FINDING AND CLASSIFYING LUNG CANCER IMAGE PROCESSING

ABSTRACT

Lung Cancer is a disease of uncontrolled cell growth in epithelium of the lung. Bio-mining is a process of Extracting non-trivial, implicit, previously unknown and patterns from large amount of biological sequences. The goal here is to identify Characteristics of patient segments where average survival is significantly higher / lower than average survival across the entire dataset. Without increasing the inherent information content in data image enhancement is done. The process is carried out with Feed Forward Algorithm of Artificial Neural Network (ANN). Multi-layer neural network has been employed to achieve image compression. Several other techniques, classification techniques are used on the pre-processed data along with various mining optimisations and validations. Hence the accuracy of the output is extracted to be 91.73%. , which is the highest accurate detection of the lung cancer cells among all other algorithms.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
CLR	Common Language Runtime
CLS	Common Language Specification
CNN	Convolutional Neural Network
CT	Computed Tomography
CTS	Common Type System
DBA	Database Administrators
DICOM	Digital Imaging and Communication in Medicine
EEG	Electron encephalogram
ELCAP	Early Lung Cancer Action Program
FFA	Feed Forward Algorithm
IDE	Integrated Development Environment
JPEG	Joint Photographic Expert Group
KNN	K Nearest Neighbor algorithm
MPEG	Moving Picture Experts Group
MRI	Magnetic Resonance Imaging
.NET	Network Enable Technologies
SQL	Structured Query Language

SVM	Support Vector Machine
UML	Unified Modeling Language
VB	Visual Basic
WDNN	Weighted Distance Nearest Neighbor
WEKA	Waikato Environment for Knowledge Analysis
XML	eXtensible Markup Language

LIST OF FIGURES

CHAPTER 1

1. INTRODUCTION

1.1 Image Processing

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

- ❖ Importing the image with optical scanner or by digital photography.
- ❖ Analysing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs.
- ❖ Output is the last stage in which result can be altered image or report that is based on image analysis.

1.2 Purpose of Image processing:

The purpose of image processing is divided into 5 groups. They are:

1. Visualization – Observe the objects that are not visible.
2. Image sharpening and restoration – To create a better image
3. Image retrieval – Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
5. Image Recognition – Distinguish the objects in an image.

1.3 Types of Image Processing:

The two types of methods used for Image Processing are Analog and Digital Image Processing. Analog or visual techniques of image processing 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 image processing 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.

1.4 Characteristics of Image Processing:

Before going to processing an image, it is converted into a digital form. Digitization includes sampling of image and quantization of sampled values. After converting the image into bit information, processing is performed. This processing technique may be, Image enhancement, Image restoration, and Image compression.

1.4.1 Image enhancement:

It refers to accentuation, or sharpening, of image features such as boundaries, or contrast to make a graphic display more useful for display & analysis. This process does not increase the inherent information content in data. It includes gray level & contrast manipulation, noise reduction, edge crispening and sharpening, filtering, interpolation and magnification, pseudo coloring, and so on.

1.4.2 Image restoration:

It is concerned with filtering the observed image to minimize the effect of degradations. Effectiveness of image restoration depends on the extent and accuracy of the knowledge of degradation process as well as on filter design. Image restoration differs from image enhancement in that the latter is concerned with more extraction or accentuation of image features.

1.4.3 Image compression:

It is concerned with minimizing the number of bits required to represent an image. Application of compression are in broadcast TV, remote sensing via satellite, military communication via aircraft, radar, teleconferencing, facsimile transmission, for educational & business documents, medical images that arise in computer tomography, magnetic resonance imaging and digital radiology, motion, pictures, satellite images, weather maps, geological surveys and so on.

- Text compression – CCITT GROUP3 & GROUP4
- Still image compression – JPEG
- Video image compression – MPEG

1.5 Advantages of Image Processing:

- The processing of images is faster and more cost-effective. One needs less time for processing, as well as less film and other photographing equipment.
- It is more ecological to process images. No processing or fixing chemicals are needed to take and process digital images. However, printing inks are essential when printing digital images.
- When shooting a digital image, one can immediately see if the image is good or not.
- Copying a digital image is easy, and the quality of the image stays good unless it is compressed. For instance, saving an image as jpg format compresses the image. By resaving the image as jpg format, the compressed image will be recompressed, and the quality of the image will get worse with every saving.
- Fixing and retouching of images has become easier. In new Photoshop 7, it is possible to smoother face wrinkles with a new Healing Brush Tool in a couple of seconds.
- The expensive reproduction (compared with rastering the image with a repro camera) is faster and cheaper.
- By changing the image format and resolution, the image can be used in a number of media.

CHAPTER 2

2. LITERATURE SURVEY

2.1 Title: Classification of Pancreas Tumor Dataset Using Adaptive Weighted k-nearest Neighbor Algorithm

Author: Mahmut KAYA, Hasan Şakir BİLGE

K nearest neighbor algorithm is a widely used classifier. It benefits from distances among features to classify the data. Classifiers based on distance metrics are affected from irrelevant or redundant features. Especially, it is valid for big datasets. So, some of features can be weighted with higher coefficients to reduce the effect of irrelevant or redundant features. It suggests adaptive weighted k nearest neighbor algorithm to increase classification accuracy. This algorithm uses t test which is one of the feature selection to weight features. Classification accuracy is increased from 74.14% to 86.57% for k=3 neighbors and Euclidean distance metric thanks to the proposed method. Classifying algorithms based on distances are simple and effective methods. These algorithms classify datasets by examining distances among features. Neighbor distances among samples are considered to classify a test sample. Each feature is evaluated by equal importance to find nearest neighbor. However, this situation is not practically valid, because each feature does not have same importance on dataset. For this reason, different information about features should be evaluated for classifiers based on distance metrics. It is very important to decide to discriminatory powers of features. While features which have high discrimination values can increase the success rate of classification, features with lower discrimination values can adversely affect the success. Datasets which contain a great number of features have redundant or irrelevant features. These features can cause to incorrect classification. K nearest neighbor considers equally all of features without exception. For this reason, firstly, it sort features depending on their discriminatory power. Then features are weighted depending on their t test values

according to our approach. While t test value uses normally to select feature selection, it benefit from t test to improve classification accuracy of k nearest neighbor. According to results, classification accuracy is increased from 74.14% to 86.57% using Euclidean distance and k=3. It is quite a few change in classification accuracy, if the exponent value or weighted value increases. The proposed method also offers dimensionality reduction thanks to t test. In this wise, time complexity can be reduced.

2.2 Title: Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm

Author: E. Parvinnia a,*, M. Sabeti a, M. Zolghadri Jahromi b, R. Boostani

Electroencephalogram (EEG) signals are often used to diagnose diseases such as seizure, alzheimer, and schizophrenia. One main problem with the recorded EEG samples is that they are not equally reliable due to the artifacts at the time of recording. EEG signal classification algorithms should have a mechanism to handle this issue. It seems that using adaptive classifiers can be useful for the biological signals such as EEG. In this paper, a general adaptive method named weighted distance nearest neighbor (WDNN) is applied for EEG signal classification to tackle this problem. This classification algorithm assigns a weight to each training sample to control its influence in classifying test samples. The weights of training samples are used to find the nearest neighbor of an input query pattern. To assess the performance of this scheme, EEG signals of thirteen schizophrenic patients and eighteen normal subjects are analyzed for the classification of these two groups. Several features including, fractal dimension, band power and autoregressive (AR) model are extracted from EEG signals. The classification results are evaluated using Leave one (subject) out cross validation for reliable estimation. The results indicate that combination of WDNN and selected features can significantly outperform the basic nearest-neighbor and the other methods proposed in the past for the classification of these two groups. Therefore, this method can be a complementary tool for specialists to distinguish schizophrenia disorder. The main problem in the classification of EEG signals is the quality of the recorded signal, which can be different during the experiment. These unwanted disturbances cannot be controlled since many activities are going on at the same time in the brain. Changes in the environment can distract the attention of the patient at the time of recording the EEG signal, which directly

affects the quality of the recorded signal. In this paper, WDNN is applied for EEG signal classification task. This classifier assigns a weight to each training sample that controls its influence in classifying test samples. When a large weight is assigned to a training sample, it will increase its influence in classifying many samples. In contrast, reducing the weight of a training sample will decrease its effect in classification task. To show the effectiveness of WDNN for biological signals, EEG signals of eighteen normal subjects and thirteen schizophrenic patients are analyzed with the objective of classifying these two groups. The EEG signals are recorded in the Centre for Clinical Research in Neuropsychiatry, Perth, Western Australia. Several features like Higuchi fractal dimension, band power and AR coefficients are extracted from EEG signals. Our results showed that this scheme could improve the generalization accuracy for EEG signal classification task. Therefore, this classifier can be a complementary tool for specialists to distinguish schizophrenia disorder. For our future work, we decide to use preprocessing methods such as wavelet or principal component analysis instead of using the raw signals. Also, we decide to modify WDNN to assign weights to the features as WDNN is changed to feature weighing algorithm.

2.3 Title: A Study on Classification Algorithms and Performance Analysis of Data Mining using Cancer Data to Predict Lung Cancer Disease
Author: E. SathiyaPriya, Dr. S. Mary venila

In the field of Healthcare, cancer diagnosis is the challenging problems and also many of the research has focused to improve the performance to get a satisfactory results in the particular area. To diagnose a Lung cancer is a difficult tasks in medical research. To overcome this challenging task, the many researchers use data mining techniques were applied to predict the many type of diseases. In this research we studied and make comparison of different classification to classify and predict the lung cancer disease. In this research work using Naïve Bayes classification algorithm, SVM (Support Vector Machine) algorithm, KNN algorithm, and J48 algorithm. By using a classification algorithm it produces a different result for the lung cancer datasets in this research. The quality of the result has measured depends on correct and incorrect instances that are correctly classified by a classification techniques. In some cases even in the advanced level Lung cancer patients does not show the symptoms associated with the Lung cancer. There are many patients did not know Lung cancer disease in early stage, because the lack of awareness. So, the prevention of lung cancer is needed in reducing life losses. By this experiment results we can clearly predict the lung cancer disease, which can be used to warn the people before to avoid the unwanted drinking habits, smoking, intake of contaminated food, obesity etc., for prevent from the lung cancer disease. In this study of classification Techniques used in several dataset is help to improve lack of awareness in the Lung Cancer patients. The experiment results has show three different classification algorithms gives a better performance on selected one Lung cancer dataset. When we applying the same classification techniques for another Lung cancer dataset, the result has shows all algorithms performs significantly better performance. From this study, based on the results that no single classifier is better than other. But

the results from this data mining tool, we cannot consider the results has surely as perfect. There is chance to get a results might changes and vary in a similar datasets related to cancer disease are classified on different tools like tanegra, rapid mining etc., because that are the latest data mining tools within the data mining.

2.4 Title: Study of Classification Algorithm for Lung Cancer PredictionAuthor: Dr.T.Christopher*P*, J.Jamera banu

Lung cancer remains the leading cause of cancer related mortality for both men and women and its incidence is increasing worldwide. Lung cancer is the uncontrolled growth of abnormal cells that start off in one or both Lung. The earlier detection of cancer is not easier process but if it is detected, it is curable. We analyzed the lung cancer prediction using classification algorithm such as I Bayes, Bayesian network and J48 algorithm. Initially 100 cancer and noncancerous patients' data were collected, preprocessed and analyzed using a classification algorithm for predicting lung cancer. The dataset have 100 instances and 25 attributes. The main aim of this paper is to provide the earlier warning to the users and the performance analysis of the classification algorithms. Data mining in health care management is not analogous to the other fields due to the reason that the data existing here are heterogeneous in nature and that a set of ethical, legal, and social limitations apply to private medical information. The experiment has been performed using WEKA tool with several data mining classification techniques and it is found that the I Bayes algorithm gives a better performance over the other classification algorithm such as Bayesian and J48.Lung cancer prediction system can be further enhanced and
expanded. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules.

2.5 Title: Predicting Lung Cancer Survivability using SVM and Logistic Regression Algorithms

Author: Animesh Hazra, Nanigopal Bera, Avijit Mandal

One of the most common and leading cause of cancer death in human beings is lung cancer. The advanced observation of cancer takes the main role to inflate a patient's probability for survival of the disease. This paper inspects the accomplishment of support vector machine (SVM) and logistic regression (LR) algorithms in predicting the survival rate of lung cancer patients and compares the effectiveness of these two algorithms through accuracy, precision, recall, F1 score and confusion matrix. These techniques have been applied to detect the survival possibilities of lung cancer victims and help the physicians to take decisions on the forecast of the disease. One of the major and frequent bases of cancer deaths globally in terms of both instance and transience is lung cancer. The main reason behind the increasing of deaths from it is detecting the disease lately and faults in effective treatment. So, the early detection is needed to save lives from this disease. The survivability rate of lung cancer can be predicted with the help of modern machine learning techniques. Accordingly, it would be clever to determine the survival possibilities among the patients. In this study data cleaning, feature selection, splitting and classification techniques have been applied for predicting survivability of lung cancer as accurately as possible. This project reveals that logistic regression classifier gives the topmost accuracy of 77.40% compared to support vector machine classifier which gives 76.20% accuracy. Also, the logistic regression classifier gives maximum classification accuracy concerning every different classifier. This work can further be enhanced by modifying logistic regression classifier which gives highest accuracy. With the help of machine learning methods it is really difficult to diagnose the different medical conditions of a lung cancer patient and prediction of conditions are also more critical in nature. It is a challenging task in machine learning and data

mining fields to construct a specific and computationally efficient classifier for medical applications. This can be a great future scope of this research. For big datasets how these classification algorithms behave, that is another future scope of this project. Moreover the identification of particular stage of lung cancer can be done in near future. Another prospect of this research is the time and space complexity analysis of different classification algorithms on medical datasets which can be explored in the forthcoming work.

CHAPTER 3

3. SYSTEM STUDY

3.1 EXISTING SYSTEM

In existing system, several algorithms are available for predicting frequent patterns. Its purpose is to find the association among the item sets in the biological datasets. It is necessity to predict the dominant amino acids for causing the lung cancer and to take immediate preventive actions. The introduction of the concept of reversible watermarking in the Barton patent, several methods have been proposed. Among these solutions, most recent schemes use Expansion Embedding (EE) modulation, Histogram Shifting (HS) modulation or, more recently, their combination. One of the main concerns with these modulations is to avoid underflows and overflows. Indeed, with the addition of a watermark signal to the image, caution must be taken to avoid gray level value underflows (negative) and overflows (greater than for a bit depth image) in the watermarked image while minimizing at the same time image distortion.

3.1.1 Disadvantages

- The modulation is overflow and underflow.
- The image distortions are more.
- Image is not protected in correct way.
- Allows discontinuity in data protection.
- Time complexity
- Space complexity
- Cost Expenses
- Accuracy Level

3.2 PROPOSED SYSTEM

The proposed system will classify the lung nodule into benign (non-cancerous) or malignant (cancer). The classification will produce a function that is sorting out a data item into one of several predefined classes. 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. To detect and classify early nodules employed mining approach to extract these nodules from the datasets. In these regions, directly applying HS on pixels may be more efficient and of smaller complexity than applying it on prediction-errors. Because, the histogram maxima corresponds to the null gray value; capacity is maximized and underflows simply avoided by shifting pixel value to the right, i.e. by adding a positive gray value. This should allow us to optimize the compromise capacity/image distortion.

3.2.1 Advantages

- Gives More accuracy
- Time complexity is less
- Prediction through nodules gives more consistency.
- Cost is lesser than other

3.3 SYSTEM REQUIREMENTS

3.3.1 HARDWARE REQUIREMENTS

PROCESSOR	:	Pentium IV 2.2 GHZ
HARD DISK	:	SATA 150 GB
RAM	:	1 GB
MONITOR	:	15VGA Color
MOUSE	:	Ball/Optical
CD – DRIVE	:	WRITE 53X

3.3.2 SOFTWARE REQUIREMENTS

OPERATING SYSTEM	:	Windows 7
FRONT END	:	VB.NET
BACK END	:	MICROSOFTSQL SERVER

3.4. SOFTWARE DESCRIPTION

3.4.1 .NET TECHNOLOGY

Microsoft .NET is a set of Microsoft software technologies for rapidly building and integrating XML Web services, Microsoft Windows-based applications, and Web solutions. The .NET Framework is a language-neutral platform for writing programs that can easily and securely interoperate. There's no language barrier with .NET: there are numerous languages available to the developer including Managed C++, C#, Visual Basic and Java Script. The .NET framework provides the foundation for components to interact seamlessly, whether locally or remotely on different platforms. It standardizes common data types and communications protocols so that components created in different languages can easily interoperate. ".NET" is also the collective name given to various software components built upon the .NET platform. These will be both products (Visual Studio.NET and Windows.NET Server, for instance) and services (like Passport, .NET My Services, and so on).

C# could theoretically be compiled to machine code, but in real life, it's always used in combination with the .NET framework. Therefore, applications written in C#, requires the .NET framework to be installed on the computer running the application. While the .NET framework makes it possible to use a wide range of languages, C# is sometimes referred to as THE .NET language, perhaps because it was designed together with the framework. C# is an Object Oriented language and does not offer global variables or functions. Everything is wrapped in classes, even simple types like int and string, which inherits from the System. Object class.

3.4.2. Visual C# Express

C# can be written with any text editor, like Windows Notepad, and then compiled with the C# Command line compiler, csc.exe, which comes with the .NET framework. However, most people prefer to use an IDE (Integrated Development Environment), and Microsoft offers several options for this. Their flagship is Visual Studio, which can be used to work on every possible aspect of the .NET framework. This product is very advanced, and comes in several editions. Visual Studio is not exactly cheap, and might even be too advanced for hobby programmers.

With .NET framework 2.0, Microsoft introduced the so-called Express versions, targeted at hobby programmers and people wanting to try .NET, and they continued this tradition with the later release of .NET 3.0 and 3.5. The Express versions only work for one language, like C# or VB.NET, and miss some of the really advanced features of Visual Studio. However, they are free and will work just fine for learning the languages, which is why we will use it for this tutorial. Hello, world!

If you have ever learned a programming language, you know that they all start with the “Hello, world!” example, and who are we to break such a fine tradition? Start Visual C# Express (introduced in the last chapter), and select File -> New project… From the project dialog, select the Console application. This is the most basic application type on a Windows system, but don’t worry, we won’t stay here for long. Once you click Ok, Visual C# Express creates a new project for you, including a file called Program.cs.

3.5. SQL SERVER

3.5.1. Definition:

Microsoft SQL Server is a powerful relational database management system catering to high-end users with advanced needs. Along with Oracle, Microsoft SQL Server is widely regarded as one of the two main full-featured database systems on the market today.

3.5.2 Microsoft SQL Server:

Microsoft SQL Server offers tight integration with the Back Office series of server products. Check out this collection of SQL Server links about migrating to SQL Server, performance tuning, development and more!

3.5.3 Structured Query Language (SQL)

The Structured Query Language (SQL) forms the backbone of most modern database systems. These links provide the best resources on the Net for neophytes and expert database administrators alike!

SQL Server Management

Applications Manager SQL Server Monitoring software capability helps DBAs monitor the performance and availability of production databases. It is an agentless monitoring solution that provides out-of-the-box performance metrics for ensuring the SQL Server runs efficiently.

SQL server monitoring has the ability to connect to the database source & monitor various system table column values, collect data, etc. and also notify through alerts, if the database system properties are beyond a given threshold.

Some of the components that are monitored in MS SQL database are:

- Memory Usage
- Buffer Manager Statistics

- Connection Statistics
- Cache Details
- Lock Details
- SQL Statistics
- Latch Details
- Access Method Details
- Database Details
- Scheduled Jobs

SQL Server Monitoring Capabilities

- Out-of-the-box monitoring of MS SQL availability and performance.
- Monitors performance statistics such as memory usage, buffer manager statistics, connection statistics, cache details, sql statistics, etc., Alerts can be configured for these parameters.
- Based on the thresholds configured, notifications and alerts are generated if the SQL Server or any specified database within the server is not accessible. Actions are executed automatically based on configurations.
- Performance graphs and reports are available instantly. Reports can be grouped and displayed based on availability, health, and connection time.

CHAPTER 4

4. SYSTEM DESIGN

4.1 DATA FLOW DIAGRAM

LEVEL 0

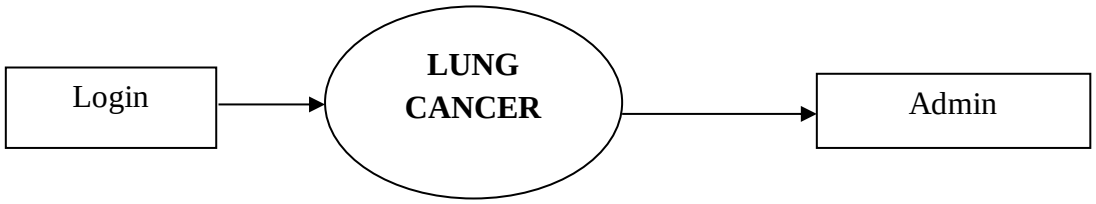

Fig. 4.1.1. Level 0 Flow chart

LEVEL 1

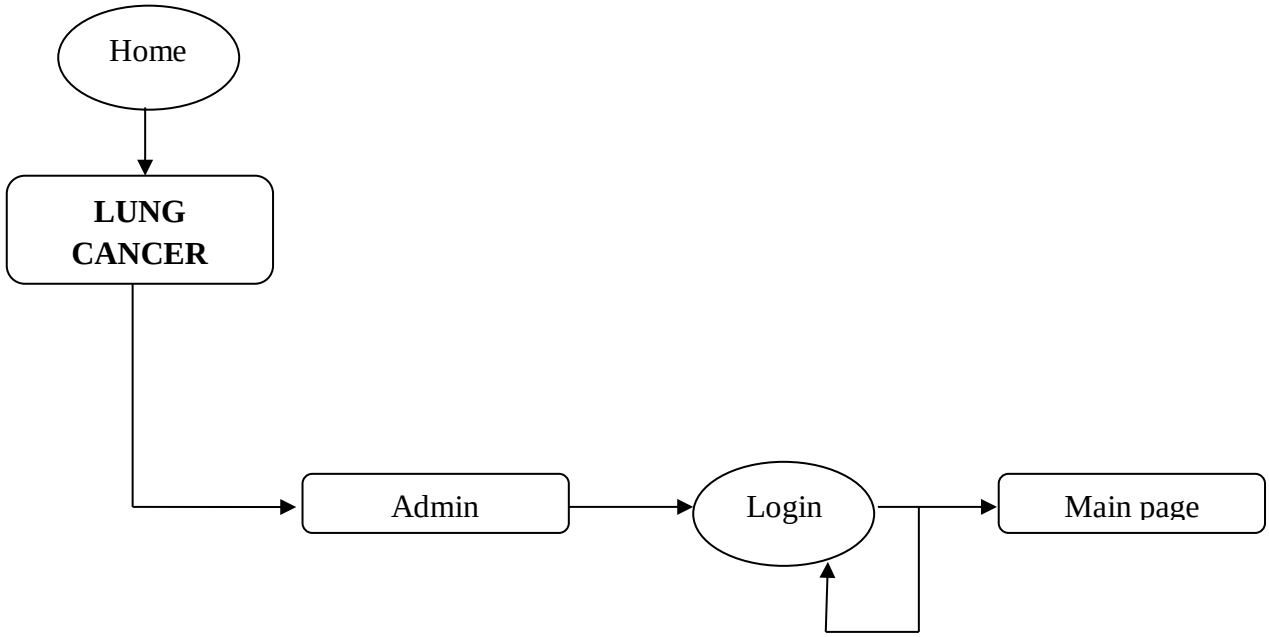

Fig. 4.1.2. Level 1 Flow chart

LEVEL 2

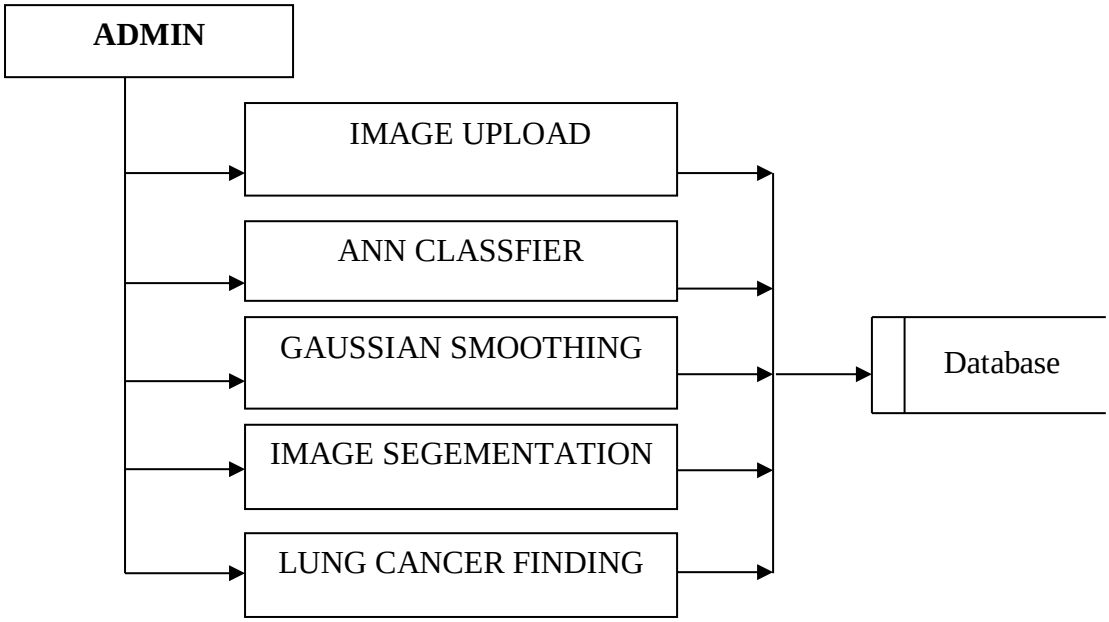

Fig. 4.1.3. Level 2 Flow chart

4.2 UML DIAGRAM

USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.

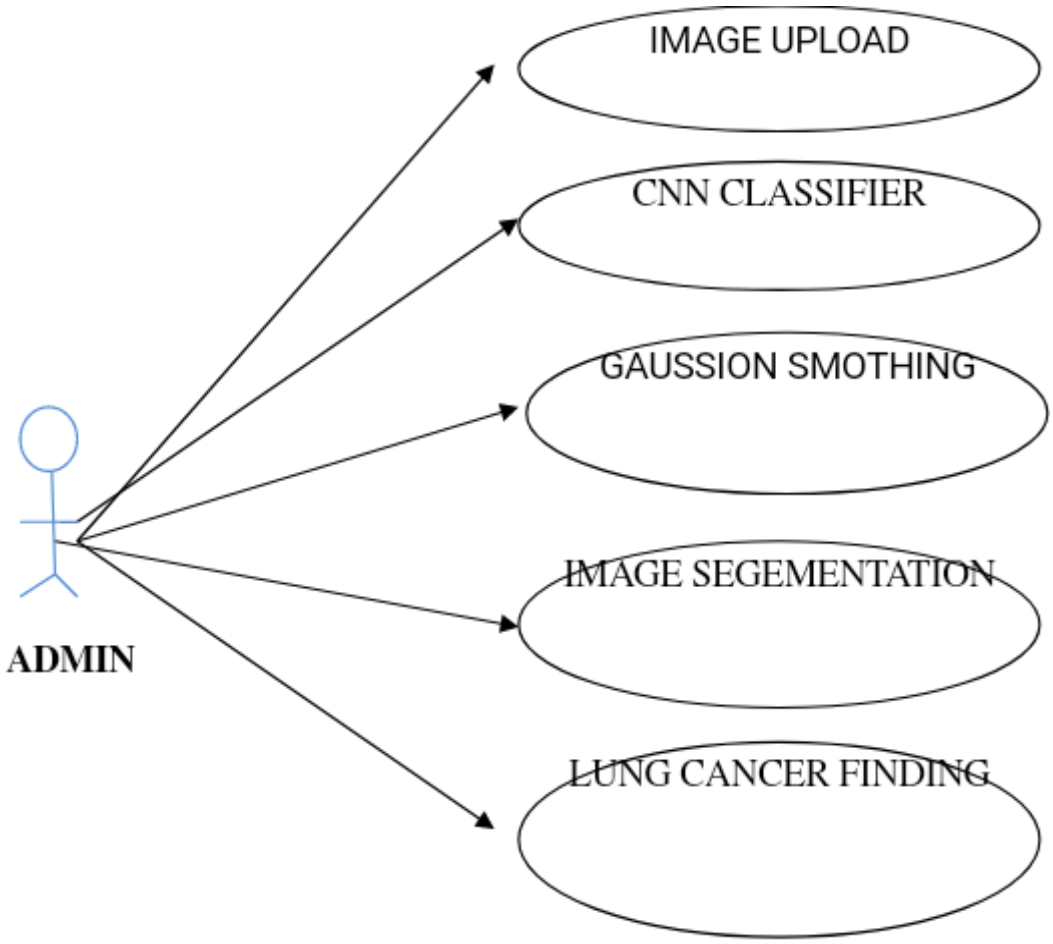

Fig4.2.Use Case Diagram

CHAPTER 5

5.SYSTEM ARCHITECTURE

5.1 OVERALL SYSTEM ARCHITECTURE

A system architecture or systems architecture is the conceptual model that defines the structure, 31ehaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning abou4.2t the structures and behaviors of the system.

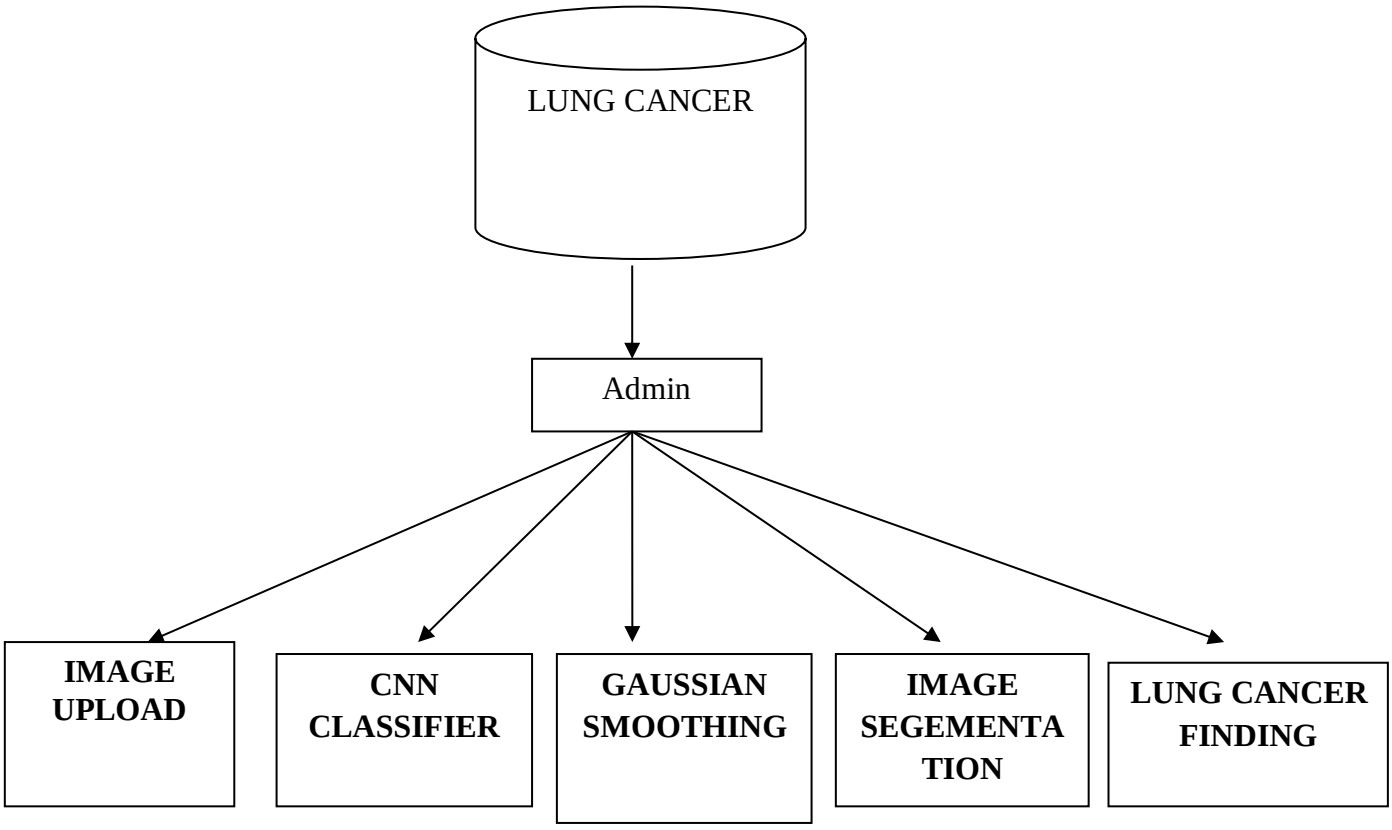

Fig 5.1 System Architecture

CHAPTER 6

6. SYSTEM IMPLEMENTATION

6.1 MODULES

- Database Collection
- Image Upload
- Ann Classifiers
- Gaussian Smoothing Technique
- Lung Cancer Finding

6.2 Database collection

In order to perform classification of lung cancer, images from various stages of lung cancer is required. CT images are used due to the variety of advantages they provide as mentioned earlier. The images are strictly in DICOM (Digital Imaging and Communication in Medicine) format in order to maintain medical standardization. These images are obtained from NIH/NCI Lung Image Data Consortium and Early Lung Cancer Action Program (ELCAP).

6.3 Image Upload

In this Upload image module, MRI images were been chosen to upload their image. After Uploading Cancer Detection process will be done to find Caner in scanning parts of the image by using an Ann algorithm. The system is capable of analyzing scan report by means of a laser excitation system and a scanning machine. The laser creates gradients inside the part under inspection, and the Scanning Machine observes accuracy and diffuses inside the part.

6.4 ANN Algorithm

Artificial neural networks (ANN) or connectionist systems are computing systems vaguely .Tasks suited for supervised learning are pattern recognition (also known as classification) and regression (also known as function approximation).ANN is a complex adaptive system which can change its internal structure based on the information pass through it.

- Supervised learning: This strategy involves a trainer which is smarter than the network.
- Unsupervised learning: This strategy is used when there is not example data set with known answer.
- Reinforcement learning: This strategy makes decision based on feedback from environment

6.5 Feed Forward Algorithm

The feed forward neural network was the first and simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. In this project multi-layer neural network has been employed to achieve image compression. The proposed technique breaks down large images into smaller windows and applies FFA to these windows as a pre-process technique. The input pixels will be used as target values so that assigned mean square error can be obtained, and then the hidden layer output will be the compressed image.

6.6 Lung Cancer Finding

Lung cancer is the most widespread and dangerous cancer in the world according to stage of discovery of the cancer cells in the lungs, so the early finding of the disease plays an essential role to avoid serious advanced stages, which helps to reduce its degree of distribution. Lung Cancer Finding System (LCFS) uses convolution filters with Gaussian pulse to smooth the cell images. It is finding the how much of lungs affected by the lung cancer. The color and contrast of the images are enhanced. Then the nuclei in the images are segmented by a method called thresholding. All of these are simple digital image processing techniques. Following that, LCFS employs morphologic and colorimetric techniques to extract the features from the images of the nuclei. The extracted morphologic features include the perimeter of the region, roundness of the area, and rectangleness of the nucleus. The extracted colorimetric features include the color components: red, green, blue; illumination, saturation, variation in the red and blue components, and the fraction of blue component in the nucleus.

7.CHAPTER

7.1 CONCLUSION

Lung cancer is one of the major causes of death in men and women. In some cases even in the advanced level Lung cancer patients does not show the symptoms associated with the Lung cancer. There are many patients did not know Lung cancer disease in early stage, because the lack of awareness. So, the prevention of lung cancer is needed in reducing life losses. By this experiment results this can clearly predict the lung cancer disease, which can be used to warn the people before to avoid the unwanted drinking habits, smoking, intake of contaminated food, obesity etc., for prevent from the lung cancer disease. In this study classification Techniques used in several images is a help to improve lack of awareness in the Lung Cancer patients. The experiments results have shown **ANN** and **Feed Forward** algorithms give a better performance on selected one Lung cancer dataset through **Image Processing**. When we apply the same classification techniques for another Lung cancer dataset, the result has been significantly better performance. This experiment can be extended by applying additional range of classification algorithms or use a proposed algorithms on additional range of datasets from huge medical database and dataset has from various domains in future.

7.2 FUTURE ENHANCEMENTS

In Future work it will describe the performed accurate image and will focus on the contribution to solve coordination problems when executing this method is fragile as any modifications will impact the watermark. Even though some solutions have already been proposed questions about watermark robustness are largely open. This is one of the upcoming challenges.