IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

A Study on Lung Nodule Segmentation and Classification using Supervised Machine Learning Techniques

K. Narmada[1*], G.Prabakaran[2], R.Madhan Mohan[3]

[1]Department of Computer Science Engineering, Annamalai University, Annamalai Nagar, Tamilnadu, India.
[2,3]Assistant Professor, Department of Computer Science Engineering, Annamalai University, Annamalai Nagar, India.

[*]*Corresponding Author: narmadhacseau@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 20/Dec/2018, Published: 31/Dec/2018

***Abstract*—** Lung cancer is one of the dangerous and life taking disease in the world. However, early diagnosis and treatment can save our life. Although, CT scan imaging is best imaging technique in medical field, it is difficult for doctors to interpret and identify the cancer from CT scan images. Therefore computer aided diagnosis(CAD) can be helpful for doctors to identify the cancerous cells accurately. Many computer aided model using image processing and Machine Learning Technique(MLT) has been researched and developed. The main goal of this research work is to evaluate the various computer-aided model, analyzing the current best model and finding out their limitation and drawbacks and finally proposing the new model with improvements in the current best model. The model utilized that lung cancer detection model were sorted and arranged on the basis of their detection accuracy. The model were developed on each step and overall limitation, drawbacks were pointed out. It is found that some has low accuracy and some has higher accuracy, but not nearer to 100%. Therefore, this research targets to increase the accuracy towards 100%.

Keywords—Image Processing, Data Mining, Segmentation, Classification, Lung Cancer, Prediction

I. INTRODUCTION

In the most recent years, the amount of information that it can extract from the data has rapidly increased. Machine learning is not just about storing large amounts of data, but it is part of Artificial Intelligence(AI). Artificial Intelligence is the analysis of the computer programs to perform tasks that usually require the human intervention, such as decision making. Making the right decision for a specific problem is the main factor for achieving our goals. For this purpose, many machine learning techniques are utilized for both classification and regression problems. Classification is used when the prediction goal is a discrete value or a class label. When the prediction goal is continuous, regression is the suitable method to use.

A. Machine learning

The essence of machine learning is to compile the data it can observe with the experience that the program learns to generate the information that it can make use of it. For example, the process of differentiating the valid emails from the spam emails. The input will be some documents or words which included in the emails and the output should be 'yes' or 'no' that indicating the email is spam or not-spam respectively, but it do not have an algorithm to accurately identify the spam emails. Machine Learning Techniques (MLT) provides a solution for this task which we provide examples of the emails that we manually labelled with spam or valid and the program can automatically learn to distinguish between them (Alpaydin, 2014).

B. Supervised learning

The main goal of supervised learning is to learn a mapping between the input and the output whose correct values provided by a supervisor. There are two main types of supervised learning, classification and regression, where there is an input K and output K, and the main role is to find a mapping between the input and the output. In classification, the task is to assign the training input to one of the predefined classes.

The approach that is followed for the prediction technique is depend on systematic study of the statistical factors, symptoms and associated with Lung cancer. Non-clinical symptoms and risk factors are some of the generic indicators of the cancer diseases. Initially parameters for the pre-diagnosis are collected by interacting with the expert of pathological, clinical and medical oncologists.

A. Lung cancer symptoms:
The following are the generic lung cancer symptoms [7].
i. A cough that does not go away and gets worse over time
ii. Coughing up blood or bloody mucus.
iii. Chest, Shoulder, or Back Pain that doesn't go away and often is made worse by deep hoarseness
iv. Weight loss and loss of appetite
    a. Wheezing
    b. Increase in volume of sputum
v. Fatigue and Weakness
vi. Repeated problems with pneumonia or bronchitis
vii. Repeated respiratory infections, such as bronchitis or pneumonia
viii. Fatigue and weakness Shortness of breath
ix. New onset of wheezing
x. Swelling of the neck and face
    a. Clubbing of the fingers and toes. The nails appear to bulge out more than normal.
xi. Paraneo plastic syndromes which are caused by biologically active substances that are secreted by the tumor.
    a) Fever
    b) Hoarseness of Voice
xii. Loss of Appetite
xiii. Puffiness of Face
xiv. Nausea and Vomiting

B. Lung cancer risk factors:
    a. Smoking:
        i. Beedi
        ii. Cigaratte
        iii. Hukka
b. Second-hand smoke
c. Radon exposure
d. High dose of ionizing radiation
e. Occupational exposure to mustard gas Chloro Methyl Ether, Inorganic Arsenic, Chromium, Nickel, Radon Asbestos
f. Air Pollution

II. LITERATURE REVIEW

Ibrahim M. El-Hasnony et al. [16] denoted a system to classify the breast cancer.This system is combined of three methods. In order to pre-process the data FRFS(Fuzzy Rough Feature Selection) is used to handle the data which are missed. To make the data cluster clustering algorithms used and features are reduced by the fuzzy algorithm through feature selection and also the features which are reduced is merged. The classification of data set is done by the D-KNN (Discernibility Nearest Neighbor) Classifier. At last the performance is evaluated and the data set is taken from the UCI repository and this model is examined under that dataset. By using K-means clustering algorithm with k-value the dataset are pre-processed for noise containing data and missing data, for this process WEKA tool and miner(Rapid Miner) is used for clustering utilization. The reduction of the clustered data is done by that feature selection algorithm. The reduced features are combined together to form a new dataset, atlast, the classification is achieved by the classifier(D-KNN). This model can classify the instances of the new dataset with accuracy of the new classifier is up to 98%. The above classification accuracy can be increased efficiently by this model.

Gopala Krishna Murthy Nookala et.al.,[17] examined the performance analysis and evaluation of different data mining algorithms used to predict cancer classification. From the acquired results, it is shown that the performance of a classifier depends on the data set, especially on the number of attributes used in the data set and one should not rely completely on a algorithm for their study.

Christopher T and J.Jamerabanu[3] described a study on mining lung cancer data for improve disease prediction value by using ant colony optimization techniques. This study used data mining techniques to detect the risk factors of lung cancer and to classify the smokers and non-smokers, who are all caused by the lung cancer. The proposed method used data mining algorithms like as Naïve Bayes, J48 and Decision Tree(DT) and is implemented using WEKA analysis tool. Naïve Bayes algorithm is performed more better based on the time and the developed method should be utilized to improve the quality of healthcare of lung cancer patients.

Rajeswararao D, et.al.,[7] analysed the performance of classification algorithms using various healthcare datasets. Carcinoma, Breast Cancer and Cardio Vascular Disease datasets are used for analysis. The algorithms such as FT, LMT, Random Forest and Simple CART are taken for analyse the performance in WEKA Analysis tool.

Moloud Abdar, et. al., [4] described the comparison of data mining algorithms in prediction of heart diseases. There are five algorithms including Decision Tree(DT), Neural Network(NN), Support Vector Machine(SVM) and k-Nearest Neighbour(K-NN), logistic regression are used for classification and comparison. After the implementation, attributes of the dataset are classified into “with” and “without” heart disease prediction. Based on the investigated methods, decision tree(DT) has achieved the best performance than other algorithms.

Durgalakshmi R and Mannar Mannan J[5], defined on prognosis of blood carcinoma utilizing data mining techniques. To analyse the possibilities of leukaemia’s presence, complete blood count (CBC)and peripheral smear are taken. For the study, the data was prepared based on the recommendations and discussions with medical experts and oncologists. Along with the data mining algorithms the semantic knowledge using ontology is utilized for predicting the blood cancer. Naïve Bayes and J48 algorithms are

utilized for classification of data set. After classifying and clustering the data semantic relationships are examined using ontology.

Neha Panpaliya, Neha Tada[5] This research paper summaries that utilizing the combination of Neural Network(NN) classifier along with binarization and GLCM will increase the accuracy of lung cancer detection process. By utilizing this system will also reduce the cost and time required for cancer detection and also if the patient is not detected with the lung cancer the system will proceed further for the prediction.

Aggarwal, Furquan and Kalra [4] proposed a model that provides classification between nodules and normal lung anatomy structure. The method extracts geometrical, statistical and gray level characteristics. LDA is utilized as classifier and optimal thresholding for segmentation. The system has 84% accuracy, 97.14% sensitivity and 53.33% specificity. Although the system detects the cancer nodule, its accuracy is still unacceptable. No any Machine Learning Techniques(MLT) has been used to classify and simple segmentation techniques is used. Therefore, combination of any of its steps in our new model does not provide probability of improvement.

Jin, Zhang and Jin [5] used convolution neural network as classifier in his CAD system to detect the lung cancer. The system has 84.6% of accuracy, 82.5% of sensitivity and 86.7% of specificity. The advantage of this model is that it uses circular filter in Region of Interest (ROI) extraction phase which reduces the cost of training and recognition steps. Although, implementation cost is reduced, it has still unsatisfactory accuracy.

Sangamithraa and Govindaraju [6] uses K- mean unsupervised learning algorithm for clustering or segmentation. It groups the pixel dataset according to certain characteristics. For classification this model implements back propagation network. Features like entropy, correlation, homogeneity, PSNR, SSIM are extracted using gray-level co-occurrence matrix (GLCM) method. The system has accuracy of about 90.7%. Image pre processing median filter is used for noise removal which can be useful for our new model to remove the noise and improve the accuracy.

Roy, Sirohi, and Patle [7] have proposed a system to detect lung cancer nodule using fuzzy interference system and active contour model. This system utilizes gray transformation for image contrast enhancement. Image binarization is performed before segmentation and resulted image is segmented using active contour model. Cancer classification is improved using fuzzy inference method. The main features like area, mean, entropy, correlation, major axis length, minor axis length are extracted to train the classifier. Overall, accuracy of the system is 94.12%. Counting its limitation it does not classify the cancer as benign or malignant which is future scope of this proposed model.

Ignatious and Joseph [8] developed a system using watershed segmentation. In pre processing it uses Gabor filter to enhance the image quality. It compares the accuracy with neural fuzzy model and region growing method. Accuracy of the proposed is 90.1% which is comparatively higher than the model with segmentation using neural fuzzy model and region growing method. The advantage of this model is that it uses marker controlled watershed segmentation which solves over segmentation problem. As a limitation it does not classify the cancer as benign or malignant and accuracy is high, but still not satisfactory. Some changes and contribution in this model has probability of increasing the accuracy to satisfactory level.

Gonzalez and Ponomaryvo[9] have planned a system that classifies lung cancer as benign or malignant. The system uses the priori information and Housefield unit(HU) to calculate the ROI. Shape features like area, eccentricity, circularity, fractal dimension and textural features like mean, variance, energy, entropy, skewness, contrast, and smoothness are extracted to train and classify the support vector machine(SVM) to identify whether the nodule is benign or malignant. The advantage of this model is that it classifies cancer as benign or malignant, Future Trends of Data Mining in Predicting the Various Diseases in Medical Healthcare System, Shubpreet Kaur[10] designed their paper has applying data mining in the medical field is an incredibly challenging mission in the medical profession. They characterizes widespread process that demands thorough understanding of needs of the healthcare.

A Critical Study of Classification Algorithms for LungCancer Disease Detection and Diagnosis by N.V. Ramana Murty[16]. In this paper they conducted an experiment to the analysis has been performed using WEKA tool with several data mining classification techniques and they found that the Naive Bayesian algorithm gives a better performance in all aspects over the other classification algorithms.

A comparative study of data mining classification techniques using lung cancer data by Tapas Ranjan Baitharu[1]. In their paper they conducted an experiment to find the impact of lung cancer data on the performance of different classifiers.
A comparative study of recent trends on cancer disease prediction using data mining techniques, by Satyam Shukla et.al [30]. This authors described their paper they applying data mining techniques like Rank based method in which reversal pairs are area and they can be easily to independent of samples by the help of which cancer diagnosis became easy.

A study on mining lung cancer data for increasing or decreasing disease prediction value by using ant colony optimization techniques designed by J.Jamera banu[4]. In their paper has successfully performed with several data mining classification techniques and they believed that the data mining can significantly help in the Lung Cancer research and ultimately improve the quality of health care of Lung Cancer patients.

In 2016, Mayuri et.al. [12] proposed lung nodule detection through concentric level partition, feature extraction and classification. Concentric level partition was constructed by an improved quick shift superpixel formulation, FS3 feature set included SIFT, MR8+LBP and multi orientation HOG was generated. Support Vector Machine classifier and Probabilistic latent semantic classifier were designed to classify the lung nodule type. This methodology was done by ELCAP public database with improved performance accuracy of 81.45%.

In 2015, Suganya et. al. [11] presented a survey on classification techniques of lung nodules. Generally, thresholding and Robust segmentation techniques are used in the segmentation process. Then, feature set containing MR8 (Maximum Response) +LBP (Local Binary Patterns), Sift Descriptor and MHOG (Multiorientation Histogram of Oriented Gradients) are used for SVM classification on ELCAP public database.

A survey on early detection and prediction of lung cancer written by Neha Panpaliya, Neha Tada[**5**], in their paper they conclude that using the combination of neural network classifier along with binarization and GLCM will increase the accuracy of lung cancer detection process. By using this system will also decrease the cost and time needed for cancer detection and also if the patient is not finded with the lung cancer the system will proceed further for the prediction process.

Early detection of lung cancer risk using data mining described by Kawsar Ahmed, Abdullah-Al-Emran[9], in their paper they shows experimental result are separated into two sections. The significant frequent patterns discover and another is represents prediction tools to predict Lung Cancer. They using a data from data warehouse, the significant patterns are extracted for Lung cancer prediction.

Model's classification of benign or malignant using Support Vector Machine(SVM) can be useful in our new model. Analyzing the literature reviews, on the basis of accuracy and advantages of the steps used, the system proposed by Ignatious and Joseph [8] is current best solution.

In image pre processing it uses Gabor filter to enhance the image and uses marker controlled watershed method for segmentation and detects the cancer nodule. This model also extracts the features like area, perimeter, and eccentricity only of the cancer nodules. It shows the comparison with other previously proposed models and highlights its accuracy 90.1% which is higher than of those. Even the system is current best solution, it has some limitations. They are highlighted below. (i) Only few features has been extracted for cancer nodules (ii)No preprocessing like noise removal, image smoothing which can probably assists in increasing the detection of nodules accurately has been implemented (iii)No classification as benign or malignant of extracted cancer has been performed.

2.1 SUPERVISED MACHINE LEARNING TECHNIQUES

A. Support Vector Machine(SVM)

Support Vector Machine (SVM) is one of the most powerful training techniques for supervised learning. Support Vector Machine was first introduced by Vapnik in 1992 (Boser, Guyon, &Vapnik, 1992). It was used for many applications for classification, regression and feature selection. In classification, support vector machine(SVM) determines an optimal separating hyperplane using the concept of margin which is the essence of the SVM. The margin is the distance between the hyperplane and the closest points to it on either side, which we want to maximize for better generalization. There is a tradeoff between maximizing the margin and minimizing the number of the misclassified examples. There are some bounds that govern the relation between the model performance and its capacity. This can be used to balance the trade-off between the model bias and the model variance.

B. Advantages and Disadvantages of Support Vector Machine

The major advantage of the Support Vector Machine(SVM) is that the training is relatively easy. It scales relatively well to the high dimensional data. The trade-off between the model complexity and the error can be controlled easily. It can deal with both continuous and categorical data. It captures the nonlinear relationships in the data. No assumptions are required regarding to the data structure, because it is a non-parametric technique. The prediction accuracy is very high and it provides a good generalization performance. It delivers a unique solution, because the optimization problem is convex, which means that it has a unique minimum value. It is robust and able to deal with data that contains errors. One of the major disadvantages of the Support Vector Machine(SVM) is the difficulty to interpret unless the features are interpretable. It can be computationally high cost and it needs a good kernel function. Its lack of transparency in results because it is a non-parametric method (Jakkula, 2006) (Auria & Moro, 2008).

C. Artificial Neural Network(ANN)

ANN is a set of connected input and output network in which weight is associated with each connection. It consists of one

input layer, one or more intermediate layer and one output layer. Learning of neural network is performed by adjusting the weight of connection. By updating the weight iteratively performance of network is improved. On the basis of connection ANN can be classified into two categories: feed-forward network and recurrent network. Feed forward neural network is the network in which connections between units do not form cycle whereas in recurrent neural network connection form cycle [9]. The behavior of neural network is affected by learning rule, architecture, and transfer function. Neurons of neural network are activated by the weighted sum of input. The activation signal is passed through transfer function to produce a single output of the neuron. Non linearity of network is produced by this transfer function. During training, the inter connection weight are optimized until the network reaches the specified level of accuracy. It has many advantages like parallelism, less affected with noise, good learning ability [10]

D.Strengths and Weakness of Artificial Neural Network

The strengths and weakness of the artificial neural network can be summarized in the following points (Basheer & Hajmeer, 2000):

1. Strengths:

• It can be used to solve linear and nonlinear programming problems.
• No prior knowledge of the process that generating the data is required for the artificial neural network to be applied.
• The ability to learn from provided examples makes them powerful and flexible; it means that the neural network learns, it does not have to be re-programmed.
• It has been successful for solving many classification, clustering and regression problems.

2. Weakness:

• It should never be viewed as a panacea to all the real-world problem, because other techniques are powerful in their own.
• The success of the model depends on the quantity of the data.
• Lack of clear guidelines for which artificial network architecture is optimal because the process is involving trial and error.

E. Decision Tree(DT)

Decision tree is one of the most widely used classifiers in statistics and machine learning. Decision tree is a hierarchical design that implements the divide-and-conquer approach. It is a nonparametric technique used for both classification and regression analysis. It can be directly converted to a set of simple if-then classification rules. It's straightforward representation makes the reader able to interpret the result and easy to understand. This section presents the basic features of the decision tree method for classification.

Strengths and Limitations of Decision Tree

1. Strengths: (Rokach & Maimon, 2005; Moore, 2001; Leung, 2007; Timofeev, 2004).

• Decision Tree is a self-explanatory tool since it has a simple schematically representation that can even be followed by the non-professionals.
• Decision Tree can easily be converted to a set of classification rules which are often comprehensible for the reader.
• Decision Tree is a non-parametric tool, therefore it does not require any functional form specification.
• Decision Tree(DT) can easily handle outliers and missing values.

Limitations of Decision Tree

• Decision Tree can be computationally expensive.
• Decision Tree can easily overfit the data, but in practice there are several tools to avoid overfitting, such as post-prune and pre-prune.
• In practice, decision tree is widely used for classification and less appropriate for estimation tasks in regression.

F. K-Nearest-Neighbours(K-NN)

K- Nearest-Neighbor is an example of instance-based leaning and it is often used for classification where the task is to classify the unseen examples based on the database stored. The observations are presented in a d - dimensional space, where K is the number of attributes or characteristics which the observation has. Given a new point, it is classified according to its similarity to the rest of the data points stored in the model by some similarity measures. The algorithm decides the class of the new point by picking the K closest points to the new example and takes the most common class among them by the majority vote to be the class of the new point. If K=1, the new point will be classified according to the 1-nearest neighbor data point to the new example. If K=2, the class of the new point will be chosen among the 2-nearest neighbors and voting would not help. Voting helps starting from K=3, classifying the new point based on the most common class among these nearest neighbors (Sutton, 2012; Larose, 2014; Tan, Steinbach, & Kumar, 2005). The intuition behind using the nearest neighbors can be clarified by the following saying "If it walks like a duck, quacks like a duck, and looks like a duck, then it's probably a duck."(Tan et al., 2005).

Advantages and Disadvantages of K-Nearest-Neighbour Algorithm

K-Nearest-Neighbour algorithm is like the rest of the techniques, it has its advantages and disadvantages (Bhatia et al., 2010; Cunningham & Delany, 2007).

A.Advantages

-The training phase is very fast and its cost is zero.
-Simple and easy for implementation.
- It can deal with the noisy data.

B.Disadvantages
- It is computationally very expensive.
- It is very sensitive to the irrelevant features.
- It is a lazy algorithm (it takes more time to run).
- It needs huge memory to store all the training examples.

G. Bayesian based classification
Bayesian Classifier predicts class membership functions, i:e: the probability that a given data belongs to a particular class (Han and Kamber, 2006). Classification is based on the assumption that effect of a feature value of a given class is independent of the values of other features; this is termed as `class conditional independence'. This assumption simplifies the computational process. Bayes Network learns using various search algorithms and quality measures. Genetic Algorithm(GA) is used as search algorithm during the classification. Base class for Bayes network classifier provides network structure, conditional probability distributions and therefore facilitates Bayes network learning algorithm.

IV. PROBLEM STATEMENT

Lung Cancer is the second leading cancer being developed globally. Purely 16% of lung cancer cases are diagnosed at an early stage. The five-year endurance rate is only 4 percent, statistically, the 5-year survival rate for patients can be improved from an average of 4% up to 49%. Near the beginning detection increases the chances for winning treatment. Hence, apart from remedial solutions some data mining solution needs to be incorporated for resolving the death causing issue. Several risk that often come across are as follows:
- Due to huge data availability information is not fully reliable and is incomplete.
- Cancer disease complication goes undiagnosed due to lack of good diagnostic level.
- Cancer remains undiagnosed due to no clear information, explanation and insufficient diagnostic criteria.
- Present cancer diagnostic methods are very expensive and time consuming.
- Difficulties are faced in identifying the types of cancer due to wrong entries in record.
- Lack of understanding of symptoms and risk factor from patient.

The proposed work presents a detailed survey of recent trends and technologies from data mining domain that are applicable to medical diagnosis.

A. Weakness
• It should never be viewed as a panacea to all the real-world problem, because other techniques are powerful in their own.
• The success of the model depends on the quantity of the data.
• Lack of clear guidelines for which artificial network architecture is optimal because the process is involving trial and error.

V.EXISTING MODEL

Lung cancer is one of the major cancer types in the men as well as in women. Smoking is the main causes for the development of Lung Cancer. Overall 20% and 80% lung cancer occurs in Non-Smokers and Smokers respectively. Currently, lung cancer is detected by using Chest X-Ray, CT scan, PET CT and Bronchoscopy etc. by the health professional.

VI. PROPOSED SYSTEM

In this new system some techniques are used related to the image processing. Techniques are crucial to the task of medical image mining, Lung Area Segmentation, Data Processing, Feature Extraction, Classification using Modified SVMs classifier. The different learning experiments were performed on two different data sets (Large and Small datasets), created by means of feature selection and SVMs trained with different parameters; the results are compared and reported in the next paper.

VII. CONCLUSION AND FUTURE WORK

Prevention of lung diseases is low in India, especially in rural, did not notice at early stage, because of lack of awareness. In this research paper here proposing a system which can predict the diseases based on the input symptoms provided by the user and help them to analyze their health status, so people can take some precautions as per the result. It could help doctors to know the health state of the patient and based on that manual diagnosis of the disease can also be easily possible. In Future work, have planned to conduct experiments on real time large health datasets to predict all the diseases and compare algorithm with other data mining algorithm. Continuous data can also be used.

REFERENCE

[1] Krishnaiah V, Narsimha G, Chandra NS. "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques". International Journal of Computer Science and Information Technologies. 2013; 4(1):39-44.
[2] Ahmed K, Emran AA, Jesmin T, Mukti RF, Rahman MDZ, Ahmed F. "Early Detection of Lung Cancer Risk Using Data Mining". Asian Pacific Journal of Cancer Prevention. 2013;14(1):595-97.
[3] Thangaraju P, Barkavi G, Karthikeyan T. "Mining Lung Cancer Data for Smokers and Non-Smokers by Using Data Mining Techniques". International Journal of Advanced Research in Computer and Communication Engineering. 2014; 3(7):7622-5.
[4] Pandya R, Pandya J. "C5.0 algorithm to improved decision tree with feature selection and reduced error pruning". International Journal of Computer Applications. 2015; 117(16):18-51.

[5] Er.Tapas Ranjan Baitharu, Dr.Subhendu Kumar Pani, "A Comparative Study of Data Mining Classification Techniques Using Lung Cancer Data", International Journal Of Computer Trends And Technology (IJCTT) – Volume 22 Number 2–April 2015.
[6] Jaimini Majali, Rishikesh Niranjan, Vinamra Phatak, Omkar Tadakhe, "Data Mining Techniques For Diagnosis And Prognosis Of Cancer", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3, March 2015
[7] Hilal Almarabeh, "A Study of Data Mining Techniques Accuracy for Healthcare", International Journal of Computer Applications, Volume 168 – No.3, June 2017.
[8] J.Jamera banu, "A Study on Mining Lung Cancer Data for Increasing or Decreasing Disease Prediction Value by Using Ant Colony Optimization Techniques", Special Issue Published in Int. Jnl. Of Advanced Networking and Applications (IJANA) Page 150.
[9] Neha Panpaliya, Neha Tadas, Surabhi Bobade, Rewti Aglawe, Akshay Gudadhe, "A Survey On Early Detection And Prediction Of Lung Cancer", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.1, January- 2015, pg. 175-184.
[10] .G. Krishnaveni, Prof. T.Sudha, "A Novel Technique To Predict Diabetic Disease Using Data Mining – Classification Techniques", International Journal of Advanced Scientific Technologies, Engineering and Management Sciences (IJASTEMS, Volume.3,Special Issue.1,March.2017
[11] K. Arutchelvan, "Prognosis of Lung Cancer Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016
[12] N.V. Ramana Murty, "A Critical Study of Classification Algorithms for LungCancer Disease Detection and Diagnosis", International Journal of Computational Intelligence Research, pp. 1041-1048
[13] Durairaj M, Deepika R, "Comparative Analysis of Classification Algorithms for the Prediction of Leukemia Cancer", International Journal of Advanced Research in Computer science and Software Engineering, Volume 5, Issue 8, August 2015.
[14] P. Thamilselvan, "An Enhanced K Nearest Neighbor Method To Detecting And Classifying Mri Lung Cancer Images For Large Amount Data", International Journal Of Applied Engineering Research Issn 0973-4562 Volume 11, Number 6 (2016) Pp 4223-4229
[15] Suganya, A., Mohanapriya, N. and Kalaavathi, B. 2015. "Lung Nodule Classification Techniques for Low Dose Computed Tomography (LDCT) Scan Images as Survey". International Journal of Computer Applications, 131(14), pp.12-15.
[16] Zhang, F., Song, Y., Cai, W., Lee, M.Z., Zhou, Y., Huang, H., Shan, S., Fulham, M.J. and Feng, D.D. 2014. "Lung nodule classification with multilevel patch-based context analysis". IEEE Transactions on Biomedical Engineering, 61(4), pp.1155-1166.
[17] Aggarwal, T., Furqan, A., & Kalra, K. (2015) "Feature extraction and LDA based classification of lung nodules in chest CT scan images." 2015 International Conference On Advances In Computing, Communications And Informatics (ICACCI), DOI: 10.1109/ICACCI.2015.7275773.
[18] Jin, X., Zhang, Y., & Jin, Q. (2016) "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network." 2016 9Th International Symposium On Computational Intelligence And Design (ISCID). DOI: 10.1109/ISCID.2016.1053.
[19] Sangamithraa, P., & Govindaraju, S. (2016) "Lung tumour detection and classification using EK-Mean clustering." 2016 International Conference On Wireless Communications, Signal Processing And Networking (Wispnet). DOI: 10.1109/WiSPNET.2016.7566533.
[20] Roy, T., Sirohi, N., & Patle, A. (2015) "Classification of lung image and nodule detection using fuzzy inference system." International Conference On Computing, Communication & Automation. DOI: 10.1109/CCAA.2015.7148560.
[21] Ignatious, S., & Joseph, R. (2015) "Computer aided lung cancer detection system." 2015 Global Conference On Communication Technologies (GCCT), DOI: 10.1109/GCCT.2015.7342723.
[22] Rendon-Gonzalez, E., & Ponomaryov, V. (2016) "Automatic Lung nodule segmentation and classification in CT images based on SVM." 2016 9Th International Kharkiv Symposium On Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW).

Authors Profile

Mrs.K.Narmada Pursuing PhD in the Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar from 2016.She has published one research paper in refereed journal.

Dr. G. PRABAKARAN working as Assistant Professor in Dept. of Computer Science and Engineering, Annamalai University, Annamalai Nagar. He has published two papers in Nationa Journals and four papers in International Refreed Journals.He has published more than 8 papers in National Conference and 13 Papers in Internbational Conferences. He has 17 years of teaching experience and 4 years of Research Experience.He is a member of CSI and ISTE. His main research work focuses on Image processing, computer graphics.

Dr. R. Madhan Mohan working as Assistant Professor in Dept. of Computer Science and Engineering, Annamalai University, Annamalai Nagar. He has published two papers in Nationa Journals and six papers in International Refreed Journals.He has published more than 4 papers in National Conference and some papers in Internbational Conferences. He has 19 years of teaching and Research experience. His main research work focuses on Wireless Ad hoc networks.

