COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 138 (2017) 31–47

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Automated classification of Pap smear images to detect cervical dysplasia

Kangkana Bora [a,*], *Manish Chowdhury* [b], *Lipi B. Mahanta* [a], *Malay Kumar Kundu* [c], *Anup Kumar Das* [d]

[a] *Department of Centre for Computational and Numerical Sciences, Institute of Advanced Study in Science and Technology, Guwahati 781035, Assam, India*

[b] *KTH, School of Technology and Health, Hälsovägen 11c, SE-141 57 Huddinge, Stockholm, Sweden*

[c] *Department of Machine Intelligence Unit, Indian Statistical Institute, 203 B.T.Road, Kolkata 700108, India*

[d] *Ayursundra Healthcare Pvt. Ltd, DMB Plaza, Lachit Nagar, Guwahati 781007, Assam, India*

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form
19 August 2016
Accepted 4 October 2016

Keywords:
Pap smear
MSER
Ripplet transform
Ensemble classification

ABSTRACT

Background and objectives: The present study proposes an intelligent system for automatic categorization of Pap smear images to detect cervical dysplasia, which has been an open problem ongoing for last five decades.

Methods: The classification technique is based on shape, texture and color features. It classifies the cervical dysplasia into two-level (normal and abnormal) and three-level (Negative for Intraepithelial Lesion or Malignancy, Low-grade Squamous Intraepithelial Lesion and High-grade Squamous Intraepithelial Lesion) classes reflecting the established Bethesda system of classification used for diagnosis of cancerous or precancerous lesion of cervix. The system is evaluated on two generated databases obtained from two diagnostic centers, one containing 1610 single cervical cells and the other 1320 complete smear level images. The main objective of this database generation is to categorize the images according to the Bethesda system of classification both of which require lots of training and expertise. The system is also trained and tested on the benchmark Herlev University database which is publicly available. In this contribution a new segmentation technique has also been proposed for extracting shape features. Ripplet Type I transform, Histogram first order statistics and Gray Level Co-occurrence Matrix have been used for color and texture features respectively. To improve classification results, ensemble method is used, which integrates the decision of three classifiers. Assessments are performed using 5 fold cross validation.

Results: Extended experiments reveal that the proposed system can successfully classify Pap smear images performing significantly better when compared with other existing methods.

Conclusion: This type of automated cancer classifier will be of particular help in early detection of cancer.

* *Corresponding author.* Department of Centre for Computational and Numerical Sciences, Institute of Advanced Study in Science and Technology, Guwahati 781035, Assam, India. Fax: 0361-2273063.
E-mail address: kangkana.bora89@gmail.com (K. Bora).

http://dx.doi.org/10.1016/j.cmpb.2016.10.001
0169-2607/

1. Introduction

Cervical cancer starts in the cells lining the cervix (the lower part of the uterus). Normal cells of cervix gradually develop pre-cancerous changes first and then develop into cancer. Cervical cancer is the second most prevalent cancer after breast cancer among women, which is more prevalent in developing countries. It is stated that cervical cancer can normally be effectively treated if it is detected at an early stage [1]. Papanicolaou test (abbreviated as Pap test) is a method of cervical screening used to detect potentially pre-cancerous and cancerous process in the cervix. Pap test has made cervical cancer one of the most preventable cancers which can be used for its early detection [2]. The test involves collection of cells lining the transformation zone (where the outer squamous cervical cells meet the inner glandular endocervical cells), preparation of smears from the collected cells, staining them with Papanicolaou stain and analyzing them under microscope to detect any abnormalities. The entire process is time consuming, costly and involves observer biases. So an automated system for screening of Pap smear images will be of particular help for the pathologist. This automated classifier is designed with the following purpose—(i) to reduce observer bias, (ii) to reduce laborious task of pathologist by sieving out the normal cases so that they can concentrate more on suspicious cases, as doctors normally spend lots of time in looking at normal cases specially when camp based screenings are performed, (iii) to quantify the features which may be interpreted by doctors in visual terms which are subjective to the concerned doctor and (iv) effective in terms of cost involved.

A number of commercial systems are available for Pap smear screening like PAPNET, ThinPrep imaging system etc., but application of those advanced techniques are limited because of the cost involved and presence of technical and linguistic gap. Many developing countries still have to rely on conventional screening techniques including India where this study has been carried out. So design and delivery of low cost but highly efficient screening system is on high demand.

Literature related to Pap smear image analysis can be categorized based on four factors—(1) the single cell level or smear level consideration, (2) types of segmentation algorithms adopted for identifying ROI, (3) types of features observed and (4) classification processes analyzed. If one considers the types of level used for analysis, it can be observed that previous researches are generally based on two types of analysis i.e. cell level study [2–5] and smear level study [6–11]. In cell level study each image contains a single cell which is either normal or abnormal cervical cell. If one considers smear level then it includes not only multiple cervical cells but also many unwanted debries which may be present in the background.

Another important characteristic on which literature can be categorized is based on segmentation techniques adopted for identifying ROI. Attempts have been made to segment the ROI based on Water immersion technique [6], Genetic algorithm [12], Fuzzy C Mean [13], Watershed segmentation [14], Joint Optimization technique [7] etc. But all the techniques were somehow incomplete in addressing some of the major issues like debries removal and consideration of adequate numbers of indigenous samples to be tested.

Literature can be also categorized based on different features used to quantify the dysplastic changes in a Pap smear image. Chen et al. [15] extracted 13 features which can describe nucleus size, nucleo cytoplasmic (N/C) ratio, shape and texture. But they have not studied the color features. Genctav et al. [16] performed their work on shape features. Guan et al. [17] and Camargo et al. [18] studied color features using DFT and MPEG 7 descriptors respectively. Using DFT accurate frequency information can be obtained but we loose time information and it has bad convergence. On the other hand in MPEG descriptors employ Discrete Cosine Transforms which has the disadvantage of neglecting correlation from the pixels of neighboring block resulting to undesired blocking artifacts affecting reconstruction of images. The present approach overcomes these drawbacks by using Ripplet Type 1 transform. Moreover consideration of small numbers of features may lead us to select an unwanted class during classification. That is why 121 low level features which can quantify the dysplastic changes were observed for this study.

Classification is the last but most important stage in designing any Decision Support System. Literature reflected that Pap smear classification is of two types—cell classification [13,18,19] and smear classification. Cell classification mainly focuses on single cell level study where researchers aim at classifying an individual cell into Normal Class (includes Normal Superficial (NS), Normal Intermediate (NI) and Normal Columnar (NC) cells), Abnormal Class (includes Light Dysplasia (LD), Moderate Dysplasia (MD), Severe Dysplasia (SD) and Carcinoma-in-situ (CIS)). Smear level classification is difficult as compared to cell level classification as it basically focuses on classifying the whole image/sample to Normal, LSIL (Low grade Squamous Intraepithelial Lesion), HSIL (High grade Squamous Intraepithelial Lesion) and Carcinoma-in-situ (according to the Bethesda system of classification) [20].

1.1. *Challenges and motivations*

A Pap smear sample contains 1000 to 10,000 cells and it is a tedious and time consuming task to manually analyze the cells under a microscope. In a Pap smear there may be different types of cervical cells present like normal superficial cells, normal intermediate cells, normal parabasal cells, normal basal cells and endocervical or normal columnar cells. Along with different types of cervical cells other types of cells called inflammatory cells and Red Blood Cells (RBC) are common which are termed as debris present in digitized image. When cervix is affected by cancer these cervical cells undergo various morphological changes which include changes in terms of shape, color and texture which in turn are termed as “features” by the researchers. But to properly study the changes in cervical cells one has to remove these debris from the image. So some of the challenges that has to be faced during designing such automated model for Pap smear classifications are staining quality, overlapped cytoplasm, presence of debris and poor contrast of some images. We aim to address some of the above mentioned challenges and attempt to design an automated classifier of Pap smear images for early diagnosis of cervical cancer. Another challenge in medical imaging is unavailability of hospital based real indigenous images. Genctav et al. [16] mentioned about the difficulty in collecting indig-

enous data for testing. To avoid ethical issues involved with these type of studies most of the researchers performed their work on publicly available limited data. But to check consistency of the system developed and for deliverable output to society one needs to perform experiments on real time data. The system based on publicly available data may not be efficient for region based real samples.

The proposed approach aimed at designing a robust fully automated cervical cancer classifier based on real Pap smear images through efficient segmentation, feature extraction and classification. This automated system can be of particular help to the pathologist in early detection of cervical cancer.

1.2. *Contributions*

The main contribution of the paper is as follows.

1. **Database generation**: The system is evaluated on real indigenous Pap smear images generated at Ayursundra Healthcare Pvt. Ltd. and Dr B. Borooah Cancer Institute, Guwahati, India following all ethical protocols. The system is trained and tested on two generated database. One database consists of single cervical cells and other database contains smear level images. Single cell database contains 1610 single cervical cells and smear level database contains 1320 images. Images are properly labeled with the help of two certified pathologists and the region of interest is selected, which is used for ground truth preparation.
2. **Fully automated segmentation technique**: A fully automated segmentation technique is used to extract the Region of Interest (ROI) i.e. the nuclei so that changes in nuclei during cancer can be studied in terms of quantitative features. Segmentation method used in this paper is efficient in dealing with inflammatory cells, RBCs, poor staining and poor contrast of images, which are some of the major challenges faced by many researchers.
3. **In-depth analysis of features**: During feature extraction phase 121 low level features were studied which are able to quantify dysplastic changes from a smear. Low level features quantify the features like shape, texture and color. Shape features were extracted using proposed segmentation technique. Color information hidden in an image was quantified using Generalised Gaussian Density (GGD) features of Ripplet Type I transform, it was also used for extracting texture features. Some other texture features were defined using first order statistics of the intensity histogram and second order statistics of Gray Level Co-occurrence Matrix (GLCM).
4. **Cell as well as smear level classification using ensemble technique**: As no single classifier exists whose accuracy is ideal (i.e. 100%), so taking decision with the help of a single classifier may not be the perfect approach. That is why ensemble classification technique is used to construct the decision model in this contribution. Ensemble classifier can be beneficial as it is a process of consulting several experts before taking a final decision [21]. In doing so we weigh the decision of three well known classifiers Least Square Support Vector Machine (LSSVM), Multilayer Perceptron (MLP) and Random Forest (RF), and integrated them through weighted majority voting to reach final decision which is presumably the most efficient one. For cell classification two and three level classification was performed. For three level classification class 1 includes normal squamous (superficial, intermediate, parabasal cells) and columnar (endocervical) cells. Class 2 includes LSIL cases. Class 3 includes HSIL including Squamous Cell Carcinoma (SCC) cases. In two level classification last two classes are combined together and termed as abnormal class. For smear classification data are divided into three level classes (NILM, LSIL and HSIL including SCC). These classifications reflect the established Bethesda pathological classification system used for examining the cytological changes in cervical neoplasia. Extensive experiments show that the proposed system is efficient and effective for Pap image classification both at cell and smear levels. Chankong et al. [13] clearly mentioned that there is no classification result reported for smear classification. So another major contribution is that this work is presenting smear level classification where ensemble classification is used and a superior recognition rate is achieved.
5. **Reproducibility**: All the experiments are reproducible using MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States). All techniques are provided with the hope of usage in future research, comparisons via peer researchers and further improvement of the system by including more features and more sophisticated techniques.

The rest of the paper is organized as follows. Section 2 describes methods proposed and adopted in this study. Section 3 demonstrates experimental results along with discussion. Finally conclusions are drawn in Section 4.

2. Materials and methods

2.1. *Summary of the algorithm*

Fig. 1 shows the block diagram of the proposed work. Work was completed in 5 phases both at cell and smear levels. In phase 1 image database was prepared. Both cell as well as smear level database was generated for experiments. This phase was completed during October 2013–October 2015. Phase 2 is for segmentation where region of interest is extracted. Phase 3 is for feature extraction where 121 low level features were extracted which can quantify shape, texture and color of a smear. Phase 4 is for final feature set design and phase 5 is for classification where ensemble classification is employed for final decision delivery. Same methodologies were used in both cell as well as smear level. The final output classes of the system reflect the degree of dysplasia present in an image, details of these classes are explained in Table 1.

2.2. *Data generation and characteristics*

1. **Generated Database**: For this study Pap smear images were generated at Ayursundra Healthcare Pvt. Ltd and Dr. B. Borooah Cancer Institute, Guwahati, India. Experienced doctors collected the sample from 132 patients and prepared the slides. The processing and staining of Pap smear was done at their respective laboratories. Conventional Pap

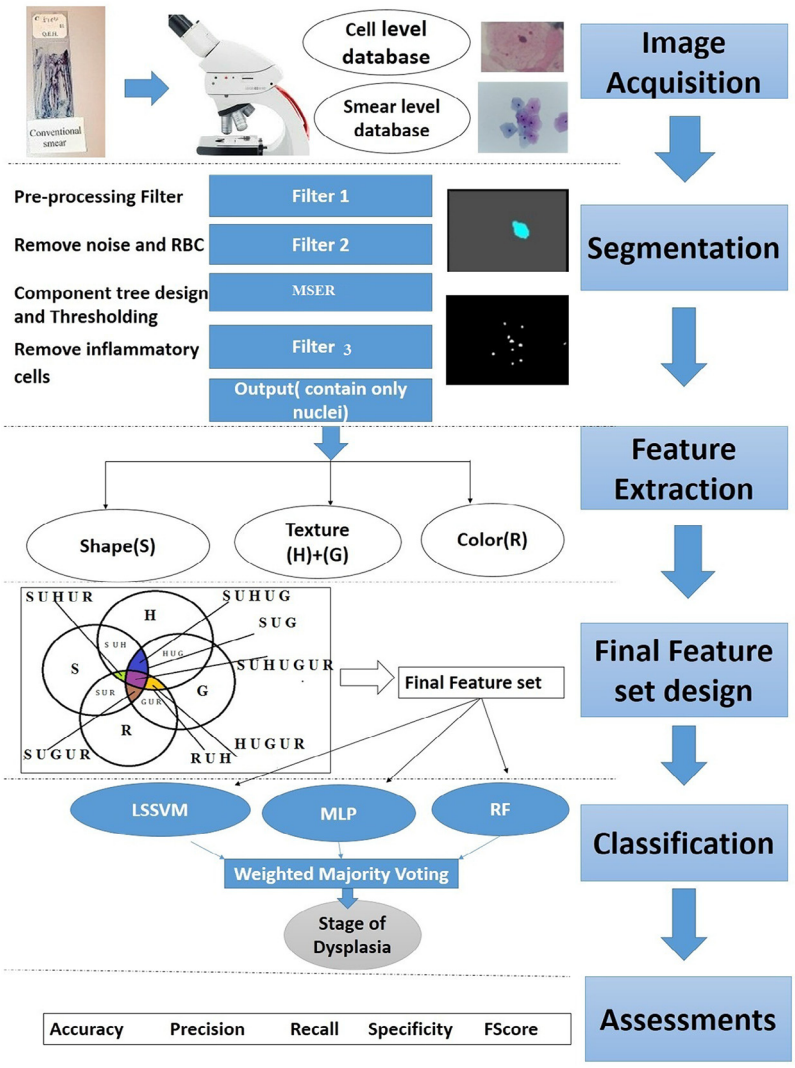


Fig. 1 – Overview of the proposed work.

smears are considered for the study as this is the screening technique which is generally used in the region where this study has been carried out. The smears were collected from patients following all ethical protocols. Images were captured by us using Leica ICC50 HD microscope using 400× resolution and 24 bits color depth. Twenty images were captured per slide. Digitized images captured were then reviewed by experienced certified cytopathologists and ten best images were selected per slide. This comprised the smear level database containing 1320 images (132 patients and 10 images per patient). Ground truth is then prepared by two pathologists by marking the area of concern in Leica photo editing software. Along with the images the corresponding original reports of the patients were also collected which are later used for labeling of the images. Image details are provided in Table 1. Characteristics of normal cell (NILM) and cells in LSIL and HSIL stage are well explained in Ref. [20].

For cell level study single cell database containing 1610 images were generated. Cytopathologist annotated the cervical cells and accordingly single cell repositories were designed by cropping out the cells.

2. **Herlev Pap smear dataset**: Database is downloaded from the url: http://labs.fme.aegean.gr/decision/downloads. This dataset contains 917 single cell images belonging to 7 different categories with resolution of 0.202 μm per pixel. The details are available in Table 1. Since we are following the Bethesda system of classification that is why Mild dysplasia is considered under LSIL and Moderate along with Severe Dysplasia is included in HSIL.

2.3. *Nuclei identification through segmentation*

Segmentation is the only tool to extract the ROI from an image. Main focus of current work is to find a segmentation technique to identify the nuclei which can be applied to all kinds of images collected.

Segmentation work of present study is inspired by Maximally Stable Extremal Region (MSER) technique introduced by Matlas [22]. In MSER the regions are defined exclusively by the intensity functions in the region and the outer border, this leads to many key characteristics of the regions which make them useful. Over a large range of thresholds, the local binarization

Table 1 – Details of images collected.

Own database: Cell level	Details
Resolution	400×
Size	Any (cropped image)
NILM	1000 (class 1 image)
LSIL	400 (class 2 image)
HSIL including SCC	210 (class 3 images)
Total image	1610
Own database: Smear level	**Details**
Resolution	400×
Size	2048*1536,2107 kB
NILM	796 (class 1 image)
LSIL	247 (class 2 image)
HSIL including SCC	278 (class 3 images)
Total image	1320
Herlev University database	
Normal image	242
Mild dysplasia	182
Moderate dysplasia	146
Severe dysplasia	197
Carcinoma in situ	150
Total image	917

is stable in certain regions. A comparision to other region detector is well explained by Mikolajczyk et al. [23]. MSER consistently resulted in the highest score through many tests (region density, region size, viewpoint change, scale change, blurring effects and light change), proving it to be a reliable region detector, which inspires us to work on MSER for segmentation purpose. MSER is a method for blob detection of images. It extracts a number of co-variant regions from an image. These connected regions are called MSERs. It considers those regions which stay same through a wide range of thresholds. According to Matlas, definition of stable region is as follows.

Let $A_1, A_2, \ldots, A_n$ be a sequence of nested extremal regions i.e. $A_i \subset A_{i+1}$. Maximally regions A_i^* are maximally stable if $a(i) = |(A_{i+\Delta} - A_{i-\Delta})/A_i|$ has a local minimum i^* (|.| denotes cardinality).

But MSER technique alone cannot extract the ROI. It is not applicable to all real images given as an input to the system. The reason behind this is the poor staining quality of some of the samples which force to select some unwanted areas during segmentation. Indeed, in real life one needs to deal with more then 2000 images, so in such cases it will be inefficient.

In the present work MSER is integrated with Discrete Wavelet Transform (DWT) and morphological operations to achieve efficient segmentation result. The technique used is shown in Fig. 2 as a block diagram. Three filters are used for this purpose.

Filter 1: This filter is used as a pre-processing filter and used to improve the quality of poorly stained and poor contrast images. In the first step low–low (LL) filter of DWT was applied in the input image. LL filter contains the approximate information of the image taken as an input. Simple Haar wavelets are considered for this purpose. Main reason behind using this wavelet is to deal with staining problem of smears. As can be observed from collected Pap smear images poor staining is one of the major challenges in Pap smear segmentation. Pretesting of this technique was performed on 50 collected images which gave good results. It will make the nucleus more prominent which will help in the segmentation process.

Filter 2: Next filter is the background removal filter which helps in removing noises and RBC present. In this phase noise is removed using median filter and background is rejected using bit plane slicing. Bit plane slicing identifies the cluster where cells are present. After that RBC were extracted using K-Mean color segmentation approach considering the fact that RBC can take only red color. Presence of RBC is another problem during segmentation, so before going for nucleus segmentation RBC

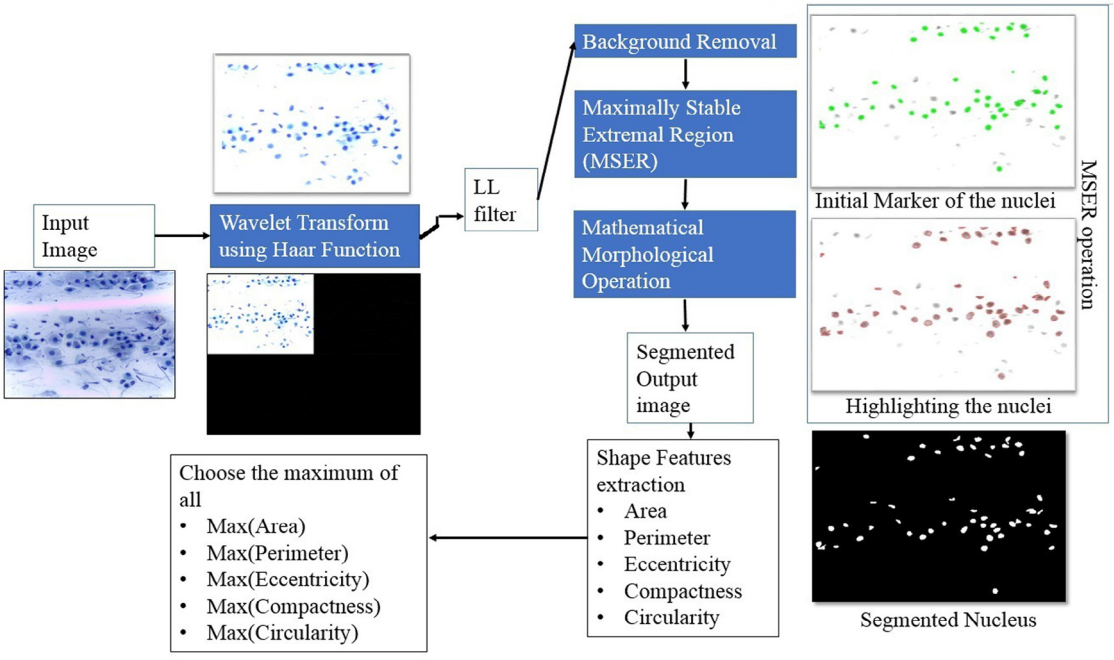


Fig. 2 – Overview of proposed segmentation technique.

has to be removed so that accuracy of the segmentation increases. Then channel identification is performed which extracted nine channels from the image, which included red, green, blue, cyan, magenta, yellow, hue, saturation and intensity channels. The channel containing the maximum information is considered for the study. Pre-testing is performed on 50 images considering all channels separately and it was observed that intensity channel performs better during segmentation as compared to other channels and is considered for further processing.

MSER: Component tree was designed for this purpose. For each node (which represents one connected component) it calculates the value ψ using the following formula:

$$\psi\left(R_i^g\right) = \frac{\left|R_i^{g-\Delta}\right| - \left|R_i^{g+\Delta}\right|}{R_i^g} \quad (1)$$

|.| indicate cardinality. R_i^g is a region which is obtained by thresholding at a gray value g and Δ is a stability range parameter. $R_i^{g-\Delta}$ and $R_i^{g+\Delta}$ are the extremal regions that are obtained by moving upward and downward respectively in the component tree from region R_i^g until a region with $g-\Delta$ and $g+\Delta$ is found [24].

Then find the connected component R_i having local minimum, this region is termed as maximally stable region.

Filter 3: Then segmented output using MSER was subjected to many morphological operations including opening and filling of holes. Next important step during morphological operation was to eliminate inflammatory cells. For this purpose area, perimeter and circularity ratio of 50 individual inflammatory cell were observed. Circularity ratio (CR) is calculated using following equation:

$$Circularity\,(CR) = A/new_A \text{ where } new_A = P^2/4\Pi,$$

A = Area of nucleus and P = Perimeter of nuclei

It was found that their values lie approximately in certain ranges (i.e. $CR \leq 0.71$). In this approach objects falling inside that particular ranges were eliminated. Pre-testing of this technique was performed on 50 Pap stained images, results showed a big improvement in the output. That is why this step is integrated in the process as a filter for removing inflammatory cells. Inspiration comes from the work of Malm et al. [25]. Output image of this filter contains only the cervical cells.

2.4. Feature extraction

Feature extraction is a process which can be used to transfer microscopically observed visual parameters to quantitative values. To differentiate among different classes, a total of 121 features comprising five shape descriptor, 50 textures and 66 Ripplet descriptors were extracted from a Pap smear image.

2.4.1. Shape features

The shape descriptors used for the study are Area, Perimeter, Eccentricity, Compactness and Circularity of the nucleus which composed the vector f^I_{Shape} as shown in Equation 2.

$$f^I_{Shape} = \{Area, Perimeter, Eccentricity, Compactness, Circularity\} \quad (2)$$

The formula used are as follows:

1. Area (A): Represents the number of pixels used to represent the object.
2. Perimeter (P): Represents the number of pixels used to represent the boundary of the object.
3. Eccentricity $(ECC) = \frac{u}{v}$ where u and v represent the minor and major axess of the object.
4. Circularity $= \frac{A}{new_A}$ where $new_A = \frac{P^2}{4\pi}$ [26].
5. Compactness $= \left(\frac{P^2}{2 * A}\right)$

These features mainly aimed at studying the nucleus enlargement, irregularity in nuclear membrane and dysmorphism of the nuclei.

2.4.2. Texture features

Out of 50 texture features 6 texture descriptors ($f^I_{Histogram}$) were extracted using first order statistics of histogram which includes features like Mean, Variance, Kurtosis, Skewness, Energy and Entropy. Remaining 44 texture features were extracted using GLCM analysis and designed the vector f^I_{GLCM}. These include the second order texture features like Angular Second Moment, Homogeneity, Correlation, Sum of Squares, Inverse Difference Moments, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation 1, Information Measure of Correlation 2, Autocorrelation, Contrast, Cluster Prominence, Cluster Shade, Dissimilarity, Maximum Probability, Variance, Sum Homogeneity and Sum of Energy [27,28]. Both minimum and maximum of above GLCM values along $\theta = 0°, 45°, 90°, 135°$ were considered for this study. Chromatin pattern and existence of nucleoli have an effect on textural information.

2.5. Ripplet descriptors

Ripplet descriptors ($f^I_{Ripplet}$) were used to quantify both color as well as texture features. Main aim of color features is to check the presence of hyperchromasia and cytoplasmic changes. Conventional transforms like Fourier transform (FT) and wavelet transform suffer from discontinuities such as edges and contours in images. To address this problem, Xu et al. proposed a new MGA-tool called Ripplet Transform (RT) [29]. RT is a higher dimensional generalization of the Curvelet Transform (CVT), proficient of demonstrating images at different scales and directions. To attain anisotropic directionality, CVT uses a parabolic scaling law i.e., $width = length^2$. From the perception of microlocal study, the anisotropic attribute of CVT promises for resolving 2D singularities along C^2 curves. Moreover, RT delivers a new tight frame with sparse demonstration for images with discontinuities along C^d curves. Also RT optimizes the scaling law with addition of two parameters, i.e., support (c) and degree (d) [29]. CVT is just a singular case of RT with $c = 1$ and $d = 2$. The anisotropy capability of representing singularities along subjectively shaped curves of RT is due to these new parameters c and d [30]. The images in the database prior to RT decomposition are transformed from RGB to YCbCr color space. This ensures that the textural characterizations of the images are independent of the

color characterization. RT decomposition over the intensity plane (Y) characterizes the texture information, while the RT decomposition over chromaticity planes (Cb and Cr) characterizes color. Texture and color information is extracted by using RT on each color plane with 4 level (1, 2, 4, 4) decomposition. This decomposition configuration produces 11 (= 1 + 2 + 4 + 4) subbands for each image of the database for each color plane. As there are three color planes, so altogether we get 33 (= 3 × 11) subbands for each image of the database. The distribution of ripplet subband coefficients are then modeled with Generalized Gaussian Density (GGD) which is defined as,

$$p(x;\alpha,\beta)=\frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|x|/\alpha)^{\beta}} \tag{3}$$

where x is the ripplet subband coefficients and $\Gamma(.)$ is the gamma function, i.e., $\Gamma(z)=\int_0^{\infty}e^{-t}t^{z-1}dt$, $z>0$. Here the scale parameter α models the width of the probability distribution function (PDF) peak (standard deviation), while the shape parameter β is inversely proportional to the decreasing rate of the peak. These two parameters need to be estimated for feature vector creation. As Maximum Likelihood (ML) estimator is best suited for estimating heavy-tailed distribution like GGD for both small and large samples, we have used ML estimator in our proposed scheme.

For the sample set $x=(x_1,x_2,x_3,\ldots,x_k)$, x_i is the ripplet coefficients at the ith subband, and $i\le L$, the ML estimator is defined as [31]

$$L(x;\alpha,\beta)=\log\prod_{i=1}^{L}p(x_i;\alpha,\beta) \tag{4}$$

GGD parameters are defined with the following equations, which have a unique root in probability

$$\frac{\partial L(x;\alpha,\beta)}{\partial\alpha}=-\frac{L}{\alpha}+\sum_{i=1}^{L}\frac{\beta|x_i|^{\beta}\alpha^{-\beta}}{\alpha}=0 \tag{5}$$

$$\frac{\partial L(x;\alpha,\beta)}{\partial\beta}=\frac{L}{\beta}+\frac{L\Psi(1/\beta)}{\beta^2}-\sum_{i=1}^{L}\left(\frac{|x_i|}{\alpha}\right)^{\beta}\log\left(\frac{|x_i|}{\alpha}\right)=0 \tag{6}$$

where $\Psi(.)$ is the digamma function, i.e., $\Psi(z)=\Gamma'(z)/\Gamma(z)$. α has a unique, real, positive solution and can be obtained from Eq. (5) by fixing $\beta>0$:

$$\hat{\alpha}=\left(\frac{\beta}{L}\sum_{i=1}^{L}|x_i|^{\beta}\right)^{1/\beta} \tag{7}$$

Substituting this into (6), the shape parameter β is the solution of the following *transcendental* equation

$$1+\frac{\Psi\left(1/\hat{\beta}\right)}{\hat{\beta}}-\frac{\sum_{i=1}^{L}|x_i|^{\hat{\beta}}\log|x_i|}{\sum|x_i|^{\hat{\beta}}}+\frac{\log\left(\frac{\hat{\beta}}{L}\sum_{i=1}^{L}|x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}}}{\hat{\beta}}=0 \tag{8}$$

which can be solved numerically using the Newton–Raphson iterative procedure. Therefore, we obtain two features from a ripplet subband (l). Considering s (= 3l) as the total number of subbands for an image I, we obtain 2s dimensional ripplet GGD feature vector. The final GGD based feature vector of an image I in the database is as follows:

$$f^{I}_{\text{Ripplet}}=[\alpha_1,\beta_1,\alpha_2,\beta_2,\ldots,\alpha_s,\beta_s] \tag{9}$$

where α_l and β_l are the GGD parameters computed from the coefficients of the lth ripplet subband and $1\le l\le s$.

2.6. *Ensemble classification using weighted majority voting*

Ensemble classifiers are the choice of many researchers now a days and applied in many fields including pattern recognition and bioinformatics. Bolon-Canedo et al. [32] tried to study ensemble classifiers as an application to microarray data classification in the field of bioinformatics. Sun et al. [33] applied the concept of ensemble classification on imbalanced data classification and obtained a superior classifier compared to conventional one. Some more application of this technique in bio-medical field is available in Refs. [34, 35]. This inspired us to apply the concept of ensemble classification in our work. The strategy in ensemble system is to consult as many classifiers as possible (considering the complexity) and weigh their decision such that the final decision improves upon the performance of a single classifiers. To increase the efficiency of the proposed algorithm, ensemble classifier has been designed using three well known classifiers i.e., Multilayer Perceptron (MLP), Random Forest (RF) and Least Square Support Vector Machine (LSSVM). This individual classifiers were selected because of their diversity in performance. The final feature sets are given as an input to the ensemble classifier and output generated by system is the particular class of the sample. For this approach two level and three level classification of dysplasia is considered, details are available in Table 1. The descriptions of individual classifiers are widely available in literature [30,36–38].

For this approach weighted majority voting technique is applied to weight the decision of each classifiers. Any strategy for designing ensemble classifier focus on individual classifiers diversity [21]. Decision of ith classifier is given as $\alpha(t,j)\epsilon(0,1), t=1,2.....T$ and $j=1,2.....C$ where T is the number of classifiers and C is the number of classes. The tth classifiers choose class w_j then $\alpha(t,j)=1$ and 0 otherwise. A weight w_t to each classifier is assigned to estimate the performance. Here the weight is computed as follows:

$$w_t=\log\frac{1}{\beta_t},\ where\ \beta_t=\frac{\gamma_t}{1-\gamma_t} \tag{10}$$

γ_t is the weighted training error of each classifier. The classifiers whose decision are combined through weighted majority voting will choose class, if

$$\sum_{t=1}^{T}w_t\alpha_{t,j}=max_{i=1}^{C}\sum_{t=1}^{T}w_t\alpha_{t,j} \tag{11}$$

3. Results and discussion

3.1. *Evaluation of segmentation technique*

Measure used for assessing the segmentation techniques is segmentation accuracy. Formula for segmentation accuracy for cell level segmentation is as follows:

$$Accuracy\,(Cell-level) = Avg\left(\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}\right) \quad (12)$$

where $i = 1, 2, \ldots n$, n is indicating the total number of images. TP is the total count of pixels which is correctly classified as nucleus by segmentation process. TN is the total count indicating those pixels which are classified as correct background. FP indicate those pixels classified as background but they are actually nucleus in the ground truth. On the other hand FN is the count of pixels which are classified as nucleus during the process but actually they are background pixels in the ground image. Average of the accuracy count for all 1610 images is considered as final accuracy.

For smear level segmentation the ground truth was prepared by pathologist by marking the region to be identified. But visual inspection is performed to calculate accuracy for smear level segmentation as it is difficult to manually segment 1320 images which contain more than 10,000 cells. So accuracy is calculated as follows:

$$Accuracy\,(Smear-level) = Avg\left(\frac{TP_i' + TN_i'}{TP_i' + TN_i' + FP_i' + FN_i'}\right) \quad (13)$$

where $i = 1, 2 \ldots m$, m indicating 1320 number of total images. TP' indicates the total number of regions correctly segmented by the proposed technique, TN' indicates the correctly segmented background (in this case is considered as constant, as there is always one background therefore $TN' = 1$), FP' indicates the regions which are not marked by the pathologist but detected as nuclei by the system and FN' is the total regions which are marked by pathologist but unable to segment by the process.

Accuracy of this technique was compared with existing segmentation techniques and results are shown in Fig. 3. From the results obtained it can be stated that the technique used in this paper is efficient in terms of accuracy and stability as compared to Otsu, Entropy, K-Mean, FCM [13] and Morphological Reconstruction [10] based on segmentation-accuracy. Thresholding techniques like Otsu and Entropy are fast in performance but not able to perform well in case of small invariant cluster segmentation. Clustering techniques like K-Mean and FCM are limited in the sense that one has to pre-define the number of clusters. Morphological Reconstruction [10] used watershed segmentation which was seen to perform over-segmentation in many abnormal image segmentation. These five segmentation techniques were tested in 9 different channels of the input image i.e red, green, blue, cyan, magenta, yellow, hue, saturation, intensity channels and the channel giving good segmentation output is considered for comparison. Different channels used during segmentation for above mentioned techniques are displayed in Fig. 3.

Fig. 4(a,b) displays some visual result of the segmentation method used in both cell and smear level. Three middle column of Fig. 4 shows the output of three filters used in the process. First filter used is DWT LL filter which mainly deals with staining problem and handle images captured in different ranges of light in a microscope. The output image as shown in the figure will have a homogeneous background and nuclei regions will be more prominent which will reduce the rate of misclassification. It is clear from the figure that all the five input images have different background color with different contrast levels which is mainly coming from different ranges of light used during the capture of images. Second filter used is the background removal filter, this filter (third column) removes RBCs present. This filter is not required in cell level segmentation as the concerned image only contains the cervical cell. Final filter deals (fourth column) with the inflammatory cells where most of them gets removed during the process and output contains only the ROI. All objects having circularity less than 0.71 were detected as inflammatory cells or unwanted regions and get removed during the process (for more details please refer to Section 2.3).

3.2. Feature extraction and final feature set design

Next, attempt was made to study which feature set carries maximum information of malignancy and give good classification accuracy both at cell level and smear level. For this purpose ($15 = 2^4 - 1$) subsets of the feature set f^I_{Shape}, $f^I_{Histogram}$, f^I_{GLCM} and $f^I_{Ripplet}$ were considered. Accuracy was tested using LSSVM, MLP and RF for all individual subsets. For this purpose dataset was randomly divided into two independent subsets, i.e training set (contain 70% of data) and test set (contain 30% of data).

Tables 2 and 3 show the accuracy of classification considering all the subsets. Two class classification mentioned in the table indicates that the sample is divided into either Normal or Abnormal class for cell level. In case of smear level classification (three level classification) sample is divided into one of the three classes of Normal, LSIL and HSIL (including SCC). It can be observed from the results that the subset $\{f^I_{Shape}, f^I_{Histogram}\}$ carries maximum information of malignancy and giving highest accuracy during cell level classification using all the three individual classifiers. Whereas the subset $\{f^I_{shape}, f^I_{Histogram}, f^I_{GLCM}, f^I_{Ripplet}\}$ carries maximum information for smear level classification, indicating the need of additional color and texture information for dysplasia detection from smear. Therefore these feature sub-sets were finally considered as the final feature set for ensemble classifier design. It also signifies that all the features i.e. shape, texture and color, are equally important for smear classification and carrying maximum information for malignancy detection and highly associated with each other. Whereas only Shape and Histogram features are sufficient enough for cell level classification (Table 2).

Therefore final feature sets considered for cell and smear classification are as follows:

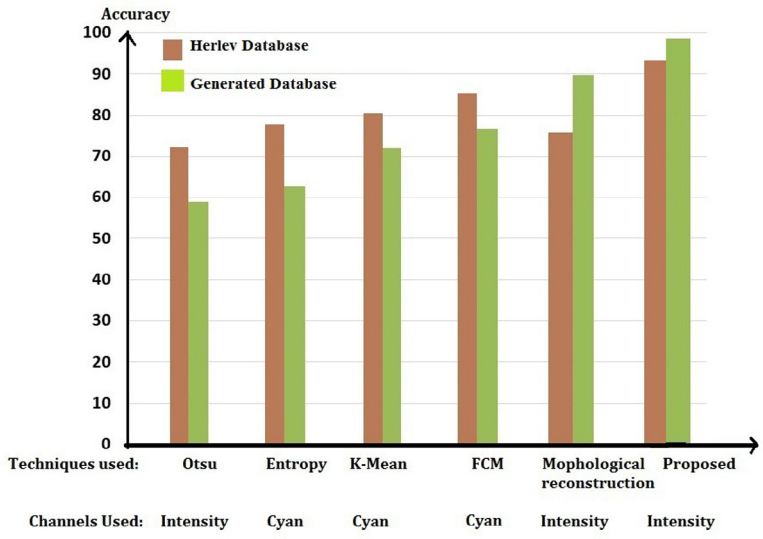
$$f^I_{Final-cell} = \{f^I_{Shape}, f^I_{Histogram}\} \quad (14)$$

$$f^I_{Final-smear} = \{f^I_{Shape}, f^I_{Histogram}, f^I_{GLCM}, f^I_{Ripplet}\} \quad (15)$$

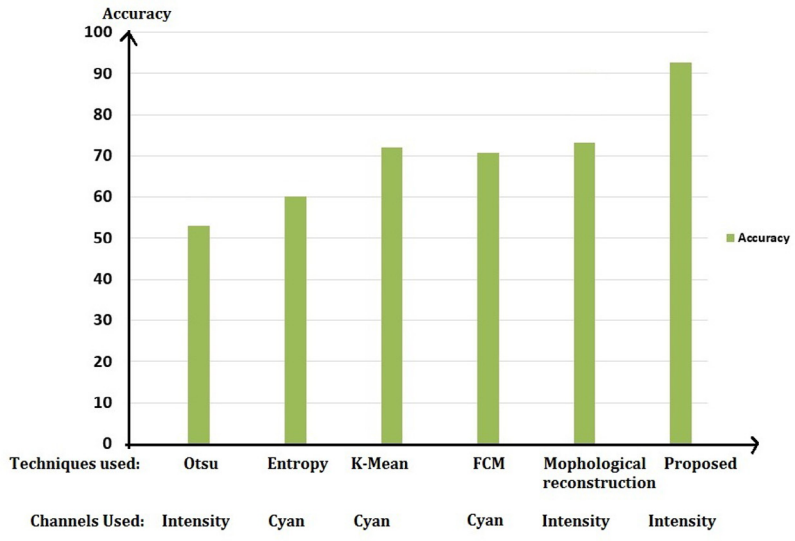
3.3. Evaluation of ensemble classifier performance

For assessment of the classification technique five measures are taken into consideration. These measures are listed in Table 4.

The accuracy of classification was evaluated by computing the number of correctly recognized class examples (true positives), the number of correctly recognized samples that do



(a) Cell Level Segmentation

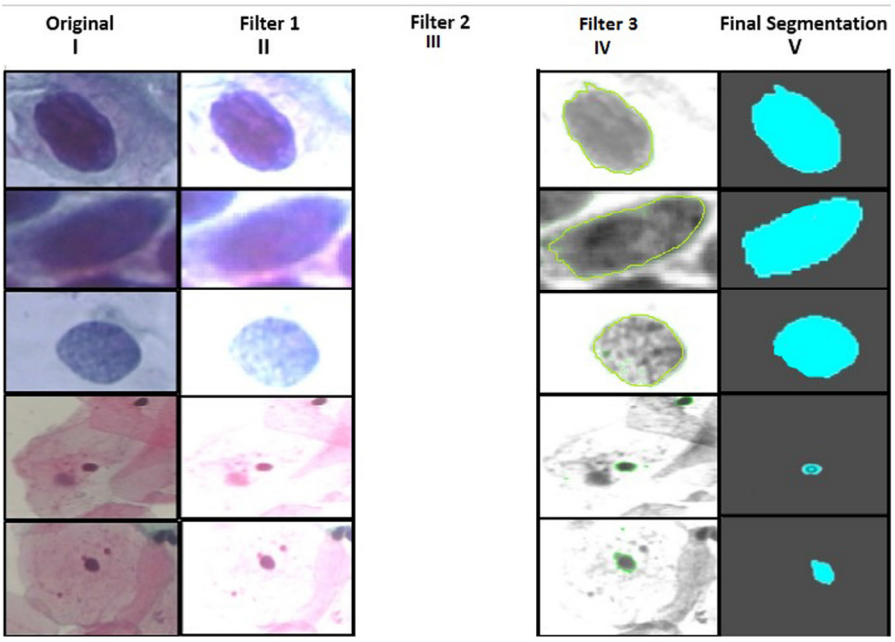


(b) Smear Level Segmentation

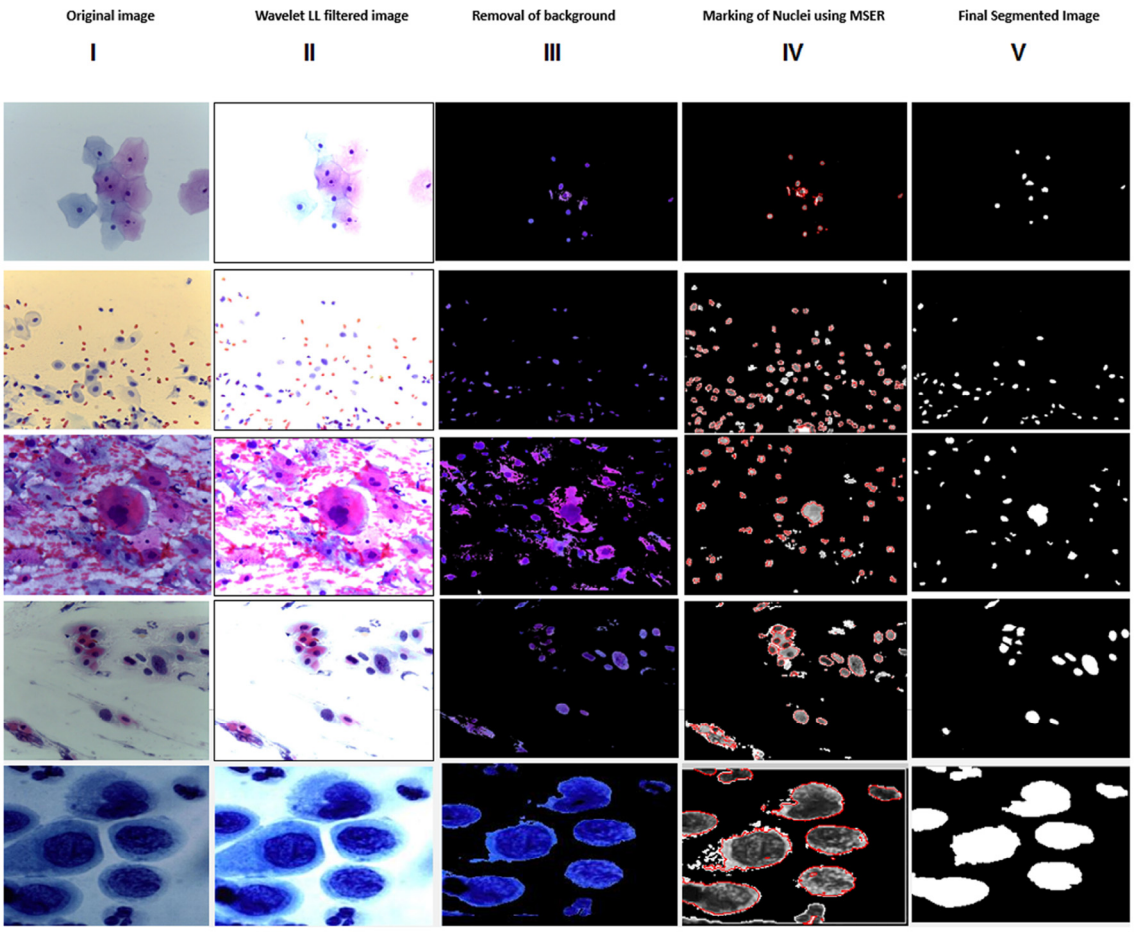
Fig. 3 – (a, b) Comparison of different segmentation techniques using cell and smear level images respectively.

not belong to the class (true negatives) and examples that either were incorrectly assigned to the class (false positives) or that were not recognized as class examples (false negatives) [39]. The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined. The higher the accuracy, the higher the rate of truly classified classes. The value of recall reflects the rate of miss-classification i.e., the higher the value of recall, the lower will be the rate of misclassification. On the other hand, precision reflects the rate of false positives; the lower is the rate of false positives, the higher is the precision. F-measure is a harmonic mean of recall and precision. The value of F-measure will be high only when both the recall and precision are high, i.e. only when the miss-classification rates as well as rate of false positives both are low.

For performing experiments features in the final feature matrix were randomly divided into five subsets. Then 5 fold cross validation is used to get the optimal parameter for LSSVM, MLP and RF. In each iteration four folds are used for training the model while the rest one was used for validation. The procedure was repeated five times using an individual fold for validation. The classification accuracy of the LSSVM, RF and MLP classifiers and a comparison are summarized in Fig. 7. For cell level classification feature set $f^{I}_{Final-cell}$ was considered and for smear level classification $f^{I}_{Final-smear}$ was used. Five experiments were performed for assessment of classifiers.



(a) Cell Level Segmentation



(b) Smear Level Segmentation

Fig. 4 – (a, b) Visual output of segmentation process using cell and smear level respectively.

Table 2 – Feature subset selection for generated single cell dataset.

Features subsets	Generated single cell database					
	Two level classification					
	LSSVM		MLP		RF	
	Accuracy	Time	Accuracy	Time	Accuracy	Time
{S}	95.86	0.374	96.41	58.944	83.11	48.48
{H}	91.93	0.329	92.94	77.859	81.49	51.79
{R}	70.19	0.400	66.20	47.795	66.57	68.91
{G}	83.23	0.374	84.73	58.898	75.69	37.68
{S,H}	**96.27**	0.362	**98.88**	39.820	**83.12**	47.11
{S,R}	60.66	0.380	65.47	40.473	55.71	89.56
{S,G}	85.09	0.363	83.77	56.980	76.92	31.27
{H,R}	61.49	0.404	65.47	55.687	51.66	22.57
{H,G}	89.10	0.370	89.60	122.137	79.61	90.99
{R,G}	62.11	0.417	77.78	58.770	61.61	35.79
{S,R,H}	62.53	0.398	72.52	44.670	55.77	35.11
{S,R,G}	62.78	0.344	75.62	33.356	61.91	44.12
{S,H,G}	89.97	0.246	90.11	45.679	70.68	61.79
{H,R,G}	65.56	0.211	76.88	32.784	55.78	35.45
{S,H,R,G}	88.36	0.253	90.79	48.998	78.15	87.21
Features subsets	Three level classification					
	LSSVM		MLP		RF	
	Accuracy	Time	Accuracy	Time	Accuracy	Time
{S}	91.72	0.282	93.16	59.663	75.88	98.42
{H}	82.82	0.218	90.70	72.724	73.18	82.86
{R}	78.88	0.288	80.77	68.929	79.57	101.39
{G}	81.99	0.280	80.58	55.687	73.92	100.22
{S,H}	91.55	0.231	94.11	101.473	80.11	77.92
{S,R}	95.48	0.268	95.33	101.233	83.18	87.11
{S,G}	90.72	0.265	91.67	97.761	79.91	88.71
{H,R}	87.99	0.303	82.79	80.911	68.52	103.29
{H,G}	85.51	0.633	79.89	76.111	79.11	48.67
{R,G}	88.82	0.314	87.55	100.231	71.89	39.11
{S,R,H}	95.52	0.301	94.71	100.239	80.31	47.67
{S,R,G}	94.72	0.355	81.88	59.720	75.16	61.44
{S,H,G}	89.72	0.288	80.72	102.534	72.98	78.53
{H,R,G}	79.99	0.414	80.11	78.981	66.67	35.40
{S,H,R,G}	**97.55**	0.253	**97.48**	164.181	**84.11**	110.60

$S = f^I_{Shape}$, $H = f^I_{Histogram}$, $G = f^I_{GLCM}$ and $R = f^I_{Ripplet}$.
Bold values indicate the best performances of each classifier.

- *Experiment 1*: Two level classification of generated single cell dataset. Total number of images is 1610.
- *Experiment 2*: Three level classification of generated single cell dataset. Total number of images is 1610.
- *Experiment 3*: Two level classification of Herlev dataset. Total number of images is 917.
- *Experiment 4*: Three level classification of Herlev dataset. Total number of images is 917.
- *Experiment 5*: Three level classification of generated smear level dataset. Total number of images is 1320.

Average accuracies of LSSVM, RF and MLP are used to generate weights for the ensemble classifier. Performance of ensemble classifier compared with individual classifiers is shown in Fig. 8.

Ensemble classifier results were compared with other three classifiers which are found in existing literature for Pap smear classification. These include SVM [10,13,15,18], Bayesian [13] and KNN [18]. Results obtained (as shown in Table 5) proved that ensemble classifier performs better as compared to the mentioned classifiers for all levels for classification.

Next, the system was evaluated on generated database containing 1320 real Pap smear images and classification was performed on smear level using feature set $f^I_{Final-smear}$. Fig. 9 shows that the ensemble classifier is giving better performance (98.11% accuracy and 98.33% precision) as compared to individual classifiers LSSVM, MLP and RF in smear classification. Its also performs well compared to SVM, Bayesian and KNN classifiers (as shown in Table 5). Fig. 9 shows different precision, specificity, recall and F-score values associated with each class of the ensemble classifier. All the measures are in the range of 0.9–1.0 reflecting high significance of the methods adopted for this approach. High precision and recall value of ensemble classification indicate that the misclassification rate is low for ensemble method as compared to individual classifiers.

Table 3 – Feature subset selection for generated single cell dataset.

	Generated smear level dataset					
	Three level classification					
	LSSVM		MLP		RF	
	Accuracy	Time	Accuracy	Time	Accuracy	Time
{S}	85.67	0.2061	82.00	53.8683	67.98	5.0382
{H}	61.33	0.4847	42.66	83.2635	52.67	5.61267
{R}	86.83	0.2166	88.00	1005.8887	71.50	5.8912
{G}	68.00	0.2092	64.00	250.1491	59.67	5.4989
{S,H}	86.83	0.2575	71.33	71.8826	55.67	4.9864
{S,R}	90.83	0.2239	93.00	195.7888	68.80	5.5665
{S,G}	89.51	0.4227	92.33	587.0897	65.58	5.4636
{H,R}	88.17	0.2244	83.00	540.3698	59.95	5.3320
{H,G}	78.00	0.2182	84.66	327.482	60.10	5.2210
{R,G}	90.83	0.2239	89.66	1222.5613	57.95	5.9674
{S,R,H}	90.17	0.2177	93.33	1189.8556	68.80	5.4683
{S,R,G}	92.33	0.2166	88.66	558.786	69.17	5.9116
{S,H,G}	91.17	0.2149	84.66	600.2015	61.20	5.3455
{H,R,G}	88.33	0.2333	90.00	2935.89	58.95	6.0326
{S,H,R,G}	**95.17**	0.2450	**94.33**	526.8349	**78.15**	6.2711

$S = f^{I}_{Shape}$, $H = f^{I}_{Histogram}$, $G = f^{I}_{GLCM}$ and $R = f^{I}_{Ripplet}$.
Bold values indicate the best performances of each classifier.

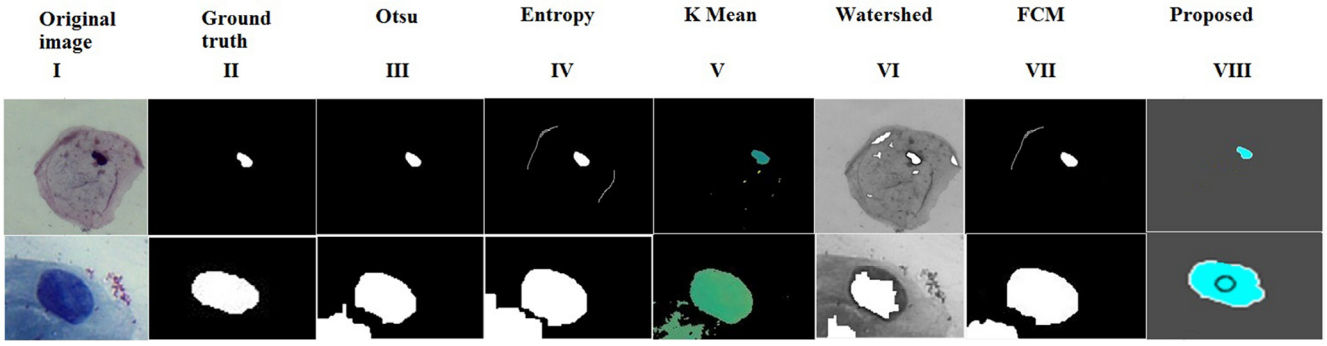


Fig. 5 – Comparison of different algorithms at cell level segmentation.

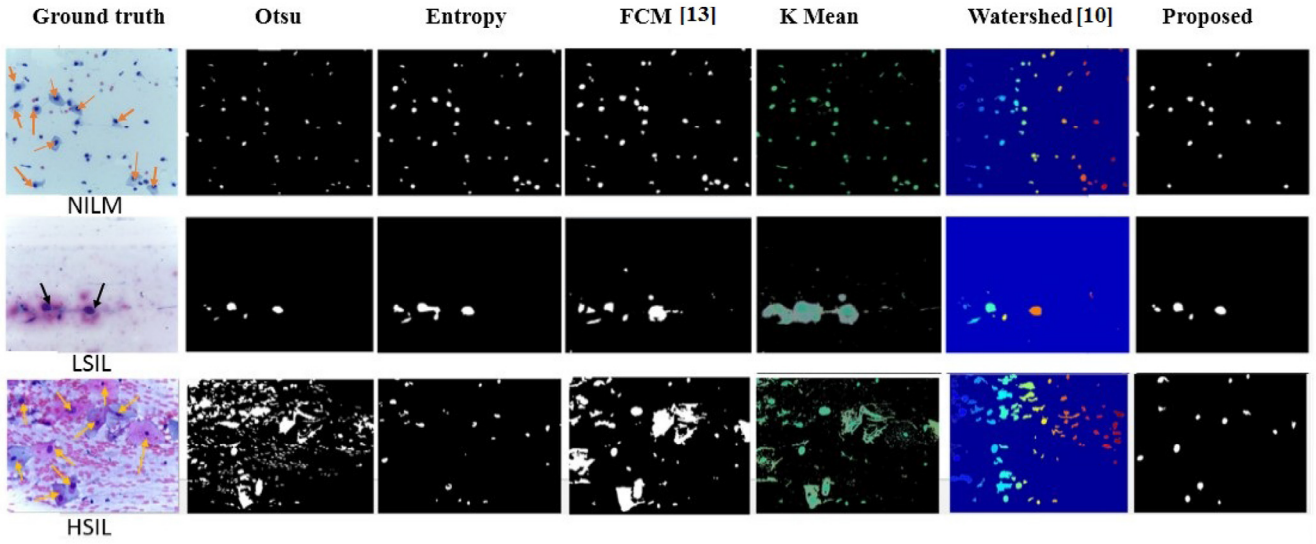


Fig. 6 – Comparison of different algorithms at smear level segmentation.

Table 4 – Measures for efficiency evaluation of classification tp_i = true positive of class C_i, tn_i = true negative of class C_i, fp_i = false positive of class C_i, fn_i = false negative of class C_i, l = number of classes.

Assessments	Formulas
Average accuracy	$\frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l}$
Precision	$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}$
Recall	$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}$
Specificity	$\frac{\sum_{i=1}^{l} \frac{tn_i}{tn_i + fp_i}}{l}$
F score	$2 * \frac{Precision * Recall}{Precision + Recall}$

3.4. *Comparison based on cell level classification*

The proposed approach was compared with 2 exiting approaches [13,19]. In order to have a fair comparison a balanced number of samples in training and test set was considered. For comparison purpose above two techniques were reproduced and tested using Herlev database and generated single cell database. Comparison results are displayed in Table 6. Results show that ensemble classifier outperforms the existing two techniques.

3.4.1. *Comparison based on smear level classification*

According to Chankong et al. [13], there is no classification result reported on smear level. Because of the unavailability of proper documentation and benchmark database, this technique could not be compared but we have developed this technique in such a way that this work can be reproduced whenever needed and can be compared in future. Although comparison with existing segmentation technique for smear level segmentation is provided in Figs. 3 and 4. Also comparison with existing classifiers is available in Table 5.

4. Conclusion

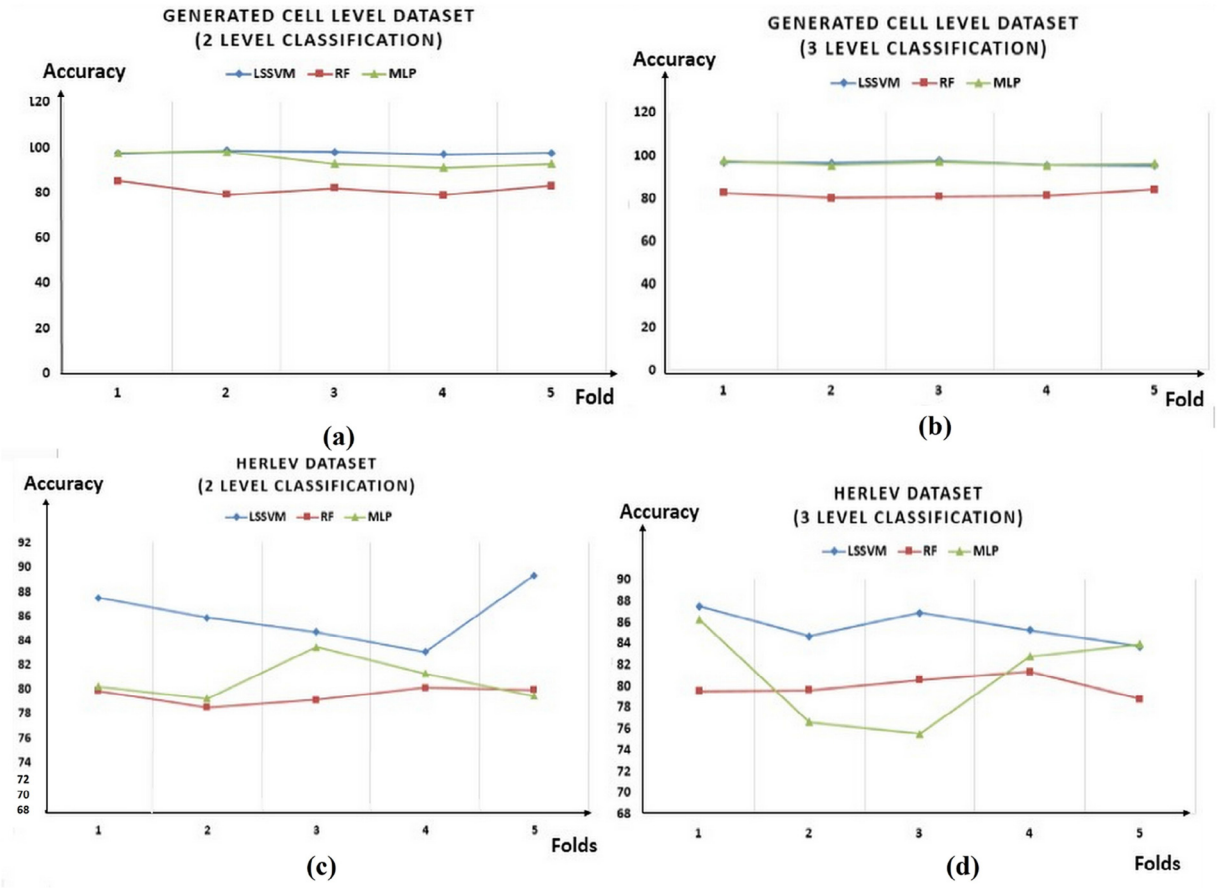
The conventional method of obtaining Pap smear is still the most economic and prolifically used technique, both for actual patients and for early diagnostic camps held time to time by government and non-government organizations. So development of an efficient and robust automated screening system for Pap smear would be useful for all underdeveloped and developing countries, if not for developed countries too. This approach lacks in addressing confounding factors like age and menstrual cycle which can also influence the output of the system. But further studies will include these confounding factors as new additional features. We will also focus on increasing the size of database so that system accuracy can be made consistent in dealing with images having various degrees of dysplasia. Testing the system on ThinPrep images will also include in the upcoming studies.

In this study ensemble classifier is used for Pap smear image classification which gives performance with an accuracy of 98.11% and precision of 98.38% in smear level and 99.01% in cell level using generated database for three level classification. Using Herlev database it achieves an accuracy of 96.51% (2 class) and 91.71% (3 class). Three popular individual classifiers LS-SVM, MLP and RF are used in designing of ensemble classifiers. Three features i.e. Shape, Texture and Color are analyzed properly. It is proven that all the features are very important for classification of Pap smear samples. Single feature cannot provide high accuracy during classification. Final class

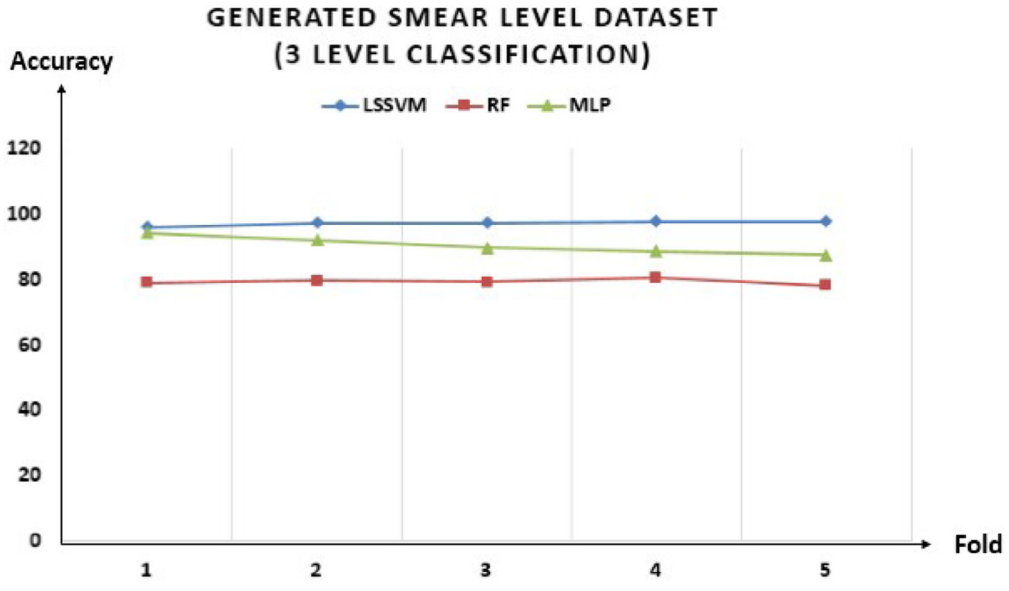
Table 5 – Comparison of ensemble classifier with existing classifiers.

Database	Classifier	Level	Accuracy	Precision	Recall	Specificity	F-score
Herlev	*Ensemble*	2	**96.51**	**96.88**	**98.96**	**89.67**	**93.13**
		3	**91.71**	**91.83**	**89.41**	**94.84**	**90.51**
	Bayesian	2	87.98	90.63	97.78	60.42	72.50
		3	72.67	73.81	73.22	84.76	70.90
	SVM	2	95.20	93.75	97.93	87.50	90.52
		3	89.64	89.66	86.56	93.55	87.94
	KNN	2	89.39	81.36	80.49	81.82	90.00
		3	72.00	69.70	67.69	83.65	68.55
Single cell	*Ensemble*	2	**99.07**	**99.80**	**99.50**	**98.90**	**99.35**
		3	**99.01**	**98.25**	**98.61**	**99.53**	**93.82**
	Bayesian	2	89.75	92.39	87.70	91.00	91.69
		3	83.23	76.30	80.23	91.35	77.97
	SVM	2	96.50	98.57	96.50	96.56	97.52
		3	94.92	93.06	93.03	96.97	93.04
	KNN	2	91.51	92.36	87.70	93.92	93.13
		3	79.09	72.60	73.05	87.62	72.78
Smear	*Ensemble*	3	**98.11**	**98.38**	**97.87**	**99.35**	**98.11**
	Bayesian	3	85.98	81.60	85.69	92.98	83.33
	SVM	3	96.14	95.16	95.09	97.83	95.09
	KNN	3	80.30	74.01	73.98	88.77	73.97

Bold values indicates the best performance of the classifier used for comparison.



(a) Cell Level classification



(b) Smear Level classification

Fig. 7 – Result of individual classifiers performance using 5 fold cross validation.

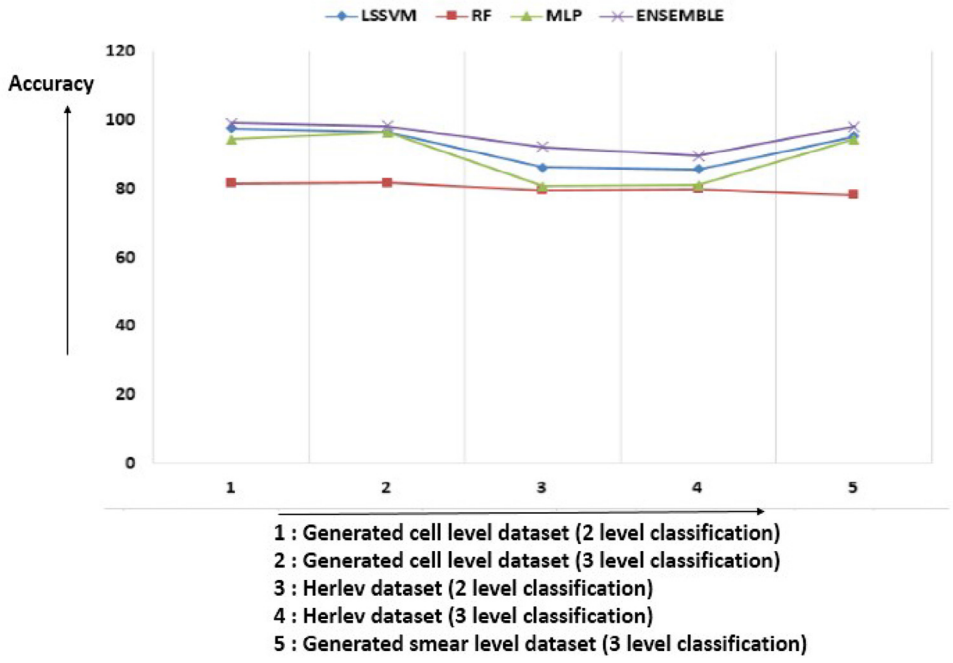


Fig. 8 – Comparison of ensemble classifier performance with LSSVM, RF and MLP using different databases.

of the image is obtained by fusing the outputs of the individual classifiers using weighted majority voting. Ensemble classifier gives better result as compared to three individual classifiers. All results indicate that the method that is proposed is very promising in detecting the degree of dysplasia present in cervical lesion from Pap smear images. This type of automated system can reduce the time consumed in manual observations, removes observer bias and improves efficiency.

Acknowledgements

This work has been funded by the INSPIRE Fellowship Scheme of DST, Govt. of India, awarded to the research scholars which is implemented through Institute of Advanced Study in Science and Technology (IASST), an autonomous institute under DST, Govt. of India. The authors would like to thank the funding authority and Director of implementing Institute. Malay K. Kundu

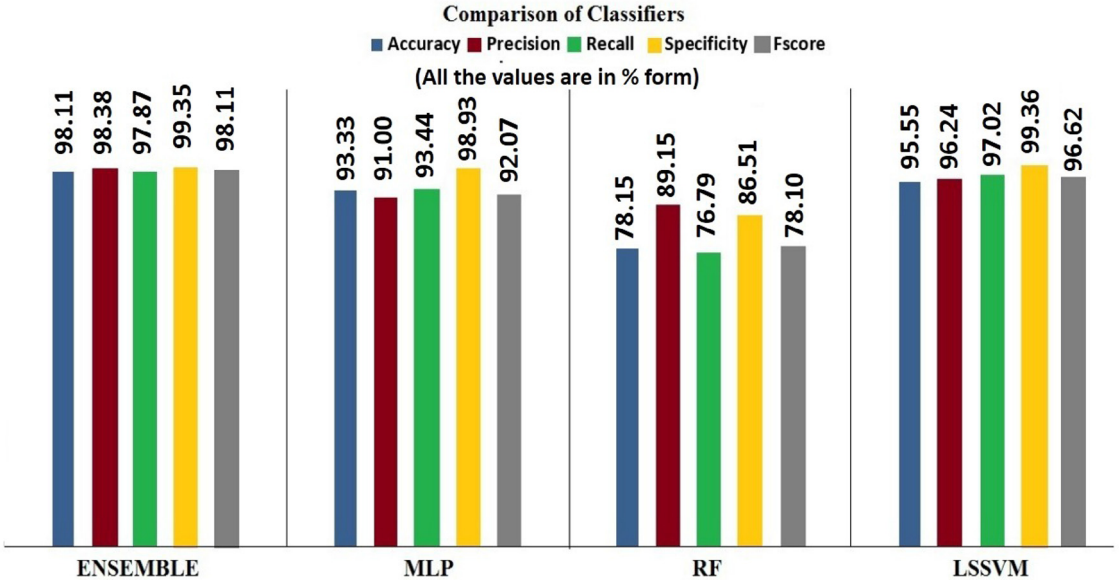


Fig. 9 – Output of different classifiers for smear level classification.

Table 6 – Comparison with existing approach.

	Approach 1 [19]	Approach 2 [13]	Proposed
	Herlev dataset		
Feature size	20	9	11
Database size	917	917	917
LSSVM	**92.11**	82.15	91.09
MLP	70.71	81.00	**81.71**
RF	74.00	**78.00**	75.25
SVM	93.41	**95.36**[a]	95.20
KNN	90.90	**94.99**[a]	89.39
BAYESIAN	**90.16**	89.38[a]	87.98
ENSEMBLE	93.75	91.63	**96.51**
	Generated single cell dataset		
Feature size	20	9	11
Database size	1610	1610	1610
LSSVM	94.18	83.27	**96.27**
MLP	92.54	90.17	**98.88**
RF	85.91	**86.72**	83.12
SVM	92.11	93.15	**96.50**
KNN	84.91	89.99	**91.51**
BAYESIAN	**90.70**	81.50	89.75
ENSEMBLE	97.23	95.33	**99.07**

[a] Values which are stated in the paper and bold values indicate the maximum accuracy.

acknowledges the Indian National Academy of Engineering (INAE) for their support through INAE Distinguished Professor fellowship. The authors would also like to thank Dr. J. D. Sarma, Department of Pathology, Dr. B. Borooah Cancer Institute, Guwahati, India for his valuable guidance. This work has the consent of all the co-authors and authorities of the institute, where this study has been carried out and there exists no conflict of interest anywhere. This study was approved by Ethical Committee for Human Studies of IASST with registration number ECR/248/Indt/AS/2015 of Rule 122D, Drugs and Cosmetic Rule, 1945. We would also like to thank the editor in chief, associate editors and the reviewers of the journal for the time spent in reading/handling the manuscript and for the useful feedback.

REFERENCES

[1] A. Labeit, F. Peinemann, A. Kedir, Cervical cancer screening service utilization in UK, Sci. Rep. 3 (1–9) (2013).
[2] M.H. Tsai, Y.K. Chan, Z.Z. Lin, S.F. Yang Mao, P.C. Huang, Nucleus and cytoplast contour detector of cervical smear images, Pattern Recognit. Lett. 29 (9) (2008) 1441–1453.
[3] S.J. McKenna, I.W. Ricketts, A.Y. Cairns, K.A. Hussein, Cell searching with a neural net, in: Third Iris Neural Networks Conference, The Queen's University of Belfast, Northern Ireland, 1993, pp. 205–215.
[4] K. Li, Z. Lu, W. Liu, J. Yin, Cytoplasm and nucleus segmentation in cervical smear images using radiating gvf snake, Pattern Recognit. 45 (4) (2012) 1255–1264.
[5] P. Bamford, B. Lovell, Unsupervised cell nucleus segmentation with active contours, Sig. Process. 71 (2) (1998) 203–213.
[6] P. Bamford, B. Lovell, A water immersion algorithm for cytological image segmentation, in: APRS Image Segmentation Workshop, Sydney, Australia, 1996, pp. 75–79.
[7] Z. Lu, G. Carneiro, A.P. Bardley, An improved joint optimization of multiple level set function for the segmentation of overlapping cervical cells, IEEE Trans. Image Process. 24 (4) (2015) 1261–1272.
[8] Z. Li, K. Najarian, Automated classification of pap smear tests using neural networks, in: Neural Networks, 2001 Proceedings, IJCNN'01, International Joint Conference on, IEEE, Washington, DC, 2001, pp. 2899–2901.
[9] E. Bak, K. Najarian, Brokway, efficient segmentation framework of cell images in noise environments, in: Proceedings of the 26th Annual International Conference of IEEE EMBS, IEEE, San Francisco, CA, USA, 2004, pp. 1–5.
[10] M.E. Plissiti, C. Nikou, A. Charchanti, Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering, IEEE Trans. Inf. Technol. Biomed. 15 (2) (2011) 233–241.
[11] M.E. Plissiti, C. Nikou, Combining shape, texture and intensity features for cell nuclei extraction in pap smear images, Pattern Recognit. Lett. 32 (6) (2011) 838–853.
[12] N. Lassouaoui, L. Hamami, N. Nouali, Morphological description of cervical cell images for the pathological recognition, Int. J. Med. Health Biomed. Bioeng. Pharm. Eng. 1 (5) (2007) 307–310.
[13] T. Chankong, N. Theera-Umpon, S. Auephanwiriyankul, Automatic cervical cell segmentation and classification in pap smears, Comput. Methods Programs Biomed. 113 (2) (2014) 539–556.
[14] M.E. Plissiti, C. Nikou, A. Charchanti, Watershed-based segmentation of cell nuclei boundaries in pap smear images, in: Information Technology and Applications in Biomedicine (ITAB), 10th IEEE International Conference on, IEEE, 2010, pp. 1–4.
[15] Y.F. Chen, P.C. Huang, H.H. Lin, L.E. Wang, Semi-automatic segmentation and classification of pap smear cells, IEEE J. Biomed. Health Inform. 18 (1) (2014) 94–108.
[16] A. Genctav, S. Aksoy, S. Onder, Unsupervised segmentation and classification of cervical cell images, Pattern Recognit. 45 (12) (2012) 4151–4168.
[17] T. Guan, D. Zhou, Y. Liu, A novel rgb Fourier transform-based color space for optical microscope image processing, Robot. Bioinform. 1 (16) (2014) 1–6.
[18] L.H. Camargo, G. Diaz, E. Romero, Pap smear cell image classification using global mpeg7 descriptor, Diagn. Pathol. 8 ((Suppl. 1):S38) (2013) 1–4.
[19] J. Jantzen, J. Norup, G. Dounias, B. Bjerregaard, Pap-smear Benchmark Data For Pattern Classification, NiSIS, 2005, pp. 1–9.
[20] W. Gray, Diagnostic Cytopathology, Churchill Livingstone, 2010.
[21] R. Polikor, Ensemble based system in decision making, IEEE Circ. Syst. Mag. 6 (3) (2006) 21–44.
[22] J. Matlas, O. Chum, M. Urban, T. Padla, Robust wide baseline stereo from maximally stable extremal regions, in: British Machine Vision Conference (BMVC), 2002, pp. 384–393.
[23] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, T. Kadir, L. VanGool, A comparison of affine region detectors, Int. J. Comput. Vis. 65 (1–2) (2005) 43–72.
[24] M. Donoser, H. Bischof, Efficient maximally stable extremal region (mser) tracking, in: Computer Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Conference on, IEEE, 2006.
[25] P. Malm, B.N. Balakrishnan, V.K. Sujathan, R. Kumar, E. Bengtsson, Debris removal in pap-smear images, Comput. Methods Programs Biomed. 3 (1) (2013) 128–138.
[26] M. Yang, K. Kpalma, J. Ronsin, Pattern Recognition, IN-TECH, 2008, Ch. A Survey of Shape Feature Extraction Techniques.

[27] R. Haralick, K. Shanmugam, I. Dinstein, Texture features for image classification, IEEE Trans. Syst. Man Cybern. 3 (6) (1973) 610–621.
[28] L.K. Soh, C. Tsatsouli, Texture analysis of sar sea ice imagery using gray level co-occurrence matrices, IEEE Trans. Geosci. Remote Sens. 37 (2) (1999) 780–795.
[29] J. Xu, L. Yang, D. Wu, Ripplet new transform for image processing, J. Vis. Commun. Image Represent. 21 (7) (2010) 627–639.
[30] M. Chowdhury, M.K. Kundu, Comparative assessment of efficiency for content based image retrieval systems using different wavelet features and pre-classifier, Multimed. Tools Appl. 74 (24) (2015) 11595–11630.
[31] M.N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, IEEE Trans. Image Process. 11 (2) (2002) 146–158.
[32] V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzas, An ensemble of filters and classifiers for microarray data classification, Pattern Recognit. 45 (1) (2012) 531–539.
[33] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, Y. Zhou, A novel ensemble method for classifying imbalanced data, Pattern Recognit. 48 (5) (2015) 1623–1637.
[34] J.J. Yang, J. Li, R. Shen, Y. Zeng, J. He, J. Bi, et al., Exploiting ensemble learning for automatic cataract detection and grading, Comput. Methods Programs Biomed. 124 (2016) 45–57.
[35] A. Ozcift, A. Gulten, Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms, Comput. Methods Programs Biomed. 104 (2016) 443–451.
[36] A. Betker, T. Szturm, Z. Moussavi, Application of feedforward backpropagation neural network to center of mass estimation for use in a clinical environment, in: Engineering in Medicine and Biology Society, 2003, vol. 3, Proceedings of the 25th Annual International Conference of the IEEE, 2003, pp. 2714–2717.
[37] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[38] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neur. Process. Lett. 9 (3) (1999) 293–300.
[39] M. Sokolova, G. Lapalme, A systematic analysis of performance measure for classification task, Inf. Process. Manag. 45 (4) (2009) 427–437.