



Classification of Lung Cancer Stages using Quantitative Image and Genomic Biomarker

Akshay Bhor¹, Aditya Likhar², Azhar Maner³, Aditya Patil⁴, Dipti Chaudhari⁵
 BE Student^{1,2,3,4}, Assistant Professor⁵
 Department of Computer Engineering
 Dr. DY Patil Institute of Technology, Maharashtra, India

Abstract:

Lung cancer could be a malady that happens attributable to the uncontrolled cell growth in tissues of the respiratory organ. It's terribly difficult to sight it in its early stages as its symptoms seem solely within the advanced stages. The aim is to automate the classification method for the first detection of carcinoma. It includes classification algorithmic rule i.e. Neural Network and for improvement GA (Genetic Algorithm) is employed. Analysis would be done on the idea of properly classified sample information. By mistreatment X-raying (CT) pictures, a laptop motor-assisted detection theme accustomed phase respiratory organ tumors and computed tumor connected image options. All CT pictures were viewed at a laptop digital computer by one in every of four inquiring radiologists. Pictures were viewed at commonplace respiratory organ, soft tissue, and bone window settings. The steps for detection of carcinoma start with method of accretive CT pictures. This CT pictures area unit additional processed; mistreatment coaching and testing strategies options area unit classified mistreatment artificial neural network. This classification helps in evaluating the results of the input CT image.

Keywords: Computer aided diagnosis, Image features, Quantitative image feature analysis, GA (Genetic Algorithm), computed tomography (CT).

I. INTRODUCTION

Lung Cancer could be a noteworthy reason for Mortality within the western world as exhibited by the hanging factual numbers distributed systematically by the Yankee carcinoma Society. They demonstrate that the 5-year survival rate for patients with respiratory organ malignancy will be increased from a traditional of fourteen % up to forty nine % if the complaint is analyzed and treated at its initial stage. Healthful footage as an important piece of therapeutic determination and treatment were specializing in this footage permanently.

These footage incorporate success of hid information that victimized by doctors in deciding on contemplated selections around a patient. Then again, removing this vital shrouded information could be a basic 1st stride to their utilization. This reason evokes to utilize info dig systems skills for productive learning extraction and realize hid respiratory organ. Mining Medical footage includes various procedures. healthful data processing could be a promising zone of procedure insight connected to a consequently break down patients records going for the speech act of latest info valuable for restorative alternative creating.

Affected info is predicted not simply to increment actual determination and effective infection treatment, in addition to boost security by decreasing blunders. The systems during this paper prepare the advanced X-beam midriff movies in 2 classes: standard and strange. the normal ones are those depiction a solid patient. The irregular ones incorporate style of respiratory organ

tumor; we are going to utilize a typical arrangement technique specifically SVMs and neural systems.

II. LITERATURE SURVEY

1) Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer

AUTHORS: ZakariaSulimanZubi and RemaAsheibaniSaad,
 Lung cancer is a disease of uncontrolled cell growth in tissues of the lung, 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, a measure for early stage lung cancer diagnosis mainly includes those utilizing X-ray chest films, CT, MRI, etc. Medical images mining is a promising area of computational intelligence applied to automatically analyzing patient's records aiming at the discovery of new knowledge potentially useful for medical decision making. Firstly we will use some processes are essential to the task of medical image mining, Data Preprocessing, Feature Extraction and Rule Generation. The methods used in this paper work states, to classify the digital X-ray chest films into two categories: normal and abnormal. The normal state is the one that characterize a healthy patient. The abnormal state including the types of lung cancer; will be used as a common classification method indicating a machine learning method known as neural networks. In addition, we will investigate the use of association rules in the problem of x-ray chest films categorization. The digital x-ray chest films are storied in huge multimedia databases for a medical purpose. This multimedia database provides a great environment to apply some *image recognition methods* to

Extract the useful knowledge and then rules from the mentioned database. These rules that we could get using image recognition methods, will help the doctors to decide important decisions on a particular patient state.

2) A Fully Automated Method for Lung Nodule Detection From Postero-Anterior Chest Radiographs

AUTHORS: PaolaCampadelli, Elena Casiraghi, and Diana Artioli,

In the past decades, a great deal of research work has been devoted to the development of systems that could improve radiologists' accuracy in detecting lung nodules. Despite the great efforts, the problem is still open. In this paper, we present a fully automated system processing digital poster-anterior (PA) chest radiographs that starts by producing an accurate segmentation of the lung field area. The segmented lung area includes even those parts of the lungs hidden behind the heart, the spine, and the diaphragm, which are usually excluded from the methods presented in the literature. This decision is motivated by the fact that lung nodules may be found also in these areas. The segmented area is processed with a simple multi scale method that enhances the visibility of the nodules, and an extraction scheme is then applied to select potential nodules. To reduce the high number of false positives extracted, cost-sensitive support vector machines (SVMs) are trained to recognize the true nodules. Different learning experiments were performed on two different data sets, created by means of feature selection, and employing Gaussian and polynomial SVMs trained with different parameters; the results are reported and compared.

3) An Approach for Discretization and Feature Selection of Continuous-Valued Attributes in Medical Images for Classification Learning

AUTHORS: Diana Artioli, RemaAsheibaniSaad,

Many supervised machine learning algorithms require a discrete feature space. In this paper, we review previous work on continuous feature discretization and, identify defining characteristics of the method. We then propose a new supervised approach which combines discretization and feature selection to select the most relevant features which can be used for classification purpose. The classification technique to be used is Associative Classifiers. The features used are Harlick Texture features extracted from MRI Images. The results show that the proposed method is efficient and well-suited to perform preprocessing of continuous valued attributes.

4) Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques

AUTHORS: V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra. 2013

Cancer is the most important cause of death for both men and women. The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. A disease that is commonly misdiagnosed is lung cancer. Earlier diagnosis of Lung Cancer saves enormous lives, failing which may lead to other severe problems causing sudden fatal end. Its cure rate and prediction depends mainly on the early detection and diagnosis of the disease. One of the most common forms of medical malpractices globally is an error in

diagnosis. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, we briefly examine the potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bays and Artificial Neural Network to massive volume of healthcare data. The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not “mined” to discover hidden information.

III. PROPOSED SYSTEM

Lung cancer is that the leading reason behind tumor-related deaths among the globe. At identical time, it looks that the speed has been steady increasing. Malignant euplastic disease is caused by the uncontrolled growth of tissues among the viscous. the American cancer society estimates that 213, 380 new cases of malignant euplastic disease among the U.S square measure diagnosed and 100 and sixty, 390 deaths as a results of malignant euplastic disease will occur in 2007. Tobacco smoking is that the main behind all cases. Lung cancer is that the expansion of a growth, referred to as a nodule that arises from cells lining the airways of the system. The detection of malignant euplastic disease has been a tedious task in medical image analysis over the past few decades. among the health trade, chest X-rays square measure thought-about to be the foremost wide used technique for the detection of malignant euplastic disease. The image processing consists of mainly four steps.

These steps are

1. Image Enhancement
2. Segmentation
3. Feature Extraction
4. Result.

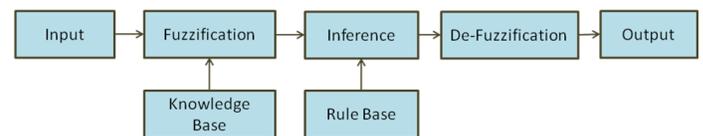


Figure.1. System Design

Mathematical model

Mathematical model of the proposed system

INPUT:-

Let S is the Whole System Consists:

Let S is the Whole System Consist of

$S = \{I, P, O\}$

Where,

I = input.

$I = \{U, Q\}$

U = User

$U = \{u_1, u_2, \dots, u_n\}$

Q = Query

$Q = \{q_1, q_2, \dots, q_n\}$

P = Process

$P = \{MBWA, WA, GLCOMA, SVM\}$

MBWA =Marker-Based Watershed Algorithm.

WA = Watershed Algorithm.

GLCOMA = Grey level co-occurrence matrix algorithm.
SVM = Support Vector Machine
OUTPUT: The predicted result will be the output of the system

IV. CONCLUSION

In this paper, completely different phases of image process were applied on respiratory organ Nodules. From these completely different image process techniques, the fuzzy filter can give the economical demising. Segmentation done by marker based mostly watershed algorithmic rule, provides varied region of image. GLCM is employed to extract the various options of image and that takes less time for generating the result. This results ar well-versed SVM Classifier, that classifies the nodules as benign or malignant. SVM classifier provides ninety two.5% accuracy.

V. ACKNOWLEDGMENT

We might want to thank the analysts and also distributors for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

VI. REFERENCES

- [1].akariaSulimanZubi and RemaAsheibaniSaad, Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer Recent Researches in Arti cial Intelligence, Knowledge Engineering and Data Bases, Libya, 2007.
- [2].aola Campadelli, Elena Casiraghi, and Diana Artioli, A Fully Automated Method for Lung Nodule Detection From Postero-Anterior Chest Radiographs,In Proc. of IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 25, NO. 12, DECEMBER 2006.
- [3].abaSheela L and Dr.V.Shanthi, An Approach for Discretization and Feature Selection Of Continuous-Valued Attributes in Medical Images for Classi cation Learning, International Journal of Computer Theory and Engineering, Vol. 1, No.2,June2009.
- [4].Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra. 2013, Diagnosis of Lung Cancer Prediction System Using Data Mining Classi
- [5]. Swensen, et al., CT screening for lung cancer: ve-year prospective experience, Radiol-ogy, vol. 235, no. 235, pp. 259-265, APR. 2005.
- [6]. Bach, et al., Computed tomography screening and lung cancer outcomes, J. Amer. Med. Assoc., vol. 297, no. 9, pp. 953-961, Mar. 2007.