International Journal of Research in Advent Technology, Vol.2, No.5, May 2014
E-ISSN: 2321-9637

Genetic Algorithm Based Classification for the Lung Needle Biopsy Images

S.T.PremKumar[1], M.Sangeetha2
PG Student[1],Assistant Professor[2]
Department of Computer Science and Engineering[1, 2],
Adithya Institute of Technology[1, 2], India
premkumarsmt@gmail.com[1],sangee_ganapathy@gmail.com[2]

Abstract-Lung cancer is one of the most common malignant cancers that spread worldwide. Nowadays, detection of lung cancers take places in advanced stages. Noveldetecting technique by CT scanning of cancer making it possible to detection in its initial stages, when it is most curable.Cancer is an irregulargrowing of cells in lungs. Normally cancer will grow from cells of the lungs, blood vessels, nerves that appear from lungs. Types of cancer which are-squamous carcinoma, adenocarcinoma, small cell cancer and nuclearatypia. Specialistconclude size and position of cancer, if it is increasing into nearby tissues region, and if it has chance to spread into lymph glands in the neck. This approaches also check for spread of cancer in lungs and its neighboring tissues.Cancer can damage the normal lungs cells by producing swelling, exerting compression on parts of lungs and increasing pressure inside the chest. Each type of cancer has individual characteristics. Thistechnique is especially for classifying dissimilar cancerous types in the lung tissue.

Keywords:lung needle biopsy, dictionary learning,genetic algorithm,hierarchical fusion.

1. INTRODUCTION

Lung cancer studiedthrough by doctors but its grading gives different decisions which may differ fromone doctor to another. The classification and accurate determination of lung cancer grade is very vital because it influences and specifies patient's treatment scheduling and finally their life. A newtechniquemultimodal sparse representation-based classification (mSRC), suggested for classifying lung needle biopsy images. In this technique data acquisition is done through the new method, the cell nuclei are mechanically segmented by itself from the input images caught by needle biopsy specimenswhich areobtainable in this research work, which is samples of the human thinking methods and the classification results are compared with some other computer-aided lung cancersanalysismethodsshowing the efficiency of the proposed methodology.

The three features modalities such as texture, color and shapeare extracted from the segmented cell nuclei from input images. After this procedure, mSRC goes through a testing level and traininglevel. The training level, three discriminative subdictionaries corresponding to three features information are togetherknowledgeable by a genetic algorithm directed multimodal dictionary learning approach.[8] The dictionary learning is used to select highest discriminative samples and encourage large disagreement amongstdissimilar subdictionaries.

In testing phase, when a novel image comes, a hierarchical fusion scheme is applied, which originally prediction of labels of all cell nuclei by fusing modalities such as color, texture and shapeare predicts the label of the image by maximum popular voting.The above cell nuclei areas can be separated into five classes which consist of four cancerous classes andoneregular class (non-cancer). Usually, lung cancer can be categorized into four types: squamous carcinoma (SC), adenocarcinoma (AC), small cell cancer (SCC), and nuclearatypia (NA). Fig. 1 shows several sample images of each of the four cancerous types. The results prove that the multimodal information isvital for lung needle biopsy image classification, this technique is particularly for classifying various cancerous types. The output of the proposed novel mSRC model present reasonably higher accuracy, and are more similar with other existing algorithms.

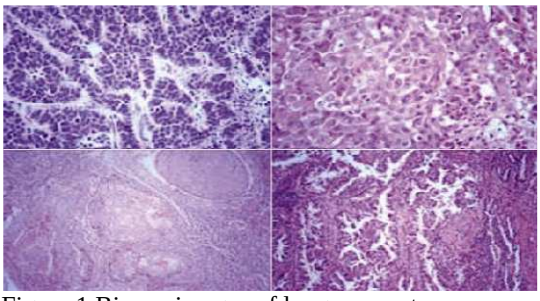

Figure.1 Biopsy images of lung cancer types.

2. LITERATURE REVIEW

The lungs are a complex tissue which containsnumerous structures, such as vessels, splits, bronchi or pleura that can be situatednear to lung nodules. Very Simple thresholding approaches are regularlyenough for the separation of solid well-circumscribed nodules, while lung nodules close to vessels or pleura needadditional complex schemes exploiting geometrical and gray level features mined from nearby structures of the lungs. Variousmethods have been suggested to summaryof the lung nodules close to vessels or pleura.

Human associated parasites create a problematic in maximumhumidcountries, producing death or physical and psychologicalillnesses. Their conclusionregularlytrusts on the visualexamination of microscopy images, with error rates that may collection from moderate to higher. [3]The problem has been addressed through computational image analysis, but only for a rare species and images free of fecal layers. In routine, fecal layers are aactualtrial for programmed image analysis.

Firstmethod exploits ellipse matching and image foresting transform for image subdivision, multiple item descriptors and their optimalgrouping by genetic software design for object symbol, and the optimum-path forest classifier for object acknowledgment. The output demonstrations that this method is a favorablemethodnear the completely automation of the enteroparasitosis diagnosis.

The effectiveness of sparse representations gained by learning a set of over complete basis or dictionary in the context of action recognition in videos. While the work focusses on distinguishingthe movement of human, [6] physicalappointments as well as the appearance of face and the suggestedmethod is fairly general and can be used to address some other classification methods

Thesuggestedapproach is computationally effective, highly accurate, and is strong against partial sealing, spatiotemporal scale variants, and to some range to view changes. This robustness is attained by manipulating the discriminative nature of the sparse representations joined with spatio-temporal motion descriptors. The fact that the descriptors are mined over multiple temporal and spatial determinations make them indifferent to scale variations. The descriptors being calculated locally make them robust besideblocking or other distortions. Features such as compressedsample can also develop the recognition accurateness but are highlycostlycomputationally [8].

The presentation of the collective system is calculated in fact and the outputcorrectness, rapidity, robustness, and anactive retinal vesselsegmentation method based on supervised classification usingacollective classifier of boosted and bagged decision trees. In this technique 9-D feature vector which contains of the vessel mapacquired from the orientation examination of the gradient vectorfield, the morphologicalrevolution, line strength measuresand the Gabor filter reaction which translatesparticulars tosuccessfully handle both standard and pathological retinas.

3. RELATED WORKS

An automaticmethod for finding brain tumor in the MRI image by means of the simple image processing methods. [8] It is decided that the system result in better detection of presence of tumor in the image. The considerable accuracy level is detected and it mines the important features which are used to identify the class of the tumor. Performance and accurateness of the designed system is found to be better. Hence it candetect the Brain Malignant cells in real time and provide exactness detection of the class of the Brain Tumor.

The system proposed in this technique can be used in future to identify the class of any type of tumor with appropriatechanges in the system. The only combination of CT, PET, and MRI etc. can increase the performance of the system. The work includesmining of the significant features for image recognition. The features minedoffer the property of the textures are stored in knowledge base. Texture features or more exactly, Gray Level Co-occurrence Matrix (GLCM) features are used to differentiate between standard and irregular brain tumors [8]. Dissimilar features are enlisted as:

Contrast

$$\boldsymbol{R_1 = \sum_{i,j=0}^{N-1} p(i,j) * (i-j)^2} \quad \textbf{(1)}$$

Angular Second Moment

$$\boldsymbol{R_2 = \sum_{i,j=0}^{N-1} P_{i,j}^2} \quad \textbf{(2)}$$

One of the most vital problems in the segmentation of lung nodes in CT imaging rises from possible accompanimentshappeningamong nodules and other lung structures, such as vessels or pleura. The problematic of vessels additions by suggesting an automatic correction technique applied to an initial rough segmentation of the lung nodule. [12]

Inverse Difference Moment

$$\boldsymbol{R_3 = \sum_{i,j=0}^{N-1} \frac{p(i,j)}{1+(i+j)^2}} \quad \textbf{(3)}$$

Entropy

$$\boldsymbol{R_4 = \sum_{i,j=0}^{N-1} p(i,j) * \left[-ln\left(p(i,j)\right)\right]} \quad \textbf{(4)}$$

Input images are occupied as input which is applied in the equitation (1), (2), (3), (4), (5). Different MRI examples are collected and given as input for the query phase. Database is a grouped database containing 70dissimilarimages characterized into 4 classes. Any automatic, computerized calculation of disease needsestablishing of healthy standards against which a test subject can be equated. A high quality, carefully designed image database of healthy subjects could be of value to many clusters for creation of healthyatlases, for calculation of disease, and for calculation of the effects of both gender and healthy aged.

Table 1Features Extracted

Angle	**ASM**	**Contrast**	**Entropy**	**IDM**	**Dissimilarity**
0^0	4578	2341894	-76.9528	3.0731	29370
45^0	4190	3236016	-53.3858	3.0464	34456
90^0	4804	1964810	-145.0278	13.089	25346
135^0	4156	2901768	-51.9995	3.0697	32744

All images were divided for the existence of disease. Images containSC, AC, SC, and NA which goal to make this database effort efficiently. Table 1describes the features mined from affected MRI. The features are calculated by using GLCM in four dissimilar angles (0º, 45º, 90º, and 135º).The finalgoal in image processing applications is to extract significantfeatures from the image data, from which a graphic, informational, or reasonableview can be attained by the device.

Dissimilarity

$$R_5 = \sum_{i,j=0}^{N-1} p(i,j) * |(i-j)| \quad (5)$$

An image preprocessing step goals to obtain the distinctcell nuclei from the caught images by segmenting the cellnuclei from the background. The image preprocessingstep is with the followingstages: smoothing the imagesthrough Gaussian kernel, segmenting the images by means of Otsu'salgorithm, and labeling the associatedwith cell nuclei regions. Thereason of adopting Otsu's algorithm is that the difference betweenthe cell nuclei region and background is large enough to besimply separated see sample images in Fig.3. Fundamentally, thesegmentation results can meet scientific requirements accordingto the pathologist's suggestions.

4. PROPOSED METHOD

Usually, the suggested method contains three phases (Fig. 2). The first phase is the data acquisition procedure, whichpurpose is to extract the features for cell nuclei in lung needle biopsyimages. Later this process and thetechniqueunder goes over the rest of the two training and testing phases.In the training phase, the newidea of dictionary in the pattern recognition/computer vision, where dictionaryresources a collection of elements or words or feature vectors.

The features extracted from the three modalities such as color, texture and shapeof every single cell nucleus in the data acquisitionprocedure, which build three original subdictionaries on color, texture and shapeby collecting the corresponding feature vectorsof individual cell nuclei.

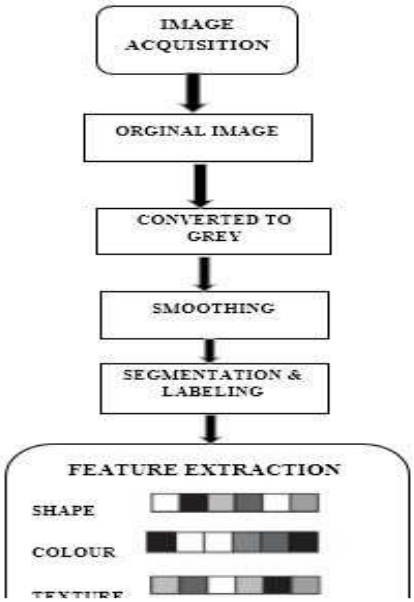

Figure.2 System Architecture

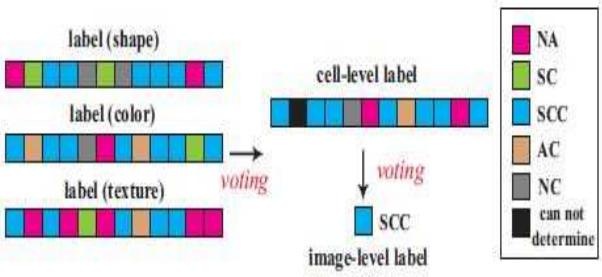

Figure.3 Extraction of features and the label of each testing image is determined by voting on the cell-level label.

A feature extraction step goal to extract features for distinctcell nuclei from three modalities color, texture and shape respectively. Exactly, as shown in Table 2,shape-based (9), color-based (11), and texture-based(16) features aremined. For the Fourier descriptor, the alphabet "i" in thebracket means that only use the second, third, and fourthcoefficients. The first coefficient is the mean value of bordercoordinates, which is regularlymeasured as unusable in featurerepresentation. In the texture-based features, the alphabet "j" in a bracket means that which calculate the four directions $(0^0, 45^0, 90^0, 135^0)$ for the four features energy, entropy, contrast,and divergence, thereforetotally 16 texture-based features areextracted.

Table 2features are extracted from every cell nucleus region

Color (11)	Shape (9)	Texture (16)
R,G,B	height	energy (4)
gray variance	circumference	contrast (4)
H,S,I	elongation	entropy (4)
gray mean	area	divergence (4)
central moment	width	
feature gray	circularity	
IOD	Fourier descriptor (3)	

5. EXPERIMENTAL SETUP AND RESULT ANALYSIS:

The proposed real time lung cancer system is estimated under certain hardware, software constraints and necessities. The hardware platform used throughout the estimation was constructed on an AMD A8 Elite Quad-Core @ 2.10 GHz processor, 4 GB of DDR3 RAM the software development has been supported out in MATLAB on Microsoft Windows 7 Ultimate 64-bit operating system.

5.1 Image Set Details:

The image set contains 270 needle biopsy images of five dissimilar classes: 50 standard or normal (NC) images, 90 AC images,55 SC images, 45 SCC images, and 30 NA images. Each image is labeled by skilled pathologists. After the cell nuclei segmented, total 4350 cell nuclei which including the overlapping cell nucleus regions are take out from all the images. The quantity of segmented cell nuclei in every image differs from 4 to 80.

Table 3validation on genetic algorithm-based multimodal dictionary learning

Methods	F1 score	TNR	Precision	Recall	Accuracy
shape only	0.405	0.867	0.376	0.394	0.511
color only	0.525	0.895	0.524	0.571	0.622
texture only	0.414	0.860	0.383	0.521	0.519
Shaped D.L	0.429	0.879	0.412	0.360	0.548
colorD.L.	0.543	0.890	0.551	0.543	0.627
textureD.L	0.428	0.878	0.437	0.473	0.556
mSRC	0.862	0.962	0.834	0.913	0.867
SCT	0.620	0.923	0.613	0.755	0.726

5.2 Evaluation Metrics:

The accuracy, precision, recall, F1 score, and true negative rate (TNR) are employed for calculating the multiclassification results. The accurateness is calculated as the sum of correctly categorized images separated by the amount of total images. For precision, recall, F1 score, and TNR calculate not only the classdetailed values of each particular class NA, SCC, SC, NC, and AC, but also the mean values by averaging totally the parallel class detailed values.

Table 4 Comparison with related methods

Methods	F1 score	TNR	Precision	Recall	Accuracy
KSRC [37]	0.804	0.953	0.782	0.843	0.830
LapRLS 36]	0.657	0.907	0.533	0.538	0.625
MCMI-[16]	0.563	0.899	0.585	0.564	0.608
ESRC [13]	0.777	0.940	0.730	0.884	0.800
mcSVM [7]	0.576	0.921	0.598	0.577	0.674
mSRC	0.862	0.962	0.834	**0.913**	0.867
mSRC (rw)	**0.866**	0.963	**0.846**	0.901	**0.881**
mSRC (tn)	0.846	**0.967**	0.841	0.858	0.867

Also the unique population selection principleselecting the chromosomes through the maximum K/2 suitability totals in every duplication and consider supplementary population collection criteria such as tournament and rangeroulette wheel range.

6. PERFORMANCE ANALYSIS

To validate the performance with various numbers of training images which is randomly sample a certain quantity of lung needle biopsy images as training images, and the rest ones are used as testing images. For the training images, the said limitations can be successfully learned by relating the leave-one-out testing. The attained parameters will be used for construction the mSRC classifier, and finally the usesof the mSRC to categorize the testing images

6.1 Evaluating the Hierarchical Fusion Strategy:

Notice that the output of mcSVM, KSRC, ESRC, and mSRC by means of the hierarchical fusion strategyovertake individuals of MCMIAdaBoost and LapRLSwithout using hierarchical fusion strategy, since both MCMI-AdaBoost and LapRLS essentially belong to image-level analysis techniques. In MCMI-AdaBoost, the space between each two images is calculated by Hausdorff distance. In LapRLS, the k-

means clustering algorithm is applied to approximately partition the training cell nuclei into some clusters.

6.2 Evaluating the Performance of SRC:

The same hierarchical fusion strategy is assumed in mcSVM, KSRC, ESRC, and mSRC. The only variance among ESRC and mcSVM is the classifiers used for the single-modal classification. ESRC uses SRC though mcSVM uses SVM. Conversely, ESRC acquires better classification performance than mcSVM, which authorizes the efficiency of SRC on single-modal classification.

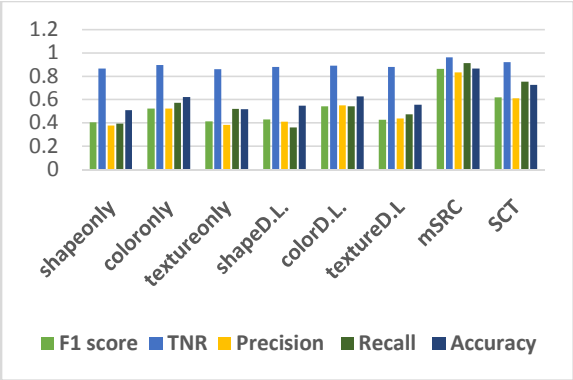

Figure.4 Comparison of Various Features

Obviously, thistechnique exceeds the related works on practically completely the listed quantities. Also as suggested in the accuracy of NC is measured as a very significant value, because a higher exactness of NC specifies a lower possibility of the item that the system will categorize a malignant image NA, SCC, AC, and SC into a normal image. In Fig.3, the accurateness of NC is greater than 0.9, which is a sufficient value in experimental practice.

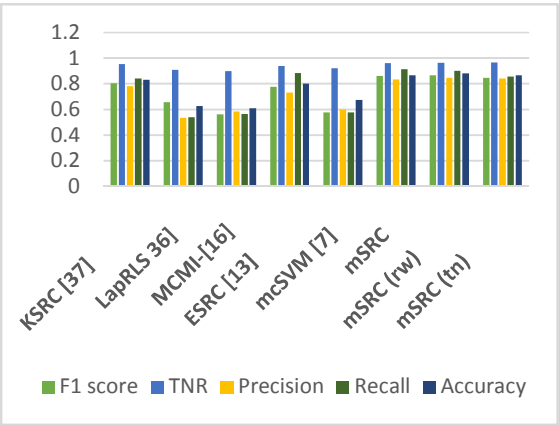

Figure.5 Comparison of Various Technique

7. CONCLUSION

The novel technique is suggested mSRC for categorizing the lung needle biopsy images. mSRCgoal is to rise the classification performance, which exactly for the images of various cancerous types. The genetic techniquegoal is to select the topmost discriminative samples for every single distinct modality as well as to assurance the huge diversity among different features. From the perception of experimental practice, misclassifying a cancerous image as a standard or normal one will be considerably more serious than misclassifying a standard or normal image as a cancerous one, Forthcoming work will examine how to implement the technique on the image set through various class relations, i.e., considering the ratio of malignant nuclei in every image. Similarly the multimodal data is broadly obtainable in medical image exploration due to the numerous data acquisition methods.

REFERENCE

[1]. Bhattacharya, V. Ljosa, J.-Y. Pan, M.R. Verardo, H. Yang, C. Faloutsos, and A.K. Singh, “Vivo: Visual Vocabulary Construction for Mining Biomedical Images,” Proc. IEEE Fifth Int’l Conf. Data Mining, Nov. 2005.

[2]. AnantMadabhushi*, Michael D. Feldman, Dimitris N. Metaxas, John Tomaszeweski, and Deborah Chute, “Automated Detection of Prostatic Adenocarcinoma From High-Resolution Ex Vivo MRI”, IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 24, NO. 12, DECEMBER 2005.

[3]. C.D. Stylios, P.P. Groumpos, Modeling complex systems using fuzzy cognitive maps, IEEE Trans. Syst., Man, Cybern. Part A Hum. Sci. 34 (1) (2004) 155–162.

[4]. Celso T. N.Suzuki, Jancarlo F. Gomes, Alexandre X. Falcao, Joao P. Papa, and Sumie Hoshino-Shimizu, “Automatic Segmentation and Classification of Human Intestinal Parasites From Microscopy Images” in IEEE Transactions on Biomedical Engineering, VOL. 60, NO. 3, MARCH 2013.

[5]. E.I. Papa Georgiou , P.P. Spyridonos, D. Th. Glotsos, C.D. Stylios, P. Ravazoula, G.N. Nikiforidis, P.P. Groumpos , “Advanced soft computing diagnosis method for tumour grading” , in Elsevier B.V. Artificial Intelligence in Medicine (2006) 36, 59—70.

[6]. E.I. Papageorgiou, C.D. Stylios, P.P. Groumpos, An integrated two-level hierarchical decision making system based on fuzzy cognitive maps,

IEEE Trans. Biomed. Eng. 50 (12) (2003) 1326–1339.

[7]. Guo, Z., Zhang, L, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," IEEE Trans. Image Process., vol. 19, no. 6, pp. 1657– 1663, Jun. 2010.

[8]. H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed Sensing and Redundant Dictionaries," to appear in IEEE Trans. Information Theory, 2007.

[9]. J. Kim, J. Choi, J. Yi, and M. Turk, "Effective Representation Using ICA for Face Recognition Robust to Local Distortion and PartialOcclusion," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1977-1981, Dec. 2005.

[10]. Joel A. Tropp, Student Member, IEEE, "Greed is good: Algorithmic Results for Sparse Approximation ", IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 50, NO. 10, OCTOBER 2004.

[11]. John Wright, Student Member, IEEE, Allen Y. Yang, Member, IEEE, Arvind Ganesh, Student Member, IEEE, S. Shankar Sastry, Fellow, IEEE, and Yi Ma, Senior Member, IEEE, "Robust Face Recognition via Sparse Representation", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 31, NO. 2, FEBRUARY 2009.

[12]. K. Arthi, J. Vijayaraghavan, "Content Based Image Retrieval Algorithm Using Colour Models" in International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2013.

[13]. KimmiVerma, AruMehrotra, Vijayeta Pandey, Shardendu Singh, "Image processing techniques for the enhancement of brain tumor patterns", in International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering ,Vol. 2, Issue 4, April 2013.

[14]. Miao Y, Liu Z. "On causal inference in fuzzy cognitive maps". IEEE Trans Fuzz Syst 2000; 8:107—19.

[15]. Muhammad MoazamFraz∗, Paolo Remagnino, Andreas Hoppe, BunyaritUyyanonvara, Alicja R. Rudnicka, Christopher G. Owen, and Sarah A. Barman , "An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 59, NO. 9, SEPTEMBER 2012.

[16]. Prof. Vikas Gupta, Kaustubh S. Sagale, "Implementation of Classification System for Brain Cancer Using Backpropagation Network and MRI" INTERNATIONAL CONFERENCE ON ENGINEERING, NUiCONE-2012, 06-8DECEMBER, 2012.

[17]. S. Santini and R. Jain, "Similarity Measures," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 9, pp. 871-883, Sept.1999.

PREMKUMAR.S.T, obtained his B.E. degree in Computer Science and Engineering from Anna University, 2011, Coimbatore, India and pursuing his M.E degree in Computer Science and Engineering at Anna University, Chennai, India. He has published 7 research articles in various conference proceedings. His research area of interest is Image processing and specialization on classification for lung cancer from lung needle biopsy images.
E-mail: premkumarsmt@gmail.com

M. SANGEETHA obtained her BE degree in Computer Science and Engineering from Anna University, Coimbatore, India. She obtained ME in computer science and engineering from Sri Krishna College of Engineering and Technology, Coimbatore, India. Her research interest focus on Data mining and image processing.
E-mail: sangee_ganapathy@gmail.com