



# Automated characterization and counting of Ki-67 protein for breast cancer prognosis: A quantitative immunohistochemistry approach

Tushar Mungle<sup>a</sup>, Suman Tewary<sup>a</sup>, Indu Arun<sup>b</sup>, Bijan Basak<sup>b</sup>,  
Sanjit Agarwal<sup>b</sup>, Rosina Ahmed<sup>b</sup>, Sanjoy Chatterjee<sup>b</sup>,  
Asok Kumar Maity<sup>c</sup>, Chandan Chakraborty<sup>a,\*</sup>

<sup>a</sup> School of Medical Science & Technology, IIT Kharagpur, Kharagpur, West Bengal, India

<sup>b</sup> Tata Medical Center, New Town, Rajarhat, Kolkata, West Bengal, India

<sup>c</sup> Midnapur Medical College and Hospital, Midnapur, West Bengal, India

## ARTICLE INFO

### Article history:

Received 5 October 2015

Received in revised form

21 October 2016

Accepted 3 November 2016

### Keywords:

Breast cancer

Immunohistochemistry

Ki-67 protein

Fuzzy C-means

k-means

Proliferation index

## ABSTRACT

Ki-67 protein expression plays an important role in predicting the proliferative status of tumour cells and deciding the future course of therapy in breast cancer. Immunohistochemical (IHC) determination of Ki-67 score or labelling index, by estimating the fraction of Ki67 positively stained tumour cells, is the most widely practiced method to assess tumour proliferation (Dowsett et al. 2011). Accurate manual counting of these cells (specifically nuclei) due to complex and dense distribution of cells, therefore, becomes critical and presents a major challenge to pathologists. In this paper, we suggest a hybrid clustering algorithm to quantify the proliferative index of breast cancer cells based on automated counting of Ki-67 nuclei. The proposed methodology initially pre-processes the IHC images of Ki-67 stained slides of breast cancer. The RGB images are converted to grey, L\*a\*b\*, HSI, YCbCr, YIQ and XYZ colour space. All the stained cells are then characterized by two stage segmentation process. Fuzzy C-means quantifies all the stained cells as one cluster. The blue channel of the first stage output is given as input to k-means algorithm, which provides separate cluster for Ki-67 positive and negative cells. The count of positive and negative nuclei is used to calculate the F-measure for each colour space. A comparative study of our work with the expert opinion is studied to evaluate the error rate. The positive and negative nuclei detection results for all colour spaces are compared with the ground truth for validation and F-measure is calculated. The F-measure for L\*a\*b\* colour space (0.8847) provides the best statistical result as compared to grey, HSI, YCbCr, YIQ and XYZ colour space. Further, a study is carried out to count nuclei manually and automatically from the proposed algorithm with an average error rate of 6.84% which is significant. The study provides an automated count of positive and negative nuclei using L\*a\*b\* colour space and hybrid segmentation technique. Computerized evaluation of proliferation index can aid pathologist in assessing breast cancer severity. The proposed methodology, further, has the potential advantage of saving time and assisting in decision making over the present manual procedure and could evolve as an assistive pathological decision support system.

© 2016 Elsevier Ireland Ltd. All rights reserved.

\* Corresponding author. School of Medical Science & Technology, IIT Kharagpur, Kharagpur, 721302, India. Fax: 91 3222 282221.  
E-mail address: [chandanc@smst.iitkgp.ernet.in](mailto:chandanc@smst.iitkgp.ernet.in) (C. Chakraborty).

<http://dx.doi.org/10.1016/j.cmpb.2016.11.002>

0169-2607/© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

Breast cancer, one of the most common cancers among the women, has become a rising cause of concern owing to its increased mortality rate. It accounts for 22.2% of all new cancer diagnoses and 17.2% of all cancer deaths occurring in India [1]. The increasing incidence of breast cancer draws attention to the treatment efficacy which relies primarily on its early diagnosis. In breast cancer management, clinicopathologic parameters like tumour size, histological type, Nottingham grade and lymph node status are traditionally used to determine outcome of the disease. However, with advances in molecular studies, it became evident that breast cancer is a heterogeneous disease and breast cancers presenting at the same stage or of the same histological type may not have the same underlying biology or clinical behaviour. Since multi-parameter molecular assays are expensive and not widely available, current management of breast cancer takes into account, in addition to the traditional clinicopathological parameters, testing for IHC expression of three established biomarkers estrogen receptor (ER) and progesterone receptor (PR) and Human Epidermal Growth Factor 2 (HER2) Receptor along with Ki-67 labelling index as a surrogate of molecular classification of breast cancer into luminal types A and B, Her-2 enriched and triple negative breast cancer based on which important treatment decisions are made [2,3]. Ki-67 is a nuclear protein expressed in all phases of cell cycle except G0 phase and has proven to be a prognostic and predictive marker in breast cancer [4]. High values of Ki-67 scores predict the benefit of addition of cytotoxic chemotherapy in breast cancer. Hence, Ki-67 score in a hormone receptor positive disease is used to differentiate luminal B type of breast cancer with high Ki-67 scores, which may benefit from chemotherapy from luminal A type breast cancer which have low proliferation and is strongly endocrine responsive [3]. Manual scoring techniques used for assessing the expression of these biomarkers are time consuming and there is substantial variability in the results. Integration of hardware (microscope) and software (analysis software) for imaging and automatic counting of Ki67 nuclei and others biomarkers have evolved over a period of time. Slidepath Tissue IA system and Aperio Scanscope by Leica Biosystems are used in several studies for image analysis and evaluation [5–7]. Affordability of these systems in rural places of India becomes very tough due to financial constraints. Thus, having low cost affordable image analysis software to assist pathologists to determine the Ki-67 labelling index and help the clinicians in treatment decision making process becomes crucial. In view of this, the proposed work is an essential initial step towards making a potentially employable and affordable system.

Analysis of the results for Ki-67 staining is challenging, mainly due to the lack of uniformity and consistency among scoring standards. In addition to this, the inter-observer variability problem remains a hurdle [8,9]. The length and tedium of manual cell counting are major impediments in correctly estimating the number of proliferative cells. Various digital image analysis methods [10–13] have been proposed to address this problem. The three broad steps followed in image analysis

are image pre-processing, cell segmentation, and post processing.

One of the motivations of our study and an important aspect in image analysis is the selection of a colour space as there is no mutual belief in the image processing community to which colour space is the best suited for respective application. The preference of one colour space over the other is a very subjective issue depending upon the application. Choosing a colour space requires an extensive comparative study of effectiveness and accuracy of the end result. Studies have described the use of particular colour space depending upon the application relevance. Further details regarding colour spaces and image segmentation can be found in Refs. [14,15] and has been also described in this study under section 2.2. Five colour transformations namely, normalized RGB, HSV, I1I2I3, L\*a\*b\*, and YCbCr performances have been evaluated via Support Vector Machine (SVM), C4.5, and Radial Basis Function (RBF) neural network to classify pizza toppings [16]. The work showed SVM having best classification accuracy with HSV, while RBF-NN with YCbCr, HSV or L\*a\*b\* and C4.5 with L\*a\*b\* and I1I2I3; finally confirming SVM along with HSV being the best. An efficient face detector using skin colour detection under a certain condition has been accomplished by using ten colour spaces along with K-means for classification. At pixel level, a quantitative analysis was done by considering true positive, false positives and false negatives [17] where in HSV colour space won over the rest. Ruiz-Ruiz et al. [18] used hue to propose Hue-Saturation (HS) and Hue (H) as a new colour space to improve the environmentally adaptive segmentation algorithm in terms of time complexity. In Ref. [19], spectral clustering has been used with different colour spaces to select the colour space best suited for the application, which was then followed by pixel labelling. The study in Ref. [20] involves the use of HSV, CMY, YUV and RGB colour spaces for clustering based segmentation. The authors showed that the CMY colour space is appropriate for ignorance based clustering and entropy based clustering techniques. A comparative study of RGB, HSV, CMY, XYZ and YUV colour space models based on automatic GrabCut colour segmentation technique is given in Ref. [21]; stating RGB to be best for segmentation. Niazi et al. have used CIE L\*a\*b\* colour space for entropy based quantification of Ki67 nuclei [22]. The discussed bibliography suggests the conflicting closure about the pertinence of different colour spaces available in the context of image segmentation.

Another significant aspect in image analysis and motivation for our study is to emphasize on image segmentation. Segmentation—a technique for cell detection—is applied to digitized images where different regions of interest are obtained depending on their application [12,14,23–27]. Some other widely and commonly used segmentation techniques include watershed-based, edge detection, clustering, thresholding etc. Of all these, clustering is one of the most widely used technique due to its simplicity, and is applied to segment cells in different modalities. K-means clustering [28] based image segmentation applications can be seen in Refs. [29–31]. Fuzzy C-means [32] is the generalization of k-means clustering, which is mainly used for fuzzy image segmentation [33–35]. Manual segmentation which is routinely carried out by pathologists is a tedious and time consuming task and thus calls for high throughput automatic segmentation methods which can

expedite the process. However, given the complexity of cellular structures, accurate segmentation becomes difficult and serves as a challenge in image processing.

Our work aims at providing a more accurate and less time-consuming diagnostic method to pathologists for quantification of Ki-67 expression. The proposed method first classifies the Ki-67 positive and negative nuclei via hybrid segmentation technique, and thereafter counts the number of segmented nuclei to give the appropriate score. In the present methodology, colour transformation is carried out, where the original image in the RGB colour space is transformed into various other colour spaces. We use images having various colour variations, obtained from an in-house histology digital image database, to determine which colour space model best captures relevant pathological artefacts. Morphological operations are then applied to polish the segmented nuclei followed by filtering out small undesired regions. Further, statistics are obtained by comparing the results of segmentation which helped us to decide on the optimal colour space for our analysis. The total number of positive and negative nuclei is calculated using a labelling algorithm. And finally, average error rate is computed from manually and automatically computed Ki-67 scores for the images.

## 2. Material and methods

In this section, we describe the procedure followed to digitize the slides with the help of inbuilt software used with microscope and the motivation behind the usage of different colour spaces followed by the proposed methodology.

### 2.1. IHC image acquisition

Ki-67 stained invasive breast cancer tissue slides were collected from two eminent hospitals viz., Tata Medical Centre (TMC), Kolkata and Midnapur Medical College and Hospital, West Bengal, India. The study has been approved by hospital administration for data collection [Study Protocol Ref. No.

EC/GOVT/07/14]. Additionally, institutional ethical clearance was taken from the Indian Institute of Technology Kharagpur, India [Ref. No. IIT/SRIC/SAO/2015] for data handling and computational works. The patients diagnosed with breast cancer were notified about the study research and informed consent was obtained from the patients at the hospital upon their willingness. Diagnostic core biopsies of breast cancer diagnosed patients regardless of age, grade or stage of disease were included in the study. Also, counting was done only in the invasive carcinoma component. Carcinoma in-situ component and stromal cells which may be positive for Ki-67 IHC were excluded from counting. For the current study, we obtained a total of 47 patient's slides during August 2014 to June 2015. All slides were prepared and stained using standardized techniques in an accredited laboratory, and were assessed by a trained and experienced pathologist. Leica DM750 microscope with camera model ICC50 is used for the study. Digitized images are grabbed for the study using LAS EZ image acquisition software accompanied with the microscope.

During all image acquisition procedure, we calibrated our image acquisition system for normalizing the image. A blank glass slide was observed under microscope where background was calibrated. With respect to this background intensity, normalized images were captured. The slides are initially scanned at a magnification of 10× to identify regions of positive stained nuclei (i.e. hotspots—area where positive stained nuclei population is evident). The selection of region of interest or hotspots varies from individual to individual. The standard norm is to select or identify at least three different regions. Each region is then examined under 40× lens magnification with 10× eyepiece producing effective magnification of 400×. The camera was set in auto exposure settings with gamma, gain, and exposure values optimized automatically for the grabbed images. Tagged Image File Format (tiff) without compression format of the image is used for the study. The size of image is 2048 × 1536 pixels with 1 pixel = 0.064 μm. A total of 215 images are obtained from the 47 patients slides. Fig. 1 shows sample images being used in the study. The images contain two types of stained nuclei: the brown nucleus indicating Ki-67 positively-stained cells

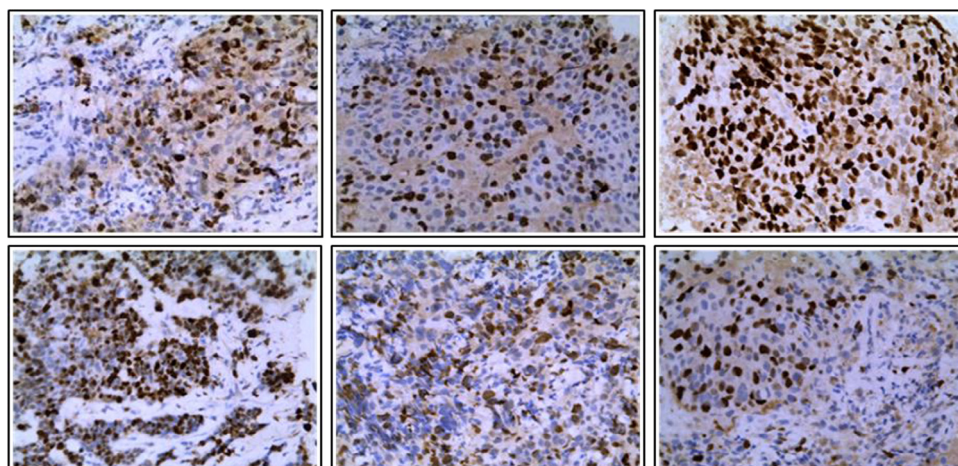


Fig. 1 – Sample Ki-67 stained histology sample images at 40× magnification.



and the blue nucleus showing Ki-67 negatively-stained cells. The pathologist utilizes this information to study the proliferation rate of tumour in the patients.

## 2.2. Colour space models

A colour space model is a mathematical model that defines organization of colours that in turn dictates the range of colours that can be represented using the model. The specification of these colours is based on human eye perception. Different colour space models are used to characterize the colours in digital images [14] which aid in segmentation. According to their characteristics, different colour space families, viz. *primary spaces*, *luminance-chrominance spaces*, *perceptual spaces* and *independent axis spaces* [15], are linear/nonlinear transformations of the RGB colour space. The most common colour space used for images is RGB, wherein at the pixel level, the colour components used are red (R), green (G) and blue (B). In our study, we make use of six different colour space models, namely RGB, XYZ, CIE  $L^*a^*b^*$ , HSI, YCbCr, and YIQ (NTSC), which are described in the following sections.

### 2.2.1. RGB

Red, green and blue are three basic primary colours that the human eye can perceive and therefore forms one of the primary colour spaces, denoted as (R, G, B). The additive mixing of these colours can be used to produce any available colour in the visible spectrum. The common usage of this colour space can be seen in photography, television, scanners, cameras etc. This model uses *light (luminance)* along with chrominance as a means to produce the desired colour. Due to this undesired dependency on luminance, one can make use of the luminance independent component commonly known as the normalized component, and is defined below:

$$n_r = \frac{R}{R+G+B} \quad (1)$$

$$n_g = \frac{G}{R+G+B} \quad (2)$$

$$n_b = \frac{B}{R+G+B} \quad (3)$$

$n_r$ ,  $n_g$ , and  $n_b$  are respectively the normalized values representing the chromatic components red, green and blue.

### 2.2.2. XYZ

To overcome the shortcomings of the RGB colour space, CIE (Commission International de l'Eclairage) proposed the XYZ colour space [15]. This device-independent colour space addresses the perceptual uniformity and human observer psychophysical need. Three basic primaries defined here are X, Y, and Z which can represent a mix of cone response curves, luminance, and quasi-equal to blue stimulation respectively, and is computed as a linear transformation of the RGB space as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

### 2.2.3. CIE $L^*a^*b^*$

CIE developed the CIE  $L^*a^*b^*$  colour space overcoming the drawbacks of CIE XYZ colour space by means of a non-linear transformation. This model dissociates the luminance component from the chrominance information and encompasses the complete colour gamut beyond the visible one. Application wise, this model encompasses a wide range from colourimetry, printing, photography, dyes, printing ink and paper, paint etc. The following is the conversion of the XYZ to  $L^*a^*b^*$  [15]:

$$L^* = \begin{cases} 116 * \sqrt[3]{\frac{Y}{Y^w}} - 16 & \text{if } \frac{Y}{Y^w} > 0.008856 \\ \frac{903.3 * Y}{Y^w} & \text{if } \frac{Y}{Y^w} \leq 0.008856 \end{cases} \quad (5)$$

where  $X^w$ ,  $Y^w$ , and  $Z^w$  are the tri stimulus values of the reference white.

$$a^* = 500 * \left( f\left(\frac{X}{X^w}\right) - f\left(\frac{Y}{Y^w}\right) \right) \quad (6)$$

$$b^* = 200 * \left( f\left(\frac{Y}{Y^w}\right) - f\left(\frac{Z}{Z^w}\right) \right) \quad (7)$$

$$\text{where } f(x) = \begin{cases} \sqrt[3]{x} & \text{if } x > 0.008856 \\ 7.787x + 0.138 & \text{if } x \leq 0.008856 \end{cases}$$

### 2.2.4. HSI

Human eye perception for RGB colour space is not so evident in practical scenarios, and goes in accordance with *hue*, *saturation* and *intensity*; hence HSI colour space. HSI colour space quantifies the colour information according to perceptual data. *Hue (H)* represents colour type or perceived colour; *Saturation (S)* describes the amount of white mixed with hue or purity of colour; and *Intensity (I)* expresses the luminance, i.e. the amount of light or brightness of the colour. The RGB to HSI transformation can be achieved using the following equations

$$H = \cos^{-1} \left[ \frac{0.5[(R-G) + (R+B)]}{\sqrt{(R-B)(G-B) + (R-G)^2}} \right] \quad (8)$$

$$S = 1 - \frac{3}{R+G+B} [\min(R+G+B)] \quad (9)$$

$$I = \frac{1}{3}(R+G+B) \quad (10)$$

### 2.2.5. YCbCr

In the YCbCr model, Y represents luminance information and Cb/Cr denotes the difference between blue/red colour and a reference value. Mostly used in video and image

processing, this colour space is an encoding of the RGB colour space in which colour depends solely on the RGB primaries. One reason why YCbCr is preferred over RGB in some applications is the luminance independent property which yields better results. The transformation from RGB to YCbCr is expressed as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (11)$$

Note that luminance and chrominance are separate components.

### 2.2.6. YIQ (NTSC)

The NTSC (National Television System Committee) colour TV system is implemented using the YIQ colour space [15]. It consists of luminance (represented by Y) and chrominance information (represented by I and Q). The advantage of YIQ is its capability to separate grey scale information from colour data which helps in representing the same image/signal for both black-and-white and colour. The linear transformation to convert RGB to YIQ is as follows:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (12)$$

The motivation behind experimenting with the use of different colour models for our work is to find the colour model best suited to deal with the colour disparity in the images obtained from the database of breast cancer patients developed by Biostatistics and Medical Informatics (BMI), IIT Kharagpur lab members in collaboration with TMC. The clustering problem of blue and brown nuclei adds to the uncertainty of the choice. Therefore, it would be good to estimate a reliable and optimal colour space that is tuned to achieve the desired accuracy for our proposed hybrid segmentation algorithm.

## 2.3. Proposed methodology

This section describes the justifications of the colour space selection problem presented in the previous section. It also describes various steps of the proposed hybrid clustering based segmentation technique followed by post processing to clean up the final output.

### 2.3.1. Colour space conversion

Digital histopathology microscopic RGB images were converted into six colour spaces, viz. grey, CIE  $L^*a^*b^*$ , HSI, YCbCr, YIQ and XYZ. The output images, in the various colour models, are then given as an input to the hybrid clustering segmentation method described in the next section.

### 2.3.2. Hybrid clustering based segmentation

The Hybrid Clustering Algorithm is presented in two stages. Stage 1 uses a Fuzzy C-means clustering technique which

uses a thresholding technique to eliminate spurious information. Stage 2 then uses a K-Means clustering technique on the result to identify the Ki-67 positive and negative nuclei. The two stages of the Hybrid Clustering Algorithm are elaborated below:

**2.3.2.1. Stage 1.** The input image is the single channel 2D image. Fuzzy C-means (FCM) clustering is applied on the image data to segment out the probable region of interest which is masked from the original image containing the positive and negative nuclei and some low contrast region. In FCM, fuzzy membership is associated with every pixel. Let  $I_{m \times n} = \{I(i,j), 0 \leq i, j \leq m, n\}$  be an image of dimension  $m \times n$  which is to be partitioned into “c” clusters. The cost function defined for the iterative optimization is defined by

$$J = \sum_{j=1}^{m \times n} \sum_{i=1}^c (\mu_{ij}^{m'}) \|I_{i,j} - V_i\|^2 \quad (13)$$

where  $\mu_{ij}$  represents the membership of the pixel  $j$ th for  $i$ th cluster. The parameter  $m'$  is the weighing exponent controlling the fuzziness of the partition in question. In this work, we set  $m' = 2$ .  $V = \{v_1, v_2, \dots, v_c\}$  are the cluster centres. Since, the region of interest is Ki-67 positive and negative nuclei only, we used two clusters ( $c = 2$ ). FCM has high or low thresholding and in our study, we have selected the higher threshold to segment the image into two clusters. The first cluster gives background and low contrast region and the second cluster produces Ki67 positive and negative nuclei. The membership function signifies the probability of a pixel it may belong to. The main idea behind this probability is its dependence on the distance between each pixel and cluster centres. Each iteration updates the membership function value and cluster centre as

$$\mu_{ij} = 1 / \sum_{k=1}^c \left\{ \frac{\|I_{ij} - V_i\|}{\|I_{ij} - V_k\|} \right\}^{(2/m'-1)} \quad (14)$$

and

$$V_i = \frac{\sum_{j=1}^{m \times n} \mu_{ij}^{m'} I_{ij}}{\sum_{j=1}^{m \times n} \mu_{ij}^{m'}} \quad (15)$$

The convergence of the algorithm is ensured by equating the two consecutive values for the cost of membership values or cluster centres. The advantage of using FCM at the first stage is to remove the low contrast region below threshold value. This results in extracting the region of interest along with few low contrast regions.

**2.3.2.2. Stage 2.** The second stage now has much less divergence in the blue and brown colour spectrum giving more promising nuclei characterization results. The blue channel of the image created from the previous step is used to further cluster the image regions into positive nuclei, negative nuclei, and low contrast regions by using K-Means clustering. Finally, the images are reconstructed from the output of the K-Means clustering. The K-Means clustering algorithm works as follows.

### K- means Algorithm

**Step1.** Initialize cluster centres from data points (Image  $I_{m,n}$ ) randomly.  $C = \{c_1, c_2, \dots, c_k\}$

**Step2.** Calculate the distance between data point and cluster centre, and based on shortest distance place the data point into respective cluster.

**Step3.** Recalculate the cluster centres based on the above clusters formed.

**Step4.** Repeat steps 2 and 3 until the algorithm converges to give same value for cluster centres in two consecutive steps.

The membership function used to calculate the distance is as follows

$$J = \sum_{j=1}^{m \times n} \sum_{k=1}^l \|I_j - \mu_k\|^2 \quad (16)$$

where  $I_j$  is the pixel in the image and  $\mu_k$  is the cluster centre for  $k$ th cluster (mean of all data points in the  $k$ th cluster).  $J$  at each iteration reaches a minimum with a guarantee that eventually it will reach a minimum which will not further reduce. The value of  $k$  depends upon the user and in our case upper value of  $k$  i.e.  $l = 3$ , as the histology images roughly consist of three regions: brown, blue, and rest (stain marks and low contrast region).

Fig. 2 presents a pictorial representation of the proposed hybrid clustering segmentation method.

#### 2.3.3. Post processing

The need of improving the quality of segmented images in order to assess the quantitative analysis can be addressed with the help of mathematical morphological operations in image processing. Morphological operations, as discussed in Ref. [14], offer a way to post process (eliminating shape abnormalities) the digital images by tailoring the structuring element in addition to preserving the objects in context. In our case, segmented nuclei output achieved by means of the segmentation method contains some unwanted artefacts. Various operations such as dilation, erosion, opening/closing, region filling were used on the output binary image.

The segmented Ki-67 positive nuclei contain some minor artefacts which are removed with the help of area opening operation with area below 50 pixels. The area was decided by calculating the average area of artefacts for the test images. In case of Ki-67 negative nuclei, erosion, area-opening, and dilation followed by region-filling were performed sequentially. This series of operations was carried out because the background and blue nuclei region colours fall under the same blue-colour spectrum which led to the presence of a low level region.

Once, the cells are obtained, we need to differentiate cancerous and stromal cells. Stromal cells are elongated and tubular in structure which are different from circular cancerous cells. For this, we calculate the structural properties of nuclei like

aspect ratio and circularity. We consider the threshold value of aspect ratio and circularity as 0.4 and 0.7 respectively.

#### 2.3.4. Proliferation rate

Proliferation Index is defined as the percentage of positively stained cells to total number of malignant cells counted [4] and calculated as

$$\text{Proliferation Rate} = \frac{\text{Total number of positive nuclei}}{\text{Total number of nuclei}} \quad (17)$$

The ratio is calculated by counting the positive and negative nuclei with the help of the labelling algorithm [36]. A connected component based approach is chosen for labelling and counting the nuclei. Eight neighbourhood connected components select the detected blob based on the neighbourhood such that the separate blobs are labelled spatially.

The overall proposed methodology with sub-stage results are shown in Fig. 3.

Along with the segmentation of nuclei and proliferation index calculation, ground truth was generated with the help of collaborating hospitals' pathologists. The results were compared for correction and robustness of the algorithm which is discussed in the following section.

## 3. Results and discussion

This work recognizes the time consuming and lengthy evaluation process of calculating the proliferation index manually and proposes a new technique based on a two-stage hybrid approach that better characterizes Ki-67 nuclei. It has been observed that the methodologies followed in state of the art techniques [31,33] produce extra spurious results. These methodologies use single stage clustering, which does not correctly identify positive and negative Ki-67 nuclei. The segmented region produced covers the nucleus as well as the background. This in turn affects the overall accuracy in determining the proliferation index. Fig. 4, below, depicts the results of using the methods discussed in Ref. [31]. Similar results are found using the methods proposed in Ref. [33].

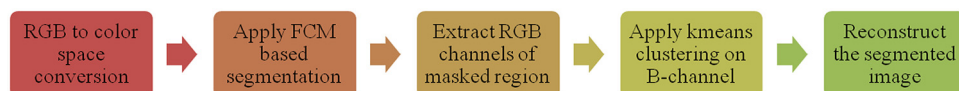
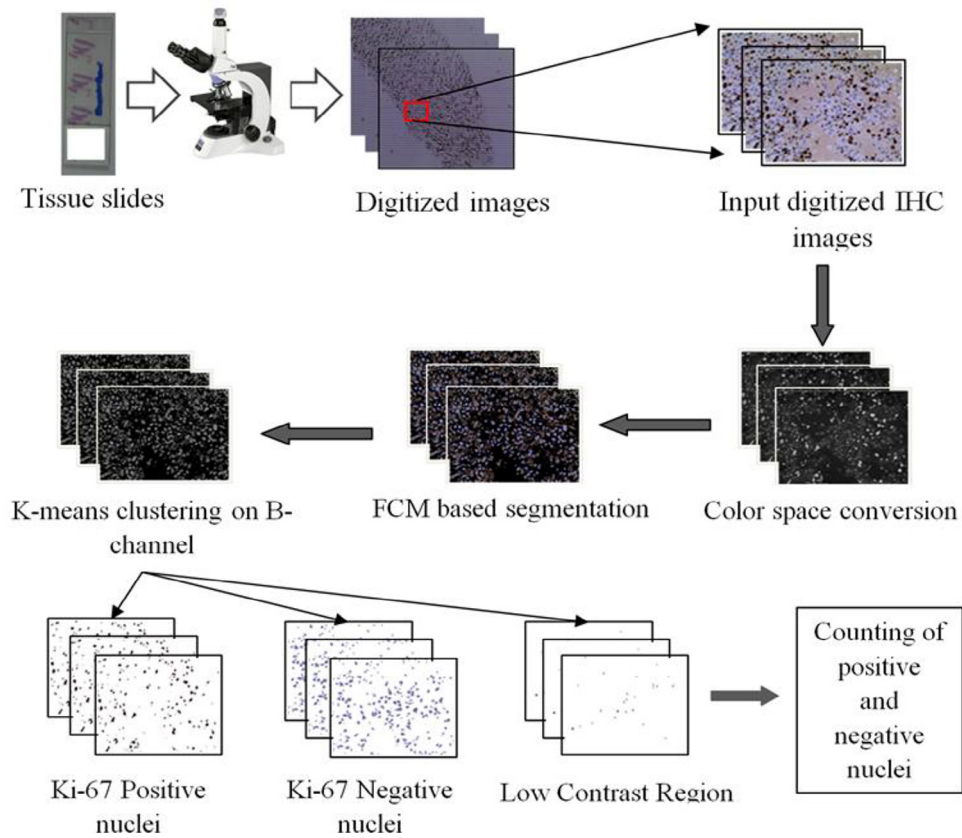


Fig. 2 – Proposed hybrid clustering segmentation method.



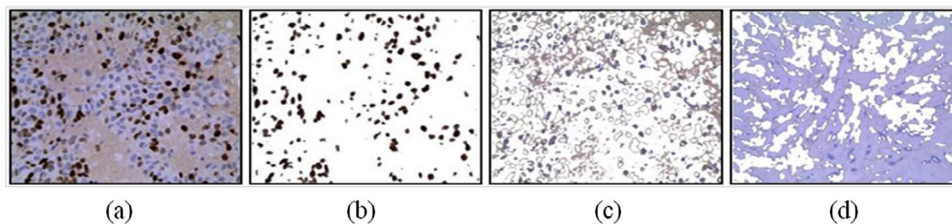
**Fig. 3 – Overall computational framework of the proposed methodology.**

This motivates us to look for alternative techniques to better distinguish between the types of cells. This work proposes the use of a hybrid clustering algorithm for the segmentation of positive and negative Ki-67 nuclei. The proposed methodology addresses the problems encountered in Refs. [31,33] with an aim to provide better results. Moreover, the algorithm proposed does not require any extra parameter to feed into the software apart from the image to generate the output. The widely used web based software ImmunoRatio [10] is used to visually get the analysis. However, it cannot be compared since it involves manual intervention of selecting approximate cell dimension. Although, we give the results from the ImmunoRatio software for one sample image in Fig. 5 below. The software asks for the input as size of nuclei. Fig. 5 shows segmentation results which are computed by selecting different sizes of nuclei. For Fig. 5(c), the size of nuclei was chosen to be much larger than the normal size producing wrong segmentation

results. The Fig. 5(c) shows very few nuclei segmented out, which in turn would affect the Ki67 proliferation rate. However, the quantitative results comparing proposed methodology and ImmunoRatio software along with manual counting are described in Table 5 for robustness of the algorithm.

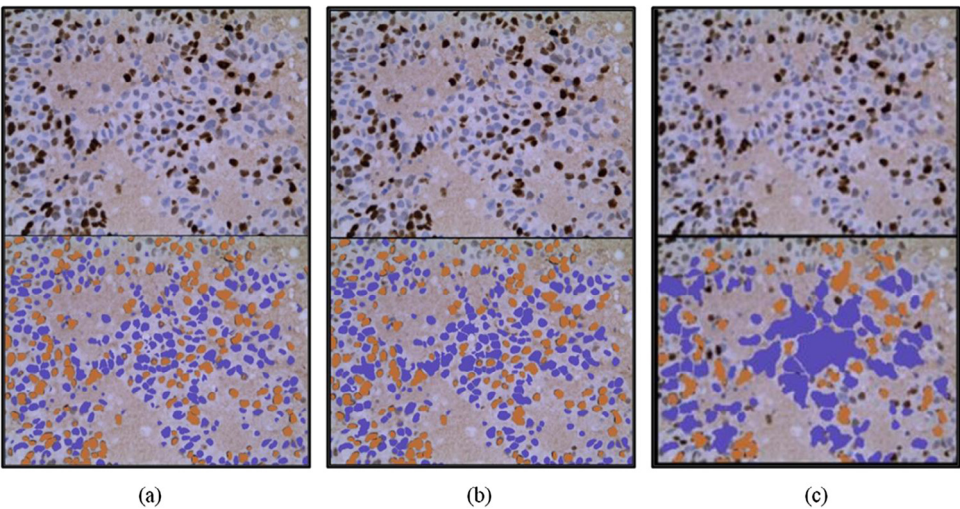
The development environment for the proposed algorithm is MATLAB™ 2014b version using the DELL™ Optiplex™ 990 system with 2nd generation Intel™ core i5 processor and 4 GB RAM. The approaches mentioned in some of the literatures focuses on working on a different colour space than RGB. The RGB colour space is transformed into various colour spaces for a given input image. The presence of appropriate ground-truth observations assisted us in the preliminary validation of output data in the form of comparisons. The result for one image is shown in Fig. 6.

Additionally, Ki-67 immunohistochemical images were collected from Midnapur Medical College and Hospital, India to



**Fig. 4 – Experimental results for state of art in Ref. [31]. (a) Original Image. (b, c) Segmented Ki-67 positive and negative nuclei respectively. (d) Low contrast region.**





**Fig. 5 – Results from ImmunoRatio [10] software. (a) Selection of normal (average) size of nuclei to produce the segmentation result; (b) Selection of larger (above average) size of nuclei to produce the segmentation result; (c) Selection of big size of nuclei to produce the segmentation result.**

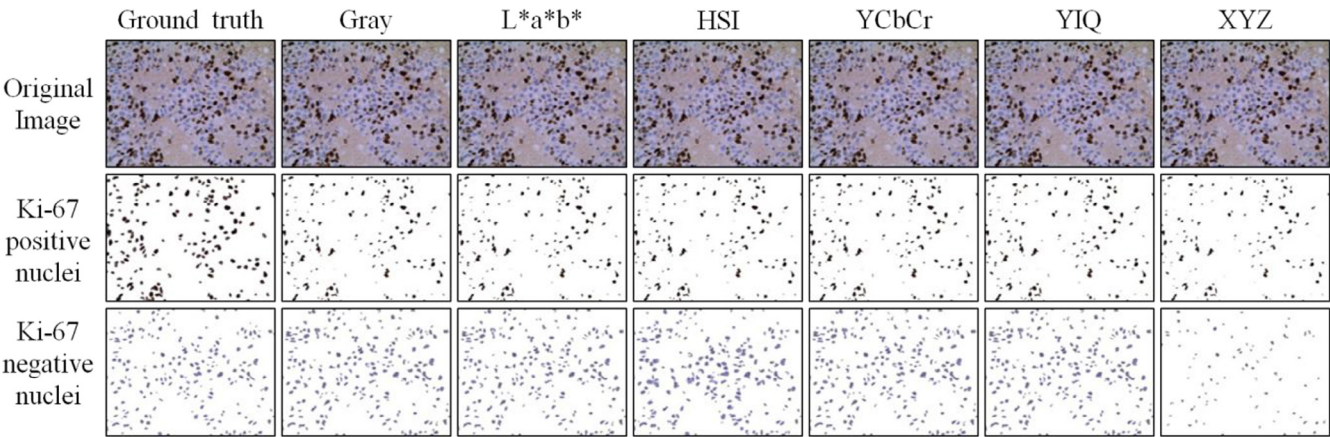
validate the proposed methodology and the segmented results are cited in Fig. 7. Varying colour intensity in molecular images due to different resources viz., hospitals/laboratories/clinics etc. leads to the subjective visualization of nuclei. It can be observed in Fig. 7 that image index-1 shows more of a bluish background, whereas image index-3 is more of a brownish background. In these two different situations, the proposed methodology works as:

- (a) The difference between the background and nuclei is significant irrespective of staining. Thresholding would remove all the background details.
- (b) Once the threshold image is obtained, considering the blue channel, it becomes easy for k-means clustering to separate out blue and brown nuclei along with the low contrast region.

The results similar to Fig. 6 were experimentally obtained for five images and used to count the Ki-67 positive and

negative nuclei using labelling algorithm, which led to statistical evaluation of ground truth and the segmented output. Tables 1 and 2 below show the counting of brown and blue nuclei with respect to different colour spaces.

Statistical evaluation of the above results is performed to choose an optimal colour space. In statistics, the confusion matrix is a reliable tool that can be used to calculate performance measures and assess accuracy [37,38]. The correctness of the experiments is carried out based on confusion matrix values. True positive (TP) is defined as the identification of the expected results in comparison to Ground truth (GT). GT is the plain expected absolute result desired as outcome of the experiment. False negative (FN) is the value of output present in the ground truth not observed in the test. False positive (FP) is defined as the results detected in the experiments which are undesired i.e. absent in the ground truth. Based on the confusion matrix terms, we calculate these values for the above calculated number of nuclei. Tables 3 and 4 show the only values of F-measure [37] calculated with the help of GP, TP, FN, and FP.



**Fig. 6 – Detection of Ki-67 positive and negative nuclei based on proposed methodology using different colour space.**



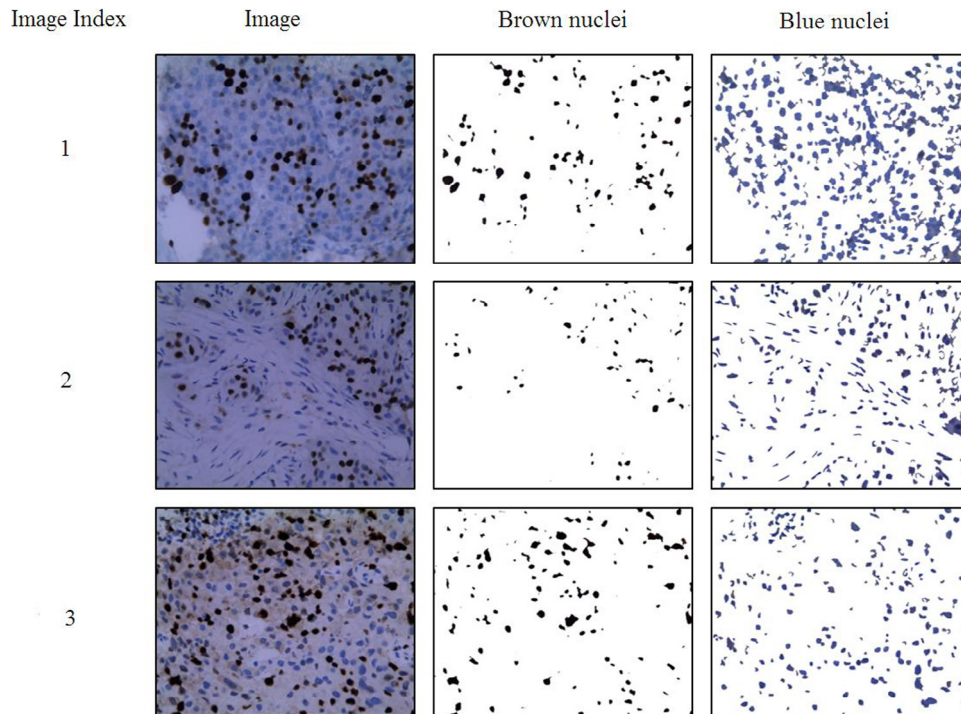


Fig. 7 – Segmentation results of images obtained from second lab.

Considering F-measure from the above results and the bar plot plotted (with the help of Tables 1 and 2) below in Fig. 8, we conclude that for the brown nuclei count, the use of  $L^*a^*b^*$  and HSI shows more promising results over the others. However, we observe that the average F-measures of  $L^*a^*b^*$ , HIS, YIQ and YcbCr are equivalent within an acceptance tolerance; therefore, the use of either of these colour spaces can be considered. Similarly, for blue nuclei, the average F-measure of YIQ and  $L^*a^*b^*$  leads over all others. Therefore, we use  $L^*a^*b^*$  (having the highest overall average F-measure, 0.8847) as the optimal

colour space for analysing histopathology images. From this observation, we can conclude that to segment both positive and negative nuclei, the  $L^*a^*b^*$  colour space is the optimal choice. Hence we use the results of segmentation using the  $L^*a^*b^*$  colour space to calculate the proliferation rate.

We chose 30 random patient sample images to test our algorithm. The calculation is done manually by an expert pathologist also to calculate the average error rate. Table 5 shows the calculation of proliferation rate which is the fraction of Ki-67 positive nuclei among all nuclei. The nuclei count from the analysis using  $L^*a^*b^*$  colour space is presented below.

The Auto Counting average error ( $AC_{Error}$ ) in calculating proliferation rate is found to be 6.84% between manual and proposed methodology. The results in column  $AC_{Error}$  of Table 5 depicts that deviation with respect to manual counting results is due to different natures of images. Difference in proliferation index range above approximately 10% is mainly due to presence of stromal cells in large number. The dilemma in excluding stromal cells with respect to cell structure is one of the major factors. In these cases the population of stromal cells is substantial when compared to cancerous cells which makes it difficult and dubious for an expert and for software to differentiate these two types of cells. The average errors  $AC_{Error}$  and  $IR_{Error}$  cannot be compared as, in ImmunoRatio, selection of cell size, manually, may lead to different results for same image. However, we have selected appropriate cell size in ImmunoRatio software to evaluate the proliferation index. The average error rate in ImmunoRatio software is higher and it may increase further with improper cell size selection which limits the usage of the software.

In India, especially in the rural areas, hospitals have restricted facilities; hence, affordable tools with automated cell counting in histology images could be applied in regular

Table 1 – Ki-67 positive nuclei count for different colour spaces

|         | Ground truth | Grey | $L^*a^*b^*$ | HSI | YcbCr | YIQ | XYZ |
|---------|--------------|------|-------------|-----|-------|-----|-----|
| Image 1 | 158          | 141  | 141         | 143 | 141   | 141 | 123 |
| Image 2 | 172          | 159  | 159         | 160 | 159   | 159 | 141 |
| Image 3 | 197          | 186  | 182         | 176 | 186   | 186 | 162 |
| Image 4 | 230          | 202  | 202         | 199 | 202   | 202 | 194 |
| Image 5 | 212          | 195  | 185         | 197 | 195   | 195 | 185 |

Table 2 – Ki-67 negative nuclei count for different colour spaces

|         | Ground truth | Grey | $L^*a^*b^*$ | HSI | YcbCr | YIQ | XYZ |
|---------|--------------|------|-------------|-----|-------|-----|-----|
| Image 1 | 211          | 214  | 207         | 190 | 214   | 212 | 80  |
| Image 2 | 214          | 235  | 235         | 211 | 235   | 236 | 87  |
| Image 3 | 262          | 244  | 241         | 201 | 253   | 254 | 120 |
| Image 4 | 158          | 211  | 196         | 130 | 207   | 209 | 136 |
| Image 5 | 146          | 192  | 196         | 170 | 192   | 189 | 161 |

**Table 3 – Statistical evaluation of Ki-67 positive counted nuclei**

|         | Grey |    |    |           | L*a*b* |    |    |           | HSI |    |    |           |
|---------|------|----|----|-----------|--------|----|----|-----------|-----|----|----|-----------|
|         | TP   | FN | FP | F-measure | TP     | FN | FP | F-measure | TP  | FN | FP | F-measure |
| Image 1 | 132  | 26 | 9  | 0.8829    | 135    | 23 | 6  | 0.9030    | 139 | 19 | 4  | 0.9236    |
| Image 2 | 154  | 18 | 5  | 0.9305    | 155    | 17 | 4  | 0.9366    | 151 | 21 | 9  | 0.9096    |
| Image 3 | 172  | 25 | 14 | 0.8982    | 176    | 29 | 6  | 0.9096    | 169 | 28 | 7  | 0.9062    |
| Image 4 | 190  | 42 | 12 | 0.8756    | 193    | 39 | 9  | 0.8894    | 190 | 40 | 9  | 0.8858    |
| Image 5 | 185  | 27 | 10 | 0.9091    | 180    | 32 | 5  | 0.9068    | 188 | 24 | 9  | 0.9193    |
| Average |      |    |    | 0.8993    |        |    |    | 0.9091    |     |    |    | 0.9089    |

|         | YcbCr |    |    |           | YIQ |    |    |           | XYZ |    |    |           |
|---------|-------|----|----|-----------|-----|----|----|-----------|-----|----|----|-----------|
|         | TP    | FN | FP | F-measure | TP  | FN | FP | F-measure | TP  | FN | FP | F-measure |
| Image 1 | 138   | 20 | 5  | 0.9169    | 139 | 19 | 5  | 0.9205    | 121 | 37 | 3  | 0.8582    |
| Image 2 | 155   | 17 | 4  | 0.9366    | 154 | 18 | 5  | 0.9305    | 138 | 34 | 3  | 0.8818    |
| Image 3 | 174   | 23 | 12 | 0.9086    | 173 | 24 | 13 | 0.9034    | 157 | 40 | 5  | 0.8747    |
| Image 4 | 190   | 40 | 12 | 0.8796    | 187 | 43 | 15 | 0.8657    | 183 | 47 | 11 | 0.8632    |
| Image 5 | 180   | 32 | 15 | 0.8845    | 183 | 29 | 12 | 0.8993    | 174 | 38 | 11 | 0.8766    |
| Average |       |    |    | 0.9053    |     |    |    | 0.9039    |     |    |    | 0.8709    |

practice. Various medical and pathological groups at different locations communicate with each other to discuss and put in efforts to analyse the results. Differences in training and inter observer variability are also factors that must be taken into

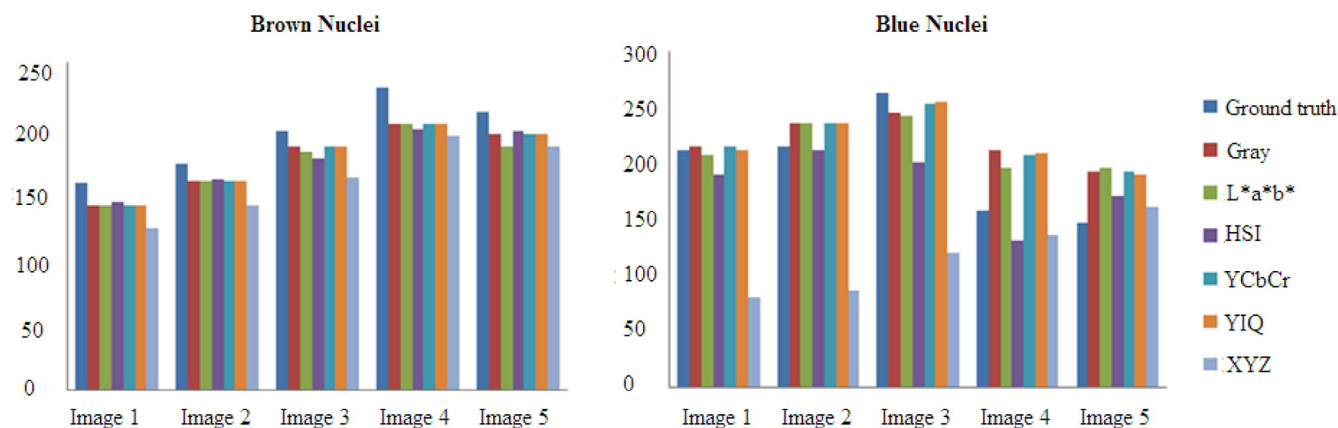
consideration while designing tools. Taking this into consideration, there is an urgent need to design algorithms that address the above challenges. The present methodology provides simplicity along with low computation time. The nuclei

**Table 4 – Statistical evaluation of Ki-67 negative counted nuclei**

|         | Grey |    |    |           | L*a*b* |    |    |           | HSI |    |    |           |
|---------|------|----|----|-----------|--------|----|----|-----------|-----|----|----|-----------|
|         | TP   | FN | FP | F-measure | TP     | FN | FP | F-measure | TP  | FN | FP | F-measure |
| Image 1 | 188  | 23 | 26 | 0.8847    | 187    | 22 | 20 | 0.8990    | 185 | 26 | 5  | 0.9227    |
| Image 2 | 208  | 6  | 27 | 0.9265    | 211    | 3  | 24 | 0.9399    | 189 | 25 | 22 | 0.8894    |
| Image 3 | 236  | 26 | 8  | 0.9328    | 232    | 30 | 9  | 0.9225    | 191 | 71 | 10 | 0.8251    |
| Image 4 | 136  | 22 | 75 | 0.7371    | 134    | 24 | 62 | 0.7571    | 119 | 39 | 11 | 0.8264    |
| Image 5 | 113  | 33 | 79 | 0.6686    | 134    | 12 | 62 | 0.7836    | 125 | 21 | 45 | 0.7911    |
| Average |      |    |    | 0.8300    |        |    |    | 0.8604    |     |    |    | 0.8509    |

|         | YcbCr |    |    |           | YIQ |    |    |           | XYZ |     |    |           |
|---------|-------|----|----|-----------|-----|----|----|-----------|-----|-----|----|-----------|
|         | TP    | FN | FP | F-measure | TP  | FN | FP | F-measure | TP  | FN  | FP | F-measure |
| Image 1 | 191   | 20 | 23 | 0.8988    | 200 | 11 | 12 | 0.9456    | 60  | 161 | 20 | 0.3987    |
| Image 2 | 212   | 2  | 23 | 0.9443    | 211 | 3  | 25 | 0.9378    | 70  | 144 | 17 | 0.4651    |
| Image 3 | 246   | 16 | 7  | 0.9553    | 242 | 20 | 14 | 0.9344    | 85  | 177 | 35 | 0.4450    |
| Image 4 | 142   | 16 | 65 | 0.7781    | 140 | 18 | 69 | 0.7629    | 126 | 32  | 10 | 0.8571    |
| Image 5 | 120   | 26 | 72 | 0.7101    | 123 | 23 | 63 | 0.7410    | 135 | 15  | 26 | 0.8682    |
| Average |       |    |    | 0.8573    |     |    |    | 0.8643    |     |     |    | 0.6068    |

**Fig. 8 – Quantitative assessment using bar plot for counted brown and blue nuclei.**

**Table 5 – Proliferation rate calculation (result from L\*a\*b\* colour space)**

| Image              | Auto counting (AC)    |                       |                                   | Manual counting       |                       |                                   | AC <sub>ERROR</sub><br>( $ A_{PR} - M_{PR} $ ) | Immuno-ratio (IR)                  | IR <sub>ERROR</sub><br>( $ IR_{PR} - M_{PR} $ ) |
|--------------------|-----------------------|-----------------------|-----------------------------------|-----------------------|-----------------------|-----------------------------------|--|------------------------------------|---|
|                    | Positive nuclei count | Negative nuclei count | Proliferation index ( $A_{PR}$ %) | Positive nuclei count | Negative nuclei count | Proliferation index ( $M_{PR}$ %) | (%)  | Proliferation index ( $IR_{PR}$ %) | (%)   |
| 1                  | 107                   | 153                   | 41.53                             | 158                   | 148                   | 51.63                             | 10.10  | 22.80                              | 28.83   |
| 2                  | 132                   | 184                   | 41.72                             | 175                   | 208                   | 45.69                             | 03.97  | 16.40                              | 29.29   |
| 3                  | 148                   | 187                   | 44.17                             | 171                   | 162                   | 51.35                             | 07.18  | 15.10                              | 36.25   |
| 4                  | 148                   | 130                   | 53.23                             | 188                   | 135                   | 58.20                             | 04.97  | 21.50                              | 36.70   |
| 5                  | 144                   | 94                    | 60.50                             | 165                   | 125                   | 56.89                             | 03.61  | 48.70                              | 08.19   |
| 6                  | 97                    | 144                   | 40.24                             | 135                   | 210                   | 39.13                             | 01.11  | 30.30                              | 08.83   |
| 7                  | 73                    | 118                   | 38.21                             | 80                    | 180                   | 30.76                             | 07.45  | 15.70                              | 15.06   |
| 8                  | 54                    | 181                   | 22.97                             | 65                    | 241                   | 21.24                             | 01.73  | 13.40                              | 07.84   |
| 9                  | 144                   | 112                   | 56.25                             | 166                   | 133                   | 55.51                             | 00.74  | 31.10                              | 24.41   |
| 10                 | 37                    | 93                    | 28.46                             | 50                    | 170                   | 22.72                             | 05.74  | 14.60                              | 08.12   |
| 11                 | 29                    | 76                    | 27.61                             | 53                    | 123                   | 30.11                             | 02.50  | 05.00                              | 25.11   |
| 12                 | 29                    | 88                    | 24.78                             | 45                    | 245                   | 15.51                             | 09.27  | 04.80                              | 10.71   |
| 13                 | 87                    | 130                   | 40.09                             | 98                    | 178                   | 55.05                             | 14.96  | 24.30                              | 30.75   |
| 14                 | 74                    | 84                    | 46.83                             | 172                   | 138                   | 55.48                             | 08.65  | 49.10                              | 06.38   |
| 15                 | 23                    | 66                    | 25.84                             | 50                    | 90                    | 35.71                             | 09.87  | 23.50                              | 12.21   |
| 16                 | 36                    | 169                   | 17.56                             | 38                    | 228                   | 14.28                             | 03.28  | 10.30                              | 03.98   |
| 17                 | 38                    | 55                    | 40.86                             | 57                    | 130                   | 30.48                             | 10.38  | 20.00                              | 10.48   |
| 18                 | 39                    | 69                    | 36.11                             | 65                    | 156                   | 29.41                             | 06.70  | 23.70                              | 05.71   |
| 19                 | 143                   | 23                    | 86.11                             | 153                   | 12                    | 92.72                             | 06.61  | 67.70                              | 25.02   |
| 20                 | 53                    | 82                    | 39.25                             | 63                    | 132                   | 32.30                             | 06.95  | 27.50                              | 04.80   |
| 21                 | 63                    | 159                   | 28.37                             | 68                    | 196                   | 25.75                             | 02.62  | 18.60                              | 07.15   |
| 22                 | 36                    | 169                   | 17.56                             | 42                    | 222                   | 15.90                             | 01.66  | 10.30                              | 05.60   |
| 23                 | 55                    | 138                   | 28.49                             | 57                    | 223                   | 20.35                             | 08.14  | 08.80                              | 11.55   |
| 24                 | 29                    | 38                    | 43.28                             | 47                    | 65                    | 41.96                             | 01.32  | 03.90                              | 38.06   |
| 25                 | 81                    | 113                   | 41.75                             | 82                    | 174                   | 32.03                             | 09.72  | 10.90                              | 21.13   |
| 26                 | 22                    | 24                    | 47.82                             | 32                    | 124                   | 20.51                             | 27.31  | 09.70                              | 10.81   |
| 27                 | 53                    | 117                   | 31.17                             | 71                    | 182                   | 28.06                             | 03.11  | 12.20                              | 15.86   |
| 28                 | 81                    | 113                   | 41.75                             | 85                    | 194                   | 30.46                             | 11.29  | 10.90                              | 19.56   |
| 29                 | 53                    | 117                   | 31.17                             | 60                    | 204                   | 22.72                             | 08.45  | 12.20                              | 10.52   |
| 30                 | 32                    | 64                    | 33.33                             | 58                    | 153                   | 27.48                             | 05.85  | 22.40                              | 05.08   |
| Average error rate |                       |                       |                                   |                       |                       |                                   | 6.84   |                                    | 16.13   |

texture properties or other characteristics could be used as prior information in algorithms to make the tool more robust.

Breast cancer diagnosis also includes detection of ER, PR, and HER2 receptors in stained tissue. Integration of detection and analysis of these procedures at the same place and time would be of great help to pathologist and doctors for formation of patients' reports and in treatment.

#### 4. Conclusion

In conclusion, this work proposes a hybrid segmentation methodology that detects positively and negatively stained nuclei in histopathology images followed by calculation of proliferation index. The proposed methodology has been found to be more accurate in this computation as compared to previously proposed techniques. The procedure suggested herein provides a better classification in terms of segmentation accuracy, time consumption and simplicity. It further simplifies the rigorous job of manually discriminating between cancerous and non-cancerous cells which may lead to false positive results. The F-measure value suggests that the L\*a\*b\* colour space is the optimal choice for segmenting Ki-67 positive and

negative nuclei using the proposed hybrid clustering segmentation method. The average error in automatic calculation of proliferation rate is found to be 6.84% which strengthens the case for using the proposed methodology. The problem of colour variances (basically illumination effect on the image and staining), when slides are obtained from different labs, may also be addressed with the help of staining normalization method. Advancements can be made by integrating image analysis system with other automated predictive/prognostic procedures to develop a complete diagnosis system.

#### Acknowledgement

We would like to thank the Ministry of Human Resource Development, Government of India for supporting and financially aiding the study under the SSLS-research grant, 4-23/2014 -T.S.I. dated: 14-02-2014, An IIT Kharagpur Initiative.

#### REFERENCES

- [1] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, et al., GLOBOCAN 2012 v1.0, Cancer incidence and



- mortality worldwide: IARC cancerbase no. 11 [Internet]. Lyon, France: International Agency for Research on Cancer, 2013. Available from: <http://globocan.iarc.fr>. (Accessed 2 September 2015).
- [2] K.H. Allison, Molecular pathology of breast cancer, *Am. J. Clin. Pathol.* 138 (2012) 770–780.
  - [3] A.S. Coates, E.P. Winer, A. Goldhirsch, R.D. Gelber, M. Gnant, M. Piccart-Gebhart, et al., Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015, *Ann. Oncol.* 26 (8) (2015) 1533–1546.
  - [4] M. Dowsett, T.O. Nielsen, R. A'Hern, J. Bartlett, R.C. Coombes, J. Cuzick, et al., Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group, *J. Natl. Cancer Inst.* 103 (2011) 1656–1664, <http://dx.doi.org/10.1093/jnci/djr393>.
  - [5] Z.M.A. Mohammed, D.C. McMillan, B. Elsberger, J.J. Goings, C. Orange, E. Mallon, et al., Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer, *Br. J. Cancer* 106 (2012) 383–388, <http://dx.doi.org/10.1038/bjc.2011.569>.
  - [6] S. Nilsson, C. Möller, K. Jirstrom, A. Lee, S. Busch, R. Lamb, et al., Downregulation of miR-92a is associated with aggressive breast cancer features and increased tumour macrophage infiltration, *PLoS ONE* 7 (2012) e36051, <http://dx.doi.org/10.1371/journal.pone.0036051>.
  - [7] S. Fasanella, E. Leonardi, C. Cantaloni, C. Eccher, I. Bazzanella, D. Aldovini, et al., Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies, *Diagn. Pathol.* 6 (Suppl. 1) (2011) S7, <http://dx.doi.org/10.1186/1746-1596-6-S1-S7>.
  - [8] V. Manucha, X. Zhang, R. Thomas, The satisfactory reproducibility of the Ki-67 index in breast carcinoma, and its correlation with the recurrence score, *Clin. Cancer Investig. J.* 3 (2014) 310–314, <http://dx.doi.org/10.4103/2278-0513.134483>.
  - [9] R. Shui, B. Yu, R. Bi, F. Yang, W. Yang, An interobserver reproducibility analysis of Ki67 visual assessment in breast cancer, *PLoS ONE* 10 (2015) e0125131, <http://dx.doi.org/10.1371/journal.pone.0125131>.
  - [10] V.J. Tuominen, S. Ruotoistenmäki, A. Viitanen, M. Jumppanen, J. Isola, ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67, *Breast Cancer Res.* 12 (4) (2010) R56.
  - [11] R. Muñoz, R. Gonzalo, Easing scoring in ER and Ki-67 breast cancer histopathological images, 2012. <http://e-archivo.uc3m.es/handle/10016/16161>.
  - [12] F. Xing, H. Su, J. Neltner, L. Yang, Automatic Ki-67 counting using robust cell detection and online dictionary learning, *IEEE Trans. Biomed. Eng.* 61 (2014) 859–870, <http://dx.doi.org/10.1109/TBME.2013.2291703>.
  - [13] C. Ko, C. Tsai, C. Lin, K. Liao, A computer-aided diagnosis system of breast intraductal lesion using histopathological images, 2014. 212–217.
  - [14] R.C. Gonzalez, R.E. Woods, Digital image processing, third, 2007. <http://dx.doi.org/10.1049/ep.1978.0474>.
  - [15] L. Busin, N. Vandenbroucke, L. Macaire, Color spaces and image segmentation, *Adv. Imaging Electron Phys.* 151 (2008) 65–168, [http://dx.doi.org/10.1016/S1076-5670\(07\)00402-8](http://dx.doi.org/10.1016/S1076-5670(07)00402-8).
  - [16] C.J. Du, D.W. Sun, Comparison of three methods for classification of pizza topping using different colour space transformations, *J. Food Eng.* 68 (2005) 277–287, <http://dx.doi.org/10.1016/j.jfoodeng.2004.05.044>.
  - [17] J.M. Chaves-González, M. Vega-Rodríguez, J. Gómez-Pulido, J.M. Sánchez-Pérez, Detecting skin in face recognition systems: a colour spaces study, *Digit. Signal Process.: Rev. J.* 20 (2010) 806–823, <http://dx.doi.org/10.1016/j.dsp.2009.10.008>.
  - [18] G. Ruiz-Ruiz, J. Gómez-Gil, L.M. Navas-Gracia, Testing different color spaces based on hue for the environmentally adaptive segmentation algorithm (EASA), *Comput. Electron Agric.* 68 (2009) 88–96, <http://dx.doi.org/10.1016/j.compag.2009.04.009>.
  - [19] L. Busin, J. Shi, N. Vandenbroucke, L. Macaire, Color space selection for color image segmentation by spectral clustering, 2009. *Signal Image Process. Appl. (ICSIPA)*, 2009 IEEE Int. Conf. (2009) 262–267. <http://dx.doi.org/10.1109/ICSIPA.2009.5478603>.
  - [20] A. Jurio, M. Pagola, M. Galar, C. Lopez-Molina, D. Paternain, A comparison study of different color spaces in clustering based image segmentation, 2010. 532–541.
  - [21] D. Khattab, H.M. Ebied, A.S. Hussein, M.F. Tolba, Color image segmentation based on different color space models using automatic GrabCut, 2014.
  - [22] M.K.K. Niazi, M. Pennell, C. Elkins, J. Hemminger, M. Jin, S. Kirby, et al., Entropy based quantification of Ki-67 positive cell images and its evaluation by a reader study, 2013. 86760I. <http://dx.doi.org/10.1117/12.2007909>.
  - [23] W. Yan, X. Feng, A watershed based segmentation method for overlapping chromosome images, 2010. 2nd Int. Work. Educ. Technol. Comput. Sci. ETCS 2010. 1 (2010) 571–573. <http://dx.doi.org/10.1109/ETCS.2010.107>.
  - [24] C. Jung, C. Kim, Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization, *IEEE Trans. Biomed. Eng.* 57 (2010) 2600–2604, <http://dx.doi.org/10.1109/TBME.2010.2060336>.
  - [25] V. Anari, R. Amirfattahi, Enhancement and segmentation of IHC images of meningioma tumor through adaptive vector directional filter and K-means clustering algorithm, 2011.
  - [26] H. Sharma, N. Zerbe, D. Heim, S. Wienert, H.-M. Behrens, O. Hellwich, et al., A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images, *Proc. 10th Int. Conf. Comput. Vis. Theory Appl.* 37–46 (2015) <http://dx.doi.org/10.5220/0005247900370046>.
  - [27] K. Sirinukunwattana, D.R. Snead, N.M. Rajpoot, A novel texture descriptor for detection of glandular structures in colon histology images, 2015. 9420, 94200S. <http://dx.doi.org/10.1117/12.2082010>.
  - [28] J.B. MacQueen, Kmeans some methods for classification and analysis of multivariate observations, *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* 1 (1967) (1967) 281–297, doi:citeulike-article-id:6083430.
  - [29] P. Filipczuk, M. Kowal, A. Obuchowicz, Automatic breast cancer diagnosis based on K-means clustering and adaptive thresholding hybrid segmentation, *Adv. Intell. Soft Comput.* 102 (2011) 295–302, [http://dx.doi.org/10.1007/978-3-642-23154-4\\_33](http://dx.doi.org/10.1007/978-3-642-23154-4_33).
  - [30] A. Gautam, White blood nucleus extraction using K-Mean clustering and mathematical morphing, 2014. 549–554.
  - [31] H.Z. Al-lahham, R. Alomari, H. Hiary, V. Chaudhary, Automating proliferation rate estimation from breast cancer Ki-67 histology images, *SPIE 8315, Medical Imaging Comput. Diagnosis.* 1 (2012) 7, <http://dx.doi.org/10.1117/12.911009>.
  - [32] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, *Comput. Geosci.* 10 (1984) 191–203, [http://dx.doi.org/10.1016/0098-3004\(84\)90020-7](http://dx.doi.org/10.1016/0098-3004(84)90020-7).
  - [33] V. Anari, P. Mahzouni, R. Amirfattahi, Computer-aided detection of proliferative cells and mitosis index in immunohistochemically images of meningioma, 2010 6th

- Iran, Conf. Mach. Vis. Image Process. MVIP 2010 0–4 (2010) <http://dx.doi.org/10.1109/IranianMVIP.2010.5941151>.
- [34] E. Nadernejad, S. Sharifzadeh, A new method for image segmentation based on Fuzzy C-means algorithm on pixonal images formed by bilateral filtering, *Signal Image Video Process.* 7 (2013) 855–863, <http://dx.doi.org/10.1007/s11760-011-0274-0>.
- [35] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, T.-J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, *Comput. Med. Imaging Graph.* 30 (2006) 9–15, <http://dx.doi.org/10.1016/j.compmedimag.2005.10.001>.
- [36] R.M. Haralock, L.G. Shapiro, *Computer and Robot Vision*, 1st ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1991.
- [37] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, F-Score and ROC: a family of discriminant measures for performance evaluation, *Adv. Artif. Intell. (Lecture Notes Comput. Sci.)* 4304 (2006) 1015–1021, [http://dx.doi.org/10.1007/11941439\\_114](http://dx.doi.org/10.1007/11941439_114).
- [38] S. Tewary, A. Akula, R. Ghosh, S. Kumar, H.K. Sardana, Hybrid multi-resolution detection of moving targets in infrared imagery, *Infr. Phys. Technol.* 67 (2014) 173–183, <http://dx.doi.org/10.1016/j.infrared.2014.07.022>.