A New and Efficient Approach to Cell Segmentation and Tumor Detection in Histopathological Images

Hanae Moussaoui
Applied Sciences and Emerging Technologies Laboratory (LSATE), ENSA of Fez, Sidi Mohamed Ben Abdellah University of Fez, Morocco
hanae.moussaoui1@usmba.ac.ma

Nabil El Akkad
Applied Sciences and Emerging Technologies Laboratory (LSATE), ENSA of Fez, Sidi Mohamed Ben Abdellah University of Fez, Morocco
nabil.elakkad@usmba.ac.ma

Mohamed Benslimane
Innovative Technologies Laboratory, EST of Fez, Sidi Mohamed Ben Abdellah University of Fez, Morocco
mohamed.benslimane@usmba.ac.ma

Walid El-Shafai
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia | Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt
welshafai@psu.edu.sa

Layal Kazma
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
lkazma@psu.edu.sa (corresponding author)

Ahmad Taher Azar
College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia | Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia
aazar@psu.edu.sa

Received: 17 November 2025 | Revised: 28 December 2025 | Accepted: 3 January 2026

 DOI: https://doi.org/10.48084/etasr.16320

ABSTRACT

Detecting and segmenting cells or tumors in histopathological images is a very challenging task. This study presents a novel technique for detecting tumors and segmenting cells in histopathological images. The EBHI-SEG dataset contains 5170 images of six types of tumor differentiation stages and the corresponding ground truth images. Normalization was applied using a specific target image. Then, the Gaussian blur technique was applied to reduce the noise in the original image. The Otsu's thresholding method was applied to obtain binary images, followed by morphological operations. The results obtained were evaluated using the Jaccard index, Intersection Over Union (IOU), Precision, Recall, F1-score, and Structural Similarity Index (SSIM), showing satisfactory results. A comparison with six other well-known methods showed that the proposed approach provides promising and sufficient results.

Keywords-histopathological images; histology; image segmentation; tumor detection; biomedical image analysis; Otsu thresholding; Gaussian filter; morphological operations

I. INTRODUCTION

After the invention of the microscope in the late 16th century, histopathology became an important field of research. However, it was not until the 19th century that technological developments allowed for a thorough examination of tissues. Clearer observation of cellular components and structures was made possible by innovations such as the tissue staining techniques developed by Paul Ehrlich and others. Histopathological analysis is now more accurate and efficient due to advances in automated systems and digital imaging over time. Recent developments in the field of medical science to improve the accuracy and utility of histopathological images include the introduction of computerized pathology and machine learning image analysis. The diagnosis of diseases, especially malignancies, is the most important use of histological imaging [1]. Pathologists can find abnormalities that point to disease by using a microscope to study the morphology of cells and the architecture of tissues.

Histopathological studies help identify cancer types, distinguishing benign from malignant tumors, and provide details on the prognosis and possible treatment options. Understanding both malignant and normal biological changes requires careful examination of the microscopic structure of tissues. This component is crucial for medical research and teaching, as it sheds light on how diseases affect bodily tissues and cause symptoms and problems seen in patients. Moreover, the severity and scope of a disease can be determined through histopathological analysis of tissue samples, which is essential for therapy planning. For instance, in the treatment of cancer, the histological grade and stage of the tumor can influence whether surgery, chemotherapy, radiation therapy, or a combination of these treatments is best for the patient. Eventually, histopathological techniques are essential for biomedical research to understand the causes of the disease, evaluate the effects of experimental treatments, and develop new medical remedies and diagnostic tools.

The literature has extensively chronicled the various problems the field of histopathology encounters despite its progress. The subjectivity involved in image interpretation is a significant gap that can cause variation in pathologists' diagnoses. In addition, manual histopathology image interpretation is labor- and time-intensive, slowing down throughput in busy diagnostic environments. Furthermore, some uncommon or overlapping histological characteristics can make a diagnosis more difficult, highlighting the need for more exact and consistent diagnostic standards. A crucial stage in the analysis of histology data is histopathological image segmentation, which involves splitting a digital image into several segments (groups of pixels) to make its representation simpler, more meaningful, and easier to examine. The act of identifying and quantifying diverse regions within a tissue sample is crucial, with examples of these regions including areas of necrosis, tumor boundaries, and distinctions between different cell types. Histopathological image segmentation is achieved using a variety of methods, from conventional image processing techniques to sophisticated deep learning and machine learning methods. This study presents an effective technique for histopathological image segmentation that uses a hybridization of some image segmentation techniques and shows promising results for various types of cells.

II. RELATED WORKS

Machine learning and deep learning advances have greatly benefited the evolution of histopathology image segmentation from manual and semi-automatic procedures to more complex automated systems. As they advance, these technologies raise the bar for what is possible in the field of digital pathology. In [2], an innovative weakly supervised system was proposed for segmenting images in digital histopathology (WESUP), identifying complicated semantics and taking advantage of hierarchical information by utilizing over-segmentation and transfer learning. WESUP is an end-to-end learning architecture that involves superpixel label generation, hierarchical feature representation, and deep dynamic label propagation. For fine-grained prediction, pixel-wise hierarchical feature representation and classification are used during the inference stage.

In [3], a new Hierarchical Conditional Random Field (HCRF)-based Gastric Histopathology Image Segmentation (GHIS) method was proposed to localize abnormal regions (cancer) in gastric histopathological images obtained by an optical microscope. In the test set, this technique achieved segmentation accuracy, recall, and specificity of 78.91%, 65.59%, and 81.33%, respectively. In [4], a feature Pyramid-based image semantic segmentation technique was proposed (ResNet50-GICN-GPP). First, the image was resampled using the patch sampling approach, which also reduced the size of a single sample and increased the total number of training samples. Second, the ResNet50 learning feature placement information was used to create the entire convolution network. Multi-level features were combined using the GICN structure and deconvolution network. Finally, the GPP structure was coupled to investigate the multi-scale semantic information to address the issue that the GICN structure can lose small items. This approach achieved an average segmentation accuracy of 63%.

In [5], cellular structures in histopathological images were segmented using a Convolutional Neural Network (CNN) in conjunction with the simple linear iterative clustering superpixel segmentation method. This approach consisted of two primary steps. Cellular and non-cellular superpixel segmentation was first accomplished using the SLIC superpixel approach as a pre-segmentation algorithm. The final segmentation of the entire image was then obtained by classifying those superpixels using a deep learning approach based on CNNs. This system achieved an overall classification accuracy of 0.9876 for the superpixels. In [6], a U-net model, based on MobileNetV2, was used to segment the nuclei regions from histopathological images of Triple Negative Breast Cancer (TNBC). This approach achieved a Jaccard index of 59% on the TNBC nuclei segmentation dataset.

In [7], a histopathological image segmentation technique was based on object-shaped kernels that allow the identification and analysis of cell nuclei regions. The multi-stage process of this technique involved image contrast augmentation, nucleus region extraction, nucleus centroid labeling, nuclei area refining, and complicated nuclei separation. Three common datasets of histopathology images, stained with Hematoxylin and Eosin (H&E), and a newly developed dataset were used to gauge the effectiveness of this approach, using Hausdorff distance, Jaccard index, and F1 score. In [8], a multi-task learning strategy used information from several independent data sources to segment and classify nuclei, glands, lumina, and other tissue areas.

In [9], the UNet and SegNet models were combined to develop a pixel-wise segmentation technique for colon cancer WSIs. A multi-step training technique was presented for the sparse annotation of histopathology images. UNet and SegNet achieved up to 76.18% and 81.22% accuracy when deployed and tested in various training settings, such as data augmentation and transfer learning. Additionally, this study used the CRC-5000, nct-crc-he-100k, and Warwick datasets to test the models and training techniques, where SegNet attained corresponding accuracies of 98.66%, 99.12%, and 78.39%. In [10], a revolutionary end-to-end deep learning architecture was proposed to address shape-variability and nuclei-touching issues in the segmentation of nuclei in histopathology images. Unlike UNet, the suggested NucleiSegNet leveraged salient information and extracted high-level spatial information by using a modified attention mechanism and a recently proposed robust residual block, respectively. The study in [11] addressed the issue of background and foreground segmentation in histopathological images of H&E-stained skin specimens. The proposed method was compared with previous methods, GrabCut and FESI, demonstrating its superiority with a Jaccard index of 0.929 vs. 0.776 and 0.695.

Despite great progress in histopathological image segmentation, there are still several unanswered questions and problems that need to be resolved. These gaps are frequently caused by the special difficulties associated with medical imaging data, the significant risks associated with medical diagnosis, and the rapid pace with which technology is developing [12]. The lack of huge and well-annotated datasets is one of the main obstacles. Medical image annotation is costly, time-consuming, and requires specialized knowledge. Moreover, histopathological images can differ significantly depending on the scanner and the staining procedures employed. When applied to different image modalities, models trained on data from one image modality frequently perform poorly on another. Therefore, there is a need for robust models that can generalize to many kinds of staining techniques and imaging apparatuses without requiring a lot of retraining. However, significant computational obstacles arise when dealing with and analyzing high-quality WSIs, which can have gigabit sizes. It is necessary to develop effective algorithms that are capable of processing these large images quickly without sacrificing accuracy [13], especially for use in clinical settings when time is of the essence. Furthermore, artifacts such as folds, rips, dust, and uneven staining are frequently seen in histopathological images. Reliable analyses require robust algorithms capable of robustly handling images with faults of this kind or automatically detecting and correcting these artifacts. This study discusses how to overcome these challenges by applying some wisely chosen image-processing techniques.

III. THE PROPOSED METHOD

In medical image analysis, histopathological image segmentation is an intricate and vital task that forms the basis for the diagnosis, investigation, and management of diseases such as cancer. Despite technological progress, the field continues to face many important obstacles. The proposed method is a hybridization of many image processing techniques [13] that provides promising results. The reasons for choosing image processing techniques over deep or machine learning methods are several, starting with the fact that pathologists can more easily notice microscopic characteristics that are essential for a correct diagnosis when using image processing techniques to improve the clarity and detail of histopathological images. For precise tissue analysis and disease classification, methods that include thresholding, edge detection, and morphological procedures aid in the delineation of certain tissue structures, such as cell borders, nuclei, and connective tissues. In addition, compared to deep learning models, traditional image processing methods frequently use simple algorithms that can be implemented with minimal programming complexity. These techniques may be more manageable and accessible for organizations or academics who lack substantial machine-learning experience. Additionally, traditional approaches are appropriate for environments with restricted IT infrastructure, since they often demand less processing power instead of high-end GPU resources. Furthermore, when real-time or near-real-time analysis is needed, conventional image processing may be faster than training and executing deep learning models. Eventually, compared to machine learning techniques [14], traditional image processing techniques are less likely to overfit, particularly when there is a lack of training data. This suggests that they may generalize more effectively when presented with fresh, unviewed images that are consistent with the training set. Since the main objective of this research is to help make histopathological image segmentation easier for pathologists, traditional approaches are frequently easier to integrate into current digital pathology systems without requiring significant overhauls or disruptions.

The first step of the proposed method is preparing the input image using normalization to enhance its quality. Then, a Gaussian filter is used to minimize noise and enhance image details. Subsequently, Otsu's thresholding and morphological operations are applied to obtain the final segmentation results.

A. Dataset

This study used the EBHI-SEG dataset [15], which contains 5170 images of six types of tumors' differentiation stages along with their corresponding ground truth images of 224×224 pixels in PNG format. This dataset has six types of images: polyp, normal, low-grade intraepithelial neoplasia, high-grade intraepithelial neoplasia, adenocarcinoma, and serrated adenoma. Table I displays some images from the EBHI-SEG dataset.

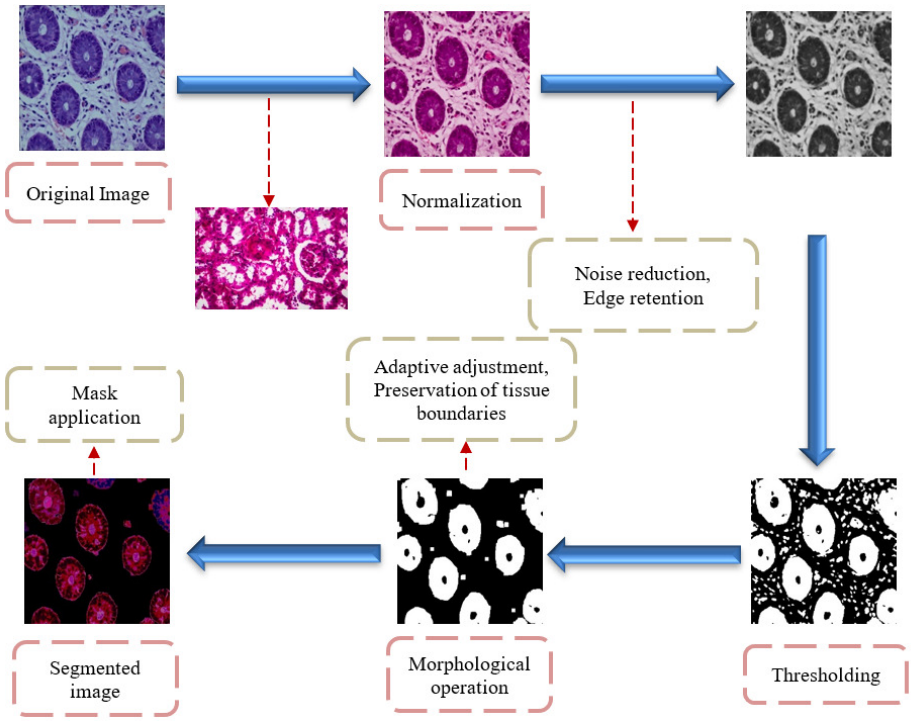
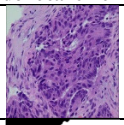
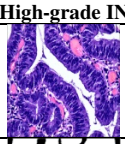
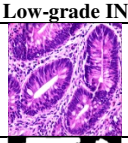
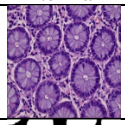
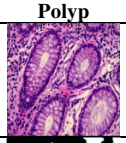
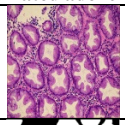
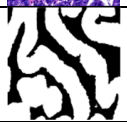
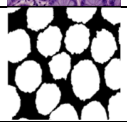
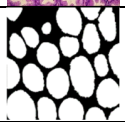


Fig. 1. The proposed method.

TABLE I. THE EBHI-SEG DATASET.

	Adenocarcinoma	High-grade IN	Low-grade IN	Normal	Polyp	Serrated adenoma
Original image						
Ground truth						

B. Normalization

Using normalization techniques, images are adjusted to improve consistency in color representation across samples. Maintaining consistency is essential to guarantee the accuracy and dependability of subsequent image analysis processes, such as segmentation or classification, when applied to various data sets. On the other hand, normalization [16] contributes to the clarity and ease of analysis of the tissue patterns and structures.

Image processing algorithms can identify contours, edges, and other morphological elements required for precise segmentation more efficiently by standardizing color and intensity. Similarly, staining artifacts and low contrast can make features appear blurred or indistinct, but normalization can enhance them. By improving these characteristics, segmentation algorithms can more easily and accurately divide the image into meaningful segments that correspond to various types of tissue or disease states. Another important step in the normalization process is the selection of a target image. In histopathological image analysis, selecting a target image for normalization is an important task that should not be performed at random [17]. To ensure uniformity in intensity and color among various samples, specific target image attributes are defined. When choosing a target image, several factors must be considered to ensure that the image is representative of the dataset and best suited to deliver precise results in later studies, such as segmentation. Among these characteristics is the target image quality, which needs to be good and free of artifacts such as dust, scratches, uneven staining, or insufficient focus. For this research, and after many trials, the image in Figure 2 was selected as a target image.

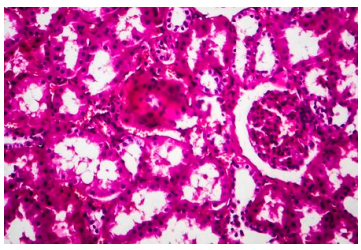


Fig. 2. Target image used in the normalization process.

This study used the Reinhard normalization [18], also known as Reinhard color normalization. The foundation of Reinhart normalization is readjusting an image's color channel average and standard deviation to be equivalent to those of a reference image. This method assumes that color changes in histopathological images are mostly caused by variations in stain concentrations and lighting, rather than by biologically significant variances. This approach seeks to standardize these color metrics to a reference/target image to minimize these non-biological variances across a series of images.

For every color channel [19], the following steps are applied to mathematically represent the Reinhard normalization technique:

- Step 1: The RGB color space of the images is first transformed into the Lab color space, which separates the lightness (L) and color components (a and b), which translate into blue-yellow and green-red color dimensions.
- Step 2: Compute the standard deviation and the mean for the source image as well as the image used as a reference for each channel (L, a, and b) in the Lab color space.
- Step 3: The normalization process involves taking the mean of each channel in the source image and scaling it using the images used as references. Then, the reference image's mean is finally added. Each pixel's formula for every channel can be expressed as follows:

$$L' = \left(\frac{\sigma_{r,L}}{\sigma_{s,L}}\right) \times (L_s - \mu_{s,L}) + \mu_{r,L} \quad (1)$$

$$a' = \left(\frac{\sigma_{r,a}}{\sigma_{s,a}}\right) \times (a_s - \mu_{s,a}) + \mu_{r,a} \quad (2)$$

$$b' = \left(\frac{\sigma_{r,b}}{\sigma_{s,b}}\right) \times (b_s - \mu_{s,b}) + \mu_{r,b} \quad (3)$$

where L_s, a_s, and b_s are the source image's pixel values in the Lab color space, μ_s, and σ_s are the original image's mean and standard deviation, and μ_r, and σ_r are the reference image's mean and standard deviation.

- Step 4: The Lab color space image is normalized and then transformed to its original RGB color space for further employment or display.

This technique is especially significant for histopathological image analysis [18], as it produces more consistent images that pathologists can visually evaluate and computer systems can automatically analyze. The consistency of color representation improves the accuracy and certainty of examinations and research findings by making diagnostic traits related to the morphology of tissues and structures identical across different slides.

C. Gaussian Filter

This is one of the most important tools in image processing to smooth local intensities and suppress noise, particularly when it comes to image segmentation. A Gaussian function is used for smoothing to reduce noise and improve the shape of elements within images before segmentation. Using the Gaussian filter effectively requires an understanding of its workings and how it is applied in image segmentation. A kernel, also called a mask, illustrates a Gaussian function, characterized by a bell-shaped curve, and is applied by convolution involving an image and a set of kernels or filters. The convolution has the main impact of smoothing, which aids in lowering the level of detail and noise in images. Moreover, the Gaussian function is especially useful in this situation because of its basic qualities [20]:

- Due to the nature of the Gaussian function, pixels nearer the kernel's center are given more weight in the average. Compared to other averaging filter types, such as the box filter, this produces an effect of smoothing that is more natural and maintains the edges better.
- The Gaussian function's standard deviation (σ) governs how widely it spreads. While a smaller σ concentrates on a smaller region, conserving more detail, a greater σ increases the effect of smoothing by adding a wider region of the image with an average process for each pixel.

The 2D Gaussian formula defines the Gaussian filter mathematically:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4)$$

where x and y indicate the vertical and horizontal distances, respectively, from the central pixel, and σ is the standard deviation, which establishes the blur's spread. Therefore, for an accurate and trustworthy assessment of microscopic images of tissues, Gaussian filters are essential in the image segmentation process of histopathological images [20].

There are several reasons for employing the Gaussian filter. The first is that during the imaging process, noise of many types might be introduced into histopathological images. Hence, Gaussian filters improve the smoothing process of these images by reducing the effect of noise and preventing tissue components from being incorrectly classified during segmentation. Second, the ability of Gaussian filtering to maintain edges while smoothing an image makes it very valuable. The boundaries of tissue compartments and cellular structures are essential for precise segmentation and diagnosis in histopathology. Therefore, preserving these borders guarantees that crucial biological boundaries are preserved during the preprocessing stage. In addition, the Gaussian filter improves the quality of the whole image, facilitating the identification and segmentation of various tissue features by automated systems and pathologists. Overall, smoother images facilitate more consistent and dependable segmentation results.

In this study, the size of the Gaussian kernel utilized for the blur was defined using the tuple (5, 5), indicating the kernel's dimensions: 5 pixels wide by 5 pixels high. This choice was made because it can sufficiently reduce the noise while preserving the features in the data. Since the amount of image blurring is determined by the size of the kernel [21], the blur effect will be more pronounced with a larger kernel. To guarantee that there is a legitimate central pixel, the kernel size needs to be a positive odd integer. In addition, the provided σ_x (standard deviation in the x direction) is 0. Since the amount of blurring is determined by σ, a higher σ causes the pixels to spread out more widely, which increases the amount of

blurring. OpenCV uses the formula $(ksize[i] - 1) * 0.5/2 + 0.5$ to obtain the average σ using the kernel size, where $ksize[i]$ is the kernel's width or height. Setting it to 0 makes the function dynamically adjust based on the supplied kernel size by automating the process of calculating the standard deviation.

D. Otsu's Thresholding

The smoothed image is then used to produce a binary segmentation. The conversion of a gray-level image to a binary representation, or clustering-based image thresholding, can be performed automatically with the help of the well-liked and effective Otsu's thresholding technique [22]. This approach determines the ideal threshold separating the two classes so that their total spread (intra-class variance) is as small as possible after assuming that the image consists of two classes of pixels (background and foreground) that follow a bimodal histogram. Otsu's thresholding involves certain steps [23]:

- Step 1: First, the pixel intensity values' histogram is calculated using Otsu's approach. Every intensity level's probability is calculated using this histogram.
- Step 2: The algorithm initializes the probability of both classes (background and foreground) and computes the histogram's overall mean level.
- Step 3: With Otsu's approach, every potential threshold that could separate the image into the background and foreground is tested iteratively. The pixels are separated into those above (foreground) and below (background) for each threshold t.
- Step 4: Compute the Variance, Means, and Weights:
  - Weights (w) represent the probability of the two classes divided by the threshold t. The sum of the likelihood of each bin in the histogram above and below t is used to compute them.
  - Means (μ) determine each class's mean levels.
  - Calculate the variance (σ^2) for every class.
- Step 5: The within-class variance, the weighted sum of the variances for these two classes, is computed by the algorithm. This amount is calculated for every potential threshold.
- Step 6: Otsu's method aims to determine the threshold value at which the within-class variation is reduced. It can be expressed mathematically as maximizing the between-class variance, which is determined by subtracting the within-class variance from the overall variance.
- Step 7: The ideal threshold is determined by calculating the least within-class variance (or, conversely, the highest between-class variation) that it produces. The mathematical expression of the within-class variance is as follows:

$$\sigma_{\omega}^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (5)$$

where σ_0^2 and σ_1^2 are the variances of these two classes, σ_{ω}^2 is the within-class variance at threshold t, and ω_0 and ω_1 are the weights for each class.

Since Otsu's thresholding technique is so effective and reliable at automatically determining the best threshold for binarization, it is a great tool for segmenting histopathological images. This is especially helpful in medical imaging, in which analysis and diagnosis depend heavily on accurately segmenting characteristics such as cells, tissues, and diseased areas. Due to its ease of use, effectiveness, and reliable automatic threshold determination, Otsu's thresholding is still a popular and useful method in the field of histology. As such, it is a useful tool in the toolkit for digital pathology.

E. Morphological Operations

After applying Otsu's thresholding, the presence of artifacts, gaps, and some unrefined boundaries was noticed. Thereby, a set of image processing methods that manipulate images according to their forms, known as morphological operations, is needed. By applying a structural element to an input image, these processes produce an identically sized output image. During a morphological operation, every pixel in the image is altered based on the values of nearby pixels. Different effects can be obtained by selecting the structural element correctly. These methods are very helpful for the preprocessing stages of image analysis, such as feature extraction, image improvement, shape refinement, and noise reduction, especially for binary images [24]. There are four morphological operations.

1) Erosion

Pixels on object borders are lost due to erosion. Erosion's main impact is reducing the dimension of foreground objects by eroding the borders of foreground pixel regions, usually white pixels. The minimal value of every pixel surrounding the input pixel determines the value of the resulting pixel. In a binary image, the output pixel is zero if any of the neighboring pixels are also zero.

$$(f \ominus B)(x,y) = min_{(s,t)\in B} f(x-s, y-t) \quad (6)$$

2) Dilation

Dilation enlarges an object's boundaries by adding pixels. It causes items to enlarge in the opposite way that erosion does. The highest value of every pixel in the vicinity of the input pixel makes up the value of the final pixel. In a binary image, the resultant pixel is 1 if any nearby pixel is 1.

$$(f \oplus B)(x,y) = min_{(s,t)\in B} f(x+s, y+t) \quad (7)$$

3) Opening

This is a dilation that comes after erosion to keep the size and form of larger objects in the image while removing little ones (often tiny white dots) from the foreground. To restore the dimension of the larger objects, the image is dilated after it was first degraded to remove small objects.

$$(f \circ B) = (f \ominus B) \oplus B \quad (8)$$

4) Closing

Closing is an erosion that comes after a dilatation, employed on objects to fill up tiny holes and dark spots while preserving their size and form.

$$(f \bullet B) = (f \oplus B) \ominus B \quad (9)$$

In the above equations, f is the image, and B represents the structuring element. The structuring element B can have several sizes and forms, such as a disk, square, or rectangle. The decision is based on the application's particular needs, including the size and form of the characteristics that the operation intends to amplify or reduce. In the domain of histopathology image analysis, morphological operations are essential tools, especially for the segmentation of microscopic tissue images. These operations contribute to the segmentation output's refinement, noise reduction, and precise delineation of relevant features, including cells and intracellular components.

This study employed the opening technique, which can eliminate small-scale noise while maintaining the size and shape of bigger image structures. A 5×5 square structuring element was applied using two iterations to remove the isolated small artifacts while keeping the tissue morphology unaffected. For instance, this can remove tiny, white-noise spots that are not connected to the tissue structure.

IV. EXPERIMENTS, RESULTS, AND DISCUSSION

The efficacy and accuracy of segmentation techniques can be evaluated using several measures. This study employed the Jaccard index, Precision, Recall, F1-score, and the Structural Similarity Index (SSIM), which all showed satisfactory and very promising results.

A. Performance Metrics

1) Jaccard Index

The Jaccard index calculates how diverse and similar sample sets are. It can be expressed as the intersection's size divided by the sample sets' union's size. This metric assesses how much the ground truth labels or annotations and the segmentation results overlap. The mathematical formula of the Jaccard index can be written as follows:

$$J(A,B) = \left|\frac{A \cap B}{A \cup B}\right| \quad (10)$$

where A and B are two sets, $|A \cap B|$ represents the number of pixels present in the two sets (in this case, images), and $|A \cup B|$ represents the overall number of distinct pixels found in both images.

2) Precision

Precision is a crucial parameter that is frequently employed in conjunction with recall, also referred to as sensitivity, and accuracy. It is especially crucial when the cost of a false positive is high. Precision evaluates how well a segmentation or a classification algorithm predicts favorable results. It is the percentage of accurate identifications, and is also known as the value of positive predictions. The mathematical formula for precision can be written as follows:

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

where True Positives (TP) is the number of accurate positive estimates, and False Positives (FP) is the quantity of false positive estimates, which are the cases where a negative class is mistakenly classified as positive by the model.

3) Recall

Recall is a crucial evaluation parameter for histopathology image segmentation, sometimes referred to as sensitivity or true positive rate. Recall is a metric used in medical imaging, specifically in histopathology, to evaluate the accuracy of a segmentation model in identifying all pertinent diseased or tissue-specific regions in microscopic tissue images. The mathematical formula of the Recall can be written as follows:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

where False Negatives (FN) indicate the number of pixels or regions that belong to the tissue type or disease class but were missed by the model.

4) F1-Score

F1-score is an important evaluation metric in many fields, particularly medical imaging and, more especially, image segmentation in histopathology. It stands for the harmonic mean of recall and precision. The F1-score is beneficial in histopathology because it offers a metric that strikes a balance between segmentation precision and recall, both essential for ensuring accurate and thorough examination of medical images. The mathematical formula for F1-score can be written as follows:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

5) Structural Similarity Index (SSIM)

SSIM is an advanced metric for comparing the similarity of two images. Although standard metrics such as Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) compute absolute errors, SSIM takes into account changes in brightness, contrast, and structural information [25], which are more in line with how humans see the world. Thus, SSIM is especially helpful in tasks such as image segmentation, especially in histopathology, where precise diagnosis depends on an image's quality and structural integrity.

For SSIM, a pair of windows, x and y, of the same size $N \times N$, which are usually taken from the test image and the corresponding reference image, is defined. The formula for SSIM is as follows:

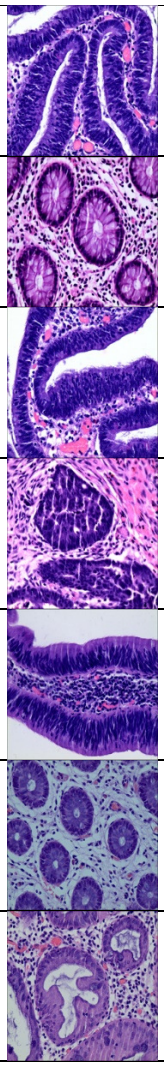
$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (14)$$

where μ_x and μ_y represent the average pixel values of both images x and y, σ_x^2 and σ_y^2 are x and y variances, σ_{xy} represents the x and y covariance, and c_1 and c_2 are two factors that stabilize the weak denominator division, where $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$, with L being the pixel-values' dynamic range, $k_1 = 0.01$ and $k_2 = 0.03$.

B. Results and Discussion

To ensure that the proposed method performs well on all histopathology images, it was applied to all seven types existing in the dataset: Adenocarcinoma, High-grade IN, Low-grade IN, Normal, Polyp, and Serrated adenoma. Table II displays the results obtained by applying the proposed method.

TABLE II. THE OBTAINED SEGMENTATION RESULTS.

Original	Ground Truth	Normalization	Gaussian filter	Otsu's thresholding	Morphological operations	Segmented image
						

The proposed method achieved good segmentation results by applying the normalization process using a specific target image (Figure 2). Normalization techniques are crucial to improve image consistency, contrast, and visibility, mitigate illumination variations, reduce inter-patient variability, facilitate collaboration, and prepare data for deep learning models, making them essential preprocessing phases in histopathology image segmentation workflows. These methods help produce segmentation outcomes that are more trustworthy and accurate, helping in disease diagnosis and therapy. Reinhard color normalization has many advantages that justify its selection, as it preserves tissue information, improves generalization, mitigates batch effects, makes multicenter studies easier, and helps with preprocessing for deep learning models. This method normalizes color distributions across images, increasing the consistency and reliability of segmentation algorithms, and advancing our understanding of disease pathophysiology.

Afterward, in histopathological image segmentation, employing a Gaussian filter following normalization has several advantages, including noise reduction, edge retention, increased image smoothness, enhanced segmentation efficiency, robustness to perturbations, and preparation for additional analysis. Gaussian filtering advances our knowledge of disease pathology by improving image quality and optimizing images for segmentation algorithms, leading to more accurate and consistent segmentation results. Following normalization and Gaussian filtering, Otsu's thresholding provides several advantages, including global thresholding, adaptive segmentation, computational efficiency, robustness to noise, handling intensity variations, and automatic threshold selection. Histopathology researchers can improve the accuracy and reliability of segmentation results, enhance their understanding of disease pathology, and improve diagnostic accuracy in clinical practice by combining these preprocessing and segmentation techniques.

In addition, histopathology image segmentation benefits from implementing morphological opening at the end of the preprocessing pipeline in several ways, including adaptive adjustment, artifact reduction, noise elimination, preservation of tissue boundaries, smoothing connectivity improvement, and noise reduction. Researchers studying histopathology can obtain more reliable and accurate segmentation findings by including a morphological opening in the segmentation method. This makes it easier to analyze and understand tissue structures for diagnostic and investigative purposes. Finally, at the end of the proposed technique, a mask was created to store the segmented cells with the same colors as the original image for better visuality.

The choice of the five metrics, namely the Jaccard Index, Precision, Recall, F1-score, and SSIM, to evaluate the proposed method was not arbitrary. The Jaccard index was utilized to calculate the spatial overlap between the predicted and ground-truth tumor regions. However, precision and recall determine false positive and false negative errors, respectively, which are of significant importance in tumor region detection problems. Additionally, the F1-score provides an equal weight for both precision and recall. SSIM was used to calculate the structural similarity of the predicted tumor region with the ground-truth region regarding the preservation of tissue structure. These metrics show that the proposed method achieved satisfactory results, as shown in Table III and Figure 3.

TABLE III. NUMERICAL RESULTS

Images of Table II	Jaccard Index	Precision	Recall	F1-score	SSIM
Image 1	87.07%	97.12%	89.38%	93.09%	60.76%
Image 2	58.88%	87.5%	64.29%	74.12%	54%
Image 3	89.64%	93.47%	95.62%	94.53%	76.48%
Image 4	78.61%	93.98%	82.77%	88.02%	67.10%
Image 5	70%	73.78%	92.31%	82.01%	54.69%
Image 6	85.14%	90.64%	93.34%	91.97%	77.80%
Image 7	63.54%	88.10%	69.51%	77.71%	50%

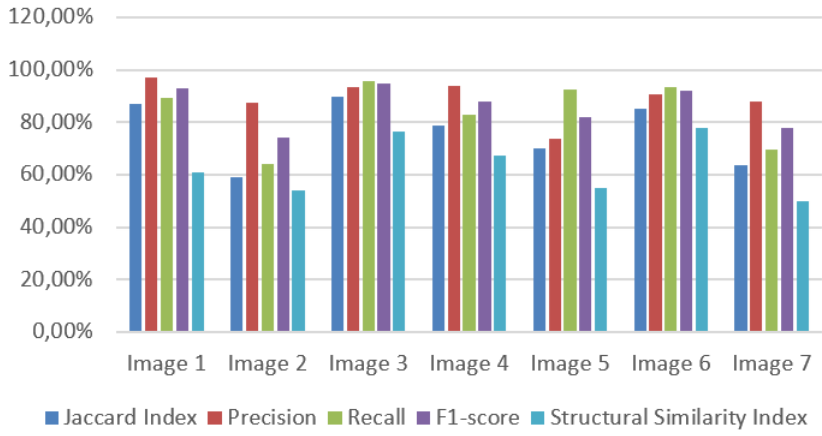


Fig. 3. Graphical representation of the results obtained.

TABLE IV. COMPARISON WITH OTHER METHODS.

Methods	Dataset	Jaccard index	Precision	Recall	F1-score
Proposed method	EBHI-SEG dataset	89.64%	97.12%	93.34%	94.53%
[26]	Colorectal histopathological private images (biopsies) from 351 biopsy specimens at Yeouido St. Mary's Hospital	69%	-	-	-
[27]	TCGA (The Cancer Genome Atlas)	-	75%	75%	-
[28]	GlaS and MoNuSeg	-	-	-	75.08%
U-Net [29]	EBHI-SEG dataset	88.6%	89.9%	98.3%	93%
Seg-Net [29]	EBHI-SEG dataset	88.6%	91.6%	97.7%	93.9%
MedT [29]	EBHI-SEG dataset	80.8%	89.6%	95%	91.9%

Table IV shows a comparison of the proposed method with other existing methods in the literature. In [26], a compressed domain image segmentation method was based on Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA). Following neural network inference for each tile, a Wavelet Weighted Ensemble (WWE) based on the Inverse DWT (IDWT) was used to reconstruct the entire prediction image. The resolution of the normal image (20× magnification) needed to be reduced (10× magnification) to learn the experimental ROI size without compression. However, it is impossible to avoid losing high-frequency components throughout this operation. In [27], an analysis pipeline used a quality control procedure to segregate nuclei in whole-slide tissue images of various types of cancer using a CNN along with Watershed. This method began by creating random polygons that could potentially be nuclear masks. The textures

and colors of authentic tissue were then used to build the first synthetic patch. The first synthetic patch was then improved to make it look more lifelike. The synthetic patch's sample weight was calculated along the way, showing how realistic it was. Finally, sample weights, improved synthetic patch, and initially created nuclear masks were used to train a segmentation network. The method in [28] involved a Quick Attention Module, an Encoder-Decoder Network, and a Multi Loss Function that combined the Binary Cross Entropy (BCE) Loss and the Focal Loss Dice Loss. In [29], the same dataset was used. As can be observed in Table IV, the proposed method achieved better results in terms of Jaccard index, Precision, and F1-score compared to relevant deep learning methods, demonstrating its robustness.

V. CONCLUSION

This study presented a robust method for histopathology image segmentation, which is a hybridization of some image processing techniques: normalization, Otsu's thresholding, Gaussian filter, and the opening morphological operation. The proposed method was applied to the EBHI-SEG dataset, which contains several histopathology images showing cancer areas. The results of the proposed method, in terms of Jaccard index, Precision, Recall, F1-score, and SSIM, demonstrated satisfactory and promising results. In addition, a comparison with other existing methods showed that the proposed method outperforms many of them. The proposed method improves tissue analysis's accuracy, effectiveness, and repeatability, greatly benefiting a range of clinical and scientific uses, as numerous pathological characteristics, including calcifications, fibrosis, necrosis, blood vessel segmentation, counting the number of immune cells, segmenting and recognizing mitotic figures, and ROI segmentation, are crucial for a deep tissue examination. In summary, this work offers a straightforward and computationally effective segmentation technique that can be applied to histopathological tumor identification. Unlike approaches that make copious use of data in image processing or tumor cell segmentation, the proposed technique is straightforward and uses parameters that can be easily replicated.

DATA AVAILABILITY

All data is available upon request from the corresponding author.

ACKNOWLEDGMENTS

This paper was derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 13382-psu-2023-PSNU-R-3-1-EI-. The authors would like to acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia, in paying the article processing charges of this publication. This research is supported by the Automated Systems and Computing Lab (ASCL), Prince Sultan University, Riyadh, Saudi Arabia.

REFERENCES

[1] R. Gurumoorthy and M. Kamarasan, "Breast Cancer Classification from Histopathological Images using Future Search Optimization Algorithm and Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12831–12836, Feb. 2024, https://doi.org/10.48084/etasr.6720.

[2] Z. Chen *et al.*, "Weakly Supervised Histopathology Image Segmentation With Sparse Point Annotations," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1673–1685, Feb. 2021, https://doi.org/10.1109/JBHI.2020.3024262.

[3] C. Sun *et al.*, "Gastric histopathology image segmentation using a hierarchical conditional random field," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1535–1555, Oct. 2020, https://doi.org/10.1016/j.bbe.2020.09.008.

[4] P. Qin, J. Chen, J. Zeng, R. Chai, and L. Wang, "Large-scale tissue histopathology image segmentation based on feature pyramid," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, Aug. 2018, Art. no. 75, https://doi.org/10.1186/s13640-018-0320-8.

[5] L. Khrissi, N. El Akkad, H. Satori, and K. Satori, "Clustering method and sine cosine algorithm for image segmentation," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 669–682, Mar. 2022, https://doi.org/10.1007/s12065-020-00544-z.

[6] A. Kanadath, J. A. A. Jothi, and S. Urolagin, "Histopathology Image Segmentation Using MobileNetV2 based U-net Model," in *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India, June 2021, pp. 1–8, https://doi.org/10.1109/CONIT51480.2021.9498341.

[7] Y. Kurmi, V. Chaurasia, and N. Kapoor, "Design of a histopathology image segmentation algorithm for CAD of cancer," *Optik*, vol. 218, Sept. 2020, Art. no. 164636, https://doi.org/10.1016/j.ijleo.2020.164636.

[8] L. Khrissi, N. E. Akkad, H. Satori, and K. Satori, "An Efficient Image Clustering Technique based on Fuzzy C-means and Cuckoo Search Algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 423-432, 2021, https://doi.org/10.14569/IJACSA.2021.0120647.

[9] A. Ben Hamida *et al.*, "Deep learning for colon cancer histopathological images analysis," *Computers in Biology and Medicine*, vol. 136, Sept. 2021, Art. no. 104730, https://doi.org/10.1016/j.compbiomed.2021.104730.

[10] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, and J. Kini, "NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Computers in Biology and Medicine*, vol. 128, Jan. 2021, Art. no. 104075, https://doi.org/10.1016/j.compbiomed.2020.104075.

[11] L. Khrissi, N. El Akkad, H. Satori, and K. Satori, "A Performant Clustering Approach Based on An Improved Sine Cosine Algorithm," *International Journal of Computing*, pp. 159–168, June 2022, https://doi.org/10.47839/ijc.21.2.2584.

[12] X. Li *et al.*, "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches," *Artificial Intelligence Review*, vol. 55, no. 6, pp. 4809–4878, Aug. 2022, https://doi.org/10.1007/s10462-021-10121-0.

[13] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Computers in Biology and Medicine*, vol. 128, Jan. 2021, Art. no. 104129, https://doi.org/10.1016/j.compbiomed.2020.104129.

[14] H. Ziang, J. Zhang, and L. Li, "Framework for lung CT image segmentation based on UNet++." arXiv, Jan. 05, 2025, https://doi.org/10.48550/arXiv.2501.02428.

[15] N. Kumar *et al.*, "A Multi-Organ Nucleus Segmentation Challenge," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, May 2020, https://doi.org/10.1109/TMI.2019.2947628.

[16] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, and P. H. N. de With, "Histopathology Stain-Color Normalization Using Deep Generative Models," presented at the Medical Imaging with Deep Learning, July 2022.

[17] Md. Z. Hoque, A. Keskinarkaus, P. Nyberg, and T. Seppänen, "Retinex model based stain normalization technique for whole slide image analysis," *Computerized Medical Imaging and Graphics*, vol. 90, June

2021, Art. no. 101901, https://doi.org/10.1016/j.compmedimag.2021.101901.

[18] S. Vijh, M. Saraswat, and S. Kumar, "A new complete color normalization method for H&E stained histopatholgical images," *Applied Intelligence*, vol. 51, no. 11, pp. 7735–7748, Nov. 2021, https://doi.org/10.1007/s10489-021-02231-7.

[19] Z. Yildirim, E. Hancer, R. Samet, M. T. Mali, and N. Nemati, "Effect of Color Normalization on Nuclei Segmentation Problem in H&E Stained Histopathology Images," in *2022 30th Signal Processing and Communications Applications Conference (SIU)*, Safranbolu, Turkey, May 2022, pp. 1–4, https://doi.org/10.1109/SIU55565.2022.9864814.

[20] F. Pérez-Bueno *et al.*, "Blind color deconvolution, normalization, and classification of histological images using general super Gaussian priors and Bayesian inference," *Computer Methods and Programs in Biomedicine*, vol. 211, Nov. 2021, Art. no. 106453, https://doi.org/10.1016/j.cmpb.2021.106453.

[21] I. Hirra *et al.*, "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling," *IEEE Access*, vol. 9, pp. 24273–24287, 2021, https://doi.org/10.1109/ACCESS.2021.3056516.

[22] M. T. Nyo, F. Mebarek-Oudina, S. S. Hlaing, and N. A. Khan, "Otsu's thresholding technique for MRI image brain tumor segmentation," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 43837–43849, Dec. 2022, https://doi.org/10.1007/s11042-022-13215-1.

[23] A. T. H. Al-Rahlawee and J. Rahebi, "Multilevel thresholding of images with improved Otsu thresholding by black widow optimization algorithm," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 28217–28243, July 2021, https://doi.org/10.1007/s11042-021-10860-w.

[24] J. Jose *et al.*, "An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion," *Biomedical Signal Processing and Control*, vol. 66, Apr. 2021, Art. no. 102480, https://doi.org/10.1016/j.bspc.2021.102480.

[25] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, June 2022, Art. no. 210, https://doi.org/10.1186/s13104-022-06096-y.

[26] H. Kim *et al.*, "Deep learning-based histopathological segmentation for whole slide images of colorectal cancer in a compressed domain," *Scientific Reports*, vol. 11, no. 1, Nov. 2021, Art. no. 22520, https://doi.org/10.1038/s41598-021-01905-z.

[27] L. Hou *et al.*, "Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types," *Scientific Data*, vol. 7, no. 1, June 2020, Art. no. 185, https://doi.org/10.1038/s41597-020-0528-1.

[28] S. Wazir and M. M. Fraz, "HistoSeg: Quick attention with multi-loss function for multi-structure segmentation in digital histology images," in *2022 12th International Conference on Pattern Recognition Systems (ICPRS)*, Saint-Etienne, France, June 2022, pp. 1–7, https://doi.org/10.1109/ICPRS54038.2022.9854067.

[29] L. Shi *et al.*, "EBHI-Seg: A novel enteroscope biopsy histopathological hematoxylin and eosin image dataset for image segmentation tasks," *Frontiers in Medicine*, vol. 10, Jan. 2023, https://doi.org/10.3389/fmed.2023.1114673.