Chapter 14

Introductory Approach towards Analysis of Histopathological Biopsy Slide of Breast Cancer

14.1. Introduction

Still today, pathologists mount tissue slices on glass slides, treat them with appropriate stains and examine them through a microscope. Despite advances in staining techniques, it’s a process that has changed little over the last twenty years. Interpreting what they see is a time-consuming and requires a great deal of skill and experience which leads to a huge work load and expensive process. In practice it has been found that a large amount of time is wasted because majority of the samples are normal and do not required any further investigation. The huge work load always increases the possibility of false positive and true negative interpretation by the experts. The major obstacle in this manual processing is significant shortage of skill and experienced pathologist in all over the world especially the developing countries.

Image processing techniques can play an important role towards the interpretation of histopathological slides like breast biopsies. Several researches have been started worldwide for automatic interpretation of histopathological slide. Most of them are very
much successful to assist the expert to interpret the abnormalities. Due to the growth in computational speed and advanced image processing algorithm digital pathology is now a reality especially for identification of abnormalities.

The proposed method uses simple well established technique for isolation of insignificant portions of the slide images by colour polarisation as a pre-processing step. The simplicity of algorithm leads to less computational time and thus suitable tool to assist experts for automated real-time breast cancer diagnosis.

14.2. Histology of Breast Cancer

Image processing techniques can play an important role in interpreting images of breast biopsies. The breasts are made up of fat, connective tissue and gland tissue divided into lobes. A network of ducts spreads from the lobes towards the nipple. Younger women have more glandular tissue in their breasts, which makes them dense. Once a woman is past her menopause, the glandular tissue is gradually replaced by fat, which is less dense. A 'tail' of breast tissue leads into the armpit (axilla). The armpits have many lymph glands, also known as lymph nodes. They are part of the lymphatic system. The lymphatic system is made up of a network of lymph glands, connected throughout the body by tiny tubes called lymph vessels. Lymph glands are serious in cancer care because the lymph vessels nearest to lymph glands can carry cancer cells that have broken away from a tumour. A patient with initial stage of cancer will have no cancer cells in any of the lymph glands and they are most likely to survive.

Cancer can develop from normal cells that go through abnormal mutation that eventually transformation those to malignant cells which reproduce out of control. Many breast cancers grow from a sequence that initiates with an increase in the number of breast cells
(hyperplasia) to the beginning of unusual breast cells (atypical hyperplasia) followed by carcinoma in situ (noninvasive cancer) and finally, invasive cancer. Not all breast cancers necessarily follow the same pattern. The speed of advancement for those that do is highly inconstant. It is also evident that some cancers may never progress beyond in situ disease [278]. Before further progress to describe the proposed method, it is important to understand the types and characteristic features of the different categories of breast cancer that can be interpreted by the biopsies.

Figure 14.1. Breast cancer cell, photographed by a scanning electron microscope (Source: Website of Training Assistant, Continuous Quality Improvement Unit, Every Woman Counts, a Continuous Quality Improvement Project, State of California)

Lobular Carcinoma in Situ (LCIS) - Lobular cancer in situ (LCIS) means that there are cells change inside the breast lobes. This is not cancer. A patient having LCIS means that patients have an increased risk of getting breast cancer in the future.

Ductal Carcinoma in Situ (DCIS) - Ductal cancer in situ (DCIS) means that cells inside some of the ducts of breast have started to turn into cancer cells. These cells are all inside the ducts and have not started to spread into the surrounding breast tissue. So, there is very little chance that any of the cells have spread to the lymph nodes or elsewhere in the body.

Invasive Ductal Breast Cancer (IDBC) - Invasive Ductal Breast Cancer is the most common type of breast cancer. Between 70 and 80 out of every 100 breast cancers diagnosed are of
this type. A ductal carcinoma of the breast is a cancer that is growing in the cells that line the ducts of the breasts.

Invasive Lobular Breast Cancer (ILBC) - About 1 in 10 breast cancers diagnosed are of invasive lobular carcinoma. Invasive lobular cancer is most common in women between 45 and 55 years old. It is possible for men to get invasive lobular breast cancer but this is very rare.

Beside the aforesaid four types of breast cancer there are several other categories of invasive breast cancer that have been reported namely Tubular carcinoma, Medullary carcinoma, Mucinous carcinoma, Metaplastic carcinoma, Invasive cribriform carcinoma, Invasive papillary carcinoma and Invasive micropapillary carcinoma. These categories have
their own identifiable characteristic features and originated in specific part of the breast. Instead of invasive breast cancer, there are also some other types of cancers and are also developed in breast like Inflammatory breast cancer, Paget’s disease of the nipple, Phyllloides tumours etc.

14.2.1 Biopsy
Mammography is highly accurate but like most medical tests, it is not perfect. On average, mammography will detect about 80% - 90% of the breast cancers in women without symptoms. Testing is somewhat more accurate in postmenopausal than in premenopausal women [181]. The small percentage of breast cancers that are not identified by mammography may be missed out as mammography uses x-ray machines designed especially to image the breasts. Breast MRI also requires special equipment for breast imaging. Higher-quality images are produced by dedicated breast MRI equipment than by general purpose machines designed for head, chest or abdominal MRI scanning. However, many hospitals and imaging centers do not have dedicated breast MRI equipment available. It is important that screening MRIs are done at facilities that are capable of performing an MRI-guided breast biopsy at the time of the exam if abnormalities are found. Otherwise, the scan must be repeated at another facility at the time of the biopsy.

Initial mammographic or MRI images themselves are not usually enough to determine the existence of a benign or malignant disease with certainty. If a finding or spot seems suspicious, the radiologist may recommend further diagnostic studies. Sometime interpretations of mammograms can be difficult because a normal breast of different women may vary significantly. Also, the appearance of an image may be compromised if there is powder or salve on the breasts or if they have undergone breast surgery. Some
breast cancers are difficult to visualise. Not all cancers of the breast can be seen on mammography [220].

**Fine Needle Aspiration Cytology (FNAC)**

Fine Needle Aspiration Cytology (FNAC) is a method of collecting samples for cytological examination of a solid tissue or tumour with the help of a fine needle. This is an easy and simple method to collect the specimen. No anesthesia is required and no special preparation is required. It has high degree of sensitivity and specificity in the diagnosis of most of the lesions. It is highly used in breast cancer today [195]. Most of the studies have confirmed correct results in about 96% of the tests, which is very significant. It is only used in solid tumours of breast tissue.

**Core Biopsy**

A core biopsy is a procedure where a needle is passed through the skin to take a sample of tissue from a mass or lump. The tissue is then examined under a microscope for any abnormalities. Core biopsy may be performed when a suspicious lump is found, for example a breast lump or enlarged lymph node or if an abnormality is detected on an imaging test such as x-ray, ultrasound or mammography.

Core biopsy is a more invasive procedure than fine needle aspiration biopsy; however, it is quicker and less invasive than a surgical biopsy. In some cases, the result of a core biopsy will prevent the need for surgery to take place.

In radiologically positive cases of carcinoma breast, FNAC sensitivity is 61% and core biopsy sensitivity is 97%. In radiologically undiagnosed cases, FNAC sensitivity is 53% but core biopsy sensitivity is 96%. Automated core needle biopsy has superior diagnostic power than FNAC [195].
**Surgical Biopsy**

If the patient has a breast lump and wants it verified, a surgical biopsy is a good way to get a definite diagnosis. This type of breast biopsy removes the largest size of tissue sample, as compared to any type of needle biopsy. In some cases, the entire mass and a margin of healthy tissue may be removed. The tissue will be examined in a pathological lab right away to ensure that it is an accurate sample and to get a diagnosis. Surgical breast biopsy takes the largest tissue sample and has the highest accuracy rate of all biopsy methods.

A pathology lab can use two methods to study the tissue sample. The quickest method is called "frozen section" or cryosection. The tissue is rapidly frozen and sliced with a special blade into a section thin enough to see through. A permanent section method is a more thorough process, using special chemicals to get information from the tissue slide [2].

The slides are then put under the microscope for extraction of characteristics features to establish the abnormalities. The sample under microscope may be digitised using some specialised high resolution cameras for interpretation of sample using digital image processing methods. This will lead to higher accuracy and more efficiency.

**14.2.2 Digital Diagnosis of Histopathological Slides**

In all the cases human-error is one of most important factor, which cannot be eliminated completely, especially when it is detecting cancer, which can cause death of a patient. However, there are many more additional advantages of digital diagnosis which are stated below.

- Decrease the human-error under microscopic investigation of histopathological slide.
• Distribution and sharing of the digital images of histopathological slide to remote location is much more easy and time saving in respect of sending glass slide for expert’s opinion.

• Preservation of digital images of histopathological slide for future references is quite simple in contrary to the glass slides as they can break, deteriorate over time or can be lost.

• Further most significantly, a slide can be under the lens of only one microscope at a time, creating logistical problems for consultations, conferences and training sessions. A digitised slide image is portable for a second opinion and for other consultation purposes [207].

• The increased demand is not currently being matched by the supply of skilled and experienced pathologists, who remain in short supply in most parts of the world. Therefore, a real need for digital technology that improves the efficiency of pathology labs, reduce investigation time and helps pathologists with their diagnoses [171].

• In pathological investigation, chemical reagents are used which are highly toxic, costly and dangerous. The said process will be reduced by using digital image processing. So, it is not only effective and less costly but also eco-friendly too.

• Destruction and recycling of these highly toxic slides is also very costly and hazardous but in the said process it is absolutely not required.

• Reduce investigation cost drastically.

Moreover, it is not sublimating the traditional process but it is providing aid to it for further accuracy and improvement.
14.3. Literature Review

The histopathological slide image analysis of digital microscope is a young subject with enormous potentiality. There are several interesting works that have been carried out by different research groups. But most of the research is concentrated in the field of blood sample analysis and decision support system. In the field of tissue culture or more specifically cancer detection by biopsy slide is highly dependent on human intervention. Automatic CAD system development in this field has not initiated much. There are some researches but they are mostly theoretical in nature. In this chapter, few of the researches are discussed below which are dealing with this related field.

Cigdem Gunduz et al., 2004, reported a computational method that modelled a type of brain cancer using topological properties of cells in the tissue image. They constructed the graphs based on the locations of cells within the image. They used the Waxman model in their experiment [93].

C. Cagatay Bilgin et al., 2007, classified [28] the breast cancer tissues using graph theory. Image segmentation approach was used and Euclidean Distances were calculated between vertices. Considering the cell locations generated cell Graphs. These approaches toward automatic detection of cancer actually failed because the types of cancers identified were more complicated.

Lin Yang et al., 2007, introduced a Grid-enabled CAD to perform automatic analysis of imaged histopathology breast tissue specimens [149]. More than 100,000 digitised samples (1200 × 1200 pixels) were processed on the Grid. They analysed results for 3744 breast tissue samples, which were originated from four different institutions using diaminobenzidine (DAB) and hematoxylin staining. Both linear and nonlinear dimension
reduction techniques were compared, and the best one was applied to reduce the
dimensionality of the features. The results shown that the Gentle Boosting using an eight
node CART decision tree as the weak learner provided the best result for classification. The
algorithm has an accuracy of 86.02% using only 20% of the specimens as the training set.

Claus Bahlmann et al. presented a computationally efficient method for analysing H&E
stained digital pathology slides with the objective of discriminating diagnostically relevant
vs. irrelevant regions. Such technology is useful for several applications: (1) it can speed up
computer aided diagnosis (CAD) for histopathology based cancer detection and grading by
an order of magnitude through a triage-like pre-processing and pruning. (2) It can improve
the response time for an interactive digital pathology workstation (which is usually dealing
with several GByte digital pathology slides), e.g., through controlling adaptive compression
or prioritisation algorithms. (3) It can support the detection and grading workflow for
expert pathologists in a semi-automated diagnosis, hereby increasing throughput and
accuracy. At the core of the presented method is the statistical characterisation of tissue
components that are indicative for the pathologist’s decision about malignancy vs.
benignity, such as, nuclei, tubules, cytoplasm, etc. [12].

Philippe Belhomme proposed a computer aided diagnosis system dedicated to virtual
microscopy based on stereology sampling and diffusion maps. The original strategy is
presented, combining stereological sampling methods based on test grids and data
reduction methods based on diffusion maps, in order to build a knowledge image database
with no bias introduced by a subjective choice of exploration areas. The practical
application of the exposed methodology concerns virtual slides of breast tumours [22].
14.4. Proposed Method

The proposed method is an introductory pre-processing work towards the confirmation of breast cancer using histopathological slide of biopsy. Digital slide image analysis is a complex work and required long time to establish viable algorithm to detect the abnormalities in a slide without further human intervention. The proposed work is preliminary initiative towards the same. The proposed method does not identify the abnormalities but as a pre-processing step it enhances the histopathological slide, so that, if abnormalities present, it will become more noticeable. It will make easy the search for abnormalities by the experts of this field.

In the proposed work, free Tissue Blocks downloaded from OriGene Technologies [202] are used as dataset. In the experiments, breast cancer tissues from different patients and non-cancerous breast tissues from different normal females are considered. The sample images are 24-bit bitmap image with the size of 640X480 Pixels.

14.4.1 Grey Scale Conversion

The sample images provided by the OriGene technologies are coloured image. The coloured image is more informative but at the same time it will increase the complexity of the method towards analysis of the image. The 24-bit bitmap can produce almost 16,777,216 number distinct colours shades which are virtually impossible to handle. So, conventionally almost all the medical image processing algorithms use the grey shade image to reduce the complexity of the method and preserve maximum information within the image to investigate.

The grey image contains 256 numbers of grey shades to represent an image. The proposed method analysed different alternative methods to convert the colour image to grey scale
image. The most important factor here is to preserve the relative colour distance within the image colour domain. The Euclidean distance is the perfect method to covert the colour image to respective grey shade image. The relative colour distance will be perfectly preserved in the colour palette.

Initially a grey colour palette i.e. GP is generated. It is known that the value of red, green and blue of grey colour palette is same for a particular instance of intensity of colour. The algorithm iteratively read the colour pixel (P) from the sample image (I) and split the colour information in terms of Red, Green and blue.

\[ P = \sum_{i=1}^{h} \sum_{j=1}^{w} I_{i,j} \]  

(1)

Now the Euclidean distances (ED) are calculated for all the colour shade in the GP.

\[ ED_i = \sqrt{\sum_{n=1}^{255} (GP_i - P_R)^2 + (GP_i - P_G)^2 + (GP_i - P_B)^2} \]  

Where \( n=255 \)  

(2)

The grey shade with lowest distance \( (ED_i) \) is selected and value of the grey shade is propagated to red, green and blue component of the pixel, ensuing the conversion from colour to grey scale image. The iteration will terminate when all the pixels of the colour image are converted to grey scale.

**Algorithm: Grey Scale Conversion**

**GREY-SCALE-CONVERSION (I, Height, Width)**

- \( I \rightarrow \) Sample Image
- \( P \rightarrow \) Single Pixel
- **Loop** \( i \leftarrow 1 \) to Height
  - **Loop** \( j \leftarrow 1 \) to Width
    - **Do**
      - \( P.R \leftarrow I_{i,j}.R \)
      - \( P.G \leftarrow I_{i,j}.G \)
      - \( P.B \leftarrow I_{i,j}.B \)
      - \( ED \leftarrow 0 \)
      - \( MaxV \leftarrow 0 \)
      - \( MinV \leftarrow 255 \)
    - **Do**
      - \( P.R \leftarrow I_{i,j}.R \)
      - \( P.G \leftarrow I_{i,j}.G \)
      - \( P.B \leftarrow I_{i,j}.B \)
      - \( ED \leftarrow 0 \)
      - \( MaxV \leftarrow 0 \)
      - \( MinV \leftarrow 255 \)
// Comparing Euclidean distance

Loop k 0 to 255

Do
  If ED ≤ SQRT((k- P.R)^2 + (k- P.G)^2 + (k- P.B)^2)
    Then R ← k
    \(I_{i,j}.R \leftarrow R\)
    \(I_{i,j}.G \leftarrow R\)
    \(I_{i,j}.B \leftarrow R\)

// Calculating the minimum and maximum intensity value

If k ≤ MinV
  Then MinV ← R
If k ≥ MaxV
  Then MaxV ← R

Return I

Complexity Analysis of Algorithm

The proposed method is implemented using three loops. The outer loop is for height, inner
loop is for width and the inner most loop is to traverse the grey scale palette. Assuming
that height = width = n and grey palette is k. So, the running time of the algorithm is k.n^2.
But the grey palette i.e. k is always 256, so, it is a constant. Hence, the running time of the
algorithm is order of n^2.

14.4.2 Colour Polarisation

The initial objective of the process is to eliminate the irrelevant object from the slide to
make it more enhanced and clear. The final objective is to merge the similar objects in slide
by colour polarisation technique. The initial objective will be achieved by increasing the
contrast level of the slide by using a constant (Δ) with the intensity shades. The final
objective will be accomplished by colour polarisation using a threshold (k) determined by
the central tendency of image. For simplicity, the average of highest (MaxV) and lowest
(MinV) intensity of colour can be used for thresholding which are already derived in
previous algorithm.
Finally, the inverse image is generated due to better understanding of the abnormalities present in the sample image.

\[ p_{i,j} = 255 - p_{i,j} \]  \hspace{1cm} (4)

**Algorithm: Colour Polarisation**

**COLOUR-POLARISATION** \( (I, \text{Height, Width, MaxV, MinV}) \)

1. \( I \leftarrow \text{Sample Image} \)
2. \( P \leftarrow \text{Single Pixel} \)
3. \( k \leftarrow (\text{MaxV} + \text{MinV}) / 2 \)

**Loop** \( i \leftarrow 1 \) to \( \text{Height} \)

**Do**  **Loop** \( j \leftarrow 1 \) to \( \text{Width} \)

**Do** \( \text{Val} \leftarrow i_{i,j} \)

// Enhancing the contrast

**If** \( \text{Val} \geq 128 \)

**Then** \( \text{Val} \leftarrow \text{Val} \times \Delta \)

**If** \( \text{Val} > 255 \)

**Then** \( \text{Val} \leftarrow 255 \)

**Else** \( \text{Val} \leftarrow \text{Val} / \Delta \)

**If** \( \text{Val} < 0 \)

**Then** \( \text{Val} \leftarrow 0 \)

// Merging the similar pixels by colour polarisation

**If** \( \text{Val} \geq k \)

**Then** \( \text{Val} \leftarrow 255 - \text{MaxV} \)

**Else** \( \text{Val} \leftarrow 255 - \text{MinV} \)

\( i_{i,j} \leftarrow \text{Val} \)

**Return** \( I \)

**Complexity Analysis of Algorithm**

The proposed method traverses the entire image. The outer loop is used for the height whereas the inner loop is for width. Assuming that, height = width = \( n \). Then the running time of the algorithm is order of \( n^2 \).
14.5. Experimental Result

The primary objective of the proposed method is to remove the huge amount of fat, connective tissue and gland tissue from the cancerous cells within the histopathological biopsy samples. The stage, intensity, type, future development and treatment of cancer can only be detected on the basis of orientation of malignant cell, shape of the cell and duct, density carcinogenic cells in compare with normal cells. The outputs of aforesaid algorithms are depicted in the following figures for cancerous cells within the biopsy slide and normal slide along with the histogram of the images.

![Figure 14.3. The Original Histopathological Slide showing Malignant Cells along with Histogram](image1)

![Figure 14.4. The Grey Shade Histopathological Slide showing Malignant Cells along with Histogram](image2)
Figure 14.5. Finally the Inverse Colour Polarised Histopathological Slide showing Enhanced Malignant Portion along with Histogram

Figure 14.6. The Original Histopathological Slide showing Normal Cells along with Histogram

Figure 14.7. The Grey Shade Histopathological Slide showing Normal Cells along with Histogram
Figure 14.3, 14.4 and 14.5 show the Cancerous tissue. In final Figure 14.5 all the fat, connective tissue and gland tissue are supressed. It is only showing the parts, which are important to determine the carcinoma. It is depicting the abnormal orientation of malignant cell, shape of the cell and duct. In contrast, Figure 14.6, 14.7 and 14.8, show the normal slide and all the fat, connective tissue, and gland tissues are supressed in Figure 14.8. The orientation, shapes etc. in this figure are normal in nature and showing no sign of carcinoma or abnormalities. The image histograms are showing the colour clustering on the basis of algorithms. The primary objective is to supress less or insignificant parts from the considerable portion of the slide image which is done by the contrast enhancement. For prominence and clustering the similar tissue cells colour polarisation is done. The polarisation of colour is highly proved by the aforesaid histograms provided.

14.6. Conclusion
Most of the CAD systems of breast are concentrated in detecting the abnormalities in the screening technologies. But it is also important to analyse the confirmation part of the detection process; to make it a true diagnosis system. The histopathological slide image analysis or digital microscopy has an immense potentiality for future. Several outstanding works have been done regarding blood sample analysis but in the field of biopsy slide image
analysis and similar assisting systems are rare. In this dissertation, it has been tried to amalgamate screening technology along with histopathological slide image analysis to make it a complete diagnosis tool to assist the experts. The histopathological slide analysis is described in this chapter is a simple and pre-processing step by applying well established methods but huge works are left for future research. The initial objective of the proposed method is met and results obtained are encouraging for further development.