Image Automatic Categorisation using Selected Features Attained from Integrated Non-Subsampled Contourlet with Multiphase Level Sets

U. Rajyalakshmi[#,*], S. Koteswara Rao[$], and K. Satya Prasad[!]

[#]*Aditya Engineering College, Surampalem, Andhra Pradesh* - *533 437, India*
[$]*KL University, Vijayawada, Vaddeswaram, Andhra Pradesh* - *522 502, India*
[!]*Vignan's Foundation for Science, Technology Research, Guntur, Andhra Pradesh* - *522213, India*
[*]*E-mail: rajyalakshmiuppada@gmail.com*

ABSTRACT

A framework of automatic detection and categorisation of breast cancer (BC) biopsy images utilising significant interpretable features is initially considered. Appropriate efficient techniques are engaged in layout steps of the discussed framework. Different steps include : (a) To emphasise the edge particulars of tissue structure; the distinguished non-subsampled contourlet (NSC) transform is implemented. (b) For the demarcation of cells from background, k-means, adaptive size marker controlled watershed, two proposed integrated methodologies discussed. Proposed Method-II, an integrated approach of NSC and multiphase level sets is preferred to other segmentation practices as it proves better performance. (c) In feature extraction phase, extracted 13 shape morphology, 33 textural (includes 6 histograms, 22 Haralick's, 3 Tamura's, 2 Graylevel Run-Length Matrix) and 2 intensity features from partitioned tissue images for 96 trained images. Lastly, K nearest neighbourhood and multi-class support- vector machine are implemented for categorisation of images into normal and cancerous categories using six key features. The results of methodology were tested for 24 image and analysed with pathologist results. It's analysed that proposed method-II achieved better classifier accuracy over literature techniques.

Keywords: Contour let transform; Adaptive marker controlled watershed approach; Multiphase level sets; MC-SVM classification; Biomedical and defence applications

1. INTRODUCTION

Breast cancer (BC) identification and diagnosis has for all time been a foremost concern for the pathologists and even for medical practitioners. 32 per cent of Indian inhabitants get cancer at some time in their life[1]. For Precise detection of BC, experts and medical practitioners prefer microscopic biopsy images collected under the microscope. In histopathology, BC biopsy images will be characterised into cancerous one or normal one[2,3]. Highly (40 x/100 x) magnified biopsy image provides consistent information about abnormal and normal tissues. Later the segmentation and categorisation application can be continued with other defence and military applications where we considered IRS satellite images for segmentation of required areas.

Plissiti[4], *et al.* proposed color gradient watershed transform using 90 pap-stained cervical images of resolution 1536 × 2048 pixels and obtained 6 shape, 8 texture and 3 intensity features. They utilised maximum-relevance with minimum-redundancy (MR-MR) criterion for feature selection. They handled cell level diagnosis using image Processing methodologies. Bergmeir[6], *et al.* presented a model for obtaining the local histograms and GLCM texture features. Huang and Lai[7] explained a

methodology for segmentation and categorisation methods for histology images basing on texture features and with help of SVM the highest Categorisation accuracy obtained is 92.8 per cent.

Adem Kalinli[8], *et al.* considered otsu thresholding approach with the classifiers namely *k*-nearest neighbours, radial basis neural networks, support vector machines and *k*-means clustering, naive bayes and functional trees for object extraction followed by classification. Kasmin[10], *et al.* obtained the features of BC tissue images possessing area, perimeter, solidity, convex area, orientation filled area, major axis length, ratio of cell and nucleus area, eccentricity, mean intensity of cytoplasm, and circularity. The efficacy of other classifiers such as SVM, random forest, and fuzzy *k*-means is also examined. Proposed work uses Ductal Carcinoma (DC) BC images, as > 80 per cent of BC is because of ducts. Also 40 X magnified Hematoxylin & Eosin (HE) DC images were chosen for clear cell segmentation.

Ali[5], *et al.* considered Active contour models using multiple level sets for segmentation of 14 BC histology images with resolution 512 × 512 pixels and generated shape features to obtain an accuracy more than 90 per cent. Fatakdawala[9], *et al.* proposed expectation maximisation driven geodesic active contour (EMaGAC) without and with overlap resolution using 100 breast histology images with resolution of 200 × 200

pixels and generated texture features to attain sensitivity of 86 per cent, Positive-Predicted Value of 64 per cent and overlap resolution of 90 per cent .

Kanchanamani[16], *et al.* discussed the various classifiers namely Support Vector Machine (SVM), naive bayes, k-nearest neighbour, multi-layer perceptron, and linear discriminant analysis and obtained best classification Sensitivity, Specificity and Accuracy values for SVM as 89.2 per cent, 96.4 per cent, and 92.5 per cent. Mouelhi[13], *et al.*, defined modified geometric active contour model and touching nuclei method for nuclei detection. In the paper, they extracted colour and shape features and obtained segmentation accuracy of 97 per cent.

George[11], *et al.* defined marker-controlled watersheds transform using 92 breast cytological images of 640×480 pixels and extracted 12 statistical, 10 textural and 2 intensity features to evaluate classifier sensitivity and specificity as 95.49 per cent and 83.16 per cent, respectively.

Till now, less research is engaged on segmentation of noise corrupted, less contrast BC histopathological microscopic biopsy images. Those images lead to poor and erroneous segmentation of local image intensity non-homogeneities, and ultimately failing to extract true edges. Without the optimal segmentation, the processed result cannot be further utilised and preferred for classification. This paper proposes two new segmentation methodologies for attaining optimum segmentation with highest accuracy.

2. METHODOLOGY

Automatic detection, classification of BC biopsy images from microscope is quite challenging as the images retain clusters and overlapping cells. Distinctive stages involved in methodology include the tissue image enhancement using 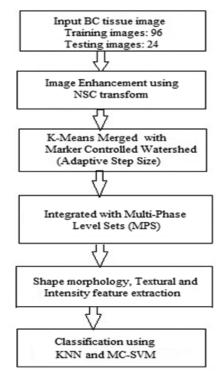non-subsampled contourlet (NSC) transform for detection of edge levels are presented. Proposed methodology is presented in Fig.1. The details of stages utilised in methodology are discussed as follows.

2.1 Image enhancement

Preprocessing necessity is to reduce noise, improve processing speed, for contrast (quality) improvement of area of interest. Biopsy images gathered from light microscope may lack in some respect like uneven staining and poor contrast, so enhancement improves the contrast between the foreground (area of interest) and background. Circular Hough (CH) transform described by Hrebien[23], *et al.* is preferred for detection of edges its result is presented in Fig.3(b). For image enhancement initially contrast-limited adaptive histogram (CLAH) equalisation[21] is used. CLAH equalised output is presented in Fig. 3(c).
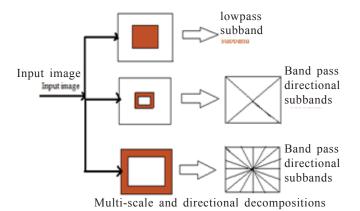

Figure 2. NSC transform filter bank structure.

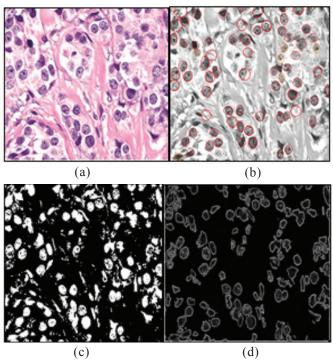**Figure 1. Distinctive stages involved in methodology.**

Figure 3. (a) BC biopsy image, (b) CH cell detected output, (c) CLAH enhanced image, and (d) edge detected image using NSC transform.

2.1.1 Non Subsampled Contourlet Transform

Wavelets became efficient alternative to Fourier methodologies in microscopic image applications because of differing window size (wide for slow and narrow for high frequencies i.e., optimised time, frequency resolution). Resultants are ¼ th size of original tissue. Smoothed image can further decompose into sublevels for further iterations. Wavelet transform being 1-D, smoothness of curves in tissue image is limited to 1-D, which leads to implementation of contourlets (a 2-D transform). Contour let is non-shift invariant because it utilises up and down samplings in Directional Filter Bank (DFB) and Laplacian Pyramid (LP) construction. It leads to NSC transform implementation[12]. Its filter bank structure presented in Fig. 2.

Contrast with contour let, NSC transform possesses (a) non-subsampled LP obtained by non-subsampled 2-channel 2-D filter banks and (b) non-subsampled DFB obtained by switching of down/up samplers in 2-channel DFB structure while filter up sampling accordingly. Non-subsampled LP (NSLP) preserves multiscale feature and Non-Subsampled DFB (NSDFB) possess directionality. Let the edge outcomes attained by NSC transform using NSLP and NSDFB be $e(p, q)$. The outcomes are better than the DWT based method as presented in Fig. 3(d). NSC decomposes BC image into approximation sub-band (that preserves the content of input image) and detail sub-bands (that stores intensity changes in all directions). The energy detail coefficients from all decomposed detail sub-band levels is formulated as follows.

$$e(p,q) = \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{J=1}^{N} \left| W_L(p,q) \right|^2 \qquad (1)$$

And/or its standard deviation (SD) from NSC decomposed BC tissue image on every directional sub-band is formulated as

$$\sigma_L(p,q) = \sqrt{\frac{1}{N \times N} \sum_{i=1,J=1}^{N} \sum \left| W_L(p,q) - \mu_L \right|^2} \qquad (2)$$

Where μ_L is the mean of Lth sub-band defined as

$$\mu_L = \frac{1}{N \times N} \sum_{i=1}^{N} \sum_{J=1}^{N} |W_L(p,q)|,$$

where p, q are the available total rows, columns of decomposed tissue image respectively, W_L is the Lth coefficient of the NSC decomposed sub-band, and LxL is the NSC decomposed sub-band. It results with energy and SD feature vectors as $f_{je} = [E_1, E_2, \ldots E_n]$ and/or $f_{j\sigma} = [\sigma_1, \sigma_2, \ldots \sigma_n]$ respectively, where 'n' defines the number of obtained directional sub-bands. Pixel energy defines its information regarding the intensity variation. NSC energies of every pixel in BC image are evaluated (considering all direction average values) to construct the energy map, which gives the distinguished energy variations contained in it.

2.2 Cell Segmentation

2.2.1 K-Means Partitioning

Pre-processing is required for separation of nuclei from its false positives. K-means[22] clustering is used for splitting the nuclei from false positives i.e. to divide N observations $(i_1, i_2, i_3 \ldots i_n)$ of image I into K areas (K ≤ N), $(A_1, A_2, A_3 \ldots A_n)$ for all A_i, and mean m_i as

$$KM = \arg\min [\sum_{j=1I}^{K} \sum_{A_j} | \ | I - m_j | \ |^2] \qquad (3)$$

The resultant of k-means is as presented in Fig. 4(b).

2.2.2 Proposed Model I (PM-I)

Reliable unsupervised Watersheds concept in image processing depends on assumption of an image I as topographic 3-D space, with intensity (elevation information) versus two spatial coordinates. To get watershed lines of I with local minima $L_i = (L_1, L_2, L_3 \ldots L_M, i=1, \ldots M)$ flooding process is initiated from $n_{min} = I_{min} + 1$ to $n_{max} = 1_{max} + 1$, where I_{min}, I_{max} are the respective lower, higher image intensity values and $c_b(L_i)$ be the points within catchment basin due to local minimum L_i. During flooding process, the union of catchment basins at n^{th} stage, $c_b(n)$, can be related as $c_b(I_{min} + 1) = T(I_{min} + 1)$, Where $T(n) = \{x \ I \ I(x) < n\}$ is the image points set with intensity less than n and its set of connected intensity components be R. For each component connected r ε R, the intersection k with $c_b(n-1)$ is evaluated as $k = r \cap c_b(n-1)$. Further $c_b(n)$ set is derived sequentially using $c_b(n-1)$ based on possibilities of k as:

- If k is vacant or if k possess one component of $c_b(n-1)$, then k belongs to an existing basin of local minimum i.e. new minimum component r is added to $c_b(n-1)$, then

$$C_b(n) = C_b(n-1) U \ r. \qquad (4)$$

- If k possess higher than one component of $c_b(n-1)$, then r belongs partially to distinct basins and the level in these basins will unite. Then watershed line is built to eradicate the overflow between distinct basins.

Application of traditional watershed using image gradient produces over segmentation The image gradient $g(p, q)$ is obtained using gray scale dilation and erosion, with structuring element (SE) s. So, Marker Controlled Watershed (MCW) method is introduced. Xiaopeng[14], *et al.* discussed that due to low SE size, opening, and closing morphology will detach bright, dark details and large SE eradicate large contours. So in PM-I altering SE size is constructed for every pixel in the tissue image $I(p, q)$ which is defined as Adaptive Marker Controlled Watershed (AMCW) method. Initially SE map, $M_s(p, q)$ is evaluated using the weighted variance, and weighting coefficient discussed by Rajyalakshmi[20], *et al.* Then all pels in $I(p, q)$ are altered using $M_s(p, q)$ with opening $(\circ)$ and closing $(\bullet)$ operations as

$$M_1(p,q) = [g(p,q) \circ M_s(p,q)] \bullet M_s(p,q) \qquad (5)$$

In this PM-I, the edge details attained by NSC transform $e(p,q)$ are combined with $M_1(p,q)$ using fusion technique and the observed results are as depicted in Fig.4(c).
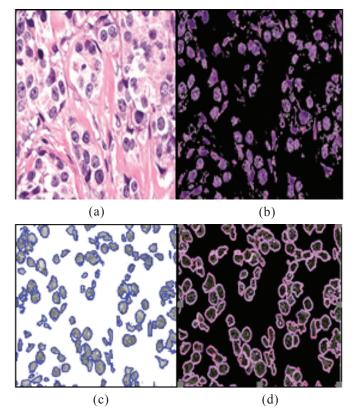
(a)   (b)

(c)   (d)

Figure 4. (a) BC biopsy image, (b) K-means result, (c) PM-I integrated result, (d) PM-II integrated result.

2.2.3 Proposed Model II (PM-II)

In this PM-II the edge extractions $e(p,q)$ of NSC transform are utilised as markers for Multiphase Level Sets (MLS) energy function shown in following equations to eliminate the local inhomogeneities[19]. NSC provides a combined result of NSLP and NSDFB. With window function and cluster mean, F(h-x) (non-negative and is 0, for $x \notin M_S$ and $\tilde{m}_j \approx \tilde{b}(h)c_j$ respectively, local clustering ε_h criteria is attained as,

- Image Model Formulation

H and E stained images captured from the camera (via microscope) are affected with intensity in homogeneities, so an observed image model *I* can be perceived as $I = \tilde{b}\tilde{R}+\eta$. Where R be the real image on a spatial domain Ω, takes approximately N distinct constants $c_1, c_2 \ldots c_N$ (with $\tilde{c} = \bigcup_{j=1}^{N} \tilde{c}_j$) partitioned into N disjoint regions $\Omega_1, \Omega_2, \ldots, \Omega_N$ with $\tilde{\Omega} = \cup_{j=1}^{N} \tilde{\Omega}_j$ and $\tilde{\Omega}_i \cap \tilde{\Omega}_j = 0 \forall j \neq k$. The component $\tilde{b}$ is defined as shading image (bias field). The tilde of $\tilde{R}$ and $\tilde{b}$ is to distinguish from its estimates R and b respectively.

-
- Formulation of Local Clustering Image-Function

As square neighborhood possesses unequal distance between corner or adjacent cells to analysis cell, it introduces directional bias that results in inaccurate segmentation. Also for high-resolution images i.e. as neighborhood to cell size ratio increases, circular neighborhood result is more efficient and accurate. So, we considered circular neighborhood o$_q$ each point $q \in \tilde{\Omega}$ with a radius p defined as $O_q \triangleq p : |p-q| \leq \rho$. The partition of the entire spatial domain $\tilde{\Omega}$ leads to a partition $\left\{ o_q \cap \tilde{\Omega}_j \right\}_{j=1}^{N}$. The values $\tilde{b}(p) \forall p$ in o$_q$ are close to $\tilde{b}(q)$, for slow varying $\tilde{b}$. Therefore,

$$I(p) \approx \tilde{b}(q)\tilde{c}_j + \eta(p) \, for \, p \, \int O_q \cap \tilde{\Omega}_j \qquad (6)$$

So, the intensities $I^j_q = \left\{ I(p) : x \int O_q \cap \tilde{\Omega}_j \right\}$ form a cluster set with centre or cluster mean $m_j \approx \tilde{b}(q)c_j, j = 1, 2 \ldots N$. Considering above equation (3) and (6), using o$_q$ for the intensities I(p) and window function (which is non- negative) F_G(q-p), local clustering criteria $\mathring{a}_q$ can be written as,

$$\varepsilon_q = \sum_{j=1}^{N} \int_{\tilde{\Omega}_j \cap O_q}^{N} F_G(q-p) |\, I(p) - \tilde{b}(q)\tilde{c}_j|^2 u_h(p) dp \qquad (7)$$

Where, u_q (p) is the membership function, u_q(p)=1, for $p \varepsilon \tilde{\Omega}_j$ else zero. Where F_G(q-p) is a truncated Gaussian function, which is stated as $F_G(v) = 1/a \left[e^{\left(-|V|^2/2\sigma^2 \right)} \right]$ for $|v| \leq p$, the radius of neighborhood.

- Integration of NSC edge details with LCIF

NSC edge detected results are combined with MLS results, to obtain nuclei boundaries which removes complete intensity in homogeneities presented in Fig. 4(d). NSC solves the problem of coarser details and also it achieves all fine details of the image. So, for cell extraction in H & E images the NSC edge details $e(p,q)$ obtained from equation (1) is integrated with LCIF to overcome local in homogeneities problem. And the initial front for MLS is provided by NSC transform. Energy formulated by integrating e(p,q) with MLS-LCIF is

$$\varepsilon_q = \sum_{j=1}^{N} \int_{\tilde{\Omega}_j \cap O_q}^{N} F_G(q-p) |\, e(p,q) - \tilde{b}(q)\tilde{c}_j|^2 u_h(p) dp \qquad (8)$$

The above equation (8), defines the integrated LCIF due to the response attained from NSC transform. Now the integrated LCIF is applied for the whole image to extract global function. Where, u_q(p) is the membership function, u_q(p)=1, for $p \varepsilon \tilde{\Omega}j$ else zero. $F_G(q-p) = 0, \, for \, p \notin O_q$.

- Formulation of Global Clustering Image-Function (GCIF)

A real image on domain Ω, chooses N distinct constants $\tilde{c}_1, \tilde{c}_2 \ldots \tilde{c}_N$ to split to disjoint N regions $\tilde{\Omega}_1, \tilde{\Omega}_2, \ldots, \tilde{\Omega}_N$ and $M_j^F(\phi)$, as membership function, that relies on number of LS functions initiated. For energy minimisation, it is formulated by representing the regions $\tilde{\Omega}_1, \tilde{\Omega}_2, \ldots, \tilde{\Omega}_N$ with LS functions.

For N=2, let ϕ be the LS function used to represent $\tilde{\Omega}$ as $\tilde{\Omega}_1 = p$ for $\phi(p) > 0$, and $\Omega_2 = p$, for $\phi(p) < 0$. To reduce ε_q in $\tilde{\Omega} \forall q$, ε_q integral must be minimised with respect to q over the image domain $\tilde{\Omega}$. In PM-II formulation, total energy ε using dependent parameters $\left(\Phi, \tilde{b}, \tilde{c}\right)$ is obtained as,

$$\varepsilon\left(\Phi, \tilde{b}, \tilde{c}\right) = \int_{\Omega}\left(\sum_{j=1}^{N}\int_{\Omega_j} F_G(q-p)|\ e(p,q) - \tilde{b}(q)\tilde{c}_j|^2 dq\right) M^F{}_j\left(\phi(p)\right)dp \quad (9)$$

Replacing,

$$\varepsilon\left(\Phi, \tilde{b}, \tilde{c}\right) = \int_{\Omega}\left(\sum_{j=1}^{N} e_j(p)\right) M^F{}_j\left(\phi(p)\right)dp \quad (10)$$

where $e_j(p) = \int F_G(q-p)|\ e(p) - \tilde{b}(q)\tilde{c}_j|^2 dq$ and

$M^F{}_j(\phi) = M^F{}_j{}^{\Omega}\left(\phi_1(q), ... \phi_f(q)\right)$

$M^F{}_j(\Phi)$ is the membership function described as $M^F{}_1(\Phi) = H^F(\Phi)$ and $M^F{}_2(\Phi) = 1 - H^F(\Phi)$ with Heaviside smoothed function, $H^F(\Phi) = \frac{1}{2}\left[1 + \frac{2}{\pi}\arctan(\phi)\right]$. For N>2, two or more LS functions are used to represent N disjoint regions $\tilde{\Omega}_1, \tilde{\Omega}_2, ..., \tilde{\Omega}_N$. When N=3, two LS functions ϕ_1 and ϕ_2 are defined using $M^F{}_1(\phi_1, \phi_2) = H^F(\phi_1)H^F(\phi_2)$, $M^F{}_2(\phi_1, \phi_2) = H^F(\phi_1)(1 - H^F(\phi_2))$, and $M^F{}_3(\phi_1, \phi_2) = (1 - H^F(\phi_1))H^F(\phi_2)$. Let LS functions be represented by a vector function $\Phi = (\phi_1, \phi_2, ... \phi_f)$.

The energy functional E_F,

$$\acute{E}^F(\Phi, \tilde{b}, \tilde{c}) = \varepsilon\left(\Phi, \tilde{}, \tilde{c}\right) + R(\Phi) \quad (11)$$

Where energy regularisation term $R(\Phi)$, defined as

$R(\Phi) = \sum_{k=1}^{f} R(\phi_k)$, for LS function $\Phi = (\phi_1, \phi_2, ... \phi_f)$. Considering regularisation $R(\Phi) = \int p(|\nabla\phi|)$ term into account, the term $E^F(\Phi, \tilde{b}, \tilde{c})$ in equation (11) can be minimised using the principle discussed in Rajyalakshmi[24], *et al.* The integrated response of NSC edge details with MLS extracts clearly the nuclei shown in Fig. 4(d). Performance evaluations for traditional K-means, AMCW, MLS, PM-I, PM-II were presented in Table.1.

2.3 Feature Extraction

Mouelhi[13], *et al.* extracted Haralick's, oriented gradients histogram, 4 colour based moment features to categorise cancerous One over BC biopsy images. A number of

Table 1. Segmentation approaches performance evaluation

Segmentation algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
K-means[22]	76.534	71.33	76.59
MCW[14]	82.678	88.63	81.92
MLS[19]	87.689	91.79	93.89
PM-I	93.741	96.27	95.79
PM-II	98.741	97.27	98.79

comprehensive discussions were reviewed in the survey[7,10,15,17] for segmentation and classification. Tables 2 (a) and 2 (b) show the features extracted in the proposed model.

These features set act as input for categorisation using supervised classifiers.

2.4 Supervised Classification using Features Set

Using the segmented features attained from PM-II, the testing data set is classified using K-nearest neighbour (KNN) classifier studied in Altman[17] and multi-class support vector machine (MC-SVM) classifiers studied in Hsu[18], *et al.* The entire classification is done using the shape, intensity and texture features elevated from 96 trained images. The response of the ROC plot for KNN and M-SVM are compared and proved that M-SVM provides better classifier accuracy over KNN.

2.4.1 KNN Classifier

KNN classifies images based on closest trained samples in feature space. Classification also relies on highest vote of neighboring points. Problem occurs while selecting the neighbors, which are closest to each sample. Shortest distance is chosen for classifying BC tissue images. Euclidean distance in-between pixel (p_1, q_1) and pixel (p_2, q_2) is

$$D(x, y) = \sqrt{(p_1 - p_2)^2 + (q_1 - q_2)^2} \quad (12)$$

2.4.2 MC-SVM Classifier

Let f_j be feature set, w be vector splitting hyperplane, then objective optimised function to be minimised for MC-SVM is:

$$O'_F = \min[\frac{1}{2}|\ w|^2 + C\left(\sum_{j=1}^{N}\zeta_j\right)^p] \quad (13)$$

Where C is the user defined parameter, and

$0 \le \zeta_j \le 1 - y_j\left(f_j.w + b\right)y_j$ label of f_j, {-1, 1},
Hyperplane normal splitting vector w is evaluated as

$$w = \sum_{j=1}^{N} c_j y_j f_j \quad (14)$$

3. RESULTS AND DISCUSSIONS

Performance evaluation of classifiers is compared using average sensitivity, classifier accuracy, and average specificity. Average Balanced Classifier Rate (BCR), geometric mean of average values of specificity and selectivity as well as Area below curve (AUC) of classifier response are chosen for evaluation. Overall proposed method performance is assessed

Table 2(a). Textural features extracted

Feature type	Feature type
33 textural (includes 6 histogram, 22 Haralick's, 3 Tamura's, 2 Graylevel Run-Length Matrix (GRLM)) features.	Six histogram features: 1. Mean, 2. Variance, 3. Third moment (skewness), 4.Fourth moment (kurtosis), 5. Entropy, 6.Energy.
	22 Haralick's features : Uniformity(Energy/Angular Second Moment), Dissimilarity, Entropy, Inertia, Contrast, Inverse difference, Homogeneity / Inverse difference moment, correlation, Cluster Prominence, Cluster Shade, Autocorrelation, Sum of Squares, Maximum probability, Sum Variance, Sum Entropy, Sum Average, Difference entropy, Difference variance, Information measures of correlation, Information measures of correlation, Inverse difference normalized, Maximal correlation coefficient, Inverse difference moment normalized.
	3 Tamura's features: Coarseness, Contrast, Directionality.
	2 GRLM features: Short and Long Run Emphasis

Table 2(b). Shape, intensity features extracted

Type	Feature
Shape based features {13}	Area, Perimeter, Compactness, Equivalent Diameter, Major axis length, Minor axis length, Orientation, Centroid, Nucleus to Cytoplasm ratio, Eccentricity, Convex area, Solidity, Extent.
Intensity feature{2}	$Intensity_{max}$ and $Intensity_{min}$, $Intensity_{mean}$

Table 3. Performance assessment of classifiers

Parameter	KNN[17]	MC-SVM [18]
Average sensitivity (%)	97.13	98.24
Average selectivity (%)	93.76	97.57
BCR (%)	95.43	97.90
Classifier accuracy (%)	99.01	99.76
AUC	0.9703	0.9743

Table 4. Stage wise execution time

Individual stage		Time in sec (mean±deviation)
Pre-Processing		2.35±0.21
Nuclei detection		3.26±1.56
Nuclei Segmentation	K-means[22]	8.53±0.83
	MCW[14]	10.62± 1.34
	MLS[19]	15.64±2.34
	PM-I	6.83 ± 0.76
	PM-II	4.75 ± 0.69
Feature Extraction		4.89±1.34
Tissue Classification	KNN	5.21±1.39
	MC-SVM	1.17±0.67

by traditional parameters namely

$$Average\ Sensitivity S_{EN} = \frac{T_P}{(T_P + F_N)}$$

$$Average\ Specificity S_{PE} = \frac{T_N}{(F_P + T_N)}$$

$$Segmentation\ Accuracy A_{CC} = \frac{T_P + T_N}{(T_P + T_N + F_P + F_N)}$$

Computational Time (C_T): Difference between the return time of the algorithm at start-up of the process ($t_{r,start}$) and the return time at the end of the process ($t_{r,end}$). i.e., $C_T = t_{r,start} - t_{r,end}$

$$Pr ecision P_r\ (PPV) = \frac{T_P}{(T_P + F_P)}$$

$$Dice\ Similarity\ Coefficient D_I = \frac{2(A \cap B)}{(A) + (B)}$$

$$F_1 score = \frac{2 x P_r x R_{cl}}{(P_r + R_{cl})}$$

where A and B respectively define the sets of algorithm and ground-truth segmented pixels,

True Positive T_p: The total number of correctly detected true tumor pixels. A detected pixel is considered as T_p when its Eucledian distance with ground truth pixel location is < 7.5 μm(30 pixels),

True Negative T_N: The total number of correctly identified false tumor pixels as non tumor pixels,

False Positive F_p: The total number of detected pixels which are not appeared in the ground –truth,

False negative F_N : The total number of non detected pixels that are appeared in the ground-truth.

Classifier accuracy and performance assessment of classifiers as represented in Table 3 and observed that the integrated segmentation approach (PM-II) of Contourlet and MLS with MC-SVM classifier provides high system accuracy and best classifier rate over traditional K-means partitioning, NSC, MCW, MLS, PM-I segmentation methodologies. The execution time for different steps performed in the discussed model was as represented in Table 4. On part of execution time also PM-II provides better evaluation when merged with MC-SVM approach.

The proposed method is compared on basis of subjective performance measures using the literature techniques as described in Table 5. Using PM-II, we obtained better results than traditional approaches. The results were also compared to ground-truth results provided by the pathologist.

Table 5. Quantitative state-of-art evaluation of proposed method using the literature techniques

Reference	Number of images	Resolution (pixels)	Segmentation algorithms used	Features extracted	Quantitative performance measures
Ali[5], *et al.*	100 histology	512 × 512	Active contour models using multiple level sets.	Shape	Hausdorff distance (HD) 5 pixels over 95 %, Mean Absolute Distance (MAD) A_{CC} >90 %
Fatakdawala[9], *et al.*	100 histology	200 × 200	Expectation maximisation driven geodesic active contour (EMaGAC) without and with overlap resolution	Texture	S_{EN} : 86 %, PPV : 64 %, HD : 2.1, MAD : 0.9, Overlap resolution : 90 %
Plissiti[4], *et al.*	90 images	1536 × 2048	Color gradient watershed transform	6 shape, 8 texture, and 3 Intensity	HD: 1.71 (mean) ± 0.54 (std).
Mouelhi[13], *et al.*	18 images with 3-amino-9-ethylcarbazole chromogen and 42 images 3_diaminobenzidine chromogen	2048 × 1360	Modified geometric active contour model and touching nuclei method	Color and shape	A_{CC} : 97 %
George[11], *et al.*	92 breast cytological images.	640 × 480	Marker-controlled watersheds transform.	12 statistical, 10 textural and 2 intensity	S_{EN} : 95.49 S_{PE} : 83.16
Proposed method	96 training and 24 testing images	760 × 570	NSC integrated with MLS.	13 shape 33 textural and 2 intensity	S_{EN} : 98.77 % S_{PE} : 97.81 % A_{CC} : 99.02 % C_T : 16.42±4.47 s PPV = 89.74 D_I = 0.95 F_1 = 0.91

4. FUTURE SCOPE

Proposed Model-II is applied to electron microscopic (biopsy) images. PM-II proves as the best combination over the discussed traditional techniques. This module can also preferred to all military applications, like remote sensing imagery, where initially need to acquire it and then understand the imagery, followed by segmentation of required area, and at last categorisation. Also applicable to 'Autonomous vehicle', an image processing advanced defence application. This vehicle possesses a computer vision module that grasps the 3D scenic images.

5. CONCLUSIONS

Automated characterisation approach was presented for BC recognition from microscopic tissue images with HE staining using clinically interpretable feature set. Optimised investigation on cell nuclei depends on enhancement and segmentation methodologies used. For image efficacy, CLAH equalisation method is chosen and biopsy image is enhanced using NSC transform for good recovery of edges. For nuclei segmentation, K-means, PM-I and PM-II techniques are utilised. AMCW retains all image details as SE size is adaptive. PM-I integrates AMCW method with NSC transform

methodology; PM-II integrates NSC result with MLS. Among all categories, anticipated segmentation method performs better area extracted result than traditional MCW, region growing, MLS, PM-I approaches for nuclei extraction. After tissue image segmentation, obtained 49 features set for 96 training tissue images. Out of respective features, we choose 11 features for characterisation. Diverse sorting methods used were kNN, and MC-SVM classifiers. Local inhomogeneities problem is also solved using NSC integrated with AMCW transform. There is hike in performance measures i.e., average specificity, accuracy, sensitivity, BCR for MC-SVM to 97.57, 99.76, 98.24, 97.90 respectively over KNN classifier. Also there is an enhancement of AUC for MC-SVM to 0.9703 compared to KNN. The segmentation and classification results were tested for 24 images and compared with the manual results taken from the pathologist. The Average sensitivity, average specificity, PPV, average accuracy of all BC histopathological images using the proposed method were obtained as 98.77 per cent, 97.81 per cent, 87.94 per cent and 99.02 per cent, respectively. The computational time required for whole process including pre-processing, detection, nuclei segmentation, and nuclei classification is 16.42 ± 4.47 s. Higher the F1-score and dice index better is the classification, those are obtained as 0.91

and 0.95, respectively. The PM-II provides better subjective evaluation compared to the literature techniques discussed.

REFERENCES

1. Ali, I.; Wani, W. A., & Saleem, K. Cancer scenario in India with future perspectives. *Cancer Ther.*, 2011, **8**, 56–70.

2. Tabesh, A.; Teverovskiy, M.; Pang, H.Y.; Vinay Kumar, P.; David Verbel; Kotsianti, Angeliki, & Saidi, Olivier. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Trans. Med. Imaging*, 2007, **26**(10), 1366–1378.
    doi: 10.1109/TMI.2007.898536

3. Madabhushi, A. Digital pathology image analysis: opportunities and challenges. *Imaging Med.*, 2009, **1**(1), 7–10.
    doi: 10.2217/IIM.09.9

4. Marina Plissiti, E.; Christophoros, N. & Antonia, C. Combining Shape, Texture and intensity features for cell nuclei extraction in pap smear images. *Pattern Recognit. Lett.*, 2011, **32**(6), 838-853.
    doi: 10.1016/j.patrec.2011.01.008

5. Sahirzeeshan, A. & Madabhushi, A. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Trans. Med. Imaging.*, 2012, **31**(7), 1448-1460.
    doi: 10.1109/TMI.2012.2190089

6. Bergmeir, C.; Silvente, M.G. & and Ben´itez, J.M. Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework. *Comput. Methods Programs. Biomed.*, 2012, **107**(3), 497–512.
    doi: 10.1016/j.cmpb.2011.09.017

7. Huang P.W. & Lai, Y.H. Effective segmentation and classification for HCC biopsy images. *Pattern Recognit.*, 2010, **43**(4), 1550–1563.
    doi: 10.1016/j.patcog.2009.10.014

8. Kalinli, A.; Sarikoc, F.; Akgun, H. & Ozturk, F. Performance comparison of machine learning methods for Prognosis of hormone receptor status in breast cancer tissue samples. *Comput. Methods Programs. Biomed.*, 2013, **110**, 298–307.
    doi: 10.1016/j.cmpb.2012.12.005

9. Fatakdawala, H.; Xu, J.; Ajay, B.; Gyan, B.; Shridar, G.; Michael, F.; John Tomaszewski, E. & Madabhushi, A. Expectation–Maximisation-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopathology. *IEEE Trans. Biomed. Eng.*, 2010, **57**(7), 1676-1689.
    doi: 10.1109/TBME.2010.2041232

10. Caicedo, J.C.; Cruz, A., & Gonzalez, F.A. Histopathology image classification using bag of features and kernel functions. *Artif. Intell. Med.*, 2009, **5651**, 126–135.
    doi: 10.1007/978-3-642-02976-9_17

11. George, Y.M.; Zayed, H.H.; Ismail Roushdy. M. & Bassant Mohamed, E. Remote computer-aided breast cancer detection and diagnosis system based on cytological images. *IEEE SYST. J.*, 2014, **8**(3), 949- 964.

doi: 10.1109/JSYST.2013.2279415

12. Cunha, L. Da.; Zhou, J. & Minh, N. Do. The non sub-sampled contourlet transform: Theory, design, and applications. *IEEE Trans. Image Process.*, 2006, **15**(10), 3089-3101.
    doi: 10.1109/TIP.2006.877507

13. Mouelhi, A.; Sayadi, M.; Fnaiech, F.; Mrad, K. & Romdhane, K.B. Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. *Biomed. Signal. Process Control.*, 2013, **8**, 421– 436.
    doi: 10.1016/j.bspc.2013.04.003

14. Wang, X.; Wan, S. & Lei, T. Brain tumor segmentation based on structuring element map modification and marker- controlled watershed transform. *J. Software*, 2014, **9**(11), 2925- 2932.
    doi: 10.4304/jsw.9.11.2925-2932

15. Haralick, R.M.; Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man. Cybern.*, 1973, **3**(6), 610–621.
    doi: 10.1109/TSMC.1973.4309314

16. Kanchanamani, M. & Varalakshmi, P. Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer. *Biomed. Res.*, 2016, **27**(3), 623-631.

17. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.*, 1992, **46**(3), 175–185.

18. Hsu, C. & Lin, C. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural. Netw.*, 2002, **13**(2), 415-425.
    doi: 10.1109/72.991427

19. Li, C.; Huang, R.; Ding, Z.; Chris Gatenby, J.; Dimitris Metaxas, N. & John Gore, C. A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. *IEEE Trans. Image Process.*, 2011, **20**(7), 2007-2016.
    doi: 10.1109/TIP.2011.2146190

20. Rajyalakshmi, U.; Koteswararao, S. & Satya Prasad, K. Supervised classification of breast cancer malignancy using integrated modified marker controlled watershed approach. IEEE Xplore, 7th IEEE International Advanced Computing Conference (IACC-2017), E-ISSN: 2473-3571.
    doi: 10.1109/ IACC. 2017.0125.

21. Etta Pisano, D.; Zong, S.; Bradley Hemminger, M.; DeLuca, M.; Eugene Johnston, R.; Muller, K,; Patricia Braeuning, M. & Stephen Pizer, M. Contrast limited adaptive histogram equalisation image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging*. 1998, **11**(4), 193–200.
    doi: 10.1007/BF03178082

22. Kanungo, T.; David Mount, M.; Nathan Netanyahu, S.; Christine Piatko, D.; Silverman, R. & Angela Wu, Y. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 2002, **24**(7), 881-892.
    doi: 10.1109/TPAMI.2002.1017616

23. Hrebien, M.; Korbicz, J. & Obuchowicz, A. Hough transform, search strategy and watershed algorithm in segmentation of cytological images. *Computer Recognition Systems: Springer-Verlag Berlin Heidelberg 2*, 2007, 550–557.

24. Rajyalakshmi, U.; Koteswararao, S. & Satya Prasad, K. Integrated novel multi-phase level sets with modified marker controlled watershed for segmentation of breast cancer histopathological images. *J. Adv. Res. Dyn. Control Syst.,* 2017, **9**(3), 47-55.

ACKNOWLEDGEMENTS

CONTRIBUTORS

Ms Rajyalakshmi Uppada is working as a Professor in Aditya Engineering College, Surampalem, Kakinada. She completed her Ph.D from JNTU Kakinada. She has published 12 journal paper and presented 6 international conference papers.
In the current study, she is responsible for the proposal of the technique, software implementation and testing the results.

Dr Koteswararao Sanagapallela is presently working as a Professor in KL University, Vijayawada, India. He has retired as a Scientist 'G', Associate Director, DRDO-NSTL. He has published 30 IEEE paper. With his thirty two year of design and development experience and expertise in the Anti Submarine warfare (ASW) Fire Control Systems for torpedoes and rocket launchers, he have created a strong edifice in the weapon technology in the country at NSTL.
In the current study, he has provided guidance for implementing the proposed technique.

Dr Satyaprasad Kodati is presently working as Sr. Professor in Electronics & Communication department & Rector, Vignan's Foundation for Science, Technology & Research, Vadlamudi, Guntur. He retired as a Professor in Electronics & Communication department and Director-IST, JNTU Kakinada, India. He has held different positions as Head of ECE Department, vice-principal, principal of UCEK, JNTUK, and as Director of Evaluation and Rector at JNTUK, Kakinada. He has published more than 70 Journal Publications and he has presented more than 50 conference proceeding.
In the current study, he has provided guidance for implementing the proposed technique.