

## Integrated Variable Marker Controlled Watershed Method with Level Sets for Semisupervised Classification

Rajyalakshmi. U<sup>\*1</sup>, Koteswara Rao S<sup>2</sup> and Satya Prasad K<sup>1</sup>

<sup>1</sup>ECE Dept, UCEK, JNTU Kakinada, Kakinada, India

<sup>2</sup>ECE Dept, KL University, Vijayawada, India

rajyalakshmiuppada@gmail.com, dlraju@gmail.com

### Abstract

*Symptomatic awareness of the Breast Cancer (BC) in the early stage is needful for treatment and also to support the radiologists during their diagnosis. In the present module, nuclei detections of BC biopsy images stained with Hematoxylin and Eosin, are done using Hough Transform and their Segmentation with Proposed Variable - Marker Controlled Watershed Method (VMCWM). Fixed size Structuring Element (SE) wipes out the dark and bright particulars while closing and opening morphology. So using weighted variance VMCWM, an altered SE map is generated to protect all image details. Total of 24 features (10 shape, 12 texture, and 2 intensity) are generated for classification using K-Nearest-Neighbor, Decision Trees, and Multi-Class Support-Vector Machine (MC-SVM) classifiers. The provided results integrated with level sets are compared with traditional methodologies and proved to be more accurate to the bench mark results. The entire module used 90 images for testing and 20 images for training obtained from database. Also cross validation is done using Leave One Out technique for all the samples.*

**Keywords:** Breast cancer tissue images, Variable Marker Controlled Watershed, Level sets, Support vector machine

### 1.Introduction

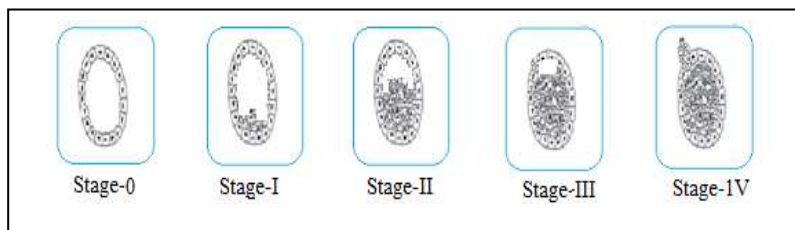
World widely, Breast Cancer (BC) is the leading (second most) fatal disease after lung cancer [1]. BC cells emerging from the milk ducts or in breast tissues can be non-invasive (remains in the duct) or invasive (that prolong to other breast tissues). For grading BC, it is essential to partition the objects of interest (namely nuclei) from stroma, back ground and cytoplasm. Morphological segmentation methods are preferred in this respect but limited due to local intensity inhomogeneities [2]. This problem has somehow solved with the initiation of Level Sets (LS) [3].

Supervised classifier methods analyses slight changes in the neighboring pixels and allots them to their groups (using their intensity and texture variations). The segmented suitable features were set to classification. This paper compares proposed method performance with different segmentation methods based on different measures like specificity, sensitivity, accuracy, and similarity indices. The tissue attained from the microscope must be stained and pre-processed as per some biopsy procedure rules reviewed in [9]. The classifier results using the segmented features were compared and retrieved that MC-SVM provides highest classifier accuracy.

## 2. Methodology

### 2.1. BC Images Used

Almost 80 % of BC that occur in women is Ductal BC [10]. Tumors due to cysts, fibro adenomas, fat necrosis, fibrocystic changes, and lactation changes are benign. For invasive malignant tumors, higher the stage (I, II, III, or IV) more will be spread to other tissues. Ductal Carcinoma (in-Situ), DCIS is a non-invasive BC which may convert to invasive BC. In stage-IV the spread extend to other lymph nodes [4]. Infiltrating ductal and lobular carcinoma, atypical papilloma and DCIS-II are malignant tumors. Figure 1 shows the stages of malignant tumors. Table 1 provides the information regarding the training and testing data.



**Figure 1. Ductal Carcinoma Cell Outlining: Stage-0(benign), Stage-I, II, III (cell spread increases with the stage), Stage-4 (Invasive)**

**Table 1. Data Description**

Images used in the work	Data	Number
BC tissue images stained with H & E	Training data	90 ( 28 normal and 62 malignant) images
	Testing data	20 images

### 2.2. Methodology

H & E (Hematoxylin and Eosin) staining makes nuclei to appear in dark blue and cytoplasm, stroma in pink color. Images with 40X magnification are collected for nucleus clear extraction [5]. Images of 10X/4X and 100X/40X are used for low and high power analysis respectively. Opening and closing by reconstruction are used to remove noise from BC tissue biopsy images (attained from microscope). Methodology of the proposed module is shown in the Figure 2. Pre-processing and nuclei detection methods are given in Section 3. Section 4 shows the nuclei segmentation techniques used. Feature extraction and classification were discussed in results section 5 followed by conclusions.

### 3. Pre-processing and cell nuclei detection

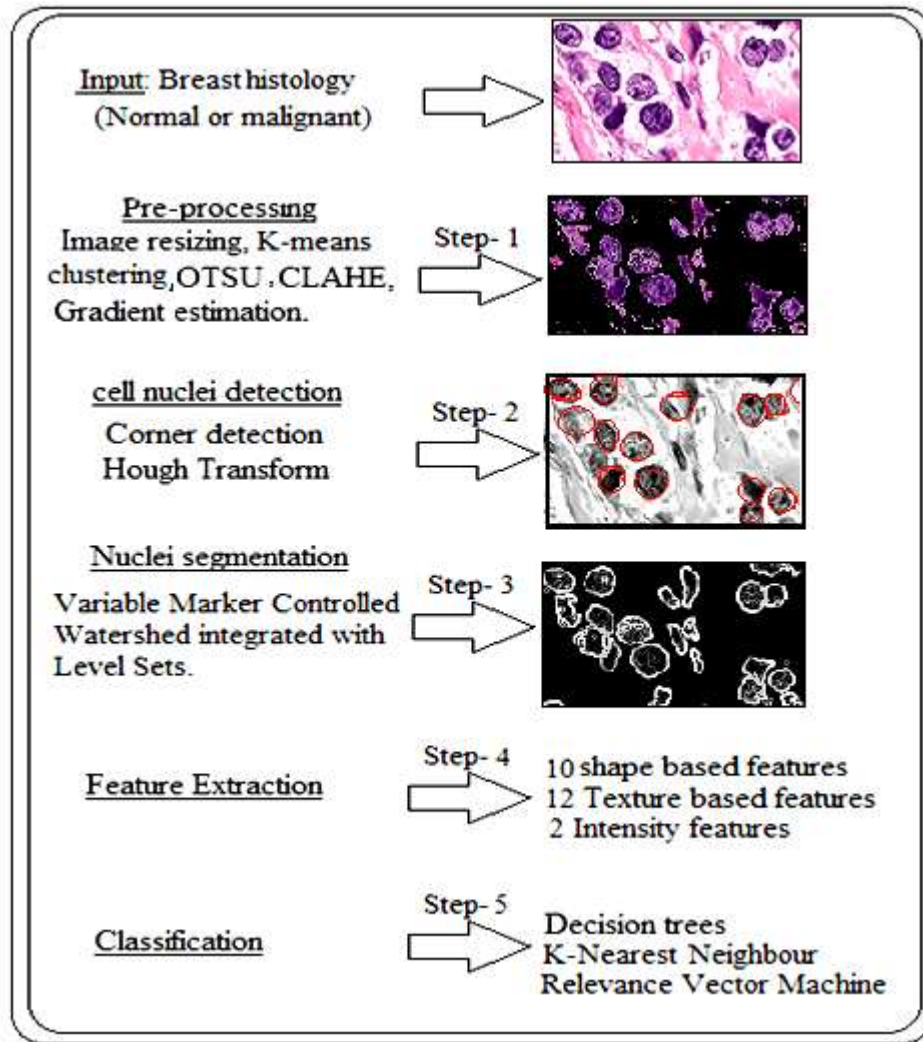
First collected images were resized to 760 x 570 pels from 2560 × 1920 pels to improve the processing speed. Then to minimize lossy compression (without direct conversion to gray scale), images were clustered with k-means technique reviewed in [6]. Image  $i$  is split into  $K$  ( $A_1, A_2 \dots A_K$ ) areas with mean  $m_j \forall A_j$  as

$$K = \arg \min \left[ \sum_{j=1}^K \sum_{i \in A_j} | | i - m_j | | ^2 \right] \quad (1)$$

Pre-processing is required to eradicate false positives like cytoplasm, background, stroma, and blood cells. Leaving luminance component (which possess image information), gray scale conversion is done by removing blue and red chrominance components. Contrast-Limited Adaptive Histogram [CLAH] Equalization upgrades the local quality [11]. Then using morphological (dilation and erosion) parameters, gradient  $g(i)$  obtained as the difference of gray scale dilation  $D_s(i)$  and erosion  $E_s(i)$  [16] is

$$g(i) = D_s(i) - E_s(i) = [i \oplus s] - [i \ominus s] \quad (2)$$

Using Structuring Element (SE)  $s$ , the dilation operator is  $\oplus$  & erosion operator is  $\ominus$ .



**Figure 2. Methodology**

Detection of cell nuclei is done using the local maximum points from Corner Detection (CD) methods [7]. Then Hough Transform using maximum points detects the nuclei locations [8]. In a discrete 2-dimensional space, CHT is stated as

$$CHT_{dispace}(S, p_0, q_0) = \sum_{p=p_0-S}^{p_0+S} \sum_{q=q_0-S}^{q_0+S} g(p, q) \delta((p-p_0)^2 + (q-q_0)^2 - S^2) \quad (3)$$

$\delta$  be the Kronecker delta,  $S$  is the circle radius.  $CHT_{dispace}$  works like accumulator by grouping the similar intensities of  $g(p, q)$  to the area (circle) present at  $(p_0, q_0)$ . OTSU

thresholding is preferred to minimize intra class and maximize inter class variances stated in [16].

## 4. Proposed Model Nuclei Segmentation

### 4.1. Region- Growing and Merging Merged with Level Sets (RGMMLS)

In RGMA, with initial assumed seed, the image is split into several regions (based on gray level intensities) or co-color regions (for RGB image). Fan *et al.*, opted seed as a homogeneous region centroid of the image [21]. Each pixel value is compared with centroid of the respective region and the variance is evaluated. That obtained variance is compared with pre-defined standard variance and pixel possessing less deviation is set to the region; else not considered. The same procedure is repeated for all pixels in the image. Let 'I' be partitioned into disjoint sub-regions  $i_1, i_2 \dots i_N$ , then it obeys the below rules:

1. ' $i_j$ ' is a linked set so that  $i_j \cap i_k = \{ \emptyset \}$ , for all  $j \neq k$ ,  $j$  and  $k = 1, 2, \dots, N$ ;
2.  $Q(i_j) = \text{true}$  for  $j = 1, 2 \dots N$ ; *i.e.*, all regions should obey the homogeneity predicate.
3.  $Q(i_j \cap i_k) = \text{false}$ , for  $j \neq k$ ; *i.e.*, pixels in separate regions must satisfy the non-homogeneity property.
4. Union of all disjoint regions generate the complete image; *i.e.*,  $I = \bigcup_{j=1}^N i_j$ .

As intensity based region method alone limits to local view and suffers with gradient problem (due to pre-defined variance). So, regions belonging to similar gray levels are merged *i.e.*,  $Q(i_j \cup i_k) = \text{true}$ , for all  $j$  and  $k = 1, 2 \dots N$ ; where  $Q$  denotes the pixel intensities. Even the obtained regions are selected as contour inputs to LS, the approach does not provide proper extraction of nuclei and possess local inhomogeneities problem.

### 4.2. Variable-Marker Controlled Watershed Method (VMCWM)

Vicent L. *et al.*, [15] stated the traditional watershed (based on edge segmentation) rule to partition the gray level image employing its gradient. Watershed is a line detaching two basins, which possess low gradient values. It suffers with severe partitioning as it relies on local minima. So, Marker Controlled Watershed (MCW) is introduced to solve segmentation problem of traditional watershed. Xiaopeng *et al.*, discussed that for large SE, bright and dark regions will be eliminated by Opening and Closing morphology [12]. And small SE does not even remove dark and bright details properly during (closing and opening) morphology. So, a variable SE size is constructed for each and every pixel of the image.

**4.2.1. SE's map Formulation [ $M_s(p, q)$ ]:** Weighted variance  $V_w(p, q)$  is obtained as:

$$V_w(p,q) = W_1(g_I(p-1,q-1) - g_I(p,q))^2 + W_2(g_I(p-1,q) - g_I(p,q))^2 + W_3(g_I(p-1,q+1) - g_I(p,q))^2 + W_4(g_I(p,q-1) - g_I(p,q))^2 + W_5(g_I(p,q+1) - g_I(p,q))^2 + W_6(g_I(p+1,q-1) - g_I(p,q))^2 + W_7(g_I(p+1,q) - g_I(p,q))^2 + W_8(g_I(p+1,q+1) - g_I(p,q))^2 \quad (4)$$

With  $W_h = \frac{1}{r_h}$ , and  $r_h$  is the weighting coefficient defined as

$$r_h = \sqrt{(p - p_h)^2 + (q - q_h)^2}$$

SE map is obtained as,

$$M_s(p,q) = \left| -\log(\alpha \times V_w(p, q)) \right|, \quad (5)$$

Where,  $\alpha$  is initially defined value for  $M_s(p, q)$ , and  $\lfloor \cdot \rfloor$  is for rounding. With  $M_s(p, q)$ , all pixels in the original image are altered to get modified image as

$$M_I(p,q)=[g_I(p, q) \circ M_s(p, q)] \cdot M_s(p, q) \quad (6)$$

• and  $\circ$  define the closing and opening manipulations respectively.

### 4.3. Integration of VMCWM with Level Sets (LS)

Out of differing segmentation approaches, LS handles any type of concavities, cavities and splitting/ merging. But, traditional LS (depend on image contrast near boundaries) cannot generate accurate output for complex images (like H & E BC images possessing local inhomogeneities problem). LS require critical parameters like seed initial location, degree of smoothness, and speed function as an input to decide the system accuracy [14].

In this approach markers obtained from VMCWM acts as input to the LS. LS approaches use moving interface initial position  $C'_0(z)$  as zero LS with dimensional function  $\phi(z,0)$  to evolve new function  $\phi(z,t)$ . The contour  $C'_0(t) \forall t$  which rely on zero LS  $\phi(z,0)$  satisfies:

$$C'_0(z)=\phi(z,0) \quad (7)$$

$$(C'(t),t)=0 \quad (8)$$

$$So,t+\nabla(C'(t),t)\frac{\partial C'(t)}{\partial t} = 0 \quad (9)$$

Where  $\frac{\partial C'(t)}{\partial t} = F$ , the normal outward vector given by  $\frac{\nabla\phi}{|\nabla\phi|}$ .

LS provide accurate outcome only when initial curve is symmetrical to the object boundary of interest. So, for cell extraction in H & E images the marker output from VMCWM is integrated with multi phase LS [17] to overcome local inhomogeneities problem. And the initial front for LS is provided by VMCWM. The integrated response of VMCWM and LS extracts clearly the nuclei shown in Figure 3. In multi phase LS a number of LS's are initiated using VMCWM boundaries.

$$M_o(p, q) = T \{ [g(p, q) \circ M_s(p, q)] \cdot M_s(p, q) \} \\ = T \{ M_I(p, q) \} \quad (10)$$

Energy formulation by integrating  $M_o$  with multi phase LS is

$$\varepsilon(\Phi, \tilde{B}, \tilde{c}) = \int \left( \sum_{j=1}^N \int F(h-x) |M_o(x) - \tilde{B}(h) \tilde{c}_j|^2 dh M_j^F(\Phi(x)) dx \right) \quad (11)$$

Where, real image on domain  $\Omega$ , chooses  $N$  distinct constants  $\tilde{c}_1, \dots, \tilde{c}_N$  to split to disjoint  $N$  regions  $\tilde{\Omega}_1, \dots, \tilde{\Omega}_N$  ( $\tilde{\Omega}_j \cap \tilde{\Omega}_k = \Phi, \tilde{\Omega} = \bigcup_{j=1}^N \tilde{\Omega}_j \forall j \neq k$ ),  $\tilde{B}(h)$ , is the bias field (estimated shaded image) and  $M_j^F(\Phi)$ , membership function, depends on the number of LS functions initiated. For 2-phase LS,  $M_1^F(\Phi) = H^F(\Phi)$ ,  $M_2^F(\Phi) = 1 - H^F(\Phi)$  and  $H^F(\Phi)$ , Heaviside smoothed function  $= \frac{1}{2} [1 + \frac{2}{\pi} \arctan(\Phi)]$ .

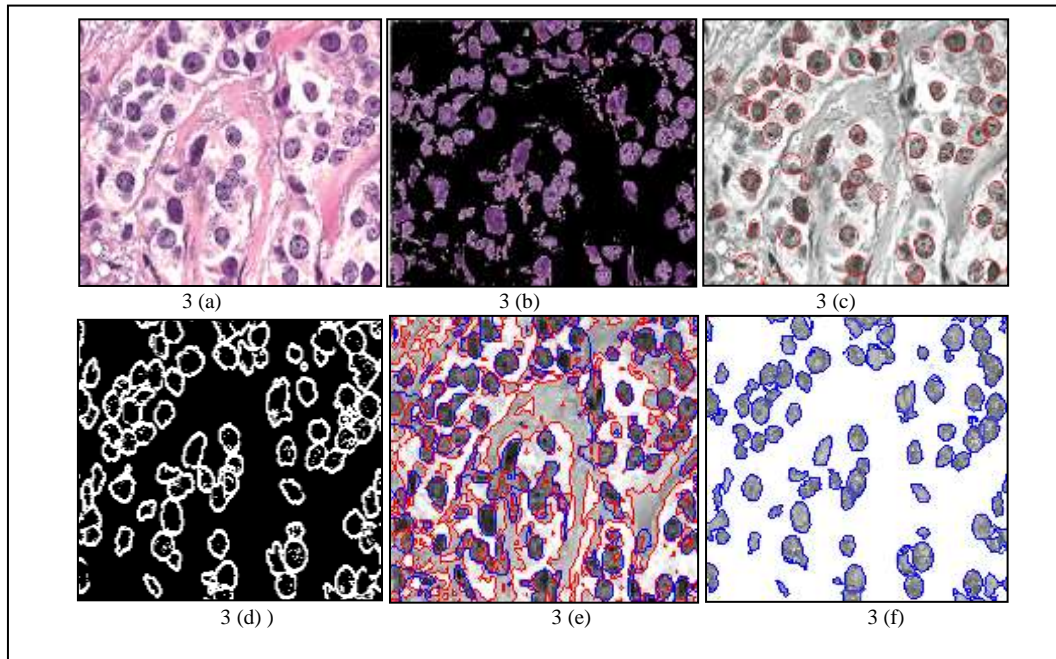
The energy functional  $E^F$ ,

$$E^F(\Phi, \tilde{B}, \tilde{c}) = \varepsilon(\Phi, \tilde{B}, \tilde{c}) + R(\Phi) \quad (12)$$

Where  $R(\Phi)$  is energy regularisation term defined as  $R(\Phi) = \sum_{k=1}^p R(\phi_p)$ , for LS function  $\Phi = (\phi_1, \phi_2, \dots, \phi_p)$  [13]. Considering  $R(\Phi)$ , the term  $E^F(\Phi, \tilde{B}, \tilde{c})$  in equation-12 can be minimised as,

$$\frac{\partial E^F}{\partial \Phi} = \frac{\partial \Phi}{\partial t} = - \left( \frac{\partial \Phi_1}{\partial t} + \dots + \frac{\partial \Phi_p}{\partial t} \right) \quad (13)$$

The overlapped nuclei problem has solved in the proposed method as shown in Figure 3(f).



**Figure 3. [3(a)] H & E Stained Breast Malignant Tissue, [3(b)] Response of MCW, [3(c)] Nuclei Detected Output, [3(d)] VMCWM Nuclei Segmented Result, [3(e)] Multi-phase LS Result [3(f)] Proposed Integrated Nuclei Segmented Result**

## 5. Results

### 5.1. Feature Extraction

Visual feature descriptors chosen include the intensity, shape and texture analysis is utilized for classification. Lot of comprehensive studies was reviewed and stated in literature [3, 6, 8, and 14] for analysis of proposed integrated segmentation module with the earlier methods. Out of all 24 features extracted only few are used for classification. Those include the nuclei shape (area, perimeter, major, minor axis length and orientation) features. Any abnormalities (for malignant breast) can be assessed with texture features. Table 2 list out the 24 features generated. Segmentation performance assessed by traditional parameters namely

1. Sensitivity (measures negative result for non-malignancy) defined as:  $\frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})}$
2. Specificity (measures positive result for malignancy) defined as:  $\frac{\text{True Negative}}{(\text{False Positive} + \text{True Negative})}$
3. Accuracy (which measures number of identified pixels out of appeared total pixels) :  $\frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})}$

4. Computational Time: Difference between the return time of the algorithm at start up of the process (tr, start) and the return time at the end of the process (tr, end).

*i.e.*,  $t_c = tr, \text{ start} - tr, \text{ end}$

Where

True Positive: The total number of correctly detected true tumor pixels.

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

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

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

**Table 2. Features Generated in the Work**

Feature type	Feature
<b>shape features</b> [10]	Area (F <sub>1</sub> ), Perimeter (F <sub>2</sub> )
	Compactness (F <sub>3</sub> )= $4\pi * F_1/F_2^2$
	Equivalent Diameter (F <sub>4</sub> ) = $\sqrt{(4F_1/\pi)}$ , Centroid (F <sub>5</sub> )
	Major axis (F <sub>6</sub> ), Minor axis (F <sub>7</sub> ) lengths
	Nucleus - Cytoplasm ratio (F <sub>8</sub> ), Orientation(F <sub>9</sub> )
	Eccentricity (F <sub>10</sub> ): deviation of segmented nucleus from that of healthy nucleus. $F_{10} = \text{Ellipse focal distance}/F_6 = \sqrt{(F_6^2 - F_7^2)}/F_6$ .
<b>Textural features</b> [12]	Mean (F <sub>11</sub> ) = $\frac{1}{kl} \sum_{m=1}^k \sum_{n=1}^l [Y_{mn}]$ k and l : number of rows and columns.
	Variance (F <sub>12</sub> ) = $\frac{1}{kl-1} \sum_{m=1}^k \sum_{n=1}^l [Y_{mn} - F_{11}]^2$ and standard deviation = $\sqrt{(F_{12})}$
	Third moment (F <sub>13</sub> ) = $\frac{1}{kl} \sum_{m=1}^k \sum_{n=1}^l [ \frac{Y_{mn} - F_{11}}{\sqrt{(F_{12})}} ]^3$
	Fourth moment (F <sub>14</sub> ) = $\{ \frac{1}{kl} \sum_{m=1}^k \sum_{n=1}^l [ \frac{Y_{mn} - F_{11}}{\sqrt{(F_{12})}} ]^3 \} - 3$
	Mean absolute deviation (F <sub>15</sub> ) = $\frac{1}{kl} \sum_{m=1}^k \sum_{n=1}^l   Y_{mn} - F_{11}  $
	Maximum probability (F <sub>16</sub> )
	Correlation(F <sub>17</sub> )= $\sum_{g_1, g_2} ( \frac{(g_1 - m_1)(g_2 - m_2) p(g_1, g_2)}{\sigma_1 \sigma_2} )$ , $m_1, m_2$ :mean and $\sigma_1, \sigma_2$ : variances of $p(g_1, g_2)$
	Contrast (F <sub>18</sub> ) = $\sum_{g_1, g_2}   g_1 - g_2  ^2 . \log p(g_1, g_2)$
	Homogeneity (F <sub>19</sub> ) = $\sum_{g_1, g_2} \frac{p(g_1, g_2)}{1 +  g_1 - g_2 ^2}$
	Energy (F <sub>20</sub> ) = $\sum_{g_1, g_2} ( p(g_1, g_2)^2 )$
	Entropy (F <sub>21</sub> ) = $-\sum_{g_1, g_2} ( p(g_1, g_2) . \text{Log } p(g_1, g_2) )$
	Haralick's correlation(F <sub>22</sub> )= $\frac{\sum_{g_1, g_2} (g_1 . g_2) . P(g_1, g_2) - m_1 . m_2}{\sigma_1 \sigma_2}$
<b>Intensity features</b> [2]	Maximum and minimum intensities (F <sub>23</sub> )
	Mean intensity (F <sub>24</sub> )

The quality of integrated approach matches with benchmark data provided by manual experts. Similarity match for integrated approach output with benchmark result is given by dice index. Table 3 shows the respective performance comparison of different segmentation approaches.

**Table 3. Performance Comparison of Segmentation Approaches**

Algorithm reference Parameter	RGMM LS [2]	VMC WM [12]	Multi-phase LS [17]	Integrated VMCWM with Multi-phase LS
<b>Sensitivity (%)</b>	73.53	84.67	88.65	95.74
<b>Specificity (%)</b>	72.33	87.63	83.46	93.27
<b>Accuracy (%)</b>	74.59	88.92	89.38	97.79
<b>Computation time (in sec)(mean ± deviation)</b>	213.09±13.63	22.832 ± 3.341	17.8309±5.341(for 500 iterations)	8.72 ± 3.762 (200 iterations)
<b>Tissue image structures extracted</b>	No separation.	Nuclei separated but suffers from local inhomogeneities	Partial separation of nuclei from stroma	Stroma and nuclei are separated. Overlapping problem of nuclei has solved
<b>Dice similarity index</b>	0.736	0.885	0.823	0.982

## 5.2. Classification

The features generated are set as input to classifiers namely Decision Trees (DT) [18], SVM (Support Vector Machine) [20] and KNN (K-Nearest Neighbor) [19]. Classification based on intensity, shape and texture features from 90 trained images were used as database for testing set of 20 images. Later Cross validation is done using Leave One Out technique (LOOT) for all the samples.

**5.2.1. Decision Trees (DT):** A decision tree (depends on Hunts algorithm) introduced in 1986 by Quinlan Ross [18] within a subfield of machine learning techniques under artificial intelligence. Construction of decision tree relies on entropy and information gain. First, Entropy, E (measure of information content) of entire data set and for every feature is calculated.

$$E(D_s) = \sum_{i=1}^n -P(x_i) \log_2 P(x_i) \quad (14)$$

where  $P(x_i)$  is the proportion of data set 'D<sub>s</sub>' belonging to the class i, and n the outcomes available in the data set. Next, Information gain G of features set is calculated as difference between entropy of complete set S before and after their split on feature F.

$$G(D_s, F) = E(S) - \sum((|D_{sv}| / |D_s|) * E(S_v)) \quad (15)$$

Where



$\Sigma$  of each value  $v$  exists for all values of Feature  $F$ .  
 $D_{sv}$  = subset of  $D_s$  for which Feature has value  $v$ .  
 $|D_{sv}|$  = number of elements in  $D_{sv}$ .  
 $|D_s|$  = number of elements in  $D_s$ .

Next the feature of highest gain is selected as root node and all remaining features as their branches. It is repeated till the possible outcomes of each and every feature is classified. For conversion of continuous data into categorized one an N-bin method (N represents the number of intervals) is used. Difference between the minimum and maximum of that column is divided into equal N bins, and each value is replaced by that interval to which it belongs.

**5.2.2. KNN classifier:** KNN classifies objects (and images) based on trained closest samples in the feature space [19]. Classification also depends on majority vote of neighbors, *i.e.*, the object is assigned to most common class among their  $k$  nearest neighbors. Minimum distance measure is utilized for classifying objects. Euclidean distance between point  $x = (a_1, b_1)$  and point  $y = (a_2, b_2)$  is

$$D(x, y) = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (16)$$

Complexity occurs in choosing the neighbors which are nearest to every sample.

**Table 4. Performance of Classifiers**

Classifier	Average sensitivity	Average specificity	Overall Accuracy	AUC
DT [18]	81.72	78.31	93.16	0.948
KNN [19]	89.68	86.42	95.34	0.967
MSVM[20]	94.37	91.87	98.11	0.972

**5.2.3. SVM Classifier:** With  $f_j$  as feature data set and  $w$  as a vector dividing hyperplane, the objective function to be reduced for non-linear SVM is

$$O'_F = \min[1/2|w|^2 + C(\sum_{j=1}^N \zeta_j)^p], \quad (17)$$

Where  $0 \leq \zeta_j \leq 1 - y_j (f_j \cdot w + b)$   
 $y_j$ , a class label of  $f_j$ ,  $\{-1, +1\}$ ,  
 $C$ , user defined parameter (chosen to evade misclassification risk).

The dual of  $O_F$  by vanishing its gradient with  $w$ ,  $b$  and  $\zeta$  for  $p=1$  (Lagrangian multiplier  $c_k$ ) is

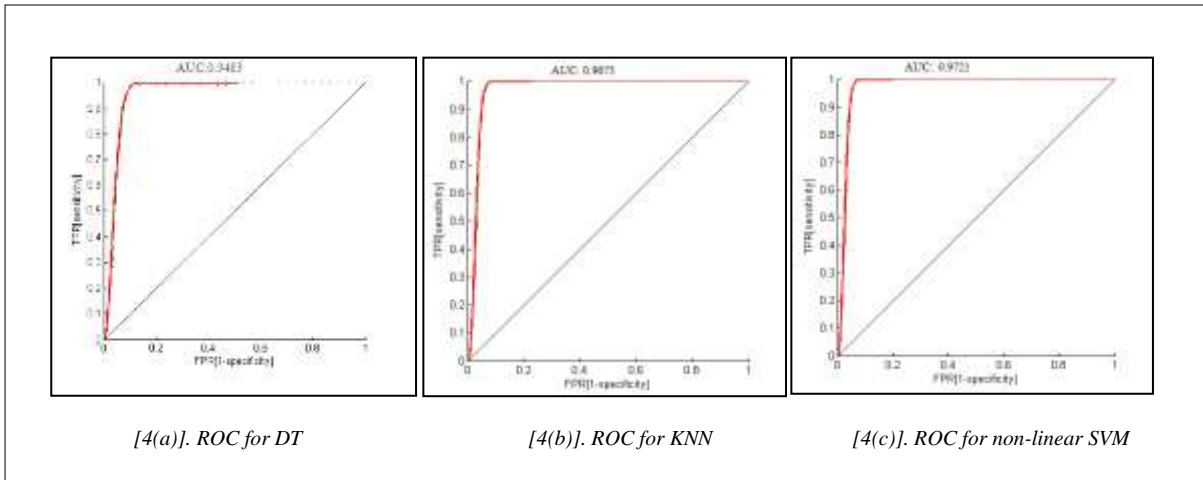
$$\text{Maximum of } O_D = \sum_{k=1}^N c_k - \frac{1}{2} \sum c_j c_k y_j y_k K(F_j, F_k) \quad (18)$$

For  $0 \leq c_k \leq C$  and  $\sum_{k=1}^N c_k y_k = 0$ ,  $K(F_j, F_k)$  is a kernel function preferred to compute the mapping functions  $\phi(F_j)$ ,  $\phi(F_k)$ . After knowing the values of  $c_j$  (known as support vectors) the trained SVM for unknown image (feature set) is defined as:

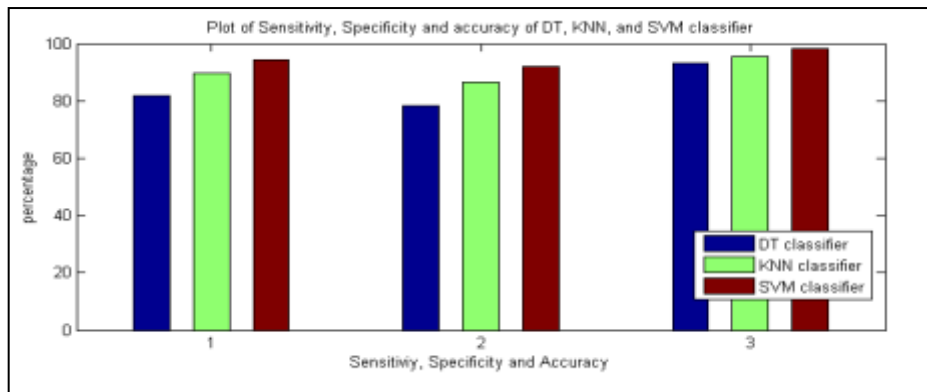
$$\begin{aligned} \text{SVM}(F_j') &= \text{sign} \left( \sum_{j=1}^N c_j y_j \phi(F_j) \phi(F_j') + b \right) \\ &= \text{sign} \left( \sum_{j=1}^N c_j y_j K(F_j, F_j') + b \right) \end{aligned} \quad (19)$$

Solving equation-(19), the hyper plane normal vector ‘w’ was obtained as (20)

$$w = \sum_{j=1}^N c_j y_j F_j$$



**Figure 4. ROC Plot**



**Figure 5. Plot of Performance Measures of Classifiers**

Figure 4 (a) show Region of Convergence (ROC) plot [i.e., TPR (True Positive Rate=sensitivity) versus FPR (False Positive Rate = 1- specificity)] for DT classifier, whose AUC (Area under Curve) is 0.9483 and a classifier accuracy of 93.16 %. Figure 4(b) shows the plot for KNN classifier, whose AUC is 0.9673 with a classifier accuracy of 95.34 %. Figure 4(c) is a SVM ROC plot, whose AUC is 0.9721 with a classifier accuracy of 98.11 %. The performance measures of these classifiers are shown in Figure 5.

## 6. Conclusions:

This paper discussed a new module which integrates VMCWM with multi-phase LS for separating nuclei from stroma and cytoplasm. Total of 24 (including shape, textural and intensity) features are recovered from the discussed method. Performance evaluation of the integrated method presents the respective increase in average specificity, sensitivity and segmentation accuracy to 93.19%, 95.74 % and 97.79 % respectively compared to VMCWM and multiphase LS. We collected 110 (90 for training, 20 for testing) BC biopsy images stained with H&E for the proposed method. Their segmented features set for classification using DT, SVM and KNN classifiers. AUC and classifier accuracy of SVM proved better than DT and KNN classifiers for the segmentation integrated

approach. Results were experimentally verified and cross-checked with the output provided by the expert.

## Acknowledgments

The authors acknowledge Department of Science & Technology, Government of India for financial support with their vide sanction reference No.100/ (IFD)/1283/2015-2016, under Women Scientist Scheme to carry out this work.

## References

- [1] World Health Organization, "Fact sheet No. 297: Cancer", <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>, (2010).
- [2] Y. Mourice George, H. Helmy Zayed, M. Ismail Roushdy and B. Mohamed Elbagoury, "Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images", IEEE Systems Journal, vol. 8, no. 3, (2014) September, pp. 949-964.
- [3] S. Osher and J. A. Sethian, "Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations", J. Comp. Phys., vol. 79, (1988), pp. 12-49.
- [4] H. Irshad, A. Veillard, L. Roux and D. Racoceanu, "Methods for Nuclei Detection, Segmentation and Classification in Digital Histopathology: A Review. Current Status and Future Potential", IEEE Reviews In Biomedical Engineering, vol. 7, (2013), pp. 1-18.
- [5] <http://www.pathologyoutlines.com/topic/breastnormal.html>.
- [6] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 881-892.
- [7] M. Trajkovii and M. Hedley, "Fast corner detection", Image and Vision Computing, Elsevier, vol. 16, (1998), pp. 75-87.
- [8] M. Hrebien, J. Korbicz and A. Obuchowicz, "Hough Transform, Search Strategy and Watershed Algorithm in segmentation of Cytological Images", Computer Recognition Systems 2, ASC 45, Springer-Verlag Berlin Heidelberg, (2007), pp. 550-557.
- [9] U. Rajyalakshmi, K. Satya Prasad and S. Koteswara Rao, "Tissue Processing, Staining And Image segmentation Of Pathological Cancer Images", International Journal of Engineering and Advanced Technology, (IJEAT), vol. 3, (2014), pp. 1-16.
- [10] M. L. Giger, N. Karssemeijer and S. G. Armato, "Computer aided diagnosis in medical imaging", IEEE Trans. Med. Imaging, vol. 20, (2001), pp. 1205-1208.
- [11] K. Zuiderveld, "Contrast limited adaptive histogram equalization", Graphics Gems IV. San Diego, CA, USA: Academic, (1994), pp. 474-485.
- [12] X. Wang, S. Wan and T. Lei, "Brain Tumor Segmentation Based on Structuring Element Map Modification and Marker- controlled Watershed Transform", Journal of software, vol. 9, no. 11, (2014) November, pp. 2925-2932.
- [13] S. Krishnamurthy, "Histogram based Morphological Edge Detector", IEEE Transaction on Geosciences and Remote Sensing, vol. 32, no. 4, (1994), pp. 759-769.
- [14] R. Malladi, J. A. Sethian and B. C. Vemuri, "Shape modeling with front propagation: a level set approach", IEEE Trans PAMI, vol. 17, no. 2, (1995), (2002) August 6, pp. 158-175.
- [15] L. Vincent and P. Soille, "Watersheds In Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, (1991), pp. 583-598.
- [16] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Syst. Man Cybern., vol. 9, no. 1, (1979) January, pp. 62-66.
- [17] C. Li, C. Xu, C. Gui and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation", IEEE Trans. Image Process., vol. 19, no. 12, (2010), pp. 3243-3254.
- [18] J. R. Quinlan, "Induction of decision trees", I. Journal of Machine Learning, vol. 1, (1986), pp. 81-106.
- [19] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression", The American Statistician, vol. 46, no. 3, (1992), pp. 175-185.
- [20] C. Cortes and V. Vapnik, "Support-vector networks", Machine learning, vol. 20, no. 3, (1995), pp. 273-297.
- [21] F. Jianping, K. Y. Yau David, E. Ahmed. K. and Walid G, "Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing", IEEE Trans on Image Processing, vol. 10, no. 10, (2001) October, pp. 1454-1466.

## Authors



**Rajyalakshmi Uppada**, is presently working as Research Associate under WOS-A, DST at JNTUK, Kakinada, India. During her 10 years of teaching experience, she has published five journal papers and presented 3 international conference papers. She is currently pursuing Ph.D from JNTU Kakinada.



**Koteswararao Sanagapallela**, is presently working as a Professor in KL University, Vijayawada, India. He has retired as a Scientist 'G', Associate Director, NSTL. He has published 30 IEEE papers. With his thirty two years 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. His salient contributions towards the goals are Submarine Fire Control System, Ship Fire Control System, Helicopter Fire Control System, Anti Torpedo Defence System, Advanced Ship Fire Control System.



**Satyaprasad Kodati**, is presently working as Sr. Professor in Electronics & Communication department and Director-IST, JNTU Kakinada, Kakinada, India. He has 32 years of teaching experience, 21 years of research experience, guided 15 Ph.D scholars and guiding 16 students for PhD. 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 40 Journal Publications and he has presented more than 50 conference proceedings.