Machine Learning for Liver Fibrosis Assessment Using Sirius Red-Stained Pathology Slides

Yuyun Pan

Master's Degree Project in
Strategic Information Systems Management
Spring term 2025
Supervisor: Qi Dang,
Department of Computer
and Systems Sciences
Fahim Ebrahimi,
University Hospital Basel
Francisco Pena Escobar,
KTH Royal Institute of Technology

Stockholm
University

Abstract

Liver fibrosis is one of the key factors influencing disease progression and prognosis in patients with metabolic dysfunction-associated steatotic liver disease (MASLD). Although techniques such as SHG/TPE imaging combined with AI have shown promising accuracy in fibrosis quantification, their high cost and limited availability make them unsuitable for widespread clinical use. At the same time, routine histological staining methods like Sirius Red remain underutilized in AI-supported fibrosis analysis.

This thesis explores a practical and accessible approach by developing a machine learning pipeline to quantify collagen from Sirius Red–stained histopathology slides. The method focuses on stain-adaptive preprocessing and HSV-based color segmentation to generate continuous collagen proportionate area (CPA) metrics. Unlike prior classification-based models, this approach does not rely on predefined staging labels but instead aims to offer consistent and interpretable fibrosis indicators.

Experimental results based on clinical liver biopsy data showed a moderate but statistically significant correlation between the extracted CPA values and traditional METAVIR fibrosis stages. In a smaller subset with SHG imaging, the method also showed comparable collagen quantification patterns. Although some stage overlaps remain, visual analysis indicated potential for separating clinically relevant fibrosis groups such as F1 vs. F3 and F2 vs. F4.

The findings suggest that AI-assisted analysis of conventional Sirius Red slides can provide a feasible alternative to more advanced imaging systems. The approach may serve as a foundation for future fibrosis assessment tools that are both interpretable and suitable for routine clinical workflows. Limitations of this study include the moderate correlation with reference standards and the reliance on imbalance data from a single clinical center.

Acknowledgements

I would like to thank my supervisors Qi Dang, Fahim Ebrahimi, and Francisco Pena Escobar for their continuous guidance and support throughout the course of this thesis. I am particularly grateful to Fahim Ebrahimi and the medical team at the University Hospital of Basel for providing access to histopathology data, valuable medical background knowledge, and expert clinical insights, which greatly contributed to the relevance and direction of this research. I would also like to thank Francisco Pena Escobar for his technical guidance on the computational aspects of this work.

Abbreviations

Abbreviation	Description
CPA	Collagen Proportionate Area
DP	Digital Pathology
F1–F4	Fibrosis stages 1 through 4
H&E	Hematoxylin and Eosin (staining method)
HSV	Hue, Saturation, Value (color space)
MASLD	Metabolic Dysfunction-Associated Steatotic Liver Disease
METAVIR	Meta-analysis of Histological Data in Viral Hepatitis (fibrosis scoring system)
ML	Machine Learning
RGB	Red, Green, Blue (color space)
ROI	Region of Interest
SHG	Second Harmonic Generation
SHG%	Collagen percentage measured via SHG
TPE	Two-Photon Excitation
WSI	Whole Slide Image

List of Figures

3.1 Collagen Quantification Workflow for Sirius Red–Stained Tiles 12

4.1 Distribution of Pipeline-Derived CPA (% Collagen) by METAVIR Stage in Group A (Without SHG Reference) 19
4.2 Distribution of Pipeline-Derived CPA (% Collagen) by METAVIR Stage in Group B (With SHG Reference) 19
4.3 Distribution of SHG-Derived CPA (% Collagen) by METAVIR Stage in Group B 20

List of Tables

4.1 Spearman Correlations Between Pipeline- and SHG-Derived Collagen Quantification and METAVIR Staging 18

Contents

Abstract iii

Acknowledgements v

Abbreviations vii

List of Figures ix

List of Tables xi

Contents xiii

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Research Gaps . . . 3
- 1.3 Research Questions . . . 3
- 1.4 Research Objective . . . 4

2 Extended Background 5
- 2.1 Digital Pathology . . . 5
- 2.2 Related Work . . . 6

3 Method 9
- 3.1 Research Strategy . . . 9
- 3.2 Data Collection Method . . . 10
- 3.3 Data Analysis Method . . . 11
  - 3.3.1 Preprocessing . . . 11
  - 3.3.2 Collagen Quantification . . . 12
  - 3.3.3 Statistical Analysis . . . 14
- 3.4 Ethical Considerations . . . 15

4 Results **17**
4.1 Correlation with METAVIR and SHG 17
4.2 Visual Analysis of Collagen Distributions 18

5 Discussion **21**
5.1 Key Findings and Interpretation of Results 21
5.2 Result Interpretation and Supplementary Analysis 22
5.3 Strengths and Weaknesses 22
5.4 Limitations and Future Work 23
5.5 Usage of AI Tools 23

Bibliography **27**

1

Introduction

1.1 Motivation

Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) is a spectrum of liver diseases characterized by hepatic steatosis associated with metabolic dysfunction, such as obesity, type 2 diabetes, or dyslipidemia. Previously known as non-alcoholic fatty liver disease (NAFLD), MASLD was officially renamed in 2023 to eliminate the term "non-alcoholic" and better reflect its underlying pathophysiology [2]. It affects approximately 30% of the global adult population [20], with its prevalence increasing from 22% in 1991 to 37% in 2019 [2]. Given the rising incidence of unhealthy lifestyles, MASLD is expected to become even more widespread in the coming years.

Clinically, MASLD encompasses both simple steatosis, a benign form, and metabolic dysfunction-associated steatohepatitis (MASH), a less common but progressive variant that can lead to fibrosis, cirrhosis, and hepatocellular carcinoma [4]. Most patients remain asymptomatic, although fatigue, malaise, and mild right upper quadrant discomfort can occur. Even among patients progressing to cirrhosis, typical symptoms may be absent. Importantly, cardiovascular disease is the leading cause of death in MASLD patients, followed by liver-related complications. Liver fibrosis progression, in particular, serves as a critical prognostic factor.

Although hepatic steatosis and fibrosis can be non-invasively evaluated by imaging modalities such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), these techniques have inherent limitations. They are influenced by patient body composition and cannot

differentiate underlying etiologies of hepatic injury. Consequently, clinicians still rely on liver biopsy as the gold standard for diagnosing MASH and staging fibrosis severity.

There is an urgent need for comprehensive strategies to raise awareness and effectively address this challenge at local, regional, and global levels. While only about 25% of MASLD patients progress to liver fibrosis, which may eventually lead to cirrhosis or even liver cancer, early detection of disease deterioration and timely intervention play a crucial role in patient prognosis.

Currently, liver biopsy remains the gold standard for assessing liver fibrosis. However, conventional pathology diagnosis relies on manual slide reading by pathologists and semi-quantitative scoring systems, which significantly contribute to inter-observer and intra-observer variability, affecting diagnostic consistency and reliability. Studies [11] have shown that the agreement rate among pathologists is only 49.9%, with a low kappa consistency index (0.409) across different fibrosis stages. The kappa consistency index measures the level of agreement between raters beyond chance, with values ranging from 0 (no agreement) to 1 (perfect agreement). A value of 0.409 indicates moderate agreement. Even in biopsy samples larger than 1.5 cm, the agreement remains limited (kappa = 0.465). Additionally, community pathologists tended to underestimate fibrosis severity in 73% of discordant cases, with 26% of patients at stage 2–4 fibrosis being wrongly diagnosed. This indicates that even though patients undergo invasive and high-risk liver biopsies, the pathology results they receive may still be inconsistent, potentially leading to incorrect clinical decisions and suboptimal treatment plans.

Several liver fibrosis staging systems are used clinically, including the Ishak, METAVIR, and Batts-Ludwig systems [3]. These semi-quantitative methods provide useful frameworks for assessing fibrosis but are inherently subjective, relying heavily on pathologist interpretation. Factors such as tissue processing, staining variability, and observer experience can all influence scoring, contributing to inconsistency and diagnostic uncertainty.

Pathology slide analysis refers to the examination of stained tissue sections under a microscope to evaluate cellular and extracellular structures. Common staining techniques include Hematoxylin-Eosin (HE) staining, which highlights nuclei and cytoplasm, and Sirius red staining, which specifically marks collagen fibers. Advanced imaging modalities such as second

harmonic generation (SHG) and two-photon excitation (TPE) imaging provide high-resolution, label-free visualization of collagen architecture based on nonlinear optical properties.

In recent years, machine learning (ML) and artificial intelligence (AI) have made significant progress in the field of pathology, and more researchers are attempting to apply ML techniques to pathology slide analysis to improve diagnostic accuracy and consistency. For example, some studies [14, 13] have combined second harmonic generation (SHG) and two-photon excitation (TPE) imaging with ML, achieving promising results in the automated assessment of liver fibrosis. However, these advanced imaging techniques have high technical and financial barriers, making them impractical for routine clinical use.

On the other hand, Hematoxylin-Eosin (HE) staining and Sirius red staining remain the most widely used and cost-effective histopathological methods today. Despite their accessibility, how to leverage ML to improve the diagnostic capability of HE and Sirius red-stained slides, reduce human variability in interpretation, and enhance clinical assessment for MASLD and liver fibrosis remains an underexplored area. This study aims to fill this gap by exploring the integration of ML in standard pathology slide interpretation, with the goal of enhancing diagnostic accuracy, reducing observer variability, and ultimately advancing the intelligent management of MASLD.

1.2 Research Gaps

In recent years, digital pathology (DP) combined with artificial intelligence (AI) has been increasingly applied in the assessment of liver fibrosis, aiming to enhance diagnostic objectivity and consistency. Several AI-powered pathology tools have demonstrated promising clinical applications. Despite these advances, many existing solutions either depend on advanced imaging modalities such as SHG/TPE or still require extensive manual annotations, limiting their clinical scalability. There remains a significant gap in developing AI-based methods that can work effectively on routine HE and Sirius red-stained slides, which are low-cost and widely available.

1.3 Research Questions

This thesis addresses the following central research question:

RQ: Can machine learning applied to Sirius red-stained histopathology images provide an effective alternative to advanced imaging methods like SHG/TPE for fibrosis staging?

To explore this question, two sub-questions are considered:

- **Sub-RQ1:** Can a machine learning pipeline applied to Sirius red-stained histopathology slides provide robust and consistent collagen quantification across staining and scanner variations?
- **Sub-RQ2:** Can collagen quantification derived from ML analysis of Sirius red-stained slides effectively separate clinically relevant fibrosis stages such as F1 vs. F3 and F2 vs. F4?

1.4 Research Objective

This study aims to develop a machine learning pipeline for quantitative collagen assessment in Sirius red–stained liver pathology slides. The objective is to enable consistent and reproducible fibrosis quantification across variable staining and imaging conditions, offering a cost-effective foundation for further analysis.

2

Extended Background

2.1 Digital Pathology

Kiran's review[12] provides a detailed account of the transition from traditional glass slides to digital pathology (DP). Whole Slide Imaging (WSI) technology utilizes high-resolution scanners to capture entire pathology slides, accurately preserving complex cellular structures and subtle morphological features while supporting multi-magnification viewing. This eliminates the limitations of traditional optical microscopy, which relies on manual observation. Additionally, digital pathology allows for remote storage and sharing, overcoming constraints related to physical space and environmental conditions such as temperature control. This advancement has significantly facilitated teleconsultation among pathologists, enhanced collaboration, and contributed to the training of the next generation of pathologists. As a result, the adoption rate of digital pathology has risen rapidly, from 30% in 2013 to 53% in 2020.

At the same time, AI-driven pathology tools have been increasingly developed. Pavone et al. [15]summarized various open-source histological segmentation software, including ImageJ, CellProfiler, Ilastik, and QuPath. As mentioned earlier, pathologists must focus on specific regions of interest (ROI) and assess morphological features for grading or staging diseases. Before AI adoption, identifying these critical regions was labor-intensive and heavily reliant on pathologists' expertise. The aforementioned tools

enable rapid ROI detection and bulk data processing, significantly improving efficiency and reducing the workload of pathologists. However, fully automated segmentation remains challenging for complex tissue structures such as the liver, and further improvements in precision are required to extract more clinically valuable features.

Beyond segmentation, AI has also been widely applied in pathology image recognition, automatic classification, disease grading, and prediction. In recent years, deep learning (DL) has achieved remarkable progress in pathology image analysis, particularly through models such as convolutional neural networks (CNNs), vision transformers (ViTs), and generative adversarial networks (GANs). CNNs automate feature extraction, reducing reliance on handcrafted features in traditional machine learning approaches, while ViTs leverage global attention mechanisms to outperform CNNs in liver and prostate cancer classification. Meanwhile, GANs play a crucial role in data augmentation, generating high-quality synthetic pathology images to address the issue of limited medical datasets. Additionally, federated learning (FL) enables multi-institutional model training while preserving data privacy, enhancing the generalizability of AI-based diagnostic systems. These advancements are driving automation in pathology, facilitating precision medicine, and fostering the development of explainable AI (XAI) in medical applications.[8]

2.2 Related Work

In recent years, several AI-powered pathology tools have been proposed for liver fibrosis assessment. This section reviews the most relevant approaches and highlights their advantages and current limitations.

Several AI-powered pathology tools have demonstrated promising clinical applications, with notable approaches summarized by Ratziu [1]:

- **AIM-NASH** [16]: Utilizing digital pathology slides stained with Hematoxylin-Eosin (HE) and Masson's trichrome (MT), this method relies on pathologist-annotated data for supervised learning. While it improves scoring consistency, it remains highly dependent on manual labeling, making it prone to inter-observer variability.

- **FibroNest** [10]: This method processes HE, MT, Picrosirius Red (PSR), and immunohistochemistry (IHC)-stained pathology slides to conduct quantitative analysis of liver fibrosis. While effective in fibrosis detection, its accuracy in fibrosis staging still requires further improvement.
- **MorphoQuant**: Combining HE, PSR, and IHC-stained slides, this method employs rule-based and morphological analysis for automated feature recognition. However, it lacks direct fibrosis scoring capabilities, limiting its application in fibrosis staging.
- **qFibrosis** [14]: Unlike the above methods, qFibrosis does not require staining but instead utilizes SHG and TPE imaging to extract a large set of quantitative parameters and establish a continuous linear fibrosis scoring model. According to Naoumov [14], qFibrosis has demonstrated exceptional performance across multiple metrics:
  - Repeatability: 86–89%
  - Inter-pathologist agreement rate: 93%
  - Retention rate within pathologists: 95%
  - Reduction in third-reader adjudication: 30% lower than conventional methods

Despite the superior diagnostic consistency demonstrated by tools such as qFibrosis, their high cost and dependence on advanced imaging hardware pose significant barriers to routine clinical implementation. To further contextualize this study within the broader landscape of fibrosis quantification using conventional histological stains, we highlight several relevant studies that illustrate the diversity of approaches. Courtoy et al.[5] developed a robust digital image analysis workflow for Picrosirius Red–stained tissues across multiple organs, with attention to collagen topology and compactness. Serdjebi et al.[18] proposed a fully automated whole-slide image analysis method for liver fibrosis quantification, demonstrating consistency with biochemical reference standards. Dao et al.[6] demonstrated the use of Sirius Red–based morphometric analysis to assess interstitial fibrosis in renal transplant samples. Sánchez-Jaramillo et al.[17] introduced an image processing pipeline for rapid fibrosis scoring based on trichrome staining, emphasizing reproducibility and speed. Farzi et al.[9] recently released Liver-Quant, an open-source toolkit that applies handcrafted feature-based analysis to automatically quantify hepatic fibrosis and detect architectural patterns.

These studies reflect the growing interest in semi-automated and interpretable fibrosis assessment methods that leverage conventional staining. In contrast to end-to-end black-box models, our method emphasizes explicit, color-space–based features that are compatible with commonly available Sirius Red–stained slides and optimized for routine clinical settings.

This study builds on the principles underlying qFibrosis by aiming to achieve robust collagen quantification from conventional Sirius red-stained slides using machine learning techniques. Rather than reproducing fibrosis staging directly, the goal is to generate accurate collagen measurements that can inform downstream fibrosis assessment.

3

Method

This chapter describes the methodological framework used to develop and validate a machine learning–based pipeline for quantitative collagen assessment in liver biopsy slides from MASLD patients. It includes the rationale behind the experimental design, data sourcing and preprocessing strategy, and the implementation of HSV-based color analysis and clustering methods to generate collagen masks. The final goal is to produce a stain-adaptive and scanner-tolerant pipeline that can support consistent fibrosis quantification across brightfield pathology images. Although deep learning-based methods are widely applied in histopathological image analysis, this study opted for a clustering-based approach due to the limited dataset size and, more importantly, the absence of a robust ground truth baseline against which the performance of deep learning models could be reliably evaluated.

3.1 Research Strategy

To achieve the research objective of developing and validating a machine learning–based pipeline for rapid collagen quantification from pathology slides, with the capability to differentiate clinically challenging fibrosis stage pairs such as F1 vs F3 and F2 vs F4, the chosen research strategy must ensure methodological rigor, reproducibility, and statistical robustness. Among

the research methodologies outlined by Denscombe[7], experimental research is the most suitable for this study, as it provides a controlled environment to systematically evaluate the pipeline's performance under consistent data conditions, optimize preprocessing and model parameters, and assess its discriminatory ability between fibrosis stages.

This strategy allows the evaluation of stage separability using statistical methods that do not require absolute value comparability, such as the Mann–Whitney U test, which assesses differences in collagen quantification values between target stage pairs within the same method. This approach is appropriate because the inherent differences in sample preparation and measurement sensitivity between second harmonic generation (SHG) imaging and the proposed pipeline preclude direct comparison of their absolute collagen values.

Alternative strategies, such as case studies, were considered but rejected. Case studies focus on in-depth qualitative analysis of individual instances and are not designed to assess algorithmic performance across large pathology datasets or to provide statistically generalizable results. Given the aim to develop a broadly applicable, reproducible computational pipeline, the experimental research strategy was deemed the most appropriate.

3.2 Data Collection Method

This section outlines the data sources and collection procedures used in this study. The primary dataset is sourced from the University Hospital Basel, Switzerland, comprising pathology slides from patients diagnosed with MASLD. This dataset is used to develop a machine learning–based pipeline capable of rapidly quantifying collagen content from pathology slides, providing an efficient and objective tool for fibrosis assessment.

The dataset includes approximately 200 patients diagnosed with MASLD, each undergoing at least two liver biopsies to monitor disease progression. Pathological slides are stained using Hematoxylin-Eosin (HE) and Picrosirius Red (PSR) to provide cytological and fibrosis-related information. These slides have been annotated by pathologists following the Ishak scoring system, covering fibrosis stages from F0 (no fibrosis) to F4 (cirrhosis). This high-quality annotation serves as the reference standard for training and validating the machine learning pipeline.

All data are obtained from the University Hospital Basel database, with Institutional Review Board (IRB) approval, and are handled following strict

ethical guidelines and anonymization protocols to ensure patient privacy and data security.

To maintain high data quality and consistency, multiple preprocessing strategies are employed. Key preprocessing steps for image data include noise reduction, normalization, and tissue segmentation to minimize the impact of variability in staining and scanning. The ultimate goal is to establish a machine learning–based pipeline that can quantify collagen content from pathology slides with accuracy comparable to, or better than, second harmonic generation (SHG) imaging in differentiating fibrosis stages.

3.3 Data Analysis Method

The data analysis pipeline in this study is designed to emulate the methodology of qFibrosis [19], adapting its principles to Sirius red-stained histopathological slides. The goal is to test whether a similar fibrosis index can be constructed using conventional staining and machine learning methods.

3.3.1 Preprocessing

Preprocessing in this study refers to preparing high-resolution Sirius Red–stained whole slide images (WSIs) for subsequent collagen quantification. Each WSI, typically ranging from several hundred megabytes to over one gigabyte in size, was divided into non-overlapping tiles of fixed dimensions to enable efficient computation.

To remove background and non-informative regions, Otsu thresholding was applied to each tile to segment tissue foreground from background. The proportion of foreground pixels was then calculated, and tiles with less than 5% tissue coverage were discarded to minimize noise and reduce storage requirements. For retained tiles, all non-foreground regions were reassigned to pure white (RGB: 255, 255, 255) to further reduce background interference in subsequent processing.

Tiles located at the edges of the WSI were processed with zero-padding to maintain consistent tile dimensions. This ensured that edge tiles containing partial tissue regions were preserved for analysis while avoiding distortion or scaling artifacts.

The remaining tissue-containing tiles were saved in JPEG format, which offers a practical balance between file size and preservation of histological detail. This tiling and filtering process ensured that only relevant image regions were retained for further collagen mask generation and quantification.

3.3.2 Collagen Quantification

The first step in collagen quantification was to convert each tile from the RGB to the HSV color space. HSV is closer to human color perception than RGB and is particularly suitable for identifying collagen in pathology slides, where collagen typically appears as bright red. In HSV, the hue (H) channel captures the dominant wavelength and thus distinguishes color type, the saturation (S) channel measures the purity of the color, and the value (V) channel represents brightness. This separation makes it easier to detect collagen's characteristic bright-red signature. In contrast, RGB encodes colors as combinations of red, green, and blue intensities without a consistent pattern for visually similar colors, which makes it less effective for this task.

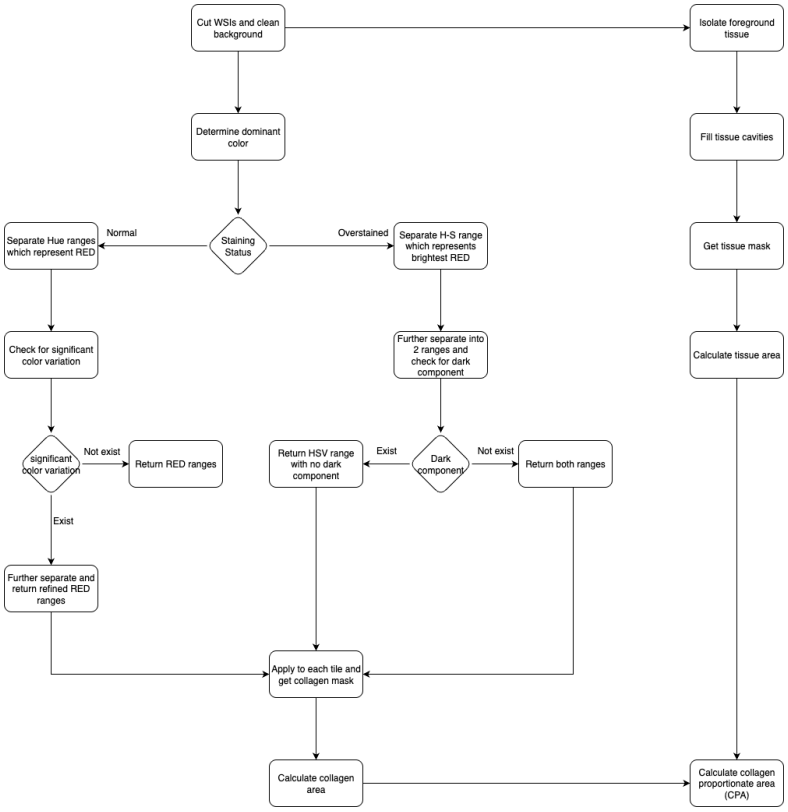

Figure 3.1: Collagen Quantification Workflow for Sirius Red–Stained Tiles

Figure.3.1 Collagen quantification workflow for Sirius Red–stained tiles. The pipeline uses hue-dominance detection (yellow vs. red) followed by iterative clustering to isolate collagen-specific hue–saturation ranges, which are combined to generate a binary mask. Collagen content is calculated

as the proportion of collagen pixels to total tissue pixels, and aggregated at the whole-slide level.

The first decision point in the workflow was the classification of the tile's dominant color. Variability in brightfield pathology staining can arise from differences in reagent quality, staining duration, and operator or mechanical factors. This variability affects the visual appearance of tissue and is one of the limitations of brightfield imaging compared to SHG, which requires no staining and thus avoids this source of inconsistency. In normally stained Sirius Red slides, liver parenchyma appears pale yellow, nuclei appear light red, and collagen appears vivid red. In abnormally stained slides, the entire tissue often takes on a red hue, collagen appears dark red, and nuclei may be dark red or black.

Considering that collagen content (red areas) varies across fibrosis stages, dominant color was determined by calculating the proportion of pixels within the yellow hue range (15–40) in HSV space. If more than 5% of the tissue foreground pixels fell into this range, the tile's dominant color was classified as yellow; otherwise, it was classified as red.

Yellow-dominant branch When the dominant color was classified as yellow, the tile was processed under the assumption of normal staining, where collagen appears as vivid red regions against a pale yellow parenchymal background. The hue channel was subjected to K-means clustering (k=5) to partition the pixels into five color groups.

In HSV, hue values are arranged on a circular scale from 0° to 360°; in the OpenCV implementation, this is scaled to 0–180. Red occupies a wrap-around position on this scale, appearing near both 0 and 180. As a result, the two clusters corresponding to vivid red tones are located at opposite ends of the hue axis. To accurately capture collagen, we therefore selected the clusters at both the low (near 0) and high (near 180) ends of the hue range.

To account for potential intra-tile color inconsistencies caused by local variations in staining or scanning, the selected red clusters were further evaluated. If inconsistency was detected—such as a wider-than-expected hue range or the presence of non-collagen colors—these pixels were subjected to a secondary clustering into five groups. This refinement allowed for more precise isolation of collagen-specific color ranges. The hue–saturation ranges corresponding to the final selected clusters were saved as the tile's collagen color profile for mask generation.

Red-dominant branch When the dominant color was classified as red, the tile was assumed to represent an abnormally stained slide, where the

overall tissue hue is reddish and collagen appears dark red. In such cases, nuclei may also appear dark red or even black, making it challenging to separate collagen from non-collagen structures using hue alone.

To improve discrimination, the first clustering step was performed in two-dimensional (H, S) space, where hue (H) captures color type and saturation (S) captures color intensity. K-means clustering with k=3 was applied, and the cluster corresponding to the brightest red—identified by high saturation and a hue near the red range—was selected as the initial collagen candidate group.

This group was then refined through a second clustering step using (H, S, V) data points, with k=2. Here, the value (V) channel was included to better distinguish between vivid collagen and darker non-collagen regions such as nuclei. To determine whether one of the subgroups corresponded to nuclei, all pixels in each subgroup were converted back to RGB. Since all points in this branch are red-hued, their green (G) and blue (B) channel intensities are low. By comparing the red (R) channel values between the two subgroups, we identified whether a subgroup contained markedly darker red or near-black pixels, indicative of nuclei.

If such a nucleus-containing subgroup was detected, it was excluded; otherwise, both subgroups were retained. The hue–saturation ranges of the final retained subgroups were stored as the collagen color profile. By applying these selected ranges to each tile, a binary collagen mask was generated, which was then used for quantitative collagen measurement.

3.3.3 Statistical Analysis

To evaluate the agreement and representativeness of the proposed collagen proportionate area (CPA) quantification pipeline against clinical fibrosis staging, all available slides were divided into two groups based on the available reference information:

- **Group A**: Slides with METAVIR staging but without SHG imaging. These slides were processed using the pipeline to compute collagen percentage (hereafter referred to as CPA), which was then compared against the corresponding fibrosis stages.
- **Group B**: Slides with both METAVIR staging and SHG-derived collagen percentages (SHG%), allowing for three-way comparisons among CPA, SHG%, and pathologist-assigned stages.

For Group A, Spearman's rank correlation coefficient (ρ) was used to assess the monotonic association between CPA and METAVIR staging.

For Group B, the above analysis was extended to include:

- Correlation between CPA and SHG%, evaluated using both Spearman's ρ and Pearson's correlation coefficient (r), to examine alignment between the proposed method and second harmonic generation imaging.

It is important to note that CPA and SHG% are derived through fundamentally different mechanisms—2D histological color analysis versus 3D optical signal response—and their absolute values are not directly interchangeable. All comparisons were therefore made independently with respect to METAVIR stage.

Finally, to investigate whether the pipeline has potential to distinguish between histologically challenging stage pairs (e.g., F1 vs. F3, F2 vs. F4), collagen percentage distributions were visualized using boxplots. Given the limited and imbalanced number of slides per stage, no formal statistical tests were performed at this stage.

3.4 Ethical Considerations

This study involves the use of pathology slides and patient clinical data, strictly adhering to ethical guidelines to ensure patient privacy and data compliance. The data originate from the University Hospital Basel, Switzerland, and have received approval from the Institutional Review Board (IRB), in compliance with the General Data Protection Regulation (GDPR). All data have been de-identified to remove personally identifiable information, ensuring privacy protection.

Due to the sensitivity of medical data, this study does not share data with third parties and restricts access to the research team only. All team members have signed data confidentiality agreements to ensure security and compliance.

During data analysis, this study follows principles of fairness and transparency. AI models are trained and evaluated on real pathology data using standardized statistical methods to mitigate algorithmic bias. Furthermore, to enhance model interpretability, the AI decision-making process will undergo visualization analysis and expert pathology review, reducing uncertainties associated with black-box models and ensuring clinical acceptability.

Overall, strict privacy protection and data security measures are implemented at all stages of data collection, storage, analysis, and publication to ensure ethical compliance while enhancing the transparency and reliability of AI models in clinical applications.

4

Results

This chapter presents the results of validating the collagen quantification pipeline (CPA) against established fibrosis assessment methods. The analysis is based on two groups: **Group A**, consisting of all Sirius Red–stained whole slide images (WSIs) with METAVIR staging but without SHG imaging, and **Group B**, a smaller subset of slides that underwent SHG imaging and thus contain both SHG-derived collagen percentage and METAVIR scores. SHG imaging provides label-free, collagen-specific contrast based on nonlinear optical properties, enabling direct extraction of collagen area without manual annotation. These SHG-derived values serve as the ground truth reference for evaluating pipeline CPA performance in Group B.

4.1 Correlation with METAVIR and SHG

To assess whether the pipeline-derived collagen percentage (CPA) reflects fibrosis severity, we computed the Spearman rank correlation coefficient (ρ) between CPA and METAVIR stage for each group. In Group B, additional correlations were computed between CPA and SHG-derived collagen percentages, as well as between SHG and METAVIR.

The results in Table 4.1 show that the CPA correlates significantly with METAVIR staging in both groups. Notably, the correlation in Group A is slightly higher than in Group B. Furthermore, the CPA also exhibits moderate correlation with SHG-based collagen percentage in Group B, supporting the biological consistency of the method across imaging modalities.

Table 4.1: Spearman Correlations Between Pipeline- and SHG-Derived Collagen Quantification and METAVIR Staging

Comparison	**Spearman** ρ	**Spearman** p
Pipeline CPA vs. METAVIR (Group A)	0.444	$<$0.0001
Pipeline CPA vs. METAVIR (Group B)	0.443	0.0006
SHG-Derived CPA vs. METAVIR (Group B)	0.362	0.019
Pipeline CPA vs. SHG-Derived CPA (Group B)	0.408	0.0016

4.2 Visual Analysis of Collagen Distributions

To visualize how collagen percentage varies across fibrosis stages, we plotted the distribution of CPA and SHG-derived values by METAVIR stage in each group. The corresponding boxplots are shown in Figures 4.1–4.3.

These visualizations confirm that CPA values tend to increase with fibrosis stage in both groups, despite some degree of overlap between adjacent stages. A similar trend is observed for SHG-derived collagen percentages in Group B. While group sizes are imbalanced and no significance testing was conducted, the overall stage-wise progression supports the use of CPA as a continuous fibrosis marker.

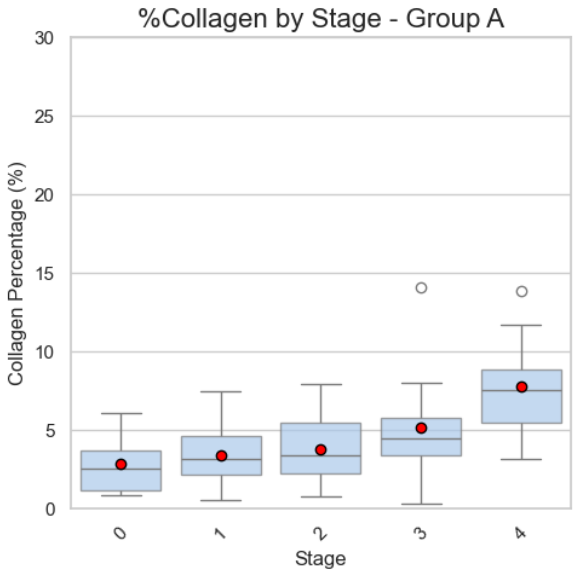

Figure 4.1: Distribution of Pipeline-Derived CPA (% Collagen) by METAVIR Stage in Group A (Without SHG Reference)

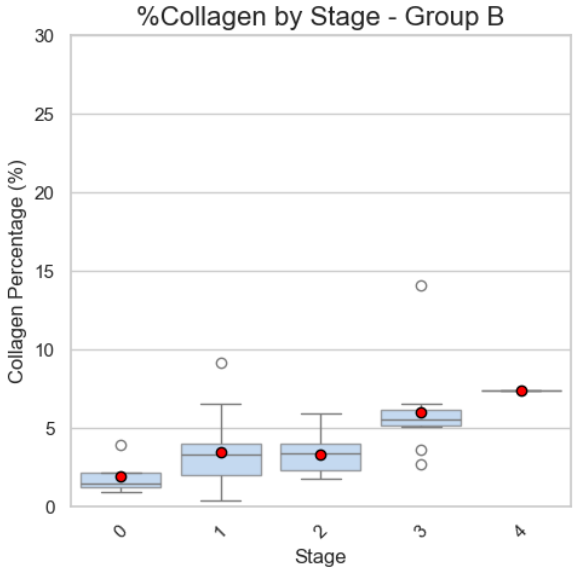

Figure 4.2: Distribution of Pipeline-Derived CPA (% Collagen) by METAVIR Stage in Group B (With SHG Reference)

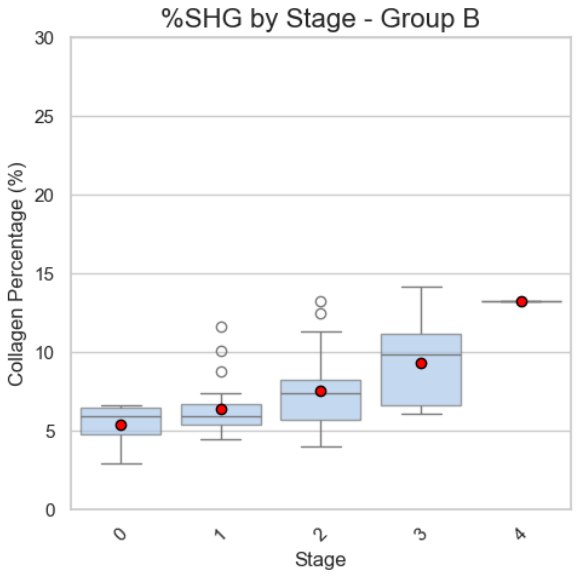

Figure 4.3: Distribution of SHG-Derived CPA (% Collagen) by METAVIR Stage in Group B

5

Discussion

5.1 Key Findings and Interpretation of Results

This study aimed to address the central research question: *Can machine learning applied to Sirius Red–stained histopathology images provide an effective alternative to advanced imaging methods like SHG/TPE for fibrosis staging?* Two sub-questions were considered to structure the analysis:

- **Sub-RQ1:** Can a machine learning pipeline applied to Sirius Red–stained slides provide robust and consistent collagen quantification across staining and scanner variations?
- **Sub-RQ2:** Can collagen quantification derived from ML analysis of Sirius Red–stained slides effectively separate clinically relevant fibrosis stages such as F1 vs. F3 and F2 vs. F4?

For **Sub-RQ1**, the most convincing evidence does not come solely from robustness against staining variations, but from the stability of the association between collagen proportion and staging across different subsets of data. In both the full set of slides without SHG imaging (Group A) and the subset with SHG imaging (Group B), the pipeline-derived collagen proportionate area (CPA) showed moderate correlations with METAVIR stage. In Group B, the CPA also demonstrated significant correlation with SHG-derived collagen, highlighting the reliability and consistency of the method.

For **Sub-RQ2**, boxplots were used to visualize CPA distributions across fibrosis stages. Although the sample size was limited and unevenly distributed across stages, preventing formal statistical testing, the visual results indicated that clinically relevant stage pairs such as F1 vs. F3 and F2 vs. F4 could be separated based on CPA values. These findings suggest that the proposed pipeline has potential to provide clinically meaningful discrimination of fibrosis severity.

5.2 Result Interpretation and Supplementary Analysis

An important observation is that SHG-derived collagen percentages were consistently higher than those obtained with the pipeline. This difference can be explained by the mechanisms of the two approaches: SHG activates collagen molecules and detects all fibrillar structures within the entire tissue thickness, whereas Sirius Red staining visualizes only the surface layer of the tissue section. Furthermore, SHG is more sensitive to newly formed fibrillar collagen, while conventional staining relies on the well-formed structure of collagen fibers, meaning that immature collagen may not be fully captured by Sirius Red.

Another point concerns the moderate correlations observed between CPA and METAVIR staging. This outcome reflects the pathological basis of METAVIR scoring, which considers not only the amount of collagen but also its architectural distribution. The presence of bridging fibrosis or nodular formations, for example, leads to a higher stage assignment (F3 or F4) regardless of the total collagen proportion in the slide. Additionally, in Group B, although CPA and SHG were both available, the paired slides were not cut from the same block at the same time. While representing the same patient and fibrosis stage, the sections are not identical, contributing further to the observed variability and limiting correlation strength.

5.3 Strengths and Weaknesses

This study demonstrates several strengths. The proposed pipeline is robust to variability in staining quality and scanner hardware, maintaining consistent performance across imaging conditions. It achieves comparable stage discrimination to SHG imaging while relying only on widely available brightfield microscopy and Sirius Red staining, making it cost-effective and

practical. In addition, the pipeline produces not only quantitative CPA values but also binary collagen masks, which provide a basis for more advanced morphological feature extraction in future studies.

Nevertheless, several limitations remain. The sample size was relatively small, and the distribution across fibrosis stages was uneven, reducing statistical power. SHG imaging, while valuable as a reference standard, was limited by its high cost, resulting in only a small subset of paired slides. Data were also sourced from a single institution and restricted to Sirius Red staining of MASLD-related fibrosis, limiting the generalizability of findings to other diseases or staining protocols.

5.4 Limitations and Future Work

Potential sources of variability—such as differences in staining intensity, tile-level artifacts, and inconsistencies introduced during slide scanning—may have influenced CPA measurements. These factors should be systematically addressed in future validations to further improve robustness and generalizability. In addition, future work should validate the pipeline on larger, multi-center datasets to confirm its stability across populations, staining protocols, and imaging systems. Extending the evaluation to fibrosis caused by other etiologies and to slides stained with alternative histological methods will further establish its broad applicability.

Moreover, the binary collagen masks generated by the pipeline offer a strong foundation for developing deep learning models that integrate both quantitative collagen proportion and structural features. Such models could improve diagnostic accuracy, support clinical decision-making, and allow more refined patient stratification by fibrosis risk. Finally, integrating multimodal data—combining histopathology with clinical and molecular information—represents a promising avenue for advancing precision medicine in liver disease assessment.

5.5 Usage of AI Tools

This thesis benefited from the use of large language models (LLMs), including ChatGPT, which were employed to assist in literature structuring, technical phrasing, and language refinement. All content was critically reviewed and finalized by the author to ensure academic integrity and consistency.

Bibliography

[1] Vlad Ratziu et al. “Artificial intelligence-assisted digital pathology for non-alcoholic steatohepatitis: current status and future directions”. In: *Journal of Hepatology* 80.2 (Feb. 2024), pp. 335–351.

[2] W. K. Chan et al. “Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): A State-of-the-Art Review”. In: *Journal of Obesity Metabolic Syndrome* 32.3 (Sept. 2023). Epub 2023 Sep 13, pp. 197–213. DOI: 10.7570/jomes23052.

[3] A. B. Chowdhury and K. J. Mehta. “Liver biopsy for assessment of chronic liver diseases: a synopsis”. In: *Clinical and Experimental Medicine* 23 (2023). Published online 22 February 2022, pp. 273–285. DOI: 10.1007/s10238-022-00799-z. URL: https://doi.org/10.1007/s10238-022-00799-z.

[4] Merck Sharp Dohme Corp. *The Merck Manual Online*. Accessed: 2025-02-20. 2025. URL: https://www.merckmanuals.com/.

[5] Pierre J Courtoy et al. “Digital Image Analysis of Picrosirius Red Staining: A Robust Method for Multi-Organ Fibrosis Quantification and Characterization”. In: *Biomolecules* 10.11 (2020), p. 1585. DOI: 10.3390/biom10111585. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709042/.

[6] Myriam Dao et al. “Usefulness of morphometric image analysis with Sirius Red to assess interstitial fibrosis after renal transplantation from uncontrolled circulatory death donors”. In: *Scientific Reports* 10.1 (2020). Published online 2020-04-23, p. 6894. ISSN: 2045-2322. DOI: 10.1038/s41598-020-63749-3. URL: https://doi.org/10.1038/s41598-020-63749-3.

[7] Martyn Denscombe. *The Good Research Guide: For Small-Scale Social Research Projects*. Fifth. Maidenhead, England: Open University Press, McGraw-Hill Education, 2014. ISBN: 978-0-335-26470-4.

[8] Naira Elazab et al. “Objective Diagnosis for Histopathological Images Based on Machine Learning Techniques: Classical Approaches and New Trends”. In: *Mathematics* 8.11 (2020), p. 1863. DOI: 10.3390/math8111863. URL: https://www.researchgate.net/publication/345436316.

[9] Reza Farzi et al. “Liver-Quant: Feature-Based Image Analysis Toolkit for Automatic Quantification and Architectural Pattern Detection of Hepatic Fibrosis”. In: *Computers in Biology and Medicine* 171 (2025), p. 107598. DOI: 10.1016/j.compbiomed.2025.107598. URL: https://www.sciencedirect.com/science/article/pii/S0010482525004007.

[10] Radina Kostadinova et al. “Digital pathology with artificial intelligence analysis provides insight to the efficacy of anti-fibrotic compounds in human 3D MASH model”. In: *Scientific Reports* 14 (Mar. 2024). DOI: 10.1038/s41598-024-55438-2.

[11] Robert M et al. “A comparison of hepatopathologists’ and community pathologists’ review of liver biopsy specimens from patients with hepatitis C”. In: *Clin Gastroenterol Hepatol* 7.3 (Mar. 2009). Epub 2008 Dec 13, pp. 335–338. DOI: 10.1016/j.cgh.2008.11.029.

[12] Kiran N. et al. “Digital Pathology: Transforming Diagnosis in the Digital Age”. In: *Cureus* 15.9 (Sept. 2023). Published online 3 September 2023, e44620. DOI: 10.7759/cureus.44620.

[13] Nikolai V Naoumov and Elaine Chng. “Second harmonic generation digital pathology with artificial intelligence: breakthroughs in studying fibrosis dynamics and treatment response”. In: *Future Medicine AI* 2.2 (Aug. 2024). DOI: 10.2217/001c.121609.

[14] Nikolai V. Naoumov et al. “Digital pathology with artificial intelligence analyses provides greater insights into treatment-induced fibrosis regression in NASH”. In: *J Hepatol* 77.5 (Nov. 2022). Epub 2022 Jun 30, pp. 1399–1409. DOI: 10.1016/j.jhep.2022.06.018.

[15] Anna Maria Pavone et al. “Digital Pathology: A Comprehensive Review of Open-Source Histological Segmentation Software”. In: *BioMedInformatics* 4.1 (2024), pp. 173–196. ISSN: 2673-7426. URL: https://www.mdpi.com/2673-7426/4/1/12.

[16] H. Pulaski, S. A. Harrison, S. S. Mehta, et al. “Clinical validation of an AI-based pathology tool for scoring of metabolic dysfunction-associated steatohepatitis”. In: *Nature Medicine* 31 (2025). Published online, pp. 315–322. DOI: 10.1038/s41591-024-03301-2.

[17] Eliana Sánchez-Jaramillo et al. “Automated Computer-Assisted Image Analysis for the Fast Quantification of Fibrosis”. In: *Biology* 11.8 (2022), p. 1227. DOI: 10.3390/biology11081227. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9404825/.

[18] Charlotte Serdjebi et al. “Automated Whole Slide Image Analysis for a Translational Quantification of Liver Fibrosis”. In: *Scientific Reports* 12.1 (2022), p. 18923. DOI: 10.1038/s41598-022-22902-w. URL: https://www.nature.com/articles/s41598-022-22902-w.

[19] Shuoyu Xu et al. “qFibrosis: A fully-quantitative innovative method incorporating histological features to facilitate accurate fibrosis scoring in animal model and chronic hepatitis B patients”. In: *Journal of Hepatology* 61.2 (2014), pp. 260–269. DOI: 10.1016/j.jhep.2014.03.006.

[20] Zobair M. Younossi et al. “The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review”. In: *Hepatology* 77.4 (Apr. 2023). Epub 2023 Jan 3, pp. 1335–1347. DOI: 10.1097/HEP.0000000000000004. eprint: https://journals.lww.com/hepatology/fulltext/2023/04000/the_global_epidemiology_of_non_alcoholic_fatty.4.aspx.