

Review on Histopathological Slide Analysis using Digital Microscopy

Sangita Bhattacharjee¹, Jashojit Mukherjee¹, Sanjay Nag¹, Indra Kanta Maitra² and Samir K. Bandyopadhyay¹

¹*Department of Computer Science and Engineering, University of Calcutta, India*

²*B. P. Poddar Institute of Management and Technology, Kolkata, India*

sangs_555@yahoo.com, jashojitmukherjee@gmail.com, sanjaynag75@gmail.com, ikm.1975@ieee.org, skb1@vsnl.com

Abstract

With the increasing research on better algorithms for segmentation of image, it is important to categories the already known technique, their conceptual basis and their overview. This paper presents a systematic survey of the algorithms used in automated cancer diagnosis based on histopathology. The various computational processes involved like pre-processing, feature-extraction, post-processing and disease detection (classifying the area of malignancy) are reviewed.

Keywords: *Colour-illumination normalization, smoothening, denoising, enhancement, thresholding, edge-based, region-based, active contour, clustering*

1. Introduction

A biopsy sample is usually taken to detect cancerous cells. The biopsy sample is sliced into thin segments and mounted over a glass slide under a microscope for determination of the extent of the disease. Approximately 80% of the biopsies out of 1 million samples are found to be benign in USA every year; pathologists are spending a lot of time in sieving through the diseased tissue which would otherwise be utilized in treating patients. With the advent of digital scanners and image processing capable computers analysis of histological slides have become easier and the different pre-processing and segmentation processes helps faster identification and characterization of a malign tissue. Digital pathology takes traditional pathology specimens – tissue samples mounted on glass slides and transform them into digital images that can be shared electronically by care providers.

While digital pathology has existed in some form since the 1980s, for most of this time the limitations have outweighed the benefits. Slow scanner speed, poor image quality, low network bandwidth and high cost have meant that digital pathology systems had no cost or efficiency advantage over the current process. In M. Gurcan *et al.*, [1], and citations of various other authors [2, 3, 4] has shown the improvement in image analysis over the past decades due to the use of digitized image analysis and the crucial role of computer-aided diagnosis (CAD). Recent innovations in digital pathology technology have radically altered this picture by transcending these earlier limitations, thus opening new horizons in anatomic pathology. O. Sertel *et al.*, [5] has proposed methodology on how digitized tissue samples are classified into histological grades under a statistical framework. The challenges in digital imaging has led to improvement in image analysis techniques resulting in better opportunities catering to the pathologist for treatment of benign tissues as observed in the citation by A. Madabhushi [6].

2. Biopsy

Biopsy is the process of collecting a sample of affected tissue for examination by a pathologist. Usually using a microscope, a pathologist examines a sample for indications and extent of disease. By definition, in biopsy, the tissue is required to be detached from the living part of the subject. Surgically removal of the entire lump or affected area is called an Excisional biopsy. The procedure of removing only the tissue samples is called an Incisional biopsy (core biopsy).

There are different methods to perform a biopsy: Scrape biopsy-Cells are scraped and then from the tissue surface according to the requirement the samples are removed either from inside the mouth or from neck of the womb (cervix). Punch biopsy - (a knife in the shape of a round) punch is used to cut and remove the tissue in the form of a disk; as a skin tissue is required to collect samples and to check malignancy. Needle biopsy –here by the use of a needle the sample of liquid (in usual cases) are removed. In case of core biopsy a needle which is quite wide is used whereas a needle that is fine in shape is used in case of fine-needle aspiration biopsy. Endoscopic biopsy - a long-thin, lighted optical instrument known as an endoscope is required for collection of the samples from deep inside the body. Endoscopes are specially made for certain purpose specific like the one for the bladder is cystoscope,the one for the kidneys are called nephroscope,one made for the bronchi (in the lungs) is known as bronchoscope, the one made for the voice box (larynx) is laryngoscope, *etc.* Capsule biopsy – it is that kind of biopsy where sample is taken from the intestines. Stereotactic biopsy – in this kind samples are extracted from human brain, using such surgeries the biopsy sites are located. Stereotactic are a three-dimensional system of coordinate used in locating minor targets found within our body. Colposcopic biopsy –this biopsy is used in evaluating a patient (female) with abnormality and having Pap smear. The samples of tissue are transferred to the lab for chemical treatments and finally various thin sections are derived after slicing up. For further observations under a microscope these sections are then studied. A person who specialises in the study of blood (haematologist) may also study these samples.

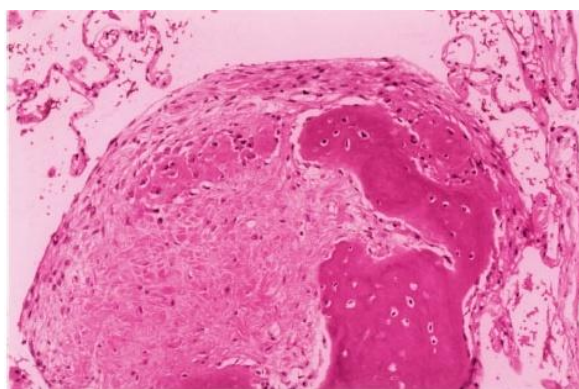


Figure 1. Image of a Biopsy Slide showing the portion where the fibrosis got transformed into the malignant portion

Slices which are cut in a thin manner are attached over a glass slide the remaining portions of the tissue are retained for future studies. For better visualization of the cells often pathologists are found to add dyes on them (they stain the tissue), this enables them to view more clearly. There are several works that have been carried out using automated techniques

in Ming Gan *et al.*, [7] that use pre-processing and segmentation phases of a histopathological slide images after the biopsy procedure is carried out.

3. Literature Review

M. Gurcan *et al.*, [1] has focused on the digitized tissue histopathology which has now become an important application of computerized image analysis as well as machine learning techniques. Increasingly, the role of (CAD) algorithms in medical imaging has escalated to a point where algorithms have been developed for disease detection, diagnosis, and prediction of prognosis that complements the opinion of the pathologist. A review on the recent development in the state of art CAD technology for digitized histopathology has been described in this paper and the scope of development of an application that can tackle the specific histopathology related problems around the world has been discussed.

J. C. Caicedo [8] has shown (CAD) for the purpose of classification and retrieval of image. Despite the diversity of interests in solving problems and variations in the histology slide images, a trend to measure some main characteristic features in images, such as tissue architectures like the colour, texture and morphology has also been presented. The main algorithms and strategies used for extraction of specific features in a histology slide image have been discussed.

C. Demir *et al.*, [2], has presented that the computational steps involved in automated CAD diagnosis is based on histopathology steps like: 1) Pre-processing of the image for determination of the area of focus, 2) Feature extraction to quantify the properties of the focal areas, and 3) classifying of the areas to focus on the presence of malignancy if present. In Step 1, the determination of focal area is preceded by denoising to improve the result. In cellular-level diagnosis, this step is categorized as nucleus/cell segmentation. Step 2 has appropriately defined representation of the areas to focus in order to provide a distinctive measure. Finally in Step 3, an automated diagnostic system for operating on the diseased tissue/cell measures has been designed. This step is responsible for estimating the accuracy of this system. The data sets have also shown comparisons of the different features and designs to prevent misleading estimation through these systems have allowed the subsets to distinguish the features for improvement of automated cancer diagnosis.

The examination of the tissue structures has been shown in the citation by A. D. Belsare *et al.*, [3], and the cell distribution in a tissue, irregularities of the shapes of cells to determine the level of malignancy and benign in a histopathology image slide is a very time consuming process that is prone to intra as well as inter observation variations. To overcome this kind of a difficulty a CAD image analysis is often required for quantitative diagnosis of tissues and cells. The applications of digital image processing techniques for histopathological image analysis to cover up the process of segmentation and classification have been discussed.

In the citation by O. Sertel *et al.*, [9] the method proposed has classified the image into either a low or high grade on the basis of the amount of cytological components present in it. For discrimination purpose the lower grades are classified as low and mid grades-an approach of colour texture analysis has been used which has modified the occurrence of grey level in an image slide. The matrix method has been used as non-linear colour quantization method with the use of self-organizing feature maps. For analysing the stained a set of dynamic colour range is used to fulfil the purpose of grey level based analysis of the texture of an image.

P.A.Hall *et al.*, [10] has shown assessment is done for identifying the three important functional types of a cell, *i.e.*, statistical, conditional renewal, and continually renewing. The practice of histopathology involves direct or indirect assessment of the cellular proliferation, the advantages and disadvantages of the better-known methods that are used for assessing cellular proliferation in histopathological slide images.

J.Kazi *et al.*, [11] commented that the last few decades have also witnessed an increasing role of electronic IT technology and the advancement in communication technology in the histopathology laboratory. Recently the major concerns lies with the increasing demands for standardization, regulation controls, and legal accountability, which have set the guidelines to help pathologists to face this kind of an increasingly complicated system. Histopathology, has witnessed its evolution from an autopsy based to the current molecular based histopathology. Incorporating these histopathological practices in particular with the functional correlation of morphological studies, has led to the change in the outlook of both histopathology and the histopathologist.

4. Methods

In the case of tissue-level feature extraction, the pre-processing step thresholds the image to identify the focal areas by eliminating the noisy regions and those with little content [11, 12]. It has been discussed the different algorithm techniques used in classification of a histopathological slide image. The basic steps like pre-processing, segmentation, feature extraction, selection extraction classification and decision-making are discussed:

- Pre-processing – this phase is required in order to improve the quality of image and make the segmentation results much more accurate. The first step involves the removal of unwanted parts in the background of the image slide. This process allows registering of the slide image so that there can be further comparison and analysis done in a standardized manner.
- Segmentation-The first row of each segment is used for locating pixel paths for traversing through the image pixels after dividing it into several sections.
- Feature extraction-the morphological characteristics of image for abnormality or classification of the image for different grades of the disease is done. The properties of individual cells without considering spatial dependency between these cells are done in this step.
- Selection extraction-Colour and texture information are obtained from the image which is combined in consideration with the distribution of components in cells and tissues which can be further segmented.
- Classification & decision making-After segmentation and feature selection the different classifiers are applied to classify images for diagnosing abnormality in them. In this step, a cell or tissue is assigned to one of the classes and then it can also be classified for malignancy level i.e. gradation of disease.

S. Petushi *et al.*, [14] has come up with the two parameters that differentiated tumour grade found in a histopathological slide image i.e. the number of cross sections as well as the density of cell nuclei and hence classification are done of the affected cell. Classifications are based on the subdivisions of a whole image slide containing a high concentration of cancer cell nuclei consistently agreed with the grade classification of the entire slide. The automated image analysis and classification makes it feasible to develop and classify the histology images based on micro-texture and reveals the grade differentiating parameters on large – scale data helping the pathologists in their decision –making process.

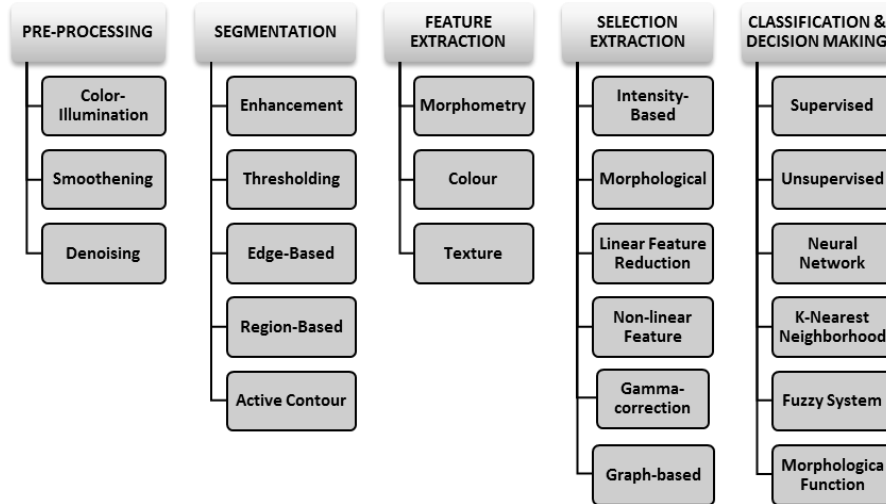


Figure 2. Schematic Diagram of Methods Related to Digital Microscopy

S. Kim *et al.*, [15] described an automatic image indexing engine that generates topical descriptors for digitized microscopic images which are available on the web. This system has shown implementation of a mapping module which could transfer the indexing terms to control vocabularies through MetaMap Transfer (MMTx) algorithm developed.

4.1. Pre-processing

The pre-processing of an image is done in three different steps; the image slides needs to be transformed so that the affected area of the image is placed at a centre position. This is one of the important steps as each image slide has both the affected as well as unaffected areas. There is a need to separate the diseased area from the rest of the regions so that further comparisons can be made for analysis so that it can be standardised for using heuristic methods.

- Colour-Illumination -Histograms based segmentations uses the peaks to find high and low values and finds clear edges after setting pixel values along the edges of the image using colours available from the colour spaces.
- Smoothing- Smoothing operators are used to smoothen out the details. For this purpose the 7×7 kernel is taken as a convolution filter and the result obtained are the values of deviation (σ) by close observation of intensity levels for all categories of image slides, thus allowing adjustments of the level of smoothening for these slides.
- Denoising-Noise is another -factor we need to eliminate from the image pair. There are different kinds of noises like high intensity noises, having high values of optical intensities exemplified as labels, scanning artefacts, tape artefacts and other shadows presenting themselves as horizontal running strips etc. Such irrelevant noise is removed so that the CAD system can be executed effectively.

4.1.1. Colour-Illumination

The capacity of visual information processing is a highly developed technique such as staining, optics and electron microscopes. The colour difference classifies specific objects to be investigated and finds the relationship among them.

According to the citation by M. Larsen [16], in a colour image, each pixel has a value or colour vector, among the most commonly used coloured vector are RGB or HSV. The illumination model is represented as: $x(\lambda) = c(\lambda) l(\lambda)$, where c =colour, x =emitted signal for a surface and l =light source of colour.

In the citation by S. Tamai [17] has shown the effective use of absolute colour values, ratios of each of the tri-stimulus colour, colour differences observed against adjacent areas and estimation of illumination data are investigated for the better performance of a pathological and cytological image slides.

4.1.2. Smoothing

In Multi-resolution graph-based analysis of histopathological whole slide images the application to mitotic cell extraction and visualization by V.Roullier et al [18], to regularize an image its main elements should be preserved and then globally smoothed using Gaussian smoothing. A weighted graph $G=(V,E,w)$ consisting of a finite set of vectors V , of N vertices and set of edges E , are used to describe a set of objects and their pair wise relationships between those objects.

According to M. Habibzadeh *et al.*, [19], to smoothen objects or reduce the surface angularities caused by irregular vertices and to reduce the visibility of certain features within the image, Gaussian smoothing is used to capture the image.

O.Sertel *et al.*, [9] used the EM (expectation maximization) algorithm to identify the underlying distribution with the use of Gaussian smoothing. To find the unknown parameters the distribution is computed, where each pixel is assigned to a particular class based on its probability. The computation of the ridge like structures is done by Gaussian smoothing to group these structures with similar orientations and represent their connected components using line segments.

4.1.3. Denoising

The search for an effective denoising technique still is a challenge at the cross-roads between the analysis which are done functionally and statistically. Despite the recent sophisticated proposed methods, most algorithms have not yet attained a desirable level of applicability and fail in general to remove fine structures or artefacts in images. Thus a general mathematical and experimental methodology will be used for the purpose of comparison and will be able to classify image algorithms for the purpose of denoising and will bring forth an algorithm with nonlocal means (NL-means) which will be addressed to preserve the structures in any type of digital image. The analysis that is done mathematically will be based on analysing the “method noise,” which will define the differences that exists between the denoised versions of the digital image from the original digital image. According to A. Buades *et al.*, [20] the NL-means algorithm under an image model that is statistically generic is optimal asymptotically. This method has been found to perform and has been compared in four separate ways; mathematically: asymptotic order of magnitude with the assumptions of regularity; secondly, the algorithm artefacts and their explanation have been used with the violations of this model; a table has been constructed to measure the distances of the version with denoise version to that of the original image. The fourth and the most powerful evaluation method is visually analysing on a natural image the method noise. One of the features in this method noise is that noise is treated like a real white noise that has produced a result with meaningful structures.

N Joshi *et al.*, [21] has shown that after computing the linear combination of the two most influential colours which are present within a neighbourhood the colour of a pixel is determined. “This two-colour prior has two major benefits”: one is that it is tuned to the

content of the particular image and the other is that it serves to decouple edge sharpness from the edge strength. The unified algorithm has been used for blurring and denoising compared to the performances of the methods that have been already worked upon in the recent past and have been specialized for such kinds of applications.

N. Kachouie [22] focused on Bayesian model which is based on a filter used which is a linear approximation filter constructed on the basis of the minimum mean-square error (MMSE) [21] and nonlinear filters have been proposed for denoising non-stationary signals, the “Wavelet Image Denoising” has been used to recover the original signal from the noisy one by removing the noise.

4.2. Segmentation

It is division of the image into parts representing real objects in the image. Techniques like Histogram based, manual, grey shade segmentation, are used to find the edges, grey shades and complex region growing to group individual pixels together into similar regions, *i.e.*, of the same intensity (shade of grey), that forms a texture, line up in a row & create a shape. Thus this segmentation leads to better separation of objects than the grey-shade-only segmentation. Segmentation of the various structures in a histopathology slide image using methods like thresholding, enhancement, and adaptive thresholding, which tends to work only on uniform image slides and does not produce a consistent result if there is considerable variation within image sets. For example, watershed algorithms pose the same problem due to variations in image sets. An object is defined as a connected group of pixels satisfying some similarity of criterion. The main focus is often on the segmentation of nuclei which explicitly uses the features of cytoplasm and has expressed need for such features as well.

The algorithm of image segmentation performs extraction of the regions by finding the similarity of an image with its predefined feature like grey-level value. In various applications, a variety of structure or texture exhibits and hence adequate segmentation cannot be done by grey-level intensities alone. Some extra features relating to the image's structure is required for segmentation of these images. Meaningful segmentation of images of skin lesions with the use of several segmentation algorithms is proposed, that uses grey-level values and features based on the textures for both types of extractions region based and boundary based. The segmentation based on intensity is derived by the use of the modified algorithm which extracts region using the pyramid-based method. The segmentation that is texture-based is derived by an algorithm which is bi-level and the processing technique is based on shifted-window method using a general level co-occurrence matrices. The results of individual segmentations obtained from different channels, representing the complete set of colour and texture information, are analysed using heuristic methods.

- Thresholding-Histograms based segmentations uses the peaks to find high and low values and finds clear edges after setting pixel values along the edges of the image using Region-based algorithm is performed before segmentation to segregate the regions in the image.
- Edge detection-Detecting edges of an image helps to find the grey scale values at the edges. This technique is used to fix edges of an image array after convolutions are performed and they produce results that vary from 0 to max grey level value.
- Enhancement-The edge detectors are used to enhance and improve the appearance of an original image into the output image.

4.2.1. Thresholding

Segmentation leads to division of the image into sub parts that constitutes the region and classification of the pixels into either: “light” or “dark”. The accuracy of this method is detrimental to the success or failure of the computer analysis procedure eventually leading the algorithms to be primarily based on two categories: firstly, to partition the image and generate the edges which are sudden changes in the intensity values. Secondly, partitioning and deriving regions in the image that are found to be the same according to predefined criteria, example of such a case is histogram thresholding. In the citation by D. Bradley *et al.*, [23] it showed thresholding as a form that takes into account spatial variations in illumination. A technique for real-time adaptive thresholding using the integral image of the input are used which is an extension of the old methods. However, thresholding gives a solution which is more robust to illumination changes in the image. This technique is sometimes suitable for processing of live video streams at a real-time frame-rate, making it a valuable tool for interactive applications that are used in reality.

The basic steps involved in thresholding are:

- Step 1. Select an initial estimate for T
- Step 2. Segment the image using T. This will help to produce two groups of pixels, consisting of all pixel with grey values $\geq T$ and consisting pixels with values $< T$.
- Step 3. Find a new threshold value i.e., $T = (\text{mean1} + \text{mean2})/2$
- Step 4. Here repetition of step 2 is done till the difference is smaller in T than a predefined parameter with successive iterations.

Each and every pixel in an image that is coloured is characterize according to the 3 RGB values and then plotting is done on the basis of colour values of each pixel and then the threshold points are found. After setting pixel values along the edges of a histopathological slide image, this way thresholding helps to lessen the edges of an image.

In Dr Md Berbar’s [24] citation on segmentation section on thresholding a region growing method has been used to threshold to a set of the values of pixels in the image to one or zero and this technique has used the routine thresholding image array to accomplish this task.

In the citation by Otsu’s [25] the thresholding algorithm has helped to determine the gradients. This way linear plotting is done which has uniquely separated the cells in three groups of normal, possibility of malignancy and malignant cells. This technique of separation has been done with cent percent accuracy due to the unique linear plots.

In Ming Gan *et al.*, [7] citation has stressed on the two modes: transparent and reflected modes in digital imaging with grey-level thresholding that determines the elimination of background influences.

4.2.2. Enhancement

Majority of images in histopathology require post-acquisition adjustments to optimize brightness, contrast, and image visibility. Thus, much research effort has been put in development of customized contrast enhancement techniques for this application.

For improvement in the visual impact and to detect the folds in the tissue in images which are whole slide by colour enhancement technique has been shown by P. A. Bautista *et al.*, [26], which applies the method of enhancing the stained image to hematoxylin and eosin (H&E) to visually improve the folds in the tissue irrespective of the variations in colour present in the images. Detecting the folds in tissue after applying enhancement methods helps

in visual improvement of folds except the nuclei which are present and stained dark than the folds, this might at times affect the level of accuracy while detection. Enhancement approach is implemented and applied in relatively simple images at high magnification showing fairly uniform-stained structures. The main objective here is to apply a similar technique but automatically, in more complex and larger-scale images, while processing further the resulting image to extract the nuclei region of interest.

The PCA Algorithm generates weighted linear combinations that are adapted to the spectral changes of the cells being imaged from one image to another and identifies a linear transformation of the original colour coordinate system such that the three axes of the new coordinate system coincide with the directions of the largest spreads of the point distribution. In the citation by G. Loukas *et al.*, [12]. PCA has also been applied successfully for the enhancement of images of stained cervical smears.

In the citation by C.F. Jiang *et al.*, [27] a comparative study, of the R/G ratio was selected as the optimal contrast enhancement method has been presented. According to M.Tahtouh *et al.*, [13] a full comparison of the finger-mark technique with existing methods, has been used as a variety of donors, surfaces, and finger-mark ages, has been discussed. Along with an examination of how it might fit into the current latent finger-mark enhancement sequences in an image with excellent ridge detail has been acquired using infrared chemical imaging.

4.2.3. Edge-Based

In B. Morse [28] citation, the continuous path of edge pixel is not required and the technique called “Hough Transform” technique has been used. Here the edge points are not on a straight line but on a circle and other shapes on an edge pixel is also considered. In this technique the matching is done of image edge points to target contour points. A table has been built that keeps records for each point with edge tangent orientation, θ , the direction, α , and distance r to the reference point. Depending on the complexity of shapes the multiple points with tangent orientation are built and edges are detected.

In the citation by Qian Yu *et al.*, [29], Edge-based methods helps in the location of boundaries according to the neighbourhood spectral variation with an algorithm of edge detections, usually high-pass convolution algorithms such as differentiation. Region extraction can be further broken down into several region growing, region dividing and the sum of edges of the objects in a slide that are shared with the other image objects.

In the citation by M.S.Brown *et al.*, [30], the curves with expected length, position, orientation at the endpoints, and edge strength are used as parameters. The lengths are first modelled in centimetres, and then they are later converted to number of pixels for any given image. “Canny edge detection is performed” since the anatomical structures fail to produce a single, continuous edge”. The image processing engine has been applied over the canny operators at multiple resolutions, with standard deviations, like SD - 3, 7, 9, and these values are used with the following algorithm to generate result by linking different combinations of edge fragments:

1. Applying gradient magnitude (strength) and phase (orientation) constraints;
2. Formation of the edge fragments from continuous sets of points;
3. Subdivision (split) of edge fragments at points of high curvature; and
4. Calculating possible linkages of edge fragments to form candidates: a link that can occur if the orientation of a line joining the end points of the fragments is within the orientation constraint range, and the overall length constraint is satisfied.

An edge based segmentation method has been used in the citation by Md. Sathik *et al.*, [31], to segment a histopathological slide image into foreground area and background area. The method proposed here, has preserved the quality of the foreground image. JPEG compression has been used as a compression technique which is widely used in this method of segmentation. In general cases the JPEG compression method uses linear quantization and threshold values for the purpose of maintaining certain quality levels in any of the image slides.

In D.Raman's [32] citation the appearance of the background, feature which is invariant to the appearance, which has been matched with an edge-based models of the image which are deformable and to obtain (soft) estimates of body part positions. Generally, these estimates gives results in poor performances because the model can be distracted by edges in the background. An iterative algorithm has been used for estimation purpose that learns a region model for each body part and also for the background. This algorithm has been initialized by the edge maps and then combining the two the result finally is has been shown for these two images.

In this citation by M Lynch *et al.*, [33], has presented a new computationally efficient three-dimensional (3-D) segmentation technique based on the detection of edges in a slide. Classification of the edges is done according to three categories: fold edges, semi step edges, and secondary edges. The 3-D slide has been sliced into different parts creating equidepth contours (EDCs). Proceeding this way; three types of critical points have been extracted from these EDCs. A subset of the edge pixels has been extracted first using these critical points. Then the edges are extended from these pixels by applying some masks over the neighbourhood critical points.

4.2.4. Cell segmentation

It divides the image into parts representing real objects in the image. Techniques like manual, grey shade segmentation, region-based, boundary -based are some of the methods discussed using edges, grey shades and complex region growing to group individual pixels together into regions if they are similar, *i.e.*, they are of the same intensity (shade of grey), form a texture, line up in a row & create a shape. This segmentation has better separation of objects than the grey-shade-only segmentation.

- Region-based approach-This method attempts to partition regions based on some general properties of an image. The common properties of an image is like, deriving from the pattern or texture of each region, the intensity values which are unique to that specific region which is later on responsible for providing an image data that is multidimensional in nature. This algorithm is performed before segmentation to segregate the regions in the image.
- Boundary-based approach- This method seeks to look for explicit or implicit boundaries between regions corresponding to different tissue types.

Thresholding is also used for cell segmentation. Although thresholding separates the cells from the background, it does not separate the overlapping cells from each other. The watershed algorithms offer a solution to this problem. Successive application of opening and closing operators is used to find the separate centres of gravity of the cells. Watershed algorithms are useful to detect the boundary lines between the touching cells. An important limitation of the watershed algorithms is over-segmentation. Applying low-pass filters prior to segmentation and letting water rise only from the marked seeds lessens the over-segmentation problem. After thresholding the pixels, deleting small objects and large areas

and filling small holes improve the segmentation results. For that, it is also possible to apply the morphological operators with different types of structuring elements such as square and octagonal ones. Here, selecting the size of the structuring element is important which is done by region-based approach. The size should be smaller than the minimum size of a cell that will be determined; all the objects smaller than the structuring element are eliminated. On the other hand, it must be large enough to eliminate the noisy areas. Another region-based approach makes use of a learning algorithm to determine whether a pixel belongs to a cell or not. The learning algorithm employs the features extracted for the pixels. Such features can be based on colour information of the pixels or their textural properties.

A geometric approach of 3D-segmentation technique as described in Sapiro *et al.*'s citation [34], an example of how an MRI slice is represented on a flat region is shown enhances the feature extraction process by detecting objects of minimal surface areas after separating into homogeneous regions.

After detecting the outlines in an image according to M.Haindl *et al.*, [35], the subsequent region growing step where the neighbouring outline curves are grouped together. Region growing based on curve segments instead of the pixels like in the classical approaches speeds up the algorithm. This approach groups pixels into regions based upon two criteria such as proximity and homogeneity of all region based methods that produces such groupings either by splitting the image, or by segmenting the regions in an image, into much smaller regions and then merging these small regions into larger ones, or splitting and merging until the criteria are maximally satisfied. In a 2-D region growing, regions that have pixels are found "four-connected", directly neighbouring each other in any of the four horizontal and vertical directions, and then the proximity with each other is analysed to assign pixel values.

Among the different techniques for segmenting of a histopathological slide image forming slices, this citation by J. S. Suri *et al.*, [36], acquired in orthogonal directions, one such technique used is the region-based techniques. This citation has focused on the state-of-the-art systems which are a fusion of boundary-based methods with the region-based method, these are also called as deformation models which are geometrically regional in nature and it has taken into consideration the partial differential equations and its paradigm as the level of framework that has been set.

In the citation by T. Blaschke [37], image segmentation, from the perspective of an algorithm, is generally divided into several categories; one such category is region-based. Segmentation provides the building blocks of object based image analysis to deal with spatially fuzzy objects or with fields. Segments are regions which are generated by one or more criteria of homogeneity in one or more dimensions respectively. Thus segments have additional spectral information compared to single pixels i.e. the mean values per band, and also median values, minimum and maximum values, mean ratios, variance, *etc.* but of even greater advantage than the diversification of spectral value descriptions of objects is the additional spatial information for objects. It is a well-known fact that semantically significant regions are found in an image at different scales of analysis and is linked to multiscale analysis concepts, even if single levels are targeted for specific applications; this concept is called 'multiscale segmentation/object relationship modelling'. This segmentation algorithm (region-based), has also been applied on a very high spatial resolution images for different landscapes.

In the citation by J. C. Caicedo *et al.*, [38], has proposed a model based on content-based retrieval for histopathological slide images. It has been able to extract features of high level that reflects the semantic content of the images. This task has been accomplished by a semantic mapper that has mapped conventional low-level features to high-level features by using the state-of-the-art machine-learning techniques.

Boundary-based approach

The trivial method to determine the boundary points is manual segmentation. A number of points are taken from the user and a closed curve is approximated from these points. In this citation by W.S.Rutkowski *et al.*, [39], an approximate closed curve is obtained by joining the arcs characterized by every three successive points or nodes representing segments. This approach leads to successful segmentation since it requires extensive user interaction. However, it is not feasible for large scale use (for a large number of cells). One of the methods that is categorised under boundary based approach is Active Contours.

Active Contours

In C. Demir *et al.*'s citation [2], the active contour model has been proposed to determine the boundary points. A function is used that gives the minimum energy level that is used to form the boundary of a cell in cell segmentation. The initialization of the contour points and the end point specification by the user helps to measure the grey-level discontinuities along the curvature of the "snake" (s) that defines the function over the snake points "s".

In the citation S. C. Agner *et al.*, [40], a new spectral embedding based active contour (SEAC) method of segmentation producing excellent results in comparison to the fuzzy c-means has been presented. This model has integrated temporal contrast enhancement information with spectral data embedding. This model resulted in strong gradients at object boundaries because the active contours are evolved on the histopathological image slide determined by embedding of the series of data in conjunction with stronger boundary gradients resulting in improved hybrid active contours. Hence in this citation description of the hybrid active contour model has been shown which is obtained by combining the boundary region driven from the spectrally embedded data space.

In the citation by A. Basavanhally [41], active contours have been utilized to fix the gradient (edge) information and automatically isolation of the nuclei in histological slides is done. This method, however, might be limited in its ability to handle variations in the appearance of nuclei by their smaller size, more circular shape, and a darker, homogeneous staining to distinguish the diseased nuclei from the other cell nuclei.

Clustering

Clustering is a type of segmentation on the basis of a particular trade native to that cluster or group. In layman's terms it is the grouping a lot of similar variables under a common head. After collecting a raw data and before carrying out the hypothesis the data is then segmented or the variables are classified into various classes according to a common features belonging to them. Before performing any kind of hypothesis on a particular data, in partitioning an unlabelled data, clustering algorithms helps to map the subset of dataset(say $X=\{x_1,x_2,\dots\}$).It has been experienced that the boundaries between the class of real objects to follow the classification.

In this approach, it will consider the clusters as sub-images that consist of a bunch of cells and their properties to determine the clustering coefficient.

In the words of S. Naqvi *et al.*, [42], clustering has been defined as "the organization of data points of multidimensional space into groups based on some similar criteria". The PCA (Principal component analysis) and FCM (Fuzzy c-mean) clustering algorithm enabled to develop the novel grading method and pattern recognition in image analysis in differentiating between the diseased tissues from the rest. The NPI (Nottingham prognostic index) it enabled

validating and determining the grading of the different tissues, which is represented by the following:

$$\text{NPI} = (0.2 * \text{tumour diameter in cm}) + \text{lymph node stage} + \text{tumour grade}$$

According to P.W Krooshof *et al.*, [4] has represented ,clustering of a large dataset of the multivariate image based on the hybrid model hierarchical clustering, which hierarchically groups representing 240 clusters up to one cluster. This helps to determine the spatial relationship between the pixels to result in stable and unique clustering for classification of the regions.

In B. Chen *et al.*'s [43] citation, he has explained the use of fuzzy c-mean clustering algorithm which has been applied on histopathological slide images to produce promising results in an unsupervised data to recognize the meaningful groups that identify similar items respectively. This leads to the ability of the algorithm to assign one data point to more than one cluster.

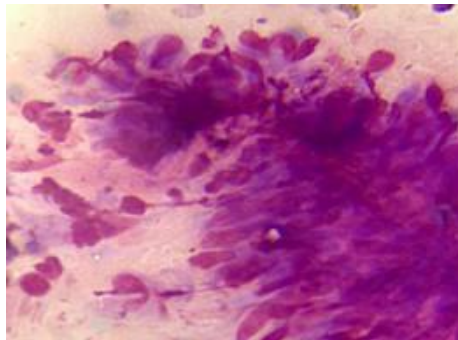


Figure 3. Image of the cell after clustering

4.3. Feature-Extraction

The morphological features provide information about the size and shape of a cell. The size is expressed by the radius, area, and perimeter of the cell. On the other hand, the shape is expressed by the compactness, roundness, smoothness, colour, and texture, length of the major and minor axes, symmetry, concavity, and perimeter. After segmentation of the image, the objects can be extracted before features of the cell are analysed.

- Colour-Initially regions of interest must be stained using some chemicals. Then a color based segmentation algorithm can be used to extract those features.
- Texture-Texture has been used in slides to characterize different kind of histology and biopsy images.
- Intensity-based-This feature of a slide is extracted from the grey-level or colour histogram of the image that provides information about the spatial distribution of the pixel.
- Morphological-The morphological features provides some vital information about the size and shape of a cell.

4.3.1. Colour

Colour based features are important in histology applications since the biologists stain tissues to highlight special structures. There are many colour spaces used to apply in analysis

of histology images. After applying some of the chemical reagents like haematoxylin and eosin (H&E stain). Others like Trichloroethylene (TRI), HNO₃, Cadmium (Cd), and aqueous solution of CdCl₂. Several changes are observed like the Haematoxylin stains the nucleus of the tissue cells dark blue to black. Eosin stains the interior of the cell surrounding the nucleus pink and the tissue outside the cell dark pink to red. The colour information from H & E stained image before grading is done leads to yielding colour spectrums of these images which have considerably limited dynamic ranges and thus obtain the colour map corresponding to the different grades (like Grade-I, Grade-II, Grade-III). Thus quantized images are used to construct the co-occurrence matrix, to label classes and thus enhance the performance. Then this approach is compared to the grey-level and uniform colour quantization approach. The stained images and their segmentation into different grades helps to conclude that the blue colour corresponds to the nuclei, cyan to cytoplasm, yellow to extracellular material and red and grey to background and red blood cells, respectively.

S.Handschuh *et al.*, [44] remarked about the method that can yield "true"-coloured high-resolution view of tissues in three dimensional format and which has the capability to show the structure which are cellular as well as sub-cellular in nature thus it becomes a powerful tool for investigations performed morphologically and also for investigations carried out for both developmental purposes and comparative uses. He concluded that the approach will bridge the vital gap that exists between three dimensional imaging that is done micro anatomically and the methods used in visualisation. This is done by framing a combination of resolutions which are histological in nature and then differentiation of the details is performed by rendering these whole samples of tissue with three dimensional rendering techniques.

4.3.2. Texture

Proper grading and characterization of the tissue is important in image segmentation before diagnosis in human body, tissues are characterized with the organization of their components i.e. their texture determines their changes in these organization of tissue. Pathologist sometimes visually examines the changes in the tissue and examination relies on the visual interpretation.

In the citation by K. P. Kannan *et al.*, [45] in his citation has described the on the arrangement and structure of cells, "GRLM based texture features" (Graph run length matrix) and "GLCM based texture features" (grey level co-occurrence matrix) which has been used to follow segmentation of the abnormal tissue. Thus, one finds the distance between the neighbouring pixels and the other has used the edges of the generated graph to extract meaningful data from the matrix.

The citation by M. M. Rama Krishnan *et al.*, [46], texture based diagnostics, describes a systematic quantitative method based on segmentation of epithelium tissues followed by pre-processing, textural feature extraction using wavelength transform method for feature selection. Finally diagnostic and classification using Bayesian approach and support vector machine that attempts for texture quantification are based on wavelet transform for early detection of disease. In Gabor's filter based texture analysis is performed on histopathological images more than one measure for classification is applied as well, such as using spatial and frequency texture features for classification by regression trees. Here, fractal dimension is used to show the textural variation in the image.

The quality of digitized images has been focussed in the citation by Kayser K. *et al.*, [47], that are estimated by spatial independence and local filtering operations that investigate in brightness, low peak given to noise ratio, full range of grey scale values, maximum gradients, equalisation of grey value distribution, and existence of grey values for thresholds.

Transforming of the RGB space into the hue-saturated-intensity (HSI) space allows detecting of colour and intensity of maxima and minima in a slide.

4.3.3. Intensity-Based

The group of adjacent cells with certain specific property values is set up between the dark and the white shades of grey coloured cells represented in a histopathological slide image. The correlation of the colour feature defines an instance of the size class, thus this way the intensity of these coloured cells determines the affected cell from its surrounding cluster of dark cells.

In the citation by S. Petushi [48], by the intensity of the pixels that are registered and then the mean is found of the neighbouring pixels. This way the characteristic grey scale intensity and the size for each structure are used to generate a labelled dataset. From this dataset the intensities of different types are formed into clusters and a graph is constructed with maximum intensity values on x-axis and y-axis before segmentation is carried out to identify the ROI, *i.e.*, segmentation divides image into several integral regions or objects. One of the aims to perform segmentation is to obtain the regions of interest (ROI) depending on the image and its characteristics. The approach is to partition an image based with abrupt variations in intensity levels into different regions, such as edges in an image and the partitioning of the image into regions that show similar intensities based on certain predefined criteria) and separate it from the background.

According to the citation of A. E. Tutac *et al.*, [49], a property named 'hasIntensity', is used to define the colour features as it appears in a histopathological slide image. A cluster is defined containing a group of adjacent cells with intensity property values setup between very dark and white limits. After the image is analysed the very dark coloured cells are identified, thus the intensity based segmentation method helps to detect the diseased cells from a slide.

In the citation by A. Basavanhally [41], the random variable denoting features the square root of area and standard deviation, helps to measure the intensity for each region. Here in the region growing algorithm the steps involved includes identification and removal of the current boundary of the pixel with low intensity values. In another step the mean intensity of the pixels are computed along with several other intrinsic properties like size, shape and their arrangements to classify a histopathological slide image.

In another citation by C. Demir *et al.*, [2], the intensity based approach is employed to calculate the intensity value of pixels to define the features in a histopathological slide image. Feature extraction has been carried on from the grey-level or coloured histogram of the image where no provision is made for information about spatial distribution of the pixel values. This method has been employed to define features of the cell and the optical density of the pixels by making use of its grey values in the images that converts the grey value of each pixel present in the nucleus to that of its corresponding optical density (OD) value, represented by

$$OD = \log_{10} \frac{\text{Grey value of the background}}{\text{Grey value of the pixel}}$$

After pre-processing according to A. D. Belsare *et al.*, [3], the segmentation process leads to separation of the objects of interest from the background by thresholding but the intensity based approach has been used in separating the regions to extract the relevant features. In case of a single cell, the intensity-based approach enables features to be extracted at the tissue level quantifying the distribution of the cells across the tissues making use of the spatial dependency or the grey-level dependency of the pixels.

4.3.4. Morphology

Several morphological features like the area, perimeter, radius, concavity, symmetry, centroid, major axis and minor axis –these are considered by the pathologists to study the characteristic of the cell or tissue. In the citation by A. Basavanhally [41], quantification of the morphological features and the arranging their structures in a histopathological slide image leading to discrimination of a cell into a particular class for the purpose of diagnosis has been shown.

These features help us to determine the roundness, smoothness and compactness from 'S'= $\{s_1, s_2, \dots\}$, the set of boundary points of the segmented cell. Some of the common features are:

- Area: the number of pixels composing the objects is considered.
- Perimeter: In the frontier of the object, the numbers of pixels are calculated.
- Centroid: the coordinates of the centre of gravity is found.
- Major axis: the lengths of major and minor axis are computed.
- Major axis direction: angular orientation of the line of best fit through the centre of gravity is measured.

4.4. Feature Selection

The feature selection techniques used to process histopathological slide images of a wide range of tissues which is oriented mainly to find the architectural features which stands on domain specific assumptions such as the ability to segment meaningful objects with particular shapes. Approaches such as analysis in image sub-blocks or global feature selection have been also been discussed. This would be useful to design strategies for image content representation of histology applications like:

- Linear feature reduction-This method is a regularized linear models approach that makes use of a partial supervision on the features a priori assumed to more relevant.
- Non-linear feature reduction-The method is used for characterizing the underlying parameterization in the distribution of shapes of the cells in a histopathology image.
- Gamma-correction-A nonlinear operation is used to code and decode luminance values in an image system.
- Graph-based-A graph-based method is used to map structures with nodes, edges, and properties that are distinct from specialized graph constructed for histopathological slide images.

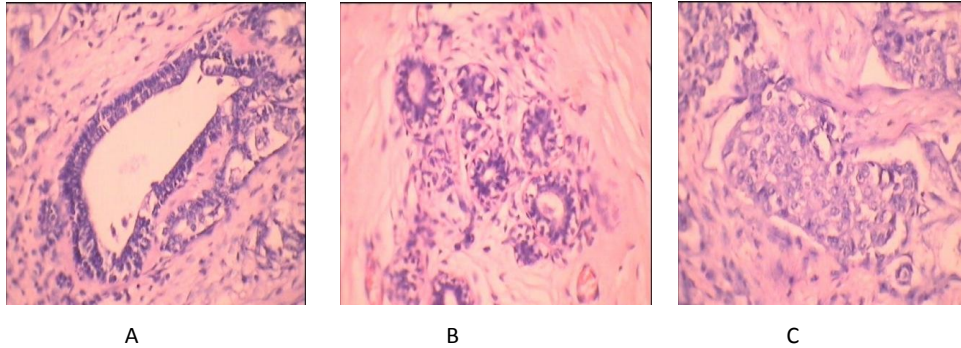


Figure 4. Showing the stained cancer images in a histopathological slide in A and B normal duct and lobules present, in C malignant cancer images are visible

4.4.1. Linear Feature Reduction

The citation by Dr. W. Cai *et al.*, [50] shows that, when the size of the database is small, a linear search can enhance performance an indexed structure helps to filter out all irrelevant images from the database. The linear quad trees one such indexing approach used. The PCA (Principal component analysis) is an optimal technique which linearly maps input data to a coordinate space such that the axes are aligned to reflect the variation in the data and the linear scaling of the image to correct size and nonlinear wrapping to minimize regional variations.

In the citation by P. J Kr Baladevbhai *et al.*, [51], the spectral graph partitioning as a linear feature reduction provides a powerful approach to image segmentation. In this method colour quantization has been used to form initial segment regions and then finally the initial regions are merged based on the distance defined by the colour and spatial information. Yonggang Li *et al.*, [52] presented a method in the year 2008 of colour image segmentation based on DPCNN. It uses the capturing characteristics of PCNN, and has a good adaptability. The method of colour image segmentation is proposed by Yonggang Li based on Double Pulse-Coupled Neural Networks (DPCNN) in their paper. First, colour quantization is performed on the colour images in the HSI colour pattern. Then the method uses the result of colour quantization as the input to form the initial segmented regions.

The citation of feature Selection by T. Drugman *et al.*, [53], focuses on the audio visual data integration and the most popular methods used are LPC (Linear Predictive Coding), MFCC (Mel Frequency Cepstral Coefficients) which resolves frequencies in a non-linear manner. Also the PCA algorithm to focus on the pixel values of the ROI and concentrate on the essential information on a few coefficients motivated the feature selection. The last step towards classification of tissues to set feature are the neural network, examples of such network are non-linear models that capture complex interactions among the input data along with noisy and irrelevant information.

According to K. Kavukcuoglu *et al.*, [54] has shown how the filters in multi stage convolution network architecture improve the recognition and detection in a histopathological slide image. Also how the nonlinear encoder produces a fast approximation and optimizes the method. The techniques used is applicable to Convolutional Networks and their variants, using multiple stages of trainable convolutional filter banks, intermingled with non-linear operations, and spatial feature pooling operations.

Alison A. Dingle [55], showed how the neural network has been used as a nonlinear filter to remove noise that lies in the same frequency range they perform better than the linear

filters. Though they are unpredictable, the simplest chaotic system is still the nonlinear recursive equations. The techniques used here is applicable to Convolutional Networks and the variants associated with them, using multiple trainable convolutional filter banks in different stages, intermingled with non-linear operations, and spatial feature pooling operations.

The citation on Image Recognition by Y-Lan Boureau [56], has concluded that choosing non linearity and pooling function is often more important than training the weights-it leads to spatial pooling over a local ROI using a linear SVM(support vector machine) improves the average pooling process and a better histogram intersection kernel of the slide. The intersection kernel used here is much more clear and better than the linear kernel when average pooling is performed. The results have been included for both to preserve the symmetry between them. Some characteristics of sparse coding may serve as starting point in the following cases, like:

- Inputs are reconstructed as a linear combination of basic functions instead of just a copy of one basic function;
- Most basic functions are rarely active across inputs;
- Basic functions tend to be active for inputs that are similar to one another;

In the image analysis by Anlei Dong *et al.*, [57], the transformation for feature dimension reduction has the characteristics like the non-linearity of the image that does not assume the distribution of class in the original feature space and that it yields a transformation in a dimension which leads to searching in a low dimensional feature space.

4.4.2. Non-Linear Feature Reduction

In the citation by C. Farabet *et al.*, [58], the likelihood that in each segment a single object is contained and the feature associated with the segments -the nonlinear mapping of each feature map is found out. This method has used a dense trainable, multi scale extractor, segmentation tree, region wise aggregation, class histogram and optimal purity cover. The steps used in this process have a complexity which is linear in the number of pixels. The bulk of all the computational works resides in the convolutional network feature extractor, resulting in the system becoming very fast. The three key contributions in observed in this paper “(1) using a multi-scale convolutional net to learn good features for region classification; (2) using a class purity criterion to decide if a segment contains a single object, as opposed to several objects;(3) an efficient procedure to obtain a cover that optimizes the overall class purity of a segmentation”. For each scale s , the convolutional network has been described as a sequence of linear transforms, interspersed with non-linear symmetric units.

The citation by F. T. Ramos [59], in his thesis presents techniques for combining non-linear dimensionality reduction with parametric learning through Expectation Maximization has been used to build general representations of natural features in a histopathological slide image. Thus, this framework is developed with non-linear dimensionality reduction techniques associated with statistical learning like the stochastic representation of objects through non-linear dimensionality reduction. A non-linear statistical model encoding a neighbourhood preserving dimensionality reduction has been proposed for representing features in any ordinary natural environments.

In the citation by G. Mesnil *et al.*, [60] , has shown the strategy of the one layer learning algorithm - a simple linear classifier with a tiny number of labelled training samples of data of k linear filter. The nonlinear method has various advantages over the linear in creating a sparse representation with zeros and helping in faster computation. This technique is

important particularly when the classifier receives this representation as input which is linear and when the number of available labelled examples is small. This is the case here with the unsupervised data.

4.4.3. Gamma Correction

It is an algorithm to code and decode the luminance (grey) or tristimulus (colour) values in still images. It is an efficient technique of adjusting the range of gamma values which are compared with each other to detect the ROI in a histopathological image slide.

In the citation by V. Veeramanikandan *et al.*, [61], gamma correction has been described as a special type of nonlinear grey-level transformation used to correct the power-law transformation phenomena which perform the transformation of an input image to its original appearance. This transformed gamma-corrected image is free from the darkness by compressing all the dark regions into bright regions. It replaces the original grey-level with the one derived, by considering gamma value (γ) > 0 , but lies between 0 and 1 (*i.e.*, $(\gamma)[0, 1]$). The underlying principle behind the gamma correction is that the intensity of the light reflected from an object is the product of the incoming illumination and the local surface reflectance. Here, the obtained image after gamma correction should be an illumination free image. The gamma value from 0 to 1 is considered to be a full log transformation which is very strong to convert the dark regions. Hence the value of can be range from [0, 0.5] by default.

In the citation by C. Tay [62], after detecting a region of interest (ROI), and then a gamma correction is applied to the red channel of the ROI. Gamma correction can change the brightness and the ratio of red, green, and blue of an image; it highlighted the cell nuclei from the background.

In the citation by K. Shibuya *et al.*, [63], after applying noise reduction was applied to suppress the electrical noise effect and the gamma correction was performed to make the image data linear in addition, to the filter treatments using a digital filter for images were applied to extract the ROI according to the diameters.

4.4.4. Graph-Based

This method helps to identify and compute quantitative metrics for tissue characterization and classification by generating different cell graphs using the positional coordinates surrounding the matrix components. The set of designed numerical operators is based on a specific dilation operator making it possible the representation of concepts like “between” or “left of” over sparse representations such as graphs then the structural analysis in a histopathological image slide is carried out.

In the citation by S. Naik *et al.*, [64], following the diseased region detection a low-level and empirical domain information, this area is initialized as a level-set curve, which is evolved to lie at the interior boundary of the nuclei surrounding the gland structure. Then the features are calculated from the boundaries that characterize the morphology of the affected regions, including area overlap ratio, distance ratio, standard deviation and variance of distance, perimeter ratio, compactness, smoothness, and area are found by this graph embedding technique. The feature space is reduced using a manifold learning scheme called "Graph Embedding" is used to embed objects that are adjacent to each other in the high dimensional feature space into a lower dimensional embedding space. Objects embedded in this low dimensional embedding space are then classified using a support vector machine (SVM) classifier as belonging to the benign epithelium is done. The efficacy of the automated segmentation algorithm by comparing the classification accuracy obtained using the automated segmentation scheme to the accuracy obtained through this technique.

The citation by I. Nakajima *et al.*, [65], showed how graph theory provides appropriate tools to apply image transformations in order to detect symmetries or repetitive arrangement of objects such as “rings”, “lines”, “stars”, etc. in an image slide. Their centres of gravity can be considered as “new objects” located within a different magnification space which can reallocate a new set of features to be applied in the diagnostic algorithm.

In the citation by K.P. Kannan *et al.*, [45], to segment the abnormal area region growing algorithm has been applied on the image in which the nodes are considered as pixels. The region based method grouped the same pixels into cluster. The proposed approach uses, Graph Run Length Matrix (GRLM) along with Grey Level Co-occurrence Matrix (GLCM) for accurate segmentation to extract the features colour graph was generated on the cytological components and then, defines a run-length matrix using the edges of the generated graph. Finally extraction of the set of texture features from the graph run-length matrix.

The problem of technical variations that arise while manually executing information from the reference patches of the original image by H. Chang *et al.*, [66] has been dealt by incorporating information from manually annotated reference patches and the local colour space of the original histopathological image slide is formulated within a multi-reference graph cut. In the approach there has been validation on manually curated samples and then an application is done on a set of data which has sections comprising of approximately 440 tissue samples which are typically ranging between 40k-by-40k pixels or larger. This representation is then regularized with the spatial smoothness constraint through the graph cut framework. Nodes and edges corresponding to special nodes known as terminals, pixels and their adjacency relationship, respectively which corresponds to the labels comprised of a full set of data that has the possibility to be assigned on to the pixel values. If the graph has two terminals, then one of the two is marked as source (S) and the other terminal is termed as sink (T) are labelled to find the results.

4.5. Classification and Diagnosis

After performing pre-processing and feature extraction, the image feature selection and classification of extracted features are required to be performed for abnormality diagnosis, if present in the image. In this step the cell or tissue is classified for malignancy level, *i.e.*, grading of disease is done. With the histological slide images algorithms like supervised, unsupervised, neural network, k-nearest, fuzzy systems, *etc.* are performed.

- Supervised & Unsupervised- It finds a set of mutually orthogonal basis functions that captures the directions of maximum variance in the data. These features may not be useful for discriminating between data in different classes.
- Neural network- Neural networks, which are composed of artificial neurons or nodes. Its structure is based on external or internal information that flows through the network.
- K-nearest neighbourhood- K-Nearest neighbour is a classification strategy that uses the training set directly to classify an input when an input is given.
- Fuzzy system-Fuzzy system includes 0 and 1 as extreme cases of truth but also includes the various states of truth in between so that, for example, the result of a comparison between two things.

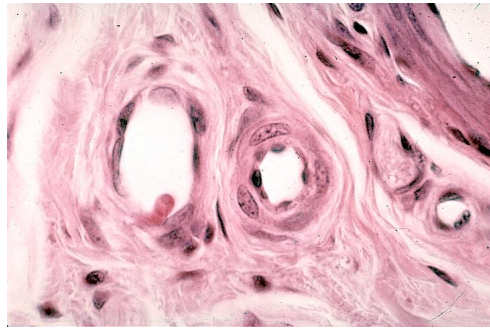


Figure 5. After classification, the cell is ready for post-diagnosis

4.5.1. Supervised

In the supervised technique the following technique is followed:

1. Operator chooses and colours the tissue sub region as training data .The supervised set consists of the classifiers in a separate dataset (say, D).
2. Train classifier D on training data.
3. Classifier D labels and colours all remaining pixels automatically.

After determining an appropriate set of features, the next step is to distinguish the malignant structures from their counterparts. In this step, a cell or a tissue is assigned to one of the classes of cancerous, benign, or healthy. As a part of diagnosis, it is also possible to classify the malignancy level of the tissues (*i.e.*, grading). In this case, the classes are the possible grades of the cancer of interest.

Another group of studies uses machine learning algorithms to learn (from data) how to distinguish the different classes from each other. Among those algorithms are the neural networks, k-nearest neighbourhood algorithm, logistic regression method, fuzzy systems, linear discriminate functions, and decision trees. The lists of the studies that use these algorithms are discussed.

In the citation by G. M. Reynolds *et al.*, [67], this method has been used as a supervised machine learning which has developed a classifier from the database of cases for which this diagnosis has already been known. However, the developing classifiers based on mixed modalities, like combining imaging information which has generated better probability of a class containing set which helps in detecting benign tissue.

This newer approach by J. A. Ozolek *et al.*'s [68],achievement inside a strategy which is supervised learning and in which the pixels of some images that are segmented has been used in training a classifier that is capable of accurately determining the class (background, tissues of interest) of each pixel images which are unlabelled of the similar kind. These algorithms used a pre-defined window size at a pixel which has given values and within the window is considered as neighbours and intensity neighbourhood vectors are formed by reordering the intensities of the pixel values into a vector for those pixels which are lying inside the window. By using this method better classification and accuracy level has been achieved significantly across the board for several tissue prototypes.

4.5.2. Unsupervised

The pattern recognition technique involves process description, feature analysis, cluster analysis and classifier design.

In the unsupervised technique the technique followed are the following:

1. Clustering the whole regions into sub regions with algorithms lemnda.
2. Lemnda colours each region with one or more colours.
3. Operator views colour region and assigns tissue labels to each.

In the citation by Koray *et al.*, [69], the input of unsupervised training has a combination of normalized contrast grayscale images obtained from the segmentation dataset. Contrast normalization has a processing feature map value which is obtained by removing the mean and then division is done by the standard deviation that is calculated around 9×9 centre region at that value all over the feature maps. Then the first layer is obtained which has a 9×9 convolution kernel. Then the systems is trained for 10 different sparse values ranging between 0.1 and 3.0. Then the second layer is found after processing the output of the first layer as input, hence the number of feature maps rounds to 256 and each feature is connected and then it proceeds to the next stages. In an unsupervised dataset clustering method helps to assign colours to each pixel, *i.e.*, what colour to be used and how to colour the pixels according to the class labels. This step of segmentation needs careful attention for improvement of the visual display issues which has greatly affected the utility of the computed output.

4.5.3. Neural Network

Artificial neural networks is characterized as a computational model with particular properties such as the ability to adapt or learn, to generalize, or to cluster or organize data, and which operation is based on parallel processing. Within the systems it is useful to differentiate the three specimens of units: the units which are namely used as input to receive data from the exterior of the neural network, the unit for output and hidden units which have their input and output signals concealed inside the neural network. The ability of this network is to represent a desired function built from a set of standard function.

In the citation by B.Krose *et al.*, [70] has shown how neural network increases the processing capacities of an image and has the ability to generalize or form clusters to an organized data. This network is configured in a way such that the application of a set of input produces a desired output. In this citation it incorporates the follow-the-leader clustering algorithm to fit the new input pattern generated into an existing class. The success of this neural network was in dynamically controlling a robot arm described by Kawato, Furukawa & Suzuki in 1987; this described a network which generated motor commands

The Neural network method allows hidden layers to be configured and promises to give better results to produce the tissue region of interest. The unsupervised methods use techniques like Bayes' classifiers with unlabelled MLE (Maximum likelihood estimator), HCM (Hard c-mean), FCM (Fuzzy c-mean), AFCM (Advanced Fuzzy c-mean) and SFCM (Semi-supervised Fuzzy c-mean). Other approaches like the Kohonen network, ART (Adaptive Resonance theory), hybrid are also used.

4.5.4. K-Nearest

One of the simplest nonparametric method for classification considered is the k-NN method where the rule classifies x , a dataset by assigning it the class label (say 'c') most frequently represented among the k nearest prototypes; *i.e.*, by finding the k neighbours with the minimum distances between x and all prototype feature points in a histopathological slide image. The error rate of this k-nearest technique to the Bayesian technique is almost infinity

.The KNN method of classifier assigns to an unclassified object of the class label, which has the majority voting amount in the set of its k nearest neighbours to obtain the best neighbouring value among all the pixels.

For consistent colouring of the ROI, accurate performance of the technique is important to be set across the tissue slides. The unsupervised techniques are comparatively slower to yield optimized result, whereas the k-nn rule is a faster technique applicable on different kinds of tissue slides in a histopathological slide image. The k-nearest neighbour rule relies on having a large number of correctly labelled samples from each tissue class. It is needed to find the k – value, the nearest neighbourhood of any unlabelled pixel vector z in X_t and measure the distance between pairs of vectors using the Euclidean distance.

In the citation by P. P. Rezaeiye *et al.*, [71], KNN rule for classifying each unknown sample of x in an n dimension space the K value finds the nearest neighbour based on a distance scale and among these. K finds nearest neighbour of the class that contains majority of attribute x that belongs to that class. If this majority is unable to make any kind of decision in finding the K nearest neighbour of x then it starts searching for a long distance from that, by repeating this process there might be uncertainty in decision making. So in their paper an introduction of K nearest neighbouring method for classification is done which is an appropriate feature extraction that could be used to improve speed and recognition ratio of classification.

In L. M. Fletcher-Heath *et al.*'s [72] citation , the fuzzy clustering algorithm is used as an initial segmentation technique for an unsupervised vector the k-nearest neighbours groups pixels based on their similarities as seen in image features then selection is done from the training data of various tissue types per slice per volume. This is a supervised technique whose accuracy depends on the representation of the operators used. When a radial line does not intersect a white or grey matter pixel in an image then the average length of their neighbouring lines which intersected white or grey matter has been chosen as a post detection of disease technique.

In the citation by K. S. Ni *et al.*, [73], an algorithm which has used interpolation techniques and which is both nonparametric and based on learning with the use of an adaptive k-nearest neighbour algorithm within a random field is proposed here. The algorithm ensured results from the image that are driven by data and that reflects “real-world” images in them, given in a training data. Thus this proposed algorithm is a dynamic one, which differs from pixel to pixel, like, small for test points with highly relevant neighbours and large otherwise. Based on the neighbours, a weighted minimum mean squared error solution is determined implicitly where the definition of filters are those which are low resolution specific images and which does not yields to the insufficient training limitations. Detailed analysis and a comparison of the existing methods is drawn parallel between concepts which are derived intuitively and includes classification of ideas which are introduced by similar other nearest neighbour algorithms that explains manifolds in low and high dimensions.

In the citation by W. E. Henley *et al.*, [74], the KNN is a standard technique of pattern recognition that attempts to incorporate class separation contained in a data. This method has been found to perform better than any other non-parametric techniques and it has achieved the lowest expected bad risk rate. A linear, logistic regression and a decision graph is used to design the set with equal proportion of good and bad risks and thus the superiority of this approach reflecting good results is observed.

4.5.5. Fuzzy System

Fuzzy-connected segmentation technique permits rapid, reliable, consistent and highly reproducible measurement of a histopathological slide image. Fuzzy approach has been found

to yield more accurate decision in comparison to their counterparts. The value of the parameters and the affinity relation from a given set is determined using fuzzy logic.

G. Moonis *et al.*, [75] has adapted the fuzzy connectedness algorithm segmentation technique to measure the volume of the disease. It attempts to capture this notion in a computational framework and uses the training facility available in a 3D view. The ability to handle different degrees of partial volume effects that the boundaries of an image are subjected to is handled by this method.

L. M. Fletcher-Heath *et al.*, [72], the fuzzy clustering algorithm is used as an initial segmentation technique for an unsupervised vector set. The FCM (fuzzy c-mean) clustering method does not require training data and provides knowledge on how well a pixel fits a cluster. It yields consistent partitions, describing the specific shape in image analysis. The edges are detected and the boundaries in an image can be obtained. It has been found the value of pixels in the cluster which enables us to display the coloured regions that fade into adjacent regions.

In the citation by M. S. Kalas [76] has developed the fuzzy logic for classification and estimation of the diseased cell affected in a histopathological slide, imaging. Its approach is found to be more accurate in decision making as compared to its counterparts. The features obtained after performing this technique are processed using a fuzzyfication layer before passing the output results to a neural network. Thus, this Research has represented an automatic recognition system for the histopathological slide images using the neuro fuzzy logic.

4.5.6. Morphological Function

Several morphological features like the area, perimeter, radius, concavity, symmetry, centroid, major axis and minor axis –these are considered by the pathologists to study the characteristic of the cell or tissue. This feature helps us to determine the roundness, smoothness and compactness from the set of boundary points of the segmented cell.

In the citation by A. Basavanhally [41], the need to quantify the morphological features and the arrangements of their structures in a histopathological slide image which discriminates a cell into a particular class after the detection of disease for further diagnosis.

5. Discussions

The success of the methods applied at various stages of segmentation, thresholding, feature extraction and classification depends on less time taken, usage of overheads as well as practical applicability of these suggested methods. Hence, discussion of the different techniques and analysing their problems and key features which can be worked upon to make this system of diagnosis better and more effective in the near future.

In the previous section three different techniques of pre-processing technique were discussed. Colour illumination method was described by M. Larsen [16] in which the authors have enhanced the features of images using absolute colour values and S. Tamai [17] had used similar method with better performance on pathological and cytological slide images. Another technique commonly used is smoothing. Gaussian smoothing is performed to achieve smoothing of image features by all the authors as described in the Section V. Roullier *et al.*, [18] used a method that preserved the main elements before applying Gaussian filter whereas M. Habibzadeh *et al.*, [19] have reduced surface angularities to reduce the visibility of some features of the said image. Sertel *et al.*, [9], have applied the EM method to establish the distribution of intensity before application of Gaussian on the image. All the denoising algorithms have preserved the structures after removing the artefacts from a digital image. A. Buades *et al.*, [20] have used the powerful ordering and evaluation method in

statistical image model than N. Joshi *et al.*'s two colour model to blur and decouple the sharpness of the edges and N. Kachouie's recovery of signal to remove noise. Dr. Berbar's [24] method of routine thresholding of pixels into two groups of 0's and 1's whereas Otsu's algorithm separated the pixels into several groups but also has achieved cent percent accuracy than other methods.

D. Bradley *et al.*, [23] showed the adaptive thresholding model which is robust and easier method to characterize pixels into RGB values to identify the edges in an image slide. Enhancement improves visualization of the stained images, G. Loukas *et al.*, [12] has been successful to identify the spectral changes observed in the cell nuclei along with the usual technique of enhancement. Both C.F. Jiang *et al.*, [25] and M. Tahtouh *et al.*, [13] have shown a comparative method of enhancement but the later technique has produced excellent ridge details by following the finger-mark enhancement in a slide image. A better approach by M. Haindl *et al.* [35] has been shown where a simple and classical method of subsequent splitting and merging of regions into smaller ones. The geometric deformation model by J. S. Suri *et al.*, [36] is less time consuming as it has been approached mathematically. W. S. Rutkowski *et al.*, [39] has shown a tedious method of boundary based segmentation which is extensive and not feasible for large scale images. C. Demir's [2] boundary based approach has used a function over the snake points but has paid specially attended the grey -level discontinuities to initialize the contour points. Hence it requires minimal energy and less manual effort whereas, A. Basavanhally's [41] model is limited in scope to fix the gradients since it is incapable to handle various shapes of nuclei to identify the diseased ones. The PCA and FCM approach by S. Naqvi *et al.*, [42] and B. Chen *et al.*, [43], has helped in better and meaningful grading of different tissues, whereas, P.W Krooshof *et al.*, [4] has presented a unique hierarchical clustering of the regions based on hybrid model.

S. Handschuh *et al.*, [44] showed a powerful tool of high resolution view of cells and tissues is helpful in differentiating the malign cells. A. Basavanhally [41] showed a faster technique identifying morphological features of a cell like the area, perimeter, etc. along with other characteristics are studied leading to cellular discrimination. By finding the distance of the neighbouring pixels features are extracted from an image, similar approach shown by K.P Kannan *et al.*, [45] through GRLM and GLCM methods but meaningful data is extracted from the matrix. M.M Rama Krishnan's texture based Bayesian approach to show the textural variation in an image but Kayser K *et al.*, [47] has detected the colour and the intensity both from a slide with preservation of the quality of the image slide as well in an effective way. S. Petushi [48] has used a graph to represent the maximum intensity of pixel values after separating it from the background to find the ROI hence this method is easy to understand. A similar approach by A. E. Tutac *et al.*, [49] but within the intensity value limits of very dark and white pixels enabled better detection of the diseased cell. A. Basavanhally [41] has attempted to remove the low intensity values by calculating the square root, standard deviation and mean intensity of pixels for each region, hence classification becomes faster. Though C. Demir *et al.*, [2] has found the OD value after finding the grey values of each pixels but it lacked provision for information about spatial distribution of pixels. The same problem is also faced in the technique of A. D. Belsare *et al.*, [3] where the intensity-based approach worked successfully but only in case of single cell.

Dr. W. Cai *et al.*, [50] has used the PCA to map input data to the coordinate axes, thus indexing has helped in filtering out the irrelevant images from the database. Y-LanBoureau [56] has shown a better method than non-linear feature models by using linear SVM but with several constraints added. Both A. Dingle [55] and T. Drugman *et al.*, [53] focused on audio-visual frequency ranges but somewhere the process of A.Dingle's has got intermingled with the non-linear operations, whereas Drugman has captured the complex interaction among

input data. The non-linear method of feature reduction by using feature maps according to C. Farabet *et al.*, [58] associated with the segments and construction of a class histogram has made this procedure very fast and efficient in providing an optimized class. F. T. Ramos [59] and G. Mensil *et al.*, [60] have used the k-linear filters to classify tiny labelled samples of data representing stochastic objects through this method. Though G. Mensil *et al.*'s method speeds up computation process but the model by F.T Ramos can be applicable in any ordinary natural environment. V. Veeramanikandan *et al.*, [61] described an efficient and special nonlinear grey level transformation technique using the range of gamma values to convert the dark regions but one limitation is that the image must be an illumination free image. K. Shibuya *et al.*, [63] has tactfully and in a skilled manner has performed gamma correction using digital filters for images to extract ROI according to the different diameters of the slides whereas this kind of skillfulness is lacking in C. Tay [62] technique that has followed the same procedure but by changing the RGB ratios to highlight the cell nuclei. The cell graphs are plotted surrounding the matrix components, to characterize the diseased tissue, S. Naik *et al.*, [64], has used the graph embedding technique to embed adjacent objects in a high dimensional feature space developing an automatic segmentation technique with cent percent accuracy. H. Chang *et al.*, [66] has carried out the graph-cut method as an adjacent relationship between the nodes and edges that can be assigned to pixels but the vagueness of this technique is that it has been validated only on a very small size of dataset.

Classification of a tissue using supervised machine learning technique is one of the traditional approaches, according to G. M. Reynolds *et al.*, [67] these classifiers are developed based on mixed modalities, combining image information leading to generation of a class with much better probability for detection. J. A. Ozolek *et al.*, [68] followed a similar technique but significantly better accuracy level has been achieved for several tissue prototypes. Pattern recognition from an unsupervised set by K. Koray *et al.*, [69] has been shown in a well organised and systematic manner by grouping the feature mapped values into several layers and into different stages. Consequently, the visual display is greatly affected by the utility of output produced. The neural network according to B. Krose *et al.*, [70] is capable to generalize into an organized dataset and the application can produce output according to a desired output, versatility, led to the success of this technique as it can dynamically control the output. The KNN technique assigns values to objects in such that it obtains the best neighbouring values among all pixels, P. P. Rezaeiye *et al.*, [71] showed similar method but has effectively handled the indecisiveness that might occur if the method failed to attain the nearest neighbourhood value. L. M. Fletcher Heath *et al.*'s [72] has paid attention to minute details after selection is done of various tissue types per slice and volume from the dataset, before choosing the average length of the neighbouring lines for post detection of the disease. K. S. Ni *et al.*, [73] in the proposed algorithm has shown a realistic approach and is dynamic in both low and high dimensions image datasets based on the weighted minimum mean squared result of neighbouring values. W. E Henley *et al.*, [74] has shown a standard technique of pattern recognition where the risk rate associated in designing the decision graph is the lowest hence, lies the superiority of this method to obtain good results. G. Moonis *et al.*, [75] has adapted the most reliable and consistent technique of fuzzy set to obtain the affinity relation in a given set and measure the volume of the disease, the capability to handle all kinds of partial volume effects that the pixel boundaries are subjected to makes this method more practical than others. M. S. Kalas [76] has shown a cent percent accurate and automatic technique of recognition through fuzzy logic leading to better decision making in comparison to its counterparts.

6. Conclusion

Many different techniques exist and continue to be developed, while the ways of detecting the diseased cell or tissues in digital microscopy can be used together, to minimize the detection time taken by pathologists. Proper segmentation techniques will enhance efficiency of pathologist in diagnosis and reduces their time for reviewing large number of tissue slide per day and help provide better patient care and advanced therapeutics. The evaluation of the different statistical methods has also been used to measure the accuracy of our detection methods. These methods can be used for histopathological image processing applications with minor modification.

Our effort in this direction has been to establish an effective and robust CAD system to determine the benign tissue or cell present in a histopathological slide image that exists not only in the morphological structures like colour, texture using the techniques like region-based and boundary-based segmentation techniques but also the changes that occur within anatomical regions. To enhance the effectiveness of this method classification into normal or abnormal tissue has been suggested using clustering, fuzzy sets and neural network. For betterment of evaluation in histology images these steps would assist the pathologists in disease diagnosis and also reduces human error. Hence, to detect any future risk from biopsies will be the future scope of analysis.

References

- [1] M. N. Gurcan, A. Can, L. Boucheron, A. Madabhushi, N. Rajpoot and Bulent Yener, "Histopathological Image Analysis: A Review", *IEEE Reviews in Biomedical Engineering*, vol. 2, (2009), pp. 147-171.
- [2] C. Demir and B. Yener, "Automated cancer diagnosis based on Histopathological Images: A systematic survey", Technical Report Rensselaer Polytechnic Institute Dept. Of Computer Science, TR 05-09, CiteSeerX β , (2005), doi: 10.1.1.61.1199, pp. 1-12.
- [3] A. D Belsare and M. M. Mushrif, "Histopathological Image Analysis using image processing techniques: An Overview", *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 3, no. 4, (2012), pp. 23-31.
- [4] P. W. Krooshof, G. J. Postma, W. J. Melessa and L. M. Buydens, "Clustering of Brain Tumors by Segmentation of Magnetic Resonance Data", *Proc. Intl. Society Magnetic Resonance Med.*, vol. 14, (2006), pp. 1019-1019.
- [5] O. Sertel, J. Kong, G. Lozanski, J. Saltz, M. Gurcan, A. Shana'ah and U. Catalyurek, "Texture Classification using nonlinear color quantization: Application to Histopathological image analysis", *Journal on Pattern Recognition Archive*, (2008) IEEE, vol. 9, no. 1, (2008), pp. 596-600.
- [6] A. Madabhushi, "Digital Pathology Image Analysis: Opportunities and Challenges", *Imaging Med.*, vol. 1, no. 1, (2009), pp. 7-10.
- [7] M. Gan and J. Wang, "Applications of Image Processing Technique in Porous Material Characterization", *Advanced Image Acquisition, Processing Techniques and Application*, vol. 3, (2012), pp. 110-117.
- [8] J. C. Caicedo, "Features for Histology Images", *Image Segmentation*, (2009), pp. 1-9.
- [9] O. Sertel, U. V. Catalyurek, H. Shimada and M. N. Gurcan, "A Combined Computerized Classification System for Whole-slide Neuroblastoma Histology: Model-based Structural Features", *Workshop on Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy*, (2009), pp. 1-12.
- [10] P. A. Hall and D. A. Levison, "Review: Assessment of cell proliferation in histological material", *Journal of Clinical Pathology*, vol. 43, no. 3, (1990), pp. 184-192.
- [11] J. Kazi and M. Mubarak, "Histopathology in present era", editorial from dept. of histopathology SIUT, Karachi, vol. 59, no. 1, (2009), pp. 1-2.
- [12] C. G. Loukas, G. D. Wilson, B. Vojnovic and A. Linney, "An Image Analysis-Based Approach for Automated Counting of Cancer Cell Nuclei in Tissue Sections", *Image Analysis-Based Automated Cell Counting*, vol. 55A, no. 1, (2003), pp. 30-42.
- [13] M. Tahtouh, J. R. Kalman, C. Roux, C. Lennard and B. J. Reddy, "The Detection and Enhancement of Latent Fingermarks Using Infrared Chemical Imaging", *Journal of Forensic Sciences*, vol. 50, no. 1, (2005), pp. 1-9.
- [14] S. Petushi, F. U. Garcia, M. M. Haber and K. A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast -cancer", *BMC Medical Imaging*, vol. 6, no. 14, (2006), pp. 1-11.

- [15] S. G. Kim, H. S. Jang and K. Ju, "Ameloblastoma: A clinical, radiographic, and histopathologic analysis of 71 cases", *Oral Surgery Oral Medicine Oral Pathology*, vol. 91, no. 6, (2001), pp. 622-649.
- [16] M. Larsen, "Colour(and multispectral) images", *Image Analysis 2010*, Dept of Basic Sciences and Environment Mathematics and computer science group, University of Copenhagen Session, vol. 3, (2010), pp. 1-22.
- [17] S. Tamai, "The Color of Digital Imaging in Pathology and Cytology", *Digital Color Imaging in Biomedicine*, vol. 10, (2001), pp. 61-66.
- [18] V. Roullier, O. Lezoray, V. T. Ta and A. Elmoataz, "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization", *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, (2011), pp. 603-615.
- [19] M. Habibzadeh, A. Krzyzak and T. Fevens, "Application of pattern recognition techniques for the analysis of thin blood smear images", *Journal of Medical Informatics and Technology*, vol. 18, (2011), ISSN 1642-6037, pp. 29-40.
- [20] A. Baudes, B. Coll and J. M. Morel, "Image Denoising Methods A New Nonlocal Principle", *Society for Industrial and Applied Mathematics*, vol. 52, no. 1, (2010), doi: 10.1137/090773908, pp. 113-147.
- [21] N. Joshi, C. L. Zitnick, R. Szeliski and D. Kriegman, "Image deblurring and denoising using color priors", *IEEE Pattern analysis and machine intelligence*, vol. 50, (2009), ISBN-978-1-4244-3992-8,doi, 10.1109/CVPR.2009.5206802, pp. 3-8.
- [22] N. N. Kachouie, "Bayesian Model Based Tracking with Application to Cell Segmentation and Tracking", A Thesis presented to the University of Waterloo in Systems Design Engineering, (2008), pp. 236-250.
- [23] D. Bradley and G. Roth, "Adaptive Thresholding using the Integral image", *Journal of Graphics Tools*, vol. 12, no. 2, (2007), pp. 3-8.
- [24] M. Berbar, "Histogram-Based Image Segmentation", *Journal of Computer Science Research, IJSIS*, (2002), pp. 1-27.
- [25] N. Otsu, "A Thresholding Selection Method from Gray-Level Histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, (1979), pp. 62-66.
- [26] P. A. Bautista and Y. Yagi, "Improving the visualization and detection of tissue folds in whole slide images through color enhancement", *Journal of Pathology Informatics*, (2010), doi: 10.4103/2153-3539.73320, pp. 1-25.
- [27] C. F. Jiang, C. Y. Wang and C. P. Chiang, "Comparative study of protoporphyr- in IX fluorescence image enhancement methods to improve an optical imaging system for oral cancer detection", *Journal Of Biomedical Optics*, vol. 16, no. 7, (2011), pp. 3-8.
- [28] B. S. Morse, "Segmentation (Edge-Based Hough Transform)", *The Journal of Supercomputing*, Lecture 14 in Brigham Young University, (2008), pp. 1-5.
- [29] Q. Yu, P. Gong, N. Clinton, G. Biging, M. Kelly and D. Schirokauer, "Object-Based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery", *Photogrammetric Engineering & Remote Sensing*, vol. 72, (2006), pp.799-811.
- [30] M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill and C. Sun, "Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images", *Computerized Medical Imaging and Graphics*, vol. 22, no. 6, (1998), pp. 463-477.
- [31] M. Sathik, K. Senthamarai and J. V. Raj, "Hybrid JPEG Compression Using Edge-Based Segmentation", *An International Journal on Signal & Image Processing*, vol. 2, no. 1, (2011), doi: 10.5121/sipij.2011.2112, pp. 3-6.
- [32] D. Raman, "Learning to Parse Images Of articulated bodies", *Intl. Journal of computer vision*, Toyota Technological Institute at Chicago, (2006), pp. 1-8.
- [33] M. Lynch and O. Ghita and P. F. Whelan, "Automatic Segmentation of the Left Ventricle Cavity and Myocardium in MRI Data", *Journal on Computers in Biology and Medicine*, vol. 36, no. 4, (2006), pp. 389-407.
- [34] G. Sapiro, R. Kimmel, V. Caselles and C. Sbert, "Minimal Surfaces based Object Segmentation", *Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, (1997), doi: 10.1109/34.588023, pp. 394-398.
- [35] M. Haindl and P. Zid, "Multimodal Range Image Segmentation", *Vision Systems - Segmentation and Pattern Recognition*, (2006), doi:10.1109/ICPR.2006.838, pp. 1-23.
- [36] J. Suri, S. Singh and L. Reden, "Fusion of Region and Boundary/Surface-Based Computer Vision and Pattern Recognition Techniques for 2-D and 3-D MR Cerebral Cortical Segmentation (Part-II): A State-of-the-Art Review", *Pattern Analysis and Applications*, vol. 5, no. 1, (2002), pp. 77-98.
- [37] T. Blaschke, "Object-based-image-analysis-for-remote sensing", *Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, (2010), pp. 2-16.
- [38] J. C. Caicedo, F. A. Gonzalez and E. Romero, "A Semantic Content Based retrieval method for Histopathological images", *Histopathology Image Classification using Bag of Features and Kernel Functions*, vol. 4993, (2008), pp. 51-60.

- [39] Rutkowski, S. Peleg and R. Azriel, "Shape Segmentation Using Relaxation", *Pattern Analysis and Machine Intelligence*, vol. PAMI 3, no. 4, (1981), doi:10.1109/TPAMI.1981.4767123, pp. 368-375.
- [40] S. C. Agner, "A Computerized image analysis framework for dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) with application to Breast Cancer", A dissertation on Programs in Biomedical Engineering submitted to the Graduate School—New Brunswick Rutgers, University of New Jersey and School of Biomedical Sciences, vol. 35, (2001), pp. 43-51.
- [41] A. Basavanthally, "Automated Image-based Detection and grading of lymphocytic infiltration in breast cancer histopathology", A thesis submitted to the School—New Brunswick Rutgers and The Graduate School of Biomedical Sciences University of Medicine and Dentistry on Program in Biomedical Engineering under the direction of Dr. Anant Madabhushi and approved in New Brunswick, New Jersey, (2010), pp. 1-39.
- [42] S. Naqvi and J. M. Garibaldi, "An Investigation into the use of Fuzzy C-Means Clustering of Fourier Transform Infrared Microscopic Data for the Automation of Breast Cancer Grading", *Intelligent Modeling and Analysis (IMA) Research Group, School of Computer Science, University of Nottingham*, (2009), pp. 1-6.
- [43] B. Chen, M. Mete and S. Kockara, "Parameter-free multi-level fuzzy c-mean clustering for unsupervised detection in histological images", *Society for design and process Science, USA, Transformative Systems Conference*, (2010), pp. 1-5.
- [44] S. Handschuh, T. Schwaha and B. D. Metscher, "Showing their true colors: a practical approach to volume rendering from serial sections", *BMC Developmental Biology*, vol. 10, no. 41, (2010), pp. 4-16.
- [45] K. P. Kannan and A. Ananthakumari, "Texture Analysis of Histopathological Images to Identify Anomalous Region", *International Journal of Management, IT and Engineering*, vol. 2, no. 8, (2012), pp. 1-10.
- [46] M. M. R. Krishnan, C. Chakraborty and A. K. Ray, "Wavelet based texture classification of oral histopathological sections", *Microscopy: Science, Technology, Applications and Education*, (2010), pp. 897-906.
- [47] K. Kayser, J. Gortler, K. Metze, T. Goldmann, E. Vollmer, M. Mireskandari, Z. Kosjerina and G. Kayser, "How to measure image quality in tissue-based diagnosis (diagnostic surgical pathology)", *Diagnostic Pathology*, vol. 3, (2008), doi:10.1186/1746-1596-3-S1-S11, pp. 2-7.
- [48] S. Petushi, C. Katsinis, C. Coward, F. Garcia and A. Tozeren, "Automated Identification of Microstructures on histology slides", *Biomedical Imaging: Nano to Macro, 2004 IEEE International Symposium*, vol. 1, (2004), ISBN- 0-7803-8388-5, doi:10.1109/ISBI.2004.1398565, pp. 424-427.
- [49] A. E. Tutac, D. Racoceanu, W. K. Leow, J. R. Dalle, T. Putti, W. Xiong and V. Cretu, "Translational Approach for Semi-Automatic Breast Cancer Grading Using a Knowledge-Guided Semantic Indexing of Histopathology Images", *Conference on Proc. Microscopic Image Analysis with Application in Biology workshop (MIAAB)*, (2008), pp. 6-12.
- [50] W. Cai, J. Kim and D. D. Feng, "Content-based medical image retrieval", Chapter 4, (2011), pp. 2-16.
- [51] P. J. K. Baladevbhai and R. S. Anand, "Color Image Segmentation Methods: A survey and Comparison", *International Journal of Communication Engineering Applications*, vol. 04, no. 01, (2013), ISSN: 2230-8520, pp. 519-525.
- [52] Y. Li, H. Yin, M. Shi and G. Yue, "A Method of Color Image Segmentation Based on DPCNN", *2008 Cyberworlds, Intl. Conference Publication on Computer Applications Research*, (2008), doi:10.1109/CW.2008.142, pp. 725-729.
- [53] T. Drugman, M. Gurban and J. P. Thiran, "Feature Selection and Bimodal Integration for Audio-Visual Speech Recognition", *School Of Engineering-STI Signal Processing Institute, Lausanne, Switzerland*, (2007), pp. 1-86.
- [54] K. Kavukcuoglu, P. Sermanet, Y. L. Boureau, K. Gregor, M. Mathieu and Y. Le Cun, "Learning Convolutional Feature Hierarchies for Visual Recognition", *Journal of Machine Learning Research*, ISSN: 2493-2537, vol. 10, (2010), pp. 3-9.
- [55] A. A. Dingle, "Processing Electroencephalograms and Chaos in Neural Networks", Thesis, Engineering in Brain-Research, University of Canterbury, Electrical and Electronic Engineering, (1992), pp. 83-100.
- [56] Y. L. Boureau, "Learning Hierarchical Feature Extractors For Image Recognition", A dissertation submitted to the Dept. of Computer science, New York University, (2012), pp. 1-195.
- [57] A. Dong, B. Bhanu and Y. Li, "Evolutionary feature synthesis for image databases", *Seventh IEEE Workshops on Application of Computer Vision 2005*, vol. 1, (2005), doi:10.1109/ACVMOT.2005.50, ISBN: 0-7695-2271-8, pp. 330-335.
- [58] C. Farabet, C. Couprie, L. Najman and Y. Le Cun, "Scene Parsing with Multiscale Feature Learning, Purity Trees, and Optimal Covers", *29th International Conference on Machine Learning in Edinburgh, United Kingdom, Edinburgh, Scotland*, (2012), pp. 1-8.
- [59] F. T. Ramos, "Recognizing, Representing and Mapping Natural Features in Unstructured Environments", A thesis submitted in The Australian Centre for Field Robotics, The University of Sydney, (2007), pp.1-209.
- [60] G. Mensil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, P. Vincent, A. Courville and J. Bergstra, "Unsupervised and Transfer Learning Challenge: a

- Deep Learning Approach”, Workshop and Conference Proceedings 7 on Unsupervised and Transfer Learning, (2011), pp.1-15.
- [61] K. Veeramnikandan, R. Ezhilarasi and R. Brindha, “An FPGA-Based Real-Time Face Detection & Recognition System across Illumination”, International Journal of Emerging Science and Engineering, vol. 1, no. 5, (2013), pp. 66-68.
- [62] C. H. Tay, Thesis submitted on “Algorithms for Tissue Image Analysis using Multifractal Techniques”, in the Dept of Computer Science and Software Engineering in the University of Canterbury, (2012), pp. 1-106.
- [63] K. Shibuya, H. Hoshino, M. Chiyo, K. Yasufuku, T. Iizasa, Y. Saitoh, M. Baba, K. Hiroshima, H. Ohwada and T. Fujisawa, “Subepithelial vascular patterns in bronchial dysplasias using a high magnification bronchovideoscope”, Intl. Journal for health professionals in all aspects of Respiratory medicine, vol. 57, no. 10, (2002), doi:10.1136/thorax.57.10.902, pp. 902-907.
- [64] S. Naik, S. Doyle, M. Feldman, J. Tomaszewski and A. Madabhushi, “Gland Segmentation and Computerized Gleason Grading of Prostate Histology by Integrating Low-, High-level and Domain Specific Information”, MIAAB Workshop, (2007), doi:10.1.1.98.6547, pp.1-8.
- [65] I. Nakajima, Y. Yagi and J. R. Gilbertson, “Digital Pathology from the Past to the Future”, special issue dedicated to Intelligent Systems for Biomedical Applications”, Journal of eHealth Technology and Application, vol. 8, no. 2, (2010), pp. 1-115.
- [66] H. Chang, B. Parvin and L. A. Loss, “Nuclear Segmentation in H&E Sections via Multi-Reference Graph Cut”, International Symposium Biomedical Imaging, (2012), pp. 1-5.
- [67] G. M. Reynolds, A. C. Peet and T. N. Arvanitis, “Generating prior probabilities for classifiers of brain tumours using belief networks”, BMC Medical Informatics and Decision Making, vol. 7, no. 27, (2007), doi:10.1186/1472-6947-7-27, pp. 1-6.
- [68] J. A. Ozolek and C. A. Castro, “Teratomas Derived from Embryonic Stem Cells as Models for Embryonic Development, Disease, and Tumorigenesis”, Journal on Embryonic Stem Cells – Basic Biology to Bioengineering, (2011), pp. 231-262.
- [69] K. Kavukcuoglu, P. Sermanet, Y. L. Boureau, K. Gregor, M. Mathieu and Y. Le Cun, “Learning Convolutional Feature Hierarchies for Visual Recognition”, Journal of Machine Learning Research, (2011), pp. 1-9.
- [70] B. Krose and P. Van der Smart, “An Introduction to Neural Networks”, Robot Arm Dynamics, vol. 8, (1996), pp. 92-103.
- [71] P. P. Rezaeiye, M. Rayani, M. I. Ghiasi and M. R. Amini, “A New method for classifying corneal data”, Indian Journal of Science and Technology, vol. 6, no. 2, (2013), pp. 3996-3998.
- [72] L. M. F. Heath, L. O. Hall, B. Goldgof and F. R. Murtagh, “Automatic Segmentation of Non-enhancing Brain Tumors in Magnetic Resonance Images”, Artificial Intelligence in Medicine, vol. 21, no. 1-3, (2001), pp. 43-63.
- [73] K. S. Ni and T. Q. Nguyen, “An Adaptable k-Nearest Neighbors Algorithm for MMSE Image Interpolation”, Journal on Image Processing, vol. 18, no. 9, (2009), doi:10.1109/TIP.2009.2023706, pp. 1976-1987.
- [74] W. E. Henley and D. J. Hand, “A k Nearest Neighbour Classifier for Assessing Consumer Credit Risk”, Journal of The Royal Statistical Society (The Statistician), vol. 45, no. 1, (1996), pp.77-95.
- [75] G. Moonis, J. Liu, J. Udupa and D. B. Hackney, “Estimation of Tumor Volume with Fuzzy-Connectedness Segmentation of MR Images”, American Society of Neuroradiology, AJNR, vol. 23, (2002), pp. 356-360.
- [76] M. Kalas, “An Artificial Neural Network for Detection of Biological Early Brain Cancer”, International Journal of Computer Applications, vol. 1, no. 6, (2010), ISSN: 0975 – 8887, pp. 18-22.

Authors



Sangita Bhattacharjee

She passed Masters in Computer Applications in the year 2009 from B.P Poddar Institute of Management and Technology (WBUT). A Research Scholar under Prof. Samir Kumar Bandyopadhyay, Calcutta University working in field of medical image processing. She has worked upon the review on digital microscopy involving pre-processing steps to detection of disease.



Sanjay Nag

He is Research Scholar at University of Calcutta, India. He obtained Master in Computer Application in 2009 from IGNOU and Microsoft Certified Professional (MCP) in 2011. He is the author of more than 10 publications in National and International Journal and Conference including IEEE, Elsevier, etc. His fields of specialization are Biomedical Image Analysis, Image Processing, Data Communication and Network, Software Engineering, *etc.*



Jashojit Mukherjee

He passed M.Sc. in Computer and Information Science from University of Calcutta in 2013 after graduating from St. Xavier's College. He is currently working under Prof. Kashi Nath Dey in the field of Medical Image Processing. In the past he has worked in areas like face recognition and stereovision.



Indra Kanta Maitra

He is the Senior System Analyst of B. P. Poddar Institute of Management and Technology and Research Scholar of University of Calcutta. He obtained Master in Computer Application (MCA) in the year 2002 from IGNOU. He received award in 96th Indian Science Congress, 2009 at Shillong for Best Poster in the section of Information and Communication Science and Technology (including Computer Science). He is a Member of Computer Society of India, Indian Science Congress Association and IEEE. He was Managing Committee Member, Nomination Committee Member and Student Coordinator (East) of CSI. He is author of two books and more than 30 publications in National and International Journal and Conference including IEEE, Elsevier, etc. His working areas of specialization are Image Processing, Network Security, Data Structure, Programming Language, Computer Organization and Architecture, etc. He is now working in the field of Biomedical Image Analysis and CAD.



Prof. Samir K. Bandyopadhyay

He is Professor of Computer Science and Engineering, University of Calcutta, India. He obtained his Ph.D. in Computer Science and Engineering in 1989 from University of Calcutta, M.Tech in Radio-Physics and Electronics in 1979 from University of Calcutta and B.E. in Electronics and Tele-Communication in 1975 from B.E. College, University of Calcutta. He is the Chairman of Science and Engineering Research Support Society (SERSC, Indian Part), Fellow of Computer Society of India, Sectional President of ICT of Indian Science Congress

Association, 2008–2009, Senior Member of IEEE, Member of ACM, Fellow of Institution of Engineers (India), Fellow of Institution of Mammographic Density Estimation and Classification Electronics and Tele-Communication Engineering, India, Reviewer of International Journals IEEE Transactions on Neural Networks, ACM and Springer Publications. His fields of specialization are Bio-medical Engineering, Mobile Computing, Pattern Recognition, Graph Theory, Image Processing, Handwritten Signature Verification, Graphical Password Verification, etc. He has 25 years of Teaching and Research experience in the Post-graduate and under-graduate studies. He published books like Data structure Using C, Addison Wesley, 2003; C Language, Pearson Publication, 2010. He is author of more than 100 publications in National and International Journals and Conferences.