Lung Nodule Feature Extraction and Classification using Improved Neural NetworkAlgorithm

Manjit Kaur[1], Dr. Kulwinder Singh Mann[2]

Research Scholar[1], Associate Professor[2], Department of Computer ScienceGuru Nanak Dev Engineering College, India

---***---

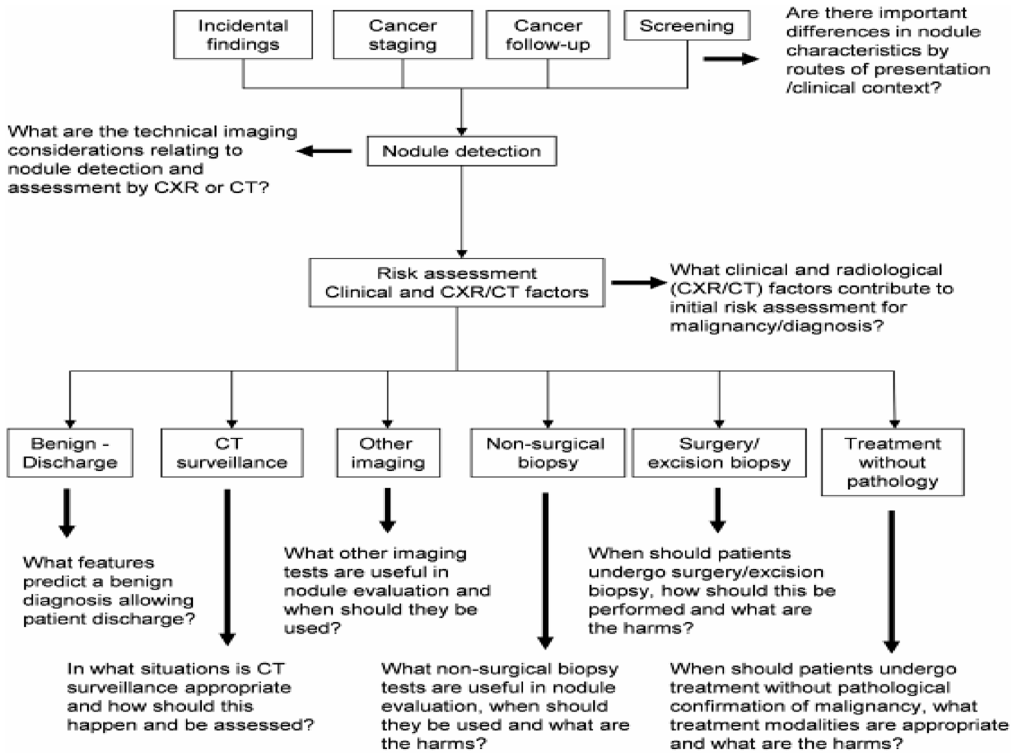
Abstract: Neural Network based techniques are being continuously evolved for classification of lung nodules detection. Different algorithms based on soft computing and hard computing has been developed to apply on nodule detection. In current research, an intelligent technique inspired from neural network algorithm have been applied and evaluated successfully to classify lung nodule. Feature extraction based on grey level covariance matrix integrated gradient (GLCMIG) technique is applied. Various texture based features like entropy, energy are extracted for nodule CT scan images then classification based on these features has been done with Intelligent Neural Network Algorithm (INNA). A contrast has been made with othertechniques to check the efficiency of the proposed method. Simulation results shows that proposed technique achieved 98.99% accuracy to classify cancer data sets which is more as compared other techniques in literature. Performance parameters like TP rate, ROC and Precision are highest for proposed method amongst other method. Hence, proposed algorithm is optimum to classify lung nodule detection.

Once a pulmonary nodule has been noticed by CT, a number of imaging modalities can be utilized to assistance supplementary notice the likelihood of malignancy. The bulk of facts involve FDG PET alongside or lacking CT. Studies have additionally assessed the utility of scintigraphic methods retaining 99m-technetium (99m Tc)-labeled compounds alongside solitary photon emission CT (SPECT), MRI encompassing diffusion-weighted (DW) and vibrant contrast-enhanced (DCE) imaging and DCE-CT. PET and PET-CT. PET-CT is a cross-sectional imaging method that provides both anatomical and useful information. It has come to be resolutely instituted in the association pathways of countless malignancies, encompassing lung cancer.

Keywords –GLCMIG, INNA, ADT, RFA, RTA

1. Introduction: -

Pulmonary nodules are well or poorly circumscribed, considering rounded constructions that materialize on imaging as focal opacities and by instituted meaning are≤3 cm in diameter and encircled by aerated lung (table 1). They might be solitary or countless and do not have associated abnormalities in the thorax, such as lymphadenopathy or pleural disease. This meaning is nowadays normally ranged to encompass nodules in link alongside the pleura. The nowadays comprehensive use of helical multi-detector line CT has made it commonplace to notice, incidentally, nodules <1 cm in diameter as well as SSNs that are slightly or wholly ground-glass opacities. These tinier nodules arguably present a larger clinical examination than their larger counterparts and are subsequently encompassed in the scope of this topic. Whereas appropriate, guidance is tailored to these disparate clusters even nevertheless it must be noted that in the works precise definitions are not always given and a collection of words are used. This case provides an adequate synopsis of the lung nodules that can furnish a momentous examination for the clinicians. The detection of pulmonary nodules is common. In populaces experiencing CT screening and at an elevated chance of lung cancer, nodules are noticed in 20–50% of people, reliant on the size of the cut-off point for delineating a nodule. The bulk of these nodules are puny and benign but a slight will be malignant and, according to the Nationwide Lung Screening Examination (NLST), competent treatment wills consequence in a reduction in mortality. It is vital to have clear guidance considering the most competent method to grasp these nodules and an assessment of how data from screening studies can be utilized to escort the method on supplementary populaces and individuals. It is acknowledged that the bulk of the facts learned from this case come from states beyond the UK and that there are potentially vital contrasts in populaces as a consequence of their geographical location.



Lung Nodule

Data mining [1] is the process of digging data for discovering latent patterns which can be translated into valuable information. Data mining usage witnessed unprecedented growth in the last few years. Of late the usefulness of data mining techniques has been realized in Healthcare domain. This realization is in the wake of explosion of complex medical data. Medical data mining can exploit the hidden patterns present in voluminous medical data which otherwise is left undiscovered. Data mining techniques which are applied to medical data include association rule mining for finding frequent patterns, prediction, classification and clustering. Traditionally data mining techniques were used in various domains. However, it is introduced relatively late into the Healthcare domain. Nevertheless, as on today lot of research is found in the literature. This has led to the development of intelligent systems and decision support systems in Healthcare domain for accurate diagnosis of diseases, predicting the severity of various diseases, and remote health monitoring.[18] Especially the data mining techniques are more useful in predicting heart diseases, lung cancer, and breast cancer and so on. The data mining techniques that have been applied to medical data include Apriori and FPGrowth, [19]unsupervised neural networks, linear genetic programming, Association rule mining, Bayesian Ying Yang , decision tree algorithms like ID3, C4.5, C5, and CART , outlier prediction technique , [20]Fuzzy cluster analysis, classification algorithm, Bayesian Network algorithm, Naive Bayesian, combination of K-means, Self Organizing Map (SOM) and Naïve Bayes, [21]Time series technique,combination of SVM, ANN and ID3, clustering and classification, SVM, , FCM, k-NN, and Bayesian Network.

II. Literature Survey

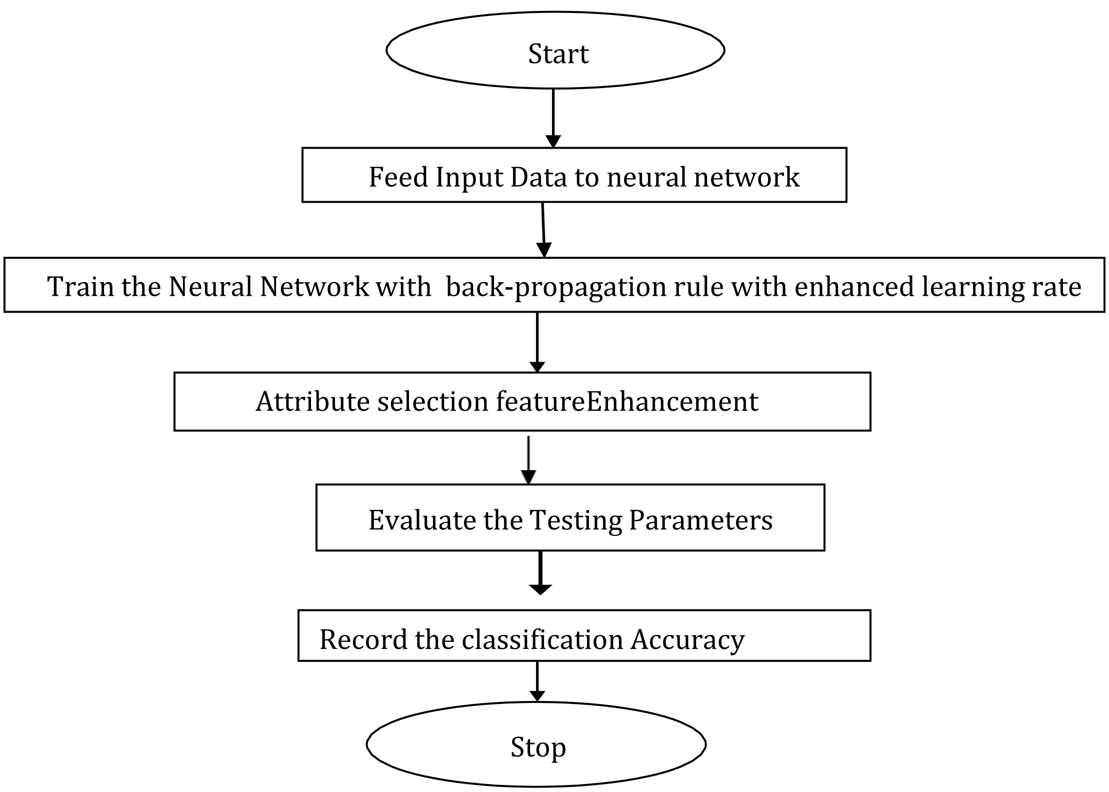
Umar et al[1] applied data mining techniques for birth outcomes. Cong et al.[2] stated that hereditary syndromes can be detected automatically using data mining techniques. Hai et al.[3] discussed medical data mining through unsupervised neural networks besides a method for data visualization. They also emphasized the need for preprocessing prior to medical data mining. Carshen et al [4] identified the need for data mining methods to mine medical multimedia content. Shariq[5] identified problems in medical data mining. The problems include missing values, data storage with respect to temporal data and multi-valued data, different medicalcoding systems being used in Hospital Information Systems (HIS). Sunil[6] explored and analyzed two programming models such as neural networks, and linier genetic programming for medical data mining. Thanh et al[7] proposed and implemented a symbolic rule extraction workbench for generating emerging rule-sets. Xiang et al.[8] explored the usage of rule-sets as results of data mining for building rule-based expert systems. Markus et al[9] proposed an algorithm for extracting association rules from medical image data. The association rule mining discovers frequently occurring items in the given dataset. Doron et al [10] proposed a classification method based on Bayesian Ying Yang (BYY) which is a three layered model. Theyapplied this model to classify liver disease through automatic discovery of medical trends. Adepele et al.[11] proposed architecture for mining geno-medical data in heterogeneous and grid-based distributed infrastructures. Cindy et al.[12] focused on decision tree data mining algorithm for medical image analysis. Especially they studied on lung cancer diagnosis through classification of x-ray images. Jeong et al.[13] presented an outlier prediction method for improving performance of classification as part of medical data mining. Jann et al[14] applied fuzzy cluster analysis for medical images. They used decision tree algorithm to classify mammography into normal and abnormal cases. Safwan et al[15] applied classification algorithm to diagnose cardio vascular diseases. For classification effectiveness they focused on two feature extraction techniques namely automatic feature selection and expert judgment. Yanwai et al[16] introduced web based data mining for the application of telemedicine. Tsang et al[17] presented an approach to integrate PSO rule mining methods and classifier on patient dataset. They used Particle Swarm Optimization technique as well. The results revealed that, their approach is capable of performing surgery candidate selection process effectively in epilepsy.

III. Proposed Method

- **Intelligent Neural Network Algorithm(INNA)**

An intelligent neural network algorithm (INNA) is a trained feed forward neural network model that maps sets of input data onto a set of appropriate outputs. An INNA consists of numerous layers of nodes in a directed graph, with each layer fully connected to the next one. Except forthe input nodes, each node is a neuron with a nonlinear activation function. INNA utilizes a learning rule called back propagation for training the network through attribute selection feature with enhanced learning rate. INNA is a modification of the conventional neural network and can distinguish data that are not linearly separable.

- **Flow Chart**



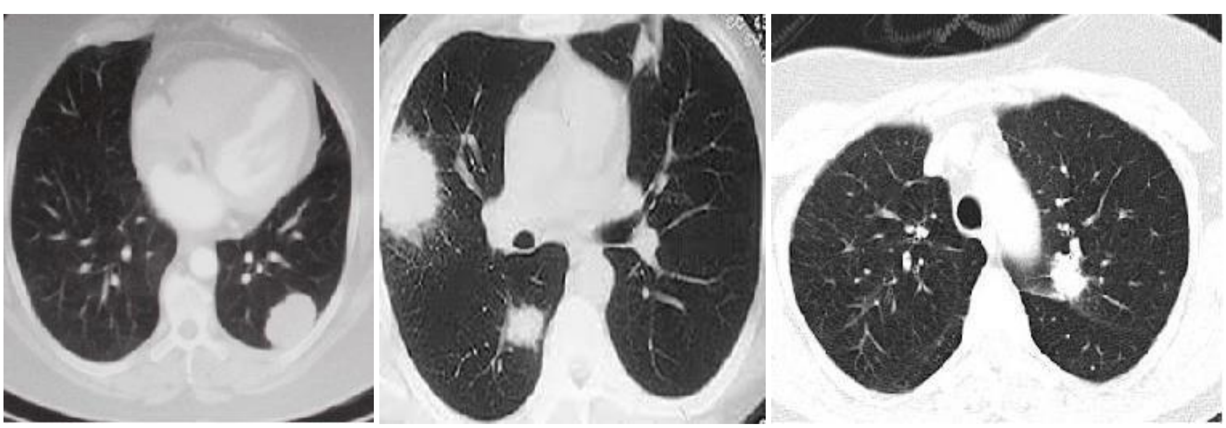
- **Lung Nodule Feature Extraction**

. The Gray Level Co-occurrence Matrix (GLCM) and associated texture feature calculations are image analysis techniques. Given an image composed of pixels each with an intensity (a specific gray level), the GLCM is a tabulation of how often different combinations of gray levels co- occur in an image or image section. Texture feature calculations use the contents of the GLCMto give a measure of the variation in intensity at the pixel of interest. Many texture based featuresare extracted from CT lung images with nodule presence using GLCM based gradient approach. Features extracted are autocorrelation, contrast, entropy, energy, Dissimilarity, cluster shade, cluster prominence. These features are selected based on their mathematical uniformity and linearity as compared to other features based on shape, location, color etc.

Autocorr	Contrast	Entropy	Dissimilr	Cshade	Cprom	Energy
23.1	0	0.55	0	-0.75	0.97	0.66
23.1	0	0.55	0	-0.75	0.97	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23	0	0.56	0	-0.75	1.3	0.65
23	0	0.56	0	-0.75	1.31	0.65
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66
23.1	0	0.55	0	-0.75	1.31	0.66

IV. Lung Nodule Image Database

The proposed system is methodically tested using different lung nodule CT image.
http://www.ncbi.nlm.nih.gov/pmc/?term=21452728[PMID]&report=imagesdocsum



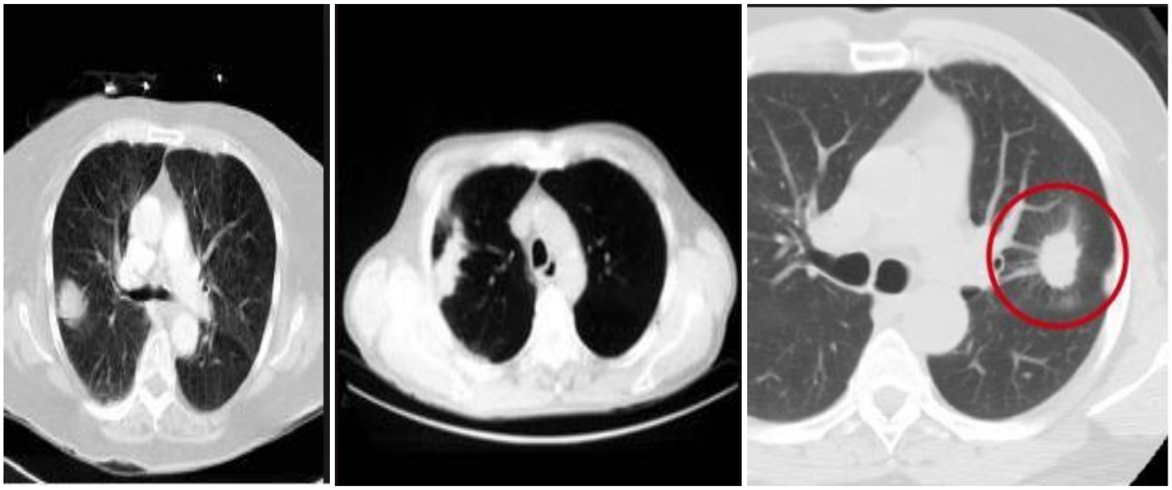


Fig. 1 CT scans Lung Nodule Images

V. Simulation Results

Comparative analysis is done for checking the effectiveness of the proposed method. As observed in Table 1. , we can see that proposed method INNA is having low computation time ascompared to other techniques.

Table 1. Computation Time for different Decision Tree Algorithms

Techniques	**Time taken to buildmodel(sec)**
Random TreeAlgorithm	0.1
Alternate Decision	0.05
Algorithm	
Random Forest Algorithm	0.2
Proposed Method	0.03

Table 2. Comparative Analysis of Error Parameters and Accuracy

Technique	Kappa statistic	Mean absolute error	Root mean squared error	Relative absoluteerror	Root relative squared error	Accuracy %
Random Tree Algorithm	0.6	0.4083	0.4618	74.85	84.67	80
RandomForest Algorithm	0.6	0.3334	0.4105	61.11	75.26	82
Alternate Decision Algorithm	0.6	0.4083	0.4618	74.85	80.45	84
ProposedMethod	**1**	**0.0898**	**0.1192**	**17.31**	**22.87**	**98**

Table 3. Accuracy by Class with RTA

Class	Recall	TP Rate	FP Rate	Precision	F- Measure	ROC Area
Non-Cancerous	0.6	0.6	0	1	0.75	0.6
Cancerous	1	1	0.4	0.714	0.833	0.6
Weighted Avg.	0.8	0.8	0.2	0.857	0.792	0.6

Table 4. Accuracy by Class with RFA

Class	Recall	TP Rate	FP Rate	Precision	F- Measure	ROC Area
Non- Cancerous	0.851	0.851	0.624	0.763	0.805	0.691
Cancerous	0.376	0.376	0.149	0.516	0.435	0.691
Weighted Avg.	0.71	0.71	0.483	0.69	0.695	0.691

Table 5 .Accuracy by Class with ADA

Class	Recall	TP Rate	FP Rate	Precision	F- Measure	ROC Area
Non-Cancerous	0.6	0.6	0	1	0.75	0.6
Cancerous	0.6	1	0.4	0.714	0.853	0.6
Weighted Avg.	0.6	0.8	0.2	0.865	0.671	0.6

Table 6 Accuracy Comparison

In table 6, we can see that proposed method achieved highest accuracy as compared to other data Mining techniques.

Technique	**Proposed Meth od**	RFA	RTA	ADA	NFR	FR	AA	ATS	AND
Accuracy %	**98%**	82%	80%	84%	95%	94%	85.3%	89.6%	88.6%

Table 7. Accuracy by Class with INNA

Class	**Recall**	**TP Rate**	**FP Rate**	**Precision**	**F- Measure**	**ROC Area**
Non- Cancerous	1	1	0	1	1	1
Cancerous	1	1	0	1	1	1
Weighted Avg.	1	1	0	1	1	1

From Table 7., we can see that True positive rate ,precision of solution, F measure and ROC value is highest for proposed technique INNA as compared to other methods as shown in table 2,3,4,5.Hence,proposed method INNA is effective in testing parameters.

VII. Conclusion

The proposed INNA based approach evolved as optimal approach to classify the cancer images with a remarkable accuracy of 98% and fast computation time of 0.03 seconds as compared to other technique and other classification methods. With such high accuracy in proposed method, it will be easy to identify the cancer and non cancer patients from different attributes of lung nodule for large data chunks where other decision tree algorithms fail to achieve high accuracy.

VIII. Acknowledgement

The authors would like to thank the management of GNDEC for providing facilities to carry outthis work.

References

[1] Umair Abdullah (2008). "Analysis of Effectiveness of Apriori Algorithm in Medical BillingData Mining". Proceedings of *IEEE*.,pp.1-5.

[2] Cong-Rui Ji and Zhi-Hong Deng. (2009). Mining Frequent Ordered Patterns withoutCandidate Generation. Proceedings of *IEEE*. , pp.1-5.

[3] Hai-Tao He and Shi-Ling Zhang. (2007). "A New method for Incremental Updating Frequentpatterns mining", Proceedings of *IEEE*, pp.1-4.

[4] Carson Kai-Sang Leung, Christopher L. Carmichael and Boyu Hao. (2007). "EfficientMining of Frequent Patterns from Uncertain Data",. Proceedings of *IEEE* ,pp.489-494.

[5] Shariq Bashir, Zahid Halim, A. Rauf Baig. (2008).," Mining Fault Tolerant Frequent Patternsusing Pattern Growth Approach". Proceedings of *IEEE* ,pp.172-179.

[6] Sunil Joshi and Dr. R. C. Jain. (2010). "A Dynamic Approach for Frequent Pattern MiningUsing Transposition of Database", Proceedings of *IEEE*, pp.498-501.

[7] Thanh-Trung Nguyen. (2010)."An Improved Algorithm for Frequent Patterns MiningProblem", Proceedings of *IEEE*, pp.503-507.

[8] Xiaoyong Lin and Qunxiong Zhu. (2010). "Share-Inherit: A novel approach for miningfrequent patterns", Proceedings of *IEEE*, pp.2712-2717.

[9] Markus Brameier and Wolfgang Banzhaf. (2001)."A Comparison of Linear GeneticProgramming and Neural Networks in Medical Data Mining", Proceedings of *IEEE ,pp.*1-10. [10]Doron Shalvi and Nicholas DeClaris., (2008). "An Unsupervised Neural Network Approach to Medical Data Mining Techniques.", Proceedings of *IEEE* , pp.1-6.

[11] Adepele Olukunle and Sylvanus Ehikioya, (2009). "A Fast Algorithm for MiningAssociation Rules in Medical Image Data", Proceedings of *IEEE,*. pp.1-7.

[12] Cindy L. Bethel and Lawrence O. Hall and Dmitry Goldgof (2007). "Mining forImplications in Medical Data." , Proceedings of *IEEE ,* pp.1-4.

[13] Jeong-Yon Shim, Lei Xu (2009). "Medical Data Mining Model for Oriental Medicine VIA By Binary Independent Factor analysis", Proceedings of *IEEE ,*. pp.1-4.

[14] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao (2001). "The Approach of Data Mining Methods For Medical Database ",Proceedings of *IEEE* ,pp.1-3.

[15] Safwan Mahmud Khan Md. Rafiqul Islam Morshed U. (2006). "Medical Image Classification Using an Efficient Data Mining Technique", Proceedings of *IEEE*, pp.1-6.

[16] Yanwei Xing, Jie Wang and Zhihong Zhao (2007). "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease". ***Proceedings of*** *IEEE*. pp.15.

[17] Tsang-Hsiang Cheng, Chih-Ping Wei, Vincent S. Tseng (2009). "Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches".Proceedings of *IEEE*. pp.1-6.

[18] Mohammad Saraee, George Koundourakis, Babis Theodoulidis. (2007). "EasyMiner: Data Mining In Medical Databases", Proceedings of *IEEE*, pp.1-3.

[19] Sam Chao(2009) "An Incremental Decision Tree Learning Methodology Regarding Attributes In Medical Data Mining". Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, pp.101-105.

[20] My Chau Tu AND Dongil Shin (2009). "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms.", Proceedings of *IEEE*. pp.1-5.

[21] Vili Podgorelec, Marjan Heriko Maribor, (2006).," Improving Mining of Medical Data by Outliers Prediction.", Proceedings of *IEEE,* pp.1-6.

[22] S. Ozekes, O. Osman and O.N. Ucan. "Nodule detection in lungs region that's segmented using genetic cellular neural networks and 3D template matching with fuzzy rule based thresholding", Vol. 9, No. 1, pp. 1-9, Feb. 2008.

[23] S.G. Armato 3rd , M. B. Altman, J. Wilkie, S. Sone, F. Li, K. Doi, and A. S. Roy. "Automated lung nodule classification following automated nodule detection on CT: A serial approach", Med. Physics, Vol. 30, No. 6, pp. 1188-1197, June 2003

[24] M. Dolejsi and J. Kybic, "Automatic two-step detection of pulmonary nodules," in Proceedings of SPIE, ser. Medical Imaging 2007: Computer- Aided Diagnosis, M. L. Giger and N. Karssemeijer, Eds., vol. 6514. SPIE, February 2007, pp. 1-12.

[25] J. S. Kim, J. H. Kim, G. Cho, K. T. Bae, "Automated Detection of Pulmonary Nodules on CT Images: Effect of section thickness and reconstruction interval", Journal of Radiology, Vol. 236, pp. 295-299, 2005