

ISSN: 2277 128X



International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Early Diagnosis of Lung Cancer with ANN, FCM and FMNN

Vinod Kumar¹ Assistant Professor

Dept. of Computer Science & Applications

Dr Kanwal Garg² Assistant Professor Dept. of Computer Science & Applications

Vijay Kher³ **Engineering Assistant** Doordarshan Kendra Kurukshetra University, Kurukshetra, India Kurukshetra University, Kurukshetra, India Pitampura, New Delhi, India

Abstract—Asian Countries like India, Pakistan, Bhutan, Bangladesh, Nepal and Sri Lanka are having their approximate total population of more than 1,500 million. In spite of enormous diversity in their demographic information, there are number of similar parameters causing to cancer. Oral cancer and lung cancer in males are noted as a few numbers, depending on its registry [15]. Uncontrolled cell progress in lung tissues causes lung cancer disease. It is the most frequent incurable malignancy in both men and women. Lung cancer can completely recover by early detection and treatment survived patient. Artificial Neural Network (ANN), Fuzzy C-Mean (FCM) and Fuzzy Min-Max Neural network (FMNN) are very effective and helpful in cancer diagnosis for its several advantages. The motive behind that the fault tolerance, flexibility, non linearity are the factors of artificial neural network. In case of FCM, it provides finest findings for overlapped data set; data point may be connected with more then one cluster centre. Non-linear separability, soft and hard decision, less training time, online adaptation is the advantages of FMNN. In this paper the author suggest using FCM and FMNN to diagnose lung cancer.

Keywords— Artificial Neural Network (ANN), Fuzzy C-Mean (FCM), Fuzzy Min-Max Neural Network (FMNN) Classification and Clustering

I. INTRODUCTION

Cancer is a primary root of adult death worldwide. In India, the International Agency for Research on Cancer (IARC) calculated indirectly that about 635000 people lost their life from cancer in 2008, approx. 8% of all estimated worldwide cancer deaths and about 6% of all deaths in India. The actual number of cancer deaths in India is likely to increase as increase the population and rates of cancer deaths are expected to rise. There are several types of cancer; particularly age-specific cancer rises due to using tobacco smoke. Thus, social and geographical distribution relating to lung cancer is differing in India. It is necessary to control and diagnose at very early stage to save from death globally [8]. The structure of the cancer cells are closely connected and are a challenging problem to detect lung cancer at early stage because symptoms come out only at later stages causing the mortality rate to be the highest among all other types of cancer [5].

Hence, early detection and early diagnosis can considerably improve the survival rate [3]. Lung cancer patient dies in large number than other cancer like breast, colon, and prostate cancers. As indicated by the latest statistics provided by whom, approximately 7.6 million deaths worldwide take place each year. Furthermore, cancer deaths are predict to go on increasing, to become around 17 million worldwide in 2030 [5].

Diagnosis and Prognosis are medical terms related to disorders. Diagnosis refers to identify a possible disease involves classification tests. Prognosis can only be done after a diagnosis is made of specific state. The prognosis predicts the death in a few short months or weeks. There are several techniques to diagnose lung cancer such as Chest Radiograph (X-Ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. Machine Learning and Data Mining are the tools for successful classification, i.e. Artificial Neural Network, Decision tree, Bayesian Network, Association Rule Mining, Clustering, Classification. Artificial Neural Networks (ANN) has been used to perform complex computations in several applications for successful detection of cancers. Such computations are done by computing nodes (neurons) and weighted connections, which work together with all data and inputs. Neurons generally operate in parallel and organized in layers, and feedback connections both inside the layer and toward adjacent layers [6].

RELATED WORK II.

An The author work on various methods for successful detection of lung cancer and [6] make a survey for classification of lung cancer by using Feed Forward Neural Network, Radial Basis Function Network, Fuzzy Neural Network and Probabilistic Neural Network. Fuzzy C-Mean Clustering Algorithm and Hopfield Neural Network are used to segment sputum color images. FCM algorithm is not good to identify the nuclei where as Hopfield neural network demonstrate better classification outcomes in extracting the cytoplasm regions and nuclei. So, Fuzzy C-Mean algorithm becomes unsuccessful in identifying the nuclei. As the medical technology become superior, it plays a good role to early diagnosis by using X-ray, CT and MRI images [5].

The diagnostic accuracy has improved by artificial neural networks. In case of lung cancer diagnosis, the main methods are biochemical diagnosis, imaging diagnosis and cytology histology diagnosis [12]. Imaging diagnosis are

Vinod et al., International Journal of Advanced Research in Computer Science and Software Engineering 3(12), December - 2013, pp. 378-383

X-ray, CT, MRI, Angiography and Interventional Radiology. One of the most widely used for the diagnosis of lung cancer, CT is an important medical imaging method.

The ANN results are compared with logistic regression by ROC (Receiver Operating Characteristics) curve analysis which is used for pattern recognition and classification. The author formulates the ROC analysis shown in Fig. 2(a), 2(b), 2(c), 2(d), 2(e), 2(f), 2(g), 2(h), 2(i), 2(j), 2(k), 2(l), 2(m) and 2(n) for finding the pattern of pulmonary diagnostic using his own constructed algorithm [5]. The receiver operating characteristic is a metric used to verify the quality of the classifiers. For each class of a classifier, ROC applies threshold values across the interval [0, 1] to outputs shows in Fig. 2(a), 2(b), 2(c), 2(d), 2(e), 2(f), 2(g), 2(h), 2(i), 2(j), 2(k), 2(l), 2(m) and 2(n).

For each threshold, two values are computed, the True Positive Ratio (the number of outputs greater or equal to the threshold, divided by the number of one targets), and the False Positive Ratio (the number of outputs less than the threshold, divided by the number of zero targets) [MATLAB documentation]. The diagnostic accuracy of ANN and logistic regression are 96.6% and 84.6% among all samples of the training group and test group. Radial Basis Neural Network is used for lung cancer screening and accuracy identification is 95.32% [3] [4]. Association Rule Mining techniques are used for the classification of hotspots from lung cancer data. By using automated association rule mining technique, hundreds of rules are to be changed on the basis of domain knowledge. This extracted knowledge provides a good approach to trace lung cancer among survivals. Segmentation can be identified by hotspot algorithm [8]. Lung Cancer Diagnosis System (LCDS) is a program compiled in MATLAB which helps to medical professional, radiologist in the analysis and evaluation of medical images as X-ray, MRI, and Ultrasound in a little time. LCDS join together the components of image processing with artificial intelligence.

The X-ray image is inputted for pre-processed in LCDS and at the end, cancerous nodule region is segmented as output image. It uncovers the grey level nodule region pixel values with an accuracy of 90% by nodes based on fuzzy system. Feature extraction phase, feature selection phase and classification phase can be determined by Machine Learning techniques [9]. To find the highest accuracy, different wavelets functions have been used in Feature Extraction and Selection. K-Nearest-Neighbour Algorithm of clustering has been developed for classification. Japanese Society of Radiological Technology successfully has done there all testing by using standard dataset of lung cancer. Nodule regions (abnormal) cases and non-nodule regions (normal) cases 130 X-ray, applied for classification and 96% accuracy levels achieved [7]. Initially, recent computer aided diagnosis system was proposed [4] for detecting lung nodules. The lung region is segmented from the CT data using adaptive threshold algorithm. This contour model is actively constructed for segmentation and eliminate lung vessel exactly in the lung region. Selective shape filter is used to detect suspicious nodules and omit renal vessel. In conclusion, nodule features is extracted. Rule-based classifier was inured to recognize true or false positive nodules with an 85% detection rate.

The author concludes that these results can help radiologist to improve the diagnosis efficiency [4]. Autofluorescence spectra of human serum are used to detect lung cancer using laser and is absorbed by electromagnetic radiation by emitting light with a longer wavelength. As a result, fluorescence of serum reflected the chemical component change. Fluorescence advantages are its sensitivity, celerity, and safety. Deficiency of this method includes low concentration of serum and low sensitivity of fluorescence. Total diagnosis accuracy of 83.3% is achieved [11]. Kernel RX-algorithm is a nonlinear, applied to detect CT images for malignant nodule, which is like to anomaly detection in military imaging applications by successfully applying RX-algorithm [11]. This is modified and applied to anomaly detection in CT images.

The author suggested for results with the help of proposed approach by using efficient technique for early lung cancer detection [2]. Even though various efforts are to be developing for new predictive models in the early detection of tumor, local failure is locally advanced in Non-Small Cell Lung Cancer (NSCLC). Yet, many cancer patients are suffered from a high local failure rate after radiotherapy. A graphical Bayesian network framework was used for predicting such local failures. Testing was done using a dataset of locally advanced NSCLC patients treated with radiotherapy. To develop predictive models of local failure in those patients, the experimental results can be demonstrated and interpreted the relationships among the different variables. Accuracy of 87.78% can be achieved by Matthew's correlation coefficient (r) of 0.74 and Spearman's rank correlation coefficient of 0.75 by cross-validation analysis [1].

An improved Group Method of Data Handling is a neural network algorithm and used for medical image diagnosis lung cancer. It can identify the characteristics of the medical images accurately. To minimize the prediction error criterion, the optimum neural network architecture is appropriate for the complexity of the medical images which is automatically organized to the Prediction Sum of Squares (PSS). Now the revised Group Method of Data Handling type neural network can be simply put on to the medical image identification. Various dissimilar output images are achieved in favour of different structural parameters by applying the conventional Neural Network. In some other cases, iterative calculations of Back Propagation Neural Network are desired to find more accurate neural network architecture. Here the author stated about the results that the revised Group Method of Data Handling type Neural Network algorithm is accurate and useful for the lung cancer medical image diagnosis [3]. The management of pulmonary disease patients, it is the important step to estimate the volume of the lungs and its viable lung tissues. The pathology is not possible to formulate the lung segmentation automatically. A statistical classifier can classify the regions and corrects the whole segmentation in the medical images. The end result of lung segmentations illustrates huge overlap and a small average boundary distance when it compared with delineations of the lung borders [12]. To detect the lung cancer in chest X-ray images founded on object shared space is finding by Parallel Immune Algorithm (IA). The author used the template matching technique and aggregates the algorithm with object shared space. Outcomes described that the algorithm is capable for diagnosing suspected lung cancer [10].

III. PROPOSED WORK

As the author surveyed many research papers and propose to implement Fuzzy Min-Max Classifier and Fuzzy C Means to the datasets of lung cancer, and uncover the problem by giving better solutions. In the supervised neural network classifier, there is a method for classification; Fuzzy Min-Max classification that forms hyperbox fuzzy sets for learning and classification. In this classification technique, the learning is an expansion and contraction process of the hyperboxes. The training data set in ordered pair is selected and locates the hyperbox which belongs to the same class including the input pattern. It every time checked the hyperbox, if it is not present there then it can formed a new hyperbox and added to Neural Network.

The Expansion Criteria is used to determine whether a hyperbox can expand or not. The min and max points are updated after the criterion satisfaction. A hyperbox expands or a new hyperbox is created, it may guide to overlap. In the end, the author approved a hyperbox Overlap Test for each input dimensions and verify the possible overlap among hyperboxes belonging to different classes. All dimensions that provide the minimum overlap value are determined and are reduced using a Contraction process [13].



Fig. 1 shows Three Layer Architecture of Fuzzy Min Max Neural Network

Fuzzy C-Mean Clustering Technique permit one part of data belongs to two or more clusters. On the basis of distance between the data point and cluster centre, the membership is allocate to each data point corresponding to each cluster centre. Membership and cluster centre are updated after the iteration.

The author explained the objective of Fuzzy C Mean Algorithm is to minimize the objective function [14]. It gives the best result for overlapped data sets. It is also allowing one part of data belongs to two or more clusters [14]. Fuzzy Min-Max Neural Network gives the advantages i.e. Non-linear separability, Online adaptation, Soft and hard decisions, Less training time [13]. The author suggested about ANN, FCM and FMNN, which may applied to standard Lung Cancer datasets one by one and the results may be noted.

The summarized description of the survey showed in the Table 1.

Sr.	Author	Datasets	Methodology	Performance			
No.							
1	Amin Moh. Roozgard Samuel	CT images 512 x 512	Kernel RX-Algorithm	Successfully Verified			
	Cheng and Hong Liu [2012]	size					
2	Hamada R. H. AI-Abs Brahim	Chest Radiographs -	Machine Learning	Classification			
	Belhaouari Samir Khaled Bashir	247 (Nodules Images -	Techniques	Accuracy - 96%			
	Shaban and Suziah Sulaiman	154 and Normal Images					
	[2012]	- 93)					
3	Tadashi Kondo, Junji Ueno	All Medical Images	Group Method of	Efficiency of			
	and Shoichiro Takao [2012]		Data Handling	diagnosis can be			
				enhanced			
4	Ankit Agrawal and Alok	SEER data - A subset	Association Rule	Patients survival time			
	Choudhary [2011]	of 13 patient attributes	Mining & Hotspots	is increased			

Table I: Different Methods Performance Advised for Lung Cancer Diagnosis

Vinod et al., International Journal of Advanced Research in Computer Science	and Software	Enginee	ring 3(12),
· · · ·	December	2012	

	December - 2013, pp. 3/8-383							
5	Fatma Taher and Rachid	Sputum Color Images -	Hopfield Neural	FCM Segmentation				
	Sammouda [2011]	100	Network and Fuzzy C-	Results - Poor				
			Mean	HNN Segmentation				
				Results - More				
				Accurate				
6	Peng Gang, Yang Xiong	Medical Images of	Immune Algorithm	Efficient for detecting				
	and Liu Li [2011]	Chest X-ray		suspected lung cancer.				
7	S. Aravind Kumar, Dr. J.	Slice Images - 685	Computer Aided	Classification				
	Ramesh, Dr. P. T. Vanathi and	(Clinical Cases - 40)	Diagnosis,	Accuracy - 90%				
	Dr. K. Gunavathi [2011]		Segmentation Fuzzy					
			Systems					
8	Xiaozhou Li, Rong Wang	Serums Data - 36	Fluorescence	Classification				
	and Ming Lei [2011]		Spectroscopy	Accuracy - 83.3%				
9	Wang Tao, Lv Jianping	Valid Data Samples -	RBFN, LLS Gradient	Accuracy - 95.32%				
	and Liu Bingxin [2011]	5,000	Descent					
10	Jung Hun Oh, Jeffrey Craft	Physical Variables	Graphical Bayesian	Accuracy - 87.78%				
	Rawan Al-Lozi and	Dataset	Network Framework					
	Manushka Vaidya [2010]	(Biomarker Proteins -						
		4)						
11	Jia Tong, Wei Ying and Wu	Thoracic CT Scans - 90	Lung Vessel	Total Detection Rate -				
	Cheng Dong [2010]		Segmentation	85%				
			Computer Aided					
			Diagnostic					
12	Thessa T.J.P. Kockelkorn Eva M.	Thoracic CT Scans	Lung Segmentation	When Automatic				
	Van Rikxoort Jan C. Grutters and	(Lung Transplantation	and Computed	Segmentation				
	Bram van Ginneken [2010]	and ILD Patients) - 12	Tomography	Methods Fail – Useful				
13	Yongjun WU, Na Wang,	CT Images of	Computer Aided	ANN Illustrate -				
	Hongsheng ZHANG Lijuan Qin	Pulmonary Nodules -	Diagnostic Scheme -	96.6 % Accuracy &				
	Zhen YAN and Yiming WU	117, (Benign -58 and	CT and ANN	Logistic Regression -				
	[2010]	Malignant -59)		84.6%				

The author compared the results, and hence suggests better solution to detect lung cancer and also proposed to test the above algorithms on standard datasets. This dataset can be normalized using Min Max method and the same will be used to further generate the Training & Testing Datasets [5].

The graphical representations of ROC (Receiver Operating Characteristics) are described here in following figures, Fig. 2(a), 2(b), 2(c), 2(d), 2(e), 2(f), 2(g), 2(h), 2(i), 2(j), 2(k), 2(l), 2(m) and 2(n) [5].



Vinod et al., International Journal of Advanced Research in Computer Science and Software Engineering 3(12), December - 2013, pp. 378-383



Fig. 2(m)

Fig. 2(n)

Table 1: Performance of various methods proposed for lung cancer diagnosis

IV. CONCLUSIONS

In this paper, the author explained all about the Lung Cancer detection survey after applied several methods and a give an innovative way for Lung Cancer detection using Artificial Neural Network, Fuzzy Min-Max Neural Network and Fuzzy C Mean. The classification methods are applied to both FMN and FCM on the X-ray 130 cancerous and noncancerous datasets available. Lung Cancer X-ray Image Datasets can be used in the future to identify the best results.

REFERENCES

- [1] Vinod Kumar, Anil Saini, "Detection system for lung cancer based on neural network: X-Ray validation performance", International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471, Vol. 2 Issue 9, Nov.-Dec., 2013, pp: (40-47)
- [2] Aminmohammad Roozgard, Samuel Cheng, and Hong Liu, "Malignant Nodule Detection on Lung CT Scan Images with Kernel RX Algorithm", Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012), 2012.
- [3] Vinod Kumar, Dr. Kanwal Garg, "*Neural Network Based Approach for Detection of Abnormal Regions of Lung Cancer in X-Ray Image*", International Journal of Engineering Research & Technology, ISSN: 2278-0181, IJERT Vol. 1 Issue 5, July 2012
- [4] Vinod Kumar, Dr. Kanwal Garg, Rakesh Kumar, "Pulmonary Nodules Diagnosis from X-Ray Imaging using Image Processing", International Conference on Recent Trends of Computer Technology in Academia, ICRTCTA, J.R.N. Rajasthan Vidyapeeth University, Udaipur, 21-23 April 2012
- [5] Vinod Kumar (Research Scholar), Dr. Kanwal Garg (Guide) "Pattern Identification by Image Processing for Early Detection of Lung Cancer", Dissertation M.Tech. Computer Science & Engineering, Dept. of Computer Science & Applications, Kurukshetra University, Kurukshetra, June-2012
- [6] Kiran S. Darne & Suja S. Panicker, "Use of Artificial Neural Networks in Medical Data Classification: A Review", International Conference on Advanced Computer Sciences, Communication and Information Technologies (ICACSIT), 2012.

Vinod et al., International Journal of Advanced Research in Computer Science and Software Engineering 3(12), December - 2013, pp. 378-383

- [7] Hamada R. H. AI-Abs, Brahim Belhaouari Samir, Khaled Bashir Shaban, and Suziah Sulaiman, "*Computer Aided Diagnosis System based on Machine Learning Techniques for Lung Cancer*", International Conference on Computer & Information Science (ICCIS), 2012.
- [8] Ankit Agrawal and Alok Choudhary, "*Identifying HotSpots in Lung Cancer Data Using Association Rule Mining*", 11th IEEE International Conference on Data Mining Workshops, 2011.
- [9] S.Aravind Kumar, Dr.J.Ramesh, Dr.P.T.Vanathi, Dr.K.Gunavathi, "Robust and Automated Lung Nodule Diagnosis from CT Images Based On Fuzzy Systems", International Conference on Process Automation, Control and Computing (PACC), 2011.
- [10] PENG Gang, YANG Xiong, LIU Li, "Parallel Immune Algorithm for Lung Cancer Detection in X-Ray Images Based on Object Shared Space", 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, 2011.
- [11] Xiaozhou Li, Rong Wang and Ming Lei, "Analysis on data for detection of Lung cancer using serum autofluorescence", International Symposium on IT in Medicine and Education (ITME), 2011.
- [8] Jia Tong, Wei Ying, Wu Cheng Dong, "A Lung Cancer Lesions Detection Scheme Based on CT Image", 2nd International Conference on Signal Processing Systems (ICSPS), 2010.
- [12] Thessa T.J.P. Kockelkorn, Eva M. van Rikxoort, Jan C.Grutters and Bram van Ginneken, *Interactive Lung Segmentation in CT Scans with Severe Abnormalities*" 2010.
- [13] Patrick K. Simpson, "Fuzzy Min-Max Neural Networks Parts 1: Classification", IEEE Transaction on Neural Network, Vol 3, No.5, 1992.
- [14] James C. Bezdek, Robert Ehrlich and William Full, "FCM: The Fuzzy C-Means Clustering Algorithm", IEEE Transactions on Computers and Geosciences, Vol 10, No. 2-3, 1984.
- [15] Moore MA, Ariyaratne Y, Badar F, Bhurgri Y, Datta K, Mathew A, Gangadharan P, Nandakumar A, Pradhananga KK, Talukder MH, Yeole BB, Sobue T. Cancer epidemiology in South Asia past, present and future. Asian Pac J Cancer Prev 2010; 11(Suppl 2): 49-66.

AUTHOR'S PROFILE



Vinod Kumar is working as an Assistant Professor in the Dept. of Computer Science & Applications, Kurukshetra University, Kurukshetra with a teaching and technical experience of sixteen years. His research interests includes Medical Image Mining, Data Mining, Web Mining and in the field of Networking. He has published various research papers in international journals and conferences. He has attended many workshops and faculty development program. He is a good counsellor and motivator to the cancer patients last eighteen years. As he was also victim of lung cancer and treated successfully by Living Foods, Self-Urine Therapy, Naturopathy, Meditation and Yoga Science.



Dr. Kanwal Garg is working as an Assistant Professor in the Department of Computer Science & Applications, Kurukshetra University, Kurukshetra with a teaching experience of twelve years. His research interest includes Data Base, Data Mining, Web Mining and Warehousing. He has published forty eight research papers in international and national journals, conferences and seminars. He has attended many workshops and faculty development programme.



Vijay Kher is working as an Engineering Assistant in the Department of Doordarshan Kendra, Pitampura, New Delhi with a professional experience of eighteen years. He has vast technical experience of GCEL 500/300/100 Watt UHF TV Transmitter, BEL 500 Watt Unmanned TV Transmitter, 20 KW NEC TV Transmitter, CATS (Computer Aided Testing Station) of Hewlet Packard along with theoretical experience of Artificial Neural Network, Fuzzy Logic, Image Processing, Simulation Concepts, Lab View Software and MATLAB. He has published several research papers in international journals and conferences.