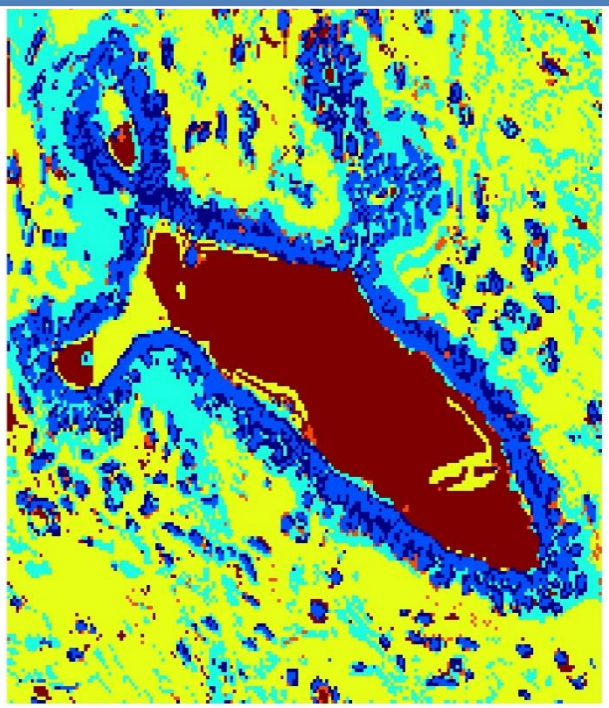
INTUITIONISTIC FUZZY BASED BREAST HISTOLOGY IMAGE ANALYSIS AND CLASSIFICATION



Dr. Amoli D. Belsare

Dr. Milind M. Mushrif

Dr. Meena A. Pangarkar

Publication Partner: IJSRP INC.

11/17/2020

Intuitionistic Fuzzy based Breast Histology Image Analysis and Classification

Dr. Amoli D. Belsare
Dr. Milind M. Mushrif
Dr. Meena A. Pangarkar

Publishing Partner:
IJSRP Inc.
www.ijsrp.org

ISSN 2250-3153
9 772250 315302

Publication Partner:

International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Preface

In this digital era, everyone is acquainted with the use of digital devices and many people use them every day. Use of digitized data has risen exponentially in the healthcare sector too & is used extensively in diagnostics, treatment & post treatment follow-ups. But there is still a vast scope for further advancements and refinements. Keeping this in mind, the present work focuses on digitizing pathological examination of a biopsy sample from a patient suspected to have Breast cancer. The study aims to be helpful in reducing the workload of the pathologist as well as help in decision making by reducing inter- and intra- observer variability.

Breast cancer is second largest cancer in India which results in a lot of mortality and morbidity in the young and productive age group women. It is a treatable and curable cancer if diagnosed and treated early, resulting in saving a lot of lives.

Thus, digitization of the microscopic images of biopsy samples was done so as to provide a helping hand to pathologist to save time and subjectivity of diagnosis. The multiple microscopic images were analyzed by use of image processing techniques and results were generated as positive or negative for Cancer. This tool can be used as a second opinion by pathologists/oncologists.

Briefly, a spatio-texture based image segmentation algorithm is developed for identifying breast duct accurately. This work is further extended by using intuitionistic fuzzy based approach to identify breast duct from digitized biopsy sample with more segmentation accuracy. Due to different staining concentrations at different labs, tissue image may have color variation and hence normalization is needed while preprocessing the image. Intuitionistic fuzzy considers the uncertainty of a pixel information and the normalization step in processing histology image can be avoided. Once the duct from breast tissue images is identified, basic binary classification of images as malignant and non-malignant has been performed & the results are compared with the state of art approaches from image processing. This work is very special as it involves the guidance from medical professionals in development of algorithms as well as development of image database for the pilot study along with ground truth data for evaluation.

I would like to thank all who directly or indirectly helped me in the development of the methods with guidance and support. I also extend my heartfelt thanks to my guide Dr. M. M. Mushrif & co guide from GMCH Nagpur Dr. Mrs. M. A. Pangarkar who constantly supported me in whole process of design, development of concept to publishing this book. At the end, I would like to thank my family members, Mother, Father, Father-in-law, Late Mother in law, my husband and my son who encouraged me constantly to work hard and supported me in everything without which this writing can’t be completed.

Publication Partner:

International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Copyright and Trademarks

All the mentioned authors are the owner of this Monograph and own all copyrights of the Work. IJSRP acts as publishing partner and authors will remain owner of the content.

Copyright©2020, All Rights Reserved

No part of this Monograph may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as described below, without the permission in writing of the Authors & publisher.

Copying of content is not permitted except for personal and internal use, to the extent permitted by national copyright law, or under the terms of a license issued by the national Reproduction Rights Organization.

Trademarks used in this monograph are the property of respective owner and either IJSRP or authors do not endorse any of the trademarks used.

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Authors

Dr. Amoli D. Belsare

Assistant Professor,
Department of Electronics & Telecommunication Engineering,
Yeshwantrao Chavan College of Engineering,
Nagpur, India

Dr. Milind M. Mushrif

Professor,
Department of Electronics & Telecommunication Engineering,
Yeshwantrao Chavan College of Engineering,
Nagpur, India

Dr. Meena A. Pangarkar

HoD,
Department of Laboratory Medicine,
National Cancer Institute,
Nagpur, India

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Table of Content

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Abstract

Researchers from the medical image analysis domain apply microscopy-based computer aided image analysis as a tool for objective, precise and quantitative disease diagnosis. In the similar context, main goal of this dissertation is the development of techniques for object level computer aided image analysis of breast histology images. In microscopic image analysis research, histology image examination is considered as gold standard. Due to complex tissue appearance in histology images, its automated analysis is challenging and need objective decision support to pathologist so as to minimize inter-intra observer variability in decision making. In this research, we propose robust, accurate breast histology image segmentation and classification techniques using inputs from expert pathologists in development process. The main contributions of the research work are fivefold. Initially, a spatio-colour similarity based super pixel generation method to capture spatial arrangement and color of duct epithelial lining lumen or solid sheets without lumen in breast histology images is proposed. Secondly, we propose to integrate this super pixel with texton: clustered vector filter responses to formulate similarity measure for Normalized Cuts (NCUT) image segmentation method. Thirdly, we extend proposed super pixel and texton generation technique to Atanassov's Intuitionistic Fuzzy Sets (AIFS) domain by inculcating hesitancy and non-membership knowledge of breast tissue appearance irrespective of staining concentration and different fixation procedure while preparing sample specimen from breast gross tissue. Using this fuzzy representation, AIFS based similarity measure is proposed to find more accurate breast duct segmentation using NCUT. Further, a new AIFS clustering based fuzzy texture map is generated to reduce computations required for AIFS texton image representation as fourth contribution. This AIFS texture map results were compared with map generated using the state of art fuzzy C-means (FCM) clustering approach and the two texture maps are integrated with AIFS super pixel separately for generating single color-texture based AIFS breast image for each integrated map. Further NCUT is performed on these resultant AIFS affinity matrices separately for segmenting duct epithelial regions in breast images and quantitative evaluation of each proposed integrated approach is performed. Finally, a texture and morphological feature based linear classifier is proposed to classify breast histology images as non-malignant and malignant images. Ground truth image database of both nonmalignant and malignant breast tissue images for evaluation of proposed algorithms was developed by expert pathologist. The experimentations and evaluations carried on these image databases demonstrate the superiority of proposed segmentation and classification algorithms.

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

1. INTRODUCTION

Automatic analysis of pathology/microscopic images is an emerging research area in recent decade. Microscopic images of diseased tissues are the foundation of diagnosis in Pathology. The information and knowledge conveyed by these images is expressed in well-defined linguistic terms and concepts. Advances in computing technology and digital image processing applications have made it possible to express these time-tested diagnostic clues in numeric terms. This digitized data creates a permanent, reproducible record, which can be used for consultation as well as teaching. Computer Aided Diagnosis (CAD) of such digitized data also provides a basis for quantitative analysis of images, which is a powerful tool in the hands of researchers and medical professionals. Preliminarily, pathology deals with examination of cells/tissues under microscope in order to study the manifestations of diseases [1-4]. It refers to the examination of invasive (histopathology) or less invasive biopsy (cytology) or surgical specimen by a pathologist, after the specimen has been processed and sample sections have been placed onto glass slides. The goal of applying image processing techniques in the histopathology field ranges from cell counting, cell type identification, classification for deriving quantitative measurements of diseased features in histopathogical images to tissue segmentation and automatic determination of presence of disease within analyzed samples [1].

Carcinoma of breast is the largest occurring disease amongst women world-wide. Because of huge burden of breast cancer cases in India, it is very important to find the strategies for its early detection so as to prevent morbidity and mortality in patients [2]. At present detection of cancer from biopsy images is carried out by expert pathologist with the examination of a diseased tissue sample under the microscope. Interpreting these pathology images requires expert knowledge, experience and observational skills. With proper training, and experience a medical professional or pathologist identifies the structural features of specific cell types, tissues and organs from microscopic images, and understand structure-function relationships at different levels. This manual process of tissue examination is error prone due to inter-intra observer variations and need human expertise for accurate diagnosis [5]. Hence, to perform automatic analysis of breast tissue structure from digitized image, we mainly focused on the development of algorithms for segmentation of duct epithelial nuclear lining lumen or without lumen using image processing techniques. This automatic image analysis provides objective diagnosis of histology images, so as to assist pathologist in decision making process. The algorithm development introduces integration of color, texture and location information of objects present in breast biopsy images in order to segment breast ducts automatically without human intervention and perform image classification for detection of breast carcinoma (malignant) images.

In this research work, duct spatial information, nuclei-lumen color and texture features of breast duct is included so as to get improved results using in graph partitioning algorithm. This integrated spatio-color-texture breast duct segmentation is further extended in Atanassov's Intuioninstic fuzzy set (AIFS) theoretic domain so as to increase segmentation accuracy along with non-membership and hesitancy values of pixels in an image. After automatic detection of breast duct using above two methods, histology breast tissue image classification for detection of malignancy using texture features is performed. Automated segmentation and classification algorithms provide robust, qualitative and objective analysis of breast histology images thereby reducing inter-intra observer variability and subjective diagnosis of breast carcinoma by pathologist. Thus, CAD for malignancy detection can be considered as assistant tool in the hands of pathologist to improve disease diagnosis and further management.

2. COMPUTER AIDED MICROSCOPIC IMAGE ANALYSIS

In literature, many techniques have been applied for histology image classification and gradation, gland & nuclei counting, cancer cell identification, so as to achieve quantifiable measurements of diseased samples through digitized histological images. Most of the histological studies were primarily concerned on distinguishing between the normal/abnormal cases and cancer grading applications for different organs like prostate [6, 7], breast [5-7, 11-13, 43], colon cancer [10, 14, 15], oral mucosa [9, 16] and Follicular lymphoma [17]. All these studies focus mainly on image segmentation, cancer feature selection and classification approaches for analyzing histopathology images of different organs with different tissue structure and morphology.

Automatic breast histopathology image segmentation involves detection of various histological structures like cancer cell nuclei, glands or ducts, lumen, tubule, nuclear pleomorphism and mitosis detection [7, 8, and 11] based on color, texture and hybrid segmentation algorithms. In literature, algorithms applied to solve these problems include mean shift algorithm, region growing, K-means clustering, expectation-maximization, graph partitioning as normalized cut, spectral clustering etc. Another area of histology image analysis research worked on breast cancer type classification followed on grade identification based on spectral clustering approach [6], multiple instance learning [19] on low optical magnification histology images. Studies in [5, 13] provide a brief review on methods applied for analysis of breast cancer histopathology images for carcinoma detection. Based on the literature studied, automated breast histology image analysis problem could be divided as boundary based, region-based segmentation for detection of individual Bloom Richardson's Gradation criteria [4] and classification of whole slide images for malignancy gradation.

Jun Xu et al. [18] applied segmentation technique based on breast duct boundaries using Geodesic Active Contours (GAC) and weighted mean shift normalized cut for detection of lumen area in beast images and the results are compared with Chan Vese model. The segmentation approach quantitatively compares the results with performance parameters like overlap, sensitivity, specificity. They have also applied the methodology for segmentation of prostate gland structure. The results in research fails for detecting weak boundaries as the work is based on edge-detection criterion.

Segmentation on the basis of region is mainly performed for detection of tumorous regions from low magnification images and this peculiar problem has been investigated by Tousn et al. [15]. They incorporated the background knowledge of colon tissue organization in relation with spatial aspect of cytological tissue components using graph run length matrix features and region growing approach for image segmentation and achieved more than90% segmentation accuracy compared to state of art JSEG approach.

Cancer grade depends on tubule and cancer nuclei count in biopsy tissue. This is one of the gold standard criteria in detection & grading of breast cancer [4]. Ajoy Basavanhally in [11], performed tubule and cancer nuclei counting using O'Callaghan neighborhood and Color Gradient based Active Contour model (CGAC). The proposed work also applied color deconvolution scheme with spatial proximity of low-level structures like lumen, cytoplasm and nuclei for correct identification of the objects of interest. Use of graph-based image features to detect lumen area achieved better segmentation results with 89% accuracy compared with manual grading of tubular density.

Naik et al. [7] investigated a level set based method initialized by Bayesian classifier for prostate and breast cancer gland segmentation in histology images. They applied individual pixel values and relationship amongst pixels to detect the objects such as nuclei and gland structure as low- and high-level features. They also incorporated the structural constraints of these objects for classification of prostate and breast cancer grading.

Numbers of computer vision algorithms have been applied successfully for automatic cancer detection and classification for high-resolution biopsy images, but analyzing large amount of tissues at high resolution is computationally too expensive. In case of automated image classification, literature shows that most of the work has been carried either on the nuclei extracted or on region of interest detected by expert pathologist for detection of

malignancy. Methods proposed in literature covers textural, morphological, and architectural features extracted from the cell nuclei or region of interest in prostate/colon/oral Mucosa histology images [6, 7, 9, 10, 14, 15, and 16]. In recent work, classification of histology images includes statistical analysis based on features extracted with machine learning algorithms to improve the results of large data. In similar work, Scot Doyel et al. [6] applied textural and nuclear architectural features of breast histology duct for analysis of cancer. Here, the spectral clustering method reduces the feature set so as to get more accurate results.

Most of the breast image classification-based research work applies textural and morphological features extracted from region of interest for malignancy grade detection [12, 19-21]. C. Loukas et. al. [19] carried work on classification of cropped region of interest from large size image using texture analysis at low magnification of 10X. The classification accuracies of 86%, 85%, and 90% for three grades of breast cancers respectively achieved by using support vector machine approach.

Zhang proposed classification of breast histology images using cropped region of interest only. They have used a kernel PCA model based one-class ensemble classifier [21] on extracted texture features using Curvelet Transform, Gray Level Co-occurrence Matrix (GLCM) and Completed Local Binary Patterns (CLBP) for breast histology images reporting 92.06% accuracy for classification of healthy tissue, tumor insitu and invasive carcinoma of breast histology image.

The concept of AIFS has been applied earlier for detecting edges so as to detect abnormalities in microscopic images by Chira [36, 37]. Hanquiang et. al. [38] and Sen et. al. [39] incorporated AIFS set theoretic approach in graph partitioning method for computer vision algorithms. But segmentation of breast histology images for malignant duct detection using AIFS is firstly introduced by us in this research. All the research discussed in literature focuses on the application of computer vision algorithms to solve any one histology image segmentation aspect and/ image analysis problem. Although the studies in literature demonstrate the promising results for tissues in which the glandular structure is present, the detection of nuclear region lining lumen or solid sheets without lumen in breast histology tissue is not much focused. Due to variation of structural appearance in tissue, staining and fixation inconsistency, same algorithm would fail to detect all required object of interest from biopsy images. Hence objective and quantitative automated analysis for detecting breast duct structure is required so as to identify small variations in the structure which may lead to carcinoma of breast. Considering these issues, focus of this research work is twofold: first, development of algorithms for image segmentation using integrated spatio-color-texture approach in graph partition and second, application of these developed algorithms on microscopic breast histology images for objective computer aided image analysis. The key contributions of the proposed work can be outlined as follows:

1. A modified super pixel generation algorithm using weighted distance based spatio-color similarity for boundary adhere breast duct segmentation is proposed initially. An algorithm for multi-scale, multi-oriented filtered image clustering based texton generation is also introduced so as to capture texture information of breast duct [42].

2. Novel integration of these two generated cues is introduced in the research pipeline to formulate affinity matrix. The affinity matrix is required for performing normalized cuts (NCUT) graph partitioning based breast histology image segmentation. Here, we focused on the use of color-texture and spatial alignment of the duct present in breast tissue image for generation of image graph. This detects breast duct automatically without human intervention. This algorithm also reduces the memory requirement for calculating similarity matrix for NCUT, as a group of similar pixels is used as vertex of graph instead of using single pixel as vertex [42].

3. Next we propose AIFS similarity measure for super pixel generation and AIFS texton generation for crude texture-based segmentation of breast ducts. Additionally, AIFS breast image analysis pipeline consists of development of two new fuzzy clustering-based texture maps by considering a multi feature textured image representation. Thus, three AIFS color-texture affinity matrices are formulated for NCUT image segmentation [44].

4. We also developed classification algorithm for breast histology images as non-malignant and malignant classes. Segmented duct image datasets from above two approaches have been used for experimentation. The process of automatic classification of breast duct images includes feature extraction, selection and then classifier is trained and tested on data. Feature extraction includes texture and morphological features of breast duct detected from histology and histopathology images. Also, a texture model-based biopsy image classification has been performed to detect malignancy automatically [45].

5. Another contribution in the proposed research is the incorporation of AIFS concept in super pixel generation, texton generation, new fuzzy texture measure and fuzzy affinity matrix calculation for normalized cut, which provides better segmentation of images. The AIFS concept has been applied firstly for automatic analysis of breast histology images as well as used in graph segmentation approach [44].

6. We have also developed real time breast histology image database and corresponding Ground Truth (GT) images for quantitative evaluation of the proposed algorithms. The GT database is created with the help of expert pathologist. All segmentation and classification results are also evaluated qualitatively by expert pathologist [42, 44].

3. BREAST HISTOLOGY IMAGE AND GROUND TRUTH DATABASE GENERATION

Breast histology images required for the research, are prepared and acquired at Department of Pathology, Govt. Medical College and Hospital, Nagpur, India with clearance from ethical committee. Motic B1 series system microscope with 10X, 40X optical magnification and 2.5µm/pixel calibration obtained from 10µm thickness of gross tissue is used for acquiring image dataset. The Hematoxylin-Eosin (H&E) stained breast tissue images are captured at resolution of 768 × 768. The digitized breast tissue image contains extraneous areas of fibro vascular stroma in addition to epithelial ducts which are region of interest (ROI) for this research. To ensure uniformity in algorithm development and evaluation, these images are resized to 200 × 200 in consultation with pathologist. Breast histology samples are processed and fixed onto a glass slide for tissue examination under light microscope and digital images are obtained.

There is scarcity of a public dataset of normal and diseased breast histopathology images in literature and also very few evaluation mechanisms for quantification of segmentation output. Hence, we contribute to the generation of breast histology image datasets and their ground truth for quantitative evaluation of proposed algorithms in consultation with expert pathologist. Following procedure is used to establish GT breast histology image database [42, 44]:

1.Non-malignant and malignant duct regions in digitized histology images were marked by the two experienced pathologists using lasso tool of Adobe Photoshop (CS5x32) and set of GT binary images are obtained.

2.Further, the regions on which both pathologists agreed were only taken as GT duct image region for both malignant and non-malignant breast histology images.

3.Generated GT breast histology image database is further presented to another panel of two expert pathologists to make final decision. This panel has given only visual suggestions on above generated GT images and refined GT image database is then used for quantitative evaluation of algorithm results. The original breast histology images and corresponding GT images are shown in Fig. 3.1.

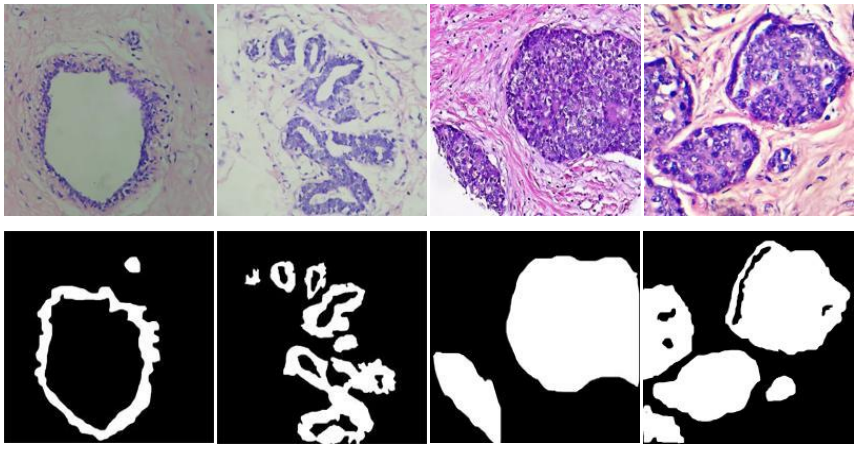


Fig. 3.1. 1st row: Original and 2nd row: GT, (a-b), (e-f) non-malignant, (c-d), (g-h) malignant Breast Histology Images

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Breast histology image and corresponding GT database consists of total 100 digitized images. In all 60 Non-Malignant (NM) and 40 Malignant (M) cases which includes about 200NM duct regions and 60 M duct regions are considered for research work. NM breast images consist of normal as well as benign duct regions. To ensure uniformity in images for research and to concentrate on ROI, only 40 NM and 30 M images are considered during entire research work in consultation with expert panel of pathologist.

4. BREAST DUCT SEGMENTATION USING SPATIO-COLOUR-TEXTURE GRAPH CLUSTERING

This section focuses on segmentation of non-malignant breast duct as well as malignant duct in digitized histology images using integration of super pixel, texton image maps to form image graph representation for NCUT approach. For super pixel image representation, we proposed weighted spatio-color distance-based similarity function. Secondly a texture model: texton is introduced by employing clustering of vectored rotation invariant filter responses. Next, we obtain an integrated spatio-color-texture mapped image using two cues derived previously by region wise intersection. A weighted distance similarity measure is then formulated and applied to graph representation for Normalized Cut (NCUT): a graph partitioning method in computer vision to obtain final segmentation result. In NCUT [22], firstly an image graph representation is obtained followed by image partition based on the dissimilarity between two parts of graph A and B; the edge is removed by iterative process to segment the image. This is termed as cut in graph theoretic language and can be calculated by using $cut(A,B) = \sum_{u \in A, v \in B} w(u,v)$. Finally, the graph is partitioned optimally by minimizing this cut value. The normalized cut criterion applied is given as:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{4.1}$$

Where, $assoc(A,V) = \sum_{u \in A, t \in V} w(u,t)$ represents the total connection of vertex A to all vertices in the graph, similarly $assoc(B,V)$ can also be defined. This criterion is used as a cut cost of total edge connections to all the nodes in the graph. The proposed segmentation algorithm is as shown in Fig. 4.1. NCUT Segmented breast duct output evaluated quantitatively using the parameters like Rand Index (RI), Global Consistency Error (GCE), Variation of Information (VI), sensitivity, specificity and accuracy [40, 41] along with qualitative evaluation by expert pathologist.

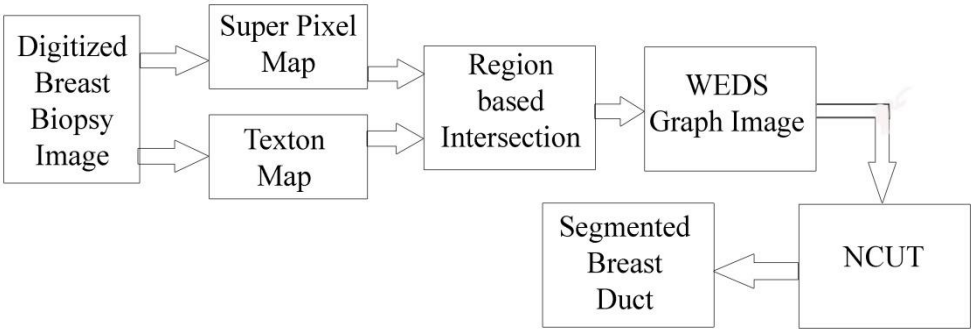


Fig. 4.1. Proposed Algorithm flow diagram

A. Super pixel Image Representation

Super pixel representation of image is a group of similar pixels to form perceptually meaningful boundary adhere regions. In literature, geometric flow [23], Turbopixel [24], SLIC [25], normalize ducts [22] and Pixel Intensity and Location Similarity (PILS) [26] methods have been applied for generating super pixel image map. In super pixels, a group of pixels acts as a single compact element which decreases complexity of images with thousands of pixels to a few hundreds of super pixels. A Simple Linear Iterative Clustering (SLIC) method by Achanta et al. [25], is most commonly used method to generate more compact super pixels for input image using spatio-color proximity-based distance.

We generated super pixels by adapting SLIC procedure. A color image I, of size $m \times n$, consisting of three primary components, red (R), green (G), and blue (B) is divided initially into equal sized super K_s pixels. Initial super pixel of size, $S = \sqrt{N/K_S}$, is considered for further analysis where $N = m \times n$. From this initial image grid, six-dimensional feature vector, $C_i = [R_i, G_i, B_i, x_i, y_i, l_i]^T$, where, $i = 1, 2, ..., K_s$ consisting of color, spatial location and label of each super pixel centroid is generated. Final super pixel image representation is obtained using iterative clustering approach, where spatial and color Euclidean distance of a pixel and corresponding cluster center C_k is calculated as: $d_C = \sqrt{(R - R_i)^2 + (G - G_i)^2 + (B - B_i)^2}$ and $d_S = \sqrt{(x - x_i)^2 + (y - y_i)^2}$. In our algorithm, arrangement of breast duct structure in tissue image depending on spatial proximity of lumen surrounded by nuclei, a single Weighted Spatio-Color Euclidean Distance (WSCED) is obtained without using any user defined constants:

$$WSCED = 0.7 * ds + 0.3 * dc \quad (4.1)$$

With extensive experimentation and importance of spatial proximity in breast histology images, we assign fixed 70% and 30% weight age to spatial and color distance respectively. Finally, by applying standard Gaussian function to image map, similarity measure is formulated as:

$$S(i, j) = \exp\left(-0.5 \left(\frac{WSCED(i, j)^2}{\sigma^2} \right) \right) \quad (4.2)$$

With rigorous experimentation σ is set at 10 and label is updated by checking neighbor boundary pixel label of each super pixel. In label assignment step, image pixel I associated with nearest cluster center in adjacent region of 2S × 2S size is searched to reduce complexity in calculations of all distances for compact super pixel generation. This super pixel segmentation of the image preserves boundary of the duct epithelial nuclei and luminal region in breast tissue and can be observed in Fig.4.2 (a).

B. Texton generation

Julesz introduced Textons in 1981 as a unit of human texture perception [27]. Initially oriented line segments on gray images are used to obtain textons, later Malik et. al. [22, 28] introduced the oriented, scaled filter response clustering to extract texton representation of image, showing texture measure of the

image. In our algorithm, we use total 67 filter kernels consisting of eight orientation, four scale Gabor filter (32 even and 32 odd), and three scale LoG filter kernels of size 11×11 for finding texton representation of the breast histology image texture. Initially a 2-D multi-scale, multi-orientation Gabor filter kernel is generated using the mother Gabor wavelet $g(x,y)$ as follows:

$$g(x,y)=\left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{{\sigma_x}^2}+\frac{y^2}{{\sigma_y}^2}\right)+2j\pi Wx\right] \tag{4.3}$$

Where σ_X and σ_y are the space constants of the Gaussian envelope along the x-axis and y-axis, and is the center frequency for mother Gabor $g(x,y)$. The children Gabor for the mother Gabor is calculated as:

$$g_{sn}(x,y)=a^{-2s}g(x',y'),\quad a>1 \tag{4.4}$$

Where, $x'=a^{-s}\left(x\cos\theta+y\sin\theta\right);\ y'=a^{-s}\left(-x\sin\theta+y\cos\theta\right)$, $\theta=n\pi/N$, $s=1,2,....,K$ and $n=1,2,....,L$ here, x and y are initial coordinates x' and y' are coordinates after rotation. a is a fixed scale factor, s is the scale parameter, θ is orientation, n is the orientation parameter, K is the total number of scales, and L is the total number of orientations used for Gabor filter. We set Gabor function parameters as follows: $W=1$, $a=2$, $\sigma_X=0.25$, $\sigma_y=3\sigma_X$, $K=4$, and $L=8$. Each scale s is selected such that the distance between each is twice. Laplacian of a Gaussian (LoG) filter kernel is also used in filter bank to detect blurry edges as well as fine details of duct epithelial nuclear region and is calculated as:

$$\nabla^2G(x,y)=\left[\frac{x^2+y^2-2\sigma^2}{2\pi\sigma^4}\right]e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4.5}$$

Where, σ is space constant of Gaussian envelope and set as 3 for all images. Further, the image $I(x,y)$ is convolved with filter bank generated to get 67 filtered outputs.

$$G_f(x,y)=I(x,y)*g_f(x,y);\quad f=1,2,...,67 \tag{4.6}$$

Next, all filter outputs are considered as a 67-dimensional feature vector for representing each pixel in image and are clustered into 25 responses for single pixel using k-means clustering. The center of each cluster is termed as texton and is the representation of a set of projection of each filter on the image and texton map is generated as shown in Fig.4.2 (b).

a. Integrated WEDS based Image graph and NCUT image Segmentation

From 100 labeled super pixel and 25 labeled texon image representations we formed single combined spatio-color-texture mapped image. To obtain this integrated image, region wise intersection of these two

cues is performed and mapped image can be seen from Fig. 4.2 (c). From this map, color-texture features: color mean, standard deviation, entropy, and smoothness for each labeled region are extracted to formulate similarity measure.

For NCUT image segmentation, generation of weighted similarity matrix using graph vertices is important step. In our algorithm we use the labeled region blocks as vertices and based on their intensity and texture features [29], required similarity matrix i.e. edge weight between vertices of graph is generated. In order to obtain final similarity matrix, initially Euclidean distances $Dist(sp),\ Dist(t),\ and\ Dist(clr)$ in terms of spatial location $(i,\ j)$, texture features $(ft(i),\ ft(j))$ and color features $(fc(i),\ fc(j))$ between integrated labeled region centroid and image pixel in neighborhood of it are extracted. Further, Spatio-Color Similarity (SCS) and Texture Similarity (TS) are calculated using these distances as:

$$SCS = \exp\left(-\left[\frac{Dist(clr)^2}{2\sigma_{clr}^2} + \frac{Dist(sp)^2}{2\sigma_{sp}^2}\right]\right) \quad \text{and} \quad TS = \exp\left(-\frac{Dist(t)^2}{2\sigma_t^2}\right) \tag{4.7}$$

Where, $\sigma_t, \sigma_{sp}, \sigma_{clr}$ are scale factor of Gaussian similarity function. These are kept small in the algorithm to find similarity between close pixels as the nuclear region is compact around lumen in normal breast duct. These two similarities SCS and TS are then combined to form weighted color texture similarity (WCTS) as used by Huchuanet. al., [30] to represent image as graph using:

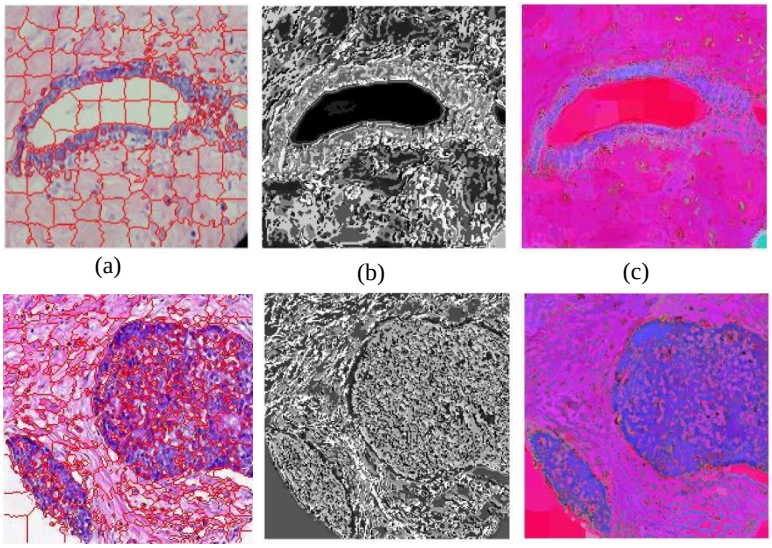
$$WCTS = \sqrt{SCS \times TS} + \alpha_t TS \tag{4.8}$$

Weighted constant α_t decides texture proximity weight in image and segments breast histology duct regions. We also propose a simple Weighted Euclidean Distance based Similarity (WEDS) measure as given by following equations:

$$Dnew = S1 \times Dist(clr) + S2 \times Dist(t) + S3 \times Dist(sp)$$

$$WEDS = \exp\left(-\frac{Dnew^2}{\sigma_{new}^2}\right) \tag{4.9}$$

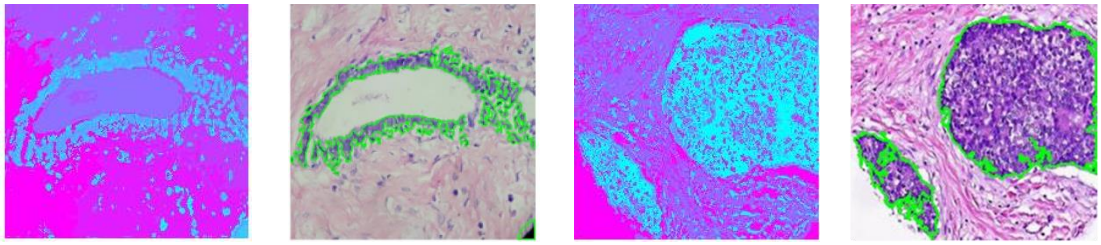
With experimentation we set the values as $S1 = 0.7,\ S2 = 0.25,\ and\ S3 = 0.05$ This weighted similarity measure is considered as input to NCUT and duct nuclear arrangement lining lumen or without lumen in breast biopsy images is detected automatically. Segmented image obtained after NCUT shows partition of image as stromal, luminal and nuclei region and can be seen from Fig.4.3 (a), (c). The detected breast ducts in non-malignant as well as malignant histology images are as shown in Fig.4.3 (b), (d).



(a) (b) (c)

(d) (e) (f)

Fig. 4.2.1st column: Sper pixel map, 2nd column: texton map, 3rd column: Integrated Spatio-color-texture map image for NM (a-c) and M (d-f) breast histology images



(a) (b) (c) (d)

Fig. 4.3.(a), (c)SCT-NCUT Clustered output, (b), (d) Segmented epithelial ducts for NM (a-b) and M (c-d) breast histology images

b. Quantitative Evaluation

The quantitative analysis of results obtained using proposed segmentation algorithm as given in Tables4.1 and 4.2. This analysis shows that incorporation of spatial, color and texture cues for graph representation of breast duct segmentation in NCUT improves results and detects breast ducts with more RI and less GCE, VI values. The Tables also depicts that the segmentation accuracy for proposed similarity measure is also increased from 90.92% to 91.79% for nonmalignant images. Also, the comparison of performance measures shows that, RI for proposed method is greater than implemented classical SLICT-NCUT approach. Also, the values of GCE and VI for the proposed method are low which depicts that our method segments breast histology images more precisely with integrated spatio-color-texture information.

TABLE 4.1: Performance Measures for Proposed SCT-NCUT and implemented SLIC-NCUT Algorithm with WEDS

Avg. Values	Images	SCT- NCUT	SLICT-NCUT
RI	NM	0.8526	0.8384
	M	0.6761	0.6642
GCE	NM	0.1055	0.1129
	M	0.4000	0.4039
VI	NM	0.6466	0.6767

	M	3.6544	3.7011
Sensitivity (%)	NM	87.89	88.95
	M	69.67	76.96
Specificity (%)	NM	92.47	90.87
	M	83.24	80.08
Accuracy (%)	NM	91.79	90.92
	M	77.26	79.84

TABLE 4.2: Performance Measures for Segmentation output using Proposed Distance Similarity Measure

Avg. Values	**Images**	**SLICT- NCUT**		**Proposed SCT-NCUT**	
		WCTS-NCUT	**WEDS-NCUT**	**WCTS-NCUT**	**WEDS-NCUT**
RI	NM	0.8043	0.8384	0.8436	**0.8526**
	M	0.6522	0.6642	0.6713	**0.6761**
GCE	NM	0.1160	0.1129	0.1127	**0.1055**
	M	0.4064	0.4039	0.3947	**0.4000**
VI	NM	0.7579	0.6767	0.6827	**0.6466**
	M	3.7223	3.7011	3.5239	**3.6544**
Sensitivity (%)	NM	93.02	88.95	87.55	**87.89**
	M	67.69	76.96	70.42	**69.67**
Specificity (%)	NM	88.14	90.87	91.89	**92.47**
	M	86.99	80.08	82.29	**83.24**
Accuracy (%)	NM	88.65	90.92	91.28	**91.79**
	M	76.55	79.84	78.56	**77.26**

5. AIFS based Color-Texture Breast Histology Image Segmentation

AIFS based color-texture spectral clustering algorithm is proposed to segment nuclear lining lumen in digitized breast biopsy images. The variation of tissue appearance in various images due to staining concentration, fixation and sectioning of gross sample, can be addressed with additional knowledge of uncertainty about color and texture of digitized breast duct region. Hence the approach of adding uncertainty is used in our work. AIFS considers non-membership and hesitancy along with membership of image pixel. Main contributions in the proposed AIFS based segmentation algorithm are as follows:

1) Developed a novel AIFS super pixel generation method to extract color-based boundary map of image,
2) Proposed two fuzzy clustering and AIFS clustering based texture maps generation to extract texture characteristics of breast duct region,
3) Developed AIFS texton generation technique to capture rotation invariant texture information from digitized breast image in fuzzy set theoretic domain and
4) AIFS based affinity matrix generation is formulated by integrating above AIFS super pixel and texture maps for image graph representation.

Proposed robust and accurate AIFS based breast histology image segmentation pipeline is as shown in Fig.5.1.

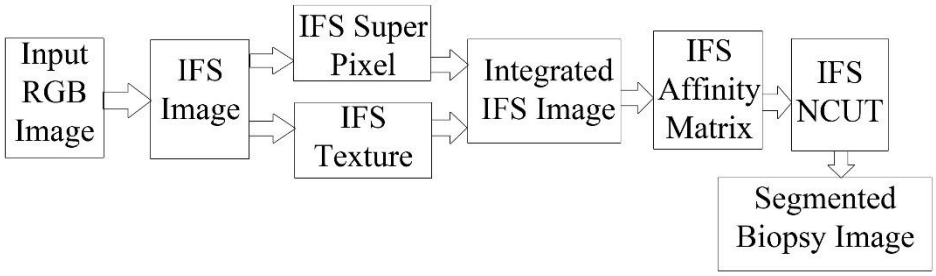


Fig. 5.1. AIFS based Breast Histology Image Segmentation

a. AIFS Notions:

Atanassov's IFS's are characterized by a degree of membership μ_A, degree of non-membership, and hesitancy π_A on a universe of discourse A in X, $X=\{x_1,x_2,.....,x_n\}$ and mathematically it is expressed as: $A=\{(x,\mu_A(x),v_A(x))\mid x\in X\}$ where, $\mu_A(x)$ and $v_A(x)$ are membership and non-membership degrees of the element to the set respectively and are functions $\mu_A(x),v_A(x):X\to[0,1]$ such that $0\le\mu_A(x)+v_A(x)\le 1$. An intuitionistic fuzzy index or the hesitation index function is given as $\pi_A(x)=1-(\mu_A(x)-v_A(x))$. It expresses lack of knowledge on the membership of $x\in A$ and $0\le\pi_A(x)\le 1$.

b. Intuitionistic Fuzzy Image Representation

An image I of size $M \times N$ pixels, whose intensity values lie between [0, L-1], is represented in AIFS domain as: $I = \left\{ \left(x_{ij}, \mu_I(x_{ij}), \nu_I(x_{ij}), \pi_I(x_{ij}) \right) \right\}$ for $i = 1, 2, ..., M$, $j = 1, 2, ..., N$ where $\mu_I(x_{ij}), \nu_I(x_{ij})$ and $\pi_I(x_{ij})$ are membership, non-membership and hesitancy values of $(i, j)^{th}$ pixel in image respectively. Membership and non-membership values at each pixel are extracted using normalized intensity level and Sugeno type complement respectively as: $\mu_I(x) = g(x)/L-1$ and $\nu_I(x) = 1-\mu_I(x)/1+\lambda\mu_I(x)$ where, $\lambda > 0, N(1) = 0, N(0) = 1$ Thus, final AIFS breast histology image is represented as: $A_\lambda^{IFS} = \{ x, \mu_I(x), (1-\mu_I(x))/(1+\lambda\mu_I(x)) \mid x \in X \}$ with hesitation degree $\pi_I(x) = 1-\mu_I(x)-(1-\mu_I(x))/(1+\lambda\mu_I(x))$.

In AIFS based automated breast histology analysis, firstly image is preprocessed using median filtering to reduce noise present in histology images and further it is represented in AIFS domain using membership, non-membership and hesitancy functions as given above. Median filtering is performed using 9 × 9 filter kernel for all images in study. In next step, the two maps: AIFS super pixel color map and AIFS texture map are generated. Further, AIFS color-texture based affinity matrix is constructed for final image segmentation using normalized cuts method. In algorithm, we have generated texture maps using three approaches. First texture map is constructed using AIFS texton model. As this texton map includes large computations over multi-scale, multi-orientation filter banks on each image pixels, block based fuzzy clustered texture map is proposed as second approach. In third method FCM based texture map is generated and the three maps are integrated separately with AIFS super pixel map to form three similarity matrices, which represents image as a graph for NCUT segmentation.

c. AIFS Super Pixel generation

A method for super pixel generation using AIFS similarity is introduced, which more accurately finds compact, boundary adhere image super pixels as compared to super pixel generated previously in chapter 4. For AIFS breast histology image A_{IFS} of size $m \times n$ pixels, S initial super pixels using rectangular grid of Ks sized blocks is obtained with $S = \sqrt{(N / Ks)}$; $N = m \times n$. To obtain final super pixel image an iterative clustering is applied using an AIFS distance D which is used in [35] between a grid center $A_{IFS}(m, n)$ and pixel $A_{IFS}(p, q)$ surrounding the grid neighborhood $\mathrm{P} \times \mathrm{Q}$ and is extracted as:

$D = \sum_{p \in P} \sum_{q \in Q} d\left(A_{IFS}(m,n), A_{IFS}(p,q)\right)$ where $d\left(A_{IFS}(m,n), A_{IFS}(p,q)\right)$ is the Euclidean distance between two pixels in initial super pixel grid calculated for each channel of AIFS color image and is given as:

$$d = \sqrt{\left(\mu_{A_{IFS}}(m,n) - \mu_{A_{IFS}}(p,q)\right)^2 + \left(v_{A_{IFS}}(m,n) - v_{A_{IFS}}(p,q)\right)^2 + \left(\pi_{A_{IFS}}(m,n) - \pi_{A_{IFS}}(p,q)\right)^2} \quad (5.1)$$

Distance of each channel is further combined to form AIFS distance D and AIFS distance based fuzzy similarity, FSIM is formulated to generate compact similar region super pixel. This image representation provides measure of alikeness between the neighborhoods of each grid center using equation:

$$FSIM = \begin{cases} e^{-D} & \|X(i) - X(j)\|_2 < r \\ 0 & otherwise \end{cases} \quad (5.2)$$

where, r is spatial threshold less than S. Centroids of each super pixel are updated with every iteration and assigned mean intensity value of super pixel. Mean squared error between new and previously calculated centroid is then obtained and algorithm converges with minimum error. Compact, boundary adhere AIFS super pixel representation of breast biopsy image is generated and can be seen from Fig. 5.3 (a), (e). Results show that breast duct boundary, lumen and fibro-vascular stroma boundary is well differentiated for final duct region detection.

d. Texture Map Generation

Texture of epithelial nuclear arrangement in the breast tissue image plays important role in examination of diseased tissue samples. Hence, texture information of the nuclear lining lumen in breast histology images is captured using AIFS based texture representations of the image. We have generated three texture maps and detail methodology for generation of these maps is described in following sub-sections.

1. AIFS Texton generation

Texton is a texture measure used in computer vision algorithms and is obtained as the centroids of clustered output of filter responses [22]. We introduced AIFS texton generation, which is extension of texton generated in our previous chapter 4, section 4.2. For generation of AIFS texton map, we use same multi-scale; multi-oriented Gabor and Laplacian of Gaussian filter kernels to extract texture information of pixel regions in AIFS image. The filtered vectors are then clustered by applying Intuitionistic Fuzzy C-Means (IFCM) clustering [31]. This results in AIFS based texton image representation in which uncertainty information of pixel to find final clusters is incorporated. In IFCM, the objective function which is to be minimized includes the term hesitancy and non-membership for color histology images as compared to conventional fuzzy C means algorithm (FCM) clustering. The modified IFCM algorithm consists of the following steps:

1) Creation of IFS image using Sugeno type intuitionistic fuzzy generator as explained in this chapter section 5.2
2) Define membership and objective function using Euclidean distance between two image pixels for each channel in AIFS breast image for clustering as:

$$D_{IFS}(A_j, v_i) = \sqrt{\sum_{i=1}^{n}\left(\left(\mu_{A_j}(x_i) - \mu_{v_i}(x_i)\right)^2 + \left(v_{A_j}(x_i) - v_{v_i}(x_i)\right)^2 + \left(\pi_{A_j}(x_i) - \pi_{v_i}(x_i)\right)^2\right)} \quad (5.3)$$

then objective function of IFCM is set using membership degree of j^{th} sample to its cluster and distance as:

$$\min\ J_m(A,V) = \sum_{j=1}^{P}\sum_{i=1}^{c}\mu_{ij}^{m} D_{IFS}^2(A_j, v_i) \quad (5.4)$$

3) The objective function is solved using Lagrange multiplier method and membership degrees are updated as:

$$\mu_{ij} = 1\Big/\left(\sum_{r=1}^{c}\left(D_{IFS}(A_j, v_i)\Big/D_{IFS}(A_j, v_r)\right)\right)^{2/m-1} \quad (5.5)$$

Also, centroids of membership, non-membership and hesitancy are updated using equation:

$$\partial L/\partial\mu_{v_i}(x_i) = \partial L/\partial v_{v_i}(x_i) = \partial L/\partial\pi_{v_i}(x_i) = 0;\quad \mu_{v_i}(x_i) = \sum_{j=1}^{P}\mu_{ij}^{m}\mu_{A_j}(x_i)\Big/\Sigma_{j=1}^{P}\mu_{ij}^{m}$$

$$v_{v_i}(x_i) = \sum_{j=1}^{P}\mu_{ij}^{m} v_{A_j}(x_i)\Big/\Sigma_{j=1}^{P}\mu_{ij}^{m};\quad \pi_{v_i}(x_i) = \sum_{j=1}^{P}\mu_{ij}^{m}\pi_{A_j}(x_i)\Big/\Sigma_{j=1}^{P}\mu_{ij}^{m} \quad (5.6)$$

4) Perform iteration till algorithm converges. If $\left(\Sigma_{i=1}^{c} d(v_i(k), v_i(k+1))/c\right) < \varepsilon$; some predefined threshold, algorithm terminates else an iterative process similar to the fuzzy C-means to solve these equations is employed in IFCM.

We use 64 even-odd Gabor filter kernels with four scale, eight orientation and 3three scale Laplacian of Gaussian filter bank on AIFS image thus forming total 201 filtered output and are clustered using IFCM to 25 clusters as shown in Fig. 5.3 (b), (f). AIFS texton results precisely depict the luminal, stroma and nuclear region based on rotation invariant filter response texture measure.

2. IFCM clustering based Texture map generation

In AIFS based texton generation, large number of filter kernels are convolved with image to obtain clustered vector filter responses. This clustered lower dimensional data, captures similar texture region characteristics in image broadly and also need more computations. Hence to improve results, AIFS texture map generation algorithm using block based IFCM clustering on multi-texture feature image as given in Fig. 5.2 is proposed in this work. Here, IFCM is applied on a group of pixels instead of individual pixel to form small fuzzy texture map and thus reduces computations required with better clustering output based on texture. Before applying IFCM a multi-texture feature image is generated from AIFS image block representation and further pixel in original AIFS image is relabeled using clustered output obtained earlier to extract final fuzzy texture map.

Fig. 5.2. Proposed AIFS Texture Map Generation Approach

For generating new texture map, initially, AIFS breast image is divided into Ks blocks of small size $b \times b$ ($b = 4$) and then multi texture feature image representation is obtained. If image size is 200×200 , then number of blocks in image is 2500. Texture features used in this algorithm are mean, standard deviation, entropy and smoothness of individual block. This block structure multi-texture feature AIFS image of size $50 \times 50 \times 4$ is obtained and then clustered into 4 clusters using IFCM as duct nuclear region, stroma region, lumen and background. Using these clustered outputs with respect to blocks, original AIFS image pixel is relabeled to obtain final AIFS based texture map as shown in Fig.5.3 (c), (g).

3. FCM based Texture image representation

Texture information of nuclear arrangement in the breast biopsy image is also captured using conventional fuzzy C means clustering on multi-texture feature image vectors to compare the results of proposed IFCM based texture map generation. FCM based texture map is generated using same procedure as proposed in previous subsection except initial FCM clustering approach used to detect duct nuclear, stroma region, lumen and background. Further, original FS image pixel is relabeled with respect to clusters obtained and final FCM based fuzzy texture map is obtained as shown in Fig.5.3 (d), (h).

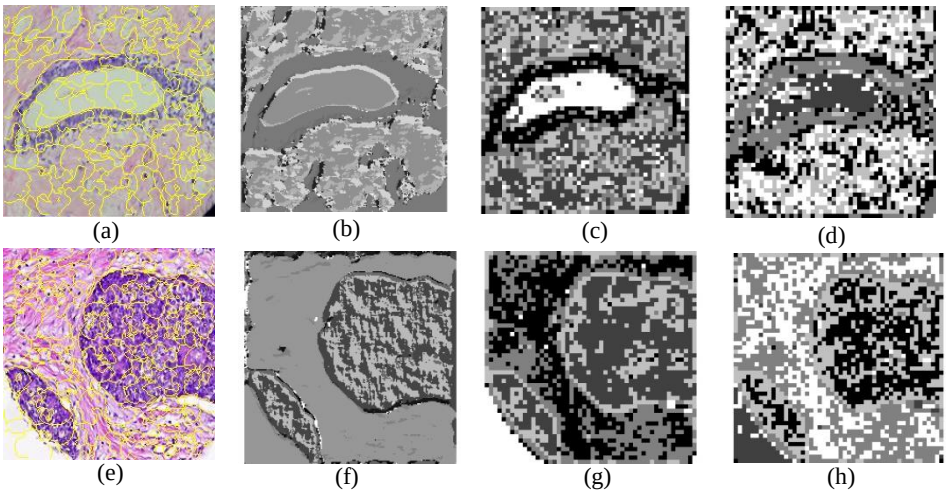


Fig. 5.3. AIFS based (a, e) super pixel, (b, f) Texton, (c, g) IFCM Texture, and (d, h) FCM FS texture NM and M Breast Histology Image Maps respectively

B. Normalized Cuts (NCUT) based breast histology image segmentation

We propose AIFS based affinity matrix calculations by considering group of pixels as a vertex instead individual pixel. For obtaining affinity matrix, an integrated AIFS color-texture mapped image is used and final nuclear arrangement lining lumen in breast histology image is detected automatically. Firstly, AIFS super pixel image with 100 labels and AIFS texture image map with 4 labels are integrated using simple region-based intersection to form single AIFS color-texture labeled image. From this integrated mapped image, color-texture features i.e. color mean, standard deviation, entropy, and smoothness of each labeled region are extracted [29]. These color-texture features are then used to obtain affinity matrix using weighted color-distance measure for AIFS image representation. Euclidean distances $Dist_{IFS}(sp),\ Dist_{IFS}(t),\ and\ Dist_{IFS}(clr)$ in terms of spatial location (i, j), texture features $(ft(i), ft(j))$ and color features $(fc(i), fc(j))$ between labeled regions of integrated mapped image are extracted respectively to calculate weighted distance as: $Dnew = S1 * Dist(sp) + S2 * Dist(t) + S3 * Dist(clr)$ and AIFS color-texture distance based Similarity (IFSim) measure is formulated using equation:

$$IFSim = \exp\left(-\left(Dnew^2\right)\Big/\left(\sigma_{new}^2\right)\right) \tag{5.7}$$

The values of $S2\ and\ S3$ considered in experiments are more than $S1$ so as to give more weight age to color-texture information of epithelial nuclear region and set as $S1 = 0.05,\ S2 = 0.15,\ and\ S3 = 0.8$. Thus, AIFS weighted similarity matrix is further used as input to final NCUT segmentation and breast epithelial duct image region for classification is obtained. To detect epithelial nuclear lining lumen in breast duct image accurately by ignoring stromal invasion, the morphological opening is used and small exterior nuclei are removed as post-processing step. Three affinity matrices are extracted with integration of super pixel and three texture maps generated. Corresponding NCUT clustered outputs are shown in Fig.5.4 (b-d) and final segmented breast ducts from these three affinity matrices are as shown in Figure 7(e-f) for M breast histology images. Fig.5.4 (a) shows GT image overlap on original image. Similar results for NM breast histology image are as shown in Figure 5.5 (a-g). Results shows that breast duct region is extracted more precisely using integrated AIFS super pixel and texture map NCUT (AIFSSTX-NCUT) approach.

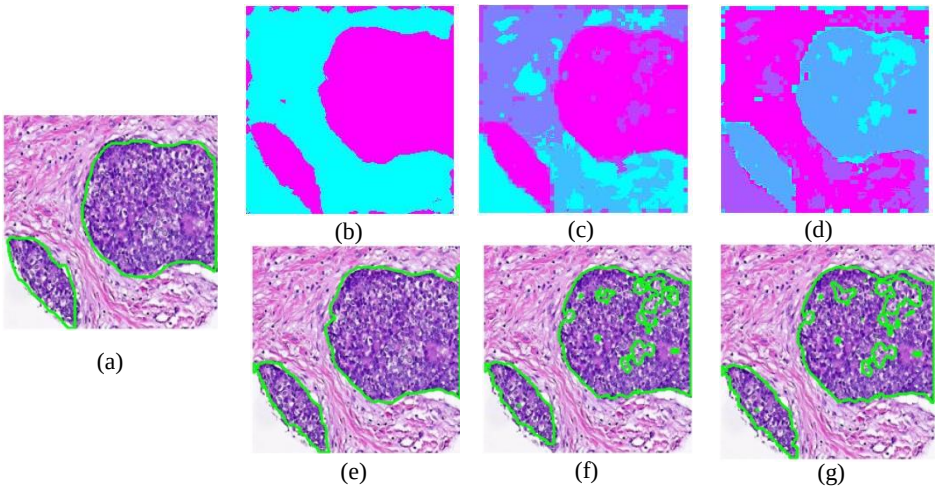


Fig. 5.4. AIFS Segmented output (a) GT overlap, (b-d) Clustered output and (e-g) Segmented breast ducts from AIFS Texton (AIFSST), IFCM Texture (AIFSSTX), and FCM FS texture (FSSTX)based NCUT M Breast Histology Images respectively

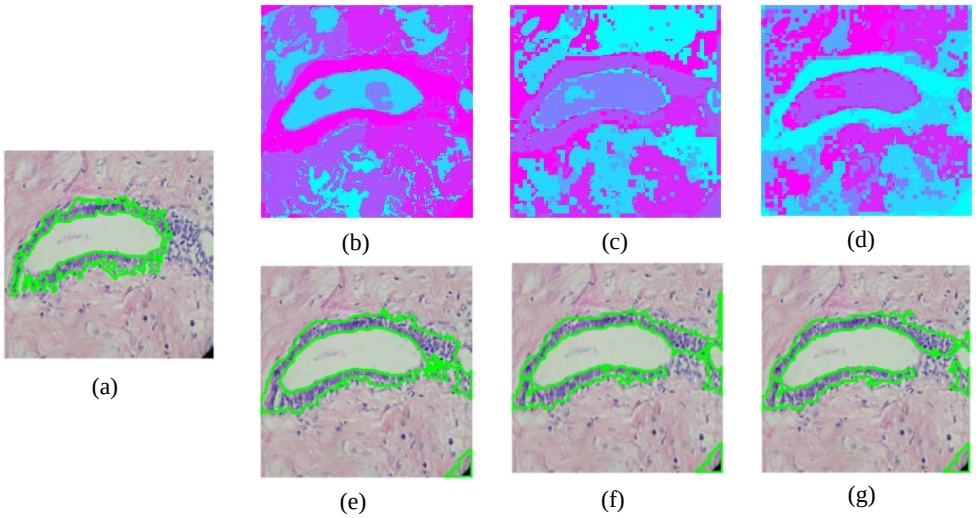


Fig. 5.5. AIFS Segmented output (a) GT overlap, (b-d) Clustered output and (e-g) Segmented breast ducts from AIFS Texton (AIFSST), IFCM Texture (AIFSSTX), and FCM FS texture (FSSTX) based NCUT NM Breast Histology Images

C. Quantitative Evaluation of Results:

Inclusion of AIFS in breast image segmentation improves qualitative results which deals with uncertainty and vagueness in the pixel intensities of breast tissue images, imitating pathologist visual image analysis procedure. Quantitative evaluation of the results concludes that our proposed integration of spatial-color-texture information in fuzzy clustering method for tissue image segmentation yields in excellent automated diagnostic results for breast biopsy diagnosis of malignancy as depicted in Table 5.1. For NM breast images RI is increased to 0.81 from 0.65 using AIFS color-texture NCUT (AIFSSTX-NCUT) approach as compared to AIFS color-texton (AIFSST-NCUT) method. Similarly, GCE and VI are found to be less which quantifies the segmentation results obtain. Also, Sensitivity (Sn), Specificity (Sp) for the proposed AIFS color-texture NCUT are improved over other two for NM and M images as shown in ROC plot from Fig.5.6 (a), (b). Segmentation Accuracy (Acc) is also improved for proposed AIFSSTX-NCUT approach and achieves 88.54% and 77.05% respectively for NM and M images. However, for M breast images the values remain similar for all approaches as the GT images used for evaluation captures duct boundary as a whole. Whereas AIFS approach finds inner duct variations as well, these are also essential for malignancy detection by pathologist. Here is a scope of improvement in algorithm by improving GT image generation for M breast histology images. Number of image dataset used for automated M breast histology image analysis is less to work on population evaluation basis.

TABLE 5.1: Average values of performance measures for NM and M breast histology images using three experiments

NCUT based on	Image Data	RI	GCE	VI	Sn %	Sp%	Acc%
AIFSST	NM	0.65	0.16	1.09	90.00	71.55	74.20
	M	0.68	0.26	1.30	73.19	83.63	77.89
AIFSSTX	NM	**0.81**	**0.13**	**0.81**	**79.48**	**90.20**	**88.54**
	M	**0.67**	**0.26**	**1.30**	**63.66**	**88.05**	**77.05**

FSSTX	NM	0.81	0.13	0.81	79.18	90.89	89.04
	M	0.66	0.26	1.32	61.79	88.23	75.61

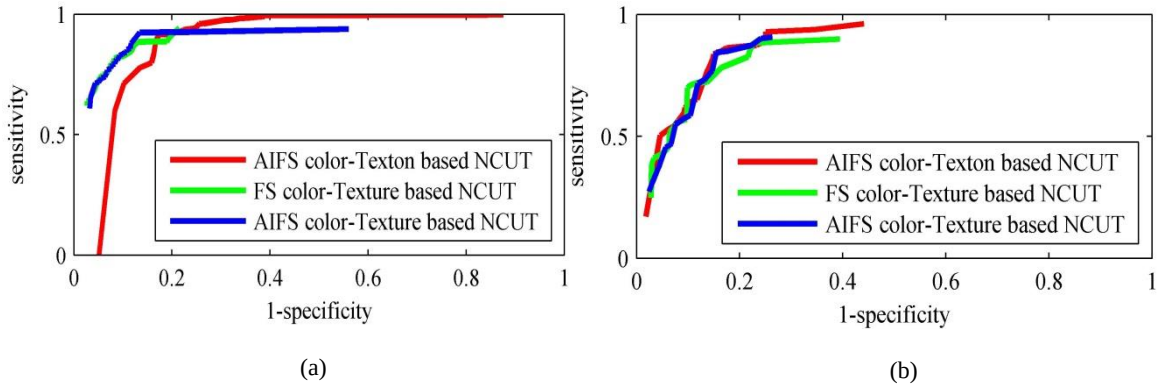


Fig. 5.6. Performance Measures for three AIFS based experiments proposed in Thesis for: (a) NM images, (d) M images

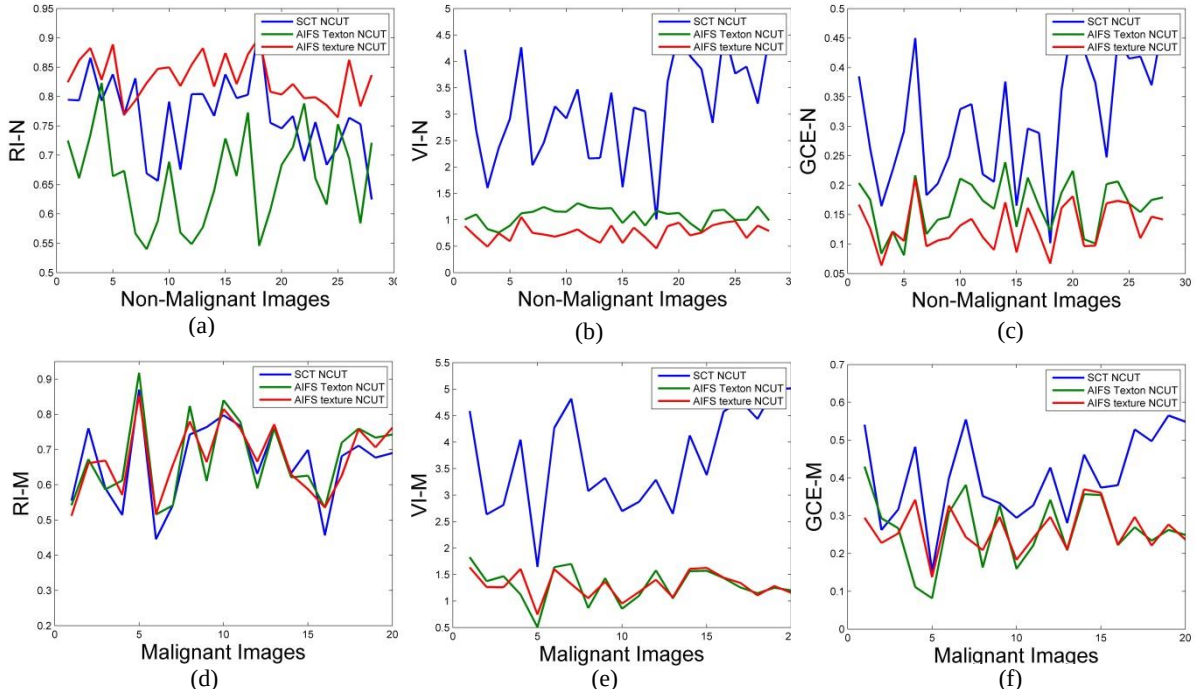


Fig. 5.7. Performance Measures for three segmentation methods proposed in Thesis for: (a-c) NM images, (d-f) M images

Three methods for segmentation of breast duct epithelial region surrounding lumen or solid sheets without lumen are proposed in this thesis. Basic proposed method only integrates spatio-color-texture maps to form affinity matrix for NCUT image segmentation. Whereas incorporation of AIFS in generation f image graph for NCUT improves segmentation results for images. Proposed AIFS texture map still improves segmentation results for breast histology images and quantitative evaluation of proposed algorithms for selected images in database is shown in Fig.5.7(a-c) for NM and (d-f) for M breast histology images. From this we can conclude that AIFSSTX based graph achieves best segmentation results in experiments carried in this research.

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)

6. TEXTURE BASED BREAST HISTOLOGY IMAGE CLASSIFICATION

Breast cancer detection from digitized breast biopsy samples and its grade assessment is crucial task for pathologists. An automated and robust method of detecting malignancy in breast cancer is required to assist pathologist. This task is challenging due to complexity of tissue architecture, appearance in non-malignancy and malignancy cases. In order to automate the breast histology classification for non-malignancy and malignancy at 40X optical magnification, we propose texture analysis of segmented breast duct images using texture and morphology features.

To accomplish task of breast image classification, segmented normal breast duct and malignant duct structure from digitized breast histology image using proposed spatio-color-texture based graph method and AIFS based segmentation methods from chapters 4 and 5 are used as input dataset. Further the texture feature analysis and texton classifier has been applied separately on these image datasets to detect presence of cancer in biopsy images automatically. Initially only texture model based texton classifier has applied on the input image data to classify them as non-malignant and malignant images. Further, linear discriminant analysis (LDA), Support vector machine (SVM) and k-nearest neighbor classifier (k-NN) classifiers have also been applied on the extracted texture, morphological feature sets of input image data and quantitative analysis of classifier performances is performed.

A. V-Z classifier for malignancy detection

Texton model classification approach is based on a statistical learning of textures which are modeled using joint distribution of multi-oriented, multi-scaled filter responses. Basic block diagram for this V-Z (Verma-Zisserman) classifier [32] is shown in Fig.6.1. The probability distribution of types of images is represented by textons. Thus, texture models are developed from training image dataset of malignant and non-malignant breast histology image class respectively. Classification of histology images is achieved by mapping test biopsy image to a texton distribution and compared this distribution to the trained class models. Total data set of 60 segmented breast biopsy images is used for classification, which is divided into two disjoint sets. First set of breast histology images is used for generation of texton dictionary and texton model for classifier whereas second set is used for classification. Each image in training dataset one is first mapped to textons for generating texton dictionary. This dictionary is reduced to single texton model per class using vector clustering of generated textons for each sample image in dataset one. After generation of texton model for both the classes of biopsy images, a texton model for a query breast histology image from dataset two is generated and is then classified to closest texton model using k-NN classifier. Only texture model-based classifier fails to classify images as structure of duct nuclear lining in breast plays a very important role in malignancy identification as can be seen from Table 6.2 for SCT-NCUT segmented image dataset. The developed classifier only considers the texture orientation of breast histology images. Hence texture-morphology feature based linear classifier is developed and compared with other state of art classifiers.

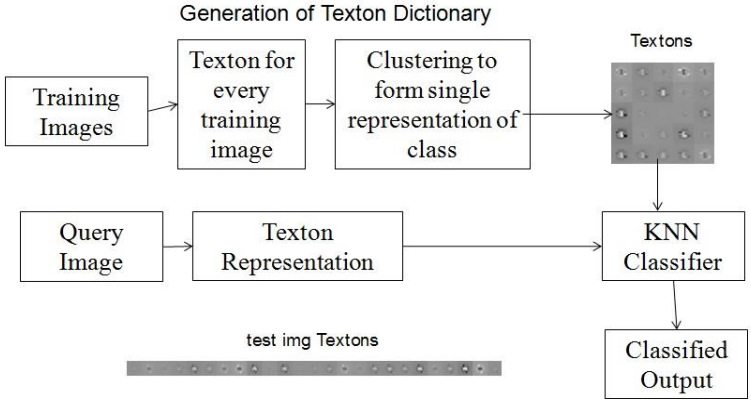


Fig. 6.1.V-Z Classifier for Breast Histology Images

B. Feature based Linear Classifier for Histology Images

In texture-morphology feature based linear classifier, segmented epithelial lining surrounding the lumen for breast histopathology images are used for two class classification. Here the segmented regions are obtained using spatio-color-texture graph segmentation method as proposed in earlier sections. Deviation of these segmented duct structure can detect abnormality in breast tissue image which is one of the parameters observed by pathologist while detecting presence of cancer. The duct image dataset used for classification can be seen from Fig.6.3. Further, texture features such as Gray Level Co-occurrence Matrix (GLCM) [33], Graph Run Length Matrix (GRLM) features [34], and morphological Euler number are extracted. Next, LDA is used to classify breast histology images. The LDA classifier performance is compared with classical k-NN and SVM classifiers. The detailed histopathology breast image classification process is shown in Fig. 6.2.

Fig. 6.2. Breast Histology Image Classification

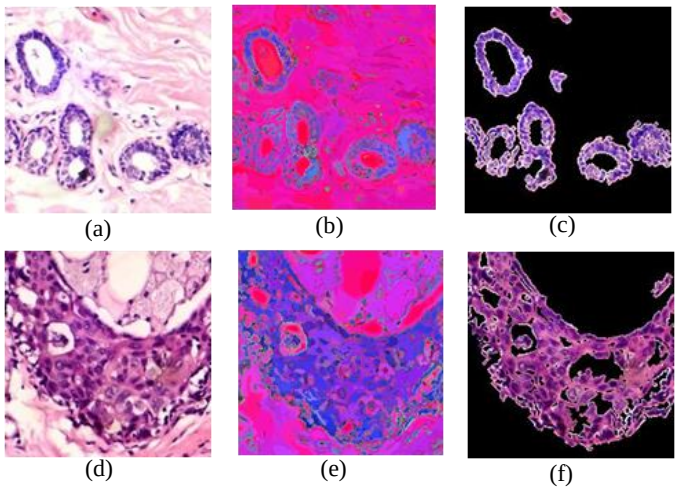


Fig. 6.3. Image Segmented output: 1st column shows original image 2nd column shows integrated spatio-color texture map and last column shows segmented breast histology duct.

C. Texture-Morphological Feature Extraction

Combination of texture & morphological features like first order intensity, GLCM, GRLM and morphological Euler number features are used for the analysis of breast histology duct images. The features used are detailed as: Mean, standard deviation, variance, coefficient of variation, are the first order intensity-based texture features. GLCM texture features such as energy, contrast and entropy at unit distance between pixels and angle of 0° are extracted

with from segmented breast histology image. Eleven GRLM texture features are used for textural analysis at $0^{\circ}, 45^{\circ}, 90^{\circ}$ and 135°. Features extracted differentiate the structural in-homogeneity for structures with similar gray values as well as detects the presence of nuclei lining which plays a significant role in identification of malignancy in breast histopathology images. Structural change in architecture of breast duct tissue image are identified using Euler number. This feature can be extracted from binary segmented breast histology image mask. In the experimentation it is observed that this feature differentiates the two-image class more prominently than others.

A set of 52 texture-morphological features are extracted from each segmented breast histology image. Among these extracted features, rank selection approach finds best 30 features which are then selected for final classification. It is observed that Morphological Euler number has highest ranking as it detects the presence of lumen in breast duct as shown in Fig.6.4 (a) and (b) shows other maximum ranked texture features. In normal breast duct structure, two epithelial nuclear layered region surrounding lumen is present and in case of malignant breast duct (breast carcinoma) is affected by presence of these epithelial nuclei inside lumen. Thus, duct structure is deviated from two-layer epithelial nuclear layer to multi-layer nuclear layer. Sometimes very less part of lumen is detected in a breast duct. This structural information is captured by Euler number.

(a) (b)

Fig. 6.4.(a)Euler number Plot, and (b) best features scatter plot for histology image

D. Breast Histology Image Classification and quantitative evaluation

Classification of these images is performed using LDA classifier. The multivariate normal distribution density function for each class is calculated first prior to the training. Results obtained from LDA classifier is then compared with k-NN and SVM classifier with radial basis kernel function. Total 70 Real time images (40: Non-malignant and 30: Malignant) are used for development of classifier. This dataset is divided into two disjoint sets consisting of 20 NM and 15 M breast duct images each for training and testing classifier. LDA classifier results illustrates its outperformance over other two methods and can be seen from Tables 6.1 and 6.2 for both the dataset used in experiments. The comparative analysis for SCT-NCUT as well as AIFSSTX-NCUT segmented dataset shows that LDA classifier outperforms over other methods and achieves 100% classification accuracy for both classes in test dataset. Whereas the SVM and k-NN classifiers achieves 80%and 40% overall accuracy for malignancy class as shown in Table 5. Correct classification rate for LDA is 100% whereas for k-NN and SVM it gives 70% and90% respectively, which quantifies the breast image classifier performance indicating higher classification accuracy for the images in dataset. The proposed texture-morphological feature based linear classification method is also evaluated using parameters like sensitivity, specificity, Positive predictive value, Negative predictive value and correct classification rate. All these metrics demonstrate better classification using LDA than other two methods. However, as the number of image database for malignancy cases is less in experimentation, we achieve 100%

classification accuracy, sensitivity and specificity in experiments. Although promising results for image classification are achieved, improvement in results of classification algorithms can be carried in future with large number of images in the experimentation as compared to current very less image database. This can be viewed as proposed analysis on population based statistical classification of breast carcinoma images and providing actual assistance to pathologist in practice.

E. TABLE 6.1: Comparison of LDA, k-NN and SVM classifier for breast histology image classification in percentage

Performance Measures	**Image Database**	**LDA %**	**kNN %**	**SVM %**
Correct classification rate	SCT-NCUT	100	70	90
	AIFSSTX-NCUT	100	60	80
Sensitivity	SCT-NCUT	100	100	100
	AIFSSTX-NCUT	100	40	100
Specificity	SCT-NCUT	100	40	80
	AIFSSTX-NCUT	100	80	60
Positive Predictive Value	SCT-NCUT	100	62.50	83.33
	AIFSSTX-NCUT	100	66.67	71.43
Negative Predictive Value	SCT-NCUT	100	100	100
	AIFSSTX-NCUT	100	57.14	100

F.

G. TABLE 6.2:Average classification accuracy for classifiers

Image Type	**Image Database**	**LDA %**	**kNN %**	**SVM %**	**V-Z Classifier %**
Non-malignant	SCT-NCUT	100	100	100	33
	AIFSSTX-NCUT	100	40	100	-
Malignant	SCT-NCUT	100	40	80	100
	AIFSSTX-NCUT	100	80	60	-

7. CONCLUSIONS AND FUTURE SCOPE

This research work describes the development of algorithms for object level computer aided breast histology image analysis. In pathology image analysis research, histology image examination is considered as gold standard. Due to subjective manual analysis and complex breast tissue appearance, its automated analysis is challenging and need objective decision support to pathologist so as to minimize inter-intra observer variability in decision making.

In this research, we presented robust breast histology image segmentation and classification algorithms using inputs from expert pathologists. First, we propose a spatio-color-texture similarity based NCUT segmentation to detect epithelial duct nuclear lining lumen regions from breast histology images. To achieve this, firstly spatial arrangement and color of duct epithelial lining lumen or solid sheets without lumen have been considered to generate compact super pixel images. Secondly, we have developed algorithm to generate textons: rotation invariant clustered vector filter responses for nuclear texture representation. Finally, we propose to integrate these two maps, namely super pixel and texton which forms single spatio-color-texture similarity measure for NCUT image segmentation method. Quantitative evaluation of proposed segmentation algorithm is performed using parameters namely, RI, GCE, VI, Sp, Sn, Acc. The average values of these parameters over entire image database is found to be 0.85, 0.11, 0.65, 87.89%, 92.47%, 91.79% and 0.68, 0.40, 3.65, 69.67%, 83.24%, 77.26% respectively for NM and M breast histology images with proposed WEDS similarity measure.

Next, we extend the super pixel and texton generation techniques to AIFS domain by inculcating hesitancy and non-membership knowledge of breast tissue appearance irrespective of staining concentration and different fixation procedure while preparing sample specimen from breast gross tissue. Using this fuzzy representation, AIFS based similarity measure is proposed to find more accurate breast duct segmentation using AIFS based spectral clustering. We also introduced AIFS clustering based fuzzy texture map is generation method to reduce computations required for AIFS texton image representation and achieve better segmentation results. This AIFS texture map is compared with conventional fuzzy C-means (FCM) clustering based fuzzy texture map. The two newly generated texture maps are integrated with AIFS super pixel separately to form two AIFS based color-texture mapped breast images which represents image graphs. Further NCUT is performed on these resultant AIFS image graphs separately for segmenting duct epithelial regions and quantitative evaluation of each proposed integrated map is performed. Performance evaluation of three AIFS NCUT segmentation experimentation shows that AIFSSTX-NCUT achieves better RI and Acc values as compared with other two experiments carried for both image types (NM and M). Moreover, it also reduces GCE and VI, which shows effectiveness of incorporation of proposed AIFS clustering based texture map and AIFS super pixel in NCUT segmentation approach.

Finally, we presented a texture and morphological feature based linear classifier to classify the segmented breast duct images to non-malignant and malignant classes. The proposed texture feature based linear classifier achieves better classification accuracy as compared with SVM and K-NN classifier. For both the data sets linear classifier achieves 100% classification accuracy. To perform quantitative evaluation of proposed algorithms, nonmalignant and malignant breast tissue images and corresponding GT image database was developed in consultation with expert pathologist. The experimentations and evaluations carried on this image database demonstrate the superiority of proposed segmentation and classification algorithms.

The research work presented in this dissertation carries entire research on portion of whole slide tissue image and could be extended to segment and analyze whole slide tissue for affected area identification and follow analysis procedure at different magnification level as pathologist does. Classification experimentation performed in the research work is carried on very a smaller number of image samples, the classifier performance could be investigated on large database so as to validate the malignancy detection at large population level and provide robust screening tool for quantitative analysis for pathologist.

Publication Partner:

International Journal of Scientific and Research Publications (ISSN: 2250-3153)

Appendixes

A. References

1. Rosai J., Rosai & Ackerman'S Surgical Pathology. Volume 2, 10th edition.

2. Rosen and Paul P., Rosen's Breast Pathology. Lippincott Williams & Wilkins, 3rd edition, 2009.

3. Koss, Leopold G., Melamed, and Myron R., Kosss Diagnostic Cytology & its histopathologic basis. Lippincott Williams & Wilkins, 2006.

4. Bloom H. J. G. and Richardson W. W., Histological grading and prognosis in breast cancer; A study of 1049 cases, of which 359 have been followed 15 years.

5. He L., Rodney L. L., Antani S, and Thoma G. R., Histology image analysis for carcinoma detection and grading. Computer Methods and Programs in Biomedicine, 107(3):538–556, 2012.

6. Doyle S., Agner S., Madabhushi A., Feldman M., and Tomaszewski J., Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In 5thIEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2008), pages 496–499, 2008.

7. Naik S., Doyle S., Agner S., Madabhushi A., Feldman M., and Tomaszewski J., Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI 2008), pages 284–287, 2008.

8. Humayun I., Veillard A., Roux L., and Racoceanu D., Methods for nuclei detection, segmentation and classification in digital histopathology: A review. current status and future potential. IEEE Reviews in Biomedical Engineering, 7:97–114, 2014.

9. Muthu M. R. K., Chakraborty C., Paul R. R., and Ray A. K., Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis. Expert Systems with Applications, 39:1062–1077, 2012.

10. Aitunbay D., Cigir C., Sokmensuer C. and Demir C.G., Color graphs for automated cancer diagnosis and grading. IEEE Transaction on Biomedical Engineering, 57(3):665–674, 2010.

11. Basavanhallya A., Yu E., Xu J., Ganesan S., Feldman M., Tomaszewski J., and Madabhushi A., Incorporating domain knowledge for tubule detection in breast histopathology using O'callaghan neighborhoods. Medical Imaging 2011: Computer-Aided Diagnosis, Proc. of SPIE, 7963:796310–1–15, 2011.

12. Cagatay B., Demir C., Nagic C., and Yener B., Cellgraph mining for breast tissue modeling and classification, In Proceedings of the 29th Annual International Conference of the IEEE EMBS, pages 5311–5314, Lyon, France, 2007.

13. Veta M., Pluim J. P. W., Paul J. V. D., and Viergever M. A., Breast cancer histopathology image analysis: a review. IEEE Transactions on Biomedical Engineering, 61:1400–1411, 2014.

14. Demir C. G., Kandemir M., and Tosun A. B. and Sokmensuer C., Automatic segmentation of colon glands using object-graphs. Medical Image Analysis, 14:1–12, 2010.

15. TosunA. B. and Gunduzdemir C., Graph run length matrices for histopathological image segmentation. IEEE Transactions on Medical Imaging, 30(3):721–31, 2011.

16. Choudhary A., Chakraborty C., Ray A. K., Muthu, M. R. K. and Paul R. R., Texture based segmentation of epithelial layer from oral histological images. Micron, 42:632AS641, 2011.

17. Sertel O., Kong J., Catalyurek U. V., Lozanski G., Saltz J. H., and Gurcan M. N., Histopathological image analysis using model- based intermediate representations and color texture: Follicular lymphoma grading. Signal Processing Systems archive, ACM, 55(1):169–183, 2009.

18. Xu J., Janowczyk A., Chandran S., and Madabhushi A., A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: Applications to histopathology image segmentation. Medical Imaging 2010: Image Processing, Proc. of SPIE, 7623:76230Y–1– 76230Y–12, 2010.

19. Loukas C., Kostopoulos S., Tanoglidi A., Glotsos D., Sfikas C. and Cavouras D., Breast cancer characterization based on image classification of tissue sections visualized under low magnification. In Computational and Mathematical Methods in Medicine, volume 2013. Hindawi Publishing Corporation, 2013.

20. Galaro J., Judkins A. R., Ellison D., Baccon J., and Madabhushi A., An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas. In 33rd Annual International Conference of the IEEE EMBS, pages 3443–3446, 2011.

21. Zhang Y., Zhang B., Coenen F., Xiao J., and Lu W., One-class kernel subspace ensemble for medical image classification. EURASIP Journal on Advances in Signal Processing, Springer Open Journal, 2014.

22. Shi J. and Malik J., Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):888–905, 2000.

23. Cevahir L. and Alatan A. A., Efficient graph-based image segmentation via speeded up turbo pixels. In Proceedings of 2010 IEEE 17th International Conference on Image Processing, ICIP, pages 3013–3016, 2010.

24. Levinshtein A., Stere A., Kutulakos K. N., Fleet D. J., Dickinson S. J., and Siddiqi K., Turbopixels: Fast superpixels using geometric flows. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(12):2290–2297, 2009.

25. Achanta R., Shaji A., Smith K., Lucchi A., Fua P., and Su S. S., SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11):1–8, 2012.

26. Xiang D., Tang T., Zhao L., and Su Y., Superpixel generating algorithm based on pixel intensity and location similarity for SAR image classification. IEEE Geoscience and Remote Sensing Letters, 10(6):1414–1418, 2013.

27. Julesz B. Textons, the elements of texture perceptions, and their interactions nature. In Pubmed, pages 290(5802):91–97, 1981.

28. Malik J., Belongie S., Leung T., and Shi J., Contour and texture analysis for image segmentation. International Journal of Computer Vision, 43(1):7–27, 2001.

29. Gonzalez R. C. and Woods R. E., Digital Image Processing. 10th Edition, Pearson Education, 2013.

30. Lu H., Zhang R., Li S., and Li X., Spectral segmentation via midlevel cues integrating geodesic and intensity. IEEE Transaction on Cybernetics, 43(6), 2013.

31. Chira T., Ray A. K., 2009, Fuzzy Image Processing and Applications with MATLAB, CRC Press, Taylor & Francis Group, LLC.

32. Varma M. and Zisserman A., A statistical approach to texture classification from single images. International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis, 62:61–81, 2005.

33. Haralick R. M., Shanmugan K., and Dinstein I., Textural features for image classification. IEEE Transactions on System, Man, and Cybernetics 8(6):610–621, 1973.

34. Tang X., Texture information in run-length matrices. IEEE Transactions on Image Processing, 7(11), 1998.

35. Mushrif M. M. and Ray A. K., A-IFS histon based multi-thresholding algorithm for color image segmentation. IEEE Signal Processing Letters, 16(3): 168–171, March 2009.

36. Chaira T. and Ray A. K., A new measure using intuitionistic fuzzy set theory and its application to edge detection. Applied Soft Computing, 8(2):919–927, 2008.

37. Chaira T., Intuitionistic fuzzy segmentation of medical images. IEEE Transactions on Biomedical Engineering, 57:1430–1436, 2010.

38. Hanqiang Liua, Feng Zhao b, and Licheng Jiao. Fuzzy spectral clustering with robust spatial information for image segmentation. Applied Soft Computing, Elsevier, 12:3636–3647, 2012. URL http://dx.doi.org/10.1016/j.asoc.2012.05.026.

39. Sen D., Gupta N., and Pal S. K., Incorporating local image structure in normalized cut-based graph partitioning for grouping of pixels. Information Sciences, 248:214–238, 2013.

40. Rand W. M., Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66(336):846–850, 1971.

41. Unnikrishnan R. and Pantofaru C. and Hebert M.,Toward objective evaluation of image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6):929–944, 2007.

42. A. D. Belsare, M.M. Mushrif, M.A. Pangarkar and N. Meshram, "Breast histopathology image segmentation using spatio-colour-texture based graph partition method", Journal of Microscopy, Volume 262, Issue 3, June 2016, Pages 260–273, Early View: Article first published online: 28 DEC 2015, 1-10, 2015, DOI: 10.1111/jmi.12361, J. microsc. 262 (3), 260-273, 2016

43. 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, August 2012, DOI : 10.5121/sipij.2012.3403

44. A. D. Belsare, M. M. Mushrif, M. A. Pangarkar "Breast Epithelial Duct Region Segmentation Using Intuitionistic Fuzzy Based Multi-Texture Image Map", INDICON 2017, IIT Roorkee, Dec 15-17, 2017

45. A. D. Belsare, M. M. Mushrif, M. A. Pangarkar, N. Meshram, "Classification of Breast Cancer Histopathology Images using Texture Feature Analysis", IEEE Region 10, TENCON 2015, 1-4 November 2015, Macau, China

Glossary

A. LIST OF FIGURES

B. LIST OF TABLES

C. LIST OF ABBREVIATIONS

AIFS	Atanassov's Intuitionistic Fuzzy Set
CAD	Computer Aided Diagnosis
FCM	Fuzzy c-Means
FS	Fuzzy Set
GCE	Global Consistency Error
GLCM	Gray Level Co-occurrence Matrix
GRLM	Gray Run Length Matrix
GT	Ground Truth
H&E	Hematoxylin and Eosin
IFCM	Intuitionistic Fuzzy c-Means
k-NN	k-Nearest Neighbor
LCE	Local Consistency Error
LRE	Local Refinement Error
LoG	Laplacian of Gaussian
LDA	Linear Discriminant Analysis
M	Malignant
NM	Non-Malignant
NCUT	Normalized Cuts
PRI	Probabilistic Rand Index
RI	Rand Index
SCT-NCUT	Spatio-Color-Texture NCUT
SLIC	Simple Linear Iterative Clustering
SLICT-NCUT	SLIC-TextonNCUT
SVM	Support Vector Machine
VI	Variation of Information
WED	Weighted Euclidean Distance

Publication Partner:
International Journal of Scientific and Research Publications (ISSN: 2250-3153)