Middle-East Journal of Scientific Research 24 (10): 3263-3267, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.3263.3267

An Early Diagnosis of Lung Cancer Disease Using Data Mining and Medical Image Processing Methods: A Survey

[1]B. Muthazhagan and [2]T. Ravi

[1]Department of IT, Kings Engineering College, Chennai, India
[2]Department of CSE, S.A. Engineering College, Chennai, India

Abstract: Many research papers have been published on Lung Cancer in medical as well as in Computer Science related journals. There are many researches have been carried out for early diagnosis of Lung cancer. Worldwide, cancer disease is considered as the killer disease and it is on the rise. There are many kinds of cancer disease found out which can affect most of the parts of human body. The world statics report reveals that Lung cancer is in the top most places of cancer related deaths. Deaths due to Lung cancer are about 1.4 million per year worldwide. The conventional and non-conventional medicated methods for Lung Cancer diagnosis may leads to inaccurate results and in turn it either delays the decision making process or wrong conclusion by the physicians. In addition to this the computer assisted diagnosis approaches which uses many computational algorithms and image processing techniques have become more helpful to predict and diagnose the Lung Cancer more precisely. According to recent statistics reports, the survival rate of lung cancer disease is only about 13 to 15 percentages. If malfunctioning cells are detected in the early stage, then the survival rate can be improved up to 50 percentages. The survival rate of Lung cancer affected person is based on the early detection of lung nodules. These research papers contribute survey of Lung cancer identification in various aspects. This research paper is intended to explore the recent research on early diagnosis of Lung Cancer using Data mining and Medical Image Processing domains.

Key words: Lung Cancer · Survey · Data Mining · Medical Image Processing

INTRODUCTION

Cancer Disease: Recent days, Cancer is becoming one of the deadliest diseases in the world. Cell growth and development is one of the important metabolisms that are happening in our body all the time. Our body has many types of cells which undergo mitosis and grow several times in a particular time period maintain normal function. At times, this cell division got affected and creates an erroneous cell or an abnormal cell. These cells split further in an unordered way and affects or spread out in to other parts of the body. These cells which split unusually are called cancers. These cells can easily spread out in to other region of the same part or organ, or other parts of the body. Cancer cells can take place in almost all parts of the body and they are characterized mainly from the place and type of the cell.

Cancer is also called as Carcinoma. Cancer cells produced from connective or supportive cells (E.g.: bone, muscle,) are known as SARCOMA and produced from blood cells (bone marrow) are called as LEUKAEMIA. It is been called with different names based on the type and pattern of the cells of the body.

Early stage diagnosis of Cancer is still a challenging task for the doctors. Genetic and Environmental factors are playing very important role in developing contemporary methods to detect and prevent cancer.

Lung Cancer: Lung cancer, also named as lung carcinoma, is a malicious lung tumor regarded as by uncontrolled cell growth in tissues of the lung. If the Lung cancer is not detected in the early stage or left untreated, the growth of the malicious cell can spread across the entire part of the lung and beyond. Most cancers that start in the lung, known as primary lung cancers, are carcinomas. There are two main types of Lung cancers. They are small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). The primary causes for

Corresponding Author: B. Muthazhagan, Associate Professor & Head, Department of IT, Kings Engineering College, Chennai, India

Lung Cancer are consumption of Tobacco and Cigarette smoking. The most common symptoms are coughing (including coughing up blood), weight loss, shortness of breath and chest pains.

Fig. 1: Normal Chest and Chest with Right Lung Cancer.

There are many research papers have been published so far to bring out various data mining and medical image processing methods to support and authenticate an early detection of Lung cancer to the physicians.

Related Work: In this paper, related work is divided into two major parts. In the first part the various Data mining methods used to find the early stage of lung cancer is discussed and in the second part types of medical images and its processing methods are discussed.

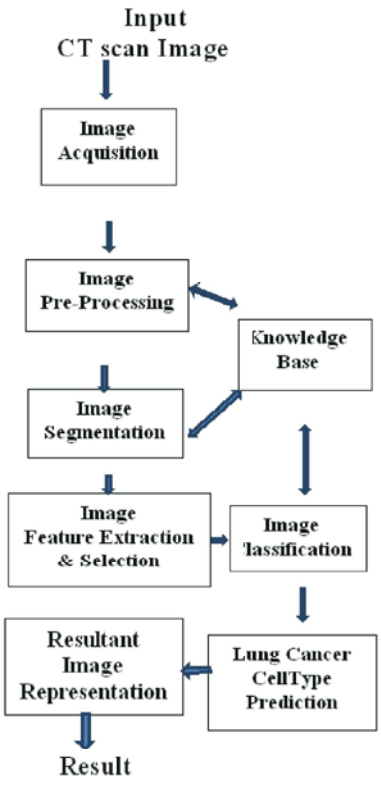

Fig. 2: Steps in Lung Cancer detection.

Data Mining Methods: Joseph A, *et al*. [1] endeavored to explain, relate and evaluate the performance of different machine learning algorithms that are useful to lung cancer prediction and prognosis. It is proved that machine learning methods are commonly used to improve the performance or predictive accuracy of most prognosis, especially when compared to conventional statistical or expert-based systems. A survey of machine learning methods used in cancer prediction showing the types of cancer, clinical endpoints, choice of algorithm, performance and type of training data are discussed in detail. The following machine learning algorithm benefits and assumptions and limitations are discussed Decision Tree, Naïve Bayes, K-Nearest neighbor, Neural Network, Support Vector Machine (SVM) and Genetic Algorithm.

N. Naveenkumar and G. Selvavinayagam [2] in this research paper, the general facts about Lung cancer and Data mining approach to Lung cancer are discussed. Importantly, the role of Decision Tree, Neural Network and Naïve Bayes algorithms are discussed. It is stated that Naïve Bayes technique is providing good accuracy results and it is the best one for detecting Lung Cancer in Clinical experts system. The prediction results are not revealed.

Fatma and Rachid [3] in this research work, the authors proposed a Modified Hopfield Neural Network (HNN) Data mining classification method and FCM clustering algorithm which has been used in color image segmentation. The segmentation results were used for Computer Aided Diagnosis (CAD) system. A preprocessing method in segmentation has been implemented to normalize the segmentation process. In this research work there were about 1000 sputum of color images have been used for testing HNN and FCM methods. As a result HNN method has shown a much better results than FCM method. But FCM is faster in cluster formation.

V. Krishnaiah *et al*. [4] proposed data mining classification techniques on Lung cancer diagnosis. The Rule set classifier, Decision Tree, Neural Network and Bayesian Network classification algorithms are used for Lung Cancer analysis. From the results it is identified that Naïve Bayes algorithm is produced better results than the other algorithms.

Juliet R Rajan *et al*. [5] proposed an unsupervised learning method is used to build an analytical model for initial detection of lung cancer. The authors used ANN technique to predict the disease. The Lung cancer was further analyzed with training resultant weight vector values.

P. Ramachandran *et al*. [6] in this research, the authors proposed a cancer risk prediction system using a multi layered method which combines both clustering and decision tree techniques. The proposed system is capable to predict lung, breast, oral, cervix and stomach and blood cancers.

Dansheng Song *et al*. [7] In this research, centrosomal numeral and morphological deviations is assessed and the degree of the difference is shown. Linear Discriminant Analysis and Support Vector Machines (SVM) with 10 fold cross validation used for classification and obtained 85%. of accuracy in the results.

Anita kumar [8] in this research paper, a study has been done on cancer perpetuation using four data mining classification algorithms. There are four Classification algorithms were used such as Random forest, CART, LMT and Naïve Bayesian. From the results it is observed that Random forest classification method outperforms the remaining classification methods for the same training dataset. The random forest method shows less absolute relative error than the other four methods whereas the Relative absolute error of LMT method is relatively high for cancer survival dataset. CART, LMT and Naïve Bayesian values for absolute relative error is greater than 50% when compared to random forest method for the same training dataset.

PatagDeoskar *et al*. [9] this research paper proposed a new algorithm which is based on ant colony based optimization technique. The proposed system consists of 3 main functionalities. The first one accepts the Lung cancer symptoms data set to find the related data from the patterns. Frequently occurring symptoms wasselected by the support count values. Then depend on the support value the ants and pheromone values are decided. Then initialize the pheromone values of cancer symptoms. The Ant Colony Optimization method provides an average accuracy of 84%.

Zakariasuliman *et al*. [10] Proposed ANN method for detection and Association rule mining data mining techniques are used in Lung Cancer in X-Ray chest images. Support vector machine (SVM) is used for image classification.

Monali Dey *et al*. [11] this survey paper provides overview on data mining algorithms that are utilized in recent Healthcare Decision Support Systems. Different medical datasets are analyzed. There are three different classification algorithms were used for classification.The algorithms are C4.5, Multilayer Perceptron and Naïve Bayes.

Medical Image Processing: NitishZulpe and VrushenPawar [12] In this research work, there are four different types of brain tumors are used and extracted the Gray-level Co-occurance matrix (GLCM) based textural features of each type and applied in two layered Feed forward Neural Network and 97.5% classification rate was obtained.

A. Gebejes and R. Huertas, [13] In this work texture is analyzed through second order statistical measurements based on the Gray-level Co-occurrence Matrix proposed by Haralick, By this method is possible to compute, along with the texture features such as Contract, Homogeneity, Dissimilarity, Energy and Entropy. The main aim of this paper is to analyze the dependency of the features.

Ada and Rajneet Kumar [14] in this research work, a hybrid technique based on feature extraction and Principal Component Analysis (PCA) is presented for Lung cancer detection in CT scan images. The CT scan image features are extracted using principal component analysis and Histogram Equalization is used for preprocessing of the images. The exact output and results are not clearly specified.

H. Mahorais, M. Zarougand L. Gabralla, [15]. This research paper review on analysis techniques used in detection of Lung cancer from CT scan images. It gives an overview on the contemporary approaches and techniques to detect Lung cancer. Comparative studies of existing approaches are also provided. In this review current detection techniques for CT images were discussed that may help researchers to have clear idea to select the correct method. It is true that lung cancer analysis techniques have been developed over the last few years. Most of the Literature reviews have been discussed elaborately, mainly in three major areas such as: preprocessing, segmentation of the lung and classification of the nodule candidates.

Anam Quadri Rashida Shujaee and Nishat Khan [16], Thisresearch paper reviews on the lung cancer detection methods using image processing. Based on recent yearsstudies the image processing techniques are used broadly in earlier detection of Lung cancer and treatment stages. And also the different types of techniques and design approaches for Lung cancer have also been discussed. In this paper, Lung cancer CT scan images are preprocessed and segmentation techniques are also implemented to get the diagnosis result. The diagnosis results and their efficiency were not discussed.

Mokhled S. Al-Tarawneh [17]. The author discussed about Lung cancer detection techniques from CT images. Here, quality of the image plays an important role on the

augmentation stage where low pre-processing techniques are used and it is based on Gabor filter implementing Gaussian rules. Then the improvised and preprocessed image that sets a base for feature extraction of images is taken and applied for segmentation principles. In this research, the main identified features for exact images comparison are pixels percentage and mask-labelling. Based on the investigational subjective assessment during the segmentation stage, Marker-Controlled Watershed Segmentation approach has produced more accurate result (85.165%) and quality than Thresholding approach (81.835%). Hence it is proved that the main detected features for accurate images comparison are pixels percentage and mask-labelling with high accuracy.

Nitin S. Lingayat and Manoj R. Tarambale, [18]. This research paper deals with the problem of developing a computer based diagnosis system to extract maximum features from the segmented wary area from the lung X-ray image. Further, these X-ray image properties can also be used to classify lung tumor as normal or malignant from theX-ray image directly. Cheat X-ray image feature extraction method recommended in this paper is less cost and less time consuming. Gray Level Co-occurrence Matrix (GLCM) is a statistical method to examine relationship of image pixels.

Disha Sharma and Gagandeep Jindal [19]. This paper proposes a Computer aided Diagnosing system to detect cancer nodules based on texture features take out from the slice of DICOM Lung CT images. In this paper, segmentation is done by Otsu thresholding algorithm and region growing techniques. About 85% of accuracy indicated by physicians and radiologists for locating malignant nodules is obtained with clinical CT images of size 2.5–7.0 mm.

Gangotrinathaney and Kanakkalyani [20]. This review article presents a number of existing techniques of medical image processing methods and their effectiveness used for prediction and analysis on CT scan images. The objective of the proposed system is to identify cancer nodules with minimum false negative rate. In general evaluation of different image processing algorithms and different classification techniques are presented. From the study, it is observed that, Thresholding segmentation approach, Neural Network classifier method and Neuro fuzzy has better results than other techniques.

Atiyeh Hashemi and Abdolhamid Pilevar, [21]. In this research paper, the author presented animage segmentation method to improvise the results from lung cancer diagnosis system by means of region growing segmentation method. For image preprocessing, to remove the noise from the CT image, the linear filtering and contrast enhancement method are used. The Fuzzy Inference System was used to classify the malignant andnormal lung nodules.A comparative diagnostic study was done between FIS and artificial neural networks (ANNs).

P. Yuvarani [22]. This survey paper provides a comparative study of various Data mining algorithms used in various stages in Lung image processing techniques and also explore different image pre-processing, segmentation, feature extraction and Classification techniques for lung cancer detection. Among the results presented SVM classifier method attained the maximum accuracy of 95.12%.

Temesguen Messay *et al*. [23] the author proposed CAD system to detect the pulmonary nodules in Lung cancer CT scan images. In the proposed system the thresholding and morphological process are combined to detect pulmonary nodules. For each nodule segment there are about 245 features and 7 fold cross validation were used. The proposed system gives 82.66% sensitivity and an average of 3 false positive.

CONCLUSION

In this paper, various lung cancer detection techniques have been discussed. This research paper is intended to explore the recent research on early diagnosis of Lung Cancer using Data mining and Medical image Processing techniques. Because in recent researches on early diagnosis of Lung cancer is viewed into two major areas, First the implementation of suitable and efficient Data mining algorithms for classification, clustering, prediction etc., Second processing of Lung cancer images that are mainly obtained from CT scan, PET scan and X-ray methods. Most of the systems have been intended to achieve the maximum accurate values with less false positive value.

Lung cancer is the one of the deadliest disease in the world that affects more number of people around the world and it is constantly increasing. To discover the Lung cancer cells,the process of finding the disease plays a vital role. Discovery and Prediction of the Lung cancer in the starting stage is very much essential to cure the disease. For this purpose and to getprecise results, the work has been divided into following steps: ImageEnhancement stage, Image Segmentation stage and Features Extraction stage and classification.

REFERENCE

1. Cruz Joseph, A. and David S. Wishart, 2006. Applications of Machine Learning in Cancer Prediction and Prognosis, A Review – Cancer Informatics, 2: 59-77.
2. Naveenkumar, N. and G. Selvavinayagam, 2015. Mining Techniques for Clinical Expert System and Predicting and Treating Lung Cancer with Big Data, International Journal of Computer Science and Engineering Communications, ISSN:2347-8586, 3(3).
3. Taher Fatma and RachidSammouda, 2010. Artificial neural network and fuzzy clustering methods in segmenting sputum color images for lung cancer diagnosis, Intl. Conf. Signal Processing, pp: 513-520.
4. Krishnaiahet, V., 2013. Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal of Computer Science and Information Technologies, 4 (1).
5. Rajan Juliet R and Jefrin J. Prakash, 2013. Early Diagnosis of Lung Cancer using a Mining Tool, International Journal of Emerging Trends in Computer Science, Special issue,
6. Ramachandran, P., N. Girija and T. Bhuvaneswari, 2014. Early Detection and Prevention of Cancer using Data Mining Techniques, International Journal of Computer Applications, 97(13).
7. Song Dansheng, Tatyana A. Zhukov and Olga Markov, 2012. Prognosis of stage i lung cancer patients through quantitative analysis of centrosomal features, IEEE.
8. Kumar Anita, 2015. A Study on Cancer Perpetuation Using the Classification Algorithms, International Journal of Recent Research in Mathematics Computer Science and Information Technology, 2(1): 96-99.
9. Deoskar Patag, Dr.Divakar Singh and Dr.Anju Singh, 2013. An Efficient Support Based Ant Colony Optimization Technique for Lung Cancer Data, International Journal of Advance Research in Computer and Communication.
10. Manzubi Zakariasuli and RemaAsheibaniSaad, 2014. Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer, Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases, pp: 32-37.
11. Dey Monali and SiddharthSwarupRautaray, 2014. Study and Analysis of Data mining Algorithms for Healthcare Decision Support System, International Journal of Computer Science and Information Technologies, 5(1): 470-477.
12. Zulpe Nitish and VrushenPawar, 2012. GLCM Textural Features for Brain Tumor Classification, International Journal of Computer Science Issues, 9(3): 3.
13. Gebejes A. and R. Huertas, 2013. Texture Characterization based on Gray-Level Co-occurrence Matrix, Conference of Infromatics and Management Sciences.
14. Ada and Rajneet Kumar, 2013. Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images, International Journal of Advanced Research in Computer Science and Software Engineering,ISSN:2277 128X, 3(3).
15. Mahersia, H., M. Zaroug and L. Gabralla, 2015. Lung Cancer Detection on CT Scan Images: A Review on the Analysis Techniques, International Journal of Advanced Research in Artificial Intelligence, 4(4).
16. Anam Quadri Rashida Shujaee and Nishat Khan, 2016. Review on Lung cancer detection using image processingtechnique, International Journal of Engineering Sciences & Research Technology, ISSN: 2277-9655.
17. AL-Tarawneh Mokhled, S., 2012. Lung cancer detection using image processing techniques, Leonardo Electronic Journal of Practices and Technologies, ISSN 1583-1078, 20: 147-158.
18. Lingayat Nitin S. and Manoj R. Tarambale, 2013. A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image, International Journal of Bioscience, Biochemistry and Bioinformatics, 3(6).
19. Sharma Disha and Gagandeep Jindal, 2011.Computer Aided Diagnosis System for Detection of Lung Cancer in CT scan Images, International Journal of Computer and Electrical Engineering, 3(5).
20. Gangotrinathaney and Kanakkalyani, 2015. Lung cancer detection system on CT images- a survey, International Journal of pure and applied research in engineering and technology, ISSN: 2319-507X, 3(9): 848-856.
21. Atiyeh Hashemi and Abdolhamid Pilevar, 2013. Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making on Fuzzy Inference System and Artificial Neural Network, I. J. Image, Graphics and Signal Processing.
22. Yuvarani Mrs. P., 2016. Analysis of Lung Cancer Detection Algorithms - A Survey, Discovery the International journal, ISSN 2278 – 5469, EISSN 2278-5450 .
23. Messay Temesguen, Russell C. Hardie and Steven K. Rogers, 2010. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery, Medical Image Analysis, 14: 390-406.