Politecnico di Torino

Master's Degree in Biomedical Engineering

Graduation Session April 2026

Quantitative Radiomics-Based Approach for Preoperative Prediction of Tumor Consistency and Intraoperative Adherence in Vestibular Schwannomas

Supervisors:
Prof. Kristen Mariko MEIBURGER
Dr. Raffaele DE MARCO
Ing. Alen SHAHINI
Ing. Bruna COTRUFO

Candidate:
Giovanni SPARACINO

Summary

Vestibular schwannomas are benign tumors arising from Schwann cells of the vestibular branch of the eighth cranial nerve and represent the most common neoplasm of the cerebellopontine angle. Their progressive growth may lead to sensorineural hearing loss, tinnitus, imbalance, and, in advanced cases, compression of brainstem structures. Management options include observation, radiosurgery, and microsurgical resection, with the aim of achieving local tumor control while preserving facial nerve function and adjacent neurological structures.

Among the factors influencing surgical strategy and postoperative outcome, tumor consistency and adherence to surrounding neural structures play a crucial role. However, conventional magnetic resonance imaging (MRI) assessment remains largely qualitative and does not directly quantify microstructural properties related to tumor stiffness and fibrous content. In this context, an objective preoperative characterization of tumor biomechanical and structural properties becomes essential for surgical planning and outcome prediction.

The aim of this thesis is to investigate whether quantitative descriptors derived from MRI can provide non-invasive information for the characterization and prediction of clinically relevant tumor properties in vestibular schwannomas. To support this objective, a dedicated and reproducible histological pipeline was developed for the automated quantification of collagen content from hematoxylin and eosin (H&E) whole-slide images. The workflow combined stain normalization using a GAN-based approach, stain deconvolution, percentile-based thresholding, chromatic filtering, and structure tensor–based orientation analysis to selectively retain organized fibrous collagen. The validity of the H&E-based collagen quantification was assessed against Masson's trichrome staining, used as histological reference standard, demonstrating a strong correlation (Pearson $r = 0.89$) and acceptable agreement between methods.

Histologically derived collagen percentage was then employed as a biologically grounded surrogate to define tumor consistency classes through an optimal cutoff strategy. Radiomic features were extracted from preoperative T2-weighted MRI and apparent diffusion coefficient (ADC) maps following standardized preprocessing, spatial alignment, and intensity normalization. Linear support vector machine

(SVM) models with stratified cross-validation were trained to predict tumor consistency, intraoperative adherence, and postoperative facial nerve outcome.

Radiomics-based models achieved good discriminatory performance for tumor consistency (AUC up to 0.82) and adherence prediction (AUC up to 0.80), while performance for facial nerve outcome prediction was moderate. Feature importance analysis indicated that texture descriptors related to signal heterogeneity and spatial organization were the most influential in model decision, suggesting that radiomic patterns may reflect underlying microstructural differences associated with collagen distribution and tissue compactness.

In conclusion, the proposed framework integrates quantitative histology and MRI-based radiomics within a biologically informed setting. The results support the potential of radiomics as a non-invasive tool for the preoperative characterization of tumor properties relevant to surgical planning. Although limited by sample size and retrospective design, this study provides a methodological and exploratory foundation for future larger-scale investigations aimed at translating quantitative imaging biomarkers into clinical decision support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Clinical background

Schwannomas are benign tumors arising from Schwann cells and may develop along peripheral and cranial nerves [1]. In the cranial setting, a particularly relevant location is the cerebellopontine angle, where vestibular schwannoma represents one of the most common entities and may involve structures that are critical for hearing and balance [1]. The clinical management of these lesions requires careful preoperative assessment and a balance between disease control and functional preservation [1].

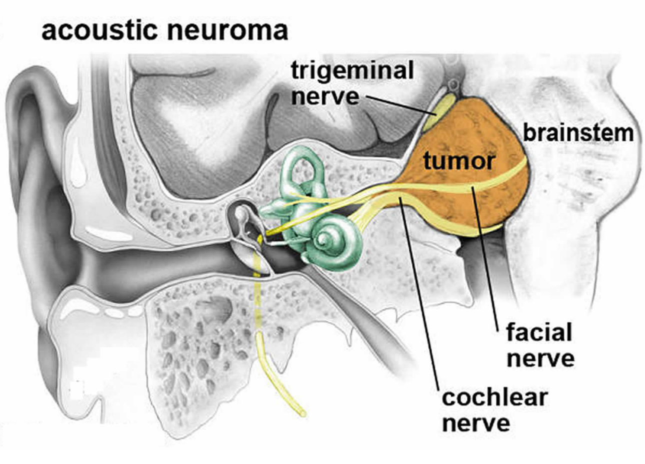


Figure 1.1: Schematic representation of a vestibular schwannoma and its anatomical relationship with the facial and cochlear nerves in the cerebellopontine angle. The close proximity between the tumor and critical neural structures highlights the relevance of preoperative tumor characterization for surgical planning.

As illustrated in Figure 1.1, vestibular schwannomas often develop in close anatomical proximity to critical neural structures, including the facial nerve and the vestibulocochlear complex [1]. From a surgical perspective, one of the most clinically relevant factors for operative planning is tumor consistency, which can vary substantially across patients [2]. In particular, lesions characterized by a firmer or more fibrous consistency may be more challenging to dissect and separate from adjacent neural structures, with potential implications for operative complexity and for the risk of postoperative deficits [2]. This aspect is especially relevant when highly functional nerves are involved, such as the facial nerve and the vestibulocochlear complex, for which anatomical and functional preservation is a primary objective.

Traditionally, vestibular schwannomas have been classified into solid and cystic subtypes based on their radiological appearance [3]. This distinction has been widely used in clinical practice and has been associated with different surgical outcomes, with cystic lesions often reported to have less favorable facial nerve preservation [3, 4]. However, the lack of a standardized definition of cystic vestibular schwannomas and the presence of conflicting evidence regarding tumor adherence and surgical complexity have highlighted the limitations of this dichotomous classification [3, 4]. Notably, increased tumor adherence has also been reported in radiologically solid lesions, suggesting that qualitative interpretation of conventional MRI alone may not fully capture the mechanical and microstructural properties relevant for surgical planning [5].

1.2 Limitations of conventional MRI assessment

Magnetic resonance imaging (MRI) represents the reference modality for the clinical evaluation of schwannomas, providing high soft-tissue contrast and detailed anatomical information [1]. Conventional MRI protocols routinely include T1-weighted and T2-weighted sequences, often complemented by contrast-enhanced imaging, which allow accurate lesion detection, volumetric assessment, and evaluation of the relationship between the tumor and surrounding anatomical structures [1]. More recently, diffusion-weighted imaging and derived apparent diffusion coefficient (ADC) maps have been increasingly investigated as additional sources of microstructural information [6].

Despite its central role in clinical practice, MRI assessment of schwannomas remains predominantly qualitative. Radiological interpretation is largely based on visual inspection of signal intensity patterns and morphological characteristics, which may provide indirect cues regarding tumor composition but do not directly quantify underlying tissue properties. As a result, clinically relevant parameters for surgical planning, such as tumor consistency, fibrous content, collagen organization, and adherence to surrounding neural structures, are not explicitly measured and are instead inferred from subjective experience or intraoperative findings [2, 6]. Consequently, tumors that appear similar on conventional imaging may exhibit markedly different mechanical behavior during surgical manipulation.

These limitations highlight a critical gap between routinely acquired MRI data and the clinical need for objective, reproducible, and quantitative biomarkers capable of characterizing tumor tissue properties preoperatively. Addressing this gap requires methodologies that go beyond qualitative image interpretation and leverage the rich information embedded within MRI data to extract descriptors more closely related to the biological and structural characteristics of the lesion [6].

1.3 Tumor Mechanical Behavior in MRI

Tumor mechanical behavior represents a clinically relevant parameter in vestibular schwannomas, particularly in relation to surgical management. Intraoperative tumor consistency directly influences debulking strategy, ease of dissection, operative time, and the risk of traction-related nerve injury [2]. Despite its importance, tumor stiffness remains primarily assessed intraoperatively and is not routinely quantified preoperatively using standard imaging protocols.

Magnetic resonance elastography (MRE) has emerged as a technique capable of directly measuring tissue mechanical properties in vivo. Recent investigations have demonstrated that biomechanical stiffness measured by MRE correlates with surgical consistency and may provide objective information regarding tumor mechanical behavior [7]. However, elastography requires dedicated acquisition protocols, additional hardware, and longer examination times, which limit its routine implementation in standard clinical practice. Consequently, the identification of surrogate imaging biomarkers capable of reflecting tumor stiffness using routinely acquired MRI sequences remains an area of active interest.

At the biological level, tumor stiffness is largely influenced by the composition and organization of the extracellular matrix (ECM). Collagen represents the principal structural component of the ECM and plays a central role in determining tissue mechanical properties, including firmness and resistance to deformation [8]. Increased collagen deposition, fiber cross-linking, and spatial organization contribute to greater stromal density and reduced tissue compliance. In intracranial tumors such as meningiomas, T2-weighted signal intensity has been shown to correlate with collagen content, with fibrous tumors demonstrating lower T2 signal due to dense collagenous matrices [9]. These observations support the concept that MRI signal characteristics may indirectly reflect underlying stromal composition.

Beyond conventional T2-weighted imaging, diffusion-weighted imaging provides additional insight into tissue microstructure. The apparent diffusion coefficient

(ADC) reflects the mobility of water molecules within tissue and is influenced by multiple factors, including cellular density, extracellular matrix composition, and structural organization. A large meta-analysis demonstrated an overall moderate inverse correlation between ADC values and tumor cellularity across different tumor types, while also highlighting substantial inter-tumoral variability [10]. Importantly, the heterogeneity of these correlations suggests that diffusion behavior is not determined solely by cell density but is modulated by additional histopathological features such as extracellular matrix content and stromal architecture.

Taken together, these observations provide a biological rationale for investigating routinely acquired MRI sequences as potential indirect markers of tumor mechanical properties. T2-weighted imaging may capture variations in water content and fibrous matrix density, while ADC maps may reflect the combined effects of cellular compactness and extracellular structural constraints on water diffusion. Although neither sequence directly measures collagen or stiffness, their sensitivity to microstructural organization supports their exploration as non-invasive surrogates of tumor mechanical behavior in vestibular schwannomas.

Within this conceptual framework, quantitative analysis of T2-weighted images and ADC maps may enable the extraction of imaging descriptors more closely linked to underlying stromal composition, thereby providing a biologically informed basis for the subsequent radiomics-based investigation of tumor consistency and related surgical endpoints.

1.4 Radiomics and histology as quantitative surrogates

Radiomics has emerged as a quantitative imaging paradigm aimed at extracting a large number of numerical features from standard medical images, with the goal of capturing tissue characteristics that are not readily appreciable through visual inspection alone [11]. By analyzing signal intensity distributions and spatial relationships between voxels, radiomic features provide descriptors of image heterogeneity, texture, and structural organization, thereby enabling a more objective characterization of lesion properties.

In the context of magnetic resonance imaging, radiomics enables the exploitation of routinely acquired sequences to derive quantitative information related to underlying tissue composition. Texture-based features extracted from T2-weighted images and apparent diffusion coefficient (ADC) maps have been investigated as potential indicators of microstructural organization, reflecting variations in cellularity, extracellular matrix composition, and tissue compactness [12, 13]. These properties are clinically relevant, as they may influence tumor consistency and adherence to surrounding neural structures, thereby affecting surgical complexity. Radiomics therefore represents a promising non-invasive approach to bridge the gap between qualitative image interpretation and the need for quantitative biomarkers relevant to preoperative decision-making.

While imaging-derived features provide indirect information, histological analysis remains the reference standard for the assessment of tissue composition at the microscopic level. In particular, collagen organization provides a biologically grounded reference for tumor mechanical behavior [7]. Quantitative histological assessment therefore serves as a biological reference for the validation and interpretation of imaging-derived descriptors associated with tumor consistency [13].

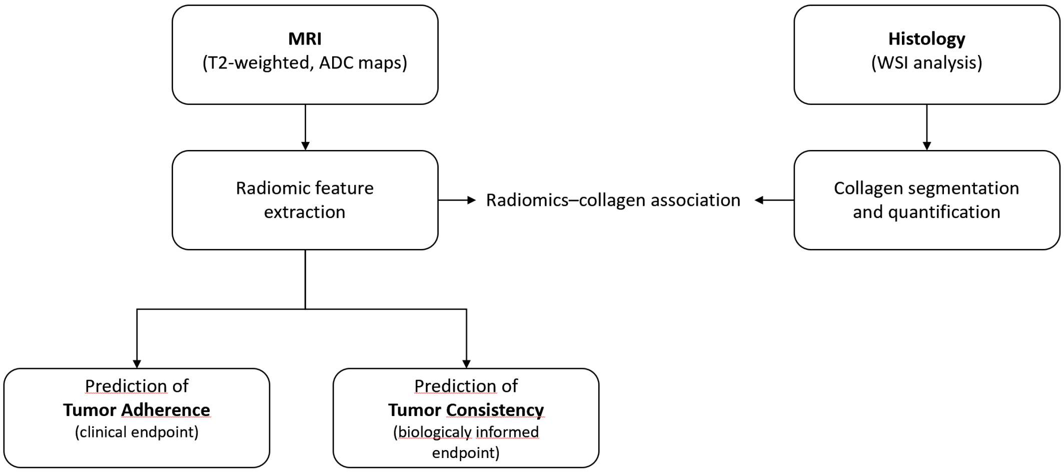


Figure 1.2: Conceptual framework of the study. Radiomic features are extracted from routinely acquired MRI sequences (T2-weighted images and ADC maps) and used for the prediction of clinically relevant tumor properties, including tumor consistency and intraoperative adherence. In parallel, quantitative histological collagen analysis is performed on whole-slide images to provide a biological reference. The association between radiomic descriptors and collagen content enables biological validation of imaging-derived features, particularly with respect to tumor consistency.

The integration of radiomics and histology enables a framework in which quantitative imaging features can be investigated in relation to objective biological markers [13]. Within this framework, histologically derived collagen content is primarily used as a biologically grounded reference to interpret and validate radiomic descriptors associated with tumor consistency. As illustrated in Figure 1.2, radiomics-based models are developed using MRI-derived features, while histological collagen quantification provides complementary biological insight. This approach establishes a structured link between imaging, tissue microstructure, and clinically relevant tumor properties.

1.5 Aim and structure of the thesis

The aim of this thesis is to investigate whether quantitative descriptors derived from magnetic resonance imaging can provide non-invasive information for the characterization and prediction of clinically relevant tumor properties in vestibular schwannomas.

To support this investigation, a first objective of the study is the design and validation of a reproducible histological workflow for the quantitative extraction of collagen content from tumor specimens. Histologically derived collagen content is employed as a biologically grounded surrogate reference for tumor tissue characterization and for the definition of clinically meaningful tumor consistency classes through discretization strategies.

Based on this biological reference, the main objective of the thesis is to evaluate the ability of MRI-based radiomic features extracted from T2-weighted images and apparent diffusion coefficient (ADC) maps to predict tumor consistency using radiomic information alone. In addition, the predictive value of radiomic descriptors is investigated for intraoperative tumor adherence as a clinically relevant extension of tumor mechanical behavior. As a secondary and exploratory objective, the potential association between radiomic features and early post-operative facial nerve outcome is also examined, acknowledging the multifactorial nature of functional recovery.

This thesis is structured as follows. Chapter 2 describes the dataset and details the methodological pipeline, with particular emphasis on the histological processing and collagen quantification workflow, as well as imaging acquisition, preprocessing, and radiomic feature extraction. Chapter 3 presents the results regarding the validation of the histological analysis and the radiomics-based predictive models for clinically relevant tumor properties. Chapter 4 discusses the findings in the context of existing literature, highlighting their clinical implications, methodological limitations, and perspectives for future research.

Chapter 2

Materials and Methods

2.1 Dataset and patient selection

The dataset employed in this thesis consists of patients affected by vestibular schwannoma who underwent surgical resection at a single tertiary referral center. The data were originally collected within the context of a previous clinical research project focused on the preoperative assessment of vestibular schwannoma consistency and its surgical implications [14].

The initial clinical cohort comprised 57 patients with histologically confirmed vestibular schwannoma.

In the present thesis, the same clinical cohort is reused with distinct and complementary objectives, specifically oriented toward the investigation of MRI-based radiomics and its relationship with histological and clinical outcomes. For this reason, the inclusion and exclusion criteria are study-dependent and vary according to the specific analysis performed.

Preoperative MRI data were available for a subset of patients:

- 45 patients had T2-weighted MRI volumes suitable for radiomic analysis,
- 40 patients had available apparent diffusion coefficient (ADC) maps.

All patients included in the radiomic analyses had corresponding histological slides available for collagen quantification. Therefore, radiological–histological integration was feasible for all subjects with MRI data.

For histological analyses, patient inclusion depended on the availability and quality of slides suitable for collagen quantification. For radiomic analyses, inclusion criteria were based on the availability of the required MRI sequences (T2-weighted volumes and/or ADC maps) and adequate image quality for segmentation and preprocessing.

However, not all patients contributed data to all classification tasks. Specific analyses (e.g., intraoperative adherence and postoperative facial nerve outcome) required the availability of corresponding clinical labels. Consequently, the effective sample size varied depending on the completeness of outcome data.

In analyses requiring the integration of radiomic and histological information, only patients meeting both sets of criteria were considered. This modular approach allowed each analysis to be performed on the largest possible subset of patients while preserving methodological rigor and data consistency.

Detailed information regarding patient selection, cohort sizes for each specific outcome, and exclusion criteria is provided in the corresponding methodological sections and in the Results chapter.

2.2 MRI data processing pipeline

This section describes the magnetic resonance imaging (MRI) data used in this study and the preparatory steps applied prior to radiomic feature extraction. The focus is on the imaging modalities employed, the strategy adopted for tumor delineation and multimodal alignment, and the operations required to ensure spatial and intensity consistency across MRI volumes [15].

2.2.1 MRI acquisition and imaging protocol

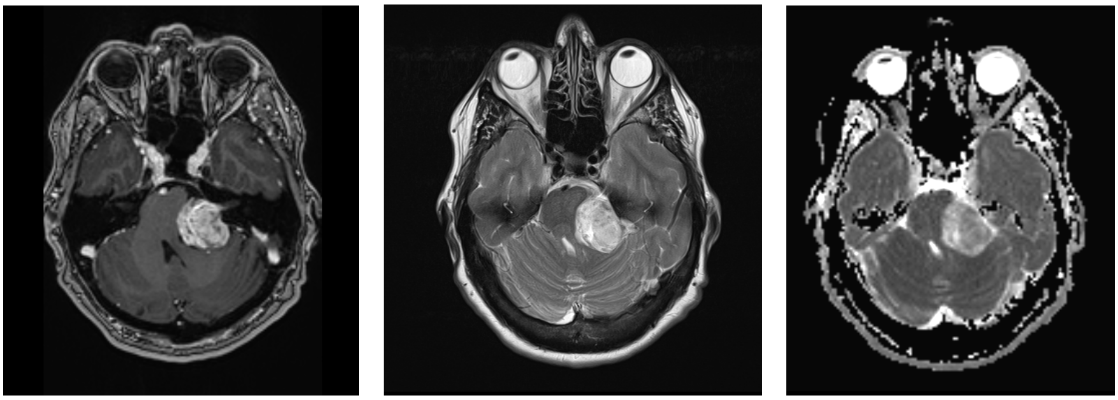
Preoperative magnetic resonance imaging (MRI) data were collected as part of routine clinical assessment for vestibular schwannoma [16]. The imaging dataset comprised volumetric acquisitions of anatomical and diffusion-related sequences that were subsequently used for tumor segmentation, image registration, and radiomic analyses.

The available MRI protocol included a high-resolution three-dimensional T1-weighted volume, as well as T2-weighted volumes and apparent diffusion coefficient (ADC) volumes derived from diffusion-weighted imaging and provided as part of the clinical dataset. From a clinical and imaging perspective, the MRI volumes used in this study provide complementary information on tumor anatomy and tissue characteristics. Specifically:

- **T1-weighted volumes**: provide high-resolution anatomical detail and clear delineation of tumor boundaries [16]. In this study, T1-weighted volumes served as the anatomical reference space for tumor localization, manual segmentation, and subsequent multimodal image registration.
- **T2-weighted volumes**: are sensitive to tissue water content and internal tumor heterogeneity, and are routinely used to characterize the internal architecture of vestibular schwannomas. T2-weighted volumes were analyzed as a

source of radiomic texture information [17].

- **ADC volumes**: provide quantitative information related to water diffusion within tissues, which has been associated with cellularity and extracellular matrix properties [18]. ADC volumes were analyzed in parallel with T2-weighted volumes to investigate complementary microstructural information.



(a) T1-weighted MRI **(b)** T2-weighted MRI **(c)** ADC map

Figure 2.1: Representative examples of the MRI volumes used in this study.

All MRI examinations were acquired preoperatively. The available imaging protocol included a 3D T1-weighted volume, whereas T2-weighted images and ADC maps were acquired as multi-slice 2D sequences. For consistency of downstream processing, all sequences were handled as three-dimensional image volumes, and were subsequently resampled and spatially aligned as described in the following sections. Since imaging was performed within routine clinical practice, the availability of specific MRI volumes varied across patients. Consequently, downstream analyses were conducted on partially overlapping patient cohorts depending on the availability and quality of the required MRI volumes.

2.2.2 Tumor segmentation

Tumor segmentation was performed on preoperative three-dimensional T1-weighted MRI volumes, which served as the anatomical reference space for all subsequent analyses. T1-weighted volumes were selected for segmentation due to their high spatial resolution and clear delineation of tumor boundaries relative to surrounding anatomical structures.

Manual segmentation of the vestibular schwannoma was carried out by an expert clinician. The tumor region of interest (ROI) was delineated slice by slice on the volumetric T1-weighted dataset, generating a three-dimensional binary mask representing the tumor volume.

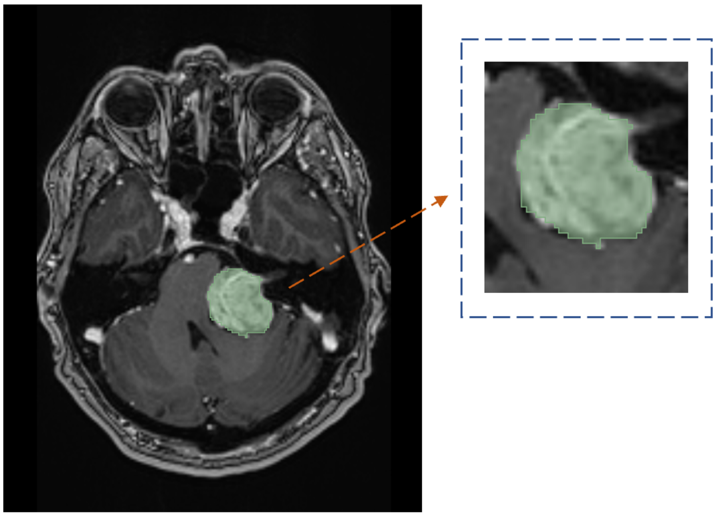


Figure 2.2: Representative axial slice of a preoperative T1-weighted MRI volume with manual tumor segmentation. The vestibular schwannoma region of interest (ROI) is overlaid in green on the anatomical image. A magnified view of the segmented tumor is shown on the right to highlight the accuracy of the manual delineation and the correspondence between the ROI boundaries and the tumor margins.

The resulting tumor mask defined the spatial extent of the lesion and constituted the reference ROI for all downstream processing steps. In particular, the T1-based segmentation provided the anatomical framework upon which multimodal image registration and radiomic analyses were subsequently performed.

Tumor segmentation was completed prior to image registration and radiomic feature extraction to ensure a consistent and anatomically accurate definition of the tumor region across all analyses.

2.2.3 Volume resampling and spatial alignment

To ensure voxel-wise correspondence between MRI sequences and the tumor region of interest, T2-weighted and apparent diffusion coefficient (ADC) volumes were spatially aligned to the anatomical reference space prior to radiomic analysis.

Volume resampling was performed using the open-source software 3D Slicer (version 5.8.1), employing the Resample Image module from the BRAINS toolkit [19]. T2-weighted and ADC volumes were resampled into the T1-weighted reference space using linear interpolation, selected to preserve intensity continuity while minimizing interpolation-related artifacts. No rigid or deformable registration was applied at this stage, as the objective was to achieve spatial correspondence through resampling to a common voxel grid.

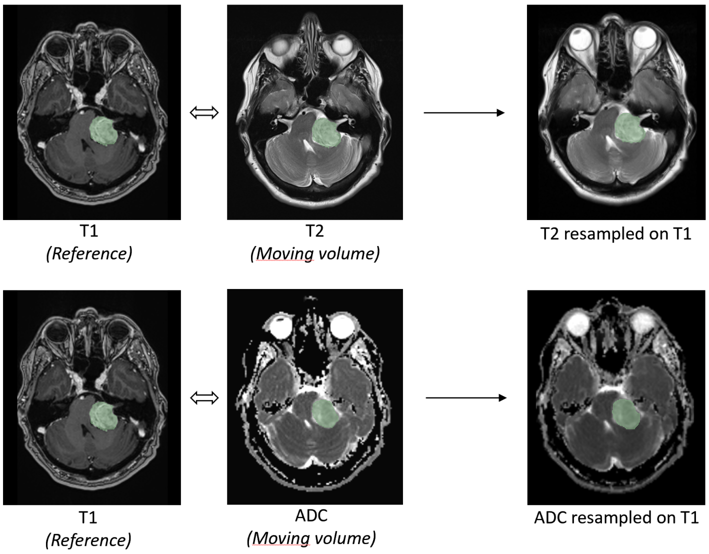


Figure 2.3: Illustration of the volume alignment workflow used in this study. T2-weighted and apparent diffusion coefficient (ADC) volumes (moving volumes) were resampled into the anatomical space of the T1-weighted MRI (reference volume). Tumor segmentation, originally defined on the T1-weighted volume, was consequently transferred to the resampled T2 and ADC volumes, ensuring voxel-wise spatial correspondence across imaging modalities prior to radiomic feature extraction.

Following resampling, a qualitative visual inspection of the spatial correspondence between volumes was performed to verify correct alignment between MRI sequences and the tumor region of interest. This verification step was conducted in collaboration with an experienced clinician, focusing on the pixel-wise correspondence of anatomical landmarks and tumor boundaries across sequences. Only cases demonstrating satisfactory visual alignment were retained for subsequent radiomic feature extraction.

This procedure ensured that radiomic features were extracted from spatially corresponding tumor regions across MRI sequences, thereby supporting the robustness and interpretability of downstream analyses.

2.2.4 Volume intensity normalization

Prior to radiomic feature extraction, intensity normalization was applied to the MRI volumes in order to reduce inter-subject variability related to acquisition settings and intensity scaling [15]. This step was performed after the spatial alignment (i.e., on T2-weighted and ADC volumes already resampled to the T1-weighted reference space).

Normalization was implemented using a dedicated Python script based on the SimpleITK library [20]. For each volume, intensity statistics were computed excluding background voxels, defined as voxels with zero intensity. This non-zero masking strategy was adopted to avoid bias introduced by large uniform background regions.

Two alternative normalization strategies were implemented (z-score and min–max scaling); in the analyses presented in this thesis, z-score normalization was applied. Specifically, for each volume the mean (μ) and standard deviation (σ) of the non-zero voxel intensities were computed and used to transform voxel intensities as:

$$I_{\text{norm}} = \frac{I - \mu}{\sigma}.$$

The normalized volumes were saved while preserving the original geometric metadata (spacing, origin, and direction) to maintain spatial consistency for subsequent processing steps.

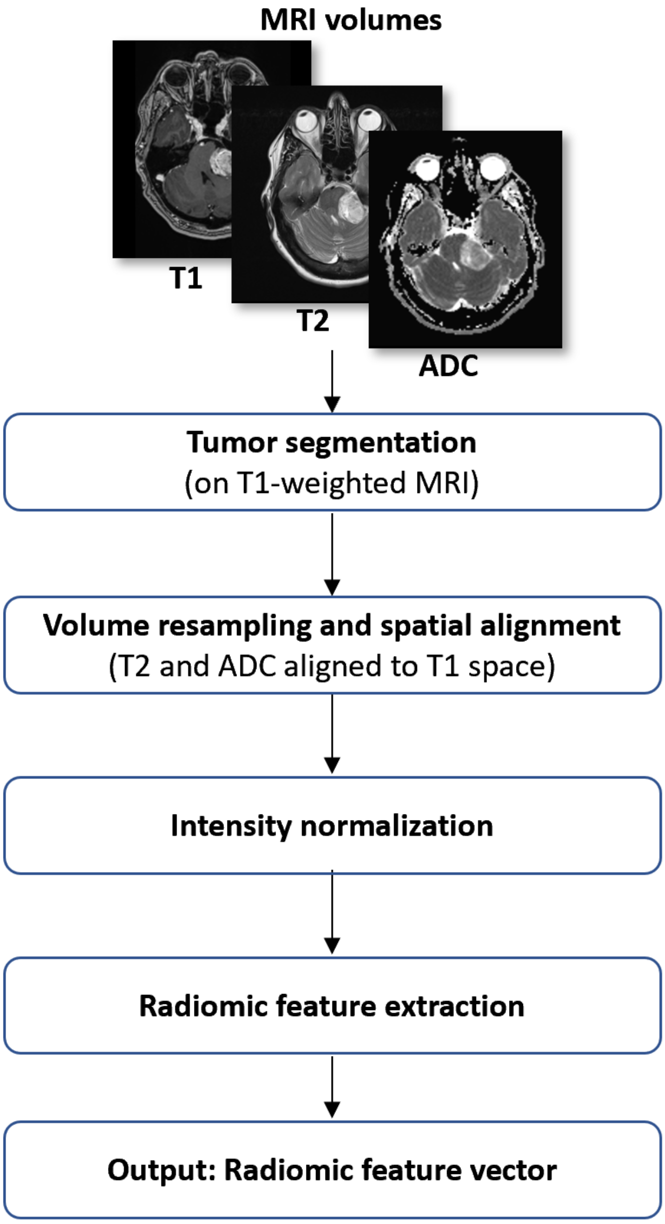


Figure 2.4: Overview of the MRI processing workflow described in Section 2.2. The pipeline starts from preoperative MRI volumes (T1-weighted, T2-weighted, and ADC maps). Tumor segmentation is performed on the T1-weighted reference image. T2-weighted and ADC volumes are subsequently resampled to the T1 reference space to ensure voxel-wise spatial correspondence across MRI sequences. After intensity normalization, radiomic features are extracted from the tumor region of interest, resulting in a radiomic feature vector used for subsequent analyses.

2.3 Radiomic Feature Extraction

This section describes the extraction of radiomic features from preprocessed MRI volumes and the statistical analyses performed to investigate their association with histologically derived collagen content. Radiomic features were computed from both T2-weighted and apparent diffusion coefficient (ADC) volumes using a consistent methodology across imaging modalities [11].

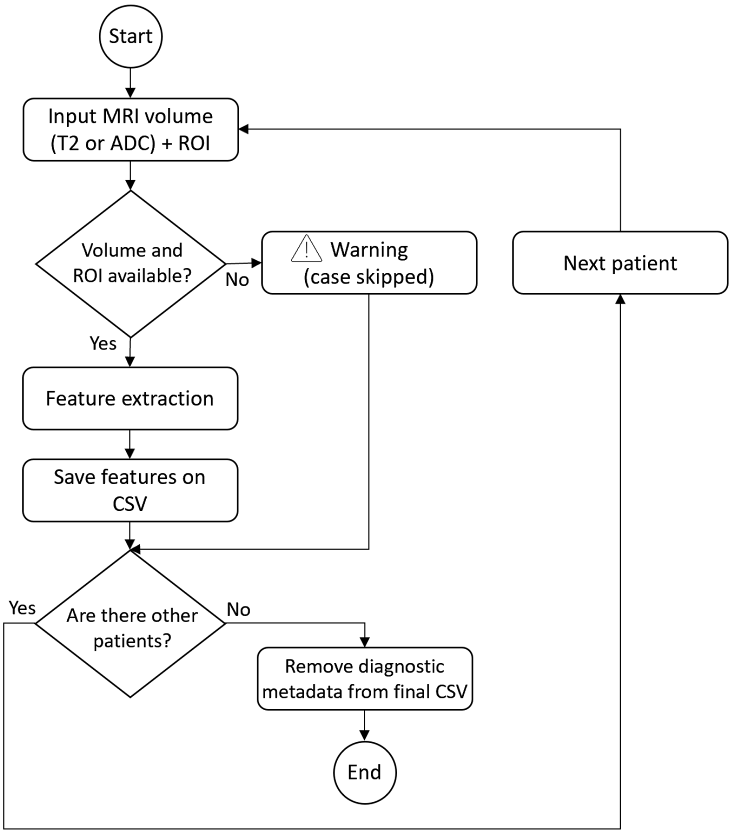


Figure 2.5: Schematic representation of the radiomic feature extraction pipeline, from input MRI volumes and ROIs to the final feature dataset.

2.3.1 Radiomic feature classes and extraction parameters

Radiomic features were extracted from the tumor regions of interest in order to quantitatively characterize intensity distribution and spatial heterogeneity within the lesion. Feature computation was performed on the original, preprocessed image volumes, and focused on intensity-based and texture-based descriptors.

The extracted radiomic features can be grouped into the following categories [15]:

- **First-order features**: describe the statistical distribution of voxel intensities within the tumor region, providing information on central tendency, dispersion, and intensity range.
- **Gray Level Co-occurrence Matrix (GLCM) features**: quantify second-order spatial relationships between pairs of voxel intensities, capturing patterns of local texture organization.
- **Gray Level Run Length Matrix (GLRLM) features**: characterize the length and distribution of consecutive voxels with identical gray levels, reflecting textural homogeneity and directional structures.
- **Gray Level Size Zone Matrix (GLSZM) features**: describe the size distribution of homogeneous intensity zones within the tumor volume.
- **Gray Level Dependence Matrix (GLDM) features**: measure the degree of gray-level dependence among neighboring voxels, providing information on local intensity complexity.
- **Neighboring Gray Tone Difference Matrix (NGTDM) features**: assess coarse texture differences by comparing each voxel with the average intensity of its neighborhood.

The complete list of extracted first-order features is reported in Table 2.1, while texture features grouped by family are summarized in Table 2.2.

Table 2.1: Extracted first-order radiomic features.

First-order features (1/2)	**First-order features (2/2)**
10Percentile	Minimum
90Percentile	Range
Energy	RobustMeanAbsoluteDeviation
Entropy	RootMeanSquared
InterquartileRange	Skewness
Kurtosis	TotalEnergy
Maximum	Uniformity
MeanAbsoluteDeviation	Variance
Mean	Median

Table 2.2: Extracted texture radiomic features, grouped by feature family.

GLCM	**GLRLM**	**GLSZM**
`Autocorrelation`	`GrayLevelNonUniformity`	`GrayLevelNonUniformity`
`ClusterProminence`	`GrayLevelNonUniformityNormalized`	`GrayLevelNonUniformityNormalized`
`ClusterShade`	`GrayLevelVariance`	`GrayLevelVariance`
`ClusterTendency`	`HighGrayLevelRunEmphasis`	`HighGrayLevelZoneEmphasis`
`Contrast`	`LongRunEmphasis`	`LargeAreaEmphasis`
`Correlation`	`LongRunHighGrayLevelEmphasis`	`LargeAreaHighGrayLevelEmphasis`
`DifferenceAverage`	`LongRunLowGrayLevelEmphasis`	`LargeAreaLowGrayLevelEmphasis`
`DifferenceEntropy`	`LowGrayLevelRunEmphasis`	`LowGrayLevelZoneEmphasis`
`DifferenceVariance`	`RunEntropy`	`SizeZoneNonUniformity`
`Id`	`RunLengthNonUniformity`	`SizeZoneNonUniformityNormalized`
`Idm`	`RunLengthNonUniformityNormalized`	`SmallAreaEmphasis`
`Idmn`	`RunPercentage`	`SmallAreaHighGrayLevelEmphasis`
`Idn`	`RunVariance`	`SmallAreaLowGrayLevelEmphasis`
`Imc1`	`ShortRunEmphasis`	`ZoneEntropy`
`Imc2`	`ShortRunHighGrayLevelEmphasis`	`ZonePercentage`
`InverseVariance`	`ShortRunLowGrayLevelEmphasis`	`ZoneVariance`
`JointAverage`		
`JointEnergy`		
`JointEntropy`		
`MCC`		
`MaximumProbability`		
`SumAverage`		
`SumEntropy`		
`SumSquares`		

GLDM	**NGTDM**	
`DependenceEntropy`	`Busyness`	
`DependenceNonUniformity`	`Coarseness`	
`DependenceNonUniformityNormalized`	`Complexity`	
`DependenceVariance`	`Contrast`	
`GrayLevelNonUniformity`	`Strength`	
`GrayLevelVariance`		
`HighGrayLevelEmphasis`		
`LargeDependenceEmphasis`		
`LargeDependenceHighGrayLevelEmphasis`		
`LargeDependenceLowGrayLevelEmphasis`		
`LowGrayLevelEmphasis`		
`SmallDependenceEmphasis`		
`SmallDependenceHighGrayLevelEmphasis`		
`SmallDependenceLowGrayLevelEmphasis`		

Radiomic feature extraction required the specification of several processing parameters within the PyRadiomics framework [21]. These parameters were selected to ensure spatial consistency, reproducibility, and sensitivity to intra-tumoral heterogeneity, and were kept constant across all patients and imaging modalities. The main extraction parameters and the rationale for their selection are summarized in Table 2.3.

Table 2.3: Radiomic feature extraction parameters required by the PyRadiomics framework and selected settings.

Parameter	Setting	Rationale
Image type	Original	Features were extracted from original images to directly characterize intrinsic intensity and texture properties.
Resampled voxel spacing	$1 \times 1 \times 1$ mm^3	Isotropic resampling ensures spatial consistency across patients and imaging modalities.
Interpolation method	B-spline	Provides smooth intensity interpolation and is commonly used in radiomic analyses.
Mask correction	Enabled	Ensures consistency between resampled images and corresponding tumor masks.
Internal intensity normalization	Disabled	Normalization was performed during preprocessing, avoiding redundant intensity transformations.
Gray-level discretization	Fixed bin width	Required for texture feature computation within the PyRadiomics framework.
Bin width	0.10	Selected to preserve sensitivity to intensity variations while limiting noise effects.

2.4 Histological image analysis and collagen quantification pipeline

This section describes the automated pipeline developed for the segmentation and quantitative analysis of fibrous collagen in digitized histological slides stained with hematoxylin and eosin (H&E). The proposed workflow was designed to robustly identify organized collagen structures while minimizing the inclusion of non-collagenous eosinophilic tissue and staining-related artifacts.

Whole-slide histological images were processed through a multi-step pipeline comprising tissue detection, color normalization, stain deconvolution, collagen segmentation, and structural analysis based on fiber orientation. Particular attention was devoted to color variability across slides and tiles, as well as to the discrimination between true fibrous collagen and visually similar histological components.

The final output of the pipeline consists of a refined collagen mask from which quantitative metrics, including the percentage of fibrous collagen relative to total tissue area, were derived. An overview of the complete histological processing pipeline is illustrated in Figure 2.6.

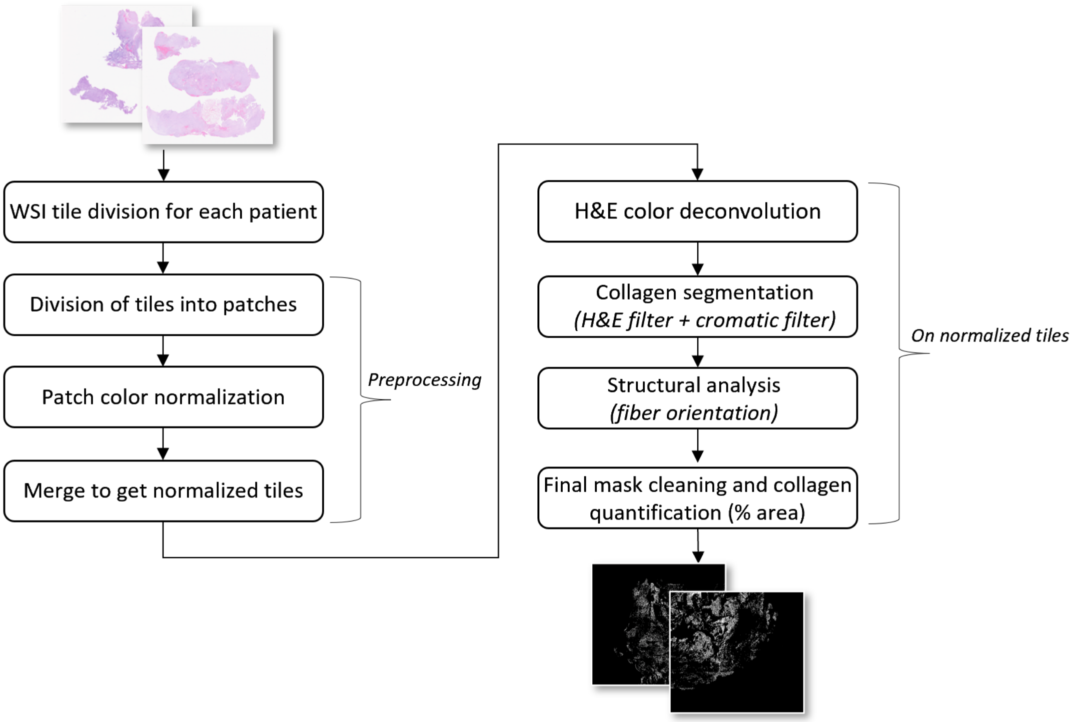


Figure 2.6: Overview of the automated histological image analysis pipeline for collagen quantification. Whole-slide images (WSIs) are processed through tile-based preprocessing and subsequent collagen segmentation and analysis on normalized tiles, leading to the estimation of collagen content at the WSI level.

2.4.1 Histological image acquisition and tiling

For the patients included in the clinical cohort described in Section 2.1, histological whole-slide images (WSIs) stained with hematoxylin and eosin (H&E) were available following surgical resection. These digitized histological slides constituted the histological counterpart of the multimodal dataset and served as the input data for the automated collagen quantification pipeline described in the following sections.

Due to the extremely high spatial resolution and large file size of WSIs, direct processing of entire slides was not computationally feasible. Therefore, a tiling strategy was adopted to enable efficient, scalable, and reproducible analysis while preserving the structural characteristics of the tissue.

Each WSI was partitioned into a fixed grid of non-overlapping tiles using a custom script implemented within the QuPath software platform [22]. The tiling strategy was designed to obtain a limited and homogeneous number of tiles per patient, specifically nine tiles per slide, representing a compromise between adequate spatial coverage of the specimen and computational feasibility in terms of memory usage and processing time. Excessive subdivision into smaller tiles was intentionally avoided to reduce fragmentation of tissue structures and to preserve sufficient contextual information within each tile.

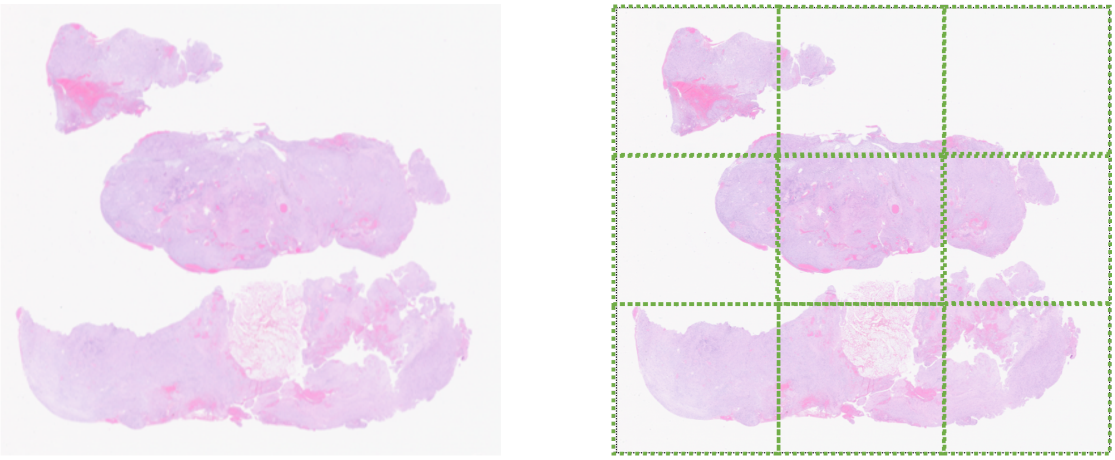

(a) (b)

Figure 2.7: Representative example of a histological whole-slide image (a) and the corresponding tiling strategy (b), illustrating the subdivision of the specimen into non-overlapping tiles for subsequent processing

Tile dimensions were predefined within the QuPath script. In cases where the spatial dimensions of the whole-slide image were not exactly divisible by the target tile size, peripheral regions were automatically handled according to the native QuPath tiling logic. This process ensured consistent tile geometry across patients while avoiding the exclusion of relevant tissue content. In particular, tile generation was configured such that meaningful tissue regions were retained and no diagnostically relevant areas were discarded as a consequence of the cropping procedure.

For organizational consistency and traceability, the generated tiles were stored in a hierarchical directory structure, with a dedicated folder assigned to each anonymized patient identifier (e.g., `T001`, `T002`, etc.), containing the corresponding set of tiles extracted from that patient's histological slide.

The resulting tiles constituted the fundamental processing units of the histological pipeline and were subsequently used as input for all downstream preprocessing, segmentation, structural analysis, and quantitative collagen assessment steps described in the following sections.

2.4.2 Color normalization using GSN-GAN

Motivation and stain variability problem

Digitized H&E whole-slide images are affected by substantial appearance variability across cases, even when acquired within the same clinical workflow. Differences in staining intensity, color balance, and illumination conditions may arise from pre-analytical factors (e.g., fixation and staining protocols) as well as scanner-dependent settings. Such variability can act as a confounding source of information for downstream quantitative analyses, potentially reducing robustness and limiting the comparability of measurements across patients [23].

This variability is illustrated in Fig. 2.8, where representative H&E whole-slide images from different patients exhibit noticeable differences in color balance and staining intensity, despite comparable tissue content.

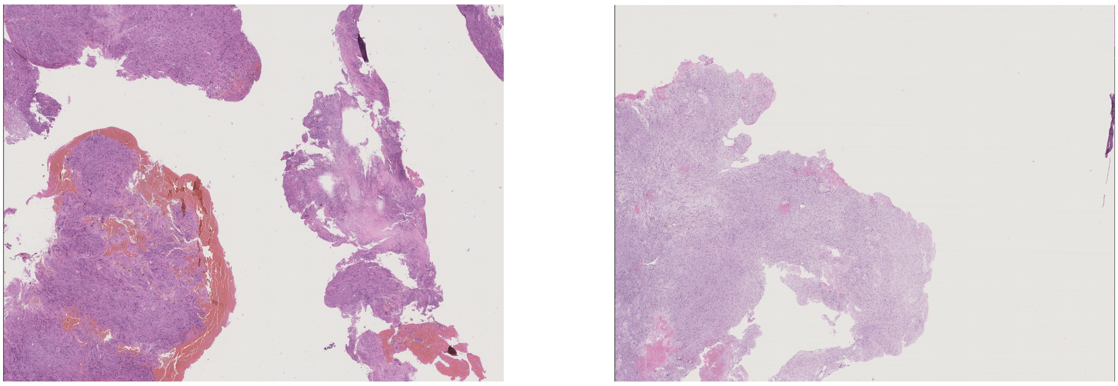


Figure 2.8: Representative H&E whole-slide images from different patients, highlighting inter-slide variability in staining appearance and color distribution that may affect quantitative image analysis.

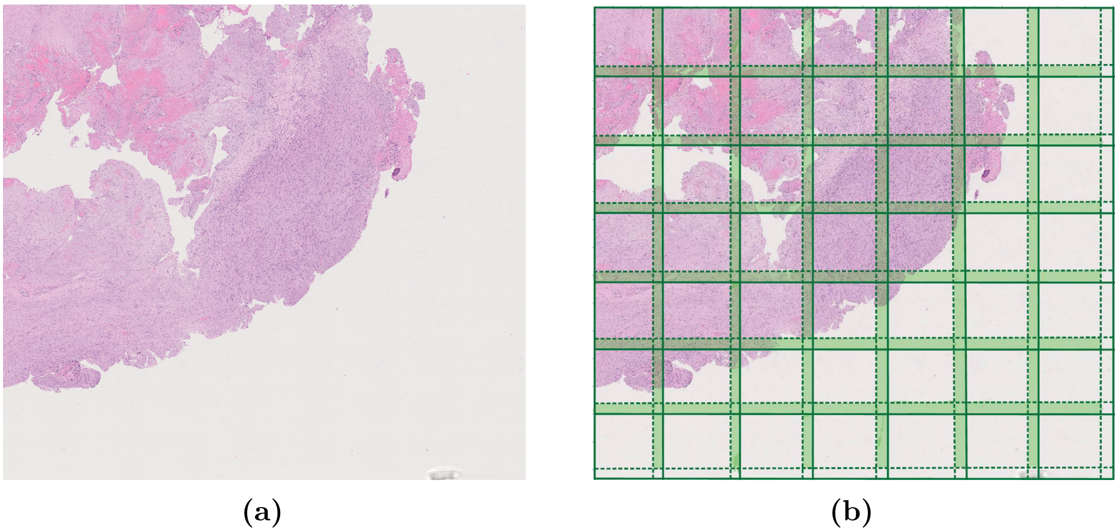
To mitigate these effects, a color normalization step was introduced prior to collagen-specific processing. The objective of normalization was to reduce irrelevant chromatic variability while preserving diagnostically meaningful tissue morphology and structural patterns. In the present workflow, normalization was performed using a generative model-based approach (GSN-GAN, [24]), selected to provide a data-driven mapping toward a consistent color appearance without requiring a handcrafted stain model.

Patch-based normalization strategy

Color normalization was applied using a patch-based strategy to ensure computational scalability and to improve local robustness to staining variability. Starting from the H&E tiles generated in Section 2.4.1, each tile was subdivided into square patches of fixed size (1024×1024 pixels) using an automated script.

To mitigate border artifacts and avoid discontinuities at patch boundaries, patches were extracted with a controlled spatial overlap of 128 pixels (i.e., an effective stride of 896 pixels). This overlap was introduced to provide contextual continuity across adjacent patches during normalization and to reduce the risk of visible seams after reconstruction.

When the tile dimensions were not exactly compatible with an integer number of strides, the sampling grid was adjusted at the borders by clamping the patch origin coordinates. In practice, the last patches along each direction were shifted as needed to remain fully within image bounds, while preserving a constant patch size across the dataset. This design ensures that normalization is performed on uniformly sized inputs. The patch extraction strategy, including patch size and spatial overlap, is schematically illustrated in Figure 2.9.



(a) **(b)**

Figure 2.9: Patch-based subdivision of a representative H&E tile. The original tile (a) is partitioned into 1024 × 1024 pixel patches with a 128-pixel overlap in both spatial directions (b). Overlapping regions between adjacent patches are highlighted in green.

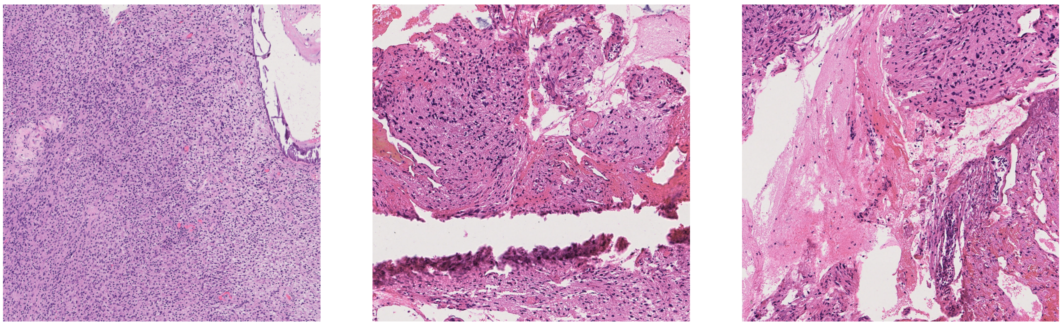
For each tile, the extracted patches were stored in a dedicated output directory, preserving the original patient-wise folder hierarchy. Patch filenames encoded their spatial origin (top-left coordinates) within the parent tile, enabling deterministic reconstruction in the subsequent merging step. In addition, a lightweight metadata file was saved for each tile, recording the original tile dimensions and the patching parameters (patch size and overlap), thereby supporting reproducible downstream reassembly of normalized tiles.

GSN-GAN model and inference configuration

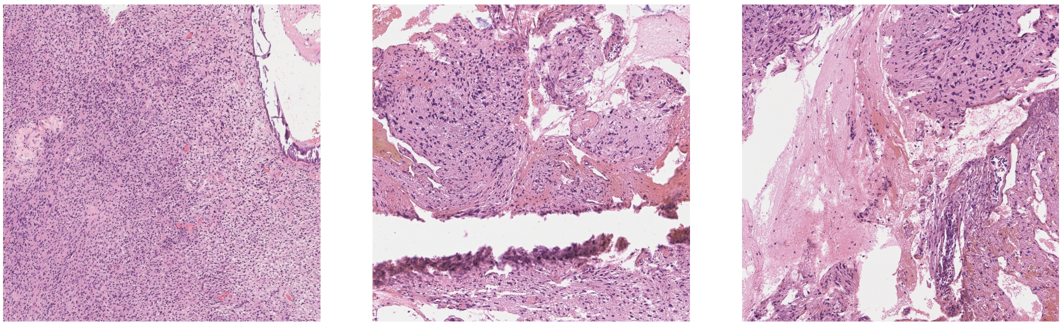
Patch-wise stain normalization was performed using the Generalized Stain Normalization GAN (GSN-GAN) proposed by Salvi et al. [24]. The method formulates stain normalization as a paired image-to-image translation problem, where an input H&E image is mapped to a corresponding normalized target while preserving pixel-level spatial correspondence. In the original work, the target domain is obtained through a heuristic normalization procedure and only visually optimal normalized samples are retained for model development, enabling the network to learn a robust and artifact-resistant mapping.

Architecturally, GSN-GAN follows a conditional GAN (Pix2Pix) design, employing a U-Net generator and a PatchGAN discriminator. The adversarial objective is complemented by a pixel-wise reconstruction loss, which enforces local consistency between source and normalized images and supports the preservation of tissue morphology during the chromatic transformation. Salvi et al. also discuss the use of the least-squares GAN (LSGAN) formulation to improve training stability and reduce common GAN failure modes, such as vanishing gradients [24].

In the present pipeline, inference was carried out on the extracted patches described in Section 2.4.2 using a customized testing script derived from the authors' framework. An example of patch-level stain normalization is shown in Figure 2.10, where representative H&E patches before and after GSN-GAN normalization are compared. The model harmonizes color distribution while preserving the underlying tissue morphology.



(a) Original H&E patches



(b) Corresponding patches after GSN-GAN color normalization

Figure 2.10: Example of stain normalization at the patch level. Representative H&E patches before (a) and after (b) GSN-GAN normalization are shown. The normalization harmonizes stain appearance across patches while preserving the underlying tissue morphology and structural details.

The script was designed to (i) process patches in a deterministic order (single-image dataset mode, no random flipping), (ii) run inference with batch size equal to one, and (iii) preserve the original patient-wise folder hierarchy when saving outputs. Specifically, each normalized patch was saved using the same filename as the corresponding input patch, while reproducing the same subdirectory structure under the output root directory, thereby ensuring full traceability and straightforward downstream reconstruction.

The generator and discriminator configurations were selected consistently with the adopted pretrained model checkpoint. Inference was performed using a U-Net generator (`unet_256`) and a `basic` discriminator, with instance normalization enabled [24]. Both input and output were RGB images (`input_nc`=3, `output_nc`=3). No resizing or additional preprocessing was applied (`preprocess none`), ensuring that normalization was executed directly on the native patch resolution. Dropout was disabled at test time (`no_dropout`) and GPU acceleration was enabled when available (`gpu_ids`=0). The pretrained weights corresponded to experiment `GSN-GAN_unet24_basic96_lsgan` at epoch 104, and the full patch set was processed by setting a high upper bound on the number of test samples (`num_test`=100000).

Table 2.4: Main GSN-GAN inference configuration used for patch normalization.

Setting	Value
Generator / Discriminator	`unet_256` / `basic`
Normalization	instance norm
Input / output channels	3 / 3 (RGB)
Patch size	1024 × 1024 px
Overlap	128 px (stride 896 px)
Preprocessing	none
Dropout	disabled
Checkpoint	`GSN-GAN_unet24_basic96_lsgan`, epoch 104
Execution	batch size 1, deterministic order, GPU enabled

A schematic overview of the GSN-GAN workflow (Figure 2.11), including the paired training paradigm and the generator/discriminator interaction, is reported in the original publication and can be used here to contextualize the adopted model within the overall pipeline.

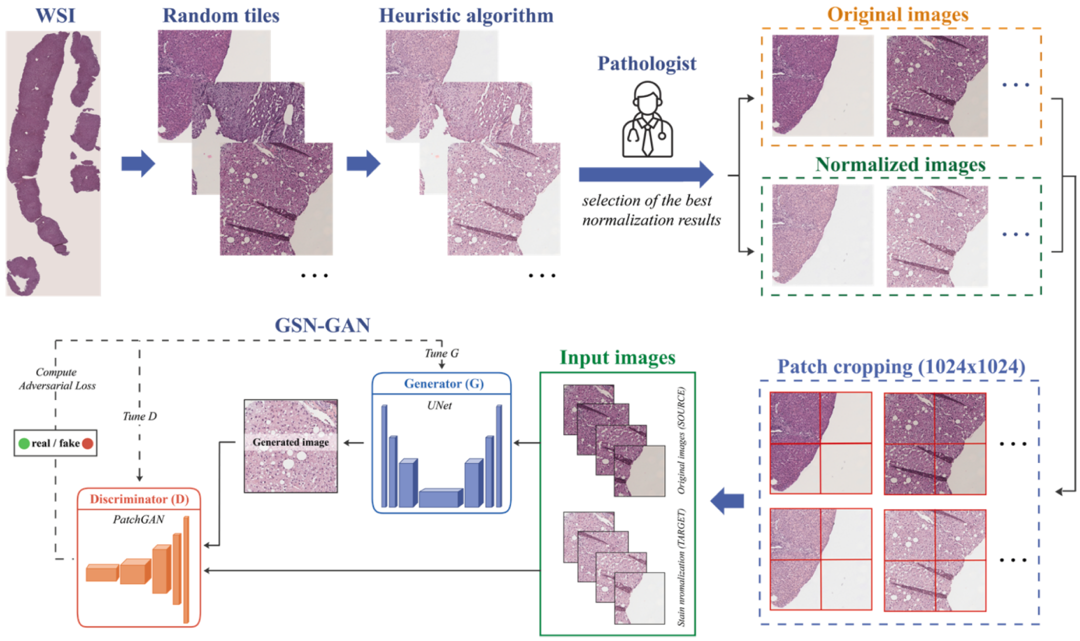


Figure 2.11: Schematic overview of the GSN-GAN stain normalization framework (reproduced from Salvi et al. [24]). The model performs paired image-to-image translation using a conditional GAN, with a U-Net generator and a PatchGAN discriminator, combined with a pixel-wise loss to preserve spatial correspondence.

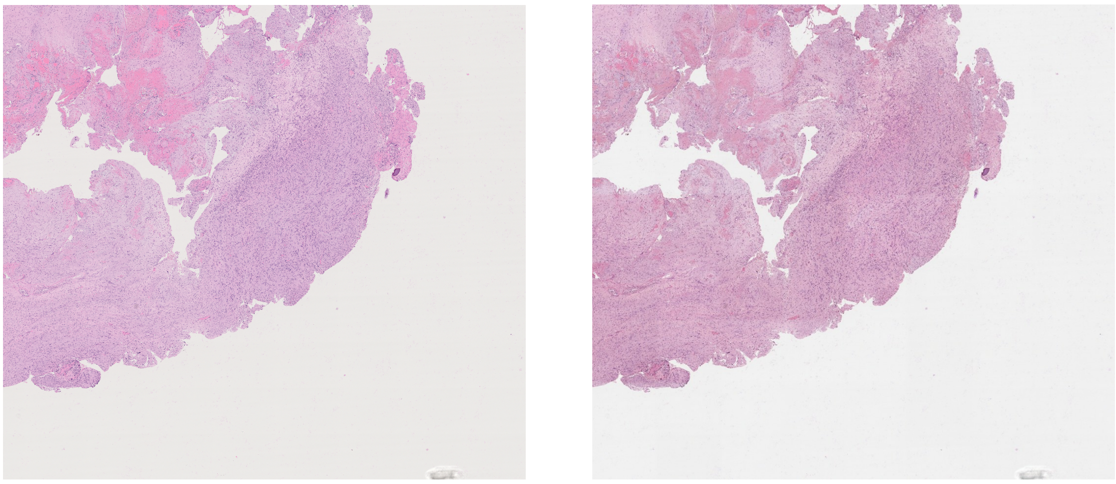
Reconstruction of normalized tiles

Following patch-level color normalization, normalized tiles were reconstructed by reassembling the corresponding normalized patches while preserving their original spatial arrangement. This step was required to restore a coherent tile-level representation suitable for subsequent elaboration steps.

Reconstruction was performed using a dedicated merging script that exploits the spatial coordinates encoded in patch filenames during the splitting stage (Section 2.4.2). For each tile, the original tile dimensions, patch size, and overlap parameters were retrieved from an associated metadata file (`_meta.txt`), ensuring consistency between the splitting and merging procedures.

Normalized patches were placed back into their original spatial locations on a blank canvas matching the original tile size. In regions where multiple patches overlapped, pixel intensities were combined by computing the arithmetic mean across all contributing patches. This averaging strategy was adopted to ensure smooth intensity transitions across patch boundaries and to minimize seam artifacts introduced by independent patch-wise normalization.

After accumulation and normalization, the reconstructed tile was cropped to the original tile dimensions to remove any redundant border regions introduced by boundary handling during patch extraction. The final output thus retained the original spatial extent and resolution of the input tile, while exhibiting reduced color variability due to normalization. An example of a reconstructed normalized tile is shown in Figure 2.12, illustrating the visual consistency between the original tile and its normalized counterpart after patch-based processing and reassembly.



(a) **(b)**

Figure 2.12: Reconstruction of a normalized histological tile from patch-based processing. The original H&E tile (a) is compared with the corresponding tile after color normalization and patch reassembly (b).

Reconstructed normalized tiles were saved following the same patient-wise directory hierarchy as the original tiles, thereby maintaining traceability and facilitating seamless integration with downstream processing steps.

Practical considerations and robustness of the approach

Several practical aspects were evaluated to ensure a robust reconstruction of normalized tiles from patch-wise GAN inference. In particular, different overlap configurations between adjacent patches were empirically assessed.

Preliminary experiments with a reduced overlap (64 pixels) occasionally resulted in faint grid-like discontinuities at patch boundaries, consistent with checkerboard artifacts commonly reported in convolutional generative models (Fig. 2.13) [25]. This phenomenon is well documented in the literature and has motivated dedicated architectural and kernel-based mitigation strategies in subsequent works.

Although kernel-based approaches were also tested in the present study, the most visually stable and consistent reconstructions were obtained using a larger spatial overlap (128 pixels) combined with simple averaging in the overlapping regions. This configuration significantly reduced boundary artifacts while preserving local tissue morphology and color consistency across the reconstructed tile.

Based on these observations, the 128-pixel overlap with averaging was adopted as the final configuration for all subsequent analyses, providing a favorable trade-off between computational efficiency and reconstruction quality.

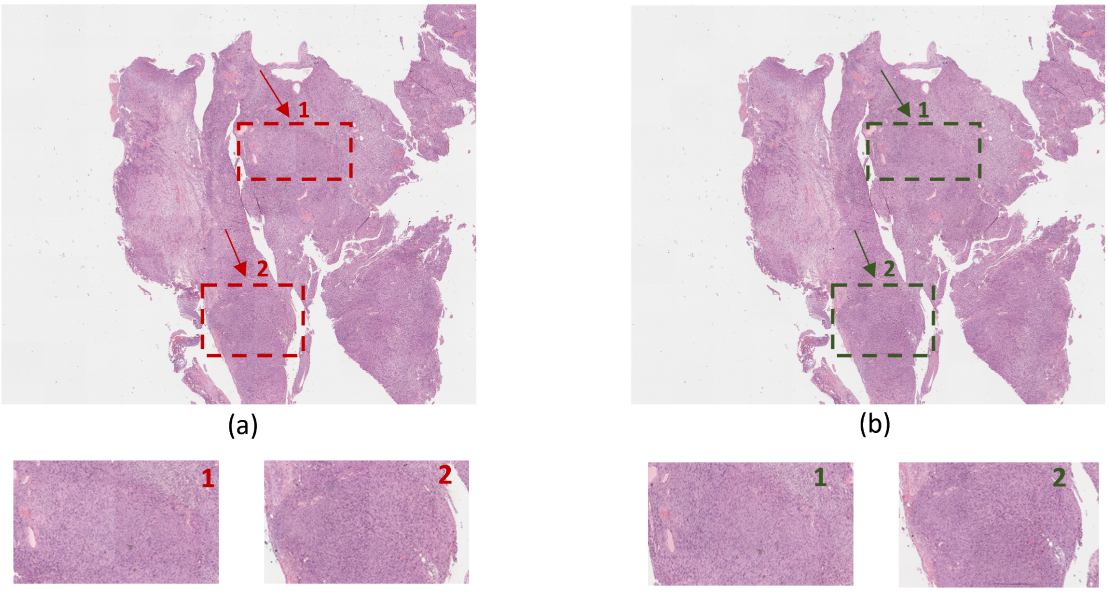


Figure 2.13: Comparison of tile reconstruction quality using two patch overlap configurations. (a) With a reduced overlap of 64 pixels, faint grid-like artifacts may appear at patch boundaries, as indicated by the red arrows and highlighted in the corresponding zoomed regions (1–2). (b) Increasing the overlap to 128 pixels and applying averaging in the overlapping regions results in smoother and more visually consistent reconstructions, as shown in the corresponding zoomed regions.

2.4.3 Tissue detection and H&E stain deconvolution

Prior to stain-specific processing, a tissue detection step was performed to exclude background regions and to restrict all subsequent computations to pixels belonging to histological material. This masking stage is particularly important in H&E tiles, where large white areas corresponding to the glass slide may otherwise bias intensity-based thresholds and percentile estimations. In the present pipeline, tissue detection was implemented directly on the RGB tile by exploiting the characteristic appearance of background pixels, which are typically associated with high brightness and very low saturation.

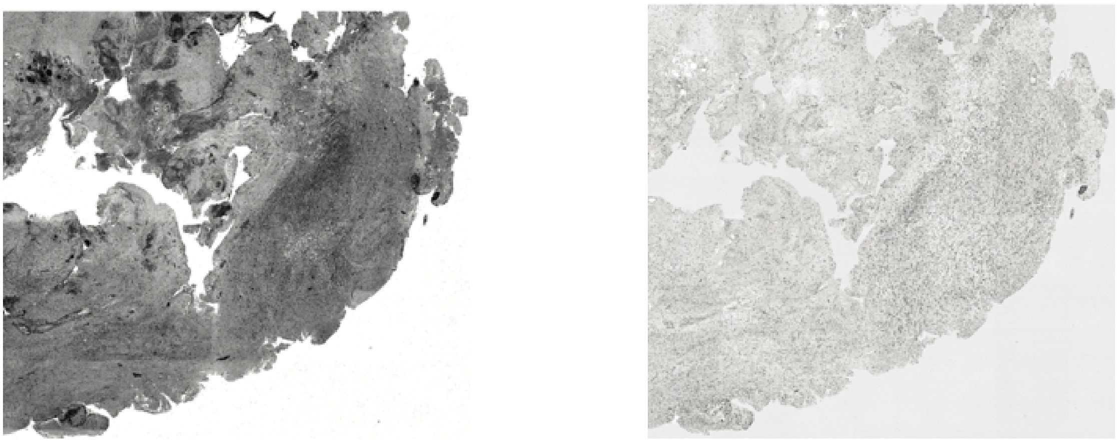
Specifically, each tile was converted to the HSV color space and a preliminary tissue mask was obtained by thresholding the Value and Saturation channels (i.e., selecting pixels with sufficiently low brightness and non-negligible saturation). The resulting binary mask was refined through morphological cleaning to remove small isolated components and to fill small holes, thus producing a compact and stable tissue region. This tissue mask (Fig. 2.14) was used as the spatial support for all subsequent percentile-based thresholds and for the computation of collagen-related metrics, ensuring that background regions did not contribute to any quantitative estimate.

(a) **(b)**

Figure 2.14: Tissue detection step used to exclude background regions before stain-specific processing. (a) Normalized H&E tile. (b) Corresponding extracted tissue mask obtained in HSV space and refined via morphological operations. The tissue mask defines the spatial support for percentile-based thresholds and quantitative estimates.

After tissue detection, stain deconvolution was applied to separate the contributions of hematoxylin and eosin from the RGB image. Color deconvolution provides a physically interpretable representation of histological staining by mapping the

RGB intensities to a stain space in which individual channels correspond to specific dye absorption components [26]. In this work, deconvolution was performed using the HED formulation, yielding independent hematoxylin-like and eosin-like channels. For numerical stability and to improve robustness across tiles, both channels were intensity-rescaled using percentile-based normalization (1st–99th percentiles computed within the tissue mask), and a mild Gaussian smoothing was applied to the eosin channel to reduce high-frequency noise prior to thresholding.



(a) **(b)**

Figure 2.15: Hematoxylin and eosin stain deconvolution of a representative normalized H&E tile. The eosin channel (a) highlights cytoplasmic and extracellular components, including collagen-rich regions, while the hematoxylin channel (b) emphasizes nuclear structures. The separation of the two stain contributions enables collagen-specific analysis to be performed primarily on the eosin component in subsequent processing steps.

From a biological standpoint, collagen-rich regions in H&E slides are primarily associated with eosinophilic structures, typically appearing as strongly pink areas, whereas hematoxylin predominantly highlights cell nuclei [27]. Consequently, the eosin channel was used as the main substrate for collagen-related segmentation steps, while the hematoxylin channel was exploited to suppress nuclear-dominated regions when required. In practice, candidate eosinophilic pixels were identified within the tissue mask using a percentile-based threshold on the smoothed eosin channel, and pixels exhibiting high hematoxylin response were excluded to reduce false positives originating from densely nucleated areas. The outputs of this stage constituted the first stain-driven candidate mask for collagen and served as input for the subsequent color-based refinement and structural filtering steps described in the following sections.

2.4.4 Collagen segmentation strategy

Following tissue detection and stain deconvolution, collagen segmentation was performed through a multi-stage strategy combining stain-driven, color-based, and structural criteria. The objective of this approach was to identify collagen-rich regions while minimizing false positives arising from non-collagenous eosinophilic structures.

The first candidate mask was derived from the eosin channel obtained through H&E deconvolution. Since collagen-rich extracellular matrix appears strongly eosinophilic, pixels within the tissue mask exhibiting high eosin intensity were selected using a percentile-based threshold computed exclusively over tissue pixels on the smoothed eosin channel. Specifically, a threshold τ_E was defined as the p_E-th percentile of the eosin intensity distribution within the tissue mask (Fig. 2.16), and pixels satisfying $E > \tau_E$ were retained as collagen candidates.

To suppress regions dominated by densely packed nuclei, pixels with high hematoxylin response were excluded using an additional percentile-based threshold on the hematoxylin channel. A threshold τ_H was computed as the p_H-th percentile of the hematoxylin distribution within the same tissue support, and pixels satisfying $H \geq \tau_H$ were discarded.

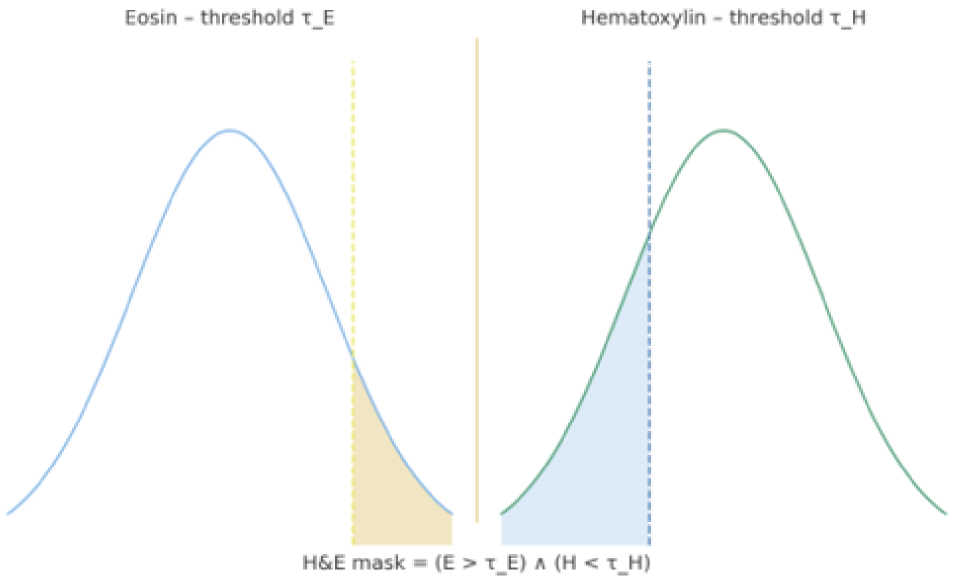
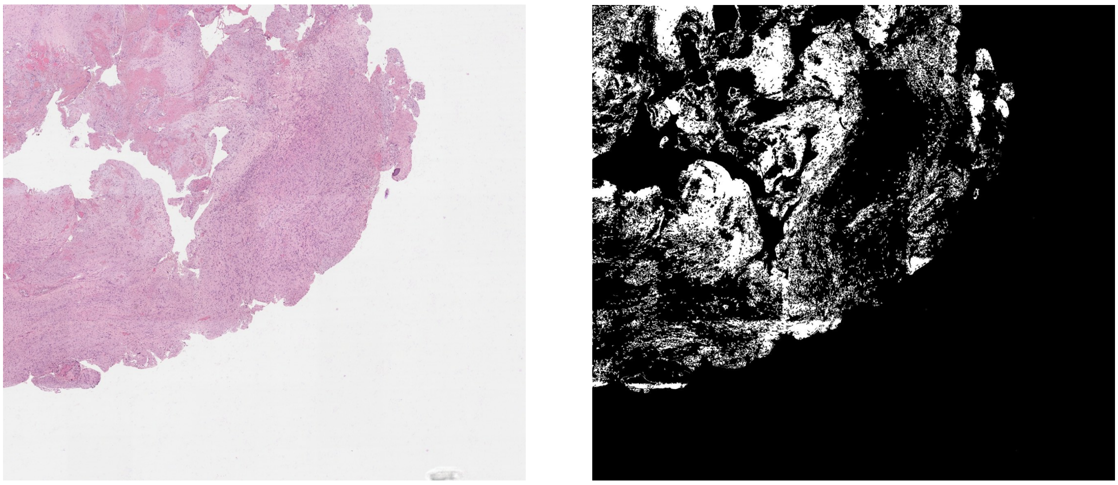


Figure 2.16: Percentile-based thresholding strategy applied to the eosin and hematoxylin channels. Thresholds τ_E and τ_H are computed from the intensity distributions restricted to tissue pixels. Collagen candidates are identified as pixels with high eosin response and low hematoxylin response, i.e., $(E > \tau_E) \wedge (H < \tau_H)$.

To further increase specificity, a chromatic filter was applied directly to the RGB image using a combined HSV and CIELab criterion. In HSV space, constraints were imposed on hue, saturation, and value to capture the characteristic pink range of eosinophilic collagen. In parallel, thresholds on the a and b components in Lab space were used to enforce consistency with reddish-pink chromatic distributions. The intersection of the HSV and Lab masks, restricted to the tissue region, yielded a color-consistent candidate mask.

A conservative logical AND operation between the stain-driven mask and the color-based mask was performed to retain only pixels satisfying both criteria. Mild morphological cleaning (removal of small objects and holes) was applied to eliminate isolated detections and to regularize region boundaries. This fused mask (Fig. 2.17) provided a robust preliminary estimate of collagen distribution.



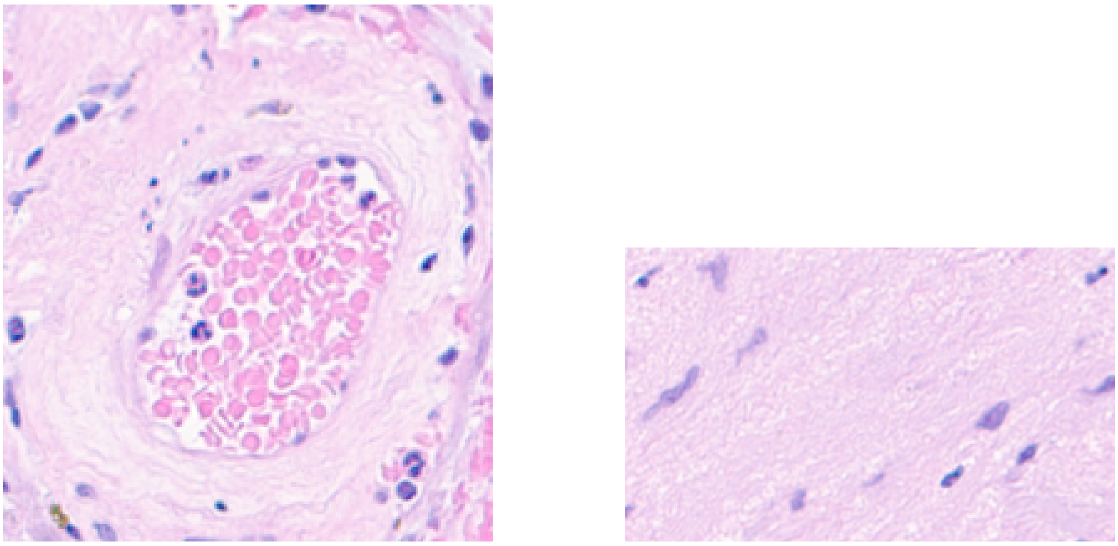
(a) **(b)**

Figure 2.17: (a) Normalized H&E tile. (b) Intermediate collagen candidate mask obtained by combining stain-driven and color-based criteria. The mask results from the intersection of the eosin-based threshold, hematoxylin suppression, and HSV/Lab chromatic constraints within the tissue region.

2.4.5 Structural analysis and fiber orientation filtering

Following stain-driven and chromatic segmentation, a structural filtering stage was introduced to improve specificity and to selectively retain regions exhibiting organized fibrous morphology consistent with collagen bundles. Although eosinophilic and chromatically compatible pixels may correspond to collagen, other extracellular or cytoplasmic components can produce similar intensity responses. Therefore, an additional structural constraint was imposed based on local orientation coherence.

Importantly, collagen in H&E slides does not only differ in chromatic appearance, but also in structural organization [27]. Collagen fibers may appear as highly aligned bundles following a dominant orientation, or as locally coherent yet globally curved structures, such as concentric perivascular arrangements (Fig. 2.18). This characteristic fibrous architecture distinguishes true collagen from disorganized eosinophilic regions, which typically lack directional consistency. Consequently, structural analysis was employed to quantify local orientation patterns and to preferentially retain pixels embedded within coherent, fiber-like configurations.



(a) **(b)**

Figure 2.18: Examples of collagen structural organization in H&E tiles. (a) concentric perivascular collagen arrangement, characterized by locally coherent but globally curved orientation. (b) more linearly aligned collagen bundles with a dominant local orientation. These patterns motivate the use of structure tensor–based analysis to quantify local coherence and orientation consistency.

Local orientation properties were estimated from the eosin channel using the structure tensor formalism [28]. Given the smoothed eosin intensity image, spatial gradients were computed and used to construct the second-moment matrix at each pixel:

$$\mathbf{J} = \begin{pmatrix} A_{xx} & A_{xy} \\ A_{xy} & A_{yy} \end{pmatrix},$$

where A_{xx}, A_{xy}, and A_{yy} denote locally averaged gradient products. The eigenvalues λ_1 and λ_2 of $\mathbf{J}$ were analytically derived and used to quantify the degree of anisotropy.

From these quantities, a coherence index was defined as

$$\text{Coherence} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \varepsilon},$$

where ε is a small constant introduced for numerical stability. The coherence measure ranges from 0 (isotropic structure) to 1 (highly anisotropic structure), and therefore provides a quantitative descriptor of local fiber organization.

In addition, the principal local orientation angle θ was estimated from the structure tensor components. To evaluate directional consistency within a neighborhood, a weighted vector representation based on $\sin(2\theta)$ and $\cos(2\theta)$ was spatially averaged using a box filter. The magnitude of the resulting vector field was normalized to obtain a directional consistency metric $R \in [0,1]$, reflecting the stability of local fiber orientation.

To suppress disorganized eosinophilic regions and retain preferentially fibrous collagen bundles, a structural gate was defined by thresholding both coherence and directional consistency:

$$\text{Gate}_{\text{struct}} = (\text{Coherence} \geq \tau_{\text{coh}}) \;\wedge\; (R \geq \tau_R).$$

Only pixels satisfying both criteria were retained. Mild morphological smoothing was subsequently applied to the structural gate in order to regularize small discontinuities.

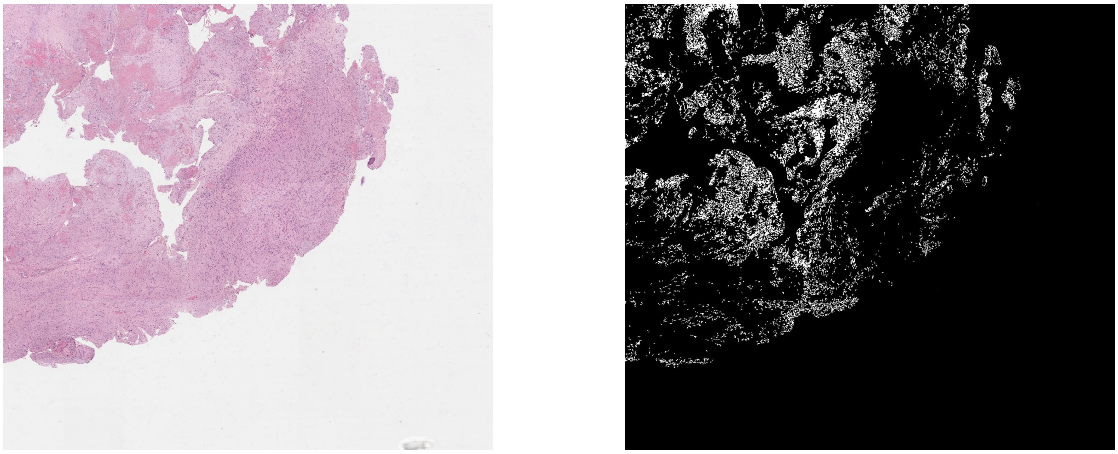
2.4.6 Final collagen mask and quantitative metrics

The final collagen mask was obtained by conservatively combining all previously defined criteria, integrating stain-driven, chromatic, and structural information into a single binary representation. The definitive collagen segmentation was defined as:

$$\mathrm{Mask}_{\mathrm{collagen}} = \mathrm{Mask}_{\mathrm{HED}\cap\mathrm{color}} \ \wedge \ \mathrm{Gate}_{\mathrm{struct}}.$$

This logical intersection ensured that only pixels simultaneously satisfying (i) eosin-based intensity constraints, (ii) chromatic consistency in HSV and CIELab spaces, and (iii) structural coherence and directional stability were retained. The resulting mask emphasized well-organized fibrous collagen while reducing false positives originating from amorphous eosinophilic regions or chromatically similar non-collagenous components.

To improve spatial consistency and eliminate residual isolated detections, a final morphological post-processing stage was applied. Small connected components were removed, small holes were filled, and mild opening–closing operations were performed to regularize region boundaries. These operations were intentionally conservative in order to preserve genuine thin collagen bundles while suppressing spurious pixel-level noise. The final collagen mask is shown in Fig. 2.19.



(a) **(b)**

Figure 2.19: Final collagen mask obtained after stain-based, chromatic, and structural filtering (b), shown alongside the corresponding normalized H&E tile (a). The mask represents the intersection of the stain-consistent, color-consistent, and orientation-based criteria, and highlights organized fibrous collagen regions while excluding background and non-fibrous eosinophilic areas.

Collagen quantification was performed by computing the fraction of collagen-positive area relative to the total tissue area:

$$\text{Collagen fraction} = \frac{A_{\text{collagen}}}{A_{\text{tissue}}},$$

where A_{collagen} denotes the area classified as collagen in the final mask and A_{tissue} represents the total tissue area within the previously defined tissue mask. Restricting the computation to tissue pixels ensured that background regions did not influence quantitative estimates.

In addition to area fraction, orientation information derived from the structure tensor was summarized over collagen-positive pixels. Specifically, a weighted mean orientation was computed using coherence as a weighting factor, providing a global descriptor of dominant collagen alignment within each tile. This metric does not replace the full spatial orientation map, but offers a compact quantitative summary of fiber organization.

For whole-slide analysis, tiles belonging to the same specimen were processed independently and subsequently aggregated. Total collagen and tissue pixel counts were summed across tiles, and the collagen fraction was computed at the slide level to obtain a robust specimen-specific quantitative estimate. This strategy ensured that local variability across tiles was preserved while providing a unified metric for downstream statistical analysis.

2.5 Validation of the Histological Collagen Quantification Pipeline

Although the proposed collagen quantification pipeline was specifically designed for hematoxylin and eosin (H&E)–stained whole-slide images, an independent validation step was required to assess its biological plausibility and quantitative reliability. In histopathology, Masson's trichrome staining is widely regarded as a reference method for collagen visualization, as collagen fibers are selectively highlighted in green tones, providing high contrast with respect to surrounding tissue components [29]. For this reason, Masson staining was adopted as a histological reference standard for validating the H&E-based collagen quantification approach.

The purpose of this validation was not to demonstrate absolute equivalence between staining modalities, but rather to evaluate whether the automated H&E-derived collagen estimates were quantitatively consistent with an established collagen-specific staining technique. The validation procedure therefore aimed to assess correlation, agreement, and systematic bias between the two methods at the slide level.

2.5.1 Validation Dataset and Pre-processing

The validation dataset consisted of 15 paired H&E and Masson's trichrome whole-slide images derived from arterial tissue sections. Although not tumor specimens, these slides were appropriate for validating the collagen quantification pipeline, as the aim was to assess the robustness and biological plausibility of H&E-based collagen extraction.

To maintain consistency with the previously described histological workflow, each whole-slide image was subdivided into a regular 3×3 grid, resulting in nine high-resolution tiles per slide. This tiling strategy allowed localized analysis while preserving computational tractability. In cases where structural inconsistencies were observed between paired H&E and Masson slides—such as missing tissue regions, sectioning artifacts, or non-corresponding fragments—affected tiles were excluded from analysis to ensure anatomical comparability between staining modalities.

All preprocessing steps applied to H&E slides (including tissue masking and tile-level processing) were replicated for the validation dataset. Collagen quantification in Masson-stained slides was performed using a dedicated computational pipeline specifically designed to exploit the chromatic properties of trichrome staining, as described in the following subsection.

2.5.2 Collagen Quantification in Masson-Stained Slides

Collagen quantification in Masson's trichrome–stained slides was performed using a dedicated computational procedure specifically designed to exploit the chromatic properties of trichrome staining. In Masson staining, collagen fibers are selectively highlighted in green tones, providing high color contrast with respect to cellular and muscular components. This property enables the definition of a color-based collagen intensity metric.

First, a binary tissue mask was generated to exclude background regions. Tissue detection was performed in the HSV color space, where pixels with high brightness and low saturation were classified as background and removed. Small isolated regions and holes were eliminated using morphological operations to obtain a spatially coherent tissue mask.

Subsequently, a collagen-specific intensity channel was computed from the RGB image. A green-dominance index was defined as:

$$C(x, y) = G(x, y) - 0.5R(x, y) - 0.5B(x, y),$$

where R, G, and B denote the red, green, and blue channels normalized to the range [0,1]. This formulation enhances pixels with predominant green content while penalizing contributions from red and blue components. Negative values were clipped to zero, and the resulting map was linearly normalized to [0,1].

Collagen pixels were then identified by applying a fixed intensity threshold exclusively within the tissue mask. The use of a fixed threshold ensured reproducibility across slides and avoided percentile-based adaptation. The resulting binary collagen mask was further refined through morphological opening, closing, and removal of small connected components to suppress isolated artifacts and stabilize region boundaries. An example of collagen extraction from a Masson-stained tile is shown in Figure 2.20.

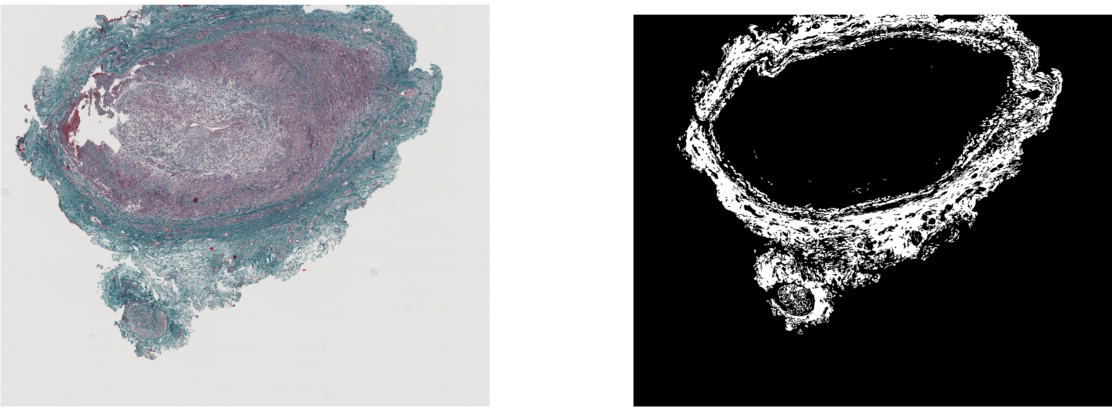


Figure 2.20: Example of collagen extraction from a Masson-stained slide. (a) Original tile, where collagen fibers appear in green tones. (b) Automatically generated binary collagen mask obtained using the green-dominance index and fixed thresholding within the tissue region. Collagen-rich regions are particularly evident around the vascular adventitia, where collagen concentration is typically higher.

Finally, collagen percentage at the slide level was computed as:

$$\%\text{Collagen} = 100 \times \frac{N_{\text{collagen}}}{N_{\text{tissue}}},$$

where N_{collagen} denotes the number of pixels classified as collagen and N_{tissue} the total number of tissue pixels.

2.5.3 Statistical Analysis for Method Comparison

Statistical analysis was performed to evaluate the agreement between collagen percentages derived from H&E-stained slides and those obtained from Masson's trichrome staining. All analyses were conducted at the slide level, using the total collagen percentage computed for each whole-slide image.

The association between the two measurement methods was first assessed using Pearson's correlation coefficient. This analysis quantified the strength of the linear relationship between H&E-based and Masson-based collagen estimates, providing an indication of overall concordance in ranking across samples.

However, correlation alone does not imply agreement. Therefore, method agreement was further evaluated using Bland–Altman analysis [30]. For each case, the difference between H&E-derived and Masson-derived collagen percentages was plotted against their mean value. This approach allowed the assessment of systematic bias and the identification of potential proportional effects across the measurement range.

In addition, the mean difference between methods was computed to estimate fixed bias, and limits of agreement were defined as the mean difference ±1.96 standard deviations of the differences. This analysis provided an estimate of the expected range of variation between the two techniques.

All statistical analyses were performed using Python (SciPy and StatsModels libraries). Statistical significance was defined at a two-sided threshold of $p < 0.05$.

2.6 Statistical Association Between Radiomic Features and Collagen Content

To investigate the relationship between MRI-derived radiomic features and tumor collagen content, a univariate correlation analysis was performed using the histologically derived collagen percentage as continuous target variable. Collagen percentage was obtained from the automated histological pipeline described in Section 2.4.

Radiomic features extracted from T2-weighted images and ADC maps were analyzed separately. Only patients with available data for both radiomics and histological quantification were included. Matching between imaging and histological datasets was performed through patient identifiers using an inner-join strategy, ensuring that only subjects with complete information were considered.

Since the distribution of both radiomic features and collagen percentage could not be assumed to be Gaussian and potential monotonic (non-linear) relationships were of interest, Spearman's rank correlation coefficient (ρ) was computed for each feature with respect to collagen percentage. Spearman correlation evaluates monotonic associations without assuming linearity or normality.

For each feature, the corresponding two-sided p-value was calculated. Given the large number of features tested, correction for multiple comparisons was applied using the Benjamini–Hochberg false discovery rate (FDR) procedure [31]. Adjusted p-values (FDR-corrected) were used to control the expected proportion of false positives.

Correlation coefficients were ranked according to the absolute value of ρ, allowing identification of features showing the strongest monotonic association with collagen content. Statistical significance was defined at a two-sided threshold of $p < 0.05$ after FDR correction.

2.7 Radiomics-Based Classification Models

2.7.1 Definition of Outcome Variables

Radiomics-based classification models were developed to predict clinically relevant tumor properties, including tumor consistency, intraoperative adherence, and postoperative facial nerve outcome.

Tumor consistency. Tumor consistency was initially defined according to the intraoperative assessment performed by the neurosurgeon. This evaluation considered surgical indicators such as the resistance of the tumor to ultrasonic aspiration and the degree of adherence to surrounding structures.

The original intraoperative classification distinguished three consistency categories (A, B, and C), reflecting increasing mechanical firmness. For the purposes of the present analysis, this semi-quantitative classification was simplified into a binary outcome by merging the intermediate class B with the higher-consistency class C, resulting in two categories: soft (A) and firm (B–C). This classification served as the clinical reference standard.

To obtain a quantitative and reproducible surrogate of tumor stiffness, the collagen percentage derived from the automated histological pipeline (Section 2.4) was used as a continuous biological marker. Collagen content represents a key determinant of extracellular matrix organization and biomechanical rigidity [8]. The relationship between collagen percentage and the surgeon-defined consistency classes was evaluated, and an optimal cutoff was determined using the Youden index [32]. This procedure identified the collagen percentage threshold that best separated soft and firm tumors according to the clinical classification.

Tumors were therefore categorized as:

- 0 (soft) for collagen percentage below the optimal cutoff,
- 1 (firm) for collagen percentage above the cutoff.

This approach allowed the transformation of a qualitative intraoperative assessment into an objective, quantitatively defined, and biologically grounded target variable. Importantly, cutoff determination was performed exclusively using histological information and did not involve radiomic features, thereby preventing information leakage between imaging data and target definition.

Intraoperative adherence. Tumor adherence to surrounding neural structures was defined according to intraoperative surgical reports and recorded as a binary variable. Unlike tumor consistency, which represents an intrinsic biomechanical property closely associated with extracellular matrix composition, adherence reflects a complex interaction between the tumor and adjacent neural structures. This phenomenon is influenced by multiple anatomical and surgical factors and cannot be directly reduced to a single histological metric. For this reason, intraoperative clinical assessment was retained as the most appropriate reference standard for adherence modeling:

- 0 (non-adherent),
- 1 (adherent).

Postoperative facial nerve outcome. Facial nerve outcome was defined as a binary endpoint based on postoperative clinical evaluation using the House–Brackmann grading system [33]. Patients with House–Brackmann grades I–II were considered to have a favorable outcome, whereas grades III–VI were classified as unfavorable. Patients were therefore categorized according to functional status, enabling formulation of a supervised classification task:

- 0 (favorable outcome),
- 1 (unfavorable outcome).

Each outcome variable was modeled independently using radiomic features extracted from MRI data.

2.7.2 Model Development and Validation Strategy

Radiomics-based classification models were developed to predict each outcome variable independently (tumor consistency, intraoperative adherence, and postoperative facial nerve outcome). Separate models were trained for T2-weighted features and ADC-derived features extracted from the segmented tumor volume as described in Section 2.3.

Preprocessing pipeline. Model training was implemented within a scikit-learn pipeline to prevent data leakage. The preprocessing steps included:

- Median imputation for handling missing values;
- Standardization using z-score normalization (zero mean, unit variance);
- Classification using a linear Support Vector Machine (SVM) [34].

Class imbalance was addressed using the `class_weight="balanced"` option within the SVM implementation. Probabilistic outputs were enabled to allow computation of receiver operating characteristic (ROC) curves and area under the curve (AUC).

Nested cross-validation. To ensure robust performance estimation and avoid optimistic bias during hyperparameter tuning, a nested cross-validation strategy was adopted [35].

The outer loop consisted of a stratified 5-fold cross-validation scheme, preserving class proportions in each fold. Within each training fold of the outer loop, hyperparameter optimization was performed using an inner 3-fold stratified cross-validation grid search.

The regularization parameter C of the linear SVM was optimized over the following grid:

$$C \in \{0.01, 0.1, 1, 10, 100\}.$$

For each outer fold, the best hyperparameter configuration identified in the inner loop was used to train the model on the training subset and generate predictions on the held-out test subset.

Out-of-fold predictions. Out-of-fold (OOF) probabilities were computed by aggregating predictions from the outer cross-validation folds. This procedure ensured that each patient received exactly one prediction generated by a model that had not been trained on that subject.

2.7.3 Performance Metrics

Model performance was evaluated using out-of-fold (OOF) predictions obtained from the outer cross-validation loop. All metrics were computed at the patient level.

Discrimination performance. The primary performance metric was the area under the receiver operating characteristic curve (AUC). The ROC curve represents the relationship between sensitivity (true positive rate) and 1–specificity (false positive rate) across all possible probability thresholds. The AUC quantifies the model's ability to discriminate between the two classes independently of a specific decision threshold. An AUC value of 0.5 indicates no discriminative ability, whereas a value of 1.0 indicates perfect discrimination.

To estimate uncertainty in the AUC, non-parametric bootstrap resampling was performed using 2000 iterations. In each iteration, patients were resampled with replacement, and the AUC was recalculated. The 95% confidence interval (CI) was derived from the 2.5th and 97.5th percentiles of the bootstrap distribution.

Threshold-dependent metrics. In addition to AUC, threshold-dependent classification metrics were computed using a fixed probability threshold of 0.5. Predicted probabilities greater than or equal to 0.5 were assigned to the positive class. The positive class was defined according to the clinically relevant condition for each outcome (e.g., firm consistency, presence of adherence, or unfavorable facial nerve outcome). Here, TP (true positives) denotes correctly classified positive cases, TN (true negatives) correctly classified negative cases, FP (false positives) incorrectly classified positive predictions, and FN (false negatives) missed positive cases.

The following performance measures were calculated:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{Sensitivity} = \frac{TP}{TP + FN},$$

$$\text{Specificity} = \frac{TN}{TN + FP},$$

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP + FP},$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN + FN}.$$

2.7.4 Feature Importance Analysis and Reduced-Feature Models

To investigate the relative contribution of individual radiomic features to model predictions and to explore their potential biological relevance, a feature importance analysis was performed for the linear SVM models.

Feature importance estimation. Given the use of a linear kernel, feature importance was derived from the absolute value of the model coefficients. In a linear SVM, the decision function is defined as:

$$f(x) = w^T x + b,$$

where w represents the weight vector and b the bias term. The magnitude of each coefficient $|w_j|$ reflects the contribution of the corresponding feature to the classification decision, after feature standardization.

Feature importance values were extracted from the trained models within each outer cross-validation fold. To obtain a stable ranking and reduce variability related to data partitioning, coefficient magnitudes were averaged across folds. Features were then ranked in descending order based on their mean absolute coefficient value.

Beyond model interpretability, this analysis aimed to assess whether highly ranked features reflected image patterns potentially associated with underlying tumor microstructure and structural organization, thereby providing insight into the biological plausibility and clinical coherence of the radiomics-based predictions across the investigated outcomes.

Reduced-feature models. To evaluate whether comparable predictive performance could be achieved using a smaller and potentially more interpretable subset of predictors, additional models were explored using only the top-ranked features. Feature subsets were defined based on the importance ranking derived from the full-feature nested cross-validation analysis.

These reduced-feature models were intended as an exploratory assessment of model parsimony and interpretability rather than as independent validation experiments. The same preprocessing steps, nested cross-validation strategy, and performance evaluation procedures described in Sections 2.6.2 and 2.6.3 were applied to ensure methodological consistency and to enable a fair comparison with the full-feature models.

Chapter 3

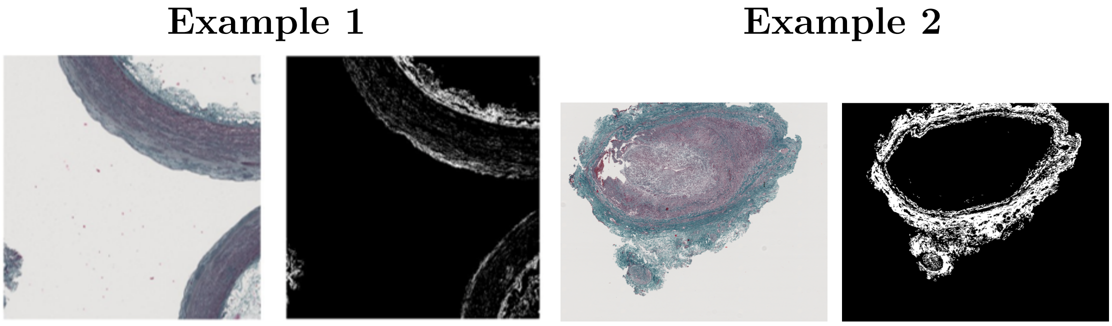
Results

3.1 Results of the Histological Validation

The methodological details of the validation procedure are described in Section 2.5. The present section reports the quantitative results of the comparison between collagen percentages derived from the automated H&E-based pipeline and those obtained from Masson's trichrome staining, used as histological reference standard.

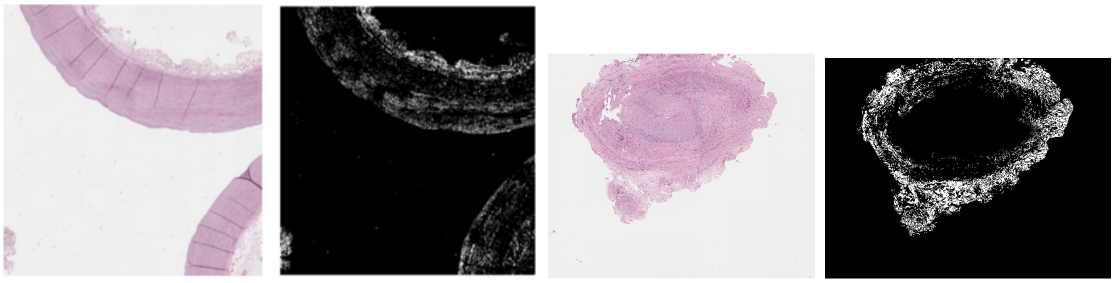
Qualitative comparison between H&E and Masson segmentation

A representative visual comparison between the two staining modalities is shown in Figure 3.1. The collagen mask extracted from H&E slides closely follows the green-stained collagen regions in the corresponding Masson slides. The overall spatial distribution of collagen-rich areas appears consistent across modalities. In Example 2, collagen surrounding a vascular structure (adventitia) is clearly visible in the Masson-stained image and corresponds to regions with the highest collagen concentration. This structure provides a particularly clear visual reference for comparing the segmentation results between the two staining modalities. Localized differences are mainly observed in regions where collagen in H&E exhibits chromatic similarity to adjacent muscular or eosinophilic tissue, leading to mild overestimation in selected areas.



Example 1 **Example 2**

Masson (original) Masson (mask) Masson (original) Masson (mask)



H&E (normalized) H&E (mask) H&E (normalized) H&E (mask)

Figure 3.1: Qualitative comparison between Masson's trichrome and H&E-derived collagen segmentation in two representative examples. For each example, the original tile and the corresponding collagen mask are reported for Masson (top row) and for H&E after color normalization (bottom row). The H&E-based masks closely reproduce the spatial distribution of collagen identified by Masson staining, with minor localized discrepancies in regions where collagen exhibits chromatic overlap with adjacent tissue.

Correlation and regression analysis

At the slide level, a strong positive linear association was observed between the two measurement methods. Pearson's correlation coefficient was $r = 0.89$, corresponding to a coefficient of determination $R^2 = 0.79$, indicating that approximately 79% of the variability in H&E-derived collagen percentage can be explained by the corresponding Masson-derived measurements.

A simple ordinary least squares regression was fitted to model the relationship between the two techniques. The estimated regression equation was:

$$\mathrm{HE}(\%) = 0.50 \cdot \mathrm{Masson}(\%) + 15.2$$

The scatter plot and regression line are shown in Figure 3.2.

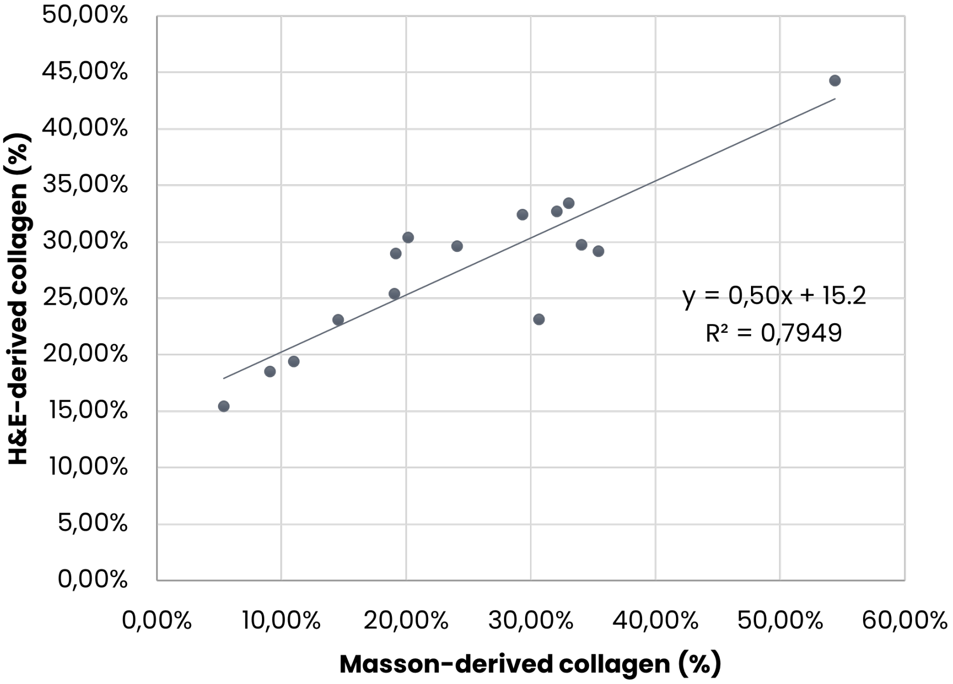


Figure 3.2: Scatter plot of H&E-derived versus Masson-derived collagen percentages at the slide level. The solid line represents the ordinary least squares regression.

Agreement analysis

Absolute error metrics were computed to quantify the magnitude of disagreement between methods. The mean absolute error (MAE) was 6.70%, and the root mean squared error (RMSE) was 7.30%, indicating a moderate average deviation between H&E-based and Masson-based collagen estimates.

Agreement between the two measurement techniques was further evaluated using Bland–Altman analysis. The mean difference (H&E minus Masson) was +2.95%, indicating a slight average overestimation of collagen content by the H&E-based method. The limits of agreement (mean difference ±1.96 standard deviations) ranged from −10.9% to +16.8%, as shown in Figure 3.3.

Visual inspection of the Bland–Altman plot suggests a trend across the measurement range, with H&E-derived estimates tending to overestimate collagen content at lower percentages and to underestimate it at higher percentages. However, most observations remained within the calculated limits of agreement.

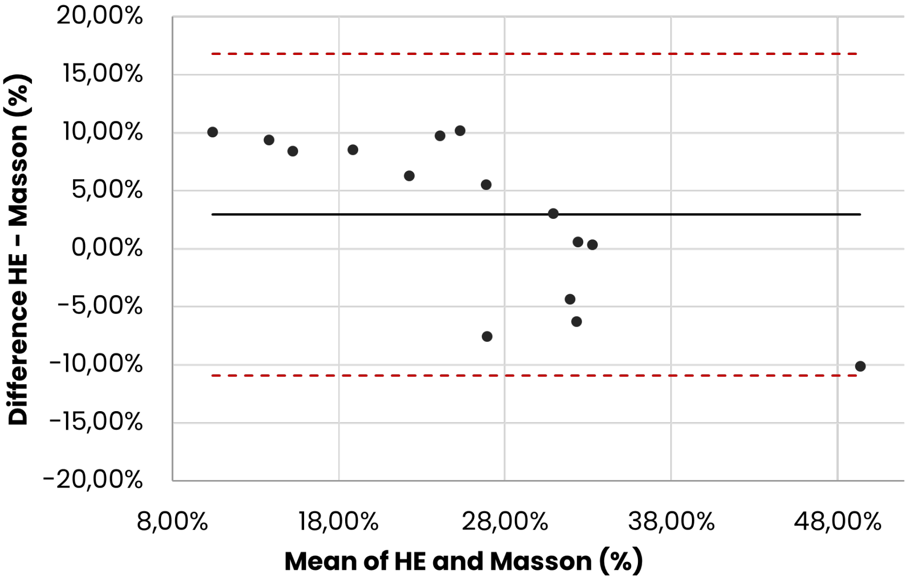


Figure 3.3: Bland–Altman plot comparing H&E-derived and Masson-derived collagen percentages. The central dashed line represents the mean difference, while the outer lines indicate the 95% limits of agreement.

Overall, these findings support a strong quantitative concordance between the automated H&E-based collagen quantification pipeline and the Masson's trichrome reference method at the slide level.

3.2 Association Between Radiomic Features and Histological Collagen Content

The statistical methodology adopted for correlation analysis is described in Section 2.3. Spearman's rank correlation was computed between radiomic features extracted from MRI and the histologically derived collagen percentage, with false discovery rate (FDR) correction applied to account for multiple comparisons.

T2-weighted MRI

In the T2-weighted cohort ($N = 45$), several radiomic features demonstrated moderate monotonic associations with collagen percentage. After false discovery rate (FDR) correction, however, none of the examined features reached statistical significance, although multiple descriptors showed trend-level associations.

The strongest correlations were observed among first-order intensity features, with absolute Spearman correlation coefficients in the range $|\rho| \approx 0.20$–0.30.

Specifically, lower-order statistical descriptors such as the 10th percentile, median intensity, mean intensity, and skewness exhibited moderate correlations with collagen content. The direction of association indicated that increasing collagen percentage was associated with systematic shifts in global intensity distribution. Although these effects did not remain statistically significant after multiple-comparison correction, their consistent directionality suggests a biologically plausible relationship between collagen deposition and global signal intensity characteristics on T2-weighted imaging.

Table 3.1: Top T2-weighted radiomic features associated with histological collagen percentage after FDR correction.

Feature	Spearman ρ	Adjusted p-value
FirstOrder_10Percentile	0.30	0.081
FirstOrder_Median	0.24	0.092
GLCM_ClusterShade	-0.22	0.104
FirstOrder_Mean	0.22	0.108
FirstOrder_Skewness	-0.20	0.110
GLDM_LargeDependenceHighGrayLevelEmphasis	0.20	0.115
FirstOrder_RootMeanSquared	0.20	0.118
GLDM_LargeDependenceLowGrayLevelEmphasis	-0.19	0.122
GLSZM_LargeAreaHighGrayLevelEmphasis	0.18	0.128
GLCM_Imc1	0.18	0.130

ADC maps

In the ADC cohort ($N = 40$), stronger associations were observed compared to T2-weighted imaging. Multiple texture-based radiomic features remained statistically significant after FDR correction, predominantly involving descriptors derived from the gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), and gray-level size zone matrix (GLSZM).

Absolute Spearman correlation coefficients ranged from $|\rho| \approx 0.33$–0.39, indicating moderate-to-strong monotonic associations between collagen content and diffusion-related textural heterogeneity.

Entropy-related descriptors and non-uniformity measures demonstrated the highest correlation magnitude, suggesting that increasing collagen percentage is associated with increased microstructural complexity in ADC maps.

Table 3.2: Top ADC-derived radiomic features associated with histological collagen percentage after FDR correction.

Feature	Spearman ρ	Adjusted p-value
GLCM_MCC	0.39	0.016
GLSZM_SmallAreaLowGrayLevelEmphasis	-0.39	0.018
GLCM_ClusterTendency	0.35	0.030
GLCM_SumEntropy	0.35	0.032
FirstOrder_Entropy	0.34	0.034
GLRLM_GrayLevelNonUniformityNormalized	-0.34	0.038
GLCM_SumSquares	0.34	0.042
GLCM_Imc2	0.33	0.045
GLRLM_RunEntropy	0.33	0.052
GLSZM_GrayLevelNonUniformityNormalized	-0.33	0.055

Comparison between modalities

Overall, ADC-derived radiomic features exhibited higher correlation magnitudes and greater statistical robustness with respect to histological collagen percentage compared to T2-weighted features.

While T2 associations were primarily driven by global intensity distribution metrics, ADC correlations predominantly involved higher-order texture descriptors reflecting spatial heterogeneity.

These findings indicate that radiomic features extracted from both T2-weighted MRI and ADC maps capture quantitative information related to tumor collagen content, with diffusion-based textural descriptors demonstrating greater sensitivity and statistical stability in relation to histologically measured fibrosis.

3.3 Classification performance

The performance of the classification framework described in Section 2.6 is reported below for each investigated endpoint.

3.3.1 Tumor consistency

T2-weighted MRI

The T2-weighted cohort included $N = 45$ patients, with 27 firm tumors (class 1) and 18 soft tumors (class 0). A total of 93 radiomic features were extracted for each case.

The T2-based model achieved an area under the ROC curve (AUC) of 0.82 (95% CI: 0.73–0.92, $p < 0.001$), indicating good discriminative performance between firm and soft tumors.

At the selected optimal operating point, the model yielded an accuracy of 0.80, with sensitivity of 0.89 and specificity of 0.75, corresponding to a balanced accuracy of 0.82. Positive predictive value (PPV) and negative predictive value (NPV) were 0.79 and 0.77, respectively.

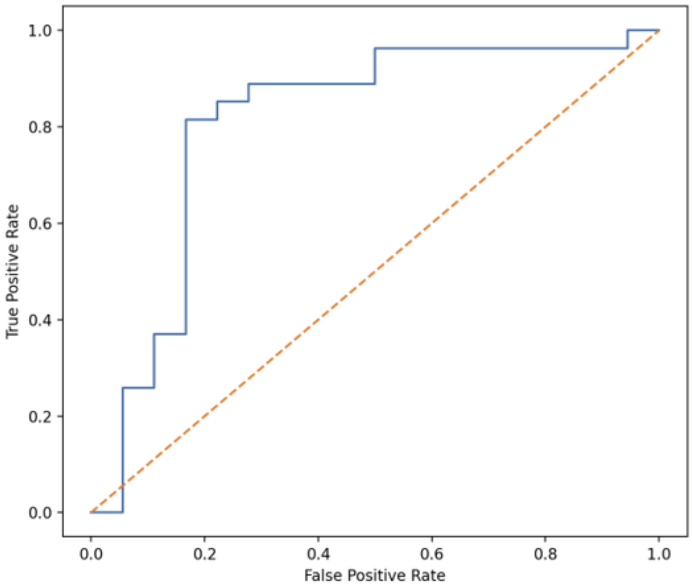


Figure 3.4: Receiver operating characteristic (ROC) curve for T2-weighted MRI-based prediction of tumor consistency.

ADC maps

The ADC cohort comprised $N = 40$ patients, including 23 firm tumors (class 1) and 17 soft tumors (class 0). As for T2-weighted imaging, 93 radiomic features were extracted per case.

The ADC-based model achieved an AUC of 0.81 (95% CI: 0.74–0.90, $p < 0.001$), demonstrating comparable discriminative ability to T2-weighted imaging.

At the optimal threshold, classification accuracy was 0.78, with sensitivity of 0.87 and specificity of 0.74. Balanced accuracy reached 0.81. PPV and NPV were 0.80 and 0.77, respectively.

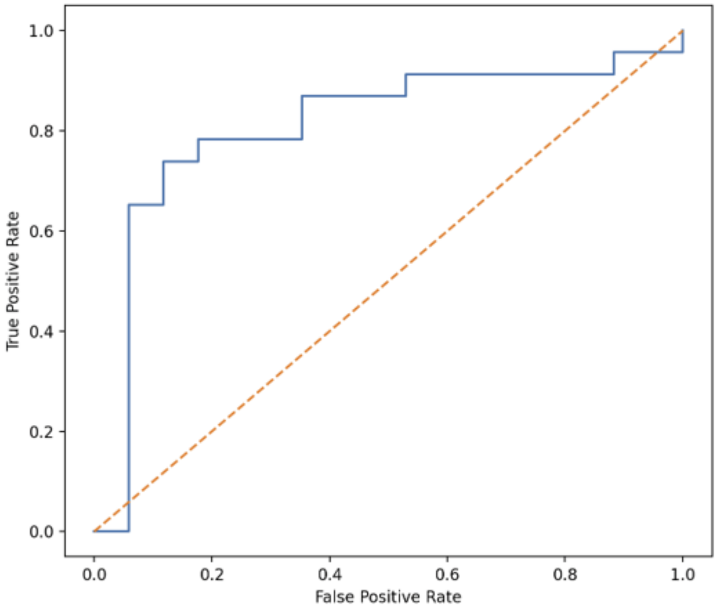


Figure 3.5: Receiver operating characteristic (ROC) curve for ADC-based prediction of tumor consistency.

Comparative analysis

Direct comparison of classification metrics revealed comparable predictive capabilities between T2-weighted and ADC-derived radiomic features (Table 3.3).

Both modalities achieved AUC values above 0.80 and balanced accuracies above 0.80, indicating good discrimination between firm and soft tumors.

Table 3.3: Summary of classification performance metrics for tumor consistency prediction.

Modality	AUC	Accuracy	Sens	Spec	Bal Acc	PPV	NPV
T2-weighted MRI	0.82	0.80	0.89	0.75	0.82	0.79	0.77
ADC maps	0.81	0.78	0.87	0.74	0.81	0.80	0.77

3.3.2 Tumor adherence

T2-weighted MRI

In the T2-weighted dataset ($N = 43$), 22 tumors were classified as adherent (class 1) and 21 as non-adherent (class 0). 93 radiomic features were available for model training and evaluation.

The resulting classifier yielded an AUC of 0.77 (95% CI: 0.63–0.90, $p < 0.001$), indicating a moderate ability to discriminate between adherent and non-adherent lesions.

At the optimal decision threshold, overall accuracy reached 0.78. Sensitivity was 0.82, whereas specificity was 0.72, corresponding to a balanced accuracy of 0.77. The positive and negative predictive values were 0.75 and 0.79, respectively.

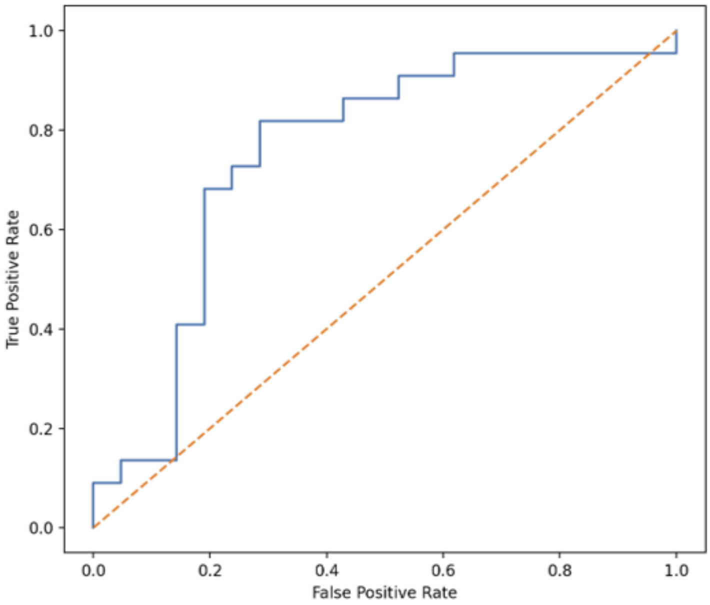


Figure 3.6: ROC curve for T2-weighted MRI-based prediction of tumor adherence.

ADC maps

For ADC imaging ($N = 39$), the cohort included 19 adherent and 20 non-adherent tumors. The same feature set (93 radiomic descriptors) was employed.

The ADC-based model achieved an AUC of 0.80 (95% CI: 0.64–0.92, $p < 0.001$), slightly exceeding the performance observed with T2-weighted imaging.

Classification accuracy was 0.75 at the selected threshold. In contrast to T2-weighted MRI, ADC demonstrated higher specificity (0.80) than sensitivity (0.70), yielding a balanced accuracy of 0.75. PPV and NPV were 0.77 and 0.73, respectively.

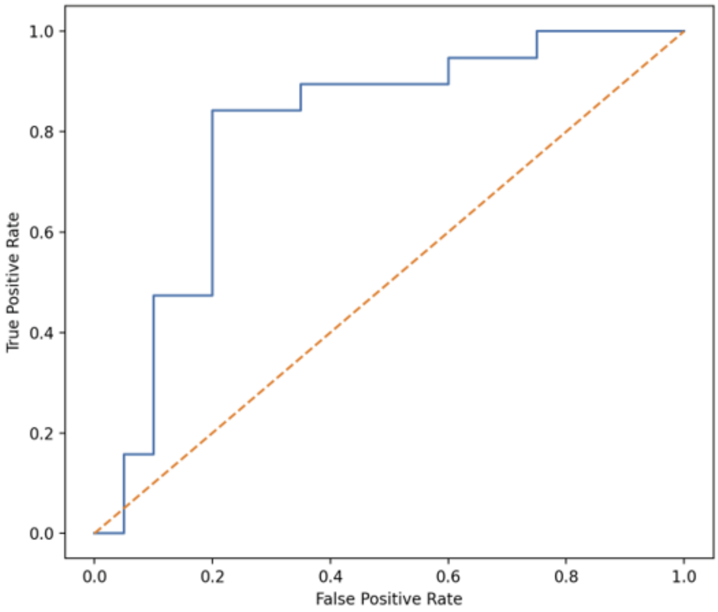


Figure 3.7: ROC curve for ADC-based prediction of tumor adherence.

Comparative analysis

Overall, both imaging modalities provided comparable predictive performance for adherence assessment. ADC maps showed a marginally higher AUC, whereas T2-weighted MRI achieved greater sensitivity. The balanced accuracy values were closely aligned between modalities, suggesting stable classification capability across imaging contrasts.

Table 3.4: Summary of classification performance metrics for tumor adherence prediction.

Modality	AUC	Accuracy	Sens	Spec	Bal Acc	PPV	NPV
T2-weighted MRI	0.77	0.78	0.82	0.72	0.77	0.75	0.79
ADC maps	0.80	0.75	0.70	0.80	0.75	0.77	0.73

3.3.3 Post-operative facial nerve outcome

T2-weighted MRI

The T2-weighted cohort comprised $N = 43$ patients, including 18 cases with unfavorable post-operative facial nerve outcome (class 1) and 25 with favorable outcome (class 0). The same 93 radiomic features previously used for consistency and adherence prediction were retained for this analysis.

The T2-based classifier achieved an AUC of 0.71 (95% CI: 0.59–0.83, $p < 0.001$), reflecting modest discriminative capability for early post-operative facial nerve status.

At the selected operating point, overall accuracy was 0.73. Sensitivity and specificity were 0.65 and 0.82, respectively, yielding a balanced accuracy of 0.74. The corresponding PPV and NPV were 0.73 and 0.70.

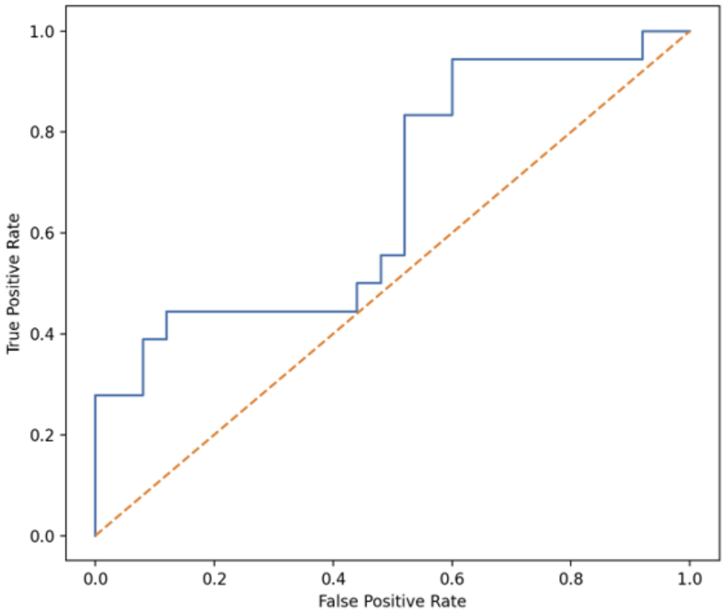


Figure 3.8: ROC curve for T2-weighted MRI-based prediction of post-operative facial nerve outcome.

ADC maps

For ADC imaging ($N = 39$), 15 patients presented an unfavorable outcome and 24 a favorable outcome. The same radiomic feature set was used.

The ADC-based model yielded an AUC of 0.61 (95% CI: 0.54–0.74, $p < 0.01$), indicating limited discriminative performance compared with T2-weighted imaging.

Classification accuracy was 0.67, with sensitivity of 0.58 and specificity of 0.71, corresponding to a balanced accuracy of 0.65. PPV and NPV were 0.55 and 0.68, respectively.

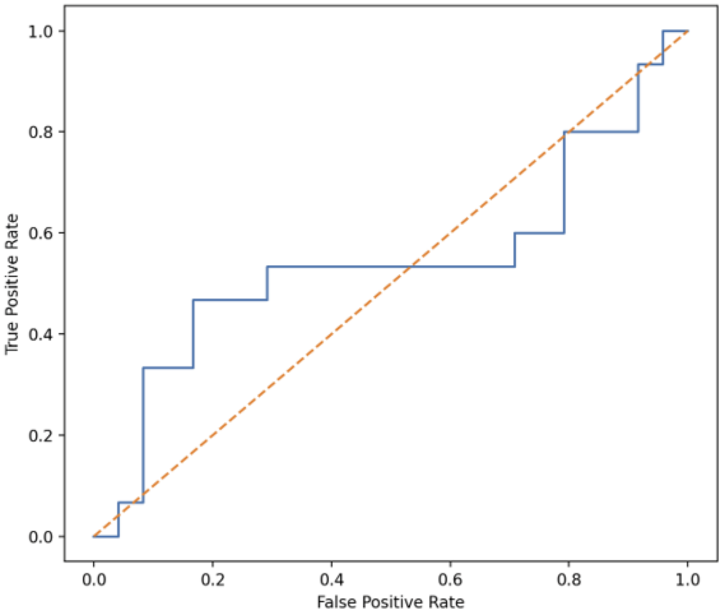


Figure 3.9: ROC curve for ADC-based prediction of post-operative facial nerve outcome.

Comparative analysis

Compared with tumor consistency and adherence prediction, classification performance for post-operative facial nerve outcome was generally lower, particularly for ADC-derived features. T2-weighted MRI demonstrated moderate discriminative ability, whereas ADC maps showed limited predictive power for this endpoint.

Table 3.5: Summary of classification performance metrics for post-operative facial nerve outcome prediction.

Modality	AUC	Accuracy	Sens	Spec	Bal Acc	PPV	NPV
T2-weighted MRI	0.71	0.73	0.65	0.82	0.74	0.73	0.70
ADC maps	0.61	0.67	0.58	0.71	0.65	0.55	0.68

3.3.4 Feature importance and misclassification patterns

To further characterize the behavior of the radiomics-based classification framework, feature importance analysis and an assessment of misclassification patterns were performed.

Global feature importance

Global feature importance was derived from the coefficients of the linear support vector machine described in Section 2.6. The ranking reflects the average absolute weight ($|w|$) associated with each radiomic feature, quantifying its overall contribution to the decision boundary. The sign of the coefficient was used to determine the direction of association with the predicted class.

Consistency prediction For T2-weighted MRI, the most influential features predominantly described signal heterogeneity and global intensity organization. Increased heterogeneity and broader intensity distributions were generally associated with soft tumors, whereas more compact and structurally organized intensity patterns characterized firm lesions.

In ADC maps, the top-ranked features mainly reflected diffusion distribution and microstructural variability. Soft tumors tended to exhibit greater diffusion heterogeneity, while firm tumors presented more asymmetric and spatially structured diffusion profiles.

Adherence prediction For adherence prediction, T2-derived features primarily captured spatial organization and textural complexity. Elevated entropy, structural irregularity, and dependency-related measures were associated with adherent tumors, whereas more homogeneous and repetitive patterns characterized non-adherent lesions.

Similarly, ADC-based features emphasized diffusion heterogeneity and non-uniformity. Higher entropy and dispersion metrics were linked to tumor adherence, supporting the hypothesis that microstructural irregularity at the tumor–nerve interface can be partially captured by diffusion-based radiomics.

Stability of reduced feature models To evaluate the robustness of the identified radiomic signatures, classification performance was reassessed using only the top 20 ranked features for each endpoint. Discriminative ability remained comparable to that obtained with the full feature set.

For tumor consistency, the AUC was 0.83 with T2-weighted MRI and 0.79 with ADC, closely matching the performance achieved using all features. Similarly, for adherence prediction, the reduced models yielded AUC values of 0.77 (T2) and 0.83 (ADC), confirming that a limited subset of highly informative features retains most of the predictive signal.

Misclassification patterns

An analysis of misclassified cases was conducted to better understand model limitations and potential complementarity between imaging modalities.

Consistency Both T2-weighted and ADC models showed a predominance of false positives (T2: 7 FP vs 2 FN; ADC: 6 FP vs 3 FN), indicating a tendency to overestimate the “firm” class. A substantial overlap of misclassified patients (6 shared cases) was observed between modalities, suggesting that certain tumors present intrinsically ambiguous radiomic profiles that remain challenging irrespective of imaging contrast.

Adherence For adherence prediction, the overall number of errors was comparable between modalities, but their distribution differed. The T2 model exhibited more false positives (6 FP vs 4 FN), whereas the ADC model showed a higher number of false negatives (6 FN vs 4 FP).

The limited overlap of misclassified cases (2 shared patients) indicates that T2-weighted MRI and ADC maps may capture partially distinct aspects of tumor phenotype, suggesting potential complementarity in adherence assessment.

Chapter 4

Discussion

4.1 Biological Interpretation of Collagen–MRI Associations

The present study demonstrated that quantitative radiomic features extracted from preoperative MRI capture imaging signatures associated with histologically measured tumor collagen content. The observed associations provide biological support for the hypothesis that radiomic descriptors reflect underlying microstructural properties of vestibular schwannomas [11].

Collagen represents a major structural component of the extracellular matrix and contributes to tumor stiffness and architectural organization [8]. Variations in collagen density and distribution are expected to alter both signal intensity patterns on T2-weighted imaging and diffusion behavior on ADC maps. The correlations identified in this study suggest that these microstructural differences are partially encoded within radiomic features.

In particular, diffusion-based radiomics exhibited stronger associations with collagen content compared with T2-weighted imaging. This finding is biologically plausible. Increased collagen deposition leads to greater extracellular matrix density and reduced water mobility, resulting in alterations of diffusion heterogeneity and spatial complexity [18]. Texture features derived from gray-level co-occurrence and size-zone matrices are sensitive to such spatial organization, capturing subtle variations in tissue architecture that are not directly appreciable on visual inspection.

Conversely, T2-weighted radiomic features showed weaker but still consistent associations, primarily driven by global intensity distribution metrics. This likely reflects the indirect relationship between collagen and T2 signal, which is influenced not only by extracellular matrix composition but also by cellularity, fluid content, and microcystic changes [16].

Overall, the concordance between histological collagen quantification and MRI-derived radiomic features supports the concept that radiomics can serve as a non-invasive surrogate of tumor microstructure. While imaging does not directly measure collagen, the identified associations indicate that structural remodeling of the tumor stroma leaves measurable signatures in routine MRI sequences.

Importantly, the reliability of the collagen quantification pipeline was supported by the validation analysis against Masson's trichrome staining. The observed agreement suggests that the automated H&E-based approach provides a reasonably accurate proxy for collagen estimation. Although minor systematic differences were observed, the overall concordance indicates that the proposed pipeline provides reliable quantitative estimates of collagen content that can be directly linked to MRI-derived radiomic features.

4.2 Prediction of Surgical Endpoints

The primary contribution of this study concerns the preoperative prediction of clinically relevant surgical endpoints using radiomic features derived from routine MRI sequences.

Tumor consistency Radiomics-based models demonstrated solid discriminative performance in predicting tumor consistency. The ability to differentiate firm from soft lesions preoperatively is of particular surgical relevance, as tumor stiffness directly influences resection strategy, operative time, and technical difficulty [2, 14].

The observed predictive performance suggests that MRI-derived radiomic features capture structural information related to the extracellular matrix organization and collagen architecture. This interpretation is coherent with the previously identified associations between collagen content and imaging descriptors. Firm tumors, characterized by higher stromal organization and denser collagen networks, likely exhibit more structured intensity and diffusion patterns, which are detectable through texture-based analysis.

The stability of model performance when restricting the analysis to the top-ranked features further supports the robustness of the identified radiomic signature. This indicates that a limited subset of biologically meaningful descriptors accounts for most of the predictive signal, reinforcing interpretability and reducing the likelihood of overfitting.

Tumor adherence Prediction of intraoperative tumor adherence showed moderate but consistent performance across imaging modalities. Although discriminative accuracy was lower than that observed for consistency, the results remain clinically relevant.

Tumor adherence to adjacent neurovascular structures is a complex phenomenon that likely depends on both mechanical properties and local microstructural irregularity at the tumor–nerve interface. Radiomic features reflecting spatial complexity and diffusion heterogeneity may partially encode this interface-level organization.

The complementary error patterns observed between T2-weighted MRI and ADC maps suggest that these modalities capture partially distinct aspects of tumor phenotype. This finding raises the possibility that multimodal integration could further improve adherence prediction in future studies.

Post-operative facial nerve outcome In contrast, prediction of early post-operative facial nerve outcome yielded more limited performance, particularly for diffusion-based models. This result is not unexpected. Unlike tumor consistency or adherence, facial nerve function after surgery is influenced by multiple intraoperative and patient-specific factors, including surgical manipulation, nerve traction, vascular compromise, and individual regenerative capacity [5].

The comparatively weaker predictive capability observed for this endpoint supports the interpretation that radiomic features primarily reflect intrinsic tumor properties rather than the full spectrum of perioperative determinants of functional outcome. Consequently, facial nerve prediction should be regarded as an exploratory analysis rather than a primary objective of the present work.

Overall, these results suggest that radiomics-based models may support preoperative surgical planning by providing objective estimates of tumor consistency and adherence. Imaging-derived features offer a structured and quantitative framework to anticipate mechanical and interface-related tumor behavior before surgery.

From a methodological perspective, the predictive models showed a coherent pattern of performance across the evaluated endpoints. The highest discriminative capability was observed for tumor consistency, which is closely related to intrinsic tumor microstructure and extracellular matrix organization. Moderate performance for adherence prediction likely reflects the partially interface-dependent nature of this endpoint, while the comparatively weaker results for facial nerve outcome further confirm that radiomic features primarily capture intrinsic tumor properties rather than the full spectrum of perioperative determinants of functional recovery.

4.3 Clinical Implications, Limitations and Future Perspectives

The results of this study suggest that MRI-based radiomics may provide quantitative information relevant to preoperative surgical planning in vestibular schwannomas. In particular, the ability to estimate tumor consistency and adherence non-invasively may complement conventional radiological assessment and support operative strategy. Nevertheless, several methodological limitations must be carefully considered.

A first set of limitations concerns the histological collagen quantification pipeline. The study was conducted on a relatively small sample size, which inevitably constrained both methodological choices and validation strength. In particular, the limited number of available slides precluded the adoption of data-intensive deep learning approaches for stain segmentation and collagen quantification. Instead, an automated threshold-based method was implemented, requiring operator-defined parameters. Although conservative filtering and color normalization were applied to enhance reproducibility, this approach remains partially operator-dependent and may introduce bias.

The limited sample size also affects the robustness of the pipeline validation. Quantitative concordance with Masson's trichrome staining was assessed on a restricted number of slides, and although strong agreement was observed, validation on only 15 specimens cannot be considered definitive. Larger cohorts are required to confirm reproducibility across broader staining variability and inter-sample heterogeneity.

Furthermore, the H&E-based pipeline showed a tendency to mildly overestimate collagen content in regions where non-collagenous structures—such as muscle fibers or similarly stained stromal components—exhibited comparable chromatic properties. While morphological constraints were introduced to mitigate this effect, complete discrimination between collagen and morphologically similar structures cannot be guaranteed.

Future developments of the histological component should therefore include validation on larger and more diverse datasets. The use of supervised deep learning segmentation models may improve tissue discrimination, provided that adequate annotated datasets are available. Generative approaches could potentially be explored to augment limited datasets; however, synthetic slide generation would require rigorous validation to ensure biological realism and avoid introducing artificial bias. Multicenter histological collections would represent the most robust strategy for strengthening the biological reference standard.

In addition, methodological developments could also include the application of histological image registration techniques to spatially align H&E and Masson's trichrome slides. Such approaches would enable more direct pixel-level correspondence between the two stainings, allowing a more precise validation of collagen segmentation results.

With respect to the radiomics-based predictive models, the relatively small cohort size also represents a significant constraint. Although repeated cross-validation was employed to reduce overfitting, the absence of external validation limits the generalizability of the findings. Larger, multicenter datasets are necessary to confirm the stability of the identified radiomic signatures.

More complex modeling strategies, including deep learning architectures, might further enhance predictive performance by capturing non-linear relationships between imaging features and surgical endpoints. However, such approaches require substantially larger datasets to avoid overfitting and loss of interpretability. Given the available sample size, the adoption of a linear support vector machine represented a deliberate balance between model transparency, statistical stability, and interpretability.

Finally, as discussed previously, post-operative facial nerve outcome reflects a multifactorial process extending beyond intrinsic tumor characteristics. Imaging-derived radiomics can capture structural tumor properties but cannot fully account for intraoperative and patient-specific determinants of functional recovery. For this reason, facial nerve prediction should be interpreted with appropriate caution and viewed as exploratory.

Future research should aim at validating both the collagen quantification pipeline and the radiomics-based predictive models in larger, independent cohorts. Integration of multimodal imaging features within unified predictive frameworks, as well as incorporation of clinical variables, may enable more comprehensive preoperative risk stratification. Such developments would help translate biologically informed radiomics into clinically applicable decision-support tools.

Bibliography

[1] Nguyen Duy and de Kantzow Lucy. «Vestibular schwannomas: A Review». In: *Applied Radiology* 3 (2019), pp. 22–27 (cit. on pp. 1, 2).

[2] William R. Copeland, Jason M. Hoover, Jonathan M. Morris, Colin L. W. Driscoll, and Michael J. Link. «Use of preoperative MRI to predict vestibular schwannoma intraoperative consistency and facial nerve outcome». In: *Journal of Neurological Surgery Part B: Skull Base* 74.6 (2013), pp. 347–350. DOI: `10.1055/s-0033-1347369` (cit. on pp. 2, 3, 61).

[3] Jitendra D. Thakur, Imran S. Khan, Christopher D. Shorter, Paul A. Gardner, Takanori Fukushima, and Bharat Guthikonda. «Do cystic vestibular schwannomas have worse surgical outcomes? A systematic review of the literature». In: *Neurosurgical Focus* 33.3 (2012), E12. DOI: `10.3171/2012.6.FOCUS12200` (cit. on p. 2).

[4] Xuan Wu, Guoqing Song, Xin Wang, Wenjun Liu, and Xiaofeng Yang. «Comparison of surgical outcomes in cystic and solid vestibular schwannomas: A systematic review and meta-analysis». In: *Neurosurgical Review* 44.4 (2021), pp. 1889–1902. DOI: `10.1007/s10143-020-01400-5` (cit. on p. 2).

[5] T. K. Y. Lee, W. S. Lund, and C. B. T. Adams. «Factors influencing preservation of the facial nerve during acoustic neuroma surgery». In: *British Journal of Neurosurgery* 4.1 (1990), pp. 5–8. DOI: `10.3109/02688699009000675` (cit. on pp. 2, 62).

[6] Cheng Yang, Daniel Alvarado, Pawan Kishore Ravindran, Max E. Keizer, Koos Hovinga, Martinus P. G. Broen, Henricus P. M. Kunst, and Yasin Temel. «MRI-Based Prediction of Vestibular Schwannoma: Systematic Review». In: *Cancers* 18.2 (2026). DOI: `10.3390/cancers18020289` (cit. on pp. 2, 3).

[7] Bailey H. Duhon et al. «Tumor biomechanical stiffness by magnetic resonance elastography predicts surgical outcomes and identifies biomarkers in vestibular schwannoma and meningioma». In: *Scientific Reports* 14 (2024), p. 14561. DOI: `10.1038/s41598-024-64597-1` (cit. on pp. 3, 5).

[8] Matthew D. Shoulders and Ronald T. Raines. «Collagen structure and stability». In: *Annual Review of Biochemistry* 78 (2009), pp. 929–958. DOI: `10.1146/annurev.biochem.77.032207.120833` (cit. on pp. 3, 40, 60).

[9] Naohiro Soyama, Jun-ichi Kuratsu, and Yukitaka Ushio. «Correlation Between Magnetic Resonance Images and Histology in Meningiomas: T2-weighted Images Indicate Collagen Contents in Tissues». In: *Neurologia Medico-Chirurgica* 35 (1995), pp. 438–441. DOI: `10.2176/NMC.35.438` (cit. on p. 3).

[10] Alexey Surov, Hans Jonas Meyer, and Andreas Wienke. «Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis». In: *Oncotarget* 8.35 (2017), pp. 59492–59499. DOI: `10.18632/oncotarget.17752` (cit. on p. 4).

[11] Philippe Lambin et al. «Radiomics: Extracting more information from medical images using advanced feature analysis». In: *European Journal of Cancer* 48.4 (2012), pp. 441–446. DOI: `10.1016/j.ejca.2011.11.036` (cit. on pp. 4, 14, 60).

[12] Hugo J. W. L. Aerts et al. «Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach». In: *Nature Communications* 5 (2014), p. 4006. DOI: `10.1038/ncomms5006` (cit. on p. 4).

[13] Nicholas A. George-Jones, Rati Chkheidze, Samantha Moore, Jing Wang, and Jacob B. Hunter. «MRI Texture Features are Associated with Vestibular Schwannoma Histology». In: *The Laryngoscope* (2020). DOI: `10.1002/lary.29309` (cit. on pp. 4, 5).

[14] Raffaele De Marco, Giovanni Morana, Silvia Sgambetterra, Federica Penner, Antonio Melcarne, Diego Garbossa, Michele Lanotte, Roberto Albera, and Francesco Zenga. «Predicting the Consistency of Vestibular Schwannoma and Its Implication in the Retrosigmoid Approach: A Single-Center Analysis». In: *Current Oncology* 32 (2025), p. 647. DOI: `10.3390/curroncol32110647` (cit. on pp. 7, 61).

[15] Alex Zwanenburg, Martin Vallières, Mahmoud A. Abdalah, et al. «The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping». In: *Radiology* 295.2 (2020), pp. 328–338. DOI: `10.1148/radiol.2020191145` (cit. on pp. 8, 12, 15).

[16] Benjamin M. Ellingson, Martin Bendszus, Jerrold Boxerman, Daniel Barboriak, Bradley J. Erickson, Marion Smits, Sarah J. Nelson, Elizabeth Gerstner, et al. «Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials». In: *Neuro-Oncology* 17.9 (2015), pp. 1188–1198. DOI: `10.1093/neuonc/nov095` (cit. on pp. 8, 60).

[17] Mark S. Shiroishi, Gloria Castellazzi, Jerrold L. Boxerman, Francesco D'Amore, Marco Essig, Thanh B. Nguyen, et al. «Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging». In: *Journal of Magnetic Resonance Imaging* 41.2 (2015), pp. 296–313. DOI: `10.1002/jmri.24648` (cit. on p. 9).

[18] Sahan M. Wijithananda, Mohan L. Jayatilake, Badra Hewavithana, Teresa Gonçalves, Luis M. Rato, Bimali S. Weerakoon, Tharindu D. Kalupahana, Anil D. Silva, and Karuna D. Dissanayake. «Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques». In: *BioMedical Engineering OnLine* 21.1 (2022), p. 52. DOI: `10.1186/s12938-022-01022-6` (cit. on pp. 9, 60).

[19] Andriy Fedorov et al. «3D Slicer as an image computing platform for the Quantitative Imaging Network». In: *Magnetic Resonance Imaging* 30.9 (2012), pp. 1323–1341. DOI: `10.1016/j.mri.2012.05.001` (cit. on p. 10).

[20] Bradley C. Lowekamp, David T. Chen, Luis Ibáñez, and Daniel Blezek. «The Design of SimpleITK». In: *Frontiers in Neuroinformatics* 7 (2013), p. 45. DOI: `10.3389/fninf.2013.00045` (cit. on p. 12).

[21] Joost J. M. van Griethuysen et al. «Computational Radiomics System to Decode the Radiographic Phenotype». In: *Cancer Research* 77.21 (2017), e104–e107. DOI: `10.1158/0008-5472.CAN-17-0339` (cit. on p. 17).

[22] Peter Bankhead et al. «QuPath: Open source software for digital pathology image analysis». In: *Scientific Reports* 7.1 (2017), p. 16878. DOI: `10.1038/s41598-017-17204-5` (cit. on p. 19).

[23] Abhishek Vahadane et al. «Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images». In: *IEEE Transactions on Medical Imaging* 35.8 (2016), pp. 1962–1971. DOI: `10.1109/TMI.2016.2529665` (cit. on p. 20).

[24] Massimo Salvi, Francesco Branciforti, Filippo Molinari, and Kristen M. Meiburger. «Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm». In: *Expert Systems with Applications* 245 (2024), p. 123105. DOI: `10.1016/j.eswa.2023.123105` (cit. on pp. 21–25).

[25] Augustus Odena, Vincent Dumoulin, and Chris Olah. «Deconvolution and Checkerboard Artifacts». In: *Distill* 1 (2016), e3. DOI: `10.23915/distill.00003` (cit. on p. 27).

[26] A. C. Ruifrok and D. A. Johnston. «Quantification of histochemical staining by color deconvolution». In: *Analytical and Quantitative Cytology and Histology* 23.4 (2001), pp. 291–299 (cit. on p. 29).

[27] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller. «Hematoxylin and eosin staining of tissue and cell sections». In: *Journal of Visualized Experiments* 20 (2008), e992. DOI: `10.1101/pdb.prot4986` (cit. on pp. 29, 32).

[28] Joachim Weickert, Wissenschaftlicher Werdegang, Steven Zucker, Allan Dobbins, Lee Iverson, Benjamin Kimia, and Allen Tannenbaum. «Anisotropic Diffusion In Image Processing». In: (Jan. 1996) (cit. on p. 33).

[29] Diantha Van De Vlekkert, Eda Machado, and Alessandra d'Azzo. «Analysis of Generalized Fibrosis in Mouse Tissue Sections with Masson's Trichrome Staining». In: *Bio-Protocol* 10.10 (2020), e3629. DOI: `10.21769/BioProtoc.3629` (cit. on p. 36).

[30] J. M. Bland and D. G. Altman. «Statistical methods for assessing agreement between two methods of clinical measurement». In: *The Lancet* 327.8476 (1986), pp. 307–310. DOI: `10.1016/S0140-6736(86)90837-8` (cit. on p. 38).

[31] Yoav Benjamini and Yosef Hochberg. «Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (1995), pp. 289–300. DOI: `10.1111/j.2517-6161.1995.tb02031.x` (cit. on p. 39).

[32] W. J. Youden. «Index for rating diagnostic tests». In: *Cancer* 3.1 (1950), pp. 32–35. DOI: `10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3` (cit. on p. 40).

[33] John W. House and Derald E. Brackmann. «Facial nerve grading system». In: *Otolaryngology–Head and Neck Surgery* 93.2 (1985), pp. 146–147. DOI: `10.1177/019459988509300202` (cit. on p. 41).

[34] Corinna Cortes and Vladimir Vapnik. «Support-vector networks». In: *Machine Learning* 20.3 (1995), pp. 273–297. DOI: `10.1007/BF00994018` (cit. on p. 41).

[35] Sudhir Varma and Richard Simon. «Bias in error estimation when using cross-validation for model selection». In: *BMC Bioinformatics* 7 (2006), p. 91. DOI: `10.1186/1471-2105-7-91` (cit. on p. 42).