Enhancing Feature Extraction from Histopathology Images of Colorectal Cancer through Comparative Analysis of Pre-processing Filters

N. Muhilarasi[1], Dr.P.Indra[2], Prof.R.Yognapriya[3]

[1]*PG Scholar., Communication Systems, Department of Electronics and Communication Engineering., Government College of Engineering Salem ,*

[2]*Assistant Professor., Department of Electronics and Communication Engineering., Government College of Engineering Salem ,*

[3]*Assistant Professor., Department of Electronics and Communication Engineering., Government College of Engineering Salem,*

---***---

Abstract - *Colorectal cancer is the second leading cause of cancer-related deaths worldwide, driven by multiple pathophysiological mechanisms such as abnormal cell proliferation, differentiation, resistance to apoptosis, local invasion by tumor cells, and distant metastasis. This malignancy poses a significant global health challenge, with early detection being paramount to improving patient outcomes. Histopathology images are the primary data source for pathologists, aiding in cancer diagnosis through detailed examination of tissue architecture, cell morphology, and abnormal growth patterns. However, these images often contain artifacts and distortions due to various factors, complicating the extraction of relevant features from colorectal cancer images. In this proposed work, filters are employed for pre-processing, comparing three different types to determine the best method based on Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) values.*

keywords: Histopathology image analysis, Progressive Switching Median filter, Wiener filter, Gaussian filter.

1.INTRODUCTION

Colorectal cancer (CRC) is the fourth most common cancer, accounting for 10% of all new cancer cases, and the fifth leading cause of cancer-related deaths, resulting in approximately 550,000 deaths annually worldwide. While 25%-50% of CRC patients are diagnosed at an early stage but later experience recurrence or metastasis, Approximately 25% of colorectal cancer cases are identified at an advanced stage, a condition that is linked to a notably poor prognosis, reflected in a disheartening five-year overall survival rate of merely 14%.. The development of CRC begins with changes in the normal colonic epithelium, including the formation of adenomatous polyps that can grow and multiply over time, accumulating genetic and epigenetic mutations. While not all malignant polyps progress to invasive cancer, those that do can spread to neighboring tissues, particularly the intestinal wall, and eventually metastasize to distant sites via the lymphatic and circulatory systems.

CRC can be analyzed using various methods, including endoscopic, laparoscopic, MRI, CT, and others. Histopathological images (HIs) are based on the morphological and architectural characteristics defined by pathologists and exhibit distinct biological structures. HIs display a high degree of visual diversity in the connected patterns of specific microscopic structures throughout the tissue area. Numerous computerized techniques have been developed to process digitized histology images across the three main computer vision tasks: segmentation, feature extraction, and classification. All these tasks use an image as input. Segmentation extends beyond object detection by identifying specific object pixels rather than using a coarse bounding box. Feature extraction, part of the dimensionality reduction process, involves dividing an initial set of raw data into more manageable groups to simplify processing. Image classification predicts the class or type of an object within the input image, providing a class label as output.

In this proposed work, dataset images undergo preprocessing to achieve high-precision images, enabling pathologists to diagnose patients with high accuracy, distinguishing between malignant and normal tissues.

2. Efficient Preprocessing Filters and Mass Segmentation Techniques for Mammogram Images

Authors: Jayesh George M, Perumal Sankar S Affiliation: Department of ECE, Vimal Jyothi Engineering College, Kannur, Kerala .Preprocessing is a crucial step in image processing techniques, aimed at enhancing image quality by reducing unwanted distortions or highlighting specific features essential for further analysis. Mammogram images, in particular, present unique challenges in interpretation, making effective preprocessing especially important. In their study, the authors evaluated various preprocessing filters for mammogram images. The median filter, when applied to noisy images, proved effective in eliminating salt-and-pepper noise while preserving the integrity of the underlying image, as demonstrated in Figure 3. However, it was less effective in addressing Gaussian noise. Conversely, the adaptive median filter was found to be highly effective for noise separation without blurring edges, making it particularly suitable for mammogram image enhancement. This filter performs better against salt-and-pepper noise compared to speckle and Gaussian noise. Additionally, the study assessed the performance of the Wiener filter on noisy images, yielding specific results discussed in the Proceedings of the 2017 IEEE International Conference on Circuits and Systems (ICCS2017).

The Effect of preprocessing filters on predictive performance in radiomics

Aydin Demircioğlu, All scans, including segmentations, were resampled to a uniform voxel size of 1 mm^3 using spline interpolation. MRI series were normalized for intensity by centering all values to have a mean of 0 and a standard deviation of 1 before processing. In contrast, CT series were not normalized since HU values were consistent across scans. Before feature extraction, several preprocessing filters were applied to the scans, including exponential, gradient, Laplacian-of-Gaussian (LoG), local binary pattern (in both two and three dimensions), logarithm, square, square-root, and wavelet filters. This work was published in *European Radiology Experimental*, with the article received on April 3, 2022, accepted on June 30, 2022, and published online on September 1, 2022.

Correction preprocessing method for cardiac scintography images using local adaptive filters

Eltayeb Wagiallah, Yasir Ahmed, Yousif Mohamed Y. Abdallah International Journal of Science and Research (IJSR) .This study introduces a robust method for estimating noise statistics in cardiac scintigraphy images. To achieve optimal image quality post-denoising, a novel low-order, local adaptive Gaussian Scale Mixture model is proposed, which addresses nonlinearities resulting from scattering. This experimental study presents a new approach for managing independently and identically distributed (IID) noise in cardiac images and aims to reduce data redundancy through advanced image processing techniques. The paper, titled *Enhancing Feature Extraction from Histopathology Images of Colorectal Cancer through Comparative Analysis of Pre-processing Filters*, is published in Volume 3, Issue 9, September 2014. It is available at www.ijsr.net, Paper ID: SEP14660, under Creative Commons Attribution CC BY 2326. The study employs Matlab version R2009 to apply median and anisotropic filtering algorithms to ten different X-ray images, calculating their Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) values.

A New Preprocessing Filter for Digital Mammograms

Peyman Rahmati, Ghassan Hamarneh, Doron Nussbaum, Andy Adler. This paper presents a computer-aided approach designed to enhance suspicious lesions in digital mammograms. The developed algorithm improves upon the widely-used contrast-limited adaptive histogram equalization (CLAHE) preprocessor filter by introducing the fuzzy contrast-limited adaptive histogram equalization (FCLAHE) filter. FCLAHE provides non-linear enhancement, effectively removing noise and intensity inhomogeneities in the background while preserving the natural gray level variations of mammographic images within the regions of interest.

Evaluation of preprocessing methods on independent medical hyperspectral databases to improve analysis

Martinez-Vega, B.; Tkachenko, M . Sensors 2022, 22, 8917. Published: 18 November This study assessed various preprocessing methods to enhance tumor detection across three hyperspectral imaging (HSI) databases related to colorectal, esophagogastric, and brain cancers. The preprocessing steps evaluated included spatial and spectral smoothing, Min-Max scaling, Standard Normal Variate (SNV) normalization, and a median spatial smoothing technique. Calibrated data refer to datasets with minimal preprocessing, where only the removal of extreme spectral bands—due to noise from the HS sensors—and glare correction were performed to avoid including problematic pixels in classification. Filtered data, on the other hand, involved applying a Gaussian filter to reduce noise. Min-Max scaling and SNV normalization were used to achieve a more uniform spectrum. Additionally, the median filter was employed to standardize pixels with varying intensities.

3. Pre-Processing Techniques

The main objective of pre-processing technique is to improve or enhance the quality of image data that removes unwanted noises or to boost some image features that are important for further processing.

These all are done by using the following techniques,

I. Progressive switching median filter

II. wiener filter

III. Gaussian filter

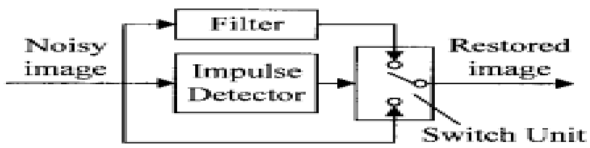
I. Progressive median switching filter:

The Progressive Switching Median (PSM) filter is a cutting-edge method for restoring severely damaged images without impulse noise. It functions as follows: Switching Scheme: To guarantee that just a portion of pixels—those recognised as carrying impulse noise—are filtered, an impulse detection technique is used before the filtering procedure.

Progressive Techniques: Using a series of iterations, both impulse detection and noise filtering are carried out gradually. By progressively enhancing the image and maintaining significant visual information, this method enhances the efficacy of noise reduction.

Progressive Techniques:

Iterative applications are used for both the noise filtering and impulse detection procedures. The noise-processed pixels in each iteration aid in the improvement of the next. Even for heavily damaged images, this progressive method guarantees improved restoration outcomes.

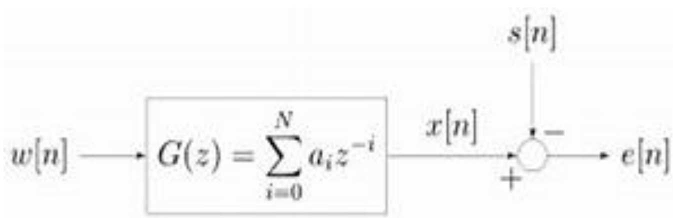


a) block diagram of switching median filter

$$\hat{f}(x, y) = \underset{(s,t)\in S_{xy}}{\text{median}}\{g(s, t)\}$$

II. Wiener filter:

The Wiener filter performs two main functions - it inverts the blur of the image and removes extra noise. This method works exceptionally well for photographs that have been blurred by a recognised low-pass filter or have experienced deterioration by a filter. It improves the overall clarity and quality of the image by successfully addressing problems caused by these distortions.



b) Block diagram of wiener filter

Wiener filtering is often performed in the frequency domain.

Steps:

- Transform the noisy image and noise to the frequency domain (e.g., using the Fourier transform).
- Compute the Wiener filter in the frequency domain:

  H(f)=SX(f)+SN(f)SX(f)

where:

- (H(f)) is the frequency response of the filter.
- (S_X(f)) and (S_N(f)) are the power spectral densities of the true image and noise, respectively.

- Inverse transform the filtered result back to obtain the restored image.

III. Gaussian filter:

- A weighted average is applied to the surrounding on the Gaussian distribution. It also functions as a realistic model to simulate the effects of a defocused lens and is especially good at eliminating Gaussian noise.
- A Gaussian kernel (or filter) defines the weights applied to neighboring pixels during the filtering process. The weights are progressively reduced as the distance from the central pixel rises, with the central pixel receiving the maximum weight.The essential parameter is the standard deviation (σ), which controls the filter's scope and smoothing effect.

Mathematical Formulation:

- The following is the impulse response of the one-dimensional Gaussian filter:

  h(t)=2πσ1e−2σ2t2

- The frequency response is the Fourier transform of this impulse response.
- In two dimensions (for images), it's the product of two such Gaussians:

$$g_{\sigma,\mu}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

DATASET AND SOFTWARE TOOL USED

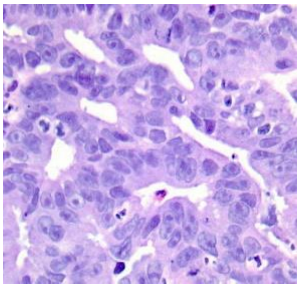
Table 1. Materials and Methods

Parameters	Value
Dataset	HISTOPATHOLOGY COLON CANCER DATASET
Dataset Number of images	5000
Tools used	MATLAB 2017a

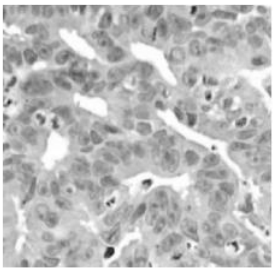
In the above-mentioned dataset, there are total of 5000 images out of which 2500 are benign tissue and another 2500 are malignant tissue images of the colon.

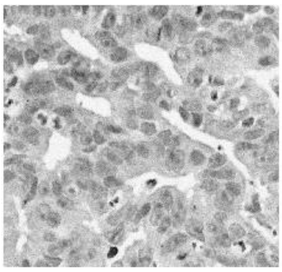
RESULTS AND DISCUSSIONS

The proposed methods were applied on different digital images and some results are given below. Each figure indicates the original image, output image of three filter. By comparison of these three pre-processing techniques, progressive switching median filter provides better results with low MSE, high PSNR and SSIM values.

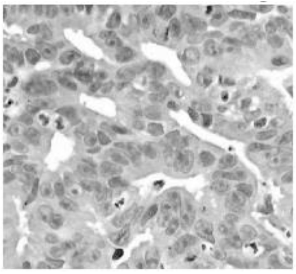


a) Original image





b) Gaussian filtered image **c) Wiener filtered image**



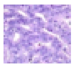
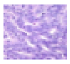
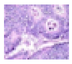
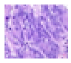
d) Median filtered image

Output images of Gaussian filter, Wiener filter, Progressive switching median filter.

COMPARATIVE PERFORMANCE OF FILTERS

To analyze the best filter for the preprocessing technique, it is important to note that the filtered image must have the least mean square error value, high Peak signal to noise ratio and high structural similarity index measure wants to be satisfied. It is know that the progressive switching median filter have satisfied all these performance criteria with the least mean square error, high structural similarity index measure and high peak signal to noise ratio.

Table 2: Comparison of MSE, PSNR and SSIM values for various images

Filters	Parameters					
Progressive switching median filter	MSE	2.3537	3.5821	5.3902	3.3221	10.9169
	PSNR	44.4133	42.5894	40.8147	42.9167	37.7498
	SSIM	0.9875	0.9827	0.9849	0.9832	0.9605
Wiener filter	MSE	386.7385	392.2611	368.2858	411.1434	400.2229
	PSNR	22.2566	22.1951	22.4690	21.9909	22.1078
	SSIM	0.5028	0.5156	0.5698	0.5078	0.5292
Gaussian filter	MSE	20.9423	31.8434	61.7409	29.6761	71.7995
	PSNR	34.9206	33.1006	30.2251	33.4067	29.5696
	SSIM	0.9240	0.8961	0.8742	0.8895	0.7949S

CONCLUSION AND FUTURE WORK

In the realm of medical imaging, various types of noise can be introduced into original images, which necessitates the use of different filtering techniques to remove these disturbances. This paper focuses on eliminating detected masses from the original images through the application of several preprocessing techniques. Among the nonlinear filters evaluated, the progressive switching median filter demonstrated superior performance, evidenced by its lower Mean Square Error (MSE), higher Peak Signal-to-Noise Ratio (PSNR), and better Structural Similarity Index Measure (SSIM) values.

Looking ahead, the further work is to segment and classify these images to accurately determine whether they contain cancerous tissue. The goal is to enhance early detection capabilities, thereby potentially reducing the mortality rate associated with cancer by improving the accuracy of predictions and facilitating earlier intervention.

REFERENCES

1) Dr. P. Indra, R.Yoganapriya, "Pre-Processing Techniques For Breast Cancer Detection In Mammography Images", International Journal of Novel Research and Development (IJNRD), ISSN: 2456-4184, pp: a79-a83, Volume 8, Issue 7 July 2023.

2) Dr. P. Indra, R.Yoganapriya, "Optimizing Melanoma Skin Cancer Analysis through Pre-Processing Strategies", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN:2455-6211, pp: 1498-1503.

3) Dr. P. Indra (2020), "Multilevel Tetro let transform based breast cancer classifier and diagnosis system for healthcare applications", in the proceedings of Journal of Ambient Intelligence and Humanized Computing ISSN 1868-5137, J Ambient In tell Human Compute. https://doi.org/10.1007/s12652-020-01755-z

4) Dr. P. Indra, (2014), "Classification of Micro Calcification in Digital Mammogram using Tetro let Transform", in the Proceedings of the International Journal of Innovative Research & Advanced Studies (IJIRS), Vol 3, Issue 8, pages:398-410, ISSN 2319-9728 on August 2014.

5) Dr. P. Indra, (2014), "Mammogram Classification based on conventional Harr Transform", in the Proceedings of the International Journal of Multidisciplinary Research and Advances in Engineering (IJMRAE), Vol 6 No IV, pages: 1-8, ISSN 0975-7074 on October 2014.

6) Dr. P. Indra, (2015), "Survey on Mammogram Image Classification CAD system" in the proceedings of Global Journal of Engineering and Applied Sciences (GJEAS) Vol 4, Issue 3, pages:54-56.

7) R. Yoganapriya , Prof .P. Indra., (2020). "Detection of masses in digital mammogram using nonlinear filters" ,International Research Journal of Engineering and Technology (IRJET) Vol 07, Issue 12.

8) R.Ramani, Dr. N . Suthanthira Vanitha, S. Valarmathy , "The Pre Processing Techniques for Breast Cancer Detection in Mammography Images", I.J. Image, Graphics and Signal Processing, 2013, 5, 47-54.

9) Nithya. K, Aruna.A, Anandakumar.H, Anuradha.B. "A Survey On Image Denoising Methodology On Mammogram Images" International journal of scientific & technology research volume 3,issue 11, november 2014 .

10) D. Sujitha Priya,Dr. B. Sarojini."Breast Cancer Detection InMammogram Images Using Region-Growing And Contour- Based Segmentation technique "International Journal of Computer & Organization Trends –Volume 3 Issue8 – Sep 2013.

11) Aydin demircioglu "the effect of preprocessing filters on predictive performance in radiomics" Demircioğlu European Radiology Experimental ,Received: 3 April 2022 Accepted: 30 June 2022 published online: 01 September 2022

12) Wagiallah 1,yasir ahmed 2, yousif mohamed y. abdallah3 , international journal of science and research (ijsr) "correction preprocessing method for cardiac scintography images using local adaptive filters"

13) Peyman rahmati1, Ghassan hamarneh2, Doron nussbaum3, and andy adler1 Stolz, W., et al. &Quot; Abcd rule of dermatoscopy-a new practical method for early recognition of malignant-melanoma.& quot; European Journal of Dermatology 4.7 (1994).

14) Ng, Vincent TY, Benny YM Fung, and Tim K. Lee & quot; Determining the asymmetry of skin lesion with fuzzy borders " Computers in biology and medicine 35.2 (2005).