Histopathological Image Analysis For Classifying Colon Cancer

Akshita Singh, Jazlyn Jose, Unnati Gupta

[1*]Computer Science Department, BML Munjal University, Gurgaon, Haryana, India.

Abstract

Background and Objectives: Colorectal cancer, often known as colon cancer or rectal cancer, is a kind of cancer that starts in the intestines. Histopathological image analysis is a branch of medicine that involves the microscopic examination and interpretation of tissue samples in order to understand the structural and cellular alterations associated with illnesses. Pathologists have traditionally depended on physical examination of stained tissue slides under a microscope to diagnose illnesses. Histopathological image analysis, however, has advanced with the advent of digital pathology to include computer-based approaches for extracting relevant information from digitized images of tissue specimens. **Material and Methods:** In this study, we conducted histopathological image analysis for the classification of colon cancer from benign to malignant using a dataset comprising histopathological images of colon tissues. The principal aim of this study was to evaluate the efficacy of two distinct feature extraction techniques: morphological feature extraction and Haralick features. This assessment was conducted within the framework of training a deep learning model using transfer learning, specifically employing the ResNet-18 architecture. The investigation delved into the comparative performance of these feature extraction methods in enhancing the capabilities of the neural network model for various applications and tasks. **Results:** The Haralick Features-based Convolutional Neural Network displayed remarkable performance, achieving 98.50% accuracy in classifying colorectal cancer histopathological images, while the Morphological Features-trained CNN demonstrated robust classification with an accuracy of 97.88% on the test set.

Keywords: Benign Tissues, Colon Adenocarcinomas Deep Learning Techniques, Image Processing Methods, Convolutional Neural Networks (CNNs).

1. Introduction

Colon cancer develops in the large intestine (colon) or the rectum (the lower part of the colon). It stands among the prevalent and serious cancers worldwide. In 2020, approximately 1.9 million individuals received a diagnosis of colon cancer. The rise in the global occurrence of malignant tumors is attributed to population growth. Although it can impact individuals of any age, it tends to be more prevalent in those between the ages of 50 and 60 [1].

Cancer mortality is expected to reach 60% by 2035 [2]. There are numerous approaches for diagnosing colon cancer, the most common of which being biopsy and laboratory analysis. Manual diagnosis takes a long time and is vulnerable to doctor disagreements due to the lack of health centers and medical staff, particularly in underdeveloped nations [3]. As a result, digital image processing and artificial intelligence approaches are used to address these issues. Non-cancerous benign tissues do not invade neighboring tissues or spread to other regions of the body. Polyps are examples of benign growths in the colon. Most polyps are benign, but some can progress to malignancy over time. Under a microscope, benign colon tissue appears normal, with well-organized and homogenous cell structures. The cells do not often exhibit cancer-like traits such as uncontrolled proliferation or invasion into adjacent tissues.

Adenocarcinomas are cancerous tumors that develop from glandular tissues. Adenocarcinomas form from the glandular cells that coat the lining of the colon in the case of colon cancer. These malignant cells grow uncontrollably, lack differentiation, and have the ability to infect neighboring tissues and organs. Adenocarcinomas can also metastasis, or spread to other areas of the body via the bloodstream or lymphatic system. Adenocarcinoma tissue often appears disorganized under a microscope, with irregular and aberrant cell structures. The cells may form gland-like structures, indicating that they originated in glandular tissues.

To this background, in this paper, we propose the CNN architecture, which will help classify between the benign and malignant tissues. Thus the contributions of the study are:

- Analysis of existing studies on histopathological image analysis of colon cancer
- Feature extraction process is thorough and encompasses both Haralick Texture Features and Morphological Feature Extraction. The extraction of Haralick texture features involves intricate statistical calculations based on

pixel intensity relationships within identified Regions of Interest (ROIs), providing insights into texture patterns crucial for classification.

- The morphological feature extraction process, including binary thresholding, skeletonization, and calculation of features like area, perimeter, elongation, compactness, and hole detection, adds a rich layer of information capturing structural nuances in histopathological images.

The paper is organized as follows. Section 2 gives the literature review. Section 3 describes the methodology and the dataset used for this study. Section 4 presents the results and discussion on the results obtained. Finally, the conclusions are presented in Section 5.

2. Related Work

In recent years, there has been substantial progress in the development of computer-aided diagnostic (CAD) systems for the detection and classification of colon cancer tissues in the field of medical image analysis. Several prominent studies have contributed to this advancement, each using unique approaches and algorithms.

Jørgensen et al. [4] utilized cell nuclei features extracted from hematoxylin and eosin stained slides to detect colon cancer tissue. The feature extraction process involved obtaining information from segmented cell nuclei structures in regions of interest (ROIs). Intensity and texture features were extracted, considering challenges posed by high cellular density. Features, including minimum, mean, and maximum intensity, along with RLM texture features, were derived from different color spaces, forming the basis for subsequent classification.Employing a Random Forest Classifier, their CAD system demonstrated a promising accuracy of 91%, with an AUC of 0.96. The optimal threshold yielded a sensitivity of 0.88 and specificity of 0.92, showcasing the potential of this approach in distinguishing normal and cancerous tissue.

Hage Chehade et al. [5] adopted a feature engineering approach for lung and colon cancer classification, achieving remarkable accuracy of 99% using XGBoost. The incorporation of SHAP interpretability enhanced transparency, providing valuable insights into feature importance for medical specialists. The study aimed to enhance disease diagnosis from histopathological images by extracting 37 features, including first-order statistics, Hu invariant moments. First-order statistics depicted brightness distribution,and Hu invariant moments highlighted geometric aspects. Recursive Feature Elimination (RFE) was employed for feature selection, retaining 12 crucial features through cross-validation for optimal classification efficiency.

Al-Jabbar and colleagues [6] introduced a hybrid system designed for histopathological analysis to detect malignancies in lung and colon cancers. The Artificial Neural Network (ANN), incorporating deep learning architectures and handcrafted features, demonstrated an accuracy ranging from 95% to 96%. Notably, optimal performance was observed when both high and low-dimensional features were combined.

In a study by Garg and Garg [7], the focus was on predicting lung and colon cancer by analyzing histopathological images using pre-trained Convolutional Neural Network (CNN) models. The methodology involved fine-tuning eight pre-trained CNN architectures for feature extraction, incorporating a modified part with Max-pooling2D, Average-pooling2D, and a flatten layer. All eight models exhibited exceptional accuracy, with a significant result indicating 100% accuracy in colon cancer classification.

Hsu et al. [8] developed a system specifically for colorectal polyp detection and classification, leveraging grayscale images and deep learning. Their CNN model achieved a high accuracy of 95.1% for polyp detection in grayscale images, surpassing RGB images in certain scenarios.

Masud et al. [9] adopted a machine learning approach for diagnosing lung and colon cancer using a deep learning-based classification framework. Their models demonstrated exceptional accuracy (100%) for colon cancer classification, with the NASNetMobile model achieving 98% accuracy.

Finally, In their work, Hasan et al. [10] introduced an automated system for detecting and characterizing colon cancer employing deep convolutional neural networks (DCNN). The DCNN-based system exhibited remarkable precision, recall, F1-score, and an AUC value of 0.998. Notably, it achieved an impressive accuracy rate of 99.80% in effectively classifying colon tissues as either benign or adenocarcinoma.

Collectively, these studies highlight the diverse approaches, ranging from traditional machine learning techniques to advanced deep learning architectures, contributing to the development of robust and reliable tools for cancer detection and classification in histopathological images. The reported high accuracies and comprehensive evaluation metrics underscore the potential impact of these methodologies in advancing the field of medical image analysis for cancer diagnosis.

3. Methodology

In our study, we present a comprehensive methodology that employs Haralick Texture Features and Morphological Feature Extraction for enhanced image classification, particularly focusing on the analysis of colon cancer tissues. Our motivation stems from the imperative to discern subtle variations in texture and structure, pivotal for accurate classification, particularly in the realm of colon cancer diagnostics.

Haralick Texture Features explore intricate texture patterns, encompassing vital metrics like angular second moment (ASM), contrast, correlation, variance, inverse difference moment (IDM), sum average, sum variance, sum entropy, entropy, and difference variance. In contrast, Morphological Feature Extraction focuses on retrieving information such as the number of holes, area, perimeter, elongation, and compactness from regions of tissue images identified as connected components in binary thresholded images.

Our methodology integrates various image enhancement techniques, incorporating Contrast Limited Adaptive Histogram Equalization (CLAHE), Gaussian Blur, and Region of Interest (ROI) through binary thresholding, contour detection, and contour drawing. This combination facilitates the extraction of Haralick Texture Features, utilizing preprocessing enhancements from CLAHE and Gaussian Blur to prepare the groundwork for effective texture-based feature extraction through Haralick analysis.

Additionally, for Morphological feature extraction, our approach incorporates CLAHE with binary thresholding, crucial for isolating pertinent morphological details within images. Subsequent steps involve skeletonization to derive morphological features, offering valuable insights into the structural characteristics of the analyzed images.

In summary, our methodology adopts a systematic approach, combining image enhancement techniques and distinct feature extraction methods, to augment the quality of image data. This concerted effort results in the extraction of meaningful texture and morphological features, facilitating a thorough analysis of visual content.

Navigating through preprocessing, statistical analysis, and ResNet-18 adaptation for classification, our holistic approach blends deep learning with nuanced features. The deliberate integration of morphological and texture-based features aims to create a synergistic foundation, enhancing the dataset for robust diagnostic advancements in histopathological imaging. This

methodology actively engages with the features, transcending conventional boundaries, and sets the stage for a nuanced exploration of their significance in medical imaging.

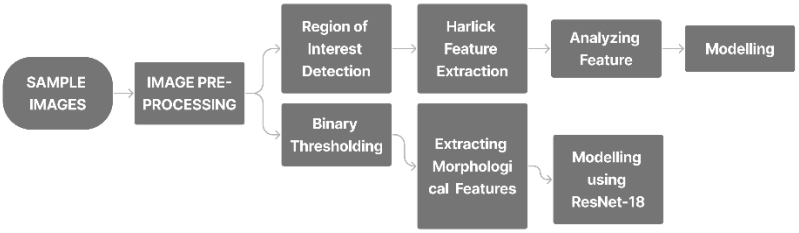

Fig. 3.1: Methodology

3.1. Data-Set

The LC25000 Lung and Colon Histopathological Image Dataset, sourced from Kaggle, constitutes a comprehensive collection comprising 25,000 high-resolution images. These images are organized into 5 classes, each containing 5,000 images, with a consistent dimension of 768 x 768 pixels and stored in JPEG file format. The dataset is particularly structured within the subfolder "colon_image_sets," which includes two additional subfolders: "colon_aca" and "colon_n." The former consists of 5,000 images depicting colon adenocarcinomas, while the latter features 5,000 images portraying benign colonic tissues. An intriguing aspect of this dataset is the inclusion of 10,000 augmented images, evenly distributed between colon adenocarcinomas and benign tissues, generated from an initial pool of 500 images. It's noteworthy that the dataset adheres to Health Insurance Portability and Accountability Act (HIPAA) standards, ensuring compliance and validation from authoritative sources in the medical field.

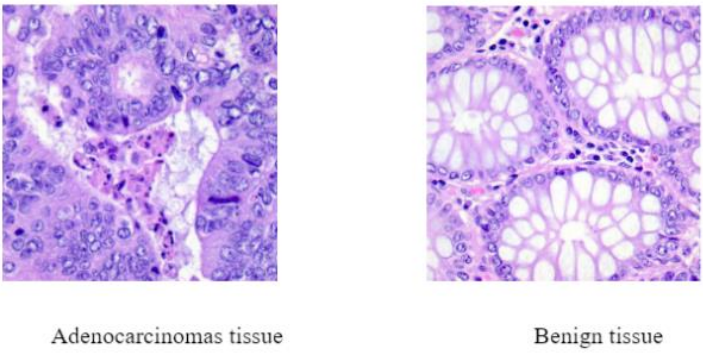

Fig. 3.2: Adenocarcinomas and Benign Tissue

3.2. Data Pre-Processing

In the pursuit of comprehensive and effective data preprocessing for the analysis of histopathological images in the context of colorectal cancer classification, a series of systematic steps were undertaken. The primary objective of these preprocessing procedures was to enhance the quality, reliability, and discriminatory power of the image dataset. The initial phase involved the conversion of color images into grayscale, a critical step in simplifying subsequent analyses and focusing on the inherent structural and textural information present in the images.

Subsequently, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to the grayscale images, further optimizing the visibility of intricate details and subtle variations in pixel intensities. This adaptive enhancement technique aims to address limitations associated with traditional histogram equalization by preventing over-amplification of noise in regions with low contrast.

Following the CLAHE application, Gaussian blur was introduced to the grayscale images, contributing to noise reduction and smoothing of pixel intensity variations. The choice of kernel size (3x3) for Gaussian blur was made to strike a balance between effective blurring and preservation of relevant image features. As with previous steps, the Gaussian blur process was executed separately for 'adenocarcinomas' and 'benign' subfolders.

A subsequent preprocessing stage involved the application of histogram equalization to further enhance image contrast. This step aimed to normalize the distribution of pixel intensities across the images, contributing to improved feature extraction and subsequent analyses. Simultaneously, the identification of Regions of Interest (ROIs) within the grayscale images was undertaken. This involved the segmentation of images through the application of binary thresholding, followed by contour detection to delineate areas of interest. The integration of these techniques facilitated the extraction of relevant anatomical structures and pathological features.

Additionally, texture features were extracted from the grayscale images using Haralick features, a set of statistical descriptors capturing textural patterns within images. This step allowed for the incorporation of texture-based information, potentially enriching the dataset with discriminative features relevant to cancer classification.

To address potential issues of missing data and prepare the dataset for machine learning applications, several subsequent steps were implemented. Missing values within numerical features were imputed using the mean of available data, ensuring the preservation of feature integrity. Categorical labels were encoded using the LabelEncoder from scikit-learn, transforming class labels into numeric representations. Furthermore, feature scaling was applied to normalize numerical features, specifically using the StandardScaler, which standardizes data to a mean of 0 and a standard deviation of 1.

An in-depth exploration of the dataset followed, encompassing an assessment of sample distributions, statistical summaries, and visualizations of feature distributions across different classes. This exploratory phase provided critical insights into the dataset's characteristics, informing subsequent modeling decisions.

Statistical tests, including t-tests, were employed to assess the significance of individual features in distinguishing between adenocarcinoma and benign samples. The identification of statistically significant features laid the groundwork for the selection of relevant variables in subsequent analyses. Effect size calculations, specifically Cohen's d, were employed to gauge the magnitude of differences between groups, aiding in the interpretation of feature significance.

3.3. Feature Extraction

Feature extraction involves transforming raw histopathological image data into meaningful numerical representations, emphasizing relevant characteristics such as texture, morphology, and statistical features. This process is essential for preparing data to train machine learning models for the classification of colon tissue types.

3.3.1 ROI Identification and Haralick Feature Extraction

This feature extraction methodology comprises a two-step procedure: Region of Interest (ROI) Identification and Haralick Texture Feature Extraction. We meticulously analyze grayscale images of 'adenocarcinomas' and 'benign' to derive meaningful insights.

Region of Interest (ROI) Identification:

Identifying the Region of Interest (ROI) is a crucial step in our approach, laying the groundwork for the extraction of Haralick Texture Features essential for the binary classification of colon cancer tissues. This segment delineates the detailed steps in the ROI identification process, highlighting the incorporation of Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques.

The application of Histogram Equalization served as the initial enhancement step. Mathematically, this process can be expressed as:

$$G(x,y) = T[f(x,y)] = \sum_{i=0}^{f(x,y)} p_i$$

where G(x,y) is the equalized intensity at position, (x,y), f(x,y) is the original intensity, and pi is the probability of occurrence of intensity i. This technique redistributes intensity values, augmenting local and global contrast.

Following HE, a critical step involves Binary Thresholding. This binary transformation facilitates segmentation, classifying pixels into two distinct intensity levels, and effectively isolates regions of interest. Contour detection is pivotal in identifying connected components within the binary thresholded image. Mathematically, contours are defined as:

$$C(x,y)=\{(x,y)|\ g(x,y)=255\}$$

The subsequent contour drawing on a copy of the original grayscale image visually accentuates the regions of interest. This enhanced visual representation is a crucial precursor to Haralick Texture Feature extraction.

In summary, the ROI identification process optimizes contrast, segments areas of interest, and visually highlights these regions, setting the stage for subsequent Haralick Texture Feature analysis in our quest for precise colon cancer tissue classification.

Haralick Texture Feature Extraction:

The essence of the feature extraction process centers on computing Haralick texture features. This technique involves capturing visual patterns within an image to articulate its texture. It employs statistical computations based on the interrelation of pixel values, providing information on attributes such as smoothness, roughness, and complexity. The Grey Level Co-occurrence Matrix (GLCM) is employed to derive Haralick characteristics, counting occurrences of adjacent gray-level pixels in an image. Haralick subsequently yields 12 values extracted from the GLCM, quantifying texture characteristics based on this matrix.

Haralick texture features are extracted through intricate statistical calculations based on pixel intensity relationships within the identified Regions of Interest (ROIs). These features encapsulate crucial information about the underlying texture patterns and include:

Angular Second Moment (ASM): Measures the local homogeneity of the image.

Contrast: Quantifies the local variations in pixel intensities.

Correlation: Reflects the linear dependence between pixel intensities.

Variance: Describes the degree of intensity variability within the image.

Inverse Difference Moment (IDM): Represents the local homogeneity of the image.

Sum Average: Provides insights into the distribution of pixel intensity sums.

Sum Variance: Measures the variability of pixel intensity sums.

Sum Entropy: Captures the disorder in pixel intensity sum distribution.

Entropy: Represents the randomness or unpredictability of pixel intensities.

Difference Variance: Quantifies the variations in pixel intensity differences.

Through the analysis of the images below, it becomes evident that a higher value for both Feature 3(correlation) and Feature 6(sum averages) correlates with a higher likelihood of the tissue being benign rather than adenocarcinoma.

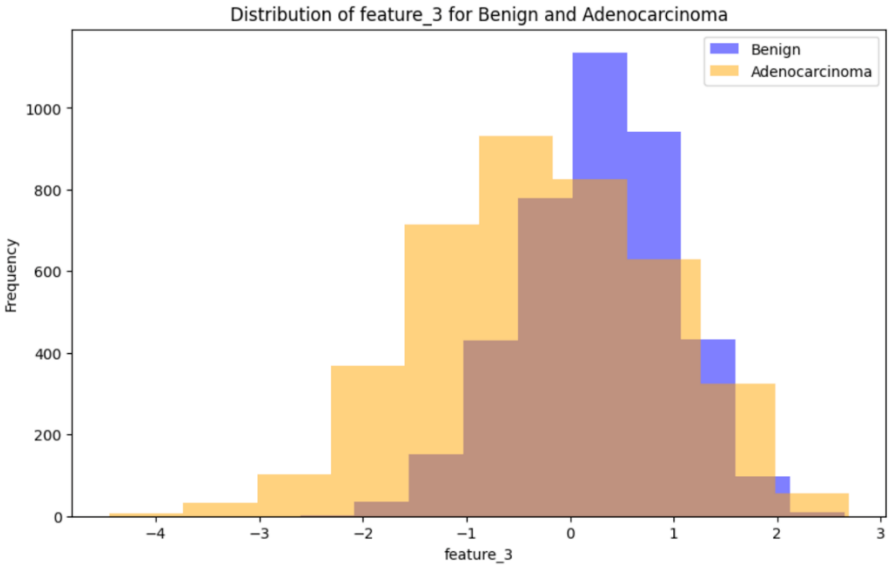

Fig. 3.3: Feature Distribution (Correlation)

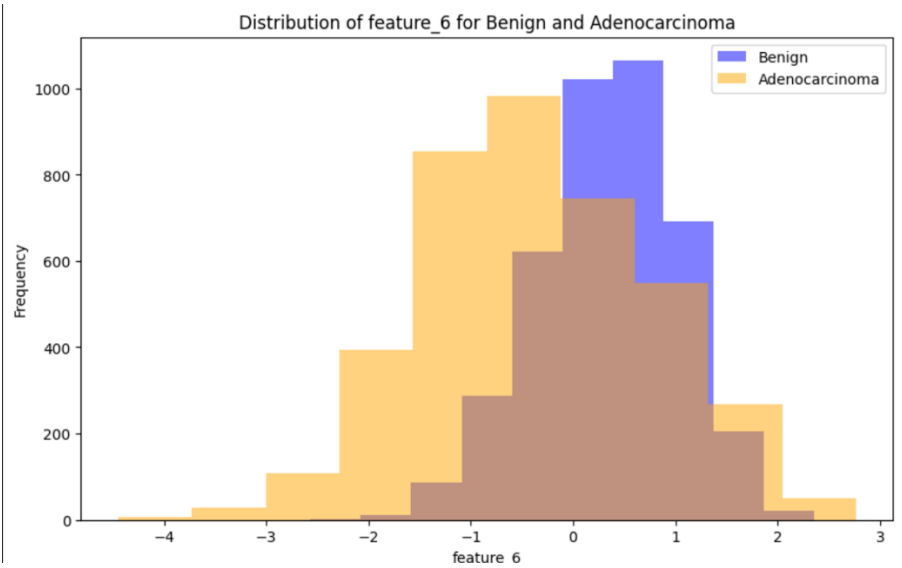

Fig. 3.4: Feature Distribution (Sum Average)

The amalgamation of these Haralick texture features serves as a powerful discriminative tool for the classification of histopathological images of colon tissues. These features encapsulate intricate textural details, empowering subsequent analyses for the precise differentiation between cancerous and non-cancerous tissues. The robustness of this methodology lies in its ability to unveil nuanced patterns, contributing significantly to the advancement of diagnostic capabilities in histopathological imaging of colon cancer.

In summary, this feature extraction methodology combines effective ROI identification techniques with advanced Haralick texture feature extraction, providing a comprehensive and discriminating set of features for the classification of colon tissue histopathological images.

3.3.2 ROI and Morphological Feature Extraction

The methodology involves a comprehensive process incorporating binary thresholding and morphological feature extraction. These steps are crucial for

capturing diverse information from histopathological images, enhancing the dataset for subsequent machine learning tasks.

Identifying structural nuances in histopathological tissue images is crucial therefore, we employ a multi-step process, incorporating morphological feature extraction, binary thresholding, and skeletonization to unravel significant characteristics for improved image-based classification.

Firstly, binary thresholding is employed to segment the histopathological images into distinct regions based on pixel intensities. This process involves applying different thresholding techniques to each color channel, resulting in multiple binary images. Each thresholded image highlights distinct structural information present in different color components.

Morphological feature extraction is then performed on the binary images, capturing information about the shape and structure of objects. The calculated features include area, perimeter, elongation, compactness, and the detection of holes. These features provide valuable insights into the characteristics of tissue structures and contribute to the development of image-based classification models for cancer diagnosis.

Skeletonization is a crucial step in morphological feature extraction for histopathological images. It is a technique that aims to capture the essential structure of objects while simplifying their representation. In the context of histopathology, skeletonization reduces complex structures, such as cellular outlines and tissue boundaries, to their primary axis. This involves transforming a binary image into a graph representation, where each pixel corresponds to a node. The resulting skeleton provides a simplified yet connected representation of structural elements, facilitating the analysis of intricate tissue patterns.

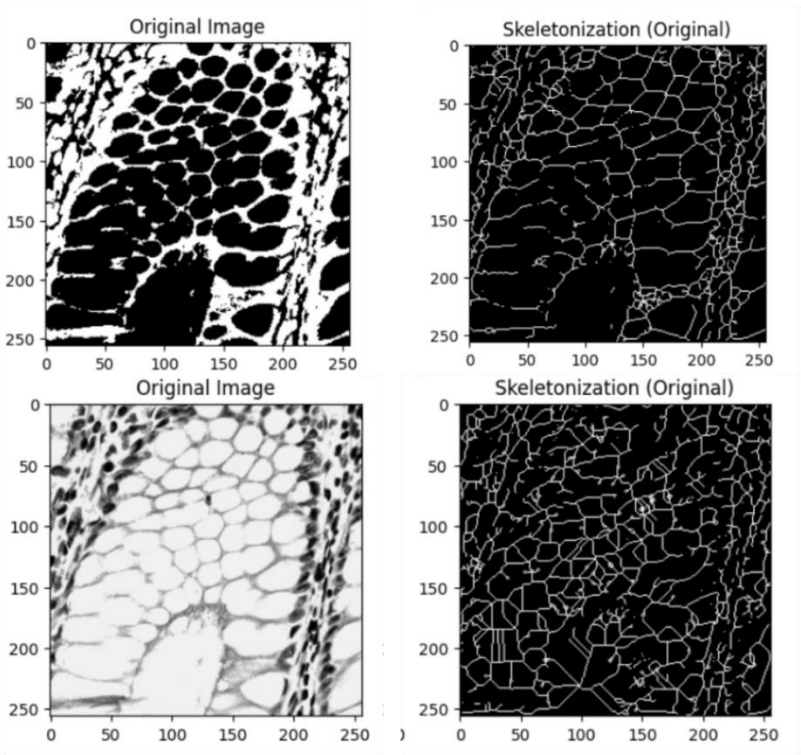

Fig. 3.5: Binary 2, Binary 3 Skeletonization

Calculating the area and perimeter of segmented regions in histopathological images yields fundamental information about object shapes. Area represents the total number of pixels within a region, providing an indication of the extent of tissue coverage. Perimeter, on the other hand, measures the boundary length of the segmented region. These metrics serve as quantitative descriptors for the spatial characteristics of tissue structures. Mathematically, area (A) and perimeter (P) are computed as A = Σ pixels and P = Σ boundary pixels, respectively. These metrics contribute to understanding the distribution and boundary complexity of histopathological features. In the tissues it was observed that Adenocarcinoma tissues exhibit larger areas, indicating a higher extent of tissue coverage. Benign tissues, however, display slightly smaller

areas. The perimeter of adenocarcinoma tissues is comparable to benign tissues, emphasizing the need for additional metrics for differentiation.

Elongation is a morphological feature that measures the deviation of an object from a perfect circle. In the context of histopathological images, elongation provides insights into the shape characteristics of tissue structures, such as nuclei. Mathematically, elongation is computed as the ratio of the sum of skeleton pixels (S) to the sum of all pixels in the original image (A). Elongation (E) is expressed as E = S / A. A higher elongation value indicates a more elongated or irregular shape, while lower values suggest structures closer to circular. Elongation serves as a quantitative measure for characterizing the morphology of histopathological elements where Adenocarcinoma tissues tend to have higher elongation values, suggesting more irregular shapes. Benign tissues, with lower elongation values, exhibit shapes closer to circular. This metric contributes valuable insights into the morphological diversity of nuclei in different tissues.

Compactness is a metric that quantifies how spread out or compact an object is. In the context of histopathological image analysis, compactness provides information about the spatial distribution of tissue structures. Mathematically, compactness (C) is calculated using the formula $C = (P^2) / (4 * \pi * A)$, where P is the perimeter and A is the area. This metric offers insights into the degree of irregularity in the shape of segmented regions. Higher compactness values indicate more compact structures, while lower values suggest greater dispersion. It was observed that Adenocarcinoma tissues demonstrate slightly higher compactness, indicating more compact structures. Benign tissues, with lower compactness values, suggest greater dispersion. Compactness, in combination with other features, aids in capturing nuances related to the spatial distribution of tissue structures.

Hole detection is a morphological feature that identifies void spaces or regions of lower intensity within structures. In histopathological images, the presence and distribution of holes can be indicative of structural complexity. Hole detection involves analyzing the topology of the binary image to identify regions with enclosed spaces. Thereby with more analysis it was found that Adenocarcinoma tissues tend to have a higher number of holes, reflecting structural complexity. Benign tissues, with fewer holes, indicate more solid and homogeneous regions. Hole detection enhances the understanding of tissue structures, contributing to the discrimination of pathological conditions.

The figure below illustrates the distribution of holes and elongation in adenocarcinoma and benign, and it is found that adenocarcinoma has more number of holes and elongation than benign tissues.

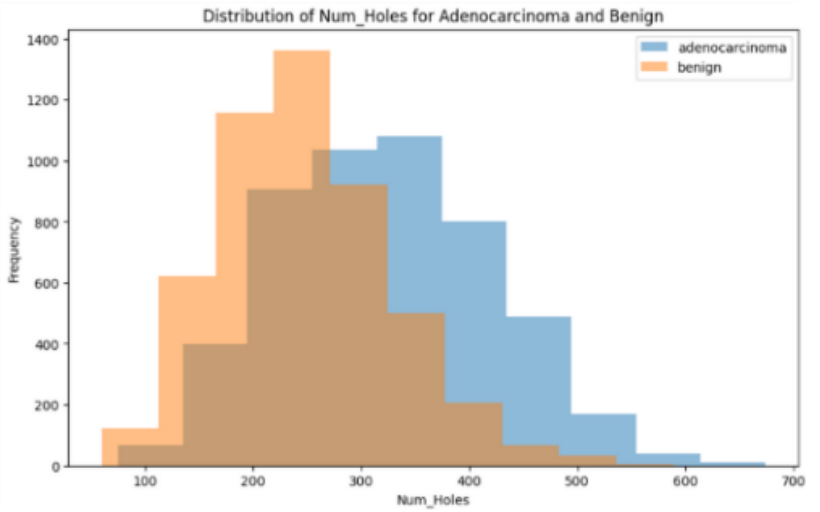

Fig. 3.6: Distribution of Num_Holes for Adenocarcinomas, Benign

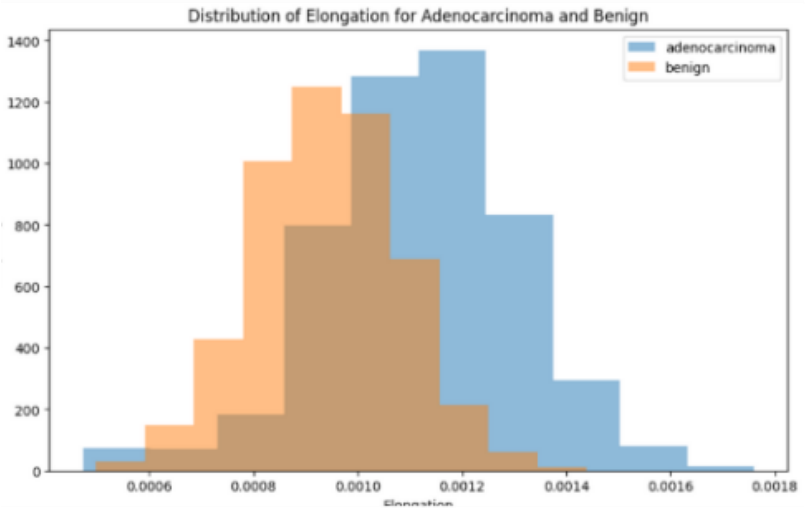

Fig. 3.6: Distribution of Elongation for Adenocarcinomas, Benign

In summary, the methodology integrates binary thresholding and morphological feature extraction to comprehensively characterize histopathological images. This multi-step approach enriches the dataset with diverse information, empowering subsequent machine learning models to effectively classify cancer types based on intricate image features.

3.4. Model Architecture

The model architecture employed in this research paper involves a Convolutional Neural Network (CNN) for the classification of histopathological images related to colorectal cancer. The preprocessing steps, outlined earlier, provided a feature-enriched dataset for training and evaluating the model. The implemented CNN utilizes a pre-trained ResNet-18 architecture, leveraging transfer learning to harness knowledge gained from a large dataset. The ResNet-18 model's layers were adapted to suit the binary classification task.

The model comprises three main components: the feature extractor, average pooling layer, and a fully connected layer for classification. The feature extractor, derived from ResNet-18, is responsible for capturing hierarchical features from the input images. The subsequent adaptive average pooling layer spatially averages the output of the feature extractor, ensuring a consistent input size for the subsequent fully connected layer. The final fully connected layer, with two output nodes corresponding to the adenocarcinoma and benign classes, facilitates binary classification.

The training process involves minimizing the Cross-Entropy Loss using the Adam optimizer. The model was trained over ten epochs, iterating through the training dataset, updating the parameters based on backpropagation, and optimizing the classification accuracy. The trained model was subsequently evaluated on a separate test set, yielding an accuracy metric indicative of its performance in discriminating between adenocarcinoma and benign samples.

The incorporation of significant texture features, identified through statistical analyses and effect size calculations, contributes to the model's ability to discern relevant patterns in the histopathological images. The integration of these features with a deep learning model aligns with a comprehensive approach to image classification, harnessing both traditional image processing techniques and advanced neural network architectures. The presented model architecture establishes a foundation for further investigations into the automated classification of colorectal cancer based on histopathological images.

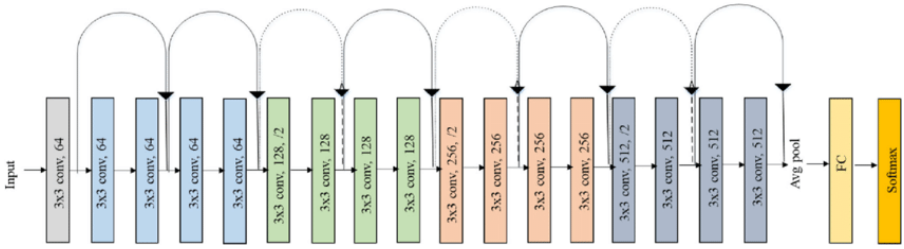

Fig. 3.7: ResNet-18 Architecture

3.5. Modeling using ResNet-18

The modeling phase involved leveraging the power of Convolutional Neural Networks (CNNs), with a specific focus on the ResNet-18 architecture. ResNet-18, a pre-trained CNN model, was employed for its proven effectiveness in image classification tasks. Transfer learning was utilized to benefit from the knowledge gained on a large dataset, providing a head start in understanding complex features within histopathological images. The ResNet-18 architecture was adapted to accommodate the binary classification task, with the last fully connected layer modified for the specific output classes—adenocarcinoma and benign. Notably, Haralick texture features, such as Contrast, Correlation, Inverse Difference Moment (IDM), Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, were seamlessly integrated with the original dataset. Similarly the Morphological features that were extracted were trained on the ResNet-18 pretrained model.

This fusion of traditional image processing features with the deep learning model showcases the versatility of the approach. The CNN model underwent training using a meticulously prepared dataset, and its performance was evaluated using standard metrics, ensuring a robust framework for the classification of histopathological images. The adaptation of ResNet-18 in this context exemplifies a powerful synergy between traditional image analysis techniques and cutting-edge deep learning architectures, culminating in a comprehensive solution for accurate and efficient medical image classification.

4. Results

The trained Convolutional Neural Network (CNN) trained on the Haralick Features achieved promising results in the classification of histopathological images related to colorectal cancer. The model was trained over ten epochs, and the training process exhibited a consistent decrease in the loss, indicating effective learning and convergence. Notably, the model achieved an impressive accuracy of 98.50% on the test set, showcasing its robust ability to correctly classify images into adenocarcinoma and benign categories. Precision was 98.04%, recall was 98.04%, recall was 99.75%, F1 score was 98.88%.

The decreasing loss throughout the training process suggests that the model effectively learned relevant features from the input images, contributing to its high accuracy on the unseen test data. The achieved accuracy of 98.50% underscores the model's potential for accurate and reliable classification, which is particularly crucial in the context of medical image analysis.

Similarly the Convolutional Neural Network (CNN) trained on the Morphological features achieved 97.88% accuracy on the test set, showcasing its robust ability to correctly classify images into adenocarcinoma and benign categories.

Method	**Model**	**Accuracy**	**TP**	**FP**	**TN**	**FN**
Haralick Feature Extraction	ResNet-18	98.50%	798	0	802	0
Morphological Feature Extraction	ResNet-18	97.88%	500	22	478	0
Feature Selection	ResNet-18	99.90%	1010	2	988	0

Both models exhibit strong performance with high accuracies, showcasing their effectiveness in the given task. Morphological Feature Extraction demonstrates impressive accuracy with only 22 false positives and no false negatives, while Haralick Feature Extraction achieves perfection by eliminating both false positives and false negatives. Although both models are excellent, Haralick Feature Extraction edges ahead with a slightly superior accuracy of 98.50% compared to Morphological feature extraction still commendable accuracy of 97.80%.

5 Conclusion

While the achieved results are promising, it is essential to acknowledge the ongoing evolution of medical image analysis and the need for continual validation on diverse datasets. Future work could explore interpretability aspects, such as attention mechanisms, to enhance the model's transparency

and facilitate trust in clinical applications. Moreover, collaboration with domain experts and validation on external datasets would further fortify the generalizability and real-world impact of the proposed methodology.

In essence, this research contributes to the burgeoning field of histopathological image analysis, demonstrating the efficacy of a comprehensive approach for colon cancer detection.

References

1. Arslan N, Yilmaz A, Firat U, Tanriverdi MH. Analysis of cancer cases from Dicle University Hospital; ten years 'experience abstract. J Clin Anal Med 2018;9(2): 102-6.

2. Araghi, M., Soerjomataram, I., Jenkins, M. A., Brierley, J. D., Morris, E., Bray, F., & Arnold, M. (2019). Global trends in colorectal cancer mortality: projections to the year 2035. International Journal of Cancer, 144(12), 2992–3000. https://doi.org/10.1002/ijc.32055.

3. Al-Jabbar M, Alshahrani M, Senan EM, Ahmed IA. Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features. Bioengineering (Basel). 2023 Mar 21;10(3):383. doi: 10.3390/bioengineering10030383. PMID: 36978774; PMCID: PMC10045080.Kim, D. H., & Lee, H. Y. (2017). Image manipulation detection using convolutional neural network. *International Journal of Applied Engineering Research*, *12*(21), 11640-11646.

4. Jørgensen, A. S., Rasmussen, A. M., Andersen, N. K. M., Andersen, S. K., Emborg, J., Røge, R., & Østergaard, L. R. (2017). Using cell nuclei features to detect colon cancer tissue in hematoxylin and eosin stained slides. Cytometry Part A, 91(8), 785–793. https://doi.org/10.1002/cyto.a.23175

5. Chehade, A. H., Abdallah, N., Marion, J., Oueidat, M., & Chauvet, P. (2022). Lung and colon cancer classification using Medical Imaging : A feature Engineering approach. Research Square (Research Square). https://doi.org/10.21203/rs.3.rs-1211832/v1

6. Al-Jabbar M, Alshahrani M, Senan EM, Ahmed IA. Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features. Bioengineering (Basel). 2023 Mar 21;10(3):383. doi: 10.3390/bioengineering10030383. PMID: 36978774; PMCID: PMC10045080.

7. Satvik Garg and Somya Garg. 2020. Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps. In 2020 3rd

Artificial Intelligence and Cloud Computing Conference (AICCC 2020), December 18–20, 2020, Kyoto, Japan. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442536.3442543

8. Hsu C-M, Hsu C-C, Hsu Z-M, Shih F-Y, Chang M-L, Chen T-H. Colorectal Polyp Image Detection and Classification through Grayscale Images and Deep Learning. Sensors. 2021; 21(18):5995. https://doi.org/10.3390/s21185995

9. Masud, M., Sikder, N., Nahid, A., & AlZain, M. A. (2021). A machine learning approach to diagnosing lung and colon cancer using a Deep Learning-Based Classification framework. Sensors, 21(3), 748. https://doi.org/10.3390/s21030748

10. Md Imran Hasan, Md Shahin Ali, Md Habibur Rahman, Md Khairul Islam, "Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks", Journal of Healthcare Engineering, vol. 2022, Article ID 5269913, 12 pages, 2022. https://doi.org/10.1155/2022/5269913