotope, bronchoscopy, etc., among which a very important measure is the so-called pathological diagnosis that analyses the specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. The lung images and Datasets are uploaded to diagnosis the lung cancer.

Preprocessing
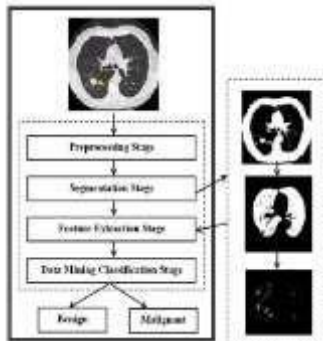
The goal of the Median filter is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters preprocessing modules are designed for a desired frequency response. Median filtering is a nonlinear operation often used in Deep Learning to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.
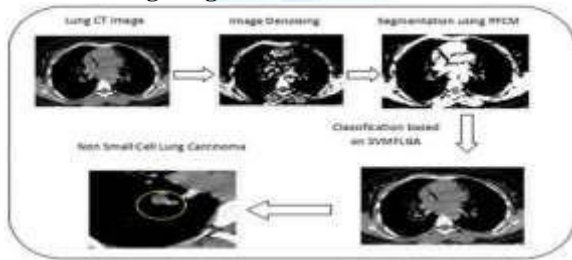
Image Segmentation

Image segmentation is an important process for the most part of image analysis consequent assignments. In particular, many of the previous methods for segmentation image report and identification depend especially on the segmentation outcomes. Segmentation separates the picture into its ingredient sections or things.

Figure 4.4 segmentation of medical images in 2D, piece by piece has many useful functions for the medical expert such as: visualization and quantity evaluation of things of attention, discovery of tumors. The goal of segmentation is to make simpler and/or modify the demonstration of the image into amazing that is more significant and easier to investigate. Image segmentation is classically used to situate things and borders (lines, curves, etc.) in images. The primary group is to separate the image based on unexpected modified intensity values, such as edges in an image. The second type is based on dividing the image into sections that are similar to the according predefined measure. CNN is an effective way of segmenting out objects in pictures containing both random noise and shading. This is illustrated both on mathematically created pictures and on some obtained from medical imaging. We segment the lung regions based on lung segmentation. Feature Extraction
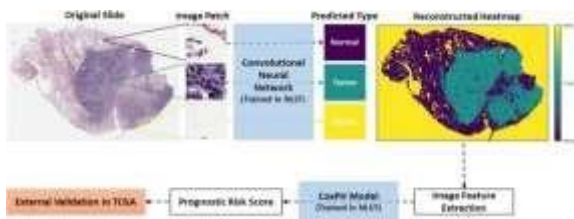

Image features Extraction stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes (features) of a given image. To predict the probability of lung cancer presence, In fact, the proposed approach does the exact opposite; it tries to create a large number of weak features and expects the classifier to weight them according to their relevance during training using above feature extraction. In general, many of the features may turn out to be irrelevant for a given application. However, our approach begins with a conservative feature set in order for it to be application-independent, at the expense of increased training

complexity. In feature extraction, we calculate the size and shape of the tumor identified by calculating the diameter value of that tumor and provides result in millimeter (mm).

CNN
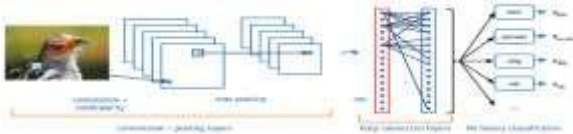
The proposed CNN architecture (shown in table 1) mainly consists of the following layers: two convolution layers which follow two max-pooling layers and one fully-connected layer with two softmax units. As shown in figure 1, the network begins with a convolution layer, in which the first convolution layer takes the image with input size of 50×50 pixels. The second convolution layer consists of 32 feature maps with the convolution kernel of 3×3. The kernel size for max pooling layers is 2×2 and the stride of 2 pixels, and the fully-connected layer generates an output of 1024 dimensions. These 10 outputs are then passed to another fully connected layer containing 2 softmax units, which represent the probability that the image is containing the lung cancer or not. Note that each convolution layer in our cnn model is followed by a rectified linear unit (relu) layer to produce their outputs. After applying these architectures, some images detected with cancerous nodules and some identified as non-cancerous.

III.RESULT ANALYSIS
MATLAB
MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include:
Math and computation
Algorithm development
Modeling, simulation, and prototyping
Data analysis, exploration, and visualization
Scientific and engineering graphics
Application development, including graphical user interface building
MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar noninteractive language such as C or Fortran.
The name MATLAB stands for matrix laboratory.
MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects. Today, MATLAB uses software developed by the LAPACK and ARPACK projects, which together represent the state-of-the-art in software for matrix computation.
MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In industry, MATLAB is the tool of choice for high-productivity research, development, and analysis.

TOOLBOXES
MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to learn and apply specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.
THE MATLAB SYSTEM
The MATLAB system consists of five main parts: Development Environment. This is the set of tools and facilities that help you use MATLAB functions and files. Many of these tools are graphical user interfaces.
It includes the MATLAB desktop and Command Window, a command history, and browsers for viewing help, the workspace, files, and the search path.
The MATLAB Mathematical Function Library. This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.
The MATLAB language: This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and objectoriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

HANDLE GRAPHICS

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and threedimensional data visualization, Deep Learning and image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete graphical user interfaces on your

MATLAB FUNCTIONS

A MATLAB "function" is a MATLAB program that performs a sequence of operations specified in a text file (called an m-file because it must be saved with a file extension of *.m). A function accepts one or more MATLAB variables as inputs, operates on them in some way, and then returns one or more MATLAB variables as outputs and may also generate plots, etc.

Some functions are

Imread() – Reading the image from the graphics file. A = imread(filename, fmt) reads a grayscale or color image from the file specified by the stringfilename. If the file is not in the current folder, or in a folder on the MATLAB path, specify the full pathname.

The text string fmt specifies the format of the file by its standard file extension. For example, specify'gif' for Graphics Interchange Format files. To see a list of supported formats, with their file extensions, use the imformats function. If imread cannot find a file named filename, it looks for a file named filename.fmt.

The return value A is an array containing the image data. If the file contains a grayscale image, A is an Mby-N array. If the file contains a truecolor image, A is an M-by-N-by-3 array. For TIFF files containing color images that use the CMYK color space, A is an M-byN-by-4 array. The class of A depends on the bits-persample of the image data, rounded to the next byte boundary. For example, imread returns 24-bit color data as an array of uint8 data because the sample size for each color component is 8 bits.

[X, map] = imread(...) reads the indexed image in filename into X and its associated colormap into map. Colormap values in the image file are automatically rescaled into the range [0,1].

IMAGE-WRITE IMAGE TO GRAPHICS FILE

imwrite(A,filename,fmt) writes the image A to the file specified by filename in the format specified by fmt. A can be an M-by-N (grayscale image) or M-by-N-by3 (truecolor image) array, but it cannot be an empty array. For TIFF files, A can be an M-by-N-by-4 array containing color data that uses the CMYK color space. For GIF files, A can be an M-by-N-by-1-by-P array containing grayscale or indexed images — RGB images are not supported. For information about the class of the input array and the output image. filename is a string that specifies the name of the output file. fmt can be any of the text strings listed. This list of supported formats is determined by the

MATLAB image file format registry.

See imformats for more information about this registry. imwrite(X,map,filename,fmt) writes the indexed image in X and its associated colormap map tofilename in the format specified by fmt. If X is of class uint8 or uint16, imwrite writes the actual values in the array to the file. If X is of class double, imwrite offsets the values in the array before writing, using uint8(X−1). map must be a valid MATLAB colormap. Note that most image file formats do not support colormaps with more than 256 entries. When writing multiframe GIF images, X should be an 4-dimensional M-by-N-by-1by-P array, where P is the number of frames to write. imwrite(...,filename) writes the image to filename, inferring the format to use from the filename's extension. imwrite(...,Param1,Val1,Param2,Val2...) specifies parameters that control various characteristics of the output file for HDF, JPEG, PBM, PGM, PNG, PPM, and TIFF files. For example, if you are writing a JPEG file, you can specify the quality of the output image.

For the lists of parameters available for each format.

MATLAB APPLICATIONS

The MATLAB Application Program Interface (API). This is a library that allows you to write C and Fortran programs that interact with MATLAB. It include facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.
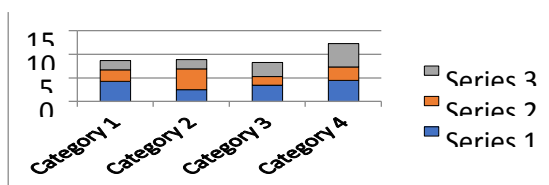
What Is Simulink?

Fig 5.1 MATLAB Table

Simulink, a companion program to MATLAB, is an interactive system for simulating nonlinear dynamic systems. It is a graphical mousedriven program that allows you to model a system by drawing a block diagram on the screen and manipulating it dynamically. It can work with linear, nonlinear, continuous-time, discrete-time, multirate, and hybrid systems. Blocksets are add-ons to Simulink that provide additional libraries of blocks for specialized applications like communications, signal processing, and power systems.

Real-Time Workshop is a program that allows you to generate C code from your block diagrams and to run it on a variety of real-time systems.

IV.LITERATURE REVIEW

A Fast and Robust Level Set Method for Image Segmentation Using Fuzzy Clustering and Lattice Boltzmann Method Author(s): Souleymane Balla

In the LSM, the movement of the zero level set is actually driven by the level set equation (LSE), which is a partial differential equation (PDE). For solving the LSE, most classical methods such as the upwind scheme are based on some finite difference, finite volume or finite element approximations, and an explicit computation of the curvature. Unfortunately, these methods cost a lot of CPU time. Recently, the lattice Boltzmann method (LBM) has been used as an alternative approach for solving LSE. It can better handle the problem of time consumption because the curvature is implicitly computed and the algorithm is simple and highly parallelizable. The LBM is used to solve the LSE. The proposed method is based on the approach of the LBM PDE solver defined. In the proposed method, using a modified CNN objective function, we design a new fuzzy external force (FEF). The method is fast, robust against noise, and efficient whatever the position or the shape of the initial contour and can detect efficiently objects with or without edges. It has, first, the advantage of the CNN which gives it the latitude to stop the evolving curve according to the membership degree of the current pixel, second, the advantages of the LSM which allow it to handle complex shapes, topological changes, and different constraints on the contour smoothness, speed, size, and shape which are easily specified, and, third, the advantages of the LBM which make it very suitable for parallel programming due to its local and explicit nature.

Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer Author(s): Nisar Ahmed Memon

A patient undergoing a CT scan rests on a movable table at the center of a donut-shaped scanner, which is about 2.4 m (8 ft) tall. The CT scanner contains an X-ray source, which emits beams of X rays; an X-ray detector, which monitors the number of X rays that strike various parts of its surface; and a computer. The source and detector face each other on the inside of the scanner ring and are mounted so that they rotate around the rim of the scanner. Beams from the X-ray source pass through the patient and are recorded on the other side by the detector. As the source and detector rotate in a 360° circle around the patient, X-ray emissions are recorded from many angles. The resulting data are sent to the computer, which interprets the information and translates it into images that appear as cross-sections on a television monitor. By moving the patient within the scanner, doctors can obtain a series of parallel images, called slices. This series of slices are then analysed to understand the three-dimensional structure of the body. In medical imaging, segmentation is important for feature extraction, image measurements, and image display. In some applications, it may be useful to classify image pixels into anatomical regions, such as bones, muscles, and blood vessels, while in others into pathological regions, such as cancer, tissue deformities, and multiple sclerosis lesions. In some studies, the goal is to divide the entire image into subregions such as the white matter, gray matter, and cerebrospinal fluid spaces of the brain, while in others on the specific structure has to be extracted, for example, breast cancer from Magnetic Resonance images.

A Segmented Morphological Approach to Detect Tumor in Lung Images Author(s): Poonam Bhayan

Image processing is one of the most growing research areas these days. In the following, we are proposing an effective scheme to detect the abnormal formation of cells in the lungs. Here we present an approach that detects the tumor from the lung image. In this proposed approach we have applied a series of operations, first to enhance the image and then to detect the tumor from the lung image. First of all, we take an image and then enhance the contrast of an image by using the contrast stretching technique to adjust its contrast. After normalizing the image, we applied a series of steps to enhance the quality of an image and to remove the noise. In this approach enhancement of an image comes first, image enhancement aims to improve the interpretability of information in images for human viewers. Image enhancement techniques can be divided into two broad categories: Spatial domain methods, which operate directly on pixels, and Frequency domain methods, which operate on the Fourier transform of an image. When image enhancement techniques are used as preprocessing tools for other image processing techniques, then quantitative measures can determine which techniques are most

appropriate. For enhancement first of all we applied Fast Fourier Transform. We applied Fast Fourier Transform (FFT) to enhance the image quality. Fast Fourier Transform is a method to calculate computations that are large enough. It works on the divide and conquers approach; it breaks N input values into N/2 values, so that it becomes easy to calculate. After the application of FFT, we used another enhancement method named histogram modelling technique; this is the technique that can be used to improve the visual appearance of an image.

A Computer-Aided Diagnosis System for Lung Cancer Detection Using Support Vector Machine Author(s): M.Gomathi and P.Thangaraj

A New CAD System for Early Diagnosis of Detected Lung Nodules is proposed. The growth rate is predictable by measuring the volumetric variation of the detected lung nodules over time, so it is important to accurately measure the volume of the nodules to quantify their growth rate over time. In this study, the author introduces a novel Computer Assisted Diagnosis (CAD) system for early diagnosis of lung cancer. The projected CAD system involves five main steps. They are Segmentation of lung tissues from Computed Tomography (CT) images, Identification of lung nodules from segmented lung tissues, A non-rigid

registration technique to align two successive LDCT scans and to correct the motion artifacts caused by breathing and patient motion, Segmentation of the detected lung nodules and Quantification of the volumetric changes. This preliminary categorization results based on the analysis of the growth rate of both benign and malignant nodules for 10 patients (6 patients diagnosed as malignant and 4 diagnosed as benign) were 100% for 95% confidence interval. The experimental results of the proposed image analysis have yielded promising results that would supplement the use of current technologies for diagnosing lung cancer.

Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles Author(s): Zhi-Hua Zhou

In this paper, based on the recognition of the power of artificial neural network ensemble, an automatic pathological diagnosis procedure named Neural Ensemble-based Detection (NED) is proposed and realized in an early stage Lung Cancer Diagnosis System (LCDS). NED utilizes an artificial neural network ensemble to identify cancer cells in the images of the specimens of needle biopsies obtained from the bodies of the subjects to be diagnosed. The ensemble used is built on a specific two-level ensemble architecture and a novel prediction diction combining method, which achieves not only a high rate of overall identification but also a low rate of false-negative identification, i.e. a low rate of judging cancer cells to be normal ones we propose an automatic pathological diagnosis procedure named NED, which utilizes artificial neural network ensemble to identify lung cancer cells in the images of the specimens of needle biopsies. The core of NED is a two-level ensemble architecture that is composed of heterogeneous ensembles that not only comprise individual networks with a different number of output units but also employ different methods to combine individual predictions. In order to improve the accuracy of false-negative identification, we also devised a novel prediction combining method named full voting that is utilized in the first-level ensemble. Through adopting those techniques, NED achieves not an only high rate of overall identification but also a low rate of false-negative identification. NED has been realized in an early stage lung cancer diagnosis system LCDS, which is being transferred into a routine examination following the analysis of Xray chest films by Bayi Hospital currently.

V.CONCLUSION

In this research, we used a vanilla CNN classifier to determine whether a CT image of lung is cancerous or non-cancerous. Before using the CNN, we preprocessed the CT image through a thresholding technique. To have performed a
thorough experiment using 16 dataset. Our obtained detection accuracy is 80%, which is better than existing methods. In future, we will perform the experiments on a large amount of data and apply more features such as nodule size, texture and position for further improvement. We will also try to apply the state-of-the-art deep CNN methods for higher accuracy and use our method on other types of cancer detection
In future, we extend this project with pattern classification using classifiers such as SVM classifier. According to the classification techniques, our work could be improved evaluating other classification algorithm as support vector machines, as well as improving the feature selection algorithm. It could be also very interesting to train the ANN in presence of noise.

REFERENCES

[1]   Singapore   Cancer   Society, http://www. singaporecancersociety.org.sg.

[2]   B. Magesh, P. Vijaylakshmi, M. Abhiram,
"Computer aided Diagnosis System for identification and classification of Lessions in
Lungs", International Journal of Computer Trends and Technology- May to June Issue 2011.

[3]   Rachid Sammouda, Jamal Abu Hassan, Mohamed Sammouda, Abdulridha Al-Zuhairy, Hatem abou ElAbbas, "Computer Aided Diagnosis System for Early Detection of Lung Cancer Using Chest
Computer Tomography Images", GVIP 05 Conference, 19-21 December 2005, CICC, Cairo, Egypt.

[4]   American Cancer Society, "Cancer Statistics, 2005", CA: A Cancer Journal for Clinicians, 55:
        10-30,   2005,
"http://caonline.amcancersoc.org/cgi/content/full
/55/1/10".

[5]   Rachid Sammouda, Jamal Abu Hassan, Mohamed Sammouda, Abdulridha Al-Zuhairy, Hatem abou ElAbbas, "Computer Aided Diagnosis System for Early Detection of Lung Cancer Using Chest Computer Tomography Images", GVIP 05Conference, 19-21 December 2005, CICC,
Cairo, Egypt.

[6]   The DICOM Standards Committee. DICOM homepage. http: //medical.nema.org/, September 2004.

[7]   B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang and L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm", journal of applied clinical medical physics, vol. 4, (2003).

[8]   A. El-Baz, A. A. Farag, PH.D., R. Falk, M.D. and R. L. Rocco,M.D.,"detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I", Information Conference on Biomedical Engineering, Egypt (2002).

[9]   Linda G. Shapiro and G.C. Stockman., Computer
     Vision: Theory and Applications. 2001: Prentice.

 [10] 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).

[11]   Penedo. M. G, Carreira. M. J, Mosquera. A and Cabello. D, "Computer-aided diagnosis: a neuralnetwork- based approach to lung nodule detection", IEEE Transactions on Medical Imaging, vol: 17, pp: 872 – 880, 1998. 879

[12]   K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N.
Moriyama and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on helicalCT images", Compute. Med. Image Graph, vol. 22, no. 2(1998),pp. 157-167.

[13]   Yamomoto. S, Jiang. H, Matsumoto. M, Tateno.Y, Iinuma. T, Matsumoto. T, "Image processing for computer-aided diagnosis of lung cancer by CT (LSCT)", Proceedings 3rd IEEE Workshop on Applications of Computer Vision, WACV '96, pp: 236 –241, 1996. [14] Kanazawa. K, Kubo. M and Niki. N, "Computer aided diagnosis system for lung cancer based on helical CT images", Proceedings of the 13th International Conference on Pattern Recognition, pp: 381 - 385 vol.3, 1996.