ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Extraction of Sputum cells using Neural Network for Early Lung Cancer Detection

Sukhjinder Kaur[1], Sukhpreet Kaur[2]

M. Tech Student, Dept. of CSE, SUSCET, Tangori, Punjab, India[1]

Assistant Professor, Dept. of CSE, SUSCET, Tangori, Punjab, India[2]

ABSTRACT: In previous years various methods exists for the detection of lung cancer like CT, X-Ray, and MRI technology. Due to the time and cost constraint, these methods does not gain high importance. So this leads to urgent need for the detection of lung cancer using cost effective technique. So this paper introduces the concept of neural network in addition to minimum variance quantization method. Neural networks plays good role in the process of classification of data or images based on obtained features. The model consists of an input layer, a hidden layer and an output layer. The network is trained with ten hidden layer and one output layer by giving five inputs. One of the most common forms of medical malpractices globally is an error in diagnosis. Minimum Variance quantization is type of clustering that clusters similar featured images. The whole simulation has been taken place in MATLAB 7.14 environment to implement the procedure.

KEYWORDS: Cancer, Minimum Variance quantization, Sputum Images, Segmentation, Thresholding.

I. INTRODUCTION

Lung cancers can be accepted as because primary reason behind cancers passing away worldwide, and it is tough in order to diagnose in their beginning simply because symptoms glimpse solely with advanced stages causing the fatality rate pace to get the very best among all the kinds of cancers [1]. A lot more persons pass away on account of lung cancers when compared with any other kinds of cancers including: breast, intestines, along with prostate cancer [2]. There may be substantial research revealing which the early discovery involving lung cancers will certainly minimize the fatality rate pace. The most recent quotations according to the most up-to-date stats given by planet well-being organization imply which about 7.6 million deaths worldwide each year because of this type of cancers. Furthermore, fatality rate from cancers are expected to stay rising, to be about 19 million worldwide in 2030 [3].

There are many approaches to prognosis lung cancers, including Breasts Radiograph (x-ray), Calculated Tomography (CT) Permanent magnetic Resonance Image resolution (MRI scan) along with Sputum Cytology [4,5,6,7]. On the other hand, a large number of approaches are expensive along with cumbersome. To put it differently, a large number of approaches are sensing the lung cancers in their advanced stages, the spot that the patient's possibility of emergency is incredibly low. Thus, there exists a fantastic requirement of a whole new engineering in order to diagnose the lung cancers in their beginning. Image running approaches supply a high quality instrument intended for bettering the guide investigation. Many medical scientists applied the investigation involving sputum solar cells intended for early discovery involving lung cancers, most recent investigation pass on quantitative information, like the size, design as well as the rate on the damaged solar [8].

In this operate; sensory system along with Neural Network is proposed in order to diagnose the lung cancers in the early point. Consequently the minimum quantization Variance rule has been proposed. Studying in neurological techniques requires modifications for the synaptic connections that exist involving the neurons. Rest of the paper is organized as, Section II describes the materials and methods used in proposed work that will contain the algorithm flowchart in addition with detailed methodology, Section III describes the results and screenshots that will contain the parameters used in addition to proposed algorithmic implementation graphs. In the end Section IV contains the conclusion and future scope of the proposed work.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

II. MATERIALS AND METHODS

The accompanying steps show the diverse steps that should be proficient with a specific end goal to portion the sputum pictures to distinguish the lung tumour.

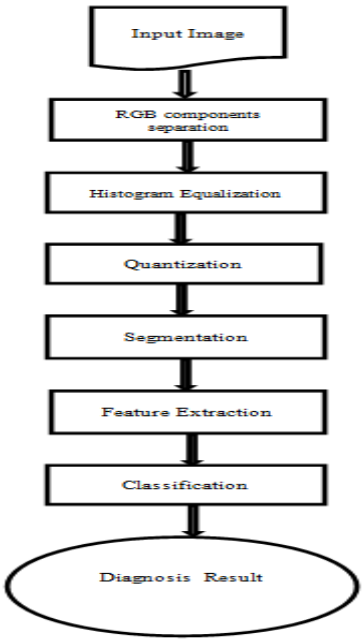


Fig. 1 Proposed algorithm

A. *Image Samples for study*

Image securing/catching of picture is the first stride of our proposed system which is finished by getting the database on the web. A database of 100 images is utilized in this work. The size of each image is 512 x 512 pixels and they are provided in RGB space. The images are obtained from sputum cytology method. The sputum cytology is an economical, non-invasive and practical method for early lung cancer detection which addresses some of the issues encountered in CT scan, and therefore can be employed as an alternative or a complement detection tool in proposed work.

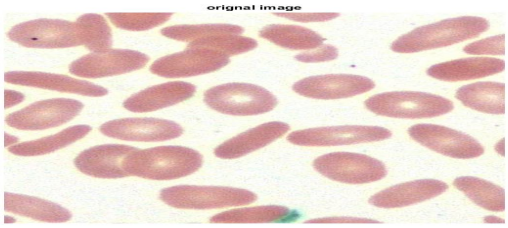


Fig. 2 Lung Cancer Sputum Image

B. *Pre-Processing*

Pre-processing improves the image in ways that increase the chances for success of the next process. Pre-processing deals with techniques for enhancing contrast, removing noise, and pictures less corrupted by patient motion can be rectified. In this work for enhancing the image quality histogram equalization was used. For implementing it first we extract the RGB components of the sputum image, and then we apply histogram equalization.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

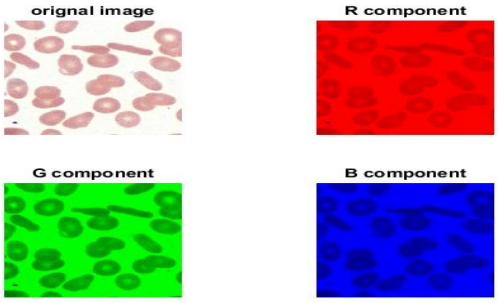


Fig. 3 RGB Components of Image

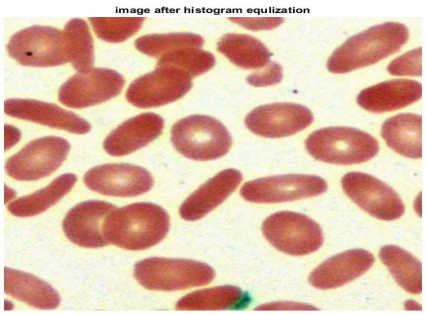


Fig. 4 Image after histogram equalization

Histogram equalization has been done to enhance the contrast of an image. In other words it is method for adjusting intensities to enhance the contrast. Histogram equalization produces the unrealistic effects in pictures but they are very useful in x-rays, satellite images as well as thermal images. Actually it allows the regions to get high contrast that has earlier low contrast.

C. *Minimum Variance Quantization*

Reducing the number of colors in an image involves quantization. The function rgb2ind uses quantization as part of its color reduction algorithm. rgb2ind supports two quantization methods: Uniform quantization and Minimum variance quantization. Minimum variance quantization, call rgb2ind and specify the maximum number of colors in the output image's colormap. The number user specifies determines the number of boxes into which the RGB color cube is divided.

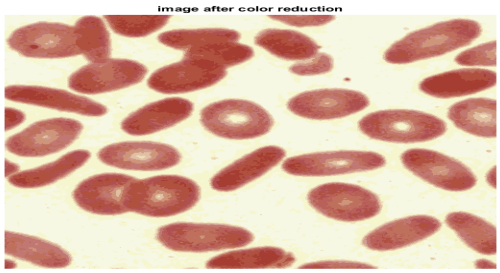


Fig. 5 Image after color reduction

D. *Segmentation*

Successful interpretation of cancer depends on accurate segmentation, so segmentation remains a challenging problem in lung cancer detection. Segmentation is the procedure of appointing a name to each pixel in a picture such that pixels with the same name share certain visual qualities. Segmentation alludes to the procedure of partitioning a computerized picture into various portions. Picture segmentation is regularly used to find questions and limits in pictures. The consequence of image segmentation is a situated of fragments that all things considered spread the whole picture. Each of the pixels in an area is comparable as for a few qualities or registered property, for example, shading, power, or composition. The division of midsection pictures is an essential stride in any framework that can prompt the early conclusion of lung tumour [18, 19].

Sputum segmentation is performed using discontinuity technique, in which image is partitioned into sub images based on sudden change in colors. The number of sub images is created according to number of colors present in image after quantization process.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

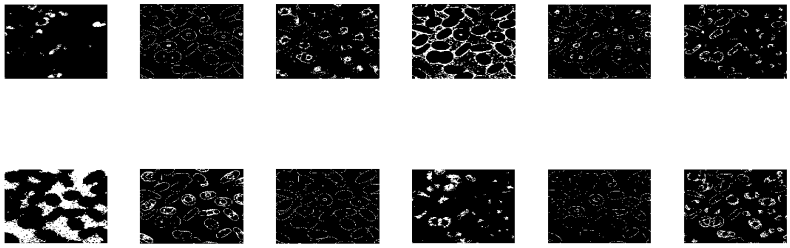


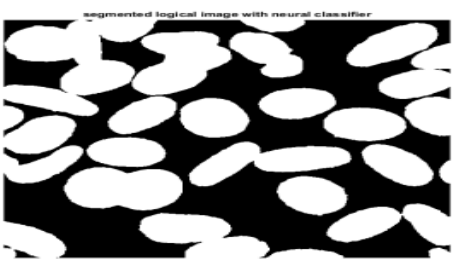
Fig. 6 Segmented images

E. *Feature Extraction*

Feature extraction is an important stage in cancer detection .This stage uses algorithms and various techniques to detect and isolate various desired portions or shapes of a given image. In the proposed method feature extraction will be done using Neural Network. Images contain large number of features that are helpful for classification [20]. It is crucial step in which feature are extracted for the recognition of an image. There are different types of features that have been extracted for image classification like auto correction, variance, sum entropy, minimum variance, standard deviation etc. [21].

F. *Classification*

In recent years, many advanced classification approaches, such as neural networks, neuro-fuzzy, k-means algorithm, expert system and SVM have been widely applied for image classification. In most cases, image classification approaches grouped as supervised & unsupervised machine learning approaches or parametric and non-parametric or hard and soft classification. The most used non-parametric classification approaches are neural networks and support vector machines. On the other hand parametric classifiers are strong. [22].



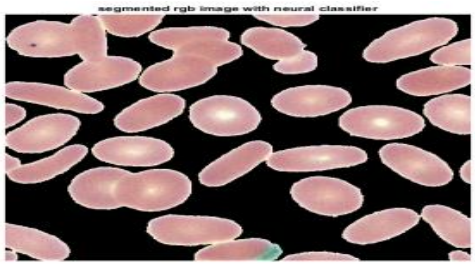


Fig 7 (a) Segmented logical image with neural classifier, 7 (b) segmented RGB image with neural classifier

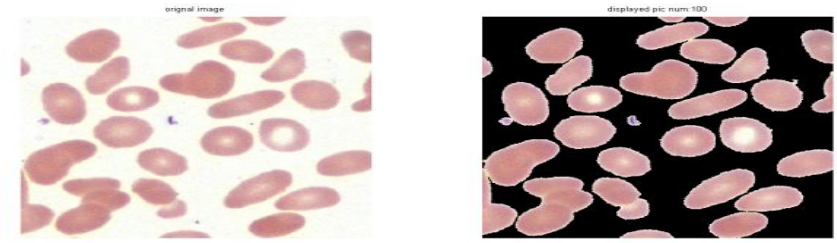


Fig. 8 original and segmented image with neural classifier

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

An ANN is a calculated model reliant on biological NNs. This one comprises of an interrelated assembly of artificial neurons and then processes data utilizing a connectionist approach to computation. Neurons are systematized into some specific layers. I/P layer encompasses merely of the original information, despite the fact that the O/P layer nodes represent the classes. Then, there might be numerous hidden layers. A key feature of neural networks is an iterative learning process in which data samples are obtainable towards the network sole at a time, as well as the weights are accustomed with the intention of forecasting the precise class tag. Reimbursements of NNs consist of their great tolerance to noisy information, in addition to their capability to categorize designs on which they have not been trained. An evaluation of benefits and drawbacks of NNs in the perspective of microarray analysis is presented. The building of the NN comprises of three kind of layers for instance I/P layer, hidden layer and output layer. The nodes in the I/P layer interconnected alongside with a number of nodes in particular hidden layer. Every single input node combined to every single node present in particular hidden layer. The nodes in some particular hidden layer may possibly associate in the direction of nodes in an additional hidden layer, or to an output layer. O/P layer comprises of one or more reaction variables. A foremost concern of the training stage is in the direction of focusing on the internal weights of the neural network which adjusted according to the transactions used in the learning process.

III. RESULTS AND DISCUSSIONS

A database of 100 images is utilized in this work. The size of each image is 512 x 512 pixels and they are provided in RGB space. Furthermore, for each image a mask was manually made as a ground truth data, dividing the images into sputum and non sputum segments. These images were obtained manually selecting the region of interest by masking the location of the corresponding pixels in binary images. A ground truth image is a binary image where one and zero corresponds to a region of interest pixels and to background pixels respectively. The ground truth images are used in comparison with the output images from the detection algorithm.
Comprehensive sets of tests are conducted to study the outcome of the proposed algorithm for detection and extraction of the cells into sputum cells and background. To evaluate the performance following quantitative measure are computed:

a) **True Positive (TP):** The pixels that are correctly classified as sputum pixels. For example a diseased person who is correctly identified as having a disease by test.
b) **False Positive (FP):** The pixels that are erroneously classified as sputum pixels. For example healthy person who is incorrectly identified as having disease by test.
c) **True Negative (TN):** The pixels that are correctly classified as non sputum pixels, for example healthy person who is correctly identified by the test.
d) **False Negative (FN):** The pixels that are erroneously classified as non sputum pixels, for example a diseased person who is incorrectly identified as being healthy by test.
e) **Sensitivity:** The sensitivity reflects the extent to which pixels classified as sputum pixels are actually sputum pixels.

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

f) **Specificity:** Specificity measures how good the background is classified.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

g) **Accuracy:** It is measured by the ratio of the total number of correctly classified pixels to the number of pixels in the image. Accuracy evaluates the overall correctly classified pixels.

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)}$$

h) **Positive Predictive Value:** The positive predictive value (PPV) or precision rate is the probability that an identified pixel is a true positive.

$$\text{Positive Predictive Value} = \frac{TP}{(TP+FP)}$$

i) **Negative Predictive Value**: The negative predictive value (NPV) is the probability that an identified pixel is a true negative.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

$$\text{Negative Predictive Value} = \frac{TN}{(FN+TN)}$$

Table I contains the sensitivity, specificity, accuracy, positive predictive value and negative predictive value results obtained by using neural network classifier with quantization. The best result is achieved by 5 to 8 color quantization operation with neural network.

Table I Results of Proposed Algorithm

No of colors	Sensitivity	Specificity	Accuracy	Positive Predictive Value	Negative Predictive Value
1	100	0	55.8231	55.8231	NaN
2	90.0118	100	95.5818	100	92.7371
3	90.0118	100	95.5818	100	92.7371
4	98.9977	100	98.0213	100	99.2932
5	98.9977	100	98.0213	100	99.2932
6	98.9977	100	98.0213	100	99.2932
7	98.9977	100	98.0213	100	99.2932
8	98.9977	100	98.0213	100	99.2932
9	98.1318	100	98.0213	100	96.8085
10	95.5843	100	98.0213	100	96.8085

Table II summarizes the performance of the proposed method and compares the result with the previous algorithm. The proposed method outperforms the thresholding method. It is found that the results from proposed algorithm are more accurate and reliable than the previous results in all cases. There is a big difference in the sensitivity since it shows the number of true positives, which means true sputum cells that are classified as sputum cells, on the other hand the specificity and accuracy are quite similar.

Table II Comparison of Thresholding method and Proposed method

Performance Measurement	Thresholding Algorithm	Proposed Algorithm
Sensitivity	81%	98.01%
Specificity	99%	100%
Accuracy	98%	98.47%

To demonstrate the results quantitatively, the error waveforms for visualization of the performance are shown as in fig.10 the sensitivity, specificity, accuracy, positive predictive value and negative predictive value for proposed algorithm.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

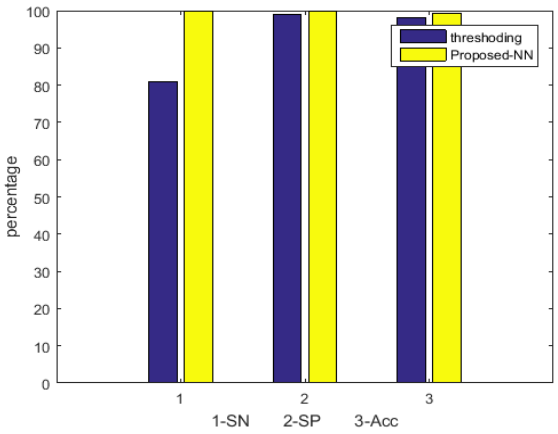


Fig.9 Performance Graph between Thresholding Method and Proposed NN based Method

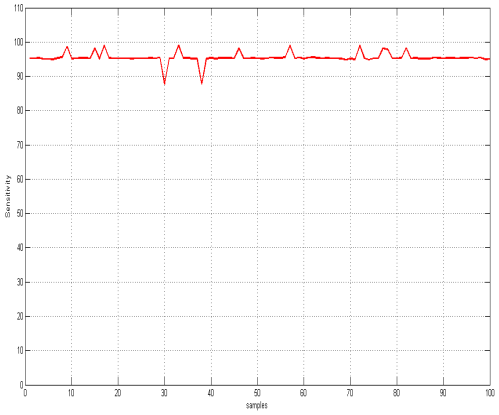


Fig. 10(a) Waveform for Sensitivity

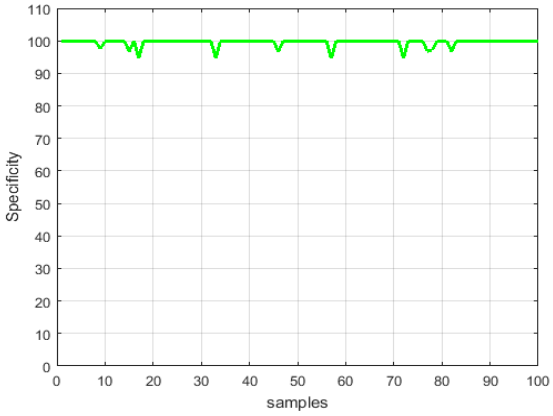


Fig. 10(b) Waveform for Specificity

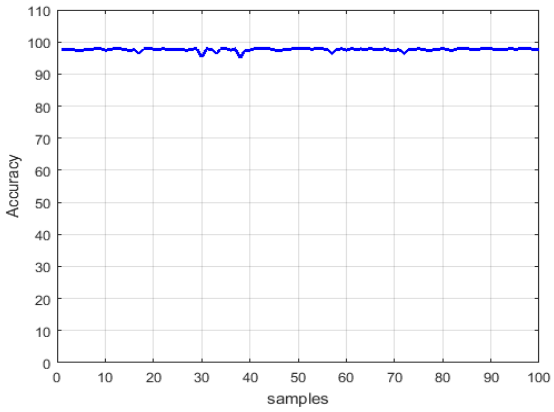


Fig. 10(c) Error waveform for Accuracy

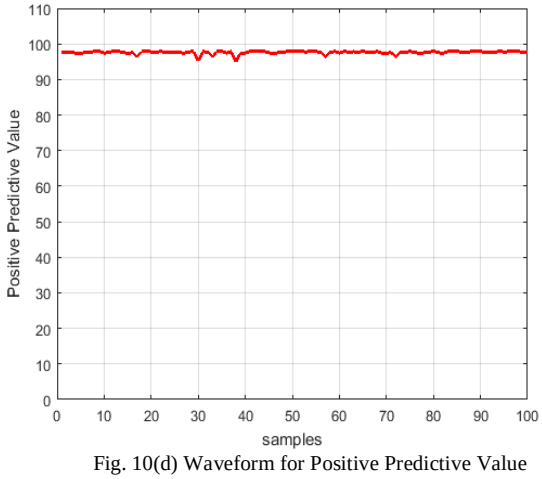


Fig. 10(d) Waveform for Positive Predictive Value

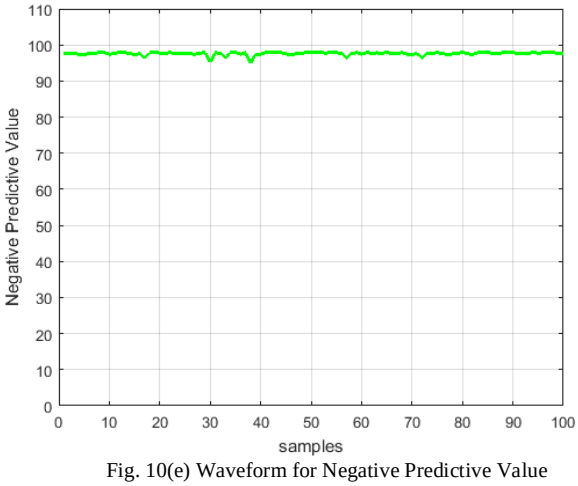


Fig. 10(e) Waveform for Negative Predictive Value

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

IV. CONCLUSION AND FUTURE WORK

As the image processing field grows day by day researcher moves towards bio medical field to emerge new techniques and to diagnose various medical diseases using automated image processing algorithms. One of them is Lung cancer segmentation correspondingly identified as cancer revealing. Till yet many researchers come up with various algorithms to detect cancer automatically from images. Successful interpretation of cancer depends on accurate segmentation, so segmentation remains a challenging problem in medical field.

This work presented an improved technique for the detection and extraction of the sputum stained cells from the background. This technique will be used as part of the analysis process for early lung cancer analysis. The NN taxonomy beside the minimum variance quantization technique has shown an improvement in all performance measurements over the previous thresholding algorithm. It has the ability to classify the cell's pixel into clear nuclei, cytoplasm and background. All these are associated with the reduction of the false negative rate while maintaining a high degree of true positive rate which was clear in the correct detection cell's pixel (TP). Moreover, it outperforms the previous algorithm by solving the problem of overlapping between the nuclei and cytoplasm regions. Overall, among one hundred sputum images, the proposed algorithm has achieved accuracy of 98.01% with high values of precision and specificity of 100% and 98.47%, respectively. The whole implementation is taken place in MATLAB R2014a environment. Future scope of this work lies in the use of any other filtering method in place of Histogram Equalization like it can be canny or median filter as they works well for all types of noise removal. Moreover we can use any other segmentation method in place of neural network it can be SVM (Support Vector Machines) or AFFN (Artificial Feed Forward Neural Network).

REFERENCES

1. Ada Rajneet Kaur,"Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images ,"IJARSSCE, Vol.3 , pp.187-190 , 2013.
2. A. D. Back, T. Chen," Universal Approximation Of Multiple Non Linear Operators By Neural Networks, Neural Computation," 2002,pp. 2561 – 2566.
3. Ankit Agrawal and AlokChoudhary,"Identifying HotSpots in Lung Cancer Data Using Association Rule Mining," IEEE, pp. 995-1002, 2011.
4. A. Bishop,"Curvature Driven Smoothing: A learning Algorithm for Feed Forward Networks," IEEE, vol. **4**, 1993,pp. 461-489.
5. Derong Liu;Zhongyu Pang; Lloyd S.R,"A Neural Network Method for Detection of Obstructive Sleep Apnea and Narcolepsy"- Based on Pupil Sizeand EEG, 2008V-19 I-2 18. D. Itchhaporia.
6. A. Wang, "Pattern recongnition: neural networks in perspective", IEEE, vol. 8, 1993, pp.5-60.
7. Frenster, J.H, "Neural Networks for PatternRecognition in Medical Diagnosis", Annual International Conference in the IEEE Engineering in Medicine and Biology Society, vol. 12, No. 3, issued 1990, pp. 1423-1424.
8. U. Javed, M. M. Riaz, T. A. Cheema and H. M. F. Zafar., "Detection of Lung Tumor in CE CT Images byusing Weighted Support Vector Machines", IEEE, pp. 113-116, 2013.
9. V.V. Thakare, P. Singhal," Neural network based CAD model for the design of rectangular patch antennas, "JETR,vol. **2(7)**, 2010.
10. Wan Hussain, Wan Ishak,"The Potential of Neural Networks in Medical Applications" -Faculty of Information Technology,Universiti Utara Malaysia, 06010 Sintok, Kedah, MALAYSIA(2002).
11. X. Yao," Evolving Artificial Neural Networks," IEEE, vol. **87**, 1999, pp. 1423-1447.
12. Zhi-Hua Zhou, Yuan Jiand,"Medical diagnosis with C4.5 rule preceded by ANN ensemble", (IEEE) Information technology in biomedicine, 2003.
13. Y. Ota, B. M. Wilamowski,"Analog Implementation of Pulse-Coupled Neural Networks, "IEEE,vol. **10**, 1999, pp. 539 – 544
14. Y. Zeng, J. Starzyk," Statistical Approach to Clustering in Pattern Recognition ,"IEEE, 2001, pp. 177 – 18.
15. Coppini, G., Diciotti, S., Falchini, M., Villari, N., Valli, G. ," Neural networks for computer aided diagnosis: detection of lung nodules in chest radiograms", IEEE Transactions on Information Technology in Biomedicine. 2003;4:344–357.
16. Patil S.A., Udupi, V.R., Kane, C.D., Wasif, A.I., Desai, J.V., Jadhav, A.N. ," Geometrical and texture feature estimation of lung cancer and TB image using chest X-ray database", in IEEE, 2009.
17. S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for kmeans clustering," Pattern Recognition Letters, vol. 25, pp. 1293-1302, 2004.
18. Charles A. Poynton (2003). Digital Video and HDTV: Algorithms and Interfaces. Morgan Kaufmann. ISBN 1-55860-792-7.
19. Nicholas Boughen (2003). Lightwave 3d 7.5 Lighting. Wordware Publishing, Inc. ISBN 1-55622-354-4.
20. R. W. G. Hunt (2004). The Reproduction of Colour (6th ed.). Chichester UK: Wiley–IS&T Series in Imaging Science and Technology. ISBN 0-470-02425-9.
21. R. C. Gonzalez and R. E. Woods, "Digital Image Processing," second ed., Prentice Hall, Englewood, Englewood, Cliffs, NJ, 2002.
22. J. Patil, and S. Jadhav "A Comparative Study of Image Denoising Techniques," International Journal of Innovative Research in Science, Engineering and Technology, vol. 2, issue 3, March 2013.