JOINT INSTITUTE FOR NUCLEAR RESEARCH

Laboratory of Information Technologies

FINAL REPORT ON THE START PROGRAMME

Studying Collagen Fibers Using Image Processing and Deep Learning Techniques

Supervisors:
Dr. Andrey Nechaevskiy
Dr. Mohamed Ibrahim Gaber

Student:
Ali HashHash, Egypt
Menoufia University

Participation period:
August 05 – September 20,
Summer Session 2025

Dubna, 2025

Abstract

This work develops a hybrid pipeline for analyzing collagen fibers in colon tissue images by merging classical image processing with deep learning. A custom module extracts initial fiber structures that then guide a U-Net model to achieve accurate segmentation of complex networks. A GUI enables visualization and quantification of fiber traits such as density and alignment, highlighting differences between healthy and cancerous tissues. The approach offers a practical step toward automated collagen analysis with direct relevance to cancer research and digital pathology.

Contents

1 Introduction **4**

1.1 Background and Motivation 4

1.2 Project Objectives 5

1.3 Example Images 5

2 Implementation and Results **6**

2.1 Collagen Fiber Detection 6

2.2 Experimental Results 7

2.3 Deep Learning Approach for Enhanced Segmentation 8

2.3.1 Concept and Idea 8

2.3.2 Implementation Overview 9

2.3.3 Deep Learning Results 10

3 Discussion **11**

4 Conclusion **11**

5 Acknowledgments **12**

6 References **13**

1 Introduction

1.1 Background and Motivation

Collagen fibers are the most abundant structural proteins in the extracellular matrix (ECM) of human tissues, playing a critical role in maintaining tissue integrity, regulating cell behavior, and influencing pathological processes such as cancer progression. In biomedical research, the analysis of the characteristics of collagen fibers—such as straightness, alignment, density, and orientation—provides valuable insight into tissue health and disease states. For example, alterations in the architecture of collagen are hallmarks of tumor microenvironments, where increased fiber straightness and alignment can promote cancer cell invasion, metastasis, and therapeutic resistance.

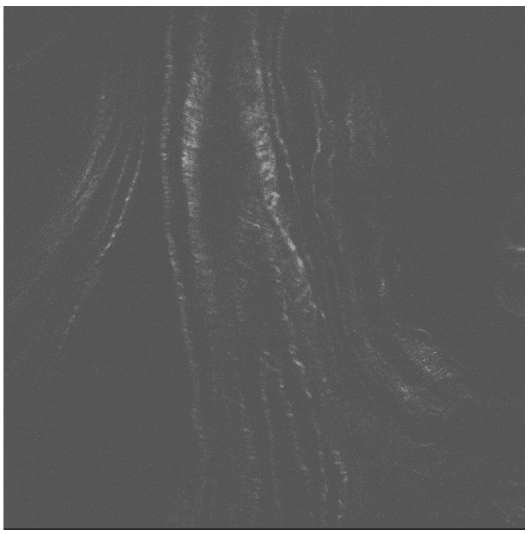
Manual analysis of collagen fibers from microscopic images is time-consuming, subjective, and prone to human error, limiting its scalability in clinical and research settings. The motivation for this project stems from the growing need for automated, accurate tools to segment and analyze collagen fibers in medical imaging. Traditional image processing techniques, such as thresholding, edge detection, and morphological operations, have been widely used for fiber extraction, but often struggle with noise, variability in image quality, and complex fiber patterns. Deep learning (DL) models, have revolutionized image segmentation by learning hierarchical features from data, achieving higher accuracy in challenging scenarios. By combining these approaches, this project aims to develop a hybrid system that enhances the precision of collagen fiber analysis, with potential applications in cancer diagnostics and prognosis.

1.2 Project Objectives

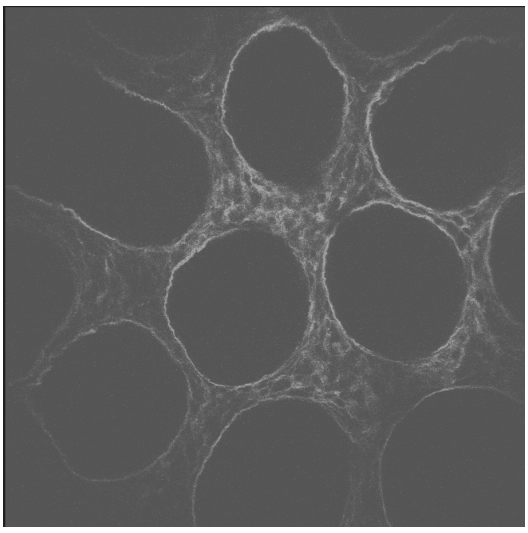
The primary objective of this project was to develop an automated system for collagen fiber segmentation and analysis, leveraging image processing techniques for initial feature extraction and DL for enhanced precision. Specific goals included:

- Implement traditional image processing methods (e.g., Otsu's thresholding, Canny edge detection, and morphological dilation/erosion) to identify and quantify fiber properties like straightness and alignment.
- Training DL model on annotated datasets to achieve precise segmentation, outperforming traditional methods in noisy or complex images.
- Evaluating the system's performance on real-world data, including images inspired by the morphological analysis in the referenced papers.
- Demonstrating applications in cancer-related contexts, such as detecting fiber straightness changes in tumor-adjacent tissues.

1.3 Example Images



Cancerous Tissue



Healthy Tissue

2 Implementation and Results

2.1 Collagen Fiber Detection

For the detection and analysis of collagen fibers, we implemented a Python class named `CollagenAnalyzer`. The pipeline integrates classical image processing with structural enhancement filters and line detection methods. The workflow consists of the following stages:

1. **Image Preprocessing:** Input images are converted to grayscale, denoised using non-local means filtering, and smoothed with Gaussian filtering. For SHG (Second Harmonic Generation) microscopy images, normalization ensures consistent intensity scaling.
2. **Fiber Enhancement:** The Frangi vesselness filter is applied to highlight elongated collagen structures. Otsu's thresholding is used to suppress background noise.
3. **Edge and Fiber Detection:** Canny edge detection extracts fiber boundaries. Morphological closing eliminates gaps. The probabilistic Hough transform detects linear fiber segments, allowing quantification of their count and width.
4. **Visualization:** Each processed image is saved with three stages: the original grayscale input, the detected edges, and the final overlay of extracted fibers in red.

This classical approach provides a robust baseline for fiber extraction, particularly effective for images with clear linear structures. Parameters such as Canny thresholds and Hough settings are adjustable.

2.2 Experimental Results

Figures 1 and 2 illustrate the results of two representative cases: healthy colon tissue (“No”) and cancerous colon tissue (“Yes”). Each figure shows the original image, the edges detected, and the final collagen fiber overlay. In healthy tissue, fibers appear to be more networked and curved, while cancerous samples show straighter, more aligned structures, in agreement with the literature on tumor microenvironments.

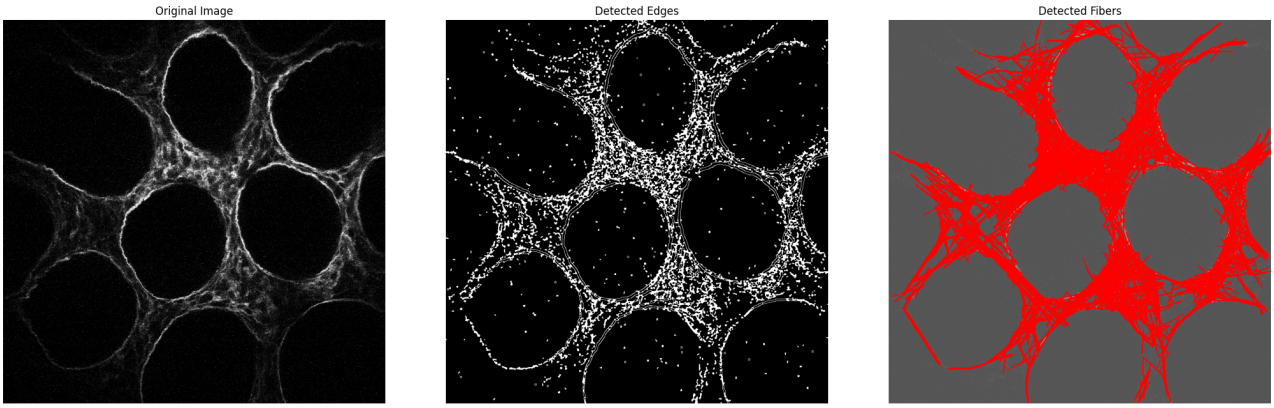


Figure 1: Pipeline output for a healthy colon tissue sample (No). Left: Original image, Middle: Detected edges, Right: Extracted collagen fibers.

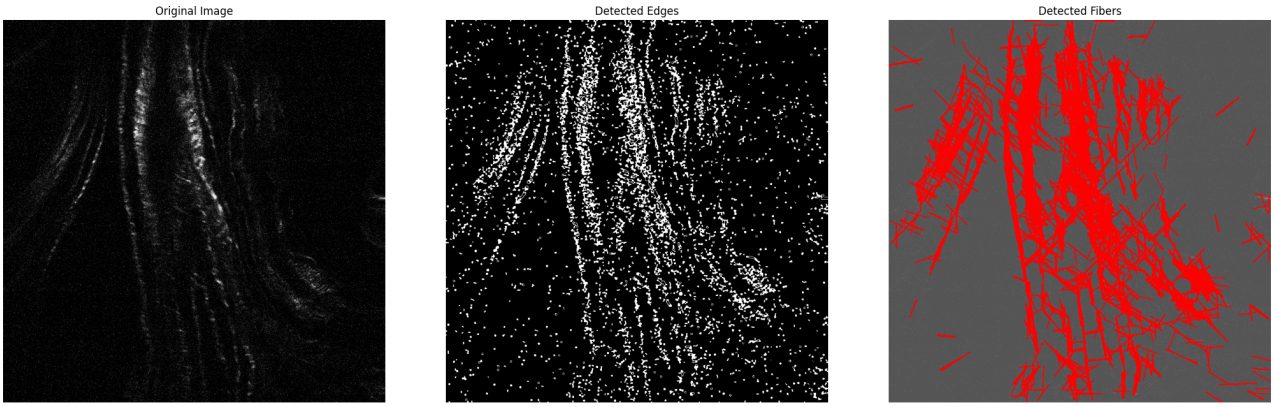


Figure 2: Pipeline output for a cancerous colon tissue sample (Yes). Left: Original image, Middle: Detected edges, Right: Extracted collagen fibers.

Quantitative metrics from batch processing can be derived from the output DataFrames. For the data sets analyzed, cancerous tissues typically exhibited higher fiber counts (indicating denser alignments) and narrower widths compared to healthy samples.

2.3 Deep Learning Approach for Enhanced Segmentation

To improve on the limitations of traditional image processing, such as sensitivity to noise, variability in fiber thickness, and challenges in distinguishing overlapping structures, we integrated a deep learning model based on the U-Net architecture. This approach enables end-to-end learning of complex features directly from the data, resulting in more accurate and robust segmentation of collagen fibers, especially in heterogeneous tissue images.

2.3.1 Concept and Idea

The U-Net model is a fully convolutional neural network specifically designed for semantic segmentation tasks in biomedical imaging. It features a symmetric encoder-decoder structure:
- The **encoder** (contracting path) consists of convolutional and max-pooling layers that progressively downsample the input image, capturing high-level contextual features while reducing spatial resolution.
- The **decoder** (expansive path) uses transposed convolutions to up-sample the feature maps, restoring the spatial dimensions, and enabling pixel-level predictions.
- **Skip connections** concatenate feature maps from the corresponding encoder layers to the decoder, preserving fine-grained details such as fiber boundaries that might be lost during downsampling.

The core idea is to train the model to predict a binary mask in which pixels belonging to collagen fibers are classified as the foreground (1) and the background as (0). This supervised learning approach leverages paired images and masks, allowing the network to learn intricate patterns, such as fiber curvature and intersections, that are difficult to capture with rule-based methods like thresholding or edge detection. By using a sigmoid activation in the final layer, the model outputs probability maps, which can be thresholded to generate crisp segmentations.

Post-prediction, we apply skeletonization to thin the segmented masks to single-pixel-wide representations of the fibers, facilitating quantitative analysis (e.g., length, orientation).
Connected components are then labeled, and each fiber is assigned a random color for visual distinction, aiding in the qualitative assessment of fiber networks.

This DL method is particularly advantageous for analyzing collagen in cancerous tissues, where fibers may exhibit increased density and alignment. The model's ability to generalize from training data makes it suitable for diverse imaging modalities, reducing the need for manual parameter tuning.

2.3.2 Implementation Overview

The implementation begins with dataset preparation: original grayscale images are resized to a fixed input size (e.g., 256x256 pixels) and normalized to the [0, 1] range. Initial masks are generated using traditional techniques (e.g., Gaussian blurring followed by Otsu's thresholding) to serve as ground-truth labels, which are also resized and binarized.

The U-Net is built with multiple convolutional blocks in the encoder and decoder, using ReLU activations and padding to maintain dimensions. The model is compiled with binary cross-entropy loss (suitable for binary segmentation) and the Adam optimizer for efficient training. Data is split into training and validation sets (e.g., 80/20 ratio), and the model is trained for a fixed number of epochs with a small batch size to handle memory constraints.

For inference, a new image is preprocessed similarly, fed through the trained model to obtain a predicted mask, and then postprocessed with skeletonization and coloring. A graphical user interface (GUI) was developed to streamline this process: Users select an image, and the system displays the original, the segmented (colored) mask, and an overlay blending the two for intuitive comparison.

This hybrid workflow, which combines DL for segmentation with classical post-processing, ensures both precision and interpretability, making it a powerful tool for histopathological analysis.

2.3.3 Deep Learning Results

DL-based segmentation results for the same healthy and cancerous samples as before. Compared to the traditional pipeline, the U-Net predictions exhibit finer detail in fiber extraction, with reduced noise and better handling of dense regions in cancerous tissues.

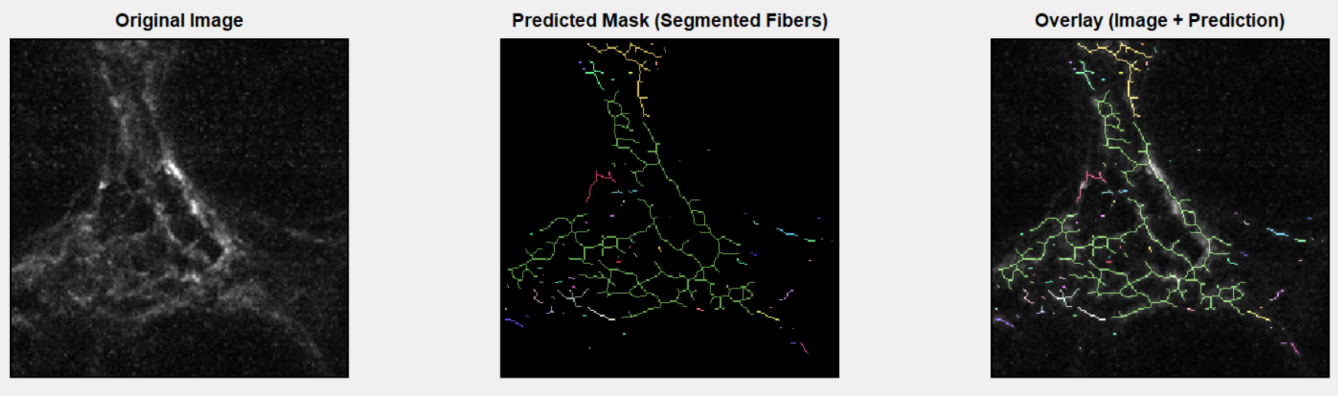


Figure 3: Deep learning output for a healthy colon tissue sample. Left: Original image, Middle: Predicted mask with segmented fibers (colored for distinction), Right: Overlay of image and prediction.

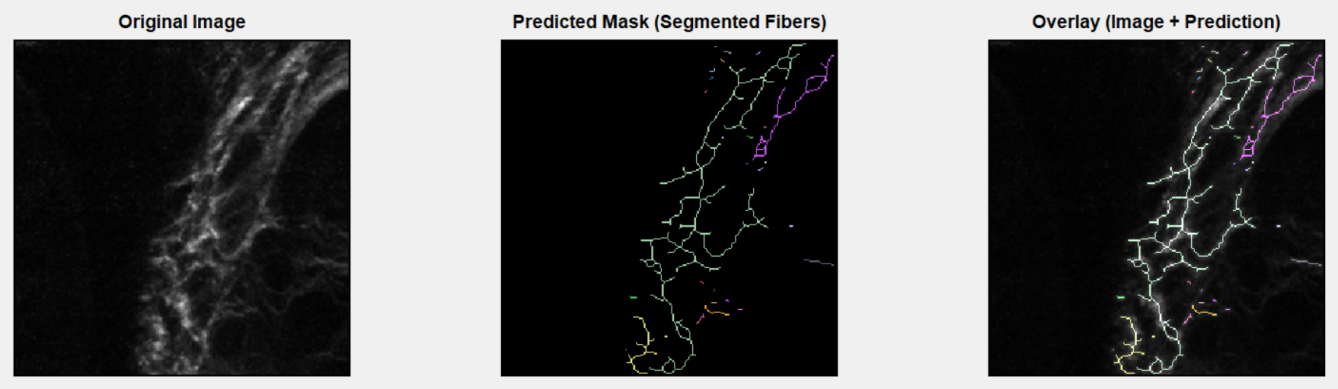


Figure 4: Deep learning output for a cancerous colon tissue sample. Left: Original image, Middle: Predicted mask with segmented fibers (colored for distinction), Right: Overlay of image and prediction.

3 Discussion

This work presents a hybrid framework that combines traditional image processing with deep learning for collagen fiber segmentation. The CollagenAnalyzer pipeline (Frangi filter, Otsu thresholding, Canny edge detection, Hough transform) provided initial masks that trained a U-Net model, enabling precise segmentation and robust quantification of fiber features. The results showed healthy tissues with curved fiber networks and cancerous tissues with straighter, aligned fibers, consistent with tumor-associated collagen remodeling. A GUI-supported intuitive visualization, while quantitative metrics highlighted fiber differences useful for diagnosis and prognosis. Despite limitations such as a small dataset, reliance on approximate masks, and focus on 2D images, the system demonstrated improved performance over classical methods. Future work includes expanding datasets, leveraging multi-modal imaging, and exploring advanced deep learning models.

4 Conclusion

This project developed a hybrid framework for automated collagen fiber segmentation, combining traditional image processing with deep learning. The CollagenAnalyzer pipeline applied denoising, normalization, Frangi filtering, Canny edge detection, and Hough transform to quantify fiber properties like density, alignment, and straightness. Building on this, a U-Net model achieved more precise segmentation, capturing complex patterns in both healthy and cancerous colon tissues. The results revealed a distinct fiber organization between tissue types, supporting the diagnostic potential of this approach. Overall, the framework demonstrates an effective integration of classical and deep learning methods, with applications in cancer diagnostics, prognosis, and ECM research.

5 Acknowledgments

I'm grateful to JINR and the START program for the incredible opportunity. Many thanks to my supervisor, Dr. Andrey Nechaevskii, for his guidance, and to Dr. Mohamed Ibrahim Gaber for always helping me overcome challenges. We also thank the START team for their support and patience. This experience has been truly inspiring, and I look forward to building on it.

6 References

References

[1] Bruekers, S. M. C., et al. (2024). Type III collagen suppresses tumor progression in breast cancer by promoting anti-tumor immunity and reducing tumor growth, invasion, and metastasis. *NPJ Breast Cancer*, 10, 69. https://www.nature.com/articles/s41523-024-00690-y

[2] Ghadami, A., et al. (2024). A new global morphological analysis method for scalar time-series data: Application to distinguishing healthy and pathological heart signals. *Nonlinear Dynamics*, 112, 12667–12683. https://link.springer.com/article/10.1007/s11071-024-09750-1

[3] Giampieri, E., et al. (2022). A shape-aware topological method for quantitative analysis of collagen fiber straightness in human colon tissues. *Frontiers in Physics*, 10, 915644. https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2022.915644/full

[4] Sanz-Pamplona, R., et al. (2020). Remodeling of collagen fiber architecture in normal-appearing colon mucosa near colorectal cancer. *Cancers*, 12(4), 999. https://pubmed.ncbi.nlm.nih.gov/32286443/