

# Predictive Diagnostic System of Infectious Lung Disease Using Breath Detection Motion

Mrs. Sanchita Shrivastava<sup>1</sup>, Rupesh Mahajan<sup>2</sup>, Minal Shahakar<sup>3</sup>

<sup>1,2,3</sup>Assistant Professor, Dr.D.Y .patil institute of technology, Pune, India, 411018,

**Abstract** – The Medical data mining is one of the major issues in this modern world. Medical problems are often there in each and every human being. Machine learning is finding ever growing interest in medical science since last decade owed to cheaper computing power and inexpensive memory making it efficient to store, process and analyse growing volumes of data. We are trying to implement an automatic hospital system for lung disease for detecting them enhanced algorithms are being designed and applied on large datasets to help discover hidden insights and correlations amongst the data elements not obvious to human. The objective is detect motion based respiration state to set timing of capturing image and after that we introduce fundamentals of machine learning methods and review system successes in detection of anatomical and cellular structures, tissue segmentation, computer-aided lung disease diagnosis and prognosis. The growing popularity of machine learning in medical diagnosis lies in the fact that learning algorithms are domain agnostic. Recent advances in machine learning are helping to identify, classify, and quantify patterns in medical images. At the core of these advances is the ability to exploit the hierarchical feature representations learned solely from data, instead of features designed by hand according to domain-specific knowledge. The main aims of the study are performing a systematic review about methods and the machine learning algorithms that are used to classify the respiratory sounds for diagnosis of lung disorders and evaluating the development of possible methods in the future.

**Keywords-** *breath detection, lung diseases, machine learning*

## I- INTRODUCTION

Lung diseases are one of the most dangerous diseases a human can ever have. It is very difficult to detect it in its early stages as its symptoms appear only in advanced stages. Also different breath-hold positions can cause a misalignment of the image between different slices, with potential detrimental effects on a variety of clinically relevant measurements (e.g. volume or shape of the left ventricle). Over the past few decades digitization of medical data and advent of medical imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), mammography, ultrasound and X-ray, have been used for the early detection, diagnosis, and treatment of diseases which caused large volumes of patient data being generated and are becoming widely available. A common outcome of this information revolution is that we are faced with daunting task of interpreting and understanding the data. It is tedious and tiresome job to manually analyse and interpret such a large corpus of data unfailingly without any error. There is a dire need for the data-driven approaches from computational sciences, often referred as data science or data analytics, to help with understanding the data. These approaches can be used to analyse medical data to extract critical health information that can help patients and doctors make better decisions. Machine learning provides methods for analysing the data and extracting key characteristics of relationships in the data by creating a computational model that best describe the data. It is now possible to quickly and automatically produce the computational models that can analyse bigger, more complex data and deliver Faster, more accurate insights or analysis even on the very large scale. Often, the hidden insights, relationships revealed by the models are not obvious to human eye or conceivable by an analyst.

## II-LITERATURE REVIEW

Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer  
 AUTHORS: Zakaria Suliman Zubi and Rema Asheibani Saad  
 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 stages the key of its cure. In general, a measure for early stage lung cancer diagnosis mainly includes those utilizing X-ray chest, 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 Pre-processing, Feature Extraction and Rule Generation. The methods used in this paper work states, to classify the digital X-ray chest \_lms 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 \_lms categorization. The digital x-ray chest \_lms are stored 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: Paola Campadelli, 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 posterior-anterior (PA) chest radiographs that starts by producing an accurate segmentation of the lung 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: Jaba Sheela L, Dr.V.Shanthi

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 Classification. 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 pre-processing 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.

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 brief examine the potential use of classification based data mining techniques such as Rule based, Decision tree, Navie 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.

#### 5) Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment Of Early Stage Lung Cancer Patients

AUTHERS: Nastaran Emaminejad, Wei Qian, Yubao Guan, Maxine Tan, Yuchen Qiu, Hong Liu, and Bin Zheng

They develop a new quantitative image feature analysis scheme and investigate its role along with 2 genomic biomarkers namely, protein expression of the excision repair cross-complementing 1 (ERCC1) genes and a regulatory subunit of rib nucleotide reductive (RRM1), in predicting cancer recurrence risk of Stage I non-small-cell lung cancer (NSCLC) patients after surgery. Methods: By using chest computed tomography Images, we developed a computer-aided detection scheme to segment lung tumours and computed tumour-related image features.

#### 6) Automatic lung segmentation based on Graph Cut using a distance-constrained energy

AUTHERS: Oluwakorede M. Oluyide, Jules-Raymond Tapamo, Serestina Viriri.

Lung segmentation serves to ensure that all the parts of the lungs are considered during pulmonary image analysis by isolating the lung from the surrounding anatomy in the image. Research has shown that computed tomography (CT) images greatly improve the accuracy of the diagnosis obtained by a physician for lung cancer detection. Therefore, inspired by the success of Graph Cut in image segmentation and given those manual methods of analysing CT images are tedious and time-consuming, an automatic segmentation method based on Graph Cut is proposed which makes use of a distance-constrained energy (DCE). Graph Cut produces globally optimal solutions by modelling the image data and spatial relationship among the pixels.

#### 7) Robust Real-Time 3D Respiratory Motion Detection Using Time-of-Flight Cameras

Authors: Jochen Penne, Christian Schaller, Joachim Hornegger, Torsten Kuwert

In this paper, they present a system that uses the fairly new and of-the-shelf Time of-Flight (ToF) technology to compute a dense Estimate of the three-dimensional respiratory motion of a Patient. The work is characterized by three key contributions. The rest is the employment of ToF sensors. Using ToF Sensors it is feasible to acquire a dense 3D surface model Of the chest and abdomen of the patient with more than 15 Frames per second. The second contribution is an algorithm to derive a surface

representation which enables the estimation? Of the 3D respiratory motion of the patient, which is sufficient? To compute 1D breathing signals for scalable specific Regions of interest like chest and abdomen. The proposed Data-driven algorithm models the chest and abdomen three dimensionally by distinct planes to different regions Of the torso of the patient. The third contribution is the possibility to derive a sub-millimetre accurate 1D respiratory Motion signal by observing the displacement of each plane.

#### 8) Breathing Analysis Using Thermal and Depth Imaging Camera Video Records

Authors: Ales Prockjka, Hanacharvota etal

The paper is devoted to the study of facial region temperature changes using a simple thermal imaging camera and to the comparison of their time evolution with the pectoral area motion recorded by the MS Kindest depth sensor. The goal of this research is to propose the use of video records as alternative diagnostics of breathing disorders allowing their analysis in the home environment as well.

### III-METHODOLOGY

Segmentation using the watershed transforms works well if you can identify, or "mark," the foreground objects and the background locations. Marker-controlled watershed segmentation follows this basic procedure:

1. Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment out.
2. Compute foreground markers. These are connected blobs of the pixels within each of the objects.
3. Compute background markers. These are the pixels that are not part of any object.
4. Modify the segmentation function so that it only has minima at the foreground and the background marker locations.
5. Compute the watershed transform of a modified segmentation function.

#### Gray-Level Co-Occurrence Matrix (GLCM)

The texture filter functions provide a statistical view of the texture based on the image histogram. These functions can provide the useful information about the texture of an image but cannot provide the information about the shape, i.e., the spatial relationships of pixels in an image.

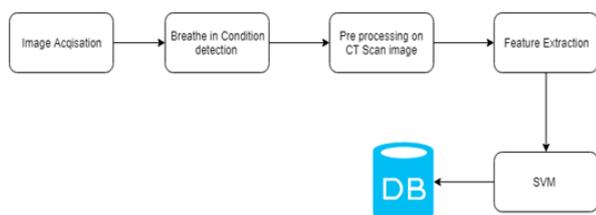
Another statistical method that considers a spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The toolbox provides functions to create a GLCM and derive the statistical measurements from it. This includes the following steps.

- Creating a Gray-Level Co-Occurrence Matrix
- Specifying the Offsets
- Deriving Statistics from the GLCM

#### IV-DESIGN

Many existing medical image processing methods rely on morphological feature representations to identify local anatomical characteristics. However, such feature representations were designed mostly by human experts, and the image features are often problem specific and not guaranteed to work for other image types. The system responds out breath motion and according to that the system will capture image on breathe in position is detected. So the resource will be utilized. After image acquisition the system perform pre-processing on image. Find out affected regions and their characteristics in form of data. This data is classified using SVM. SVM classify it as normal or diseases lung and although state-of-the-art methods use supervised learning to and the most relevant

And essential features for target tasks, they require a significant number of manually labelled training data, and the learned features may be superficial and may misrepresent the complexity of the anatomical structures. More critically, the learning procedure is often concerned to a particular template domain, with a certain number of predesigned features. Therefore, once the template or image features change, the entire training process has to start over again.



#### Step1:

The system first finds out breath motion and according to that the system will capture image on breath in position is detected. So the resource will be utilized.

#### Step 2:

After image acquisition the system perform pre-processing on image

#### Step 3:

Find out affected regions and their characteristics in form of data

#### Step 4:

This data is classified using SVM.

#### Step 5:

SVM classify it as normal or diseases lung and identify lung diseases.

#### V-CONCLUSION

For implementing and designing of Automatic Lung Disease Detection we detect breathe in conditions for capturing image so our Resource may not waste. For finding out breathe in - breathe out conditions we used different phases of image processing. From these different image processing techniques, we will provide the efficient breath in - breathe out condition. And after that we classify the lung diseases according to image classification. And it is important to understand the principles of staging, the clinical relevance of various radiological staging descriptors, and their impact on treatment decisions.

#### REFERENCES

- [1] "Zakaria Suliman Zubi and Rema Asheibani Saad" *Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer*
- [2] "Paola Campadelli, Elena Casiraghi, and Diana Artioli" *A Fully Automated Method for Lung Nodule Detection from Postero-Anterior Chest Radiographs*
- [3] "Jaba Sheela L, Dr.V.Shanthi. An Approach for Discretization and Feature Selection of Continuous-Valued Attributes in Medical Images for Classification Learning
- [4] "V.Krishnaiah,Dr.G.Narsimha,Dr.N.Subhash", *Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques Chandra*
- [5] "Nastaran Emaminejad, Wei Qian, Yubao Guan, Maxine Tan, Yuchen Qiu, Hong Liu, and Bin Zheng" *Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer Patients*
- [6] "Oluwakorede M. Oluyide, Jules-Raymond Tapamo, Serestina Viriri " *Automatic lung segmentation based on Graph Cut using a distance-constrained energy*
- [7] Jochen Penne, Christian Schaller, Joachim Hornegger, Torsten Kuwert *Robust Real-Time 3D Respiratory Motion Detection Using Time-of-Flight Cameras*
- [8] "Ales Prockjka, Hanacharvota et al", *Breathing Analysis Using Thermal and Depth Imaging Camera Video Records*