

UNIVERSITY OF ŽILINA
Faculty of Management Science
and Informatics

Histological Image Data Preprocessing

From raw data to model input

Mgr. Miroslav Mažgut
20th October 2026

Introduction to Digital Histopathology

Data Preprocessing Pipeline

Introduction to Digital Histopathology

Histopathology and Digital Histopathology

- Histopathology is the microscopic examination of biological tissue to study and diagnose the manifestations of disease.
- Digital histopathology is the practice of scanning physical tissue slides into high-resolution digital images for electronic viewing, analysis, and archiving.

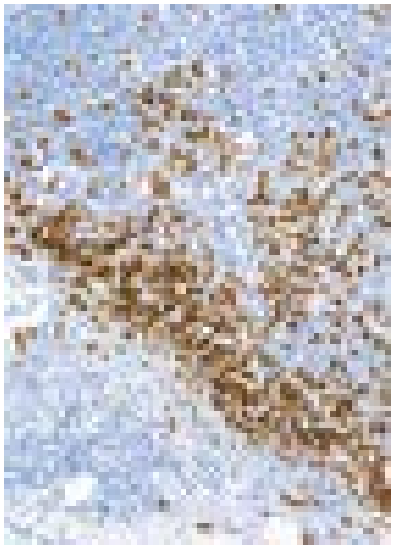


Figure: Taken from [5]

Tissue examples



(a) Hematoxylin and Eosin (H&E)



(b) Immunohistochemical staining (IHC)



(c) Tissue microarrays (TMAs)

Figure: Examples of histopathological image data. Taken from [4, 1]

Current Research Topic

- Dataset containing pairs of neighbouring tissue sections.
  - One stained with H&E.
  - The other stained with IHC targeting the Ki-67 protein.
- Ki-67 is used to calculate the proliferation index.
- Goal: Develop a model to predict the proliferation index directly from H&E images.



Figure: Taken from [3]

Raw Data

- The dataset contains 168 WSIs with a total size of 563 GB ($\pm$ 3.5 GB per WSI).
- Separate H&E-stained tissues from one another.
- Map the Ki-67 index onto the H&E-stained tissue (image registration).
- Problem: How can the dataset be processed as efficiently and automatically as possible?

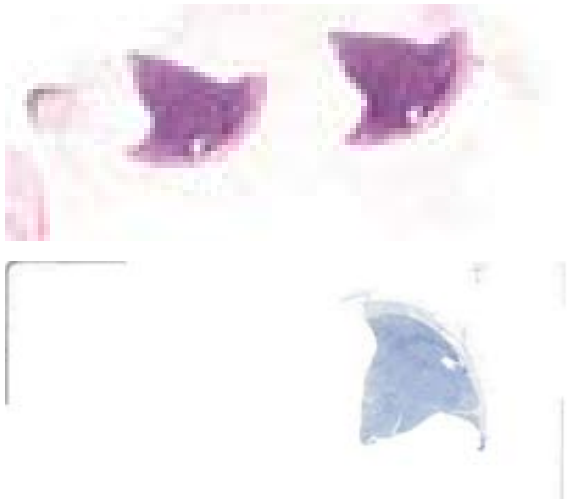
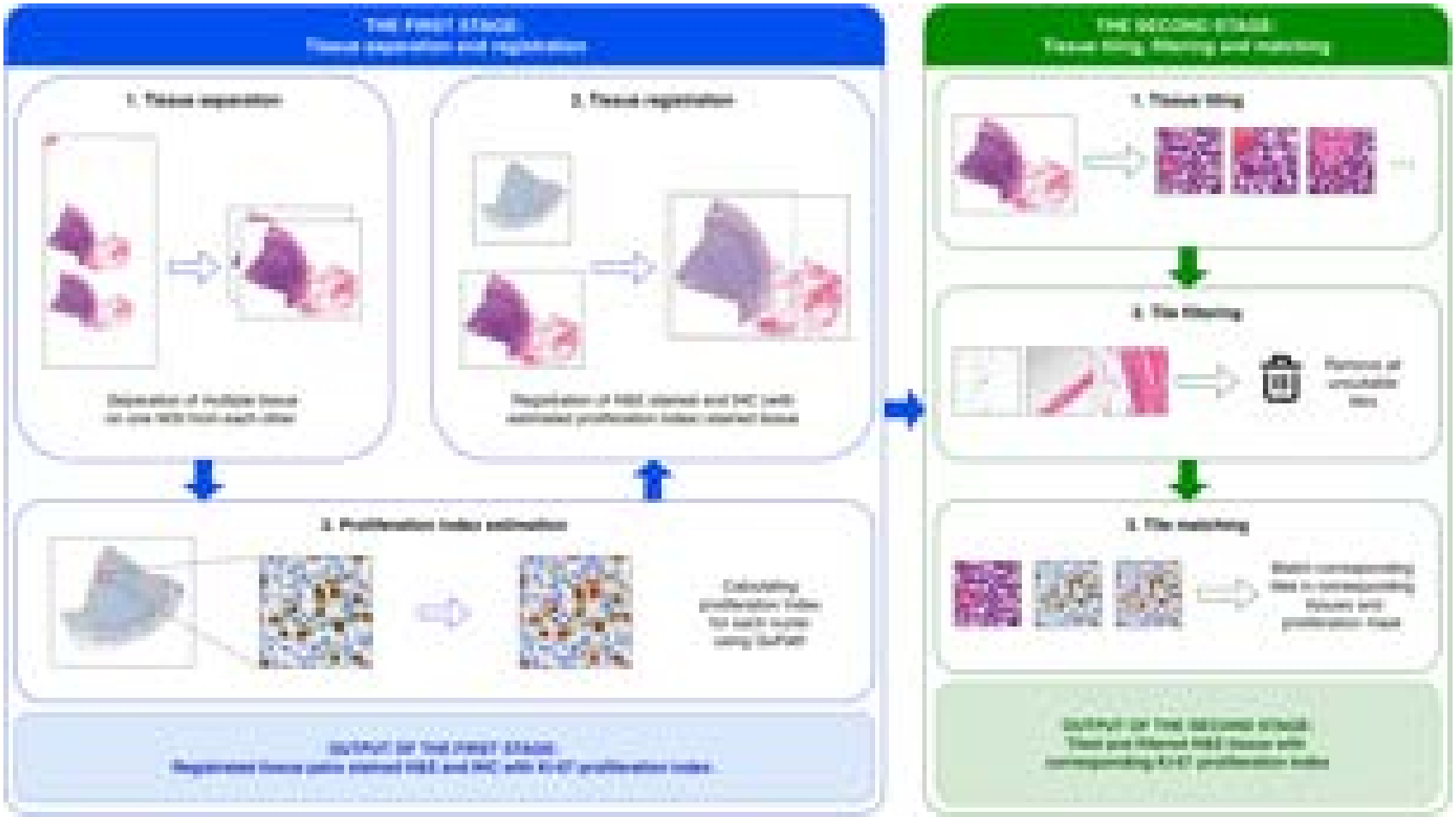


Figure: Example of a raw WSI pair stained with H&E and IHC

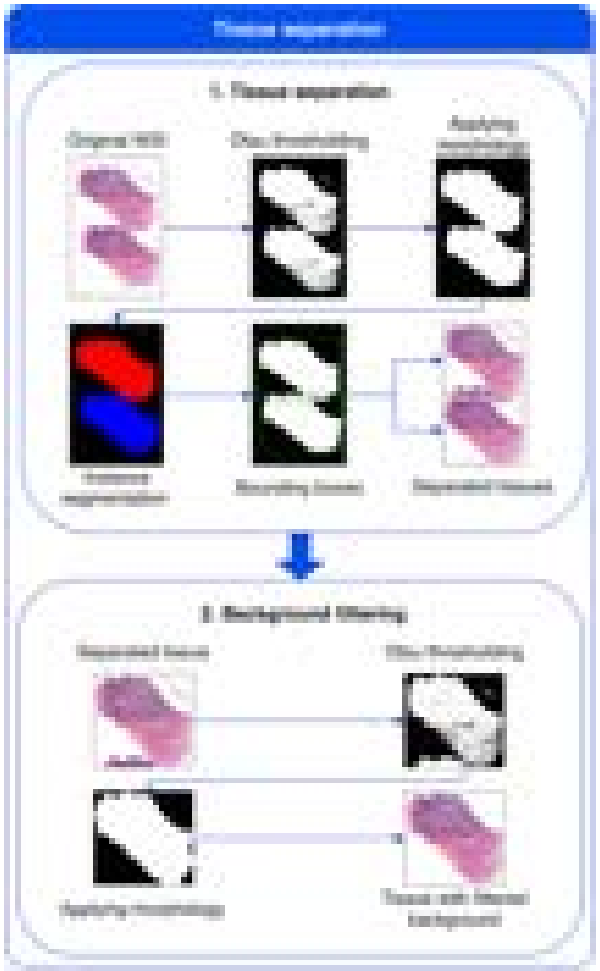
Data Preprocessing Pipeline

The proposed pipeline



Tissue Separation

- Step 1: Separate the tissues in one WSI into individual images.
- Step 2: Filter unwanted artefacts from the areas surrounding the tissues.
- Both stages use deterministic, basic image-processing techniques.

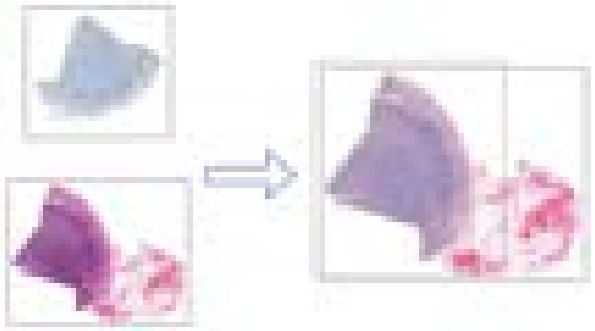


Proliferation index estimation and image registration

1. A QuPath script estimates the proliferation index on IHC-stained tissues and creates a proliferation-index mask.[a]
2. The IHC-stained tissues are then registered to the corresponding H&E-stained tissues using the Python library VALIS.
3. The same image transformation is applied to the proliferation-index mask.



(a) Proliferation index estimation



(b) Image registration

[a] The method was designed by Patrik Istvanko et al. and is to be published soon.

Tissue Tiling

1. Tissue tiling: Divide the input tissue images into small images called tiles (for example, 256×256 pixels).
2. Tissue filtering: Filter out unwanted regions (background, blur, artefacts, etc.) [2].
3. Tissue matching: Match corresponding tiles in the tissue images and the proliferation mask. Remove any tiles that do not match.

Future Work

- Train models on the preprocessed data.
- Try different preprocessing parameters (for example, tile size).
- Analyze identified deficiencies and address them in the pipeline.

Thank you for your attention!

Bibliography

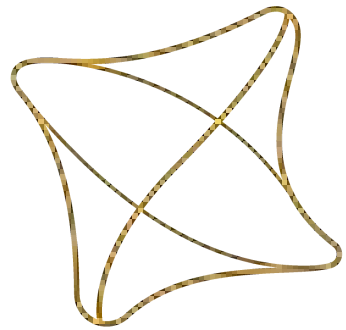
[1] Miroslav Mažgut. "Detection of Metastases in Immunohistochemically Stained Lymphatic Tissue Using Deep Learning Methods". Supervisor: Tomáš Brázdil. Bachelor's thesis. Brno, Czech Republic: Masaryk University, Faculty of Informatics, 2024. URL: https://theses.cz/id/ywrpus/ (visited on 14/06/2026).

[2] Daniela Pavlíková, Lucia Piatriková and Ivan Cimrák. "Quality Assessment and Filtering of HE-Stained Images for Machine Learning Applications". In: *Artificial Intelligence in Medicine: 23rd International Conference, AIME 2025, Pavia, Italy, June 23–26, 2025, Proceedings, Part II*. Vol. 15735. Lecture Notes in Computer Science. Cham: Springer, 2025, pp. 311–316. ISBN: 978-3-031-95840-3. DOI: 10.1007/978-3-031-95841-0_58.

[3] Dominika Petríková et al. *Histopathology pairs of hematoxylin-eosin and Ki67 stainings of testicular seminoma*. Zenodo, 2024. DOI: 10.5281/zenodo.11218961. URL: https://doi.org/10.5281/zenodo.11218961.

[4] WebPathology. *Adenocarcinoma of Prostate*. 2026. URL: https://www.webpathology.com/images/genitourinary/prostate/adenocarcinoma-of-prostate/35025 (visited on 14/06/2026).

[5] David Y. Zhang et al. "Introduction of an Integrated Pathology Image Management, Artificial Intelligence, and Reporting System". In: *JoVE* 221 (2025), e68270. DOI: 10.3791/68270.



UNIVERSITY OF ŽILINA

Faculty of Management Science and Informatics

FRI UNIZA