Multi-model deep learning pipeline and graph neural networks for skin cancer cells and tissues image analysis

Inaugural-Dissertation zur Erlangung des Doktorgrades der
Mathematisch–Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von **Lucas Sancéré** aus Bayonne, Nouvelle-Aquitaine, France

2026

Abstract

Deep learning became an increasing popular field of research in the last 25 years. It is only in 2012 with DeepBind predicting binding preferences of DNA and RNA binding proteins and in 2015 with U-net architecture segmenting neural structures from electron microscopy that we saw major applications of this emergent field to biology. Since then, academia faced an exponential increase in the application of deep learning methods to biology and medicine. MICCAI conference, The Medical Image Computing and Computer Assisted Intervention Society, one of the largest conference for machine learning applied to medicine, received 756 submitted papers in 2016 and 3,667 in 2025.

Our research is situated within this broad historical movement. This work focus on computer vision model for the analysis of medical images. More specifically on the analysis of Whole Slide Images (WSIs) of cutaneous Squamous Cell Carcinoma (cSCC). Whole Slide Images are high dimension megapixel images of tissue for which nucleus of cells are visible. They are commonly acquire in hospital's pathology departments for diagnostic or to follow treatment progress. Cutaneous Squamous Cell Carcinoma is a common type of skin cancer, highly prevalent worldwide. Our work methodology and software's are applied to this cancer as an example use-case but are design for further use on other cancer types.

We first developed Histo-Miner pipeline to segment and classify all cell nuclei from cSCC WSIs. Following this step, the pipeline is used to calculate relevant tissue features and then summarize the WSI as few key numbers for downstream analysis. We first applied our Histo-Miner software to predict therapy response of patients undergoing anti-PD1 immunotherapy through analyses of their WSI recordings. In addition to provide solid classification performance, Histo-Miner provided a list of key features responsible for therapy response and insights into the underlying biology.

We then applied Histo-Miner on WSIs cSCC from 3 clinical centers to analyze the most predictive tiles in classification of progression status (disease progression or no progression). A transformer-based multiple instance learning model was first used to classify the WSIs. Then Integrated Gradient Method was used to identify the most relevant patches and Histo-Miner was applied on these for segmentation and classification of the cell nuclei. After classification, the pipeline was employed to calculate tissue and cell based features. Several cell based features showed significantly different distribution between the two groups, for instance non-progressors maintained homogeneous tumor patterns, while progressors were characterized by a higher degree of integration between tumor cells and neighboring cell populations. Finally, using classical machine learning model XGBoost on the features calculated by Histo-Miner on the most representative patches yielded to high classification accuracy.

Lastly, we used cell graph representations and graph Transformers neural networks to improve on cell nuclei classification for the case of epithelial cells. We compared image-based and graph-based approaches on WSI cell classification, on 2 distinct scenarios. The first scenario is the classification of all epithelial cells from a singe WSI. The second scenario is the classification of epithelial cells from WSI patches of different patients. We revealed that graph Transformers with linear complexity are better performing than state of the art image-based methods on both cases. Building cell graph representations from WSI and performing classification from these graphs instead of the original image lead to improved classification performance and significantly faster training and evaluation.

Acknowledgments

I have always loved to read Acknowledgments section, wether in PhD Thesis Manuscript or in published research papers. Golden gems are always lying around, the rare occasion to learn a bit more about the researchers and people making Academia a - mostly - welcoming place. Here and there we can read a marriage proposal or a praise to the Halo video game series. Prior to my PhD life I met a man by the name of Floris, in Paris, passionate researcher collecting the best "Acknowledgments" he could find. I will unfortunately not be able to propose such delightful reading to my audience (if even this audience exists). It will be personal and LONG, so quite boring and cringe if you don't know me well. Nevertheless, I encourage any eventual reader of mine to read more"Acknowledgments", collect some incredible stories. I promise it could open new perspectives.

I would like to start by thanking the research that motivates me to continue in Academia. My 3 years Master was multi-disciplinary and during that time I was wondering which Scientific field I should focus on. One paper guided me towards Computational Biology more than any other. It was "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy" by Ounkomol et al. from the Gregory R. Johnson Lab. At that time I was spending hours in microscopy rooms in Grenoble and really loved this new application of machine learning to fluorescence imaging as well as many aspects of this paper. One year later, in the center of Paris, I assisted to the PhD defense of Peter Naylor from MINES Paris Tech. I really liked his presentation and then printed, and carefully read his PhD Thesis "From cellular phenotypes to the analysis of whole slide images; Application to treatment response in triple-negative breast cancer". These 2 works were very important for me, and as I am writing these lines, I have them next to me in the drawer. Thank you to anyone that contributed to these 2 works.

A word for Bellaïche Lab in Curie Institute Paris that welcomed me after my studies and motivated me to do a PhD. The team was multidisciplinary and everyone I met was dedicated and excellent in their work. I will not be able to mention everyone but thank you Yohannes Bellaïche, Floris Bosveld, Florencia Di Pietro, Boris Guirao, Jesus Lopaz-Gay, Victoire Cachoux, Aude Maugarny, Lale Alpar and Stéphane Pelletier. Thanks to the bakeries in the 5th Arrondissement that provided for CAKE TIMES. Also a word on my Master Thesis director Eva Faurobert and on Olivier Destaing. It was a nice time in La Tronche just before going to Paris (if I omit some bike adventures in Estonia with Jean Le Pennec and in Cyprus with Yvan Pratviel -a.k.a Pavlov-, but this is for another time). Your supervision was very kindly, you respected my lack of knowledge in Biology and my unconventional way of performing RNA transfections in the wet lab. I spent a nice time with you, it was also a great scientific - and human - community all around. A nice place surrounded by mountains and by some yellow and green drinks of the finest taste. Eva it was you telling me about this position in Paris in Bellaïche Lab. If young Master Thesis student me went to Paris and then to Cologne it is mainly thanks to you.

Thank you to all my colleagues and to my supervisor Kasia. It was a good working place and a good scientific community here in Cologne. It is too early for me to go very poetic and nostalgic about this time, as it is not even over, so I will keep it simple. Nevertheless I already know that some Lab Retreats and Conferences will always have a place in my mind. CVPR in Vancouver - very cool scientific debates, beautiful black bear; ML in PL in Warsaw, very cool talks, Pierogi of heaven; Lab retreat in Feldberd, interesting presentations in the middle of the evening and what an honor to ski with German. Thank you all for this scientific and deeply human journey of mine.

I would like to thank all my friends that were around during my PhD time, especially Edouard, Charlotte, Yvan, Alfred, Ele, Claire, Casper, Marion, Max, Alberto and Scooty.

Finally, my close family.

To Jana Tournesol, you were here all this time and more supportive than I could have imagined. I Know you overcame your own problems to stay supportive all along, so thank you for that. It is one of the 16 Reasons I love you. And we are not even in 1959. To my Papa, my Maman and my Sœur Laura, je vous aime, thank you for being around. You are the best family. It had to be written somewhere. And to my nephew about to be born, I already love you.

Use of generative AI

While specific faculty-level formatting for AI disclosure is still being refined, the author provides the following statement to ensure full transparency regarding the use of generative AI.

No generative AI was used to produce scientific methods or scientific results that are described in this work. No generative AI was used to generate figures of this manuscript nor in the research papers included within this manuscript. Generative AI was used for: synonyms finding, debug and structure LaTeX code, rephrasing, correct grammar mistakes, debug Python code, understand external Python packages, help writing Python function (sanity checked), browsing. All work that was rephrased, using generative AI or not, has been cited and is available in the Bibliography.

This work is about applied AI research (applied computer vision deep learning to medicine and biology), so non-generative AI was extensively used and created as part of the research work.

Publication Preface

The contributions presented in this thesis are based on previous publications and manuscripts.

Published Manuscripts:

[1] Lucas Sancéré, Carina Lorenz, Doris Helbig, Oana-Diana Persa, Sonja Dengler, Alexander Kreuter, Martim Laimer, Roland Lang, Anne Fröhlich, Jennifer Landsberg, Johannes Brägelmann, Katarzyna Bozek. **Histo-Miner: Deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma.** *PLoS Comput. Biol.* **22**(1), e1013907 (2026) https://doi.org/10.1371/journal.pcbi.1013907.

[2] Juan I. Pisula, Doris Helbig, Lucas Sancéré, Oana-Diana Persa, Corinna Bürger, Anne Fröhlich, Carina Lorenz, Sandra Bingmann, Dennis Niebel, Konstantin Drexler, Jennifer Landsberg, Roman Thomas, Katarzyna Bozek, Johannes Brägelmann. **Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma.** *NPJ Precis. Oncol.* **9**(1), 205 (2025) https://doi.org/10.1038/s41698-025-00997-4.

Pre-prints or Manuscripts under review:

[3] Lucas Sancéré, Noémie Moreau, Katarzyna Bozek. **Context-aware Skin Cancer Epithelial Cell Classification with Scalable Graph Transformers.** *arXiv preprint* (2026) https://doi.org/10.48550/arXiv.2602.15783.

Contribution Statement

Lucas Sancéré is the main author of [1] and [3] included in this thesis. As the main author, he took the primary responsibility for design, implementation, data collection and analysis. Lucas Sancéré is co-author of [2]. His and the co-authors' contributions to the included publications/manuscripts are described in the following using the Contributor Roles Taxonomy (CRediT)[1]:

Contribution for [1]:
Conceptualization: Lucas Sancéré, Carina Lorenz, Johannes Brägelmann, Katarzyna Bozek.
Data Curation: Lucas Sancéré, Carina Lorenz, Doris Helbig, Johannes Brägelmann.
Formal Analysis: Lucas Sancéré.
Funding Acquisition: Johannes Brägelmann, Katarzyna Bozek.
Investigation: Lucas Sancéré, Carina Lorenz.
Methodology: Lucas Sancéré.
Project Administration: Lucas Sancéré, Doris Helbig, Johannes Brägelmann, Katarzyna Bozek.
Resources: Doris Helbig, Oana-Diana Persa, Sonja Dengler, Alexander Kreuter, Martim Laimer, Roland Lang, Anne Fröhlich, Jennifer Landsberg
Software: Lucas Sancéré.
Supervision: Johannes Brägelmann, Katarzyna Bozek.
Validation: Lucas Sancéré.
Visualization: Lucas Sancéré, Carina Lorenz.
Writing – Original Draft Preparation: Lucas Sancéré, Johannes Brägelmann, Katarzyna Bozek.
Writing – Review & Editing: Lucas Sancéré, Carina Lorenz, Johannes Brägelmann, Katarzyna Bozek.

Contribution for [2][2]:
Conceptualization: Katarzyna Bozek, Johannes Brägelmann.
Data Curation: Doris Helbig, Oana-Diana Persa, Corinna Bürger, Anne Fröhlich, Sandra Bingmann, Dennis Niebel, Konstantin Drexler, Jennifer Landsberg, Roman Thomas.
Formal Analysis: Juan I. Pisula, Doris Helbig, Lucas Sancéré, Johannes Brägelmann.
Funding Acquisition: Katarzyna Bozek, Johannes Brägelmann.
Investigation: Juan I. Pisula, Lucas Sancéré.
Methodology: Juan I. Pisula, Lucas Sancéré.
Project Administration: Juan I. Pisula, Katarzyna Bozek, Johannes Brägelmann.

[1] https://credit.niso.org/
[2] The published Contributions is not written using CRediT style and is the official contribution record

Resources: Doris Helbig, Oana-Diana Persa, Corinna Bürger, Anne Fröhlich, Sandra Bingmann, Dennis Niebel, Konstantin Drexler, Jennifer Landsberg, Roman Thomas.
Software: Juan I. Pisula, Lucas Sancéré, Johannes Brägelmann.
Supervision: Katarzyna Bozek, Johannes Brägelmann.
Validation: Juan I. Pisula.
Visualization: Juan I. Pisula, Lucas Sancéré.
Writing – Original Draft Preparation: Juan I. Pisula, Doris Helbig, Katarzyna Bozek, Johannes Brägelmann.
Writing – Review & Editing: Juan I. Pisula, Doris Helbig, Lucas Sancéré, Oana-Diana Persa, Corinna Bürger, Anne Fröhlich, Carina Lorenz, Sandra Bingmann, Dennis Niebel, Konstantin Drexler, Jennifer Landsberg, Roman Thomas, Katarzyna Bozek, Johannes Brägelmann.

Contribution for [3]:
Conceptualization: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Data Curation: Lucas Sancéré.
Formal Analysis: Lucas Sancéré.
Funding Acquisition: Katarzyna Bozek.
Investigation: Lucas Sancéré.
Methodology: Lucas Sancéré.
Project Administration: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Resources: Lucas Sancéré.
Software: Lucas Sancéré.
Supervision: Noémie Moreau, Katarzyna Bozek.
Validation: Lucas Sancéré.
Visualization: Lucas Sancéré.
Writing – Original Draft Preparation: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Writing – Review & Editing: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.

List of Abbreviations

cSCC cutaneous Squamous Cell Carcinoma

DNA Deoxyribonucleic Acid

EGFR Epidermal Growth Factor Receptor

FL Federated Learning

H&E Hematoxylin and eosin

MLP Multi-Layer Perceptron

MSA Multi-Headed Self-Attention

NAS Neural Architecture Search

WSI Whole Slide Image

Contents

Chapter 1

Introduction

1.1 Cutaneous squamous cell carcinoma skin cancer

Cancer occur by a series of successive mutations in genes that regulated cellular homeostasis. Generally, dis-regulation of those genes result in alteration of cell behavior that are frequently related to uncontrollable proliferation, apoptosis, unlimited replicative potential and invasion [4]. In this work, we applied deep learning models to cutaneous Squamous Cell Carcinoma (cSCC) images. cSCC is the second most common non-melanoma skin cancer. In 2019 it accounted for around 2,4 million incidence cases and 56,000 death cases worldwide [5].

There are 3 main types of skin cancer, cutaneous Basal Cell Carcinoma, Melanoma and cSCC. cSCC is the second most aggressive (after Melanoma) [6]. The cSCC tumors, aggregate of tumor cells, develop within squamous cells located on the epidermis, the upper layer of the skin. These tumors can also invade the dermis, the inner layer of the skin and even blood vessels (**Fig.1**). Squamous cells are a subtype of epithelial cells, organized in thick layers. The mature squamous cells toward the surface are large and flattened. Their function is to protect the skin from outside environment and against interactions that could puncture the skin [7].

The main reason behind the occurrence of cSCC is immoderate exposure to ultraviolet radiation. Exposure of the skin to UV radiation causes suppression of cell-mediated immune responses, induces DNA damage, and generates reactive oxygen species, which in turn can trigger oxidative stress and cellular damage. Following high doses of UV radiation, the earliest event is the initiation of keratinocyte apoptosis, meaning the death of cells producing keratin, and a reduction in protein synthesis. Then, cell proliferation increases, that may result in epidermal hyperplasia and expansion of cells carrying UV-induced mutations. Other factors such as human papilloma-virus, chemical carcinogens, genodermatoses, inflammatory conditions, and medicament (tumor necrosis factor—α inhibitors) also hold responsible for SCC [6, 8, 9, 10, 11, 12].

While surgical excision remains the primary treatment for low-risk cSCC, the management of high-risk cSCC is still challenging and lacks standardization. With growing molecular and genetic insights, the past decade has brought several advances in the therapeutic landscape of cSCC. Among these, epidermal growth factor receptor (EGFR) inhibitors and anti–PD-1 agents (PD-1 protein on T-cells helps the immune system recognize and attack cancer cells) have emerged as treatment options for cSCC that cannot be surgically removed. PD-1 inhibitors are demonstrating significantly greater effectiveness [13].

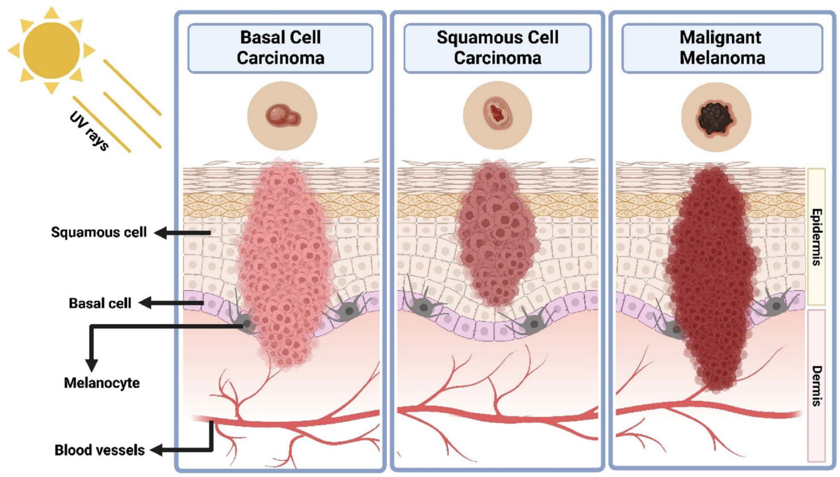


Figure 1: **Diagrammatic representation of basal cell carcinoma, squamous cell carcinoma, and melanoma** @ Figure from [6], Zeng, L. et al. 2023

An emerging line of treatment for cSCC is immunotherapy with anti-PD1 antibodies. Despite increasing literature, some aspects of this treatment remains elusive. For instance, currently there are no validated predictive biomarkers capable of identifying patients at high risk of immunotherapy failure. In our work, we build a deep-learning based pipeline that use case is to identify these patients and provide a biological explanation for these failures (see **Chapter** 2).

1.2 Histology and H&E staining

Advances in imaging techniques have democratized the use of microscopy for disease diagnosis in both medical research and clinical practice. Microscopy techniques cover all ranges of prices and complexity. Nevertheless, time-consuming analyses on expansive material are mostly developed for academic research but are impractical for hospital practices.

In contrast, Hematoxylin and Eosin (H&E) stained slides are routinely used for diagnostics in hospitals. The H&E staining allows to distinguish between cytoplasm and extracellular matrix, stained by the eosin in pink, and the cell nucleus stained in purple by the hematoxylin [14]. The scanning of the slide is performed by an image scanner including a simple light microscope. In the next paragraphs, we will describe the histological process to record such tissue images in the context of skin cancer. This process include several steps that will be detailed: **sample collection, fixation, dehydration, clearing, embedding, sectioning, staining and scanning** [15].

To **collect** the skin sample, a skin biopsy is performed on the patient. The doctor remove a small part of the skin, either by cutting, shaving or punching the sample with a circular tool. Then **fixation** is applied to the tissue collected. This process preserves biological tissues by preventing putrefaction and the destruction of cells and tissues by their own enzymes. It halts ongoing biochemical reactions and can increase the mechanical strength and structural stability of the tissue. Crosslinking fixatives, such as the ones used in for cSCC samples, preserve tissue by forming covalent bonds between proteins. This stabilizes soluble proteins by anchoring them to the cyto-skeleton and provides additional rigidity to the tissue [15, 16].

To later cut sections that will be used for scanning, the fixed tissue needs to be embedded in paraffin wax. To fulfill this objective, we first need to remove all water in the sample. We start by **dehydrating** the tissue, meaning slowly replacing the water in the sample by alcohol. After all the water is replaced by alcohol, the alcohol is replaced by xylene which is a solvent miscible with alcohol. This step is called **clearing**. Finally, as wax is soluble with xylene, we embed the tissue in paraffin wax. This step, **embedding**, is used to harden the tissue before sectioning (as fixation alone is not enough).

Sectioning is performed on a cutting device called microtome. This tools allows to cut thin sections (few nanometers of width at the thinnest) of the embedded tissue. The tissue obtained is mainly colorless. **Staining** needs to be performed in order to visualize cells and tissue parts. As previously mentioned, the H&E staining allows to distinguish between cytoplasm and extracellular matrix, stained by the eosin in pink, and the cell nucleus stained in purple by the hematoxylin. The staining solution being aqueous, the sections are re-hydrated right after sectioning, wax is replaced by water and then the section is stained. Before being imaged with the microscope, we repeat the process of dehydration and clearing.

Finally, the sample is mounted on a digital slide scanner. This scanner integrates camera and a motor to translate the slide while parts of the tissue are **scanned** and saved. After scanning, all the tiles of the tissues are stitched together to create a digital Whole Slide Image (WSI) that can be opened on open-source software. The resulting WSI is a pyramidal image composed of nested images of different resolutions. Practitioners can then study the sample at different tissue scales, from single cells to whole tissue (**Fig.2**).

The skin samples contains lots of different type of cells. Pathologist can recognize different cell types based on their morphologies and sizes (**Fig.3**). Nevertheless, WSIs often contain from 100.000 to 1 million cells. It is then impossible for pathologists to annotate all cells in an image with objective to have a satisfying overview of the tissue organization. Deep learning, in the other hand, can be used to segment and classify all cells and calculate tissue features.

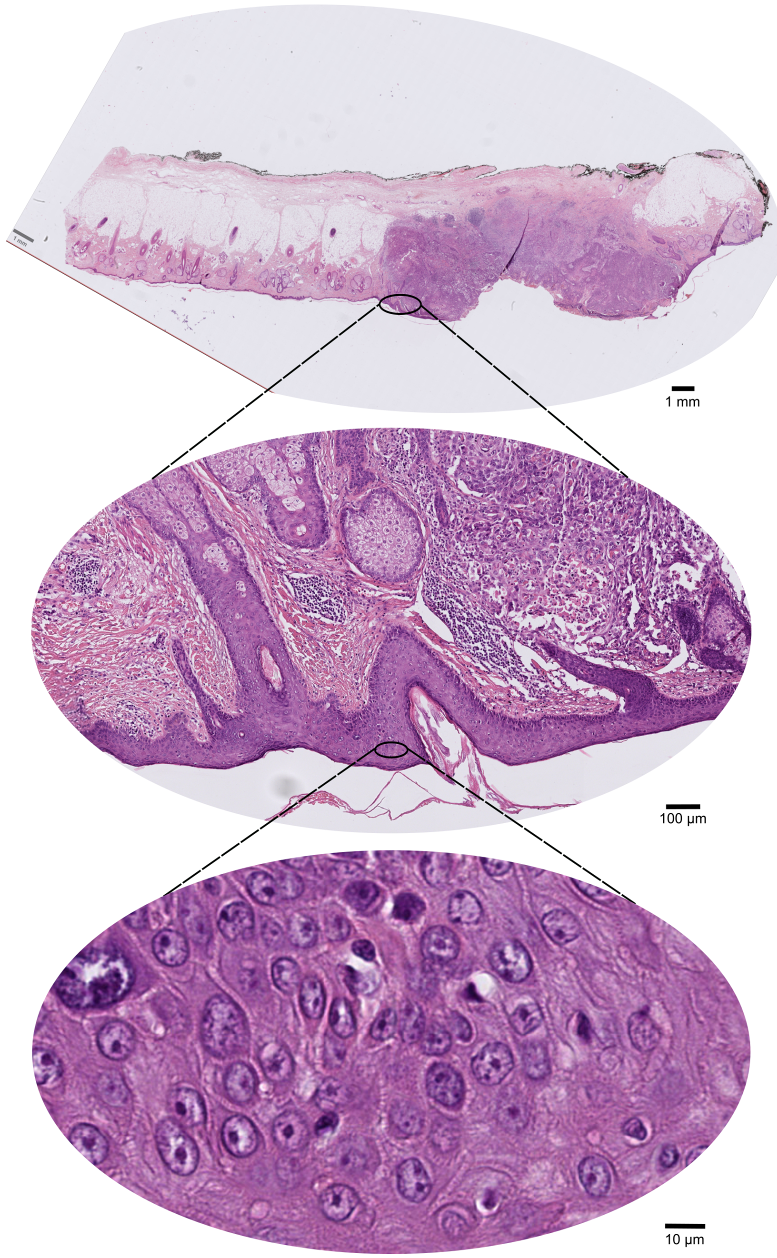


Figure 2: **H&E stained Whole Slide Image visualization** Once the Whole Slide Image of H&E stained tissue section is obtained, the file can be opened using open-source software such as QuPath [17] (see **Fig.**4). Then, user can visualize the tissue at different zoom levels. At highest zoom, rounded purple shapes correspond to cells nucleus. Looking at WSIs, pathologists can recognize different cell types, such as cancer cells and provide a diagnosis or follow treatment progress.

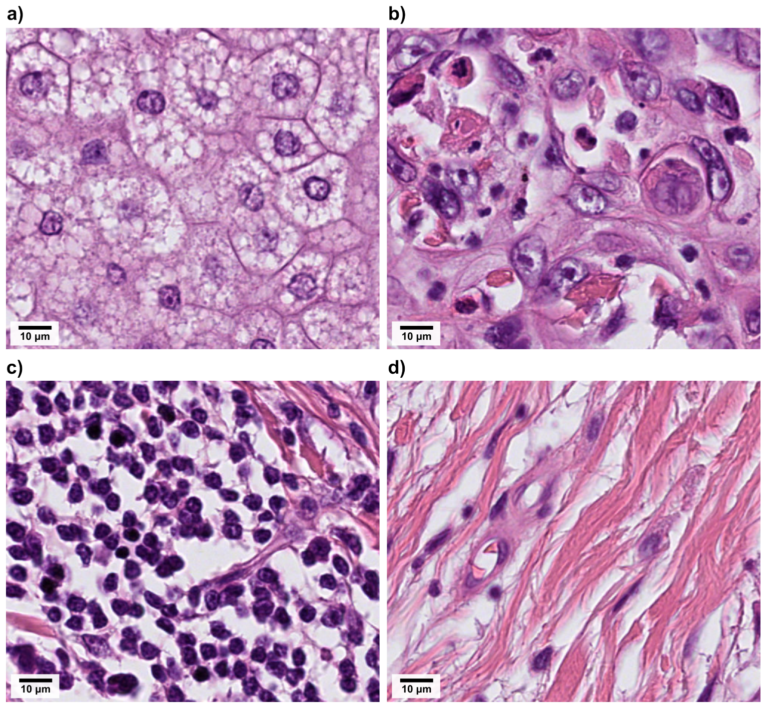


Figure 3: **Morphology diversity on WSIs** All images originate from the same WSI (**Fig.2**). The skin tissue is rich of different cell types that can be recognized by pathologist. **a)** Sweat glands cells, **b)** cancer cells and immune cells, **c)** immune cells: lymphocytes and granulocytes, **d)** stromal cells.

1.3 Deep Learning and histopathology

H&E WSIs are large images (often containing more than 1 billion pixels) rich in encoded information that are used by deep learning models for training and inference (prediction). Since the creation of U-Net model in 2015 [18], new deep learning models used to analyze all type of medical images, including WSIs, emerged and started to be profusely used. This raise of new deep learning models appeared first in academia and then quickly reached industry.

Some initiatives are grouping artificial intelligence models used for healthcare in a unique platform or programming package to ease the models usage. It is the case of project MONAI, for which more than 25 models for very specific healthcare related tasks are openly available for training and inference [19]. Among these different models, some are using whole-slide images data, for instance to detected if a patch contains tumor [19] or to segment nucleus of cells [20]. Some methods focus on extracting features from the WSIs to represent it with vectors of few key numbers. Self-supervised methods pre-trained on different modalities, including images and texts on an extremely large dataset of more than 100 000 H&E WSIs images [21], WSI-pathology reports pairs [22] or histopathology images with matched genetic profiling [23] have shown great generalization capabilities. Using the representations given by these models for downstream tasks such as classification, segmentation or captioning lead to state of the art performances.

Models like Hovernet [24] and CellViT [25] were developed to detect, segment and classify cell nucleus on WSI from different cancer types. Their architecture differ but they have similar performances depending on the training data. One can use open-source software such as QuPath [17] to visualize their inference on WSI from patients of our Cologne cohort (see **Fig.4**). These models are directly tailored to train and infer from WSI medical images. In contrast, some other models are primary designed to work with so called natural images, meaning photograph of everyday life (cars in a street, cat in a house, children playing in a garden).

Models trained and tested on natural images are sometimes used on WSIs for different tasks. Most of the models originally created to perform image classification on ImageNet [26], one of the largest natural images dataset, have been later used for WSI classification and segmentation. This is the case for Inception-v3 [27], EfficientNet [28] and Vision Transformer [29] models. Inception-v3 was used for lung cancer WSI classification [30], a light-weighted version of EfficientNet-b0 was used to classify patches within a WSI workflow [31] and hierarchical Vision Transformer architectures were built for cancer subtyping (image classification) and survival prediction [32]. In this work we used Segmenter model to segment tumor regions (see **Chapter 2**), originally trained and tested on ADE20K [33] and Pascal Context [34] dataset consisting of natural images of all types (from dog contests to bar cocktails close-up).

In our work we used, retrained, and tweaked already existing deep learning models. We will next present the inspirations behind the creation of these models and the improvement to previous work they bring.

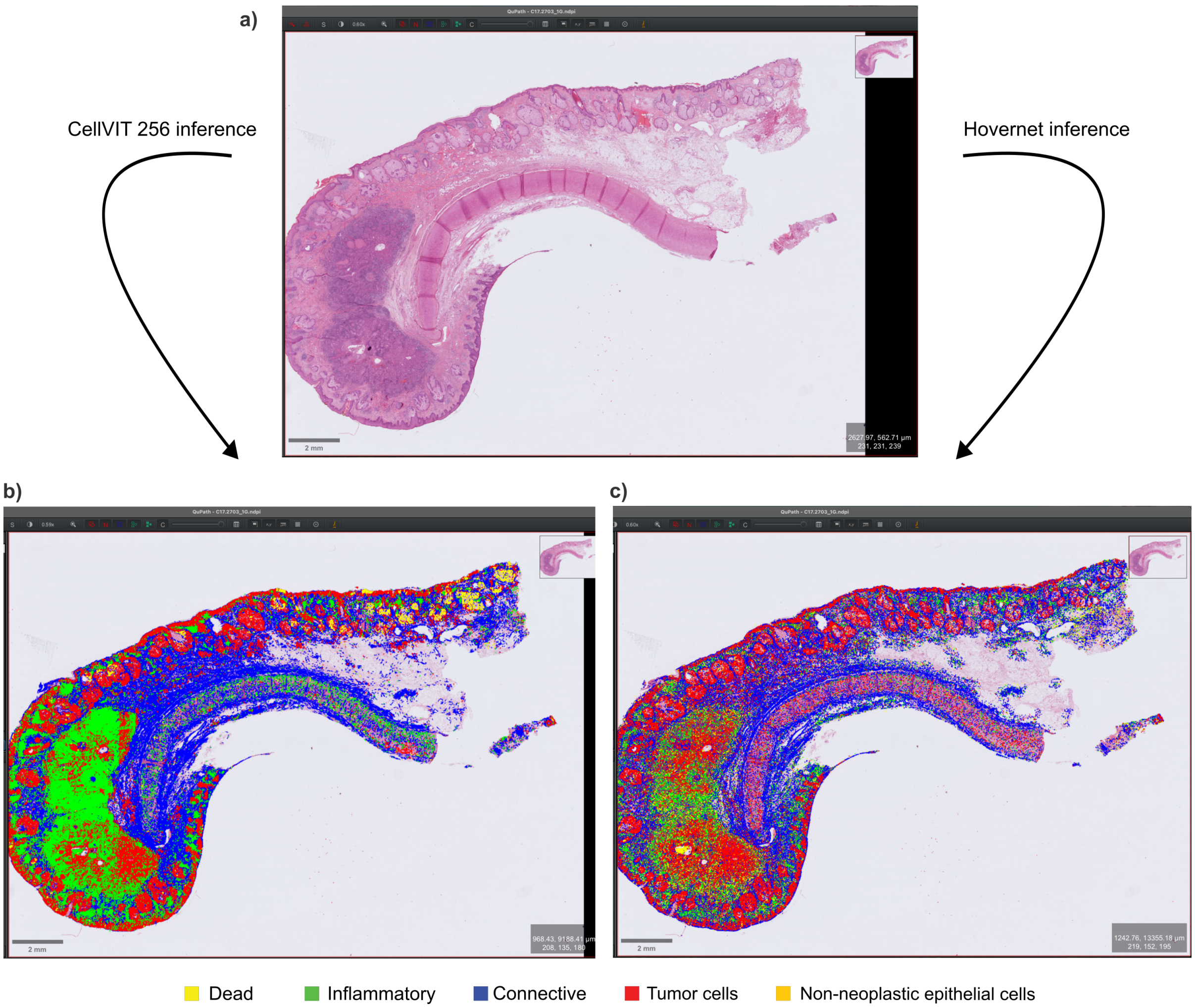


Figure 4: **Hovernet and CellVIT 256 inference visualization on QuPath** Open source software (such as QuPath) are able to load WSIs and can display inference from deep learning models. **a)** cSCC WSI from an anonymized patient of Cologne cohort. **b & c)** Inference on cSCC WSI from pre-trained CellVIT 256 (left) and from pre-trained Hovernet (right). Cell nucleus are segmented and classified following Dead, Inflammatory, Connective, tumor cells and Non-neoplastic epithelial cells classes. At this scale, segmented nucleus appeared as small dots. As we will see in **Chapter 2**, the inference from these state of the art model on cSCC WSI is not accurate, specifically for Non-neoplastic epithelial and inflammatory cells. One of the goal of the presented work is to improve on these prediction.

1.4 Presentation of main deep learning models used in this work

1.4.1 Hovernet architecture and inspirations

Hovernet is a deep learning model used to perform segmentation and classification of segmented cell nuclei on WSI patches (example of patches on **Fig.3**) The model is a convolutional neural network containing encoder and decoder parts [24].

The main inspirations for Hovernet architecture were, for the encoder part, the preactivated residual networks, such as Preact-ResNet50 [35] and for the decoder parts densely connected convolutional networks, such as DenseNet [36]. The idea behind the preactivated residual networks is to use identity skip connections (residual connections) to improve information propagation and avoid gradient vanishing, and to change location of activation function in the residual networks. Indeed, in residual networks, the activations used to affects both the identity mapping and the residual function output. The output of the l-th unit of the network is defined following **Eq. 1**:

$$x_{l+1} = f(x_l + \mathrm{F}(x_l, W_l)) \tag{1}$$

With x_{l+1} the output of the l-th unit of the network, and x_l its input, f the activation function, F the residual function and W_l the set of weights (and biases) associated with the l-th residual unit.

If we move the activation function from being directly after the element-wise addition to be first element of the residual unit it finally leads to **Eq. 2**:

$$x_{l+1} = x_l + \mathrm{F}(f'(x_l), W_l)) \tag{2}$$

With x_{l+1} the output of the l-th unit of the network, and x_l its input, f' a rearrangement of f to be the first element of the Residual Unit. F is the residual function and W_l the set of weights (and biases) associated with the l-th residual unit.

This slight change in architecture yields to an improved regularization and an eased optimization during training. DenseNet architecture, the inspiration of the decoder of Hovernet, has an increased number of connections between layers compared to traditional convolutional networks. Traditional convolutional networks possess one connection between each layers and their subsequent layers, in the case of DenseNet each layer takes all preceding feature-maps as input. This lead to a lower number of parameters to achieve the same performance as using a single convolution with larger kernel size.

In the case of WSI as input of Hovernet, an important pre-processing step performed is to generate overlapping tiles. These tiles are encoded through the encoder, similar to the Preact-Res-Net50 implementation architecture except that the total downsampling factor applied is to 8 instead of 32, to reduce loss of information useful for segmentation. The decoding part is divided into 3 different branches. The nuclear pixel (NP) branch is performing binary segmentation of the nucleus. The HoVer branch predicts the horizontal and vertical distances of nuclear pixels to their center of mass. The nuclear classification (NC) branch predicts the type of nucleus for each pixels. These branch are all mimicking DensNet architecture. Each of these branches optimized their respective set of weights using 2 different loss functions (then a total of 6 loss functions are used, (see **Chapter** 2)). Having 1 encoder and 3 decoder branches instead of 3 different networks allows to reduce training time as only one end-to-end training is needed instead of 3. Also one shared encoder improve generalization of encoding on all tasks.

Finally one last post-processing step is performed to generate an instance segmentation and classification map, where each nucleus is segmented as a specific instance with an associated ID and a specific class. Using both the binary segmentation of nucleus and the gradient of horizontal and vertical distances map, the network is generating an instance segmentation even if the nucleus are adjacent to each other. Using nuclear type predictions, the algorithm assigns each nucleus instance a class corresponding to the most prevalent predicted class within the nuclear type map at the nucleus location.

1.4.2 Segmenter architecture and inspirations

Segmenter model [37] is a state of the art model for semantic segmentation task on datasets such as ADE20K [33] and Pascal Context [34, 38]. Semantic segmentation is the task of classifying all pixels in an image based on the type of object it belongs to. Pixels belonging to different instances of the same type of object will be classified the same, contrary to instance segmentation. Segmentations that mix both instance and semantic segmentations are called panoptic segmentations. The architecture of the Segmenter network is based on Vision Transformers.

Vision Transformers are a class of Deep Learning Transformers architectures applied to images and videos datasets and related tasks such as segmentation, classification and detection [38]. Transformer model [39] was first used in Natural Language Processing tasks such as translation of a text. A transformer block consists of a multilayer perceptron of 2 layers block normalized with LayerNorm, receiving as input vector the output of a multi-headed self-attention block also normalized with LayerNorm. The multi-headed self-attention block, denoted MSA, and the multilayer perceptron block, denoted MLP, taken separately are computed following **Eq. 4** and **Eq. 5**:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{k}}\right)\mathbf{V} \tag{3}$$

$$\text{MSA}(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_\text{h})W^0 \tag{4}$$
$$\text{where head}_\text{i} = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

With $Q \in \mathcal{R}^{d\times k}$, $K \in \mathcal{R}^{d\times k}$, $Q \in \mathcal{R}^{d\times k}$ matrices of queries, keys and values, $d = 512$ chosen dimension of the output and $k = 64$. The Attention function maps a set queries and a set of key-value pairs to an output. W_i^Q, W_i^K, W_i^V, are the matrices of learnable linear projections.

$$\text{MLP}(x) = W_2^T\,\text{ReLU}\left(W_1^T x\right) \tag{5}$$

With $W_1 \in \mathcal{R}^{d\times m}$ the matrix of learnable weights of first layer, $W_2 \in \mathcal{R}^{m\times d}$ matrix of learnable weights of the second layer and ReLU the Rectified Linear Unit activation function. $d = 512$ is the chosen dimension of the output and $m = 2048$.

After the normalization of the Multi-Head Self Attention block the intermediary output obtained is called SubLayer. Then considering T the transformer block parameterized function and x the input token vector to the transformer block, the output of the transformer block follows **Eq. 6** [37, 39, 40]:

$$\text{SubLayer}(\mathbf{x}) = \text{LayerNorm}(\text{MSA}(\mathbf{x}) + \mathbf{x}, \gamma_1, \beta_1) \quad = \gamma_1 \frac{\text{MSA}(\mathbf{x}) + \mathbf{x} - \mu_{\text{MSA}(\mathbf{x})+\mathbf{x}}}{\sigma_{\text{MSA}(\mathbf{x})+\mathbf{x}}} + \beta_1$$

With $\mu_{\text{MSA}(\mathbf{x})}$ and $\sigma_{MSA(\mathbf{x})}$, respectively the mean and standard deviation of the elements of $\text{MSA}(\mathbf{x})$ and $\gamma_1 \in \mathcal{R}^d$, $\beta_1 \in \mathcal{R}^d$ learnable affine transform parameters of the normalization.

$$T(\mathbf{x}) = \text{LayerNorm}(\text{MLP}(\text{SubLayer}(\mathbf{x})) + \text{SubLayer}(\mathbf{x}), \gamma_2, \beta_2) \tag{6}$$

With $\gamma_2 \in \mathcal{R}^d$ and $\beta_2 \in \mathcal{R}^d$ learnable affine transform parameters of the normalization.

In the case of our work, the input data of Segmenter are WSIs. Pre-processing consists of splitting the image into a sequence of non-overlapping patches of 640x640 pixels, flattened into a 1D vector. These vectors are transformed into patch embeddings through a linear projection. Learnable position embeddings are added to each patch embedding to reflect the spatial organization of the patches in the image (similar to the position embeddings for words in the case of Natural Language Processing Transformers). The encoder part of Segmenter model is the Vision Transformer architecture [29], the decoder includes additional learnable class embeddings. These class embeddings are initialized randomly and assigned to a single semantic class. The model predicts one mask per class and determines the most probable class for each pixel values via softmax and argmax transformations.

1.4.3 Achieving linear complexity in Graph Transformers

Original Graph Transformers [41, 42] are limited to small and medium graphs (hundreds of nodes) for node and graph classification. Indeed the original attention mechanism requires $\mathcal{O}(N^2)$ complexity w.r.t to the number of nodes. Training on large graph is then computationally infeasible. Some initiative modified the attention mechanism in graph models to finally reach $\mathcal{O}(N)$ complexity w.r.t to the number of nodes. Resulting Graph Transformers with linear complexity can be trained and evaluated on large graphs, such as ogbn-proteins [43], Amazon-M2 [44] or pokec [45]. Here we succinctly discuss NodeFormer [46], DIFFormer [47] and SGFormer [48] all Graph Transformers with linear complexity that are evaluated in our work of **Chapter 4**.

In NodeFormer, the original softmax attention is approximated using stochastic kernel approximation [3, 49]. Following [49], we can rewrite **Eq. 3** to explicit the softmax function and refer to the $(k+1)^{th}$ layer the node embedding are calculated for. It results in:

$$\mathbf{q}_u^{(k)} = \mathbf{W}_Q \mathbf{z}_u^{(k)}, \quad \mathbf{k}_u^{(k)} = \mathbf{W}_K \mathbf{z}_u^{(k)}, \quad \mathbf{v}_u^{(k)} = \mathbf{W}_V \mathbf{z}_u^{(k)}$$

$$\mathbf{z}_u^{(k+1)} = \frac{1}{\sqrt{k}} \sum_{v=1}^{N} \frac{\exp((\mathbf{q}_u^{(k)})^\top \mathbf{k}_v^{(k)})}{\sum_{w=1}^{N} \exp((\mathbf{q}_u^{(k)})^\top \mathbf{k}_w^{(k)})} \mathbf{v}_v^{(w)} \tag{7}$$

With $\mathbf{z}_u^{(k)}$ node embedding of the k^{th} layer, $\mathbf{q}, \mathbf{k}, \mathbf{v}$ the query, key and value vectors, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$ the learnable weights at the k^{th} layer .

NodeFormer uses Random Features Map for Kernel Approximation as created in Ali Rahimi & Benjamin Recht 2007 [50]. It allows for a linear approximation of the softmax function. The node embedding of the $(k+1)^{th}$ layer is now calculated as:

$$\mathbf{z}_u^{(k+1)} = \frac{1}{\sqrt{k}} \sum_{v=1}^{N} \frac{\phi(\mathbf{q}_u^{(k)})^\top \phi(\mathbf{k}_v^{(k)})}{\sum_{w=1}^{N} \phi(\mathbf{q}_u^{(k)})^\top \phi(\mathbf{k}_w^{(k)})} \cdot \mathbf{v}_v^{(w)} \tag{8}$$

$$= \frac{1}{\sqrt{k}} \frac{\phi(\mathbf{q}_u^{(k)})^\top \left(\sum_{v=1}^{N} \phi(\mathbf{k}_v^{(k)}) \cdot (\mathbf{v}_v^{(w)})^\top \right)}{\phi(\mathbf{q}_u^{(k)})^\top \sum_{w=1}^{N} \phi(\mathbf{k}_w^{(k)})}$$

With ϕ the non-parametric random feature map from [50].

The two summation terms over N nodes are independent from node u, which means they can be re-used by all the nodes after once computation, leading to an overall $\mathcal{O}(N)$ complexity.

In DIFFormer, the exponential function in the softmax attention is replaced by its first-order Taylor expansion. This new attention layer can be efficiently computed using linear complexity thanks to re-ordering the matrix product [3, 49]. Node embedding of the $(k+1)^{th}$ layer is now then calculated as:

$$\mathbf{z}_u^{(k+1)} = \frac{1}{\sqrt{k}} \sum_{v=1}^{N} \frac{1 + (\tilde{\mathbf{q}}_u^{(k)})^\top \tilde{\mathbf{k}}_v^{(k)}}{\sum_{w=1}^{N} \left(1 + (\tilde{\mathbf{q}}_u^{(k)})^\top \tilde{\mathbf{k}}_w^{(k)}\right)} \mathbf{v}_v^{(k)} \tag{9}$$

$$= \frac{1}{\sqrt{k}} \frac{\sum_{v=1}^{N} \mathbf{v}_v^{(k)} + \left(\sum_{v=1}^{N} \tilde{\mathbf{k}}_v^{(k)} \cdot (\mathbf{v}_v^{(k)})^\top\right) \cdot \tilde{\mathbf{q}}_u^{(k)}}{N + (\tilde{\mathbf{q}}_u^{(k)})^\top \sum_{w=1}^{N} \tilde{\mathbf{k}}_w^{(k)}}$$

$$\text{where } \bar{\mathbf{q}}_u^{(k)} = \frac{\mathbf{q}_u^{(k)}}{\|\mathbf{q}_u^{(k)}\|_2} \text{ and } \tilde{\mathbf{k}}_u^{(k)} = \frac{\mathbf{k}_u^{(k)}}{\|\mathbf{k}_u^{(k)}\|_2}$$

Authors of DIFFormer tested and validated that the first-order Taylor expansion is a well-posed approximation for the original Softmax attention in the context of Graph Transformers [49]. Again, as for NodeFormer, the two summation terms over N nodes are independent from node u, which means they can be re-used by all the nodes after once computation, yielding to an overall $\mathcal{O}(N)$ complexity.

SGFormer is composed of a one-layer global attention and a shallow GNN network. Contrary to all-pair attention that incurs $\mathcal{O}(N^2)$ complexity, this simple global attention allows for $\mathcal{O}(N)$ complexity [3].

1.5 Contribution and motivations

1.5.1 Retrieve information from cSCC WSIs with Histo-Miner pipeline

With the recent improvements of both histopathology and deep learning applied to histopathology (as seen in **Chapters 1.3 and 1.4**) it is now possible to use such deep learning computer vision models to study WSIs of cSCC skin cancer. In our work of **Chapter 2**, we used different computer vision models to segment and then classify nucleus of cells from cSCC WSIs.

First, using Hovernet trained on our dataset NucSeg, we segmented and classified all nucleus of the images following 5 classes: granulocytes, plasma cells, lymphocytes, stromal cells and tumor cells. Granulocytes are cells of the innate immune system. The main function of the innate immune system is to trigger a short-term immune response to a wide broad of pathogens, such as tumors. In contrast, plasma cells and lymphocytes (here T cells) are part of the adaptive immune system, which role is to react to specific pathogen and to build an immune response on the long-term. Stromal cells in the dermis release growth factors that foster cell division and play indirect role in the inflammation response. As a result of the segmentation and classification of nucleus, all nucleus of epithelial cells, that are either healthy or tumor, are all classified as tumor. This is because the morphology of healthy and tumor epithelial is very close, so we chose to train Hovernet with only one class for both type of nucleus. Next, we use another model to discriminate between these 2 types.

Segmenter Vision Transformer model was trained on our dataset TumSeg, to segment tumor region on each image. Then, we refine the classification previously made by Hovernet to discriminate between healthy and tumor epithelial: all nucleus within tumor regions segmented by Segmenter stay as tumor epithelial but all the nucleus outside of the tumor regions classified as tumor are then classified as healthy epithelial. The rational behind this design choice is that Segmenter is segmenting at a tissue level, where it is easier for model to recognize tumor regions than comparing cells morphologies directly.

Once we have a refined segmentation and classification of all nucleus in the image, we can calculate tissue-level features that will describe the tissue composition and the distances between cells of given types. Any Histo-Miner user can then input a cSCC WSI to the pipeline and receive a feature vector descriptive of tissue to use for downstream tasks, or to save it instead of the images as the vector is only few KB, and the image roughly around 100 GB. In our study, we applied the pipeline to predict cSCC patient response to immunotherapy, using our third dataset, CPI dataset. Additionally to predict the therapy response, using feature selection method we proved that percentages of lymphocytes, the granulocyte to lymphocyte ratio in tumor vicinity and the distances between granulocytes and plasma cells in tumors are predictive features for therapy response.

1.5.2 Use of Histo-Miner to analyze factors associated with cSCC progression in the context of federated learning

While we used Histo-Miner in the case of predicting cSCC patient response to immunotherapy in **Chapter 2**, the dataset used in this case study was relatively small (45 samples of 45 patients) and all patients came from the same cohort. In the work of **Chapter 3** we applied Histo-Miner to interpret prediction of a vision transformer model trained in a federated way, implemented by McMahan et al [51], on 3 different cohorts, treated as 3 different clients on the federated learning perspective. After using Integrated Gradient Method [52] to identify most relevant patches during inference, we applied Histo-Miner to the kept patches to segment and classify nucleus and produce feature vector descriptive of the patch. In this specific case, as we are applying Histo-Miner on patches, only a subset of feature calculated within Histo-Miner are relevant. Features describing distances between cells at the WSI level, or composition of the vicinity of tumors are not applicable as not enough cells are inside a single patch.

From these observations we found biological parameters that differed between the progressor and non-progressor most relevant patches. For instance, tumor cells of non-progressor have larger nucleus size and lower nuclear eccentricity. Also, keeping the most representative patches, and using the cell-based features calculated from Histo-Miner for patch classification, we reached high prediction accuracy. As a result, we concluded that Histo-Miner features captured relevant biological parameters and variations associated with progression risk of patients.

1.5.3 Scalable Graph Transformers to improve classification between healthy and tumor epithelial

Finally, we used a new approach to improve the classification between healthy and tumor epithelial cells. Indeed, in the case of Histo-Miner (developed in previous **Chapters**), we refine the classification based on the tumor region segmented by Segmenter model. All cells classified as epithelial within the tumor region are re-classified as tumor epithelial, and epithelial cells outside tumor regions are re-classified as healthy epithelial. Nevertheless, some isolated tumor cells or small tumor regions, too small to be detected by Segmenter, can be missed by this two-models approach. In **Chapter 4** we present our work on graphs generated from WSIs and on Graph Neural Networks for binary node classification. From WSI image segmented by SCC Hovernet, we generate a graph where nodes contains feature associated to a given cells, as well as its class, and edges are connecting nodes corresponding to neighboring cells in the image.

We compared performance of Graph Neural Networks including Graph Transformers with linear complexity to state of the art computer vision models Hovernet and CellViT for binary classification of epithelial cells as healthy or tumor. We first evaluated both modalities in the case of a semi-automatically annotated cSCC WSI. We generated a cell graph with 401.943 nodes, WSI-Graph-401K, from this WSI and evaluated several Graph Neural Networks on node classification task, with nodes representing cells. We evaluated Hovernet and CellViT performance trained and tested on the annotated WSI patches and showed that the binary classification was poorer than with Graph Transformers SGFormer and DIFFormer. Additionally training time was of around 5 days for CellVit and 4 min 18 s for SGFormer on the same GPU, same number of epochs and two-thirds of the dataset. We also evaluated performance of SGFormer for different construction of the WSI-Graph-401K, with different type of node features and different levels of graph simplification.

Finally, we compared image-based and graph-based approach in the context of a dataset including several patients. Indeed, from 372 H&E patches of 84 patients, we assembled TILE-Graphs-572k dataset containing 372 medium size graphs generated from the H&E patches. Running 3 fold cross-validation with samples from the same patient kept in the same folds, we showed that DIFFormer model outperformed CellViT model in the classification task. Translating WSI images and patches into cell graphs prior to cell classification task seems to be a promising approach for future work both for classification accuracy and computational budget needed to train models.

Chapter 2

Histo-Miner: deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma

In this work we collected 3 new datasets in order to train deep learning models to ultimately generate a compact feature vector summarizing tissue morphology and cellular interactions of input WSIs. We implemented a multi-model pipeline that was tested in the clinical relevant scenario of predicting cSCC patient response to immunotherapy.

The datasets that were generated for this work all contain H&E WSIs of cSCC (either full WSI or patches) and are; **NucSeg** consisting of 47,392 nuclei labeled on 1,707 H&E non-overlapping patches of 256x256 pixels, with 40x and 20x resolutions; **TumSeg** consisting of 144 WSIs of 125 cSCC patients, for which each tumor regions were labeled (contours drawn); and **CPI dataset** consisting of 45 WSIs, each from a different patient, collected prior to receiving immunotherapy, and annotated as treatment responders and treatment non-responders.

We re-trained **Hovernet** and **Segmenter** models described in **Section 1.4** to segment and classify cell nucleus of input WSI. We implemented a 3-steps training to improve model performance. These segmentations are then used to calculate tissue-wise features that can be utilized for downstream tasks. As use case, we trained **XGBoost** classifier on the feature vector obtained to classify response to immunotherapy treatments. In addition to be able to predict if a patient may or may not respond to the immunotherapy treatment, we used feature selection algorithm **Boruta** to find interpretable predictive feature for therapy response.

NOTE: The following pages contain the latest arXiv version of the published manuscript. The published manuscript is also available in PLOS Computational Biology journal following the citation in **Publication Preface** section.

Histo-Miner: Deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma

Lucas Sancéré[1,2,3], Carina Lorenz[3,5,6], Doris Helbig[7], Oana-Diana Persa[8], Sonja Dengler[9], Alexander Kreuter[10], Martim Laimer[11], Roland Lang[11] , Anne Fröhlich[12], Jennifer Landsberg[12], Johannes Brägelmann[3,5,6,13 ⋆], and Katarzyna Bozek[2,3,4 ⋆]

[1] Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[2] Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[3] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[4] Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, North Rhine-Westphalia, Germany
[5] University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Translational Genomics, Cologne, Germany
[6] University of Cologne, Faculty of Medicine and University Hospital Cologne, Mildred Scheel School of Oncology, Cologne, Germany
[7] Department for Dermatology, University Hospital Cologne, Cologne, Germany
[8] Department of Dermatology and Allergy, School of Medicine, Technical University of Munich, Bavarian Cancer Research Center (BZKF), Munich, Germany
[9] Department of Dermatology, Dortmund Hospital gGmbH, University Witten/Herdecke, 44137 Dortmund, Germany
[10] Department of Dermatology, Venereology and Allergology, Helios St. Elisabeth Hospital Oberhausen, University Witten/Herdecke, Oberhausen, Germany
[11] Department of Dermatology and Allergology, University Hospital of the Paracelsus Medical University Salzburg, Salzburg, Austria
[12] Department of Dermatology and Allergology, University Hospital Bonn, Bonn, Germany
[13] Medical Clinic III for Oncology, Hematology, Immune-Oncology and Rheumatology, University Hospital Bonn (UKB), Germany

lsancere@uni-koeln.de, johannes.braegelmann@uni-koeln.de, k.bozek@uni-koeln.de

Abstract. Recent advances in digital pathology have enabled comprehensive analyses of Whole-Slide Images (WSIs) from tissue samples, leveraging high-resolution microscopy and computational capabilities. Despite this progress, available tools for automatic cell type identification perform poorly on skin tissue, e.g. in the classification of non-melanoma tumor cells. This is due to a paucity of labeled training data sets and high morphological similarities between tumor and non-tumor epithelial cells in the skin. Here, we propose Histo-Miner, a deep learning-based pipeline designed for the analysis of skin WSIs. To this end we generated two new datasets using WSIs of cutaneous Squamous Cell Carcinoma (cSCC) samples, a frequent non-melanoma skin cancer, by annotating 47,392 cell nuclei across 5 cell types in 21 WSIs and segmenting tumor regions in 144 WSIs. Histo-Miner employs convolutional neural networks and vision transformers for nucleus segmentation and classification, as well as tumor region segmentation. Performance of trained models positively compares to state of the art with multi-class Panoptic Quality (mPQ) of 0.569 for nucleus segmentation, macro-averaged F1 of 0.832 for nucleus classification and mean Intersection over Union (mIoU) of 0.907 for tumor region segmentation. From these output, the pipeline can generate a compact feature vector summarizing tissue morphology and cellular interactions, which can be used for various downstream tasks. As an exemplary use-case, we deploy Histo-Miner to predict cSCC patient response to immunotherapy based on pre-treatment WSIs from 45 patients. Histo-Miner predicts patient response with mean area under ROC curve of 0.755 ± 0.091 over cross-validation, and identifies percentages of lymphocytes, the granulocyte to lymphocyte ratio in tumor vicinity and the distances between granulocytes and plasma cells in tumors as predictive features for therapy response. This highlights the applicability of Histo-Miner to clinically relevant scenarios, providing direct interpretation of the classification and insights into the underlying biology. Importantly, Histo-Miner is designed to allow for its use on other cancer types and on other training datasets. Our tool and datasets are available through our github repository: https://github.com/bozeklab/histo-miner.

⋆ Equal contribution

Author Summary

Digital pathology is transforming how we study disease by turning tissue samples into high-resolution images that capture the architecture of entire tumors. However, these images are vast and complex, making it difficult to extract meaningful clinical insights without advanced computational tools. In this work, we present Histo-Miner, a framework designed to systematically analyze these images at multiple levels of detail—from one single cell to entire tissue regions. We apply this approach to cutaneous squamous cell carcinoma, a common form of skin cancer, demonstrating how large-scale tissue data can be mined for biological insights. Our method identifies and characterizes different types of cells, maps how they are organized within tumor areas, and connects these patterns to patient outcomes. Through this lens, we uncover subtle features of the tissue environment that may influence how patients respond to therapy. We find that the most informative features describe the presence and balance of different types of immune system cells, and how these cells are spatially arranged within the tissue. Beyond its immediate findings, Histo-Miner, provides openly available data and tools that aim to make large-scale tissue analysis more interpretable, reproducible, and transferable to other diseases.

Introduction

Digital pathology slide scanners and advancements in computer vision allow for automation of diagnostic pathology tasks. Hematoxylin and Eosin (H&E) staining [1] is widely used in pathology and represents a standard that both the classical and digital pathology are based on. The resulting tissue scans are called Whole-Slide Images (WSIs). Given the large size of WSIs containing thousands to millions of cells, automated methods for WSI analysis are indispensable to systematically and comprehensively quantify their content. A large panel of tasks can be performed by machine learning and deep learning models on this type of images: segmentation of nuclei and tumors in the WSIs [2,3], image classification [4], or discovery of new biomarkers [5]. Importantly, such methods automate time consuming intermediary tasks, such as cell counting, that allows the practitioner to focus on diagnosis and interpretation [6,7].

While there is a range of datasets and methods in digital pathology [8,9,10,11], few of them are dedicated to skin and non-melanoma skin tumors. Skin differs from other tissues in its unique structure, composition, and function, which presents specific challenges for digital pathology methods. The skin consists of multiple distinct layers, including the epidermis, dermis, and subcutaneous tissue, each with varying cell types, densities, and extracellular matrices. These variations lead to textural patterns and coloration that are unique to WSIs of skin. . Therefore, specialized methods tailored to the unique characteristics of skin tissue are necessary for reliable digital pathology in dermatology.

Here, we focus on cutaneous squamous cell carcinoma (cSCC) - the second most common form of non-melanoma skin cancer in the USA and widely spread worldwide [12]. While the majority of cSCCs can be cured by surgery alone, 5-10% of cSCC patients experience disease recurrence or metastases, requiring systemic treatments [13]. While the effects of chemotherapy are very limited, systemic treatment with immune checkpoint inhibition (CPI) has emerged as a promising alternative. However, up to half of patients do not respond to the immunotherapy. To date, it is impossible to predict, which patients have a high chance of response to CPI and which patients may require other/additional treatment modalities [13]. Quantitative methods for analysis of cSCC patient samples would provide more insights into the morphological variability of this

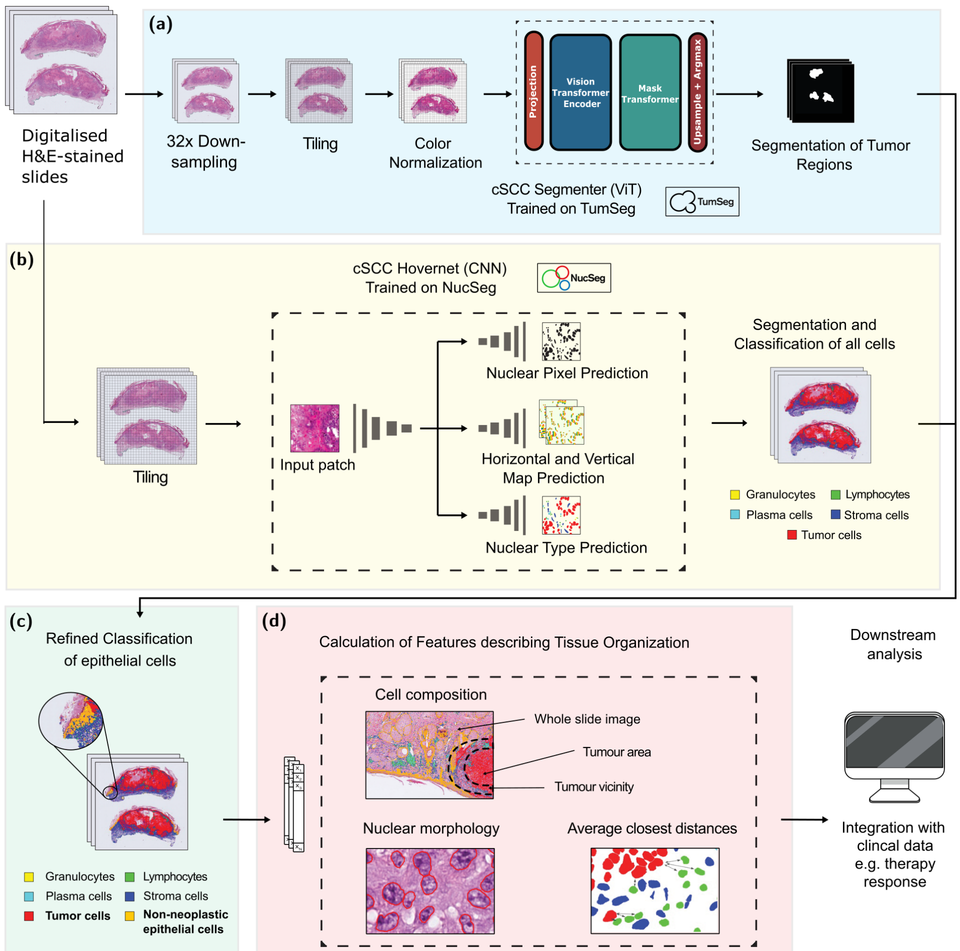


Fig. 1: Overview of Histo-Miner pipeline. The pipeline uses WSI from cSCC patient as input. **(a) & (b)** During inference, WSI images are tiled into patches and undergo pre-processing pipelines (see **Methods**). After pre-processing, SCC Segmenter performs tumor region binary segmentation on processed patches and SCC Hovernet segments and classifies cell nuclei. **(c)** Using the output of SCC Segmenter, cell classification is refined by adding a new cell class: non-neoplastic epithelial. This last result is saved in a json text file. A visualization of resulting annotations is provided in **Fig. 2**. **(d)** Using refined segmentation and classification of nuclei, together with segmentation of tumor regions, we calculate features that describe the tissue organization. Example features include e.g. percentage of lymphocytes in the vicinity of the tumor and average closest distance between tumor and lymphocyte cells. A list of all 317 calculated features is provided in **Supplementary Data**.

tumor type and could potentially allow for identification of morphological markers linked to the patient risk of progression.

We propose a deep learning-based pipeline, Histo-Miner with openly available code and datasets, for development and analysis of cSCC samples. We generated a dataset called NucSeg containing manually annotated class labels and segmentation masks of 47,392 cell nuclei from 21 WSIs of cSCC. Furthermore, we generated TumSeg dataset, containing binary segmentation masks of tumor regions in 144 WSIs of cSCC. Our pipeline performs segmentation and classification of cell types into 6 different classes (granulocytes, lymphocytes, plasma cells, stromal cells, tumor cells and epithelial cells) and tumor region segmentation, both using deep learning models trained on our datasets. Histo-Miner uses the segmentation and classification results to encode WSIs of cSCC into features describing tissue morphology, organization, and cellular interactions. The code is open-source, customizable, and each part (tumor segmentation, cell type identification) can be used separately to fit user needs. The training datasets, as well as the models weights used in the intermediary steps are publicly available (See **Data Availability** and **Code Availability** sections).

We finally tested our Histo-Miner pipeline to predict cSCC patient response to immunotherapy. Immunotherapy with anti-PD1 antibodies is the major treatment for patients with advanced cSCC, but currently no predictive biomarkers are established to identify patients with a high likelihood of therapy failure. We generated the CPI dataset including 45 skin WSIs of 45 patients, before they received immunotherapy treatment, and annotated these slides as responder or non responder to the treatment. Using the features produced by Histo-Miner we classified patient response and found interpretable features explaining model choices. These features provide insights into biological factors favoring treatment response. The CPI dataset and the feature list are publicly available (See **Data Availability** section).

Methods

Histo-Miner pipeline description

To describe the tissue organization and composition of cSCC, and obtain detailed information on the histomorphology of these tumors we developed our pipeline, Histo-Miner (**Fig. 1**). It uses both cell nuclei and tumor segmentation as first steps to quantitatively describe tumor sample morphology.

In the pre-processing pipeline of SCC Segmenter, the images are downsampled, tiled into patches and the patches are normalized using mean and standard deviation of RGB pixel values of ImageNet 1K (see our github repository for implementation details). SCC Hovernet is trained with data augmentation for model generalization [14] and then input patches don't need color normalization. To capture cellular heterogeneity, the cell nuclei are segmented and classified into: granulocytes, lymphocytes, plasma cells, stromal cells, and tumor cells. Cell nuclei segmentation and classification is performed with Hovernet convolutional network [14] trained on a manually annotated set of cSCC WSIs (NucSeg). This model shows better instance segmentation and improved Panoptic Quality (PQ) compared to other recent H&E segmentation models [14,15]. The model is open source and allows users to train and adapt it. Our pipeline additionally includes Segmenter vision transformer network to segment tumor regions. This vision transformer model outperforms other models in several benchmark tasks of instance and semantic segmentation [16,17,18]. We trained the Segmenter model on our TumSeg dataset to perform a binary pixel-wise classification tumor and non-tumor regions of the same WSIs that were used for Hovernet inference. We name the resulting trained networks SCC Hovernet and SCC Segmenter.

In the pre-processing pipeline of SCC Segmenter, the images are first downsampled, then tiled into patches and the patches are normalized using mean and standard deviation of RGB pixel values of ImageNet 1K (see our github repository for implementation details). In the pre-processing pipeline of SCC Hovernet, WSI input are only tiled into patches. In both cases, only patches with tissue are kept for prediction (at least one pixel of tissue) after tiling.

After determining tumor areas using SCC Segmenter, the results of the SCC Hovernet cell nuclei classification are updated to add a new cell class as follows: all the nuclei predicted as tumor cells outside of the predicted tumor regions are reclassified as healthy epithelial. The reason for this update is that healthy epithelial cells and tumor cells have similar morphologies and are impossible to discriminate without a broader context and information about the tissue structure (see **Supplementary Fig.1**). Example visualization of the inference of the two methods is shown in **Fig. 2**.

Results of segmentation are input to tissue analysis part of our pipeline. In this part we perform calculation of 317 features that describe and encode the tissue samples. Example features include percentages of cells of specific class anywhere in the sample as well as inside the tumor regions. For every pair of cell classes X and Y, we also calculate the average distance of the closest cell of class Y to a cell of class X inside the tumor regions. This feature describes the topology of the tissue and the interactions between cell classes.

An exhaustive list of the calculated features is available in **Supplementary Data**. We do not consider absolute numbers of cells of a given class as a feature, but cell densities and percentages, as this metric is dependent on the WSI size and not the structure of the tissue itself. All the features are stored in a light json file, which results in encoding and compressing a WSI of multiple GB into a text file of 3.7 KB. These features are a convenient WSI representation for any downstream analysis. All the different steps of the pipeline can be run separately as well as configured to fit specific needs. An example of use case, predicting response of cSCC patients to immunotherapy, is provided in the following sections.

NucSeg and TumSeg datasets descriptions

To enable segmentation and cell nucleus type classification for cSCC, we assembled 21 WSIs of H&E-stained tissue sections of 20 cSCC patients from the University Hospital Cologne. The images were acquired using a NanoZoomer Slide Scanner (Hamamatsu). In the images the nuclei contours were marked and assigned to five cell types: granulocytes, lymphocytes, plasma cells, stromal cells, and tumor cells. 1,707 H&E non-overlapping patches of 256x256 pixels, with 40x and 20x resolutions, have been manually annotated by two pathology experts. To ensure annotation consistency across the distributed workflow, ambiguous morphological patterns were subjected to joint consensus review and compared to IHC staining on validation slides. 47,392 nuclei were labeled (classified and segmented) in total (3,135 granulocytes, 12,263 lymphocytes, 3,271 plasma cells, 11,526 stromal cells, 17,197 tumor cells), see **Fig. 3a** and **Fig. 3b**.

The annotations consist of two groundtruth patches for each H&E patch. The first annotation is an instance segmentation of each nucleus. A unique value is attributed to

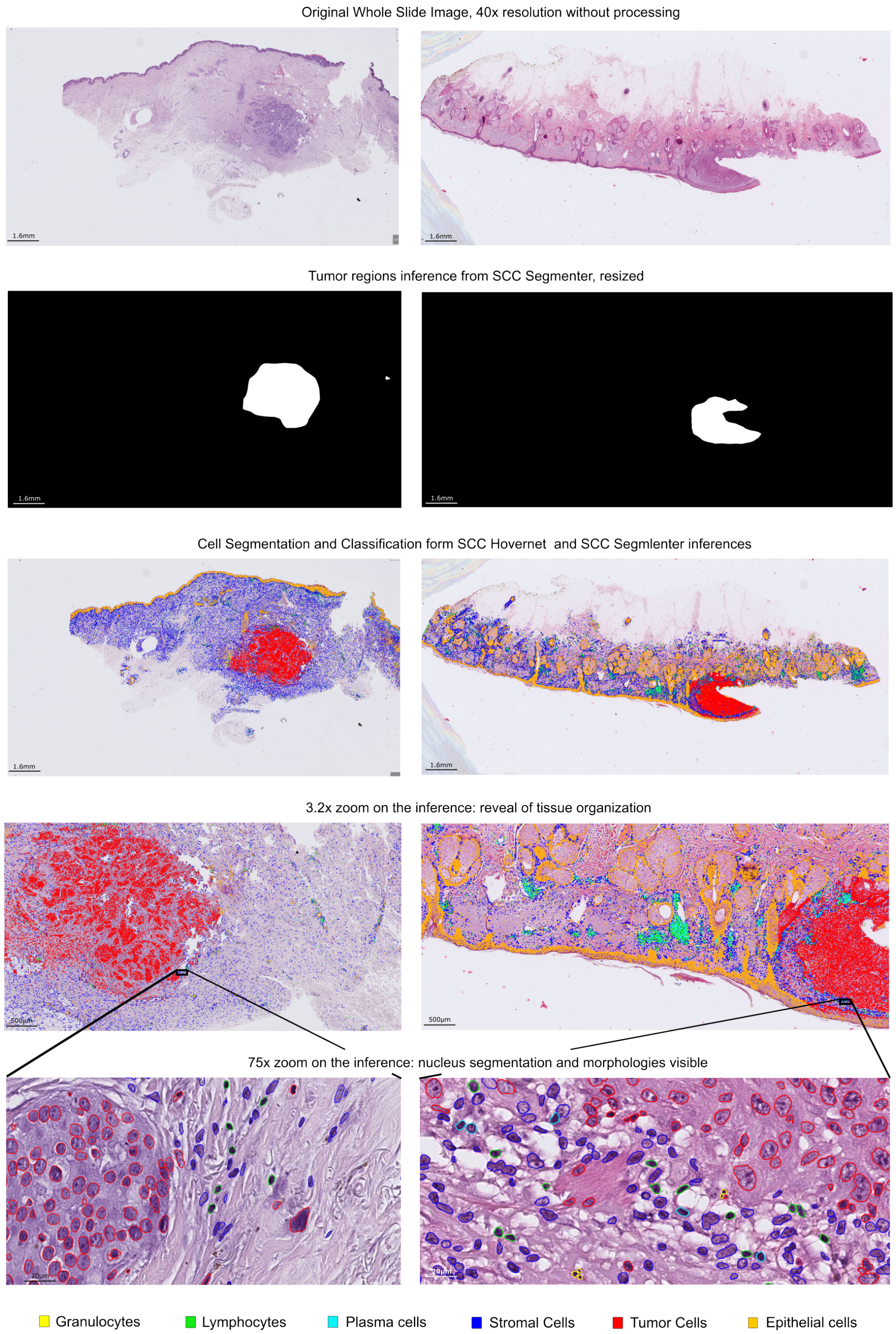


Fig. 2: Predictions of SCC Hovernet and SCC Segmenter models. The different images correspond to different steps of the Histo-Miner pipeline as depicted in **Fig. 1**. 2 WSIs of 2 different patients from 2 different cohorts are shown. The H&E staining differs between the slides showing varying hues of blue and pink. After predicting the tumor area with SCC Segmenter, Histo-Miner segments and then classifies cells into five different classes: granulocytes, lymphocytes, plasma cells, stromal cells, tumor cells. Using tumor segmentation, tumor cells detected outside tumor regions are re-classified as non-neoplastic. The cell nuclei segmentation and classification illustrate sample organization at tissue level (3x zoom), or at cell level (75x zoom). Based on segmentation results Histo-Miner calculates features describing cell-level and tissue-level tumor organization. Also in the case of damaged sample (one part of the tumor is missing in the WSI on the right), the model is not hallucinating segmentation of the remaining parts of the sample.

pixels belonging to the same nucleus instance, where every nucleus is assigned a different value. Annotation of nuclei instances allows to differentiate touching instances based on their instance pixel values. All pixels outside nucleus (background) are given the pixel value 0. The second annotation is a type map of the nucleus, in which the pixels belonging to nucleus of the same class have the same pixel values. The dataset is also available as 6,816 patches of 560x560 pixels with 70% overlap in a 5D numpy array according to the Hovernet data format requirements.

To build a tumor segmenter algorithm, we additionally assembled 144 WSIs of 125 cSCC patients from 3 medical centers - Bonn, Cologne and Munich. These WSIs are a subset of the dataset described in [19]. All tumor regions were labeled by two pathology experts. The WSIs were originally at resolution 40x and downsampled 32 times (reaching a final resolution of 1.25x) to enable to fit them into memory during model training (**Fig. 3a**). The resulting image–tumor annotation pairs constitute our TumSeg dataset, more precisely, downsampled WSIs and the binary segmentation masks of their tumor regions. We also assembled and labeled 32 slides from 25 other patients to create a test set for model evaluation (see **Results** section). Examples of WSI tumor segmentation annotations are displayed in **Fig. 3c**. . In addition, anonymized patient IDs and medical centers IDs are available as metadata.

These 2 datasets, NucSeg and TumSeg, are openly available in our Zenodo repository link.

Histo-Miner deep learning models

Histo-Miner implementation includes deep learning models trained with our custom datasets. Histo-Miner users can utilize the weights of these models to perform similar inferences on their own datasets, re-train these models through Histo-Miner implementation directly and edit model architectures and hyperparameters for further development. To fit user needs, it is possible to use only specific blocks of the pipeline - such as inference of the deep learning models - instead of using the whole process until calculation of tissue features.

We performed segmentation and classification of segmented cell nuclei using Hovernet model [14], which we selected based on its performance and ease of use. The model is a convolutional neural network containing encoder and decoder parts. The semantic segmentation of tumor region on the WSIs was achieved using Segmenter, a vision transformer model [16]. Segmenter is a collection of architectures with varying size, composed of encoder and decoder parts. In Histo-Miner pipeline we use the Seg-L-Mask/16 segmenter variant, achieving better results than the base model but requiring more GPU memory.

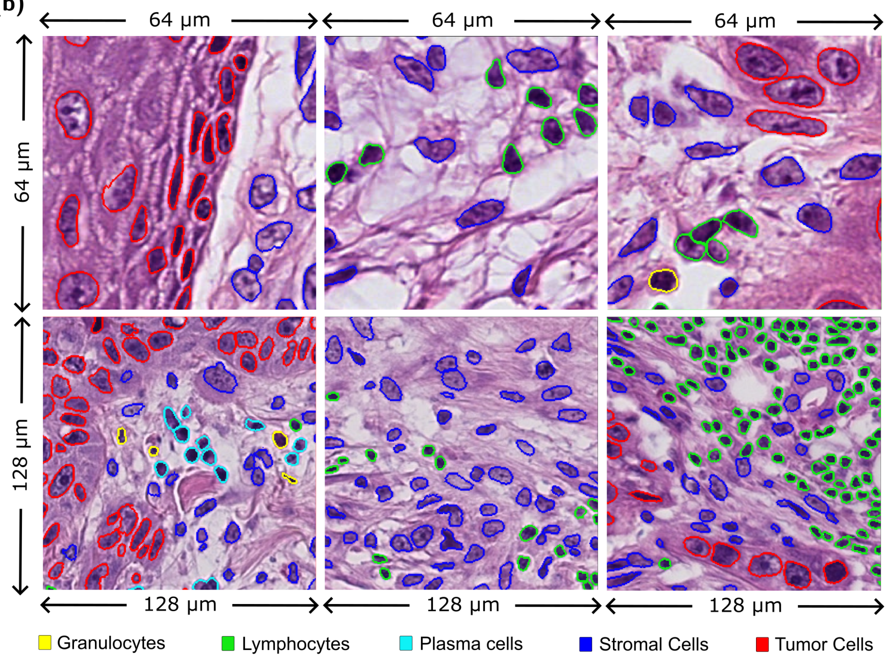
Training SCC Hovernet

To segment and classify cell nuclei into different cell classes (granulocytes, lymphocytes, plasma cells, stromal cells, and tumor cells) we trained Hovernet network with our dataset NucSeg. The training was performed on 2 80GB A100 NVIDIA GPUs (Ampere micro-architecture). We used Hovernet with the encoder pretrained on ImageNet 21k. A second pre-training of 150 epochs was performed on a not-curated dataset of H&E nucleus segmented and classified, also made openly available. This dataset resemble NucSeg, consists of the same classes, but the segmentation and classification were mostly

(a)

NucSeg		TumSeg	
Data Type	Skin H&E of cSCC	Data Type	Skin H&E of cSCC
Annotation Type	Cell Contour & Class	Annotation Type	Tumor Contour
WSIs	24	WSIs	144
Resolution	20x & 40x	Resolution	1.25x
Patients	20	Patients	12
Patches Size	256x256		
Patches	1,707		
Annotated Cells	47,392		
Granulocytes	3,135		
Lymphocytes	12,263		
Plasma Cells	3,271		
Stromal Cells	11,526		
Tumor Cells	17,197		

(b)



(c)

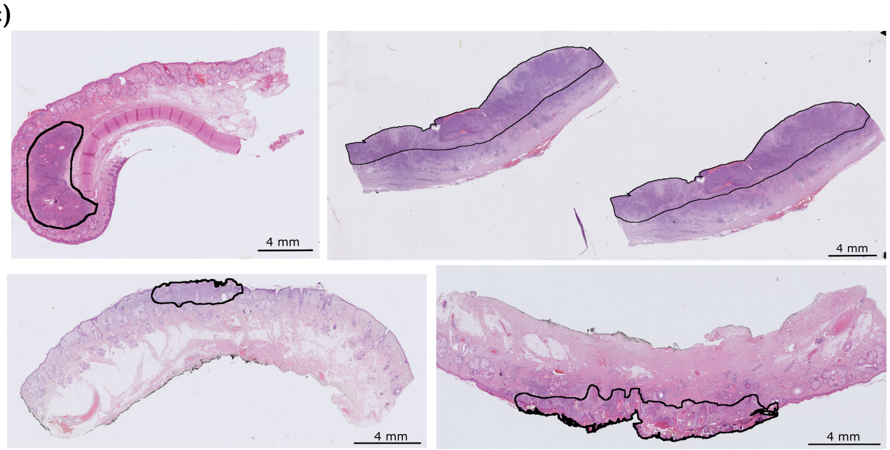


Fig. 3: Visualization of samples from NucSeg and TumSeg training datasets. **(a)** Overview of both datasets **(b)** Visualization of NucSeg training dataset. 47,392 cell nuclei from 1,707 H&E non-overlapping patches were segmented and classified. **(c)** Visualization of TumSeg training dataset. Tumor region are segmented by 2 experts on 144 WSIs.

automatized and not fully corrected by human experts. During the first 50 epochs, the encoder weights were frozen and during the following 100 epochs all weights were updated. Finally, the main training of 250 epochs (first 50 epochs with frozen encoder weights followed by 200 epochs with all weights updated) was performed on NucSeg dataset. We used Adam optimization algorithm [20] with initial learning rate γ_0 of 10^{-4} which was reduced to 10^{-5} after 25 epochs. We used the same loss functions as in the original Hovernet model. The global loss function is an addition of three losses for each of the three branches of the Hovernet Model. These branches account, respectively, for the nuclear pixel classification task, the binary segmentation of nucleus, and the horizontal and vertical distance map used to separate touching instances [14]. We describe the loss function in detail in **Supplementary Eq.1 - 4**. We used the following data augmentations during training: image flips, rotations, Gaussian blurs and median blurs according to the original Hovernet implementation.

We trained NucSeg dataset including 5,968 patches of size 540x540 pixels with 70% overlap. Our validation set contained 848 patches of size 540x540 pixels with 70% overlap. We kept the same set of hyperparameters as the original implementation [14], except that we doubled the training batch size for our last training step of 200 epochs (see above). We tested different training strategies combining network pre-trained on ImageNet 21k, pre-trained on the not-curated H&E nucleus dataset, and trained from scratch. Performance of resulting models was highest when the model was first pre-trained on ImageNet 21k, then pre-trained on the not-curated H&E nucleus dataset before fine-tuning. The choice of final model was based on maximizing the panoptic segmentation task performance.

Training SCC Segmenter

We trained Segmenter model with our dataset TumSeg. The training was performed on 2 80GB A100 NVIDIA GPUs (same as for Hovernet training). We used Vision Transformer pre-trained on ImageNet 21k [21] and fine-tuned it on our dataset TumSeg. We followed the data augmentation pipeline from the semantic segmentation library MMSegmentation [22]. It consists of random resizing of the image to a ratio between 0.5 and 2.0, random left-right flipping, and normalization of the images based on mean and standard deviation of pixel values of ImageNet 1k. We tested other normalization strategies, e.g taking mean and standard deviation of training dataset which resulted in worse model performance. We trained for 1,786 epochs (50,000 iterations) using Stochastic Gradient Descent as optimization method with learning rate following a polynomial decay scheme. Considering γ the learning rate at the current iteration number, and γ_0 the base learning rate, the decay is defined as $\gamma = \gamma_0(\frac{1-N_{iter}}{N_{total}})^{0.9}$ where N_{iter} and N_{total} represent the current iteration number and the total iteration number, respectively. We set γ_0 to 10^{-3}. We used cross-entropy without weight re-balancing as the loss function.

The model was trained on randomly chosen 115 slides of TumSeg dataset and the validation set contained the remaining 29 slides of the dataset. We performed hyperparameter grid-search to find the best set of hyperparameters as described in **Supplementary Table. 1**. The accuracy estimation was performed on the validation set due to the limited number of slides and lack of an independent test set.

Tissue Analyser

Within Histo-Miner, SCC Segmenter model segments tumor regions. SCC Hovernet model performs instance segmentation of cell nuclei and classifies them into different cell classes (granulocytes, lymphocytes, plasma cells, stromal cells, and tumor cells). Using both models' predictions, we refine the cell classification and add one more class among the possible predictions, the healthy epithelial class. In fact, healthy epithelial cells and tumor cells are hard to discriminate without a broader context and information about the tissue structure. Healthy epithelial cells and tumor cells have similar morphologies. To distinguish these two cell types, we added one refinement step: all the nuclei predicted as belonging to tumor cells by SCC Hovernet, located outside the tumor regions predicted by Segmenter, are reclassified as epithelial. The result of the classification update is visible in **Fig. 2**.

The updated cell nuclei segmentation and classification as well as the tissue segmentation are input to our Tissue Analyser part of the pipeline. Here we calculate 317 features capturing various aspects of tissue morphology and spatial organization. The features (see **Supplementary Data**) include: percentages of given cell types, composition of the tumor margin, ratio between cell types, repartition of a given cell type outside, inside and within the tumor margin. The final feature vector is a light but information-rich representation of the cSCC WSI for further downstream analyses.

One of our features is the average distance of the closest cell of a given class X - source class - to the closest cell of a given class Y - target class - inside the tumor regions. The average distance is calculated as shown in **Eq. 1** and through the following steps: **1-** We generate a rectangle that defines the initial search area. The rectangle height is 5% of the tumor bounding box height and width 5% of the tumor bounding box width. It is centered around the nucleus of the source cell class. **2-** We verify if there is at least one nucleus of the target cell class inside this search area. **3-** If there is at least one, we calculate all distances between source and target cells and keep the smallest one. If there is no nucleus of target cell class we increase the search area until we find at least one nucleus of target cell class in the search area. The search area cannot extend to other tumor areas. **4-** We perform steps 1-3 for all nuclei of the source class. A quantitative explanation is available in **Algorithm. 1** (all the memory optimization steps are skipped for readability). The increase of search area for each iteration was optimized to reduce calculation time. For computation optimization reasons, the search area is a rectangle and not a circle. Indeed, one of the main reason is that searching for coordinates in a rectangle is faster than searching for coordinates in a circle (only comparisons instead of subtractions and multiplications). In some specific cases, using a rectangle search area can lead to overestimation of the distance. Description of these cases, probability of overestimation, and bounding of overestimation are described in **Supplementary Fig. 3**. This probability of overestimation decreases drastically with the number of cells in the tumor. For instance, we can calculate that for a squared search area of side length $2r_1$ and origin 0, if N cells are in the circle of radius $\sqrt{2}r_1$ and origin 0 (the square is inscribed in this circle), the probability of overestimation is $P(error_N)_2 = 0.022$ for $N = 2$ and $P(error_N)_{10} = 2.8\text{x}10^{-4}$ for $N = 10$. These distances describe the interactions between different cell types inside the tissue. They are calculated for granulocytes, lymphocytes, plasma cells, and tumor cells to assess the organization of the tissue regarding the intensity of the immune response.

$$\bar{d}_{closest_{c_A,c_B}} = \frac{1}{n_{c_A}} \sum_{(x_i,y_i)\in E_{c_A}} \min\left(\sqrt{(x_i-x_k)^2+(y_i-y_k)^2}\right), \tag{1}$$

$$\forall (x_k,y_k) \in \left(E_{c_B} \cap E^{\lambda_0}_{(x,y)\in\mathcal{N}^2 \mid \begin{cases} 0.05\ \lambda_0 l_t - x_i \leq x \leq 0.05\ \lambda_0 l_t + x_i \\ 0.05\ \lambda_0 w_t - y_i \leq y \leq 0.05\ \lambda_0 w_t + y_i \end{cases}}\right)$$

with λ_0 verifying:

$$\forall \lambda \in \mathcal{N}^* \mid \left(E_{c_B} \cap E^{\lambda}_{(x,y)\in\mathcal{N}^2 \mid \begin{cases} 0.05\ \lambda l_t - x_i \leq x \leq 0.05\ \lambda l_t + x_i \\ 0.05\ \lambda w_t - y_i \leq y \leq 0.05\ \lambda w_t + y_i \end{cases}}\right) \neq \{\emptyset\},\ \lambda_0 = \min(\lambda)$$

where $\lambda \in \mathcal{N}^*$ coefficient of increase of sides length of search rectangle, l_t and w_t length and width of the bounding box of the tumor region, E_{c_a} set of coordinates in $\mathcal{N}^2$ of all cells centroid of class A (source class), E_{c_b} set of coordinates in $\mathcal{N}^2$ of all cells centroid of class B (target class), n_{c_A} number of cells of class A.

E^{λ} is the set of points inside the search rectangle of coefficient λ and E^{λ_0} is the set of points inside the smallest search rectangle that contains at least one centroid of cell of class B.

Algorithm 1 Minimum Distance Calculation Pseudo-Code

Require: *classjson*: file with centroid and type of all WSI's cells
Require: *maskmap*: segmentation of WSI's tumor regions
Require: *selected_classes*: classes for which we want to calculate distances, f: acceleration of increase parameter - here $f = 0.05$
Start $C(n,2) = \frac{n(n-1)}{2}$ parallel processes, $n = \mathbf{len}(selected\_classes)$
In each process a different pair source class / target class is defined
For each process:
    $allmindist = \mathbf{list()}$
    For all cells of source class:
        $dist\_list = \mathbf{list()}$
        Check tumor ID of source cell
        List all target class cells *centroid coordinates* in the same Tumor
        Calculate length l_t and width w_t of the bounding box of the Tumor
        $\lambda = 1$
        $kept\_targets = \mathbf{list()}$
        While *kept_targets* cells list is **empty**:
            Construct search zone around source cell with $length = f\lambda l_t$ and $width = f\lambda w_t$
            Append to *kept_targets* **all** target cells *centroid coordinates* inside the search area
            $\lambda = \lambda + 1$
        For all cell in *kept_targets*:
            $dist = \sqrt{(x_i-x_k)^2+(y_i-y_k)^2}$ with (x_i,y_i) coordinates of cells of source class and (x_k,y_k) coordinates of cells of target class
            $dist\_list.\mathbf{append}(dist)$
        $min\_dist = \mathbf{min}(dist\_list)$
        $allmindist.\mathbf{append}(min\_dist)$
    Put item $avgdist = \mathbf{sum}(allmindist)/\mathbf{len}(allmindist)$ in process queue
Ouput a list of all *avgdist* gathered items

Other features include ratios of cell types and percentages of cells of a given cell type in the vicinity of the tumor, in the tumor regions or in the whole WSI. The vicinity of tumor is defined as a 1mm-wide area around the tumor [23]. In the case of ratios, to limit outlayers we define ratio $\eta = \frac{\log(n_{c_A})+\epsilon}{\log(n_{c_B})+\epsilon}$. with n_{c_A} number of cells of class A, n_{c_B} number of cells of class B and $\epsilon = 10^{-3}$, smoothness parameter. The full list of features is in the **Supplementary Data**.

Feature selection to predict response of cSCC patients to immunotherapy

We collected WSIs from 45 patients (one per patient) with cSCC skin cancer, from 6 medical centers - Bonn, Cologne, Dortmund, Munich, Oberhausen and Salzburg taken before administration of immune checkpoint inhibitors treatment. 28 of them were classified as responders, showing partial response (PR) or complete response (CR), i.e. tumor shrinkage. 17 patients were classified as non-responders, showing stable disease (SD) or progressive disease (PD) states. More specifically, CR means disappearance of all lesions; PR: 50% or more in decrease of total tumor size; SD: <50% decrease and/or <25% increase of one/several tumor lesions; PD: >25% increase of one/several tumor lesions or new lesions. The classification was determined by a dedicated review of the clinical and radiological imaging by at least two observations not less than 4 weeks apart (following World Health Organization handbook for reporting results of cancer treatment). This collection of classified WSIs represents our fourth dataset, called CPI (see **Data availability** section).

We used Histo-Miner to extract tissue representative features from the WSIs of CPI dataset. Then, we trained and evaluated an XGBoost classifier [24] for the task of classifying patients in their two categories based on the feature vectors. We evaluated XGBoost classifier through 3 fold cross-validation of 2 splits, train and test (train containing 2/3rd of the data), and performed feature selection on the training split within the cross-validation runs. Due to the limited number of samples in the dataset (45) we could not perform hyperparameter search within nested cross-validation so we kept the default set of hyperparameters for all the cross-validation runs. Similarly, the low number of folds is constrained by the low number of samples. In fact, having too small validation folds would increase variance in evaluation. We used minimum Redundancy - Maximum Relevance (mRMR) feature selection method [25] which is designed to find the smallest relevant subset of features (maximum relevance) while preventing highly correlated features to be part of this subset (minimum redundancy). All the 107 features kept for analysis are describing tissue structure, no nucleus morphology features (area, circularity) were included (See **Tissue Analyser** section in **Methods** for an in-depth description of the features).

To know how many features to keep we calculated the average balanced accuracy for each training keeping $N \in [1, 107]$ features. We kept $N_{\text{best}} = 19$ features which corresponded to the best mean balanced accuracy across all runs. Even if the number of selected features N_{best} is the same for each run in the case of mRMR, the selection of features can vary. Following **Eq. 2**, we identify the most representative features by selecting those with the highest occurrence counts c_{f_k} across the selected feature sets in each cross-validation run. In the event of ties, we favor features with higher rankings in their respective sets. To do so we first record the position of the feature in its set, its rank, and calculate its pre-score as $10^{N_{best}-rank}$, so a feature ranked first would have the highest pre-score. Then we add all the pre-scores for each cross-validation groups for a given feature and take the log of this sum to obtain the final score s_{f_k}.

$$A = \text{concat}\{fv_{split_1}, fv_{split_2}, ..., fv_{split_L}\} \tag{2}$$

$$c_{f_k} = \sum_{l=1}^{L} \sum_{n=1}^{N_b} \delta(A_{l,n}, f_k)$$

$$s_{f_k} = \log(\sum_{l=1}^{L} 10^{N_{best} - rank(f_k)_{split_l}})$$

where $fv_{split_l} = [f_1, f_2, ...f_{N_b}]_{split_l}$ the vector of N_{best} selected features from split $l \in [1, L]$ of the cross-validation, $rank(f_k)_{split_l}$ the rank of feature f_k selected from split $l \in [1, L]$ and δ the Kronecker delta.

Notably, mRMR prevents redundancy in feature selection, but aggregation of features from all cross-validation runs can include highly correlated features.

Results

Nuclei segmentation and classification evaluation

Accurate segmentation and classification of nuclei is necessary, to ensure the quality of downstream analyses involving features based on cell nuclei. We used Panoptic Quality metric [14] to evaluate segmentation, as it has been shown that it is better suited for evaluation of instance segmentation than DICE2 (aggregation of DICE score for each instance) [26] or aggregated Jaccard Index (AJI) metrics [27]. Indeed DICE2 is oversensitive to small changes in the prediction and AJI is over-sensitive to failed detections as shown in [14]. Following its definition, Panoptic Quality also assess detection performance.

We compare our cell nuclei segmentation and classification model to CellViT, current state of art model for segmentation of cells in H&E-stained WSIs [28]. CellViT was originally trained on Pannuke dataset [15] a commonly used and recognized dataset for panoptic segmentation on H&E stained tissues. Pannuke contains diverse types of cancer but only few skin cancer images, and without distinction of the type of skin cancer. Additionally, the dataset does not include granulocytes and plasma cells and the Pannuke-pretrained CellViT is not able to detect them.

We trained both models on NucSeg dataset including 5,968 patches of size 540x540 pixels with 70% overlap. Our validation set contained 848 patches of size 540x540 pixels with 70% overlap (see **Method** section for more details on training dataset and training procedure). No tiles were overlapping between the train and validation sets. We compared segmentation maps from SCC Hovernet and CellViT models applied to the same validation set in **Fig. 4a**. The segmentation and classification results of each cell class are compared one by one.

SCC Hovernet outperforms CellViT$_{256}$ in the detection and segmentation tasks for all cell classes. It also performs better or same on all cells nuclei classes except stromal cells, when compared to a larger CellViT model, CellViT-SAM-B trained on NucSeg. The average Panoptic Quality for all classes, mPQ of the SCC Hovernet is also higher compared to the CellViT-SAM-B trained on NucSeg. CellViT-SAM-B was too large to be trained on our 80GB A100 NVIDIA GPU directly so we used mixed-precision during training to fit it on one GPU (as no option to split the model on different GPUs was implemented). Mixed-precision has only limited impacts on models performance has shown in [29].

We also compared the 3 models on the classification task, considering only detected and paired cells (prediction and groundtruth). We evaluate F1 score for each class (**Fig. 4a**). CellViT-SAM-B is the best classifier for 3 classes and SCC Hovernet for the 2 others, average of F1 over all classes is 0.825 for CellViT$_{256}$, 0.846 for CellViT-SAM-B and 0.832 for SCC Hovernet. A confusion matrix of SCC Hovernet classification on the validation set is shown in **Fig. 4b**.

Overall, SCC Hovernet slightly outperforms CellViT-SAM-B in segmentation and detection tasks but slightly under-performs CellViT-SAM-B in solely classification task. Nevertheless SCC Hovernet has 3 times less parameters [30] and, with its convolutional neural network (CNN) architecture, is lighter than CellViT-SAM-B, which make it easier to use for training and inference.

To additionally validate Histo-Miner, we performed immunohistochemistry (IHC) for Myeloperoxidase (Mypo; granulocytes), CD3 (lymphocytes), CD79a (plasma cells), CD10 (stromal cells) and p40 (tumor cells), as well as H&E staining on adjacent slides from the same tissue blocks. In 6 different cSCC tumors we selected 7 to 11 representative ROIs (750µm x 750µm) on H&E-stained image and predicted the number of cells of each type using Histo-Miner (**Fig. 4c**). The same regions in IHC images were classified by a board certified dermatopathologist into 4 groups (- , +, ++, +++) based on the level of IHC positivity. Comparing the number of predicted cells across IHC positivity groups, we observed a significant association of the number of predicted cells per cell type and IHC positivity of the appropriate marker (**Fig. 4d**).

Tumor segmentation evaluation

To evaluate the performance of SCC Segmenter model, we used intersection over union (IoU) between the predicted tumor regions and the groundtruth of tumor regions. The model was trained on randomly chosen 115 slides of TumSeg dataset (see **Method** section for detailed description of training procedure). The validation set contained 29 slides of the dataset. A test set composed of 32 slides from 25 patients (patients not present in the training and validation sets) was used for evaluation. No data selection was performed to generate the test set. This test set is publicly available (see **Data availability** section). The model achieves on the test set: segmented foreground intersection over union $\mathrm{IoU}_{\mathrm{tissue}} = 0.852$, mean intersection over union $\mathrm{mIoU} = 0.907$, accuracy of foreground segmentation $\mathrm{Acc}_{\mathrm{tissue}} = 0.892$, and average accuracy $\mathrm{mAcc} = 0.969$.

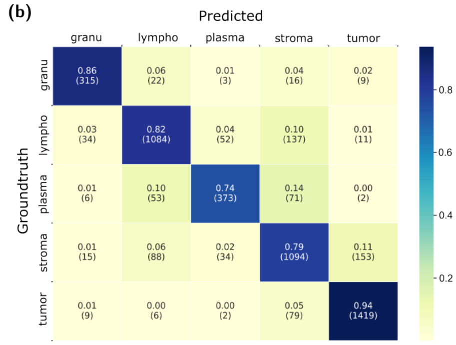
Application of Histo-Miner to predict response of cSCC patients to immunotherapy

To demonstrate the usability of Histo-Miner in a clinical scenario, we applied it to predict outcomes of patients treated with immune checkpoint inhibitors using anti-PD1 antibodies. Immunotherapy is the most effective treatment for advanced cSCC patients, but half of all patients do not respond and reliable predictive parameters do not yet exist.

The CPI dataset consists of 45 WSIs from cSCC skin cancer of 6 medical centers before they received immune checkpoint inhibitors treatment. Each WSI is taken from

(a)

		Methods		
		CellViT$_{256}$	CellViT-SAM-B	SCC Hovernet
Number of parameters		46.8M	146.1M	52.2M
Class PQ	Granulocytes	0.354	**0.464**	**0.464**
	Lymphocytes	0.524	0.521	**0.578**
	Plasma	0.460	0.502	**0.565**
	Stroma	0.582	**0.622**	0.586
	Tumor	0.603	**0.650**	**0.650**
	mPQ	0.505	0.552	**0.569**
Class F1	Granulocytes	0.825	0.844	**0.899**
	Lymphocytes	0.846	**0.853**	0.803
	Plasma	0.760	0.772	**0.818**
	Stroma	0.797	**0.824**	0.743
	Tumor	0.922	**0.935**	0.899
	meanF1	0.825	**0.846**	0.832



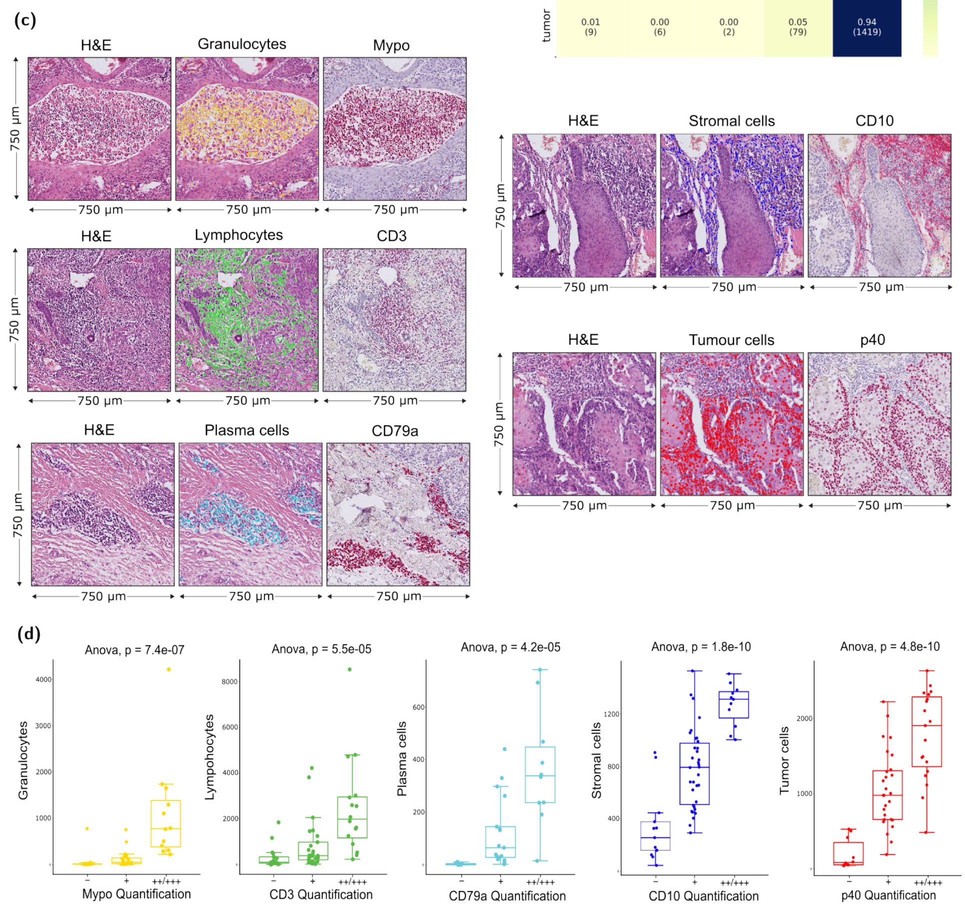


Fig. 4: Validation of SCC Hovernet. **(a)** Panoptic Quality for each cell class of SCC Hovernet, light CellViT$_{256}$ and heavy CellViT-SAM-B, all trained on NucSeg. mPQ is the average of PQ for all classes. SCC Hovernet outperforms CellViT$_{256}$ for all classes and outperforms CellViT-SAM-B general mPQ performance. CellViT-SAM-B outperforms SCC Hovernet on general classification performance. Taking into account segmentation, detection, classification tasks and weight of the models, SCC Hovernet is the preferred option. **(b)** Confusion matrix from SCC Hovernet prediction on the validation set. The most representative class, tumor cells, is accurately classified 94% of the time. The worst classification accuracy is of plasma cells: 74%. **(c)** Examples of validation via immunohistochemistry showing H&E and cell-type predictions (left and middle column) and staining for cell type markers of the same are in a consecutive section (right column) Mypo=Myeloperoxidase. **(d)** Comparison of manual cell type quantification in immunohistochemistry slides (x-axis; Mypo = Myeloperoxidase) and computationally predicted cells (y-axis) in H&E slides.

a different patient. 28 of them were classified as responders (CR and PR states) and 17 patients were classified as non-responders (SD and PD states) as displayed in **Fig. 5a**. We performed 3 fold cross-validation with XGBoost classifier [24] on the CPI dataset to predict the class of the WSIs.

In each of the runs, we performed feature selection with mRMR, as described in **Methods** section, and ranked all features from most to least predictive. We then trained 107 XGBoost classifiers on the same training split, iteratively removing the least predictive feature each time, thus going from all 107 features down to the single most predictive feature. For each of these trainings we calculated balanced accuracy on the validation split. To know how many features to keep we calculated the average balanced accuracy for each training keeping $N \in [1, 107]$ features. We kept $N_{\text{best}} = 19$ features which corresponded to the best mean balanced accuracy across all runs. We obtained mean balanced accuracy $\bar{ba}_{N_{\text{best}}} = 0.767 \pm 0.057$ across 3 fold cross-validation runs with the mean area under ROC curve $A\bar{U}C_{N_{\text{best}}} = 0.755 \pm 0.091$ (**Fig. 5b**). Interestingly, adding the morphology features to the set of features did not improve the classification - with $N_{\text{best}_{\text{w/morph}}} = 305$ features the balanced accuracy reaches $\bar{ba}_{N_{\text{best}_{\text{w/morph}}}} = 0.754 \pm 0.127$.

Other feature selection methods were tested but led to a worse performing feature set. We tested Boruta [31] method which, unlike mRMR that finds the smallest relevant subset of features, aims at finding all features carrying information usable for prediction. With Boruta the number of selected features varied for each fold. Classification of CPI WSIs using Boruta selected feature had a balanced accuracy of $\bar{ba}_{N_{BORUTA}} = \bar{ba}_{[5,0,1]} = 0.711 \pm 0.011$ across the 3 folds. For one of the fold Boruta was enable to select features, and this for any setting of maximum depth of the random tree. In the best scenario presented here the maximum depth was set to 1, the random state was set to 0 and all other hyperparameters were kept as default values. We also ranked features following Mann-Whitney U test scores [32]. We then trained 107 XGBoost classifiers on the same training split, iteratively removing the least predictive feature each time, thus going from all 107 features down to the single most predictive feature (as we did for mRMR). The best balanced accuracy was in this case $\bar{ba}_{N_{MannWhitney}} = 0.715 \pm 0.069$ with $N_{MannWhitney} = 6$.

Using mRMR feature selection, the 4 most representative features in the decreasing ranking order are: the percentage of lymphocytes among all cells in tumor vicinity, the ratio between the number of granulocytes and lymphocytes in tumor vicinity, the percentage of lymphocytes among all cells in tumor region, and the mean distances between granulocytes and plasma cells in tumor regions (**Fig. 5c)**. We discuss the interpretation of these features in the next section. The process behind feature selection is described in the **Method** section.

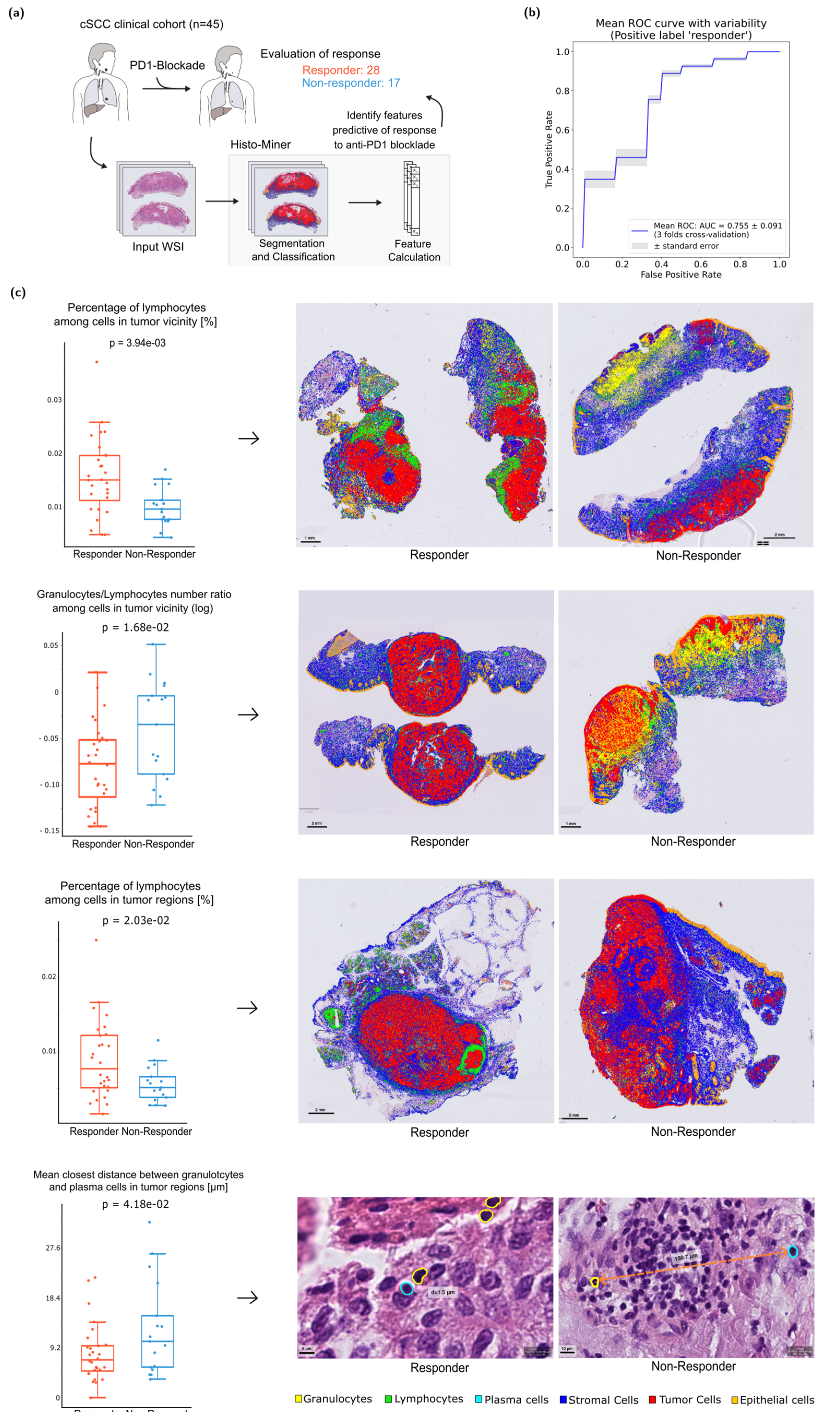


Fig. 5: Histo-Miner tested in a clinical scenario: prediction of CPI treatment response. **(a)** To test the clinical utility of Histo-Miner we assembled and processed WSIs from in total n=45 cSCC patients before checkpoint inhibition (CPI) therapy. **(b)** Mean ROC curve for the classifier keeping 19 best features and its standard error. The classifier cross-validation folds ROC curves as well as the standard deviation of the mean ROC curve are shown in **Supplementary Fig. 6**. **(c)** On the left, box plots of the 4 best features, p-value was calculated using Mann-Whitney U test. On the right, visualization of representative cases for each of the best features. For distance visualization we hide cell classes other than plasma cells and granulocytes.

Discussion

WSIs contain vast amounts of detailed information and are a rich resource for cancer diagnosis and research. Automated digital pathology methods offer possibilities to efficiently process WSIs, uncovering quantitative information as well as subtle patterns and features that may be inaccessible or indiscernible to human experts. Moreover, they enable automated cell type classification and detailed description of tissue composition and architecture. These approaches have rarely been applied to non-melanoma skin cancers like cSCC despite the fact that cSCC alone affects more than 1 Million individuals in the USA every year [12]. A major obstacle to development of automated methods for non-melanoma skin cancer slide analysis has been the high similarity between tumor cells and non-tumor skin cells. Here, we present the Histo-Miner pipeline which provides single-cell insights into skin tumor WSIs. Our methods not only generate precise and complete information on the patient biopsy composition, we also demonstrate how this information allows to predict patient outcomes and give interpretable insights into the determinants of these outcomes. Such techniques can therefore lead to discovery of previously unknown diagnostic biomarkers, leading to improved cancer detection, diagnosis, and personalized treatment strategies.

Histo-Miner performs segmentation and classification of cell nuclei using a CNN, SCC Hovernet, trained on our NucSeg dataset, as well as tumor segmentation using a vision transformer, SCC Segmenter, trained on our TumSeg dataset. We compare the performance of segmentation and classification of segmented cells task to state of the art methods, such as CellViT, and show improved Panoptic Quality of our approach. Based on classification and segmentation results, Histo-Miner creates feature vectors describing tissue composition and organization. A Histo-Miner user can choose which features to calculate to best describe their WSIs. Additionally, Histo-Miner models can be trained and adapted to other cancer types.

Given the large size of WSIs, most commonly prediction tasks in digital pathology are performed via multiple instance learning (MIL) approaches [33,34] . Via WSI tesselation and patch embedding, such approaches allow to train the model and perform the prediction on the entire WSIs directly. While it is convenient to train the prediction model end-to-end using patient-level labels that are typically available in the patient records, MIL methods offer only limited interpretability. MIL paired with attention or gradient-based mechanisms [35,36,37,38,39] allow to disentangle which WSI regions are the most predictive, however the content of these highly-predictive regions is typically assessed in a qualitative manner. In contrast, Histo-Miner represents a feature-based approach that allows for a quantitative and systematic identification of tissue characteristics that are important for prediction. Each step of the Histo-Miner pipeline from segmentation to feature selection can be visualized allowing for inspection and interpretation of the prediction results.

We demonstrate the applicability of our pipeline on a cohort of 45 patients treated with immune checkpoint inhibition, which is the major treatment modality for patients with advanced cSCC [13]. Even though our patient cohort was relatively small and collected across 6 medical centers, our pipeline was able to accurately predict CPI response in cross-validation experiments, highlighting its potential for clinical use cases. In addition, our feature-based approach points to the features driving the classification and thus to potential insights into the tumor biology. We identified four features the most predictive of CPI response, which included a higher percentage of lymphocytes within and in the vicinity of tumor regions in responders than non-responders. This result is in line with previous studies showing lymphocyte infiltration as a predictive marker of CPI response in cSCC [40]. Interestingly, we also observed a higher granulocyte to lymphocyte ratio and a higher distance between granulocytes and plasma cells in non-responders than responders. High neutrophil-to-lymphocyte ratio in the blood of cSCC patients has been associated with worse prognosis in general and suggested to correlated with decreased CPI response [41,42], but their ratio in cSCC tissues has not been described before. Similarly, very little is known about plasma cells in cSCC tumors, especially in conjunction with granulocytes. In breast cancer however, patients not responding to CPI showed high degree of granulocytes as measured by CD15 [43] positivity. In mouse models of pancreatic and squamous cell lung cancer, depletion of neutrophilic granulocytes led to reduced tumor growth [44,45]. While the interplay between granulocytes and T and plasma cells in cSCC requires experimental validation in future studies, these findings indicate that both cells of the innate (granulocytes) and the adaptive arm (T and plasma cells) of the immune system may play opposing roles in modulating CPI response in cSCC. They thereby highlight one of the strengths of our pipeline, which - in contrast to more coarse approaches - classifies immune cells into 3 subcategories and provides quantitative as well as spatial information about tumor microenvironment to identify relevant factors. It can thereby help to generate novel hypotheses for follow-up investigations.

A limitation of our approach is the requirement of expert-annotated samples (tumor regions, nucleus boundaries, and cell classes) that serve as ground truth for training and testing. It may contain human errors and introduce subjective biases that the models ultimately learn to replicate. In addition, tumor cell recognition using an additional tumor region segmentation model might lead to individual cancer cells outside of the predicted tumor regions as well as non-neoplastic epithelial cells within the predicted tumor regions being misclassified.

A difficulty we faced is the distinction between non-neoplastic epithelial and tumor cell nuclei. For the published pre-trained models we tested, morphologies of those two cell types were indistinguishable in a skin H&E-stained sample if considered in isolation. Context, such as e.g. cell localization and neighborhood, allow to distinguish them from one another. Here the combination of two deep learning models - one for cell segmentation and classification and one for tumor region segmentation - allowed us to discriminate between the two cell classes. Further studies should focus on designing a cell classifier that is able to distinguish all 6 cell classes (granulocytes, lymphocytes, plasma cells, stromal cells, tumor cells, non-neoplastic epithelial) without using an additional tumor region classifier. Such model should incorporate a broader context including surrounding cells in a patch in the prediction process.

Data availability

The checkpoints of SCC Hovernet and SCC Segmenter are available on Zenodo repository link.
TumSeg and NucSeg datasets are also available on Zenodo repository link.
The CPI image classification dataset utilized as a use-case for Histo-Miner is also available on Zenodo repository link.
Finally, TumSeg test set, SCC Segmenter inference on test set, the list of all features from Tissue Analyser and ranking of features after cross-validation with mRMR selection are available on Zenodo repository link.
All the WSIs of these datasets were anonymized and cannot be used for commercial use.

Code availability

The Histo-Miner code is self-sufficient and user friendly as it contains explanations on how to use it that are enough for user without computer science background. In addition to the core of the code, two github submodules are used, one from MMSegmentation zoo (segmenter model) and one from Hovernet network implementation. Thanks to this approach , the Histo-Miner repository is not dependent on the changes that are operated on the original repositories. Additionally, the code is much simpler if a user wants to edit it as it keeps only functions and scripts edited in the context of Histo-Miner.
The github repository is available here.

Supporting Information

S1 Material.
This supporting document contains all supplementary tables and figures cited in the main text. It includes the following sections:

- Poor precision in tumor and healthy epithelial detection from state-of-the-art pre-trained models
- SCC Hovernet loss functions
- Hyperparameters grid search for SCC Segmenter
- Probability of distance overestimation
- List of all features from Tissue Analyser
- Ranking of features after cross-validation with mRMR selection
- ROC curves of the best-feature classifier across CV folds
- Stain variability for the different cohorts of TumSeg dataset

(PDF)

Acknowledgments

The authors would like to thank Christian Knetschowsky for the annotations of segmentation maps of NucSeg dataset, and to thank Alfred Kirsch for the detailed corrections and extensive verifications brought to the calculation of the probability of closest distance overestimation.

Funding

This work was supported by the Ministry for Culture and Science (MKW) of the State of North Rhine-Westphalia [grant number 311-8.03.03.02-147635]. LS and KB were supported by the North Rhine-Westphalia return program (311-8.03.03.02-147635) and hosted by the Center for Molecular Medicine Cologne. AF was partly funded by the Deutsche Krebshilfe through a Mildred Scheel Foundation Grant (grant number 70113307). CL was partly funded through the collaborative research center grant on small cell lung cancer (CRC1399, project ID 413326622) and a project grant (grant ID BR 6949) by the German Research Foundation (DFG). JB receives funding through the collaborative research center grant on small cell lung cancer (CRC1399, project ID 413326622) and on predictability in evolution (CRC1310, project ID 325931972) by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), a Mildred Scheel Nachwuchszentrum Grant 70113307 and project funding (grant ID 70116929) by the German Cancer Aid (Deutsche Krebshilfe) and the CANTAR network (NW21-062B) funded through the program "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of Northrhine Westphalia, Germany. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. All other authors have no funding to declare.

Competing Interests

I have read the journal's policy and the authors of this manuscript have the following competing interests: JB has received research funding from Bayer AG and travel expenses from Merck KG & Bicycle Therapeutics and serves as a consultant for Bicycle Therapeutics outside of the presented work. These entities had no role in study design, data collection and analysis, publication decisions, or manuscript preparation. ODP received personal honoraria from Merck Sharp & Dohme and Almirall, received travel support from Kyowa Kirin, Pierre Fabre, Sanofi and Sun Pharma outside of the presented work and is member of the advisory board of Bristol Myers Squibb and Sanofi. These entities had no role in study design, data collection and analysis, publication decisions, or manuscript preparation. All other authors declare no conflicts of interests.

Ethics approval and consent to participate

The study was performed in agreement with the Declaration of Helsinki Institutional Review Board of the University Hospital Bonn (vote number 187/16), Ethics committee of the University Hospital Cologne (votenumbers 21–1500, 20–1082 and 22–1330-retro) and institutional review board of the TU Munich (vote number 2024–363-S-CB - 1). Need for informed consent was waived for this retrospective analysis using anonymized data.

Author Contributions

Conceptualization: Lucas Sancéré, Carina Lorenz, Johannes Brägelmann, Katarzyna Bozek.
Data Curation: Lucas Sancéré, Carina Lorenz, Doris Helbig, Johannes Brägelmann.
Formal Analysis: Lucas Sancéré.
Funding Acquisition: Johannes Brägelmann, Katarzyna Bozek.
Investigation: Lucas Sancéré, Carina Lorenz.
Methodology: Lucas Sancéré.

Project Administration: Lucas Sancéré, Doris Helbig, Johannes Brägelmann, Katarzyna Bozek.
Resources: Doris Helbig, Oana-Diana Persa, Sonja Dengler, Alexander Kreuter, Martim Laimer, Roland Lang, Anne Fröhlich, Jennifer Landsberg
Software: Lucas Sancéré.
Supervision: Johannes Brägelmann, Katarzyna Bozek.
Validation: Lucas Sancéré.
Visualization: Lucas Sancéré, Carina Lorenz.
Writing – Original Draft Preparation: Lucas Sancéré, Johannes Brägelmann, Katarzyna Bozek.
Writing – Review & Editing: Lucas Sancéré, Carina Lorenz, Johannes Brägelmann, Katarzyna Bozek.

References

1. Wittekind, D. H. Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. value and limitations of the hematoxylin-eosin stain. *Biotechnic & Histochemistry* **78**, 261 – 270 (2003).
2. van Rijthoven, M., Balkenhol, M. C. A., Silina, K., van der Laak, J. & Ciompi, F. Hooknet: multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. *Medical image analysis* **68**, 101890 (2020).
3. Shui, Z. *et al.* Unleashing the power of prompt-driven nucleus instance segmentation. *ArXiv* **abs/2311.15939** (2023).
4. Zheng, Y. *et al.* A graph-transformer for whole slide image classification. *IEEE transactions on medical imaging* **41**, 3003 – 3015 (2022).
5. Nahhas, O. S. M. E. *et al.* Regression-based deep-learning predicts molecular biomarkers from pathology slides. *Nature Communications* **15** (2023).
6. Pantanowitz, L. X. *et al.* An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *The Lancet. Digital health* **2 8**, e407–e416 (2020).
7. Viswanathan, V. S., Toro, P., Corredor, G., Mukhopadhyay, S. & Madabhushi, A. The state of the art for artificial intelligence in lung digital pathology. *The Journal of Pathology* **257**, 413 – 429 (2022).
8. Cardoso, M. J. *et al.* Monai: An open-source framework for deep learning in healthcare. *ArXiv* **abs/2211.02701** (2022).
9. Lu, M. Y. *et al.* Visual language pretrained multiple instance zero-shot transfer for histopathology images. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 19764–19775 (2023).
10. Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nature medicine* (2024).
11. Saldanha, O. L. *et al.* Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology. *NPJ Precision Oncology* **7** (2023).
12. JY, H., Y, H. & ML, R. Squamous cell skin cancer. *StatPearls. Treasure Island (FL): StatPearls Publishing* (2024).
13. Winge, M. C. G. *et al.* Advances in cutaneous squamous cell carcinoma. *Nature Reviews Cancer* **23**, 430–449 (2023).
14. Graham, S. *et al.* Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis* **58**, 101563 (2018).
15. Gamper, J. *et al.* Pannuke dataset extension, insights and baselines. *ArXiv* **abs/2003.10778** (2020).
16. Strudel, R., Pinel, R. G., Laptev, I. & Schmid, C. Segmenter: Transformer for semantic segmentation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 7242–7252 (2021).
17. Zhou, B. *et al.* Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision* **127**, 302 – 321 (2019).

18. Mottaghi, R. *et al.* The role of context for object detection and semantic segmentation in the wild. *2014 IEEE Conference on Computer Vision and Pattern Recognition* 891–898 (2014).
19. Pisula, J. I. *et al.* Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma. *medRxiv* (2024).
20. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014).
21. Steiner, A. *et al.* How to train your vit? data, augmentation, and regularization in vision transformers. *Trans. Mach. Learn. Res.* **2022** (2021).
22. Contributors, M. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation (2020).
23. Hendry, S. *et al.* Assessing tumor-infiltrating lymphocytes in solid tumors: A practical review for pathologists and proposal for a standardized method from the international immuno-oncology biomarkers working group: Part 2: Tils in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothe. *Advances in anatomic pathology* **24 6**, 311–335 (2017).
24. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, 1226–1238 (2005).
25. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 11, 785–794 (ACM, 2016).
26. Vu, Q. D. *et al.* Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in Bioengineering and Biotechnology* **7** (2018).
27. Kumar, N. *et al.* A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging* **36**, 1550–1560 (2017).
28. Hörst, F. *et al.* Cellvit: Vision transformers for precise cell segmentation and classification. *Medical image analysis* **94**, 103143 (2023).
29. Dörrich, M., Fan, M. & Kist, A. M. Impact of mixed precision techniques on training and inference efficiency of deep neural networks. *IEEE Access* **11**, 57627–57634 (2023).
30. Zhao, T., Fu, C., Tian, Y., Song, W. & Sham, C.-W. GSN-HVNET: A lightweight, multi-task deep learning framework for nuclei segmentation and classification. *Bioengineering (Basel)* **10** (2023).
31. Kursa, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *Journal of Statistical Software* **36**, 1–13 (2010).
32. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin* **1**, 80–83 (1945).
33. Xu, Y. *et al.* Deep learning of feature representation with multiple instance learning for medical image analysis. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2014).
34. Shao, Z., Dai, L., Wang, Y., Wang, H. & Zhang, Y. AugDiff: Diffusion-based feature augmentation for multiple instance learning in whole slide image. *IEEE Trans. Artif. Intell.* **5**, 6617–6628 (2024).
35. Pirovano, A., Heuberger, H., Berlemont, S., Ladjal, S. & Bloch, I. Improving interpretability for computer-aided diagnosis tools on whole slide imaging with multiple instance learning and gradient-based explanations. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, Lecture notes in computer science, 43–53 (Springer International Publishing, Cham, 2020).
36. Lu, M. Y. *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
37. Li, B., Li, Y. & Eliceiri, K. W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2021).
38. Shao, Z. *et al.* Transmil: Transformer based correlated multiple instance learning for whole slide image classication. *ArXiv* **abs/2106.00908** (2021).
39. Pisula, J. & Bozek, K. Fine-tuning a multiple instance learning feature extractor with masked context modelling and knowledge distillation. *ArXiv* **abs/2403.05325** (2024).

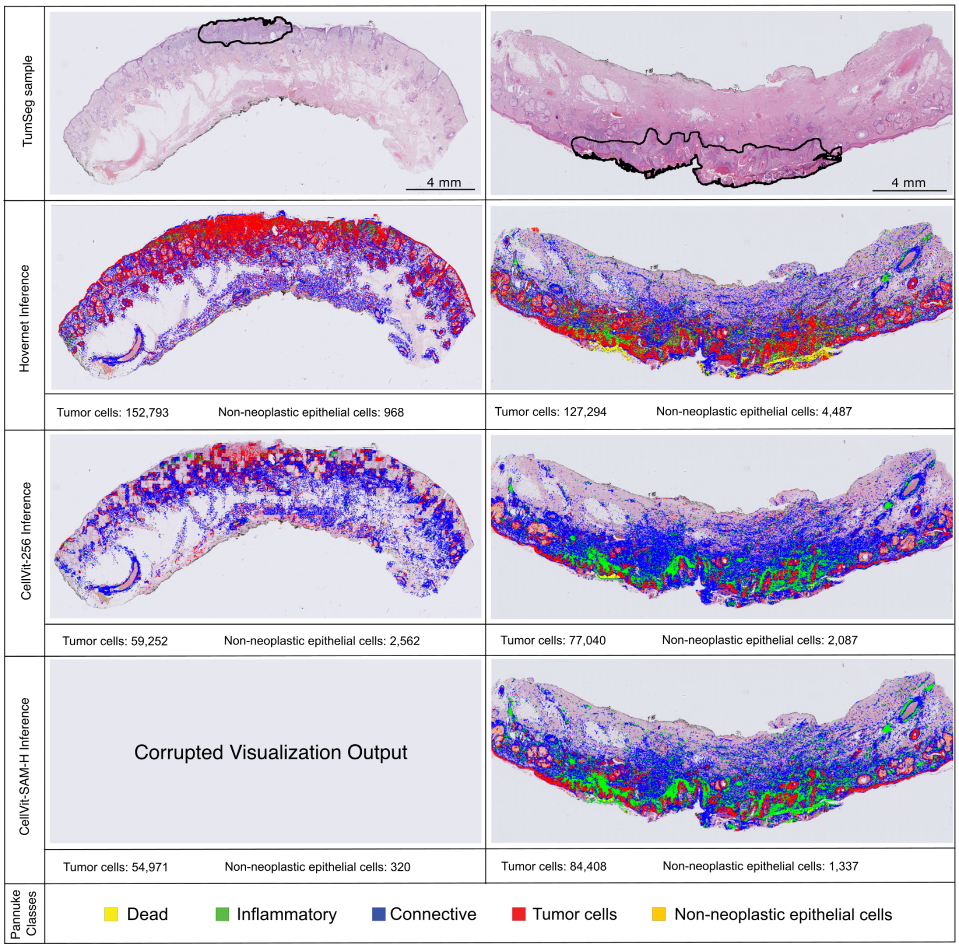
40. Ferrarotto, R. *et al.* Pilot phase ii trial of neoadjuvant immunotherapy in locoregionally advanced, resectable cutaneous squamous cell carcinoma of the head and neck. *Clinical Cancer Research* **27**, 4557–4565 (2021).
41. Seretis, K., Sfaelos, K., Boptsi, E., Gaitanis, G. & Bassukas, I. D. The neutrophil-to-lymphocyte ratio as a biomarker in cutaneous oncology: A systematic review of evidence beyond malignant melanoma. *Cancers (Basel)* **16**, 1044 (2024).
42. Strippoli, S. *et al.* Cemiplimab in an elderly frail population of patients with locally advanced or metastatic cutaneous squamous cell carcinoma: A single-center real-life experience from italy. *Front. Oncol.* **11**, 686308 (2021).
43. Wang, X. Q. *et al.* Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* **621**, 868–876 (2023).
44. Mollaoglu, G. *et al.* The lineage-defining transcription factors sox2 and nkx2-1 determine lung cancer cell fate and shape the tumor immune microenvironment. *Immunity* **49**, 764–779.e9 (2018).
45. Ng, M. S. F. *et al.* Deterministic reprogramming of neutrophils within tumors. *Science* **383**, eadf6493 (2024).

Supplementary Data

Table of Contents

Poor precision in tumor and healthy epithelial detection from state of the art pretrained models

We used Hovernet [1], CellViT-256 and CellVit-SAM-H [2] models pretrained on Pannuke dataset [3] to apply inference on 2 TumSeg samples where tumor regions are annotated by 2 experts. We show in **Supplementary Fig. 1** that the models are not able to recognize healthy epithelial, and classify most of the epithelial cells outside tumor regions as tumor cells. On the other hand, the tumor region segmentation performed by SCC Segmenter, guiding cell classification in Histo-Miner, has an average accuracy mAcc = 0.969 and mean intersection over union mIoU = 0.907 on our test set. All cells previously classified as tumor by SCC Hovernet that are outside tumor regions segmented by SCC Segmenter will be re-classified as non-neoplastic epithelial cells during Histo-Miner processing.



Suppl. Fig 1: Inference of pretrained Hovernet, CellViT-256 and CellVit-SAM-H on TumSeg samples. Application of state of the art models pretrained on Pannuke dataset and Hovernet, to two slides from TumSeg dataset for which tumor regions were annotated by two experts (marked in black). We display numbers of tumor and non-neoplastic epithelial cells detected by the different models. All inferences lead to prediction of tumor cells even outside tumor regions and within the healthy skin. Additionally, for some WSIs CellVit-SAM-H inference failed to generate an output visual, an issue described to sometimes occur with this model, probably when the number of cells to segment is too high.

SCC Hovernet loss functions

The overall training loss function L is a combination of loss functions as follows:

$$L = \underbrace{\lambda_{SCC_a} L_a + \lambda_{SCC_b} L_b}_{\text{HoVer Branch}} + \underbrace{\lambda_{SCC_c} L_c + \lambda_{SCC_d} L_d}_{\text{Nuclear Pixels Branch}} + \underbrace{\lambda_{SCC_e} L_e + \lambda_{SCC_f} L_f}_{\text{Nuclear Classification Branch}}$$

where L_a and L_b represent the regression loss with respect to the output of the HoVer branch, L_c and L_d represent the loss with respect to the output of the NP branch (Nuclear Pixels branch, corresponding to the Nuclear Segmentation), L_e and L_f represents the loss with respect to the output at the NC branch (Nuclear Classification branch), as already defined in [1]. Here all $\lambda_{SCC} = 1$.

$$L_a = \frac{1}{N} \sum_{i=1}^{N} \big(p_i(I; \mathbf{w_0}, \mathbf{w_1}) - \varGamma_i(I)\big)^2 \tag{1}$$

$$\begin{aligned} L_b = & \frac{1}{m} \sum_{i \in M} \Big(\nabla_x\big(p_{i,x}(I; \mathbf{w_0}, \mathbf{w_1})\big) - \nabla_x\big(\varGamma_{i,x}(I)\big)\Big)^2 \\ & + \frac{1}{m} \sum_{i \in M} \Big(\nabla_y\big(p_{i,y}(I; \mathbf{w_0}, \mathbf{w_1})\big) - \nabla_y\big(\varGamma_{i,y}(I)\big)\Big)^2 \end{aligned} \tag{2}$$

$$L_c = L_e = CE = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} X_{i,k}(I) \log Y_{i,k}(I) \tag{3}$$

$$L_d = L_f = DICE = 1 - \frac{2 \times \sum_{i=1}^{N} \big(Y_i(l) \times X_i(l)\big) + \epsilon}{\sum_{i=1}^{N} \big(Y_i(l)\big) + \sum_{i=1}^{N} \big(X_i(l)\big) + \epsilon} \tag{4}$$

where I input Image containing N pixels, $p_i(I; \mathbf{w_0}, \mathbf{w_1})$ regression output of HoVer branch at pixel i, w_0 and w_1 2 sets of weights. $\varGamma_i$ defines the groundtruth of the horizontal and vertical distances of nuclear pixels to their corresponding centers of mass, the horizontal and vertical components of this map are denoted $\varGamma_{i,x}$ and $\varGamma_{i,y}$ respectively (see [1] for visualization and definitions). m denotes total number of nuclear pixels within the image and M denotes the set containing all nuclear pixels. ∇_x and ∇_y denote the gradient in the horizontal x and vertical y directions respectively. Finally, X denotes the branch groundtruth, Y the branch prediction, K is the number of classes and ϵ is a smoothness constant that was set to 10^{-3}.

Hyperparameters grid search for SCC Segmenter

On **Table.1** the different hyperparameters tested are displayed. $cat-max-ratio$ corresponds to the max area ratio that could be occupied by single category for a given crop of the input image. $img-scale$ corresponds to the resizing of the image before cropping. These resizing can randomly be modified by a factor contained in [0.5, 2], following the data augmentation pipeline from the semantic segmentation library MMSegmentation [5].

Not all combinations were fully tested, some were aborted if the training loss did not decrease in the firsts epochs or if several hyperparameters in the set were already

Hyperparameters	Value Range
cat-max-ratio	[0.75, 0.85, 0.90 , 0.95, 1.0]
img-scale	[(2560, 640), (5120, 1280), (5120, 5120)]
samples-per-gpu	[4, 8]
pre-training-dataset	[ImageNet, Thomas2021 dataset [4]]

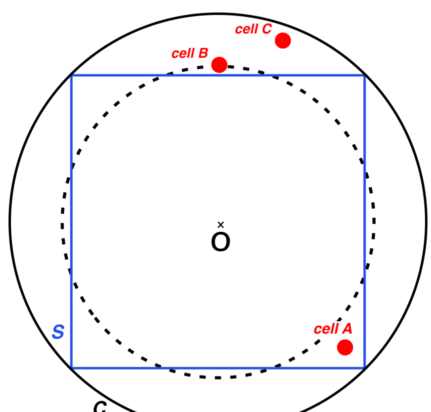
Suppl. Table 1: Hyperparameters search for SCC Segmenter

poorly performing in previous trainings. .

Set kept: {cat-max-ratio: 0.75, img-scale: (2560, 640), samples-per-gpu: 4, pre-training-dataset: ImageNet}

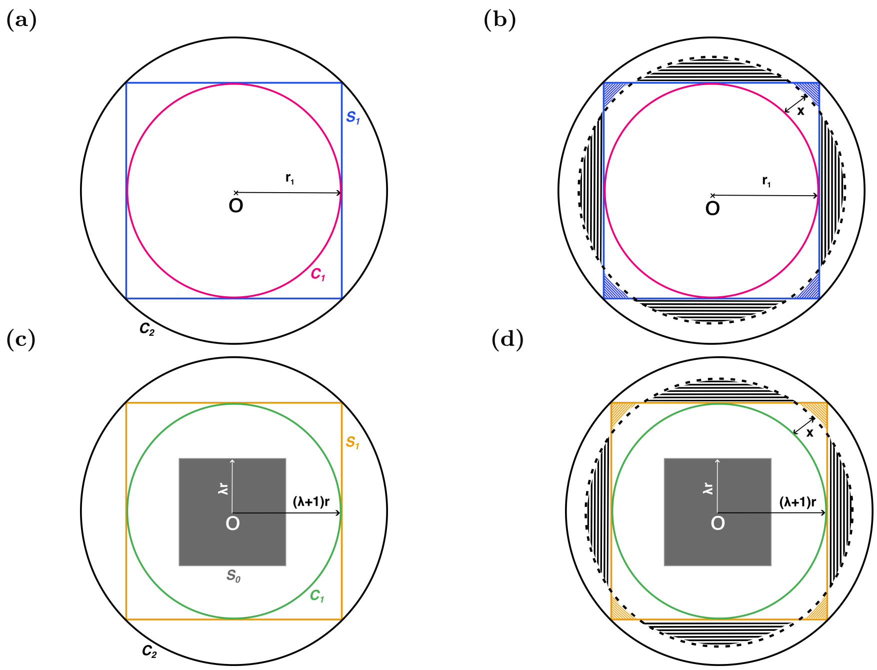
Probability of distance overestimation

In this paper, we make use of a search algorithm to estimate the average distance of the closest cell of a given class X - source class - to the closest cell of a given class Y - target class - inside the tumor regions. The principle is to define a search area (based on the tumor bounding box dimensions) centered around the nucleus of the source cell class and to check if any cells of the target class are inside. If it is the case, we calculate all the distances and keep the smallest one. If there is no cell inside the first search area, we increase the size of this area until we find at least one cell inside. Nevertheless, the closest distance found with this method is not always the actual closest distance, but sometimes it could be an overestimation. As shown in **Supplementary Fig. 2**, sometimes a cell outside the search area can be the closest but won't be taken into account.



Suppl. Fig 2: Example of distance overestimation with search area being a square. Here we have 1 cell in the squared search area S and 2 cells in $(C \setminus S)$. While calculating the minimum distance of cells from the center O, if we only consider cells in S, we will take $cellA$ as being the closest cell. In reality the closest cell is $cellB$. The distance calculated will be the distance between $cellA$ and the center, and will then be higher than the real closest distance. Nevertheless, the overestimation is bounded as expressed in **Eq. 22** with $a = 1$ as we are here in the case of a squared search area.

In this supplement we show that even if such a case can occur, it is very unlikely as the number of cells increases. We simplify the problem by taking a squared search area instead of the rectangle search area that we can have in practice. The search area is based on the tumor's bounding box shape, and tumors bounding box have low aspect ratios, so taking a square search area makes it a good approximation for evaluation of error.



Suppl. Fig 3: Probability of having distance overestimation representations, with search area being a square. **(a)** During the first search, if the search box is a square of side lenght 2r, we could have a distance calculation error if a cell is in the blue square S_1 without being in the pink circle C_1 (then in $S_1 \setminus C_1$), and another cell is in black circle C_2 without being in in the blue square S_1 (then in $C_2 \setminus S_1$). Then the cell in $C_2 \setminus S_1$ could be missed from closest distance calculation and the cell in $S_1 \setminus C_1$ considered as the closest cell. In other cases, such as having at least one cell in S_1, no distance overestimation could be made. **(b)** To be more precise, any cell in the hatched blue area inside $S_1 \setminus C_1$ monitored by x would lead for sure to an error if another cell is in the hatched black area inside $C_2 \setminus S_1$ monitored by x . For now we focused on the probability of overestimation in the case that at least one cell is found during the first search. If no cell is found, then the calculation differs for the further searching steps as visualized in **(c)** & **(d)**. In such cases, a specific new area cannot contain any cells, as none were found during prior searches, impacting the full calculation.

We consider N cells uniformly and independently distributed inside C_2 of radius r_2 and center O. We note S_1 the inscribed square in C_2 of side length $2r_1$. We here have $r_2 = \sqrt{2}r_1$. A representation of this model is shown in **Supplementary Fig. 3a**.

We are interested in computing the following conditional probability:

$$P(error_N) = P(B_N \mid A_N) \tag{5}$$

where:

$$A_N : (\text{At least one cell is in } S_1)$$

$$B_N : (\text{There is among the cells in } (C_2 \setminus S_1) \text{ a cell closer to the center } O \text{ than all cells in } S_1)$$

This probability corresponds of the probability of having an error in the distance calculation (overestimation) when at least one cell is found in the first search area C_1 of radius r_1 and center O. To compute this probability we first make use of the complementary event rule:

$$P(B_N \mid A_N) = 1 - P(\bar{B_N} \mid A_N) \tag{6}$$

where:

$$\bar{B_N} : (\text{All the cells in } (C_2 \setminus S_1) \text{ are further away to the center } O \text{ than all cells in } S_1)$$

To compute $P(\bar{B_N} \mid A_N)$, we introduce for $\forall n \in [[0, N]]$:

$$U_N(n) : (\text{Exactly } n \text{ cells are in } (C_2 \setminus S_1))$$

Using the law of total probabilities and then the definition of conditional probabilities:

$$\begin{aligned} P(\bar{B_N} \mid A_N) &= \sum_{n=0}^{N} P\big(\bar{B_N} \cap U_N(n) \mid A_N\big) \\ &= \sum_{n=0}^{N} \frac{P\big(\bar{B_N} \cap U_N(n) \cap A_N\big)}{P(A_N)} \end{aligned} \tag{7}$$

Now we note that $\forall n \in [[0, N-1]], U_N(n) \subset A_N$, so that $A_N \cap U_N(n) = U_N(n)$. We note moreover that if $n = N$, then no cells are is in the search area of S_1. So $U_N(N) \cap A_N = \emptyset$. Finally, noting that $U_N(0) \subset \bar{B_N}$ and using the definition of conditional probabilities, we have:

$$\begin{aligned} P(\bar{B_N}|A_N) &= \frac{P\big(U_N(0)\big)}{P(A_N)} + \sum_{n=1}^{N-1} \frac{P\big(\bar{B_N} \cap U_N(n)\big)}{P(A_N)} \\ &= \frac{1}{P(A_N)} \Big[P\big(U_N(0)\big) + \sum_{n=1}^{N-1} P\big(U_N(n)\big) P\big(\bar{B_N} \mid U_N(n)\big)\Big] \end{aligned} \tag{8}$$

As U_N is a binomial trial, we can calculate $P(U_N(n))$ as follows:

$$\begin{aligned} P\big(U_N(n)\big) &= \binom{N}{n} \Big(\frac{\text{area}(C_2 \setminus S_1)}{\text{area}(C_2)}\Big)^n \Big(1 - \frac{\text{area}(C_2 \setminus S_1)}{\text{area}(C_2)}\Big)^{N-n} \\ &= \binom{N}{n} \Big(1 - \frac{2}{\pi}\Big)^n \Big(\frac{2}{\pi}\Big)^{N-n} \end{aligned} \tag{9}$$

We then deduce that:

$$P\big(U_N(0)\big) = \Big(\frac{2}{\pi}\Big)^N \tag{10}$$

As A_N : (At least one cell is in S_1), $\bar{A_N}$: (No cell is in S_1) which is equivalent to $\bar{A_N}$: (All N cells are in $(C_2 \setminus S_1)$), then $\bar{A_N} = U_N(N)$ and using the complementary event rule:

$$\begin{aligned} P(A_N) &= 1 - P(\bar{A_N}) \\ &= 1 - P\big(U_N(N)\big) \\ &= 1 - \Big(1 - \frac{2}{\pi}\Big)^N \end{aligned} \tag{11}$$

We now set two new events introducing $d = r_1 + x$ as shown in **Supplementary Fig. 3b** :

$V_N(n, d)$: (There are $N - n$ cells in S_1 and the closest cell to the origin O is at a distance d from O)

And:

$\tilde{V_N}(n, d)$: (There are $N - n$ cells in S_1 and they are all located at a distance greater than d from O)

Then, using the law of total probability for continuous univariate distributions:

$$P\big(\bar{B_N} \mid U_N(n)\big) = \int_0^{\sqrt{2}r_1} P\big(\ \bar{B_N}\ \mid \big(U_N(n) \cap V_N(n, d)\big)\big)\ \rho_n(d)\ \mathrm{d}d \tag{12}$$

where $\rho_n(d)$ is the probability density function associated to the survival function $P\big(\tilde{V_N}(n, d)|U_N(n)\big)$:

$$\rho_n(d) = -\frac{\mathrm{d}P\big(\tilde{V_N}(n, d)|U_N(n)\big)}{\mathrm{d}d}$$

If $d \leq r_1$ then $P(\bar{B_N}|(U_N(n) \cap V_N(n, d))) = 1$, leading to:

$$\begin{aligned} P\big(\bar{B}_N \mid U_N(n)\big) &= \int_0^{r_1} \rho_n(d)\,\mathrm{d}d\ +\ \int_{r_1}^{\sqrt{2}r_1} P\big(\bar{B}_N \mid \big(U_N(n) \cap V_N(n, d)\big)\big)\,\rho_n(d)\,\mathrm{d}d \\ &= \Big[1 - P\big(\tilde{V}_N(n, r_1) \mid U_N(n)\big)\Big]\ +\ \int_{r_1}^{\sqrt{2}r_1} P\big(\bar{B}_N \mid \big(U_N(n) \cap V_N(n, d)\big)\big)\,\rho_n(d)\,\mathrm{d}d \end{aligned} \tag{13}$$

Knowing that all cells in $(S_1 \setminus C_1)$ are at a higher distance than r_1 from O, calculating $P(\tilde{V}(n, r_1)|U(n))$ gives:

$$\begin{aligned} P\big(\tilde{V_N}(n, r_1)|U_N(n)\big) &= \Big(\frac{\mathrm{area}(S_1 \setminus C_1)}{\mathrm{area}(S_1)}\Big)^{N-n} \\ &= \Big(1 - \frac{\pi}{4}\Big)^{N-n} \end{aligned} \tag{14}$$

Now we consider area($hatchblue$) and area($hatchblack$) as defined in **Supplementary Fig. 3b**. We also define the circle C_d of radius d ($d = r + x$).

For $d \in (r_1, r_2)$:

$$\begin{aligned} P\big(\tilde{V}_N(n,d) \,\big|\, U_N(n)\big) &= \left(\frac{\text{area}(hatchblue)}{\text{area}(S_1)}\right)^{N-n} \\ &= \left(\frac{\text{area}(S_1 \setminus C_d) + \text{area}(hatchblack)}{\text{area}(S_1)}\right)^{N-n} \\ &= \left(\frac{4r_1^2 - d^2\pi + \text{area}(hatchblack)}{4r_1^2}\right)^{N-n} \end{aligned} \tag{15}$$

area($hatchblack$) is the area of 4 circular segments of arc radius d and sagitta (d - r_1) then:

$$\begin{aligned} \text{area}(hatchblack) &= 4A(d - r_1, d)_{CS} \\ &= 4d^2 \arccos\left(1 - \frac{d - r_1}{d}\right) - 4r_1\sqrt{d^2 - r_1^2} \\ &= 4d^2 \arccos\left(\frac{r_1}{d}\right) - 4r_1\sqrt{d^2 - r_1^2} \end{aligned} \tag{16}$$

Then:

$$\begin{aligned} P\big(\tilde{V_N}(n,d)|U_N(n)\big) &= \Big(\frac{4r_1^2 - d^2\pi + 4d^2 \arccos\left(\frac{r_1}{d}\right) - 4r_1\sqrt{d^2 - r_1^2}}{4r_1^2}\Big)^{N-n} \\ &= \Big(1 + \left(\frac{d}{r_1}\right)^2 \arccos\left(\frac{r_1}{d}\right) - \frac{d^2\pi + 4r_1\sqrt{d^2 - r_1^2}}{4r_1^2}\Big)^{N-n} \end{aligned} \tag{17}$$

So:

$$\rho_n(d) = -\frac{\mathrm{d}P\big(\tilde{V_N}(n,d)|U_N(n)\big)}{\mathrm{d}d} \tag{18}$$

$$\rho_n(d) = -(N-n)\Big(\frac{1}{r_1\sqrt{1 - \frac{r_1^2}{d^2}}} + \frac{2d \arccos\left(\frac{r_1}{d}\right)}{r_1^2} - \frac{\frac{4dr_1}{\sqrt{d^2 - r_1^2}} + 2\pi d}{4r_1^2}\Big)\Big(1 + \left(\frac{d}{r_1}\right)^2 \arccos\left(\frac{r_1}{d}\right) - \frac{d^2\pi + 4r_1\sqrt{d^2 - r_1^2}}{4r_1^2}\Big)^{N-n-1}$$

To find $P(\bar{B}|U(n), A)$ we are left with calculating $P(\bar{B}|U(n), V(n,d))$:

$$\begin{aligned} P(\ \bar{B_N}\ \ |\ (U_N(n) \cap V_N(n,d)) &= \Big(\frac{\text{area}(C_2 \setminus S_1) - \text{area}(hatchblack)}{\text{area}(C_2 \setminus S_1)}\Big)^n \\ &= \Big(1 - \frac{4d^2 \arccos(\frac{r_1}{d}) - 4r_1\sqrt{d^2 - r_1^2}}{(2\pi - 4)r_1^2}\Big)^n \end{aligned} \tag{19}$$

Finally:

$$\int_{r_1}^{\sqrt{2}r_1} P(\ \bar{B_N}\ \ |\ \big(U_N(n) \cap V_N(n,d)\big)\big)\ \rho_n(d)\ \mathrm{d}d = \tag{20}$$

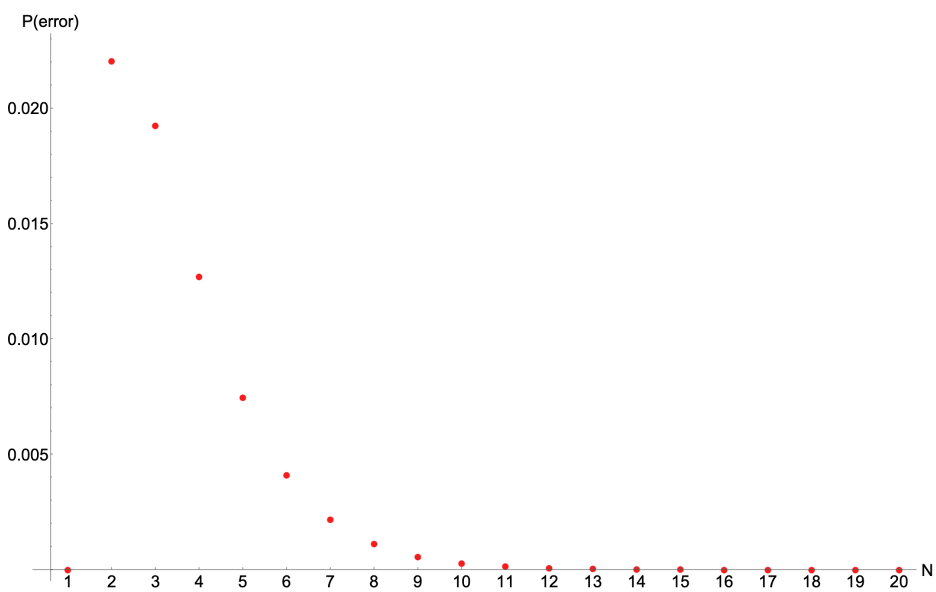
$$-(N-n)\int_{r_1}^{\sqrt{2}r_1} \Big(\frac{1}{r_1\sqrt{1 - \frac{r_1^2}{d^2}}} + \frac{2d \arccos\left(\frac{r_1}{d}\right)}{r_1^2} - \frac{\frac{4dr_1}{\sqrt{d^2 - r_1^2}} + 2\pi d}{4r_1^2}\Big)\Big(1 + \left(\frac{d}{r_1}\right)^2 \arccos\left(\frac{r_1}{d}\right) - \frac{d^2\pi + 4r_1\sqrt{d^2 - r_1^2}}{4r_1^2}\Big)^{N-n-1}\Big(1 - \frac{4d^2 \arccos(\frac{r_1}{d}) - 4r_1\sqrt{d^2 - r_1^2}}{(2\pi - 4)r_1^2}\Big)^n \mathrm{d}d$$

So to sum up:

$$P(error_N) = 1 - \frac{1}{P(A_N)}\Big[P\big(U_N(0)\big) + \sum_{n=1}^{N-1} P\big(U_N(n)\big)P\big(\bar{B_N} \mid U_N(n)\big)\Big] \quad (21)$$

where $P(A_N)$ is calculated in **(11)**, $P\big(U_N(0)\big)$ is calculated in **(10)**, $P\big(U_N(n)\big)$ is calculated in **(9)** and finally $P\big(\bar{B_N} \mid U_N(n)\big)$ is developed in **(13)**, **(14)**, and **(20)**.

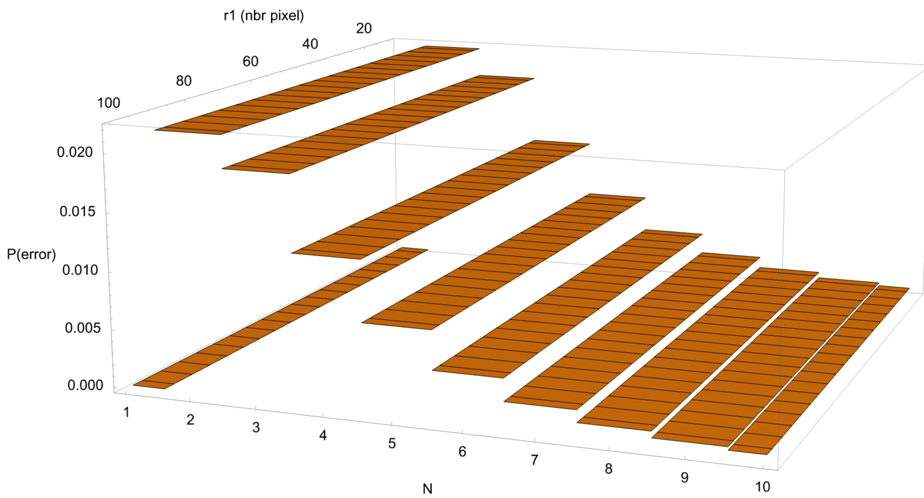
Now we can have a numerical resolution for the probability of error $P(error_N)$ as it was decomposed fully. N number of cells from 2 to 20 we can calculate $P(error_N)$ as shown in **Fig .4**.



Suppl. Fig 4: Probability of overestimation of the distance calculation in the case of a squared search area. We focus in the case of the first search, if at least one cell is found in the first squared search area. We calculate $P(error_N)$ using the symbolic expression found in **Eq. 21**. The probability of error is 0 for $N = 1$ as expected, because to consider the wrong cell we need at least 2 cells in C_2. For instance we have here $P(error_{N=2}) = 0.022$, $P(error_{N=10}) = 2.828\text{x}10^{-4}$ and $P(error_{N=20}) = 9.58\text{x}10^{-7}$.

To validate further our probability of distance overestimation, we can verify that the calculation does not depend of the value of r_1 (half length of search square side). Indeed, while calculating $P(error_N)$ with Wolfram Mathematica r_1 is given as a variable. This simple validation is displayed in **Supplementary Fig. 5**.

We provide in our github repository (available here) a Wolfram Mathematica notebook for visualization and to calculate $P(error_N)$ for any values of N.



Suppl. Fig 5: Verification of independence from r_1. We verified that the probability of error is indeed independent of r_1 (half length of search square side) while calculating numerically the probability in Wolfram Mathematica. While checking the symbolic representation of $P(error_N)$ that was found, this independence is not easy to check, especially for $P\big(\bar{B_N} \mid U_N(n)\big)$ (see **Eq. 13** & **Eq. 20**).

Finally, we can note that the relative overestimation of the distance is bounded. Going back the the case of a rectangle search area, overestimation is bounded by the aspect ratio of the tumor bounding box:

$$d_{max-overestimation} = (\sqrt{a^2+1} - 1)\frac{w}{2} \tag{22}$$

where w is the width of the rectangle and a the aspect ratio.

We covered here the case where at least one cell is found in the first search, meaning at least one cell is in S_1. In practice, distances can be overestimated in subsequent search iterations following the first. Unfortunately the calculation will be slightly different. Visualization of the cases where no cell is found inside S_1 are available at **Supplementary Fig. 3c, 3d**.

List of All Features from Tissue Analyser

All the features names are organized in a json file available in the CPI image classification dataset Zenodo repository here. We calculated two types of densities. The first one corresponds to the number of cells of a given class divided by tumor regions area (in pixel) and the second one corresponds to total area of all cells of a given class that are inside tumor regions (in pixel) divided by tumor regions area (in pixel).

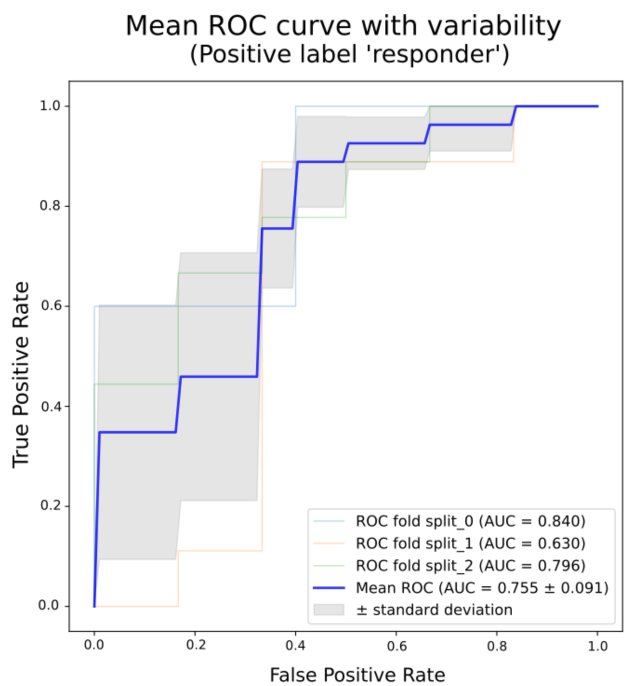
Ranking of Features after Cross-Validation with mRMR Selection

The ranking of features after the 3 fold cross-validation of XGBoost with mRMR feature selection is organized in a text file available in the CPI image classification dataset

Zenodo repository here.

ROC curves of classifier with best kept features for all cross-validation folds

We provide visualization of ROC curve for each splits.

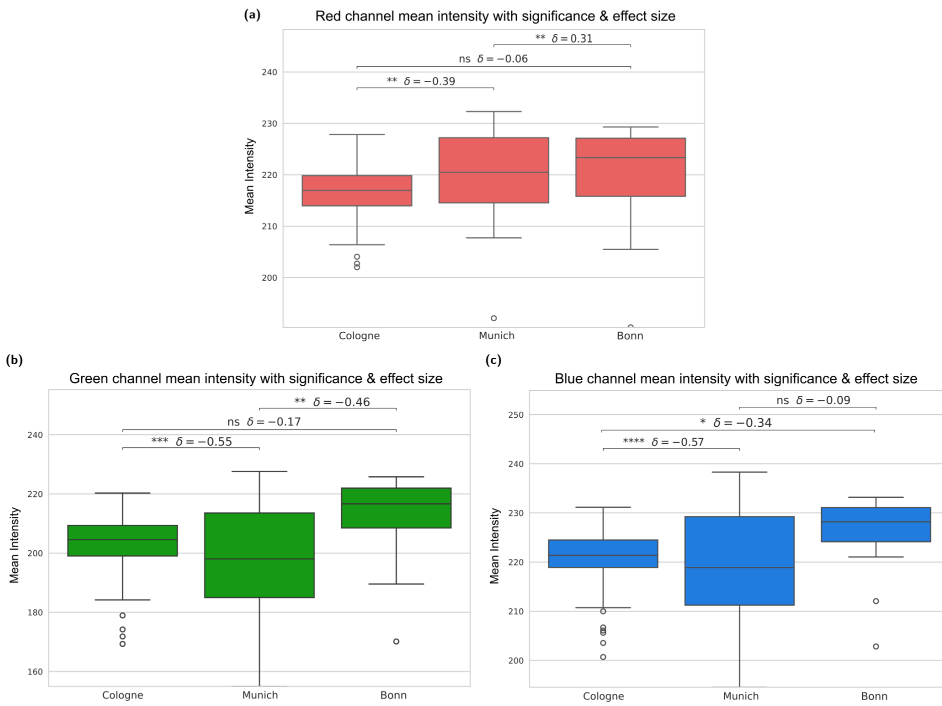


Suppl. Fig 6: Classifier cross-validation folds ROC curves.

Stain variability for the different cohorts of TumSeg dataset

We analyzed cohort-specific stain variability of TumSeg dataset by comparing RGB mean intensities between Cologne, Munich, and Bonn using the Mann–Whitney U test [6], paired with Cliff's δ [7] as a non-parametric effect size, see **Supplementary Fig. 7**.

Cologne and Bonn exhibit highly similar stain characteristics across all RGB channels, with negligible effect sizes between them (e.g., $|\delta| \leq 0.09$, ns). In contrast, Munich consistently deviates from both cohorts, showing medium to large effect sizes across channels (e.g., $|\delta| = 0.31$ in the red channel, $|\delta| = 0.46$–0.55 in the green channel, and $|\delta| = 0.34$–0.57 in the blue channel). These results indicate that Munich represents a distinct staining domain. This is the reason why color normalization step implemented within SCC Segmenter training and inference is necessary [8]. As previously stated, in the pre-processing pipeline for SCC Segmenter, the images are first downsampled, then tiled into patches and the patches are normalized using mean and standard deviation of RGB pixel values of ImageNet 1K (see our github repository for implementation details).



Suppl. Fig 7: Boxplots of Mean RGB pixel intensities for all slides from TumSeg dataset. We define the statistical significance of Mann-Whitney U test as the following: **ns**: $p \geq 0.05$; *****: $0.01 \leq p < 0.05$; ******: $0.001 \leq p < 0.01$; *******: $0.0001 \leq p < 0.001$; ********: $p < 0.0001$. While Cologne and Bonn slides have high similarities in their staining, Munich introduces a clear color-domain bias. The color normalization step from SCC Segmenter is necessary to uniform staining on all cohorts.

References

1. Graham, S. *et al.* Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis* **58**, 101563 (2018).
2. Hörst, F. *et al.* Cellvit: Vision transformers for precise cell segmentation and classification. *Medical image analysis* **94**, 103143 (2023).
3. Gamper, J. *et al.* Pannuke dataset extension, insights and baselines. *ArXiv* **abs/2003.10778** (2020).
4. Thomas, S. M., Lefevre, J. G., Baxter, G. W. & Hamilton, N. A. Non-melanoma skin cancer segmentation for histopathology dataset. *Data in Brief* **39** (2021).
5. Contributors, M. MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation (2020).
6. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**, 50–60 (1947).
7. Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.* **114**, 494–509 (1993).
8. Steiner, A. P. *et al.* How to train your vit? data, augmentation, and regularization in vision transformers. *Transactions on Machine Learning Research* (2022).

CHAPTER 3

Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma

Here we present a published work for which we used Histo-Miner pipeline as developed in **Section 2** for image classification explainability.

In this work (Pisula et al 2025 [2]), authors use Federated Learning [51] (FL) to improve classification accuracy of transformer-based multiple instance learning model applied to cSCC WSIs. The task is to classify progression status of slides - disease progression or no progression. When the transformer model is trained exclusively on one cohort (Cologne) and tested on external validation cohorts average AUROC performance is 0.65, but using FL across three clinical centers lead to an average AUROC of 0.82. In addition the work established that image-encoded information has higher discriminative power than clinical variables, as image-based models perform better in the classification task than clinico-pathological parameters.

Author used Integrated Gradient method [52] to identify the most relevant WSIs patches during classification. Part of our contribution was to employ our Histo-Miner pipeline on these patches to first, segment and classify each cell nucleus on these patches, and second, to calculate tissue and cell based features (**Fig. 4, Fig. 5, Suppl. Table 1**[3]). We discovered that many of the predictive tiles with the highest attribution score for disease progression were outside of the tumor region, whereas for patients without disease progression the predictive tiles were located within the tumors. Some cell-based features had significantly different distribution between the two groups, for instance non-progressor showed higher uniformity in the way tumor cells were distributed while progressor exhibited increased intermingling of tumor cells with other cell types. In addition, the authors found that training with federated learning and testing **XGBoost** classifier using the cell-based features calculated from Histo-Miner on the most representative patches for patch classification lead to high prediction accuracy (**Fig. 2, Suppl. Fig 10**[3])). As a result, authors concluded that Histo-Miner features captured relevant biological parameters useful for downstream tasks.

Furthermore, we took advantage of the tumor region segmentation step of Histo-Miner to cure the WSI classification dataset. In this dataset WSIs were collected from University Hospital Cologne (219 WSIs of 166 patients), University Hospital Bonn (291 WSIs of 35

[3] To access online supplementary information click or copy the DOI link that is available in **Additional Information** section

patients) and TU Munich (129 WSIs of 51 patients). Slides without any tumor tissue as predicted by Histo-Miner were filtered out. With this filter, 5 slides from the University Hospital Cologne, 158 slides from University Hospital Bonn and 16 slides from TU Munich were removed.

npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-025-00997-4

Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma

Check for updates

Juan I. Pisula[1,2,12], Doris Helbig[3,12], Lucas Sancéré[1,2], Oana-Diana Persa[4], Corinna Bürger[2,5,6], Anne Fröhlich[7], Carina Lorenz[2,5,6], Sandra Bingmann[3], Dennis Niebel[8], Konstantin Drexler[8], Jennifer Landsberg[7], Roman Thomas[5,9,10], Katarzyna Bozek[1,2,11,13] & Johannes Brägelmann[2,5,6,13]

Predicting cancer patient disease progression is a key step towards personalized medicine and secondary prevention. Risk stratification systems based on clinico-pathological criteria aim to identify high-risk patients, but accurate predictions remain challenging. Deep learning models present new opportunities for patient risk prediction, yet their interpretability has been largely unexplored. We developed a transformer-based approach for predicting progression of cutaneous squamous cell carcinoma (cSCC) patients based on diagnostic histopathology tumor slides. Our initial model showed AUROC = 0.92 on a held-out test set, with average AUROC of 0.65 on external validation cohorts. To further increase generalizability and reduce potential privacy concerns, we trained the model in a federated manner across three clinical centers, reaching AUROC = 0.82 across all cohorts, with image-based risk scores achieving hazard ratios up to 7.42 ($p < 0.01$) in multivariable analyses. Through interpretability analysis, we identified spatial and morphological features predictive of progression, suggesting that tumor boundary information and tissue heterogeneity characterize progressive cSCCs. Trained exclusively on routine diagnostic slides and offering biological insights, our model can improve secondary prevention and understanding of cSCC while enabling deployment across clinical centers without administrative overheads or privacy concerns.

Cutaneous squamous cell carcinoma (cSCC) is the second most prevalent type of non-melanoma skin cancer that is diagnosed in 1 million patients in the USA every year[1]. In the last decades, the incidence of cSCC has risen sharply and is projected to increase further[2]. Even though the majority of cSCCs can be removed by surgical excision, a relevant fraction of patients experiences disease progression by local recurrence or metastases to lymph nodes or other body sites, which is associated with poor prognosis and increased risk of death[3–6]. Due to the high incidence of cSCC, this poses a significant public health concern. Reliable predictors are thus needed to decide which patients will benefit from enhanced secondary prevention e.g. by more frequent follow-up care or additional treatments such as immuno-, chemo- or radiotherapy. Current cSCC staging systems like the American Joint Committee on Cancer (AJCC), the Brigham Women's Hospital (BWH), or the National Comprehensive Cancer Network (NCCN) staging systems aim to provide guidance on risk stratification and clinical management of cSCC patients[7,8]. However, they fall short of reliably identifying patients at high risk of disease progression. Recently, multi-gene expression signatures have been used to predict metastasis risk of cSCCs[9,10]. While these

[1]Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Köln, Germany. [2]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Köln, Germany. [3]Department for Dermatology, University Hospital Cologne, Cologne, Germany. [4]Department of Dermatology, Technical University Munich, Munich, Germany. [5]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Translational Genomics, Cologne, Germany. [6]University of Cologne, Faculty of Medicine and University Hospital Cologne, Mildred Scheel School of Oncology, Cologne, Germany. [7]Department of Dermatology and Allergology, University Hospital Bonn, Bonn, Germany. [8]Department of Dermatology, University Medical Center Regensburg, Regensburg, Germany. [9]Institute of Pathology, Medical Faculty, University Hospital Cologne, University of Cologne, Cologne, Germany. [10]DKFZ, German Cancer Research Centre, German Cancer Consortium, Heidelberg, Germany. [11]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Köln, Germany. [12]These authors contributed equally: Juan I. Pisula, Doris Helbig.[13]These authors jointly supervised this work: Katarzyna Bozek, Johannes Brägelmann. e-mail: johannes.braegelmann@uni-koeln.de

The Hormel Institute
University of Minnesota

signatures help to predict metastasis risk, they have not yet been used to predict local recurrences. In addition, they require measurement of gene expression from patient samples, which limits their potential for translation into clinical routine use.

In addition to clinical parameters such as immunosuppression, several pathological tumor features such as perineural involvement, tumor size, and invasion depth have been associated with increased risk of cSCC progression and are part of existing staging systems like the NCCN risk stratification[4–6,8]. Moreover, specific histological subtypes e.g. desmoplastic cSCC have been linked to higher recurrence and/or metastasis risk[6]. Morphology in histological specimens thus holds information on progression risk. Since deep learning has matched human experts in cancer detection and classification[11], computational pathology methods hold promise to extract information on patient progression from histopathology image data. Building robust models that offer high predictive power across data independent of their source, requires multi-institutional data sets for model training. Obtaining such data sets poses challenges regarding data governance and raises concerns about patient privacy. Federated Learning (FL) is a strategy that limits the logistic overhead and reduces privacy concerns in training a multi-center-based model[12,13]. Moreover, FL simplifies the inclusion of new patients and cohorts for further model training, which in turn facilitates model update, continuous improvement, and clinical applicability.

Here, we present a multiple instance learning transformer-based deep learning model for prediction cSCC progression risk using Hematoxylin-Eosin-(HE-) stained histopathology images acquired during routine care (Fig. 1)[14,15]. Our model, trained in a federated manner on cohorts from three clinical centers, achieved high accuracy in predicting patients at risk of disease progression, which corresponds to significant differences in progression-free survival. We developed explainability methods on our model which provide insights into the tissue areas and cell features associated with increased progression risk. Overall, we present a powerful approach that improves risk-stratification of cSCC patients and offers insights into the underlying cancer biology.

Results

Deep learning on histopathology images predicts cSCC progression risk

Even though specific pathological factors like perineural and lymphatic invasion are established parameters for risk stratification in cSCC, a systematic deep learning approach to comprehensively evaluate histopathological features for progression prediction is currently lacking. Currently, it is not clear if the progression risk of a cSCC can be inferred from a histopathology slide and which elements of the tumor and its microenvironment are decisive of disease progression. To fill this gap, we used a multiple instance learning, transformer-based classifier for the task of progression prediction from Whole Slide Images (WSIs). We trained the model in a federated manner, leveraging data from three different medical centers (Fig. 1)[14,15].

Initially, we trained our model on the Cologne cohort only ($n = 157$ patients, 214 WSIs), achieving cSCC progression status classification accuracy of 0.92 AUROC (95% CI = [0.83–1.00]) in a held-out test set from Cologne (Fig. 2A). In comparison, a multivariable logistic regression model incorporating clinico-pathological parameters associated with risk of disease progression (Suppl. Fig. 1) achieved an AUROC of 0.63 (95% CI = [0.52–0.75]) in the same prediction task and cohort (Fig. 2B). To test the robustness of our deep learning model we assembled two additional cohorts from dermatology departments at the University Hospital Bonn (Bonn cohort, $n = 35$ patients, 133 WSIs) and the Technical University Munich (Munich cohort, $n = 51$ patients, 113 WSIs). While the model trained on the Cologne cohort performed well on the Bonn cohort (AUROC = 0.90, 95% CI = [0.71–0.97]), it failed to generalize to the Munich cohort (AUROC = 0.46, 95% CI = [0.30–0.63]; Fig. 2A). A further analysis

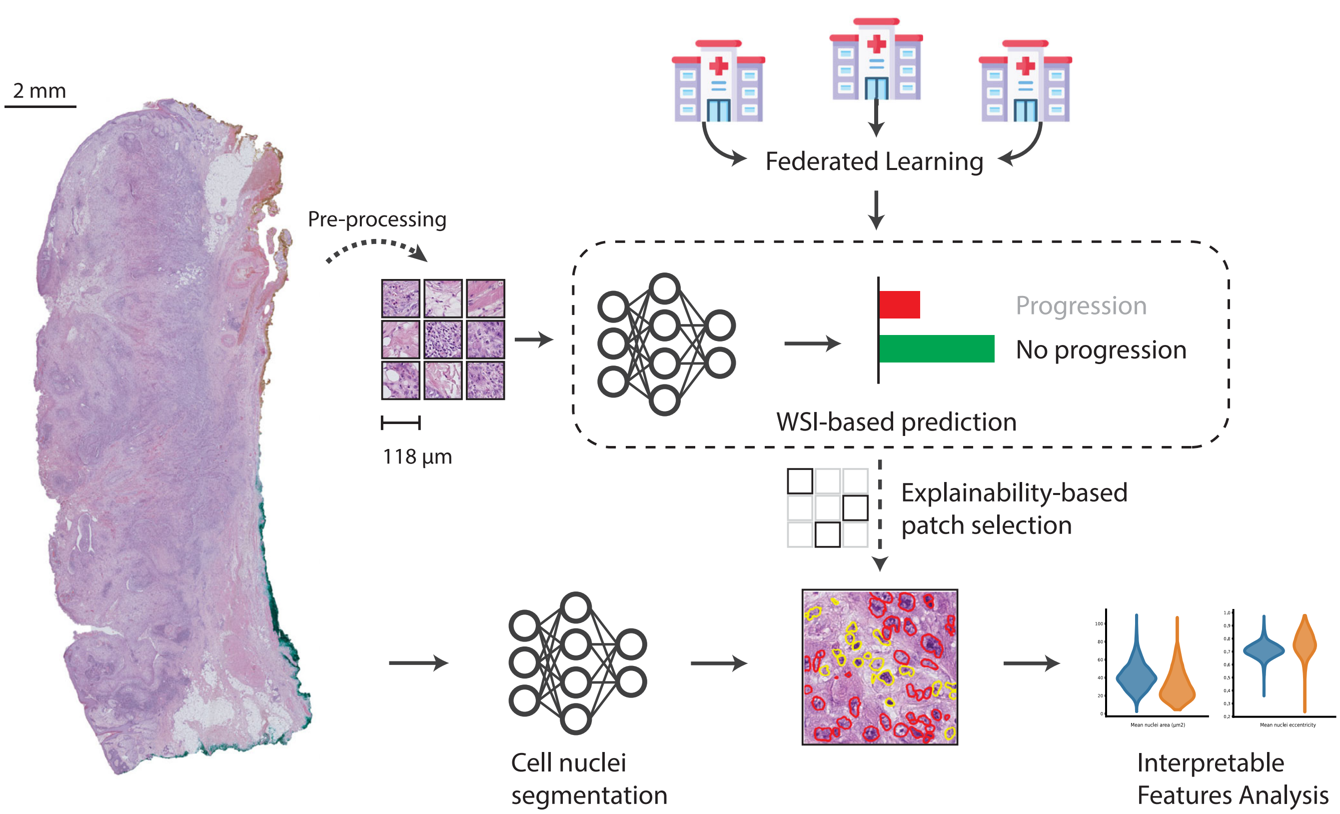


Fig. 1 | We propose a WSI-based cutaneous Squamous Cell Carcinoma (cSCC) progression prediction model, trained on data from three medical centers using Federated Learning. Beyond prediction, we investigate underlying biological features that influence our classifier. We do so by computing cellular-level features with aid of a nuclei segmentation model. We analyze these features in image regions detected as relevant for prediction outcome by Integrated Gradients, an input attribution algorithm for explainable deep neural networks.

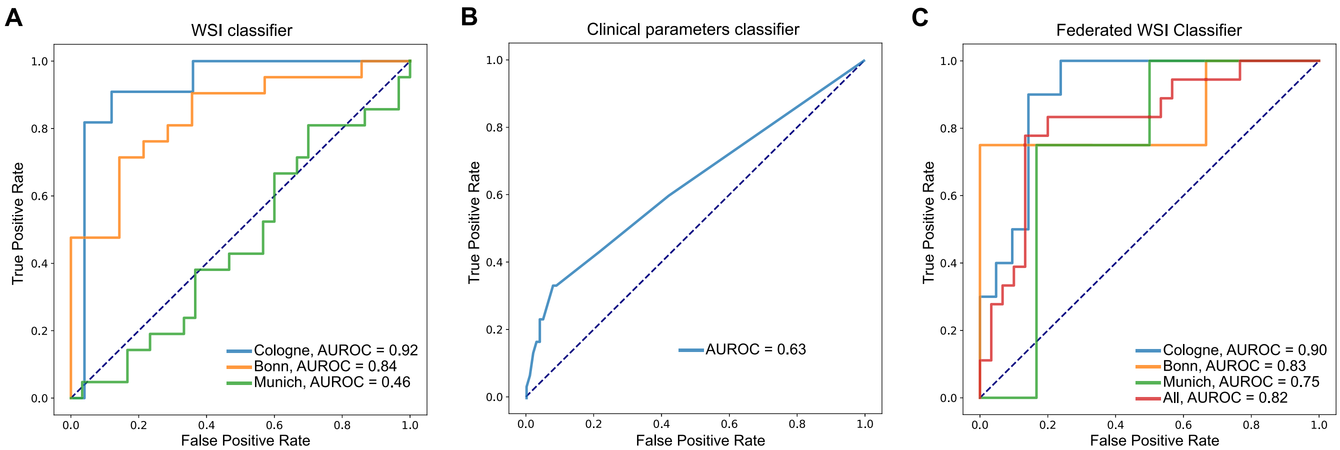


Fig. 2 | ROC curves of the classifiers. A WSI-based classifier trained exclusively on the Cologne cohort and tested on Munich and Bonn cohorts (AUROC = Area under the receiver operator curve). **B** Multivariate logistic regression model based on clinico-pathological parameters associated with progression risk in univariate analysis. Model trained and evaluated on the Cologne cohort. **C** Federated WSI-based classifier.

of this phenomenon indicates that the performance gap is due to systemic differences in image appearance rather than biological differences between the cohorts. The distributions of clinico-pathological variables shown in Table 1 appear comparable, except grading, making it unlikely that these factors contribute to the performance discrepancy. Suppl. Fig. 2 shows UMAP plots of the mean feature vector representations of the slides as computed by the EfficientNet neural network used in our pipeline[16], and by an additional CTransPath model[17]. Both plots indicate distinct visual differences between the two cohorts, potentially arising from differences in tissue processing. This highlights that variation induced by technical procedures, or distribution shift and domain adaptation problems may hamper generalizability of models trained on a single-center cohort.

Federated learning improves generalizability of image-based classification

To improve performance across cohorts, it is crucial to train deep learning models on large and diverse datasets. However transfer of patient data and histological slides across hospitals carries important logistic complexity and poses potential privacy threats. We therefore trained our model in an FL scheme on all three cohorts (Fig. 1)[12]. FL overcomes the data sharing hurdles by reducing the organizational overhead of combining different patient cohorts, since patient data can remain in the respective hospital. Model training is performed locally and only model parameters are shared between the hospitals. Moreover, it enables dynamic patient enrollment and facilitates inclusion of additional centers, which in turn increases its flexibility and the opportunities for clinical deployment. Training on the multi-institutional cohorts using the FL framework did indeed improve model performance. While AUROC on Cologne and Bonn decreased at most by 2%, performance on the Munich cohort increased by 63%, leading to prediction accuracy of AUROC = 0.82 (95% CI = [0.69–0.95]) in the complete dataset (Fig. 2C). This highlights that prediction of disease trajectories is indeed possible for cSCC patients and can be achieved with a deep learning model trained on different cohorts in a federated manner. Such prediction opens possibilities for clinical translation of the model as a tool for the identification of patients at high recurrence risk that may benefit from increased surveillance.

Image-encoded information has higher discriminative power than clinical variables

Several clinico-pathological parameters have been associated with increased risk of disease progression, such as immunosuppression, perineural involvement, tumor size, and invasion depth[4,6]. Similarly, desmoplastic cSCC histology has been linked to higher recurrence and/or metastasis risk[6]. In this experiment, we used the logit output of the deep learning models as a progression risk score, and compared it against clinico-pathological parameters available for the Cologne and Bonn cohorts. Our analyses were done separately on the Cologne cohort, to which we applied a new federated classifier trained solely on the Bonn and Munich cohorts, and on the Bonn cohort, to which we applied the original Cologne model.

From all the variables in the experiment, the image-based models' scores were the most discriminative: risk of progression was 4.2 times higher for high- compared to low-risk patients in the Cologne cohort, and 8.25 times higher in the Bonn cohort, according to univariate Cox proportional hazard models (Fig. 3A, C). These covariates were followed by perineural invasion in Cologne (Suppl. Fig. 3A, hazard ratio = 3.58, $p = 0.004$) and tumor diameter in Bonn (Suppl. Fig. 4A, hazard ratio = 5.35, $p = 0.19$).

Lastly, we joined the informative factors of our clinico-pathological parameters with the deep learning model's output to predict survival using a multivariable model. To this end we combined the deep learning models' predicted risk scores and clinico-pathological parameters that showed a p-value below 0.1 in univariate analyses in multivariable Cox regression models. These combined models showed that the image data carries more information than the clinico-pathological variables (global $p < 0.01$, Fig. 3B, D). In fact, classification of patients in the Cologne cohort as high-risk using the model trained on Bonn and Munich carries a hazard ratio of 5.96 even when adjusting for additional variables (multivariable p-value = 0.001). Similarly, patients classified as high-risk in the Bonn cohort using a model trained on Cologne only have a hazard ratio of 7.42 in a multivariate analysis (multivariable p-value = 0.01). Only the Cologne cohort model exhibits other variables that remain significant: invasion beyond subcutaneous tissue (hazard ratio = 4.53, multivariable p-value = 0.007), and tumor thickness greater than 6 mm (hazard ratio = 2.54, multivariable p-value = 0.026).

Considering that only a fraction of patients shows perineural invasion, vascular invasion, or invasion beyond subcutaneous tissue, and that clinico-pathological information is frequently incomplete, these analyses highlight the potential of our image-based model to reliably identify patients at high risk of disease progression for intensified clinical follow-up.

Explainability analyses highlight factors associated with cSCC progression

In addition to stratifying patients according to their disease progression risk, we assessed which parts of the histological images are predictive of disease progression. In these experiments, we inspected the slides of the independent test set of the federated model. We used Integrated Gradients (IGs) attributions to infer which areas in the WSIs are the most relevant for the federated model to predict the respective patient as

Table 1 | Clinicopathological characteristics of the different cSCC cohorts. *p*-values correspond to chi-squared tests for categorical variables and t-tests for continuous variables

	Cologne	Munich	Bonn	*p*
Patient characteristics				
Patient number	166	51	35	—
Sex				
Male	125 (75.3%)	—	25 (71.4%)	0.632
Female	41 (24.7%)	—	10 (28.6%)	
Age				
>80	80 (48.2%)	21 (41.2%)	17 (48.6%)	0.633
≤80	85 (51.2%)	27 (52.9%)	18 (51.4%)	
Unknown	1 (0.6%)	3 (5.9%)	—	
Age at initial diagnosis (mean ± std)	78 ± 9	79 ± 9	77 ± 9	0.549
Immunosuppression				
Yes	43 (25.9%)	6 (11.8%)	7 (20.0%)	0.137
No	123 (74.1%)	42 (82.4%)	28 (80.0%)	
Unknown		3 (5.9%)		
Tumor characteristics				
Tumor number	219	51	35	
with progress	63 (28.8%)	22 (43.1%)	14 (40.0%)	0.085
without progress	156 (71.2%)	29 (56.9%)	21 (60.0%)	
WSI (total)	219	129	291	—
with progress	66 (30.1%)	70 (52.3%)	214 (73.5%)	
without progress	153 (69.9%)	59 (47.7%)	77 (26.5%)	
WSI (detected tumor)	214	113	133	—
with progress	62 (29.0%)	61 (54.0%)	105 (79.0%)	
without progress	152 (71.0%)	52 (46.0%)	28 (21.0%)	
Invasion depth				
≤6.00 mm	149 (68.0%)	31 (60.8%)	19 (54.3%)	0.083
>6.00 mm	47 (21.5%)	16 (31.4%)	13 (37.1%)	
Unknown	23 (10,5%)	4 (7.8%)	3 (8.6%)	
Grading				
G1	137 (62.6%)	13 (25.5%)	11 (31.4%)	< 0.001
G2	35 (16.0%)	12 (23.5%)	19 (54.3%)	
G3	21 (9.5%)	12 (23.5%)	5 (14.3%)	
G4	12 (5.5%)	8 (15.7%)		
Unknown	14 (6.4%)	6 (11.8%)		
Desmoplasia				
Yes	3 (1.4%)	—	—	1
No	216 (98.6%)	48 (94.1%)	—	
Unknown	—	3 (5.9%)	—	
Perineural Invasion				
Yes	11 (5.0%)	—	—	0.368
No	208 (95.0%)	—	33 (94.3%)	
Unknown	—	—	2 (5.7%)	

progressor/non-progressor[18]. Additionally, we leveraged a separate pipeline we recently established specifically for cSCC, which performs nuclei segmentation and classification of cells into one of six cell types (granulocyte, lymphocyte, plasma, stroma, tumor, and epithelial cell)[19]. We used the cell type detection and classification to analyze the WSI regions with the highest predicted power as attributed by IGs. In the WSI regions with high IGs attribution score we calculated various features of nuclei morphology, cell type composition and spatial distribution (Suppl. Table 1).

We next performed statistical analyses of these features to gain insights into the determinants of cSCC progression. Interestingly, many of the predictive tiles with the highest attribution score for disease progression were outside of the tumor region (Fig. 4A). In fact, attribution scores were low in areas with high tumor cell density, as determined using our cell type classification pipeline (Fig. 4A, bottom left and middle)[19]. Instead, they were high at the tumor border and frequently in areas where the most common cell type was stroma (Fig. 4A, bottom right, Suppl. Fig. 5).

In contrast, for patients without disease progression, the most predictive tiles were located within the tumor and in areas with high tumor cell density. Areas outside the tumor border were, in the case of these patients, not of high value for prediction of non-progression (Fig. 4B). This highlights that different parts of histological sections contain information that distinguishes patients at high vs. low risk of disease progression and that such patient stratification needs to be based not only on the tumor but also its surroundings for adequate predictions.

Additionally, we systematically compared the cell-based features between the tiles that were regarded as most predictive for disease progression or non-progression according to their IGs scores. Numerous parameters with significantly different distributions between the two groups were detected, Fig. 5 shows a subset of the tumor-cell-related features. Non-progressors e.g. showed higher values in Average Nearest Neighbor Ratio (ANNR), indicating a higher uniformity in the way tumor cells were distributed (Fig. 5A, $p < 0.0001$), while progressors had more intermixing of tumor cells with other cell types, i.e. more heterogeneity in tissue composition (Fig. 5C, $p < 0.0001$). Moreover, tumor cells of non-progressors showed differences in their morphology compared to progressors such as larger nucleus size (Fig. 5B, $p < 0.0001$) and lower nuclear eccentricity (Fig. 5D, $p < 0.0001$). In addition, tumors of patients that later experienced disease progression showed higher degrees of nuclear dysmorphia and pleomorphism compared to non-progressors. Tumor cells from non-progressors have larger values of morphological solidity and extent (larger median, negatively-skewed distributions, Suppl. Fig. 6A–D, Suppl. Table 1), while morphological extent has a larger variance in tumor cells from progressors (Suppl. Fig. 6E, Suppl. Table 1).

Further analyses were conducted to corroborate the validity of these results. Calculating features shown in Fig. 5 separately for the tumor border and the inner tumor highlights that both regions exhibit similar feature distributions (Suppl. Fig. 7). Given the performance gap between the Cologne and Munich cohorts in the image-based progression prediction models, we investigated the biological differences between these centers according to our engineered features. Remarkably, the centers display similar feature distributions (Suppl. Fig. 8), and the features exhibit a stronger association with disease progression than with the center of origin (Suppl. Fig. 9).

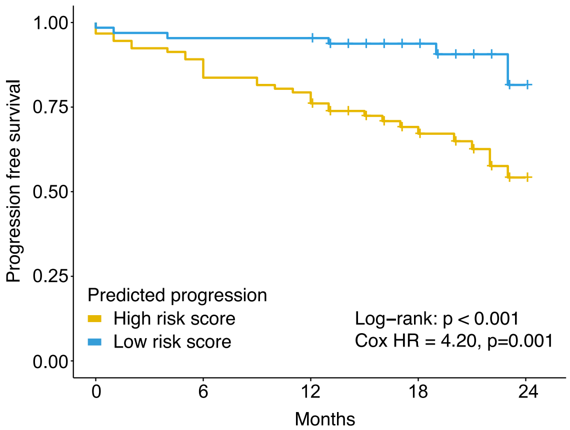
We next tested whether our cell-based features are sufficient to predict the progression/non-progression of patients based on their respective image tiles using a tree-based classification algorithm XGBoost[20]. Interestingly, using the cell-based features as input resulted in high prediction accuracy (Suppl. Fig. 10, AUROC = 0.98, 95% CI = [0.97–0.99]). This highlights that these features, which we computed using an independent pipeline, do indeed capture relevant biological parameters and variation associated with progression risk of patients. Thus the cellular and morphological features are making explicit the morphological and structural components of the tissues and cells that the deep learning model learned implicitly.

Overall, our explainability analyses indicate that tumor cell-intrinsic properties as well as composition of the microenvironment and growth patterns of the tumor are associated with the difference in prognosis and are captured by our deep learning model to accurately predict progression risk.

Discussion

Deep learning has enabled automation of the analysis of large histopathology images. These digital pathology methods not only provide fast

A

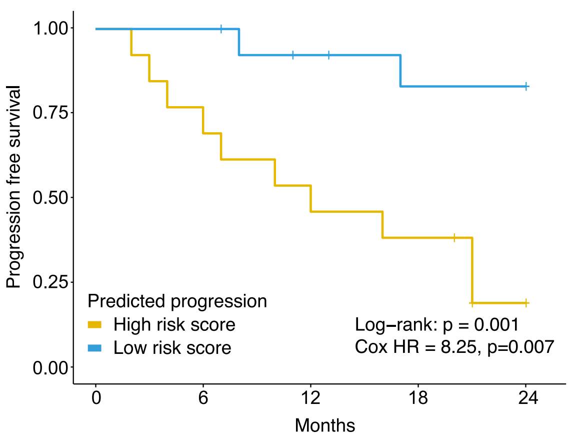


B

Variable		N	Hazard ratio (multiv.)	Hazard ratio (95% CI, multiv.)	p-value (multiv.)
Predicted risk score	Low	61			
	High	77		5.96 (2.04, 17.42)	0.001
Perineural invasion	No	132			
	Yes	6		2.98 (0.88, 10.06)	0.079
Invasion beyond subc. tissue	No	131			
	Yes	7		4.53 (1.52, 13.44)	0.007
Vascular invasion	No	136			
	Yes	2		7.13 (0.90, 56.62)	0.063
Thickness > 6mm	No	111			
	Yes	27		2.54 (1.12, 5.75)	0.026

Global model p-value < $1x10^{4}$ 1 2 5 10 50

C



D

Variable		N	Hazard ratio (multiv.)	Hazard ratio (95% CI, multiv.)	p-value (multiv.)
Predicted risk score	Low	14			
	High	13		7.42 (1.54, 35.91)	0.01
Invasion beyond subc. tissue	No	24			
	Yes	3		1.59 (0.38, 6.69)	0.53

Global model p-value = 0.004   .5 1 2 5 10 20

Fig. 3 | Comparison of federated and original deep learning models for survival prediction. A, **B** Federated model trained on Bonn and Munich cases, applied to Cologne patients. **C**, **D** Original model trained on Cologne, applied to Bonn patients. **A** Progression-free survival of Cologne patients classified as high vs. low progression risk based on federated deep learning prediction (threshold: Youden index, HR from univariate Cox regression). **B** Multivariable Cox regression for *n* = 138 Cologne patients, integrating federated deep learning risk categories with clinical parameters. **C** Progression-free survival of Bonn patients classified using the original Cologne model (threshold: Youden index, HR from univariate Cox regression). **D** Multivariable Cox regression for *n* = 27 Bonn patients, combining deep learning risk categories with clinical parameters.

and detailed insights into the cellular composition of massive WSIs[21,22], but also allow to identify patterns and anomalies that may be imperceptible to the human eye[23]. Here we present an approach that combines both: a model that detects complex, imperceptible morphological features of a tumor sample that are predictive of patient outcome with an explainability procedure to disentangle what these features are. While patient outcomes might be influenced by multifactorial clinicopathological variables and span variable development trajectories, we demonstrate that, in case of cSCC, prediction of patient progression is possible based on histological images of their tumor samples alone. Via a comprehensive and quantitative analysis of predictive regions of the tumor samples we point to consistent and repetitive patterns in tumor and tumor microenvironment morphology and organization that characterize progression and non-progression patient groups. Our model offers unmatched accuracy compared to the prediction based on clinicopathological features that were the gold standard up till now.

Our analysis combined data from three academic clinical centers: Cologne, Munich, and Bonn. The model trained on a single cohort resulted in an uneven accuracy on the remaining two cohorts, ranging from random predictions to 0.84 AUROC. Our additional analyses (Suppl Fig. 2,8,9) suggest that the performance difference observed between the Cologne and Munich cohorts is more likely due to systemic variations in image appearance, potentially stemming from disparities in tissue processing, such as reagent concentrations or processing times, rather than biological differences between cohorts. Apart from grading, the distribution of clinicopathological variables appears comparable, making it unlikely that these factors contribute to the performance discrepancy. Analysis of features extracted from cell nuclei segmentation further supports this, indicating similar distributions between centers and a stronger association with disease progression than with the center of origin. While digital pathology models require large and multi-center data for better generalization, clinical data sharing carries important administrative hurdles and data protection risks. Here we demonstrate that these difficulties can be overcome by employing an FL training scheme resulting in a model with high accuracy across all cohorts while circumventing cross-center data sharing. Our model development strategy allows for easy incorporation of additional clinical centers in the future which could potentially improve the prediction accuracy further.

Deep learning models have achieved human expert-level accuracy in standard diagnostic tasks such as tumor metastases detection and cancer subtyping[24–26]. These tasks involve detecting patterns that, while sometimes local, subtle, and difficult to notice, are known and described in pathology textbooks. In contrast, while some histopathological parameters may be indicative of disease progression[4–8], predicting patient outcomes based on WSIs remains a complex task. Numerous studies address prediction of cancer progression based on HE-stained samples of tumors across diverse

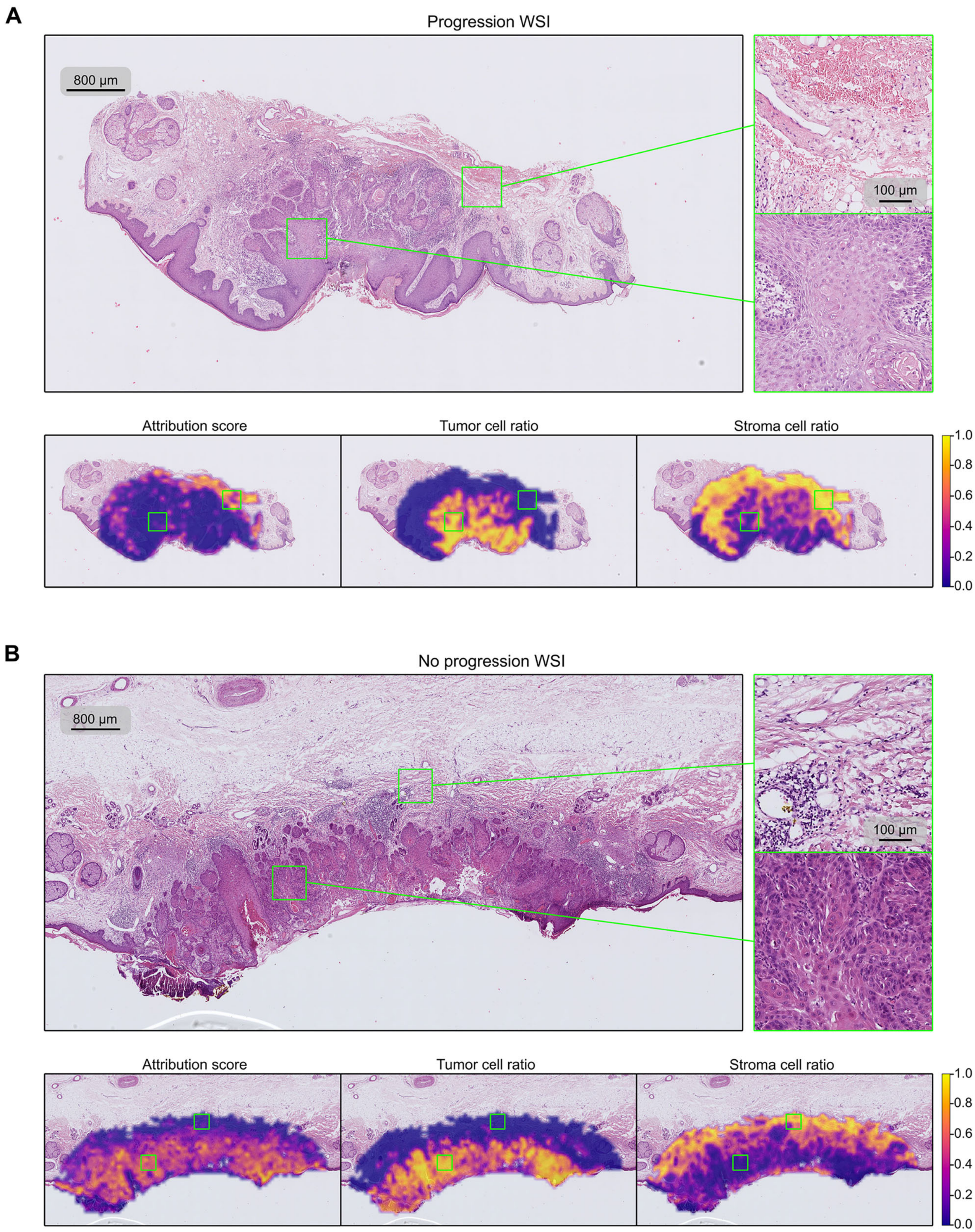


Fig. 4 | Slides and heatmaps of the patches' classifier attribution score, tumor cell ratio, and stroma cell ratio. A Slide of a progression patient, showing that the WSI-based classifier assigns higher importance to the region outside the tumor area (indicated by the tumor cell ratio heatmap). **B** Slide of a non-progression patient, where the high attribution area coincides with the tumor-cell populated areas. Colorbar indicates the slide-normalized heatmap values.

tissue types[27–32], however rarely reaching accuracy > 0.80 AUROC. Notably, combining image with clinical data has improved prediction accuracy in some studies still barely exceeding 0.80 AUROC[33–35]. In cSCC research, the work of Coudray et al. addresses the prediction of disease outcome from WSIs using a bag of visual words classifier, achieving AUROC = 0.689[36]. These examples demonstrate that prediction of patient progression is indeed difficult, and that the accuracy of our model is among the best achieved so far.

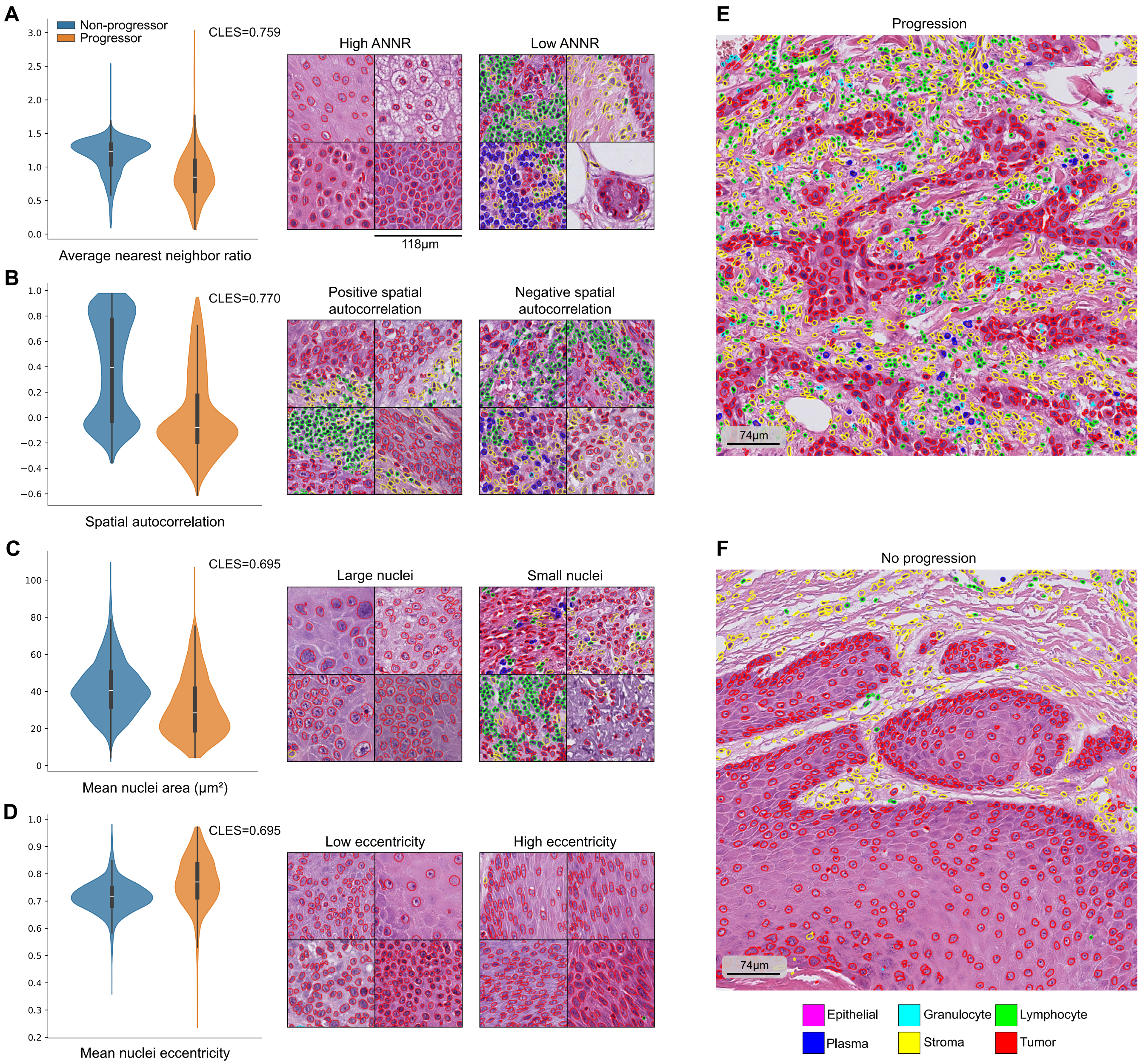


Fig. 5 | Four of the features of the tumor cells used in the analysis. **A**–**D** show violin plots and segmented image patches that illustrate these values. In general, progression-associated tumor cells cluster together (**A**), interface with other cell types (**B**), and have smaller (**C**), eccentric nuclei (**D**). These effects are not just local to image patches, but they occur in larger regions, as shown in (**E**, **F**). The displayed CLES (Common Language Effect Size) values are indicated for the group with the largest mean. All features are significantly different in both groups, with p-values < 0.0001 using Mann–Whitney U-test.

Strikingly, progression risk of a patient could be predicted based on histology images alone, exceeding by far the accuracy achieved by a model trained on clinico-pathological features. Unlike clinical parameters[7,8], or gene expression measurements[9,10], which in different clinical centers might follow different standards, be done selectively for some patients only, and come with a high cost, histology is routinely performed in cSCC diagnosis. The fact that tissue slides are available for every patient and that prediction is fast and free of additional costs, considerably increases the facility and potential of our model for clinical use. Moreover, by obviating the need for data sharing, FL greatly facilitates further model training and refinement and its extension to additional centers.

Unlike prediction based on clinical parameters, which are numeric and unambiguous, prediction based on image data is not easy to interpret. Commonly, multiple instance learning models are interpreted using qualitative inspection of image regions with high attention scores[24–26]. Here we adopt a fully quantitative and systematic approach to model interpretation in which we filter predictive patches of each patient group and statistically compare over 524 cell-based features between the two groups. Our features are based on a segmentation model specifically designed for this tumor type and capture a broad range of aspects of sample cell composition, spatial organization of the tissue, as well as nuclei morphology[19]. We point to several noticeable differences in tumor morphology between progressing and non-progressing patients.

Interestingly, the most predictive patches of disease progression were located outside of the tumor region. In contrast, in patients without disease progression, the predictive patches were inside the tumor according to our IGs-based analysis. On the level of cellular morphology and tissue architecture, tumors from patients with disease progression exhibited a higher degree of heterogeneity. Parameters quantifying nuclear morphology showed higher variability and in these patients, cells in the tumor tissues showed a less uniform distribution. Different areas in and around the cSCC tumor, as well as features of cellular morphology may play distinct roles in the propensity for local recurrence and/or metastatic spread.

Future studies in additional cohorts, ideally together with genomic and transcriptomic experiments will be instrumental to further validate our model and infer cause-and-effect relationships between morphological findings and risk of disease progression. Variability in nuclear shape has been linked to cancer and tumor grading, where higher variability correlates with greater aggressiveness. While mechanisms remain unclear, three key factors contribute to nuclear eccentricity: reduced nuclear envelope proteins (lamin A/B), chromosomal abnormalities, and mechanical forces (e.g., cytoskeletal tension or invasion)[37–42]. These factors likely interact. For example, elongated nuclei are observed in epithelial-mesenchymal transition (EMT), which increases motility and metastasis. In head and neck squamous cell carcinoma, EMT-factor Snail downregulates lamin, enhancing nuclear deformability, elongation, and metastasis propensity[43]. Nuclear deformation during migration through narrow spaces can induce DNA damage, increasing mutational load and tumor aggressiveness[37,38]. Similarly, an AI model linked nuclear area variability to aneuploidy in lung, breast, and colorectal cancer[39]. Nuclear and cellular morphology also influence signaling pathways like MAPK[40]. Multiple studies show that elongated or irregular nuclei correlate with poorer outcomes in epithelial malignancies, including cSCC, underscoring their biological relevance[41,42]. Our findings across multiple cohorts suggest high nuclear eccentricity is an intrinsic tumor trait. Further studies should explore the interplay between biomechanical properties like nuclear shape and biological processes such as EMT in cSCC aggressiveness.

In summary, our study presents an explainable, federated deep learning model that reliably stratifies cSCC patients at high risk of disease progression and identifies their characteristic morphological features. The accuracy, interpretability, and federated implementation of our model hold great promise to better understand the disease and to advance the management of cSCC patients in the future.

Methods

Patient cohorts

For the initial training cohort, all patients with a primary cSCC diagnosed and treated by excision at the Department of Dermatology at the University Hospital Cologne (Cologne cohort) between January 2009 to May 2019 were collected. For these patients we used clinico-pathological parameters based on medical records and pathology reports and performed active follow-up regarding disease progression status. Local recurrence or lymph-node/distant metastasis within 2 years after initial diagnosis was considered a progression event, and was annotated per tumor. Hematoxylin-Eosin (HE) stained slides obtained during routine work-up of surgical samples were available. The final cohort comprised 219 annotated tumors (progress $n = 63$, non-progress $n = 156$) coming from 166 patients.

From the University Hospital Bonn (Bonn cohort) patients diagnosed and treated for cSCC between March 2012 and September 2021 were included. Tumors were excised at the Department of Dermatology or the Department of Oral and Maxillo-facial Surgery and worked up histologically following standard procedures. We identified 23 primary cSCC cases with eventual disease progression (recurrence/metastasis) and randomly selected a group of primary cSCCs without disease progression. Of those, HE slides were available for 35 tumors coming from 35 patients (progress $n = 21$, non-progress $n = 14$).

For the cohort from the Department of Dermatology, Technical University Munich (TU Munich, Munich cohort) we identified patients with a primary cSCC and disease progression and assembled a random cohort of primary cSCCs without disease progression. Of those, HE slides were available for 51 tumors coming from 51 patients (progress $n = 22$, non-progress $n = 29$).

The study was performed in agreement with the Declaration of Helsinki Institutional Review Board of the University Hospital Bonn (vote number 187/16), Ethics committee of the University Hospital Cologne (vote numbers 21–1500, 20–1082 and 22–1330-retro) and institutional review board of the TU Munich (vote number 2024–363-S-CB - 1). Need for informed consent was waived for this retrospective analysis using anonymized data. Clinicopathological parameters of the cohorts are shown in Table 1.

Classification datasets

Whole-slide images (WSIs) were acquired from HE slides using a NanoZoomer Slide Scanner (Hamamatsu) at 40x resolution. In total, we collected 219 WSIs of 166 patients from the University Hospital Cologne, 291 WSIs of 35 patients from the University Hospital Bonn, and 129 WSIs of 51 patients from TU Munich. We filtered out slides without any tumor tissue according to the Segmenter model described by Sancéré et al. The final dataset used for training of the federated deep learning model comprises 214 slides from 157 patients from the University Hospital Cologne, 133 slides from 35 patients from the University Hospital Bonn and 113 slides from 51 patients from TU Munich. From this dataset, 228 slides are from patients showing cSCC progression, and 232 slides are from patients showing no cSCC progression. Data splitting is done in a stratified fashion on patient level, making 65-15-20 splits for training, validation, and testing, respectively.

Pre-processing

Each WSI is tiled into patches of 256×256 pixels at x20 magnification. Patches with less than 50% tissue are discarded, and the remaining patches are processed with an ImageNet pre-trained EfficientNet-v2-L[16], to compute its feature vector representations. The average slide of our dataset produces 11330 feature vectors.

Classification

Each WSI is treated as the sequence of feature vectors corresponding to its non-empty image patches. We use the multiple instance learning classification model described by Pisula and Bozek[44]. Following an approach similar to Lu et al.[45], a transformer model initialized with language-modeling pre-training weights is used for classification. We use a RoBERTa transformer encoder[46], and perform parameter-efficient fine-tuning by only training its normalization layers[45,47]. To reduce compute and memory footprint, we apply multi-head attention pooling at the input to shorten the length of the patch sequence. The embedding vectors from the last layer of the transformer encoder are averaged and fed to a linear layer for the final classification.

Each WSI is classified independently during model training. During inference, in cases where there are multiple slides per patient, we evaluate the model on each one and take the prediction corresponding to the slide with the biggest activation in the positive class output neuron.

We train our model with a Federated Averaging strategy for 50 rounds[12]. Adam is used as the optimizer algorithm, with a learning rate of 1.e-4, weight decay of 5.e-5, and batch size of 4. Model selection is done based on weighted validation AUROC of the three cohorts.

Classification explanation and analysis

Beyond mere disease progression prediction with a deep network classifier, we investigate the biological features that drive our classifier's decision. Our process is threefold: we detect relevant image regions responsible for the model's decision; we compute handcrafted features of the cellular composition of the image regions; and we perform the data analysis itself. This approach is described in detail below.

We use Integrated Gradients (IGs) to identify regions of a WSI that play a role in the classifier's progression prediction[18]. IGs is a deep learning explainability algorithm that attributes the prediction of a deep network to its input features. We apply IGs to our cSCC progression prediction model, to assign a positive score to image patches that contribute to the prediction of the correct class, and a negative score to patches that contribute to the prediction of the opposite outcome. By arranging the IGs attribution scores of the patches in their corresponding spatial locations in the slides, it is possible to visualize these values as heatmaps, as shown in Fig. 4.

We use the HoverNet nuclei segmentation model described by Sancéré et al. on the WSI image patches to identify their cell composition[19,21]. The model detects and classifies cell nuclei into granulocytes, lymphocytes, plasma cells, stroma cells, tumor cells, and non-neoplastic epithelial cells. Once the cells in a patch have been identified, we compute a total of 524 features that summarize the patch into a single feature vector. These features include:

- Cell type populations and ratios.
- Descriptive statistics (mean, median, variance, skewness, kurtosis, minimum, maximum) of nuclei morphology, such as the mean tumor cells nuclei eccentricity, or the variance in plasma cells nuclei area. These features were computed with the 'skimage.measure' Python package[48].
- Descriptive statistics of distances between cell nuclei, such as the median distance between stroma cells and tumor cells.
- Average Nearest Neighbor Ratio (ANNR) and Join Count analysis for each cell type.

The features from the last item are used to quantify the spatial arrangement of cells within a patch, and they capture two different aspects of it.

ANNR is used to quantify the observed pattern of distances between cell nuclei in a patch:

$$ANNR = \underline{D_O}/\underline{D_E} \quad (1)$$

where $\underline{D_O}$ is the observed mean distance between each cell and its closest neighbor, and $\underline{D_E}$ is the expected mean distance between each cell and its closest neighbor if the cells were placed randomly:

$$\underline{D_E} = \frac{0.5}{\sqrt{n/A}} \quad (2)$$

where n is the number of cells in a patch, and A is the patch area. An ANNR < 1 indicates clustering (meaning, cells in the patch are closer than a random pattern of cells), and an ANNR > 1 indicates a dispersed or regular pattern of cell nuclei. We compute the ANNR for each cell type in a patch.

Join Count analysis gives a measure of spatial autocorrelation: it describes how the values of a variable at neighboring spatial locations are similar to each other. In our case, the variable of interest is the cell type, where a positive spatial autocorrelation would mean that neighboring cells belong to the same type, and a negative spatial autocorrelation would mean that neighboring cells belong to different classes. Spatial autocorrelation is complementary to ANNR, it quantifies neighboring cell nuclei types disregarding how close or distanced they are.

Our Join Count analysis is computed for each cell type individually, in the following way:

- A patch is partitioned into a Voronoi tessellation, using the nuclei centroids as seeds for the regions.
- The regions are binary-labeled. Given a cell type, a positive label is assigned to all the cell nuclei belonging to that class, and a negative label is assigned to the remaining regions.
- The different types of joins were then counted. Two neighboring cells make a black-black (BB) join if they both are from the positive label (i.e. the cell type being currently analyzed); a black-white (BW) join is formed between two cells of opposite labels; and a white-white (WW) join happens when two cells of the negative label neighbor each other.

This procedure is done for each cell type independently, assigning the positive label (black) to the analyzed cell type and the negative label (white) to all the other cell types. Our measure of spatial autocorrelation is given by:

$$Spatial\ Autocorrelation = (J_{BB} - J_{BW})/J_T \quad (3)$$

where J_{BB}, J_{BW} and J_T are the number of BB joins, the number of BW joins, and the total number of joins, respectively. This equation is positive when the majority of joins in a patch are BB joins, indicating a positive spatial autocorrelation, and is negative when the majority of joins are BW joins, indicating negative spatial autocorrelation.

We apply IGs to all the patients in the test set, and describe their corresponding image patches as previously explained. We use in this analysis the patches coming from tumor regions detected by the Segmenter model described by Sancéré et al.[19,49], plus a surrounding tissue stripe of approximately 800µm of width next to the tumor border. From the totality of patches, we form two groups: A "positive group" of image patches coming from progression patients, which were detected to be explainable of this condition with IGs; and a "negative group" of patches coming from non-progression patients, which were detected to be explainable of this condition with IGs.

To enhance the predictive signal and avoid over-representing patients with bigger tumors, we take a slide's top 10% IGs-scored patches, and limit this quantity to 200 image patches per slide. We compare values of each feature individually between the two groups of patches. We guide our analysis by focusing on features whose values differ between the two groups with an Effect Size bigger than random. We use the Common Language Effect Size (CLES)[50], or probability of superiority, as it has no assumptions about the data distribution, and is straightforward to understand:

$$CLES = P(X > Y) \quad (4)$$

is the probability that a value sampled from group X is bigger than a value sampled from group Y. In our case, the two groups are the positive and the negative groups previously described, and we compute the CLES for each feature with brute force, by exhaustively comparing each value of one group with all the values of the same feature in the other group.

In addition to comparing the feature distributions in both groups, we tested whether the individual patches' feature vectors were sufficient to predict the progression status of their respective patients using an XGBoost classifier[20]. The patches under analysis (coming from the FL model's test set) were split into 80−20 train and test sets, and model selection was done with 3-fold cross-validation on this new train split.

Survival analysis with clinico-pathological variables

Associations of clinico-pathological variables with disease progression and survival were done for all patients with available data. Association with disease progression risk was calculated using logistic regression and reported as odds ratios. Association with survival was done using the Kaplan-Meier method with log-rank test as well as Cox proportional hazard models and reported as hazard ratios with 95% confidence intervals. For multivariable analyses, variables with $p < 0.1$ in univariate analysis were combined. Analyses were done in R statistical environment (v4.3.0).

Data availability

Data is available upon reasonable request to the authors. Code is available at https://github.com/bozeklab/cscc-response.

Received: 13 November 2024; Accepted: 4 June 2025;
Published online: 28 June 2025

References

1. Winge, M. C. G. et al. Advances in cutaneous squamous cell carcinoma. *Nat. Rev. Cancer* **23**, 430–449 (2023).
2. Keim, U. et al. Incidence, mortality and trends of cutaneous squamous cell carcinoma in Germany, the Netherlands, and Scotland. *Eur. J. Cancer Oxf. Engl. 1990* **183**, 60–68 (2023).
3. Brantsch, K. D. et al. Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. *Lancet Oncol.* **9**, 713–720 (2008).
4. Schmults, C. D., Karia, P. S., Carter, J. B., Han, J. & Qureshi, A. A. Factors predictive of recurrence and death from cutaneous squamous

cell carcinoma: a 10-year, single-institution cohort study. *JAMA Dermatol.* **149**, 541–547 (2013).
5. Thompson, A. K., Kelley, B. F., Prokop, L. J., Murad, M. H. & Baum, C. L. Risk factors for cutaneous squamous cell carcinoma recurrence, metastasis, and disease-specific death: a systematic review and meta-analysis. *JAMA Dermatol* **152**, 419–428 (2016).
6. Eigentler, T. K., Dietz, K., Leiter, U., Häfner, H.-M. & Breuninger, H. What causes the death of patients with cutaneous squamous cell carcinoma? a prospective analysis in 1400 patients. *Eur. J. Cancer* **172**, 182–190 (2022).
7. Ruiz, E. S., Karia, P. S., Besaw, R. & Schmults, C. D. Performance of the american joint committee on cancer staging manual, 8th edition vs the brigham and women's hospital tumor classification system for cutaneous squamous cell carcinoma. *JAMA Dermatol.* **155**, 819–825 (2019).
8. Schmults, C. D. et al. NCCN guidelines® insights: squamous cell skin cancer, version 1.2022. *J. Natl. Compr. Cancer Netw. JNCCN* **19**, 1382–1394 (2021).
9. Wysong, A. et al. Validation of a 40-gene expression profile test to predict metastatic risk in localized high-risk cutaneous squamous cell carcinoma. *J. Am. Acad. Dermatol.* **84**, 361–369 (2021).
10. Wang, J. et al. Transcriptomic analysis of cutaneous squamous cell carcinoma reveals a multigene prognostic signature associated with metastasis. *J. Am. Acad. Dermatol.* **89**, 1159–1166 (2023).
11. Haenssle, H. A. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol. J. Eur. Soc. Med Oncol.* **29**, 1836–1842 (2018).
12. McMahan, H. B., Moore E., Ramage, D., Hampson, S. & Arcas, B. A. Y. Communication-efficient learning of deep networks from decentralized data. *arXiv* http://arxiv.org/abs/1602.05629 (2024).
13. Ogier Du Terrail, J. et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat. Med* **29**, 135–146 (2023).
14. Maron, O., Lozano-Pérez, T. A framework for multiple-instance learning. In *Advances in Neural Information Processing Systems* (MIT Press, 2024).
15. Vaswani A., et al. Attention is all you need. *arXiv*https://doi.org/10.48550/arXiv.1706.03762.
16. Tan M., Le Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *ArXiv* https://doi.org/10.48550/arXiv.1905.11946 (2024).
17. Wang, X. et al. Transformer-based unsupervised contrastive learning for histopathological image classification. *Med. image Anal.* **81**, 102559 (2022).
18. Sundararajan, M., Taly, A., Yan, Q. Axiomatic attribution for deep networks. *arXiv* http://arxiv.org/abs/1703.01365 (2023).
19. Sancéré, L. Histo-Miner: Tissue features extraction with deep learning from h&e images of squameous cell carcinoma skin cancer. *arXiv* https://doi.org/10.48550/arXiv.2505.04672 (2024).
20. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
21. Graham, S. et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019).
22. Hörst, F. et al. CellViT: Vision transformers for precise cell segmentation and classification.*arXiv* https://doi.org/10.48550/ARXIV.2306.15350 (2023).
23. Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
24. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
25. Shao, Z. et al. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. *arXiv* https://doi.org/10.48550/arXiv.2106.00908 (2021).
26. Chen R. J. et al. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. *arXiv*https://doi.org/10.48550/arXiv.2206.02647 (2022)
27. Dietrich, E. et al. Towards explainable end-to-end prostate cancer relapse prediction from h&e images combining self-attention multiple instance learning with a recurrent neural network. In *Proc. Machine Learning for Health* 38–53 (PMLR, 2021).
28. Akram, F. et al. Artificial intelligence-based recurrence prediction outperforms classical histopathological methods in pulmonary adenocarcinoma biopsies. *Lung Cancer* **186**, 107413 (2023).
29. Wu, Z. et al. DeepLRHE: a deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images. *Front Genet*. **11**, 768 (2020).
30. Xiao, H. et al. Predicting 5 year recurrence risk in colorectal cancer: development and validation of a histology-based deep learning approach. *Br. J. Cancer* **130**, 951–960 (2024).
31. Foersch, S. et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat. Med.* **29**, 430–439 (2023).
32. Shi, Y. et al. Predicting early breast cancer recurrence from histopathological images in the Carolina breast cancer study. *Npj Breast Cancer* **9**, 1–7 (2023).
33. Howard, F. M. et al. Integration of clinical features and deep learning on pathology for the prediction of breast cancer recurrence assays and risk of recurrence. *Npj Breast Cancer* **9**, 1–6 (2023).
34. Yang, J. et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput Struct. Biotechnol. J.* **20**, 333–342 (2022).
35. Lucas, M. et al. Deep learning–based recurrence prediction in patients with non–muscle-invasive bladder cancer. *Eur. Urol. Focus* **8**, 165–172 (2022).
36. Coudray, N. et al. Self-supervised artificial intelligence predicts recurrence, metastasis and disease specific death from primary cutaneous squamous cell carcinoma at diagnosis. *Res. Sq.* **13**, rs.3.rs-3607399 (2023).
37. Shah, P. et al. Nuclear deformation causes DNA damage by increasing replication stress. *Curr. Biol.* **31**, 753–765 (2021).
38. Fan, J.-R. et al. AKT2-mediated nuclear deformation leads to genome instability during epithelial-mesenchymal transition. *Iscience* **26**, 106992 (2023).
39. Abel, J. et al. AI powered quantification of nuclear morphology in cancers enables prediction of genome instability and prognosis. *NPJ Precis. Oncol.* **8**, 134 (2024).
40. Rangamani, P. et al. Decoding information in cell shape. *Cell* **154**, 1356–1369 (2013).
41. Glazer, E. S. et al. Nuclear morphometry identifies a distinct aggressive cellular phenotype in cutaneous squamous cell carcinoma. *Cancer Prev. Res.* **4**, 1770–1777 (2011).
42. Grosser, S. et al. Cell and nucleus shape as an indicator of tissue fluidity in carcinoma. *Phys. Rev. X* **11**, 011033 (2021).
43. Chen, Y.-Q. et al. Snail augments nuclear deformability to promote lymph node metastasis of head and neck squamous cell carcinoma. *Front. Cell Dev. Biol.* **10**, 809738 (2022).
44. Pisula, J. I. & Bozek, K. Efficient WSI classification with sequence reduction and transformers pretrained on text. *Sci. Rep.* **15**, 5612 (2025).
45. Lu, K. Grover, A. Abbeel, P. Mordatch, I. Pretrained transformers as universal computation engines. *arXiv* http://arxiv.org/abs/2103.05247 (2024).
46. Liu, Y. et al. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv* https://doi.org/10.48550/arXiv.1907.11692 (2019).

47. Lialin, V., Deshpande, V., Rumshisky, A. Scaling down to scale up: a guide to parameter-efficient fine-tuning. *arXiv* http://arxiv.org/abs/2303.15647 (2024).
48. Walt, S. et al. scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).
49. Strudel R., Garcia R., Laptev I., Schmid C. Segmenter: Transformer for semantic segmentation. *arXiv*https://doi.org/10.48550/arXiv.2105.05633 (2021).
50. McGraw, K. O. & Wong, S. P. A common language effect size statistic. *Psychol. Bull.* **111**, 361–365 (1992).

Acknowledgements
A.F. was partly funded by the Deutsche Krebshilfe through a Mildred Scheel Foundation Grant (grant number 70113307). C.L. was partly funded through the collaborative research center grant on small cell lung cancer (CRC1399, project ID 413326622) by the German Research Foundation (DFG). J.B. receives funding through the collaborative research center grant on small cell lung cancer (CRC1399, project ID 413326622) and predictability in evolution (CRC1310, project ID 325931972) by the German Research Foundation (DFG), a Mildred Scheel Foundation Grant (grant number 70113307) by the Deutsche Krebshilfe and the CANTAR network (NW21-062B) funded through the program "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of Northrhine Westphalia, Germany. Both K.B. and J.I.P. were hosted by the Center for Molecular Medicine Cologne throughout this research. K.B. and J.I.P. were supported by the BMBF program Junior Group Consortia in Systems Medicine (01ZX1917B) and BMBF program for Female Junior Researchers in Artificial Intelligence (01IS20054).

Author contributions
K.B., J.B. conceived and designed the study. J.I.P., L.S., J.B. developed software and conducted the statistical analyses. D.H., O.P., C.B., A.F., S.B., D.N., K.D., J.L., R.T. provided samples, clinical data, and helped with data curation as well as interpretation of results. J.I.P., D.H., L.S., K.B., J.B. had access to all raw data of the study. J.I.P., D.H., K.B., J.B. drafted the manuscript. All authors read and approved the manuscript and participated in reviewing and editing of the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The Authors declare no Competing Non-Financial Interests but the following Competing Financial Interests: D.N. received financial support (speaker's honoraria, advisory boards, travel expense reimbursements or grants) from Abbvie, Almirall, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol-Myers-Squib, GlaxoSmithKline, Incyte, Janssen-Cilag, Kyowa Kirin, LEO Pharma, Lilly, L'Oreal/Cerave, MSD, Novartis, Pfizer, Regeneron and UCB Pharma. J.B. received research funding from Bayer and travel expenses from Merck KG and Bicycle Therapeutics outside the presented work. K.D. received financial support (speaker's honoraria, advisory boards, travel expense reimbursements or grants) from Abbvie, Bristol-Myers-Squib, Novartis, and Pierre-Fabre.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-025-00997-4.

Correspondence and requests for materials should be addressed to Johannes Brägelmann.

Chapter 4

Context-aware skin cancer epithelial cell classification with scalable graph transformers

In the previous work of **Chapter 2** and **Chapter 3** we used Histo-Miner pipeline to segment and classify nuclei from cSCC WSIs. We needed a 2-models approach to successfully classify healthy and tumor epithelial: we refined the classification performed by cSCC Hovernet based on the tumor region segmented by cSCC Segmenter model. However, this dual-model framework may overlook isolated tumor cells or micro-metastases that fall below the detection threshold of cSCC Segmenter.

In this work we evaluated Graph Neural Networks, and more specifically graph Transformers such as **NodeFormer** [46], **DIFFormer** [47] and **SGFormer** [48] in the task of node classification. From a partially annotated WSI we built a graph called **WSI-Graph** with nodes representing cell and node feature being morphology, texture (intensities and colors) and class of the cell nucleus. We trained graph-based method on the built graph and image-based method such as CellViT [25] and Hovernet [24] on the original WSI for the task of epithelial cell classification as healthy or tumor epithelial.

In the previous experiment we used a full WSI from a single patient. Unfortunately, scaling models such as CellViT and Hovernet to entire WSI datasets remains a challenge, as the computational overhead of full-slide training is currently unsustainable for multiple patients. To still compare image-based and graph-based model in a several patient context, we extracted 372 patches from 93 WSIs of 84 and built 372 graphs from these patches, consisting in **TILE-Graphs**. Once again we compared the performances of all models for epithelial cell classification.

We found that in both cases, Graph Transformers with linear complexity outperform image-based methods. Additionally, training and evaluation of graph-based approach are significantly faster than for image-based approach. We can then use graph representation of WSIs instead of original images to improve cell classification of our previous work.

Context-aware Skin Cancer Epithelial Cell Classification with Scalable Graph Transformers

Lucas Sancéré[1,2,3,4], Noémie Moreau[2,3,4], and Katarzyna Bozek[2,3,4]

[1] Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[2] Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[3] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, North Rhine-Westphalia, Germany
[4] Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, North Rhine-Westphalia, Germany
`lsancere@uni-koeln.de, k.bozek@uni-koeln.de`

Abstract. Whole-slide images (WSIs) from cancer patients contain rich information that can be used for medical diagnosis or to follow treatment progress. To automate their analysis, numerous deep learning methods based on convolutional neural networks and Vision Transformers have been developed and have achieved strong performance in segmentation and classification tasks. However, due to the large size and complex cellular organization of WSIs, these models rely on patch-based representations, losing vital tissue-level context. We propose using scalable Graph Transformers on a full-WSI cell graph for classification. We evaluate this methodology on a challenging task: the classification of healthy versus tumor epithelial cells in cutaneous squamous cell carcinoma (cSCC), where both cell types exhibit very similar morphologies and are therefore difficult to differentiate for image-based approaches. We first compared image-based and graph-based methods on a single WSI. Graph Transformer models SGFormer and DIFFormer achieved balanced accuracies of 85.2 ± 1.5 ($\pm$ standard error) and 85.1 ± 2.5 in 3-fold cross-validation, respectively, whereas the best image-based method reached 81.2 ± 3.0. By evaluating several node feature configurations, we found that the most informative representation combined morphological and texture features as well as the cell classes of non-epithelial cells, highlighting the importance of the surrounding cellular context. We then extended our work to train on several WSIs from several patients. To address the computational constraints of image-based models, we extracted four 2560×2560 pixel patches from each image and converted them into graphs. In this setting, DIFFormer achieved a balanced accuracy of 83.6 ± 1.9 (3-fold cross-validation), while the state-of-the-art image-based model CellViT256 reached 78.1 ± 0.5. DIFFormer training was also substantially faster, requiring only 32 minutes for one single cross-validation fold compared with approximately 5 days for CellViT256. Overall, these results suggest that graph-based approaches constitute a promising alternative to traditional computer vision methods for object classification tasks, such as cell classification.

Introduction

Hematoxylin and Eosin (H&E) staining [1] is widely used in pathology and serves as a standard staining method for tissue examination. The resulting scans, called Whole-Slide Images (WSIs), are high-resolution digital images of entire tissue sections captured from microscope slide. Deep learning has substantially advanced WSI analysis, primarily through convolutional neural networks (CNNs) [2,3] and more recently Transformer models [4,5,6]. Due to the large size of WSIs, their automated processing requires first splitting the image into smaller patches, extracting features from the patches, before aggregating them for prediction or segmentation of the entire slide. This approach enables analysis of the large images, however does not allow the models to capture the tissue structure as a whole as each patch contains only a small area of the image.

To explicit tissue spatial organization, recent approaches represent WSIs as graphs, where nodes correspond to image regions and edges encode spatial relationships. Different graph neural networks architecture were applied to such graphs for survival outcome prediction [7,8,9,10]. Notably, Graph Transformers models were used for lung, breast and kidney cancer WSI classification based on individual patches [11,12,13]. These models demonstrate the effectiveness of graph representations and GNNs for WSI analysis. While graphs are powerful models for representing WSIs, graphs that are based on patches rather than individual cells do not capture the detailed tissue composition and cellular interactions within it.

Graph representation in which nodes correspond to individual cells or nuclei and edges link cells based on spatial proximity or feature similarity is a rich and detailed model of a tissue sample. HistoCartography toolkit [14] was designed to build such cell graphs but is limited to individual patches of the full WSI. Such graphs were used for cell classification, through node classification with GNNs and notably Graph Transformers, however using only patch-level context [15,16,17]. Similarly, hierarchical graphs including cell representations from WSI patches were developed for tissue segmentation, cancer grading and breast cancer classification [18,19,20]. In these approaches nodes represent individual cells enabling biologically meaningful modeling, nonetheless, graphs remain restricted to local patches and do not capture entire WSIs. Achieving tissue-level representation and analysis with graphs requires incorporating all cells within a WSI and developing GNNs and their training strategies capable of operating at this scale.

Classical GNNs rely on localized message passing, which limits their ability to capture long-range dependencies and restricts their scalability to large graphs. Graph Transformer models [21,22,23] enable to address this challenge with global attention across nodes while leveraging efficient attention mechanisms. However, the quadratic complexity of the attention mechanism with regard to the input tokens, makes training on large graphs impossible. To overcome this issue new architectures of scalable Graph Transformers with linear complexity were developed such as Nyströmformer [24] and more recently NodeFormer [25], DIFFormer [26] and Simplified Graph Transformers (SGFormer) [27]. These models were used for node classification on large-graph benchmark such as ogbn-proteins [28], Amazon-M2 [29] or pokec [30]. Training such models on cell-level WSI graphs could significantly improve node classification accuracy over traditional patch-based methods.

In this work we propose a graph-based approach for tumor cell classification in cutaneous squamous cell carcinoma (cSCC), the second most common non-melanoma skin cancer worldwide [31,32]. Previous work showed that distinguishing between healthy and tumor epithelial cells in the cSCC is challenging using classical patch-based segmentation methods. The morphologies of these cells are very similar and only a broader tissue-level context capturing cell spatial organization allows to correctly classify them in these two types [33]. We tackle the challenge of discriminating between healthy and tumor epithelial cells by building large WSI graph and training scalable Graph Transformers with linear complexity. We use a previously published dataset of cSCC WSIs [33] and construct graphs at two different levels, WSI and patch and term these approaches "WSI-Graph" and "TILE-Graphs" to compare GNNs and image-based models on epithelial cell classification. We first asses the performance of GNN models on WSI-Graph and found the best combination of node features. We then evaluated image-based and graph-based approach on a several patient configuration with TILE-Graphs. To the best of our knowledge this work is the first to:

- Encode a full WSI at a single cell level as a graph to generate node classification predictions,
- Apply graphs to improve classification of epithelial cells as healthy or tumor in the cSCC skin cancer,
- Compare graph-based and image-based methods on the same underlying data represented as images and corresponding graph structures.

Methods

From WSI images to cell graphs

To apply a GNN for the classification of healthy versus tumor epithelial cells, we converted a WSI of cSCC into a structured graph representation. In this graph, each node corresponds to a cell nucleus and encodes morphological and texture features, as well as the associated cell class. Edges connect neighboring cells, thereby capturing their spatial relationships.

To obtain these nuclei and their initial labels, we used cSCC Hovernet [33], a model previously developed for cell segmentation and classification. cSCC Hovernet detects, segments and then categorizes cells into five types: granulocytes, plasma cells, lymphocytes, stromal cells, and epithelial cells (without distinguishing between healthy and tumoral cells). Using tumor regions annotated by an expert pathologist, we refined this classification by splitting epithelial cells into two subclasses: tumor epithelial and healthy epithelial. Specifically, epithelial cells located inside annotated tumor regions were relabeled as tumor epithelial, whereas epithelial cells outside these regions were relabeled as healthy epithelial. Importantly, not all cells within a tumor region are tumor cells; only those previously identified as epithelial are reassigned as tumor epithelial. As a result, each cell of the entire WSI is categorized into one of the six classes: granulocytes, plasma cells, lymphocytes, stromal cells, tumor epithelial and healthy epithelial.

Based on the cell segmentation and classification, we represent a WSI as a graph, where each detected cell nucleus forms a node. Each node $i \in \{1, \dots, N\}$ among all N nodes is associated with a feature vector $\mathbf{h}_i = [\mathbf{f}_i \parallel \mathbf{c}_i] \in \mathbb{R}^{l+6}$ where $\mathbf{f}_i \in \mathbb{R}^l$ encodes l morphological and texture features, and $\mathbf{c}_i \in \{0,1\}^6$ denotes the cell class encoded as a one-hot vector. We built the undirected, node-attributed graph $G = (\mathbf{H}, \mathbf{A})$, where $\mathbf{H} \in \mathbb{R}^{N\times(l+6)}$ is the feature matrix concatenation of each feature vector $\mathbf{h}_i$, and where $\mathbf{A} = [A_{ij}] \in \mathbb{R}^{N\times N}$ is the adjacency matrix. $[A_{ij}] = 1$ if there exits an edge connecting node i and node j, otherwise, $[A_{ij}] = 0$. We define a threshold distance r_0 and note $d_{E_{ij}}$ the Euclidean distance between the centroids of the cell nuclei in the WSI corresponding to nodes i and j in G.

If $d_{E_{ij}} < r_0$, then there is an edge between i and j and $[A_{ij}] = 1$. The resulting graph is shown in **Figure 1a**. The simplification and splitting steps are described in the following sections.

Graph simplification

We simplify the graph by first removing all nodes corresponding to cell nuclei that were detected by cSCC Hovernet but not classified (because of lower confidence). In addition, nodes of epithelial cells with no degree are more likely to be miss-classified, as epithelial cells are forming a compact group in the tissue. The nodes corresponding to these cells are also removed.

The graph is subsequently used for binary node classification, where nodes representing epithelial cells are classified as either tumor or healthy. Not all nodes are considered equally during training and inference. In particular, nodes located in the local neighborhood of tumor and healthy epithelial cells are expected to have a greater influence on the classification through message passing than nodes that are further away. To both reduce the computational complexity of the graph during training and simplify its structure —thereby limiting the number of contributing nodes in the aggregation of latent features during graph convolution— we remove nodes that are away from epithelial cells by a given number of graph edges.

We define nodes corresponding to tumor or healthy epithelial cells as anchor nodes around which the graph will be simplified. We retain all nodes that lie within a geodesic distance d_G (number of edges in a shortest path connecting 2 nodes) of at most k from at least one anchor node. We define k-max-hops as the maximum allowed geodesic distance between a cell and its nearest anchor node. We define the class indicator matrix $\mathbf{C} \in \{0,1\}^{N\times 6}$ where the i-th row $\mathbf{c}_i \in \{0,1\}^6$ encodes the class of node i. To specify which class is an anchor, we introduce a binary class-selection vector $\boldsymbol{s} \in \{0,1\}^6$ where $s_p = 1$ indicates that class p is selected as an anchor class. For each node $i \in \{1,\ldots,N\}$ we define the anchor indicator vector $\mathbf{a} \in \{0,1\}^N$ as:

$$a_i = \begin{cases} 1, & \text{if } (\mathbf{C}\boldsymbol{s})_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

The anchor indicator a_i equals 1 if and only if node i belongs to a class selected by $\boldsymbol{s}$ (here tumor and healthy epithelial). We finally introduce the node mask $\mathbf{m^k} \in \{0,1\}^N$ define as:

$$m_i^k = 1 \iff \min_{j:\, a_j=1} d_G(i,j) \le k, \qquad \forall i \in \{1,\ldots,N\}$$

d_G is computable for any pairs of nodes, but interestingly, by using the properties of the powers of the adjacency matrix one doesn't need to directly compute distances. Indeed A_{ij}^q is the number of walks of lengths q from node i to node j. From this we can deduce $d_G(i,j)$ as the smallest non negative q such as $A_{ij}^q > 0$. Another definition of $\mathbf{m}$ is:

$$m_i^k = \begin{cases} 1, & \text{if } \left[\left(\sum_{q=0}^{k} \mathbf{A}^q\right)\mathbf{a}\right]_i > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \forall i \in \{1,\ldots,N\}$$

After simplification of G we obtain the induced subgraph $G^{(k)} = (\mathbf{H^{(k)}}, \mathbf{A^{(k)}})$ with the induced adjacency matrix $\mathbf{A^{(k)}} \in \mathbb{R}^{N\times N}$ and masked features matrix

$\mathbf{H^{(k)}} \in \mathbb{R}^{N\times(l+6)}$ given by:

$$\mathbf{M^{(k)}} = \mathrm{diag}(\mathbf{m^k}),\ \mathbf{A^{(k)}} = \mathbf{M^{(k)}}\mathbf{A}\mathbf{M^{(k)}},\ \mathbf{H^{(k)}} = \mathbf{M^{(k)}}\mathbf{H} \tag{1}$$

Implementation details of WSI-Graph

From a WSI of cSCC skin cancer with annotated tumor regions we built and then simplified a graph of this tissue as described in previous sections. We also split the simplified graph into 100 non-overlapping subgraphs of similar sizes. To build the subgraphs we ran K-means algorithm on the large graph nodes' coordinates features to form K spatial clusters. We used these subgraphs to evaluate GNN's performance on node binary classification task. The process of graph building is depicted in **Figure 1a** where each node is colored according to its cell class, and spatially located using nucleus centroid coordinates. **Figure 1b** shows the graph edges. The threshold distance for edges is $r_0 = 50$ pixels corresponding to $r_0 \approx 11.5\,\mu\mathrm{m}$. This threshold ensures that neighboring nuclei within cells that are likely to interact or form structures together are connected by an edge, while not forming connections between nuclei that are far apart.

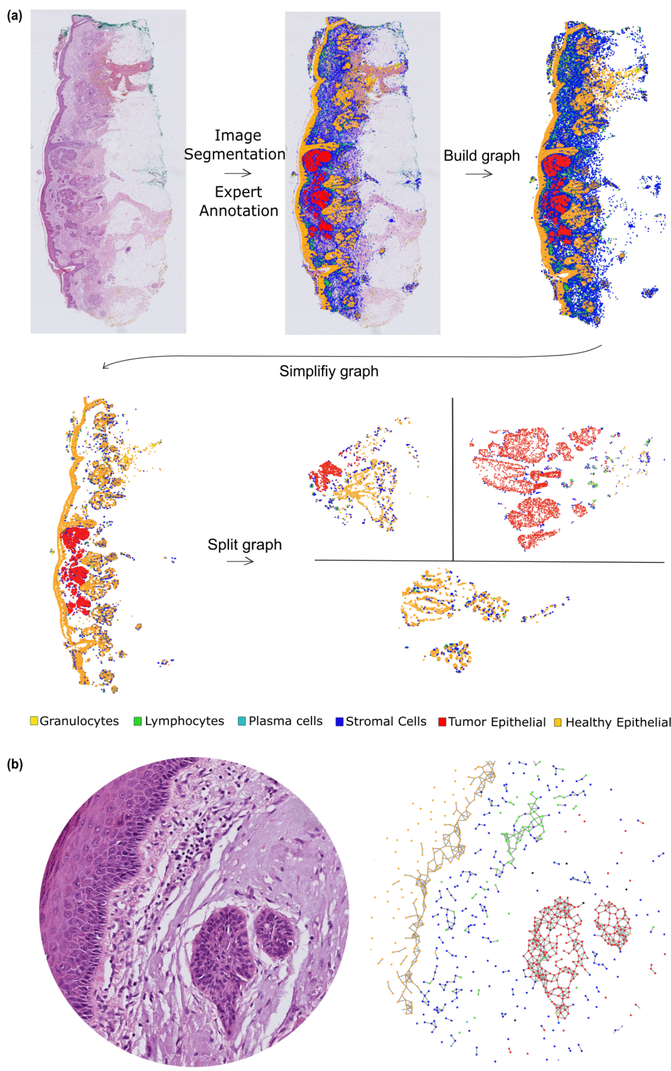
A comprehensive description of WSI-Graph before and after simplification is shown in **Figure 1c**. The 22 node features represent morphology attributes of the nucleus (7 features), the texture of the nucleus (7 features), a one-hot encoded vector of the nucleus class (6 features) and the nucleus centroid coordinates (2 features). The morphology features are: area, perimeter, eccentricity, solidity, major axis length, minor axis length and extent. The texture features are: roughness, contrast, dissimilarity, homogeneity, entropy, angular second moment and dispersion. Finally, we represent each class label as a one-hot encoded vector $\mathbf{c}_i \in \{0, 1\}^6$, where exactly one element equals 1, indicating the class membership, and all remaining elements are equal to 0.

Implementation details of TILE-Graphs

We collected 93 WSIs from 84 patients with cSCC skin cancer for which tumor regions were annotated by two experts (each expert having a subset of the data to annotate). These WSIs are a subset of already published TumSeg dataset [33], where the 93 WSIs kept here contain both tumor and healthy epithelial tissue. Contrary to TumSeg where the images were downsampled, the collected WSI are at 40x magnification. To generate TILE-Graphs we extracted two 2560x2560 pixels patches fully inside the annotated tumor regions and two 2560x2560 pixels patches containing healthy epithelial cells from each of the WSIs. We then built one graph per patch. **Figure 2a** shows the generation of 3 graphs from the dataset. Contrary to WSI-Graph dataset we did not simplify the graphs as they are smaller and all their complexity will be needed for classification. We kept the same threshold distance to generate edges as for WSI-Graph dataset: $r_0 = 50$ pixels corresponding to $r_0 \approx 11.5\,\mu\mathrm{m}$. A comprehensive description of TILE-Graphs is shown in **Figure 2b**. The node features are the same as in WSI-Graph.

Graph Transformer with linear complexity

In this work we perform binary classification of graph nodes as tumor or healthy epithelial cells. Standard Transformer graph neural networks represent state of art

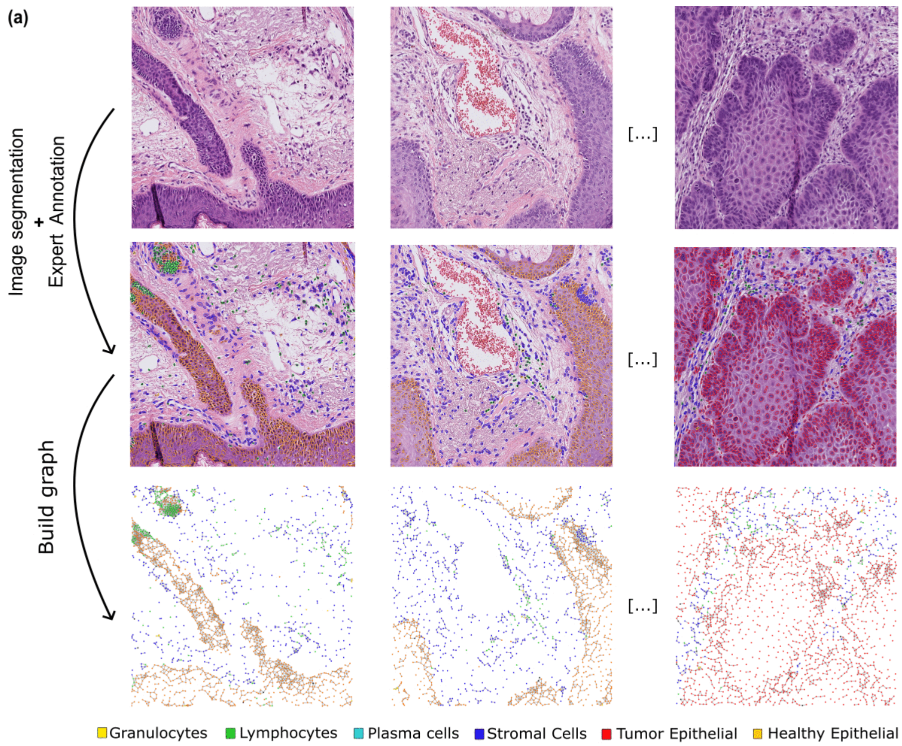


(c)

WSI-Graph	**w/o simplification**	**3-max-hops**
nodes	401,943	210,765
features/node	22	22
edges	535.878	381,660
unclassified cells	4,905	0
granulocytes	28,500	20,956
lymphocytes	40,127	6,743
plasma cells	2,746	939
stromal cells	140,565	24,691
tumor epithelial	96,738	83,620
healthy epithelial	88,362	73,816

Fig. 1: Description of WSI-Graph. (a) Steps to generate WSI-Graph. First a WSI from cSCC patient is segmented using SCC Hovernet and tumor regions are annotated by an expert to refine segmentation. From this segmentation map we build a graph, simplify it around anchor nodes and optionally split it with K-means on centroid coordinate features. **(b)** Zooming into the graph, edges generated with threshold distance $r_0 = 50$ pixels corresponding to $r_0 \approx 11.5\,\mu\text{m}$ are shown. **(c)** Number of edges, nodes, node features and instances of given cell classes before and after simplification (here $k = 3$ max-hops simplification).

(a)



(b)

TILE-Graphs	**w/o simplification**
graphs	372
tissue samples	93
patients	84
avg nodes/graph	1,539.70
features/node	22
avg edges/graph	1,950.15
unclassified cells	8,172
granulocytes	6,361
lymphocytes	67,208
plasma cells	10,309
stromal cells	182,530
tumor epithelial	192,041
healthy epithelial	106,146

Fig. 2: Description of TILE-Graphs. (a) Steps to generate TILE-Graphs. First, patches from cSCC patient sample extracted from tumor epithelial and healthy epithelial regions are segmented using SCC Hovernet. Then from these segmentation maps 372 graphs are built. **(b)** TILE-Graphs dataset statistics. It includes 372 patches from 93 samples from 84 patients. The resulting 372 graphs are then split keeping graphs of the same patients in the same split during cross-validation.

and are efficient in node classification [21,22,23], but are computationally intensive due to the quadratic complexity of the attention mechanism relative to number of nodes. This restricts their practical application to small graphs (e.g., on the order of hundreds of nodes) [23].

In our work we evaluated NodeFormer [25], DIFFormer [26] and SGFormer models [27] - Graph Transformers with linear computational complexity w.r.t number of nodes. In NodeFormer, the original softmax attention is approximated using stochastic kernel approximation, reducing the complexity to $\mathcal{O}(n)$. In DIFFormer, the exponential function in the softmax attention is replaced by its first-order Taylor expansion. This new attention layer can be efficiently computed using linear complexity thanks to re-ordering the matrix product. SGFormer is composed of a one-layer global attention and a shallow GNN network. Contrary to all-pair attention that incurs $\mathcal{O}(n^2)$ complexity, this simple global attention allows for $\mathcal{O}(n)$ complexity. In contrast to original implementation of these models, we developed an alternative training strategy which is described in the following sections.

Context-aware graph neural network classification of epithelial cells

We evaluated state of the art GNNs and Graph Transformers with linear complexity for large graphs on $k = 3$ max-hops simplified graph $G^{(3)}$ for epithelial node classification. Prior to training, continuous node features were standardized using z-score normalization computed from training set nodes and applied to both training and test sets. To prevent label leakage, the cell class features of epithelial nodes targeted for prediction were masked during training. Although the cell class feature was masked for epithelial nodes themselves, it remained available for neighboring nodes. Consequently, the graph neural network incorporates biologically meaningful contextual information, including neighboring cell class features, through message passing. The resulting epithelial node embeddings therefore depend on both intrinsic cellular features and features propagated from neighboring cells, enabling predictions based on the composition and spatial organization of the local tissue microenvironment.

In general form, node embeddings were computed as:

$$\mathbf{H^{(3)}}_{(l+1)} = \Phi_{(\mathrm{l})}\left(\mathbf{H^{(3)}}_{(l)}, \mathbf{A^{(3)}}\right) \tag{2}$$

where, $\mathbf{H^{(3)}}_{(l)} \in \mathbb{R}^{N \times d_l}$ denotes node embeddings at layer l of $G^{(3)}$, $\mathbf{A^{(3)}} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $G^{(3)}$, and $\Phi_{(\mathrm{l})}(\cdot)$ represents the architecture-specific propagation operator.

All architectures were adapted to binary node classification by replacing the output layer with a single linear unit producing one logit and corresponding probability per node:

$$z_i = w^\top h_i + b$$
$$\hat{y}_i = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}$$

where $h_i \in \mathbb{R}^d$ denotes the learned embedding of node i, $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are learnable parameters shared across nodes, $z_i \in \mathbb{R}$ is the predicted logit, $\hat{y}_i \in (0, 1)$ is the predicted probability, and $\sigma(\cdot)$ denotes the sigmoid function. Model parameters were optimized using binary cross-entropy loss computed exclusively over epithelial nodes:

$$\mathcal{L} = -\sum_{i \in V_{\mathrm{epi}}} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

where $y_i \in \{0, 1\}$ denotes the ground-truth label of node i and $V_{\mathrm{epi}} \subseteq V$ denotes the set of epithelial nodes. Finally, all model parameters were optimized using Adam algorithm [34] to minimize $\mathcal{L}$, with Adam parameters depending on model chosen.

Finally, to evaluate model performance on binary node classification we ran 3-folds cross-validation using 2 folds for training and 1 for testing. We trained for 200 epochs on one 80GB A100 NVIDIA GPU (Ampere micro-architecture) per model, without early stopping nor hyperparameter search. To train on WSI-Graph, SGFormer, NodeFormer, and DIFFormer used the hyperparameter configuration from their respective training on the large graph ogbn-proteins [28]. To train on TILE-Graph, these 3 models used the hyperparameter configuration from their respective training on the medium graph Cora [35]. We applied the default hyperparameters in all other GNNs as in [36].

Image-based approaches for epithelial cell classification

To compare GNNs' performance with image-based models, we created WSI-Graph baseline dataset consisting on the WSI used to build the graph and its segmentations and annotations. This dataset contained over 36,000 segmented patches of 256x256 pixels. We trained CellViT256 model for 200 epochs on one 80GB A100 NVIDIA GPU to segment and classify epithelial cells as tumor or healthy. We trained Hovernet for 2 distinct steps [37] of 50 epochs on two 80GB A100 NVIDIA GPUs following original implementation. Prior to training on baseline dataset, CellViT256 model was originally pretrained on 104 million 256×256px histological image patches from The Cancer Genome Atlas (TCGA) and Hovernet was originally pretrained on ImageNet [38] and the resulting weights were taken as initialization. We did not apply early stopping and used by default hyperparameter sets. We either trained the models on all the WSI patches or on patches containing at least 90% of epithelial cells. Keeping only detected cells, we calculated balanced accuracy on epithelial cell classification on 3-fold cross validation.

We also generated TILE-Graphs baseline dataset. It consists on 37,200 annotated patches generated from the 370 patches used to build the graphs of TILE-Graphs. We evaluated image-based models on TILE-Graphs baseline dataset following the same strategy as for WSI-Graph baseline dataset.

Experiments and Results

Healthy epithelial cells (also called non-neoplastic epithelial) and tumor epithelial (neoplastic epithelial) have very similar morphologies and are challenging to discriminate in cSCC WSIs. In practice, pathologists rely primarily on the global tissue architecture, overall composition, and on the surrounding cells to differentiate healthy from tumor regions. In this context, state of the art image-based model are performing poorly in classifying both cell types [33] as they process individual patches separately and lack information on the broader tissue structure.

Both convolutional network inspired models, such as Hovernet [37] or Transformer based model such as CellViT [39] have a limited input tile size they can process for training or inference that fits into GPU memory. These tiles represent only a small portion of the whole-slide image and therefore fail to preserve the broader tissue context, leading to reduced performance.

Comparison of graph-based and image-based approaches for epithelial cell classification

We compared the performance of linear-complexity Graph Transformers on large graphs, i.e SGFormer [27], NodeFormer [25], and DIFFormer [26] with that of conventional GNNs that can also scale to large graphs within a reasonable computational budget, including GCN [7], GAT [40], SGC [41], SGC-MLP [36], and SIGN [42]. To ensure consistency across models and reproducibility, we adapted all architectures to the binary node classification by replacing the final output layer with a single output unit producing one logit per node. We evaluated binary node classification performance of GNNs following 2 distinct strategies.

First, consistent with common practice [27,25,26] we randomly sampled all nodes from $k = 3$ max-hops simplified WSI-Graph (WSI-Graph$^{(3)}$) into 3 folds and applied the 3 folds cross-validation described previously. In this standard transductive evaluation protocol, randomly selected test nodes are often adjacent to training nodes, allowing their representations to be influenced by training data through neighborhood aggregation during message passing. This can lead to overly optimistic performance estimates when the goal is to assess generalization to independent or disjoint graph regions, in particular in biological tissues where cells of the same type are often grouped togehter.

To overcome this bias, we used 100 subgraphs originating from WSI-Graph$^{(3)}$ split and sample them into 3 folds. Each subgraph consists of non-overlapping nodes and edges from WSI-Graph$^{(3)}$. We ran 3-folds cross-validation and predicted all healthy epithelial and tumor cell classes in the test set in each run. This way testing on subgraphs instead of random nodes from one single large graph allows for an unbiased evaluation. The resulting balanced accuracy and standard error from both evaluations are shown in **Table 1**. Following previous findings on large graphs, SGFormer and DIFFormer are the best performing models.

We observed that state-of-the-art image-based models show lower accuracy compared to SGFormer and DIFFormer on the same dataset when represented in an alternative modality. Indeed, SGFormer and DIFFormer respectively yielded in 85.2 ± 1.5 and 85.1 ± 2.5 balanced accuracy ($\pm$ standard error) whereas best tested state of the art image-based approach CellViT256 yielded in 81.2 ± 3.0 balanced accuracy. We were unable to train CellViT-SAM-B model as, even with mixed-precision training, it could not fit into the memory of an 80GB A100 NVIDIA GPU with an acceptable batch size. Results are listed in **Table 2**. Interestingly, SGFormer was previously known as outperforming all GNNs of **Table 2** in node classification task [27], because common evaluation strategy is done under random nodes evaluation protocol. With subgraphs protocol evaluation, the performance is very similar to DIFFormer.

Method	**Subgraphs**	**Random Nodes**
DIFFormer	**85.2 ± 1.5**	91.1 ± 0.1
SGFormer	**85.1 ± 2.5**	**94.9± 0.2**
SIGN	80.5 ± 2.8	84.8 ± 0.6
GAT	80.4 ± 1.0	73.3 ± 2.0
SGCMLP	80.4 ± 2.7	79.3 ± 0.4
NodeFormer	79.0 ± 0.5	85.0 ± 1.4
SGC	76.8 ± 2.2	66.8 ± 3.1
GCN	70.4 ± 2.0	84.6 ± 1.5

Table 1: Comparison of GNN models on binary epithelial node classification performance on simplified WSI-Graph. Balanced accuracy (%) is reported as mean ± standard error over 3-fold cross-validation under subgraph-based and random node evaluation protocols and $k = 3$ max-hops simplification for the graph (WSI-Graph$^{(3)}$).

Method	**Training set**	**3-fold crossval**
Hovernet	all patches	73.7 ± 1.4
	epithelial only	79.3 ± 2.3
CellViT256	all patches	76.5 ± 1.9
	epithelial only	**81.2 ± 3.0**
CellViT-SAM-B	all patches	OOM
	epithelial only	OOM

Table 2: Comparison of image-based models on binary epithelial cell classification performance on WSI images Balanced accuracy (%) is reported as mean ± standard error over 3-fold cross-validation for different methods and training set configurations, using image-based representations..

WSI-Graph node features ablation

We conducted a feature ablation study on WSI-Graph$^{(3)}$ to evaluate the impact of node features on binary node classification performance. We evaluated SGFormer performance using the evaluation protocols described previously. Always keeping the nucleus coordinates as node features, we evaluate performance with different node features combinations. Z-score normalization was computed on the training set and applied to the test set using the same normalization parameters. In this feature ablation study, we evaluate the impact of such normalization on the performance. The results are reported in **Table 3** and show that removing any type of features and normalization leads to the poorest performance, highlighting their crucial contribution to model generalization.

WSI-Graph simplifications and impact on performance

To evaluate the impact of graph simplification of WSI-Graph on binary node classification performance, we progressively restricted the graph connectivity by limiting the maximum number of hops between nodes. This simplification reduces long-range connections and constrains information propagation during message passing. We evaluated SGFormer performance using the evaluation protocols described previously.

Node Features	z-score norm	Subgraphs	Random Nodes
morphology	no	67.8 ± 3.6	92.6 ± 0.3
morphology	yes	79.6 ± 2.4	94.0 ± 0.5
morphology & texture	no	70.6 ± 9.0	94.0 ± 0.9
morphology & texture	yes	84.2 ± 1.6	94.3 ± 0.6
morphology & cell class	no	73.6 ± 3.7	93.7 ± 0.5
morphology & cell class	yes	84.0 ± 2.8	94.5 ± 0.4
morphology & texture & cell class	no	71.4 ± 8.4	92.8 ± 1.1
morphology & texture & cell class	yes	**85.1 ± 2.5**	**94.9 ± 0.2**

Table 3: Impact of node features on binary node classification performance on simplified WSI-Graph. Balanced accuracy (%) is reported as mean ± standard error over 3-fold cross-validation for different node features under subgraph and random node evaluation protocols with SGFormer model and WSI-Graph$^{(3)}$.

Graph Simplifications	Subgraphs	Random Nodes
No simplification	82.2 ± 2.9	94.6 ± 0.1
50 max-hops	83.3 ± 3.3	94.9 ± 0.4
10 max-hops	**86.6 ± 2.2**	95.0 ± 0.2
5 max-hops	82.5 ± 1.9	94.6 ± 0.5
4 max-hops	82.1 ± 4.3	95.0 ± 0.4
3 max-hops	85.1 ± 2.5	94.9 ± 0.2
2 max-hops	81.8 ± 0.8	**95.4 ± 0.3**
1 max-hops	81.1 ± 3.9	94.7 ± 0.6

Table 4: Impact of graph simplification on binary node classification performance on simplified WSI-Graph. Balanced accuracy (%) is reported as mean ± standard error over 3-fold cross-validation for different maximum hop thresholds under subgraph and random node evaluation protocols with SGFormer model.

As shown in **Table 4**, graph simplification has a noticeable impact on performance under the subgraph-based evaluation protocol. The best balanced accuracy is achieved with a 10-hop simplification (86.6 ± 2.2), while both more restrictive configurations (1–2 max-hops) and the absence of simplification result in lower performance. This suggests that an intermediate level of connectivity provides sufficient contextual information for accurate classification while avoiding the propagation of less relevant signals. In contrast, performance under random node splits remains largely stable across all simplification levels. These results highlight that local graph structure contains most of the useful information for node classification, and that moderate graph simplification can improve robustness while reducing graph complexity.

A max-hops simplification with $k = 3$ provides a good compromise between graph sparsity and model performance; therefore, this simplification was adopted throughout this work.

Method	3-fold crossval
CellViT256	78.1 ± 0.5
DIFFormer	**83.6 ± 1.9**
NodeFormer	69.4 ± 3.3
SGCMLP	66.4 ± 1.1
SIGN	66.3 ± 1.0
SGC	63.8 ± 1.2
GCN	61.9 ± 0.9
GAT	61.6 ± 1.4
SGFormer	61.0 ± 0.8

Table 5: Comparison of graph-based and image-based models on binary epithelial cell classification performance on TILE-Graphs and baseline dataset. Balanced accuracy (%) is reported as mean ± standard error over 3-fold cross-validation for different methods and training set configurations.

Comparison of graph-based and image-based approaches on multiple patients dataset

We previously compared the performance of GNN models and state of the art image-based model for epithelial cell and node classification using a large graph derived from a WSI of a single patient. Evaluating CellViT on complete WSIs from different patients is computationally prohibitive and thus prevented a systematic training and evaluation on an entire WSIs dataset. We therefore devised another approach to enable the evaluation of our approach across a larger number of patients. Following this goal we used TILE-Graphs and its baseline counterpart to evaluate both CellViT256 and GNNs in the same 3-fold cross validation fashion as previously described. Samples from same patients were distributed into the same fold, and all models were tested using the same splits.

Our results show that DIFFormer outperforms vision Transformer CellViT256 on the binary classification of epithelial cells as healthy or tumor (**Table** 5). DIFFormer model resulted in classification with 83.6 ± 1.9 balanced accuracy on 3 fold cross-validation and CellViT256 with 78.1 ± 0.5 balanced accuracy with the same folds. Interestingly, SGFormer performed poorly on smaller graphs, even with adjusted hyperparameters. Indeed, its very light Graph Transformer architecture probably resulted in attention focused on very few nodes, and in the case of smaller graphs, these nodes are not representative enough of the overall graph structure.

Discussion

WSIs of cancer samples contain rich information for medical diagnosis and for gaining insights into tumor biology. These images contain thousands to million of cells of different types whose automated segmentation and classification can help practitioners in comprehensive assessment of sample. Due to the large size of WSIs segmentation models are trained on small image patches. Within small patches, the discrimination between healthy and tumor epithelial is difficult as healthy epithelial cells and tumor epithelial cells have very similar morphologies. For this reason, pathologists rely on the global tissue architecture, overall composition, and on the surrounding cells to differentiate healthy tissue from tumor regions. To address this limitation of the WSI segmentation models, we explored, the use of GNNs, for incorporation of broader tissue context to differentiate healthy epithelial from tumor epithelial cells. We showed that scalable Graph Transformer architectures outperform image-based approaches on this task.

In this work we show that creating graphs from medical images, where each node represent a single cell, and using scalable Graph Transformers for cell classification outperforms the image-based approaches. We started by evaluating several graph-based methods and image-based models on the same annotated WSI. Graph Transformer models SGFormer and DIFFormer resulted in 85.2 ± 1.5 and 85.1 ± 2.5 balanced accuracy, respectively ($\pm$ standard error) on 3 fold cross-validation whereas best tested state of the art image-based approach CellViT256 resulted in 81.2 ± 3.0 balanced accuracy. To further evaluate both graph-based and image-based models on several patients, we collected TILE-Graphs dataset. This dataset is composed of 372 patches from 93 WSIs of 84 different patients. The patches are converted into smaller graphs and used to train GNNs. Importantly, a graph-based approach, DIFFormer model, showed a classification performance of 83.6 ± 1.9 balanced accuracy on while state of the art image-based approach CellViT256 reached 78.1 ± 0.5 balanced accuracy. The higher accuracy of graph models in classifying the two cell types suggests that broader tissue context, encoded in graphs is important for this task.

In addition to performance, a very important advantage of graph based models is their lower computational cost. Graph representation of tissues is computationally lighter compared to raw images and can be manipulated with ease. On the TILE-Graphs image baseline dataset, training CellViT256 required approximately five days to complete a single run of 3-fold cross-validation on a 80GB A100 NVIDIA GPU but only 31 min 59 s for graph architecture such as DIFFormer. Graph-based approaches for cell classification represents therefore a powerful and efficient way for WSI analysis compared to traditional computer vision methods.

In the present study, tissue was represented as a simple undirected graph in which nodes were associated with handcrafted morphological and spatial features. Future approaches could leverage pretrained foundation models for cancer whole-slide images, such as VOLTA [43] to derive informative learned representations of cellular phenotypes and incorporate them as node attributes. Furthermore, more expressive graph formalisms, including multi-graphs or hypergraphs, could be explored to capture higher-order and multi-relational interactions between cells through alternative edge or hyperedge construction strategies.

Data availability

A Zenodo repository containing WSI-Graph and TILE-Graphs datasets as well as its image baseline datasets counterparts will be openly available at the latest at publication date of this paper in a peer-reviewed journal.

Code availability

The github repository linked to this project will be openly available at the latest at publication date of this paper in a peer-reviewed journal.

Competing Interest

Authors declare no conflicts of interests.

Author Contributions

Conceptualization: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Data Curation: Lucas Sancéré.
Formal Analysis: Lucas Sancéré.
Funding Acquisition: Katarzyna Bozek.
Investigation: Lucas Sancéré.
Methodology: Lucas Sancéré.
Project Administration: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Resources: Lucas Sancéré.
Software: Lucas Sancéré.
Supervision: Noémie Moreau, Katarzyna Bozek.
Validation: Lucas Sancéré.
Visualization: Lucas Sancéré.
Writing – Original Draft Preparation: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.
Writing – Review & Editing: Lucas Sancéré, Noémie Moreau, Katarzyna Bozek.

References

1. Wittekind, D. H. Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. value and limitations of the hematoxylin-eosin stain. *Biotechnic & Histochemistry* **78**, 261 – 270 (2003).
2. Coudray, N. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
3. Campanella, G. *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
4. Shao, Z. *et al.* TransMIL: Transformer based correlated multiple instance learning for whole slide image classification (2021). 2106.00908.
5. Chaurasia, A. K., Harris, H. C., Toohey, P. W. & Hewitt, A. W. A generalised vision transformer-based self-supervised model for diagnosing and grading prostate cancer using histological images. *Prostate Cancer Prostatic Dis.* **28**, 918–926 (2025).

6. Pisula, J. I. & Bozek, K. Efficient WSI classification with sequence reduction and transformers pretrained on text. *Sci. Rep.* **15**, 5612 (2025).
7. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations* (2017). URL https://openreview.net/forum?id=SJU4ayYgl.
8. Chen, R. J. *et al.* Whole slide images are 2D point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Lecture Notes in Computer Science, 339–349 (Springer International Publishing, Cham, 2021).
9. Liu, P., Ji, L., Ye, F. & Fu, B. GraphLSurv: A scalable survival prediction network with adaptive and sparse structure learning for histopathological whole-slide images. *Comput. Methods Programs Biomed.* **231**, 107433 (2023).
10. Zhao, L. *et al.* CoADS: Cross attention based dual-space graph network for survival prediction of lung cancer using whole slide images. *Comput. Methods Programs Biomed.* **236**, 107559 (2023).
11. Zheng, Y. *et al.* A graph-transformer for whole slide image classification. *IEEE Trans. Med. Imaging* **41**, 3003–3015 (2022).
12. Shi, Z., Zhang, J., Kong, J. & Wang, F. Integrative Graph-Transformer Framework for Histopathology Whole Slide Image Representation and Classification . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. LNCS 15011 (Springer Nature Switzerland, 2024).
13. Ramanathan, V., Pati, P., McNeil, M. & Martel, A. L. Ensemble of Prior-guided Expert Graph Models for Survival Prediction in Digital Pathology . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. LNCS 15005 (Springer Nature Switzerland, 2024).
14. Jaume, G., Pati, P., Anklin, V., Foncubierta, A. & Gabrani, M. Histocartography: A toolkit for graph analytics in digital pathology. In Atzori, M. *et al.* (eds.) *Proceedings of the MICCAI Workshop on Computational Pathology*, vol. 156 of *Proceedings of Machine Learning Research*, 117–128 (PMLR, 2021). URL https://proceedings.mlr.press/v156/jaume21a.html.
15. Lou, W. *et al.* Structure embedded nucleus classification for histopathology images (2023). 2302.11416.
16. Hassan, T. *et al.* Nucleus classification in histology images using message passing network. *Med. Image Anal.* **79**, 102480 (2022).
17. Lou, W., Li, G., Wan, X. & Li, H. Cell graph transformer for nuclei classification. *Proc. Conf. AAAI Artif. Intell.* **38**, 3873–3881 (2024).
18. Javed, S. *et al.* Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Med. Image Anal.* **63**, 101696 (2020).
19. Zhou, Y. *et al.* CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images . In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 388–398 (IEEE Computer Society, Los Alamitos, CA, USA, 2019). URL https://doi.ieeecomputersociety.org/10.1109/ICCVW.2019.00050.
20. Pati, P. *et al.* Hierarchical graph representations in digital pathology. *Med. Image Anal.* **75**, 102264 (2022).
21. Kim, J. *et al.* Pure transformers are powerful graph learners. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22 (Curran Associates Inc., Red Hook, NY, USA, 2022).
22. Ying, C. *et al.* Do transformers really perform bad for graph representation? In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21 (Curran Associates Inc., Red Hook, NY, USA, 2021).
23. Rampášek, L. *et al.* Recipe for a general, powerful, scalable graph transformer. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22 (Curran Associates Inc., Red Hook, NY, USA, 2022).
24. Xiong, Y. *et al.* Nyströmformer: A nystöm-based algorithm for approximating self-attention. *Proc. Conf. AAAI Artif. Intell.* **35**, 14138–14148 (2021).

25. Wu, Q., Zhao, W., Li, Z., Wipf, D. & Yan, J. Nodeformer: A scalable graph structure learning transformer for node classification. In *Advances in Neural Information Processing Systems (NeurIPS)* (2022).
26. Wu, Q. *et al.* Difformer: Scalable (graph) transformers induced by energy constrained diffusion. In *International Conference on Learning Representations (ICLR)* (2023).
27. Wu, Q. *et al.* Sgformer: Simplifying and empowering transformers for large-graph representations. In *Advances in Neural Information Processing Systems (NeurIPS)* (2023).
28. Hu, W. *et al.* Open graph benchmark: datasets for machine learning on graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20 (Curran Associates Inc., Red Hook, NY, USA, 2020).
29. Jin, W. *et al.* Amazon-m2: a multilingual multi-locale shopping session dataset for recommendation and text generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23 (Curran Associates Inc., Red Hook, NY, USA, 2023).
30. Lubos Takac, M. Z. Data analysis in public social networks (2012). URL https://api.semanticscholar.org/CorpusID:18659578.
31. Guo, A., Liu, X., Li, H., Cheng, W. & Song, Y. The global, regional, national burden of cutaneous squamous cell carcinoma (1990–2019) and predictions to 2035. *Eur. J. Cancer Care (Engl.)* **2023**, 1–8 (2023).
32. JY, H., Y, H. & ML, R. Squamous cell skin cancer. *StatPearls. Treasure Island (FL): StatPearls Publishing* (2024).
33. Sancéré, L. *et al.* Histo-Miner: Deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma. *PLoS Comput. Biol.* **22**, e1013907 (2026).
34. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2014).
35. Sen, P. *et al.* Collective classification in network data. *AI Magazine* **29**, 93–106 (2008).
36. Wu, Q. *et al.* Sgformer: Simplifying and empowering transformers for large-graph representations. https://github.com/qitianwu/SGFormer/tree/3578e101c701491ce068bf26a9b029d2134903be (2023). GitHub repository, commit 3578e10.
37. Graham, S. *et al.* Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis* **58**, 101563 (2018).
38. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009).
39. Hörst, F. *et al.* Cellvit: Vision transformers for precise cell segmentation and classification. *Medical image analysis* **94**, 103143 (2023).
40. Veličković, P. *et al.* Graph attention networks. In *International Conference on Learning Representations* (2018). URL https://openreview.net/forum?id=rJXMpikCZ.
41. Wu, F. *et al.* Simplifying graph convolutional networks. In Chaudhuri, K. & Salakhutdinov, R. (eds.) *Proceedings of the 36th International Conference on Machine Learning*, vol. 97 of *Proceedings of Machine Learning Research*, 6861–6871 (PMLR, 2019). URL https://proceedings.mlr.press/v97/wu19e.html.
42. Frasca, F. *et al.* SIGN: Scalable inception graph neural networks (2020). 2004.11198.
43. Nakhli, R. *et al.* VOLTA: an environment-aware contrastive cell representation learning for histopathology. *Nat. Commun.* **15**, 3942 (2024).

Chapter 5
Conclusion

5.1 Projects code

5.1.1 Open-Source code for reproducible research

Reproducibility is a main concern in academic research. A 2016 Nature's survey revealed that more than 70% of the 1,576 researchers (from all fields) that answered the survey have tried and failed to reproduce another scientist's experiments. More than half have failed to reproduce their own experiments [53]. Computer science as a whole is one of the academic field where reproducibility of the work is the easiest as long as the codes generated to produce the research results are openly available. In biology and physics, experimental outcomes often depend on complex laboratory protocols, specialized instruments, and environmental conditions that are difficult to replicate exactly. Additionally, natural variability in biological samples or physical systems can introduce unavoidable differences between experiments. In contrast, computer science experiments are executed in standardized software environments and typically rely on widely available hardware. Reproducibility is mostly limited by the access to GPU resources that are increasingly more expensive, but rising number of works tend to reduce computation budget of machine learning research experiments. An analysis of 3,700 ICLR papers conducted in 2025 (ICLR-2025 Paper Digest) shows that Efficiency, Compression & Scaling is now the third biggest machine learning research topic. It includes techniques for making large models efficient, memory optimization, and understanding scaling laws.

Unfortunately, only average 19.5% of the papers accepted to top-tier machine learning conferences in 2024 provide their code implementations [54]. Even when available, it is often not working. A 2021 large-scale study on research code quality showed that 74% of over 9000 R files tested from published research code failed to complete without error in the initial execution, while 56% failed when code cleaning was applied [55]. Additionally, code is often not reviewed in the revision process of applied machine learning journals. Semmelrock et al. [56] discussed barriers in machine learning reproducibility. They find that pressure on researchers to publish quickly prevent them to polish and work on the code and then decreases their willingness to release it, as it is often only optional.

This work provides tools aimed at ensuring the reproducibility of our findings and supporting future research applications. Concerning "Histo-Miner: Deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma" [1] and "Context-aware Skin Cancer Epithelial Cell Classification with Scalable Graph Transformers" [3] we are solely responsible for the code. For "Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma" [2] first author Juan I Pisula is responsible for the project code.

5.1.2 Histo-Miner code

Histo-Miner code is available at https://github.com/bozeklab/histo-miner.

The README file consists on the main documentation, and an end-to-end example is provided to help new user get started. To conceive the code and repository we followed Good Practices recommendations for research code such as Hastings et al. 2014 [57] and Wilson et al. 2016 [58].

5.1.3 Scalable Graph Transformers code

As stated, the github repository linked to this project will be openly available at the latest at publication date of the paper in a peer-reviewed journal. At the date of publication of this thesis manuscript, still some work is needed prior to code release. Future work should focus on comprehensively describing the repository and adding installation commands for the conda environment. Nevertheless, we can here describe the project code in its current state and list the required work for future release.

SGFormer code repository [59] with MIT license served as a base for our implementation. Our repository is organized with the following modules:

- models: we adapted all implemented GNNs models from the original repository for the task of binary node classification. We added implementation for masked training, subgraphs evaluation protocol and cross-validation as developed in the paper.

- dataset-tools: new module used to generate a cell graph from annotated WSIs or patches. Optionally, this module can be used to simplify or split the graphs following the algorithms described in the paper. Additionally a script can be run to visualize generated graphs, where centroid coordinates and cell class features of each nodes are used to plot the graph in 2D with node colored based on cell class.

- baseline-tools: new module used to generate image-based baseline dataset to train CellViT and Hovernet. Indeed, annotations and metadata must be in a CellViT and Hovernet compatible format to allow for training and inference.

- configs: in the original repository all parameters where given via command line while running the code. We created configs files compatible with Hydra [60] to improve both readability and reproducibility.

- utils: new module to store all utils function for both graph generation and GNNs training.

The code rely on PyTorch [61] and PyTorch Geometric [62] for the machine learning models.

Prior to code publication, future work should focus on:

- Providing a comprehensive README file to help user run the code (similar to Histo-Miner README),
- Provide a complete end-to-end example for user,
- Improve readability and doc-strings of all functions and scripts,
- Provide conda environment file and installation instructions.

5.2 Summary and Outlook

In this work we implemented Histo-Miner, a pipeline to detect, segment, and classify cells from cSCC WSIs. The pipeline is also designed to be adapted for other cancers WSIs. From these segmentations, tissue features are calculated and can be used for downstream tasks. One example is the study of immunotherapy treatment response. It revealed that percentages of lymphocytes, the granulocyte to lymphocyte ratio in tumor vicinity and the distances between granulocytes and plasma cells in tumors are predictive features for therapy response. We applied this method on a dataset from 3 clinical centers to study most representative WSI patches for cSCC disease progression classification and showed that non-progressive tumors maintained structural homogeneity, whereas progressive phenotypes demonstrated enhanced spatial intermingling between neoplastic cells and neighboring non-malignant population. Finally, we represented WSIs as cell graphs and use Graph Transformers model with linear complexity to improve on epithelial cell classification accuracy and model training and evaluation time.

Further work could continue on these approaches on several different ways. One could be improving architectures of the neural networks. For now Transformer architecture is the gold standard and is applied to mostly all data modalities. In computer vision there are both Vision Transformer (image-based) and Graph Transformers (graph-based). Solodskikh et al. 2023 [63] introduced Integral Neural Networks, a new family of deep neural networks based on new continuous layer representation, Albert Gu and Tri Dao 2024 [64] introduced new state space model architecture Mamba. Next research could either apply these models to medical data or create new architectures. Another aspect could be to adapt the architecture to the medical data itself instead of looking for performing foundational models. This could take the form of Neural Architecture Search (NAS) with a search space that is tailored to a given medical task. Kuş et al. 2025 [65] compared their own two NAS methods, PBC-NAS and BioNAS, across multiple biomedical image classification tasks.

Improving NAS algorithms for biomedical tasks seems to be a promising future direction. Large and open datasets for academia is also a key element to pre-train performing models. New directions of research must include initiative to generate and share medical data, at a multi-center scale. Recent work are going this direction: in the work of Hu et al. 2024 [66], researcher collected 73 different medical datasets including 12 different modalities and covering more than 20 distinct anatomical regions.

Bibliography

[1] Sancéré, L. *et al.* Histo-Miner: Deep learning based tissue features extraction pipeline from H&E whole slide images of cutaneous squamous cell carcinoma. *PLoS Comput. Biol.* **22**, e1013907 (2026).

[2] Pisula, J. I. *et al.* Explainable, federated deep learning model predicts disease progression risk of cutaneous squamous cell carcinoma. *NPJ Precis. Oncol.* **9**, 205 (2025).

[3] Sancéré, L., Moreau, N. & Bozek, K. Context-aware skin cancer epithelial cell classification with scalable graph transformers. *arXiv preprint* (2026). 2602.15783.

[4] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

[5] Guo, A., Liu, X., Li, H., Cheng, W. & Song, Y. The global, regional, national burden of cutaneous squamous cell carcinoma (1990–2019) and predictions to 2035. *Eur. J. Cancer Care (Engl.)* **2023**, 1–8 (2023).

[6] Zeng, L. *et al.* Advancements in nanoparticle-based treatment approaches for skin cancer therapy. *Mol. Cancer* **22**, 10 (2023).

[7] Nuovo, G. J. 6 - the basics of histologic interpretations of tissues. In *In Situ Molecular Pathology and Co-Expression Analyses*, 167–196 (Academic Press, San Diego, 2013). URL https://www.sciencedirect.com/science/ARTICLE/pii/B9780124159440000061.

[8] Voiculescu, V. *et al.* From normal skin to squamous cell carcinoma: A quest for novel biomarkers. *Dis. Markers* **2016**, 4517492 (2016).

[9] Pastila, R. Effects of ultraviolet radiation on skin cell proteome. In *Advances in Experimental Medicine and Biology*, 121–127 (Springer Netherlands, Dordrecht, 2013).

[10] Muller, H. K. & Woods, G. M. Ultraviolet radiation effects on the proteome of skin cells. In *Advances in Experimental Medicine and Biology*, 111–119 (Springer Netherlands, Dordrecht, 2013).

[11] Ouhtit, A., Muller, H. K., Gorny, A. & Ananthaswamy, H. N. UVB-induced experimental carcinogenesis: dysregulation of apoptosis and p53 signalling pathway. *Redox Rep.* **5**, 128–129 (2000).

[12] Ouhtit, A. *et al.* Temporal events in skin injury and the early adaptive responses in ultraviolet-irradiated mouse skin. *Am. J. Pathol.* **156**, 201–207 (2000).

[13] Jiang, R., Fritz, M. & Que, S. K. T. Cutaneous squamous cell carcinoma: An updated review. *Cancers (Basel)* **16**, 1800 (2024).

[14] Wittekind, D. H. Traditional staining for routine diagnostic pathology including the role of tannic acid. 1. value and limitations of the hematoxylin-eosin stain. *Biotechnic & Histochemistry* **78**, 261 – 270 (2003).

[15] Peckham, M., Knibbs, A. & Paxton, S. The histology guide: Credits. https://www.histology.leeds.ac.uk/credits.php. Accessed: 2025-12-02.

[16] Reipert, S., Kotisch, H., Wysoudil, B. & Wiche, G. Rapid microwave fixation of cell monolayers preserves microtubule-associated cell structures. *J. Histochem. Cytochem.* **56**, 697–709 (2008).

[17] Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

[18] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, Lecture notes in computer science, 234–241 (Springer International Publishing, Cham, 2015).

[19] Cardoso, M. J. *et al.* MONAI: An open-source framework for deep learning in healthcare. *arXiv preprint* (2022). 2211.02701.

[20] Alemi Koohbanani, N., Jahanifar, M., Zamani Tajadin, N. & Rajpoot, N. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Med. Image Anal.* **65**, 101771 (2020).

[21] Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nature Medicine* **30**, 850 – 862 (2024). URL https://api.semanticscholar.org/CorpusID:268534221.

[22] Chen, P. *et al.* WsiCaption: Multiple Instance Generation of Pathology Reports for Gigapixel Whole-Slide Images . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. LNCS 15004 (Springer Nature Switzerland, 2024).

[23] Saldanha, O. L. *et al.* Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology. *NPJ Precis. Oncol.* **7**, 35 (2023).

[24] Graham, S. *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019).

[25] Hörst, F. *et al.* CellViT: Vision transformers for precise cell segmentation and classification. *Med. Image Anal.* **94**, 103143 (2024).

[26] Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255 (2009).

[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016).

[28] Tan, M. & Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv preprint* (2019). 1905.11946.

[29] Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2021). URL https://openreview.net/forum?id=YicbFdNTTy.

[30] Coudray, N. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).

[31] Luo, W. *et al.* Frequency-based convolutional neural network for efficient segmentation of histopathology whole slide images. In *Lecture Notes in Computer Science*, Lecture notes in computer science, 584–596 (Springer International Publishing, Cham, 2021).

[32] Chen, R. J. *et al.* Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16123–16134 (2022).

[33] Zhou, B. *et al.* Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision* **127**, 302–321 (2019).

[34] Mottaghi, R. *et al.* The role of context for object detection and semantic segmentation in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014).

[35] He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. In *Computer Vision – ECCV 2016*, Lecture notes in computer science, 630–645 (Springer International Publishing, Cham, 2016).

[36] Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. *arXiv preprint* (2016). 1608.06993.

[37] Strudel, R., Garcia, R., Laptev, I. & Schmid, C. Segmenter: Transformer for semantic segmentation. *arXiv preprint* (2021). 2105.05633.

[38] Khan, S. *et al.* Transformers in vision: A survey. *ACM Comput. Surv.* **54**, 1–41 (2022).

[39] Vaswani, A. *et al.* Attention is all you need. *arXiv preprint* (2017). 1706.03762.

[40] Thickstun, J. The transformer model in equations. In *The Transformer Model in Equations* (2020). URL https://api.semanticscholar.org/CorpusID:216559335.

[41] Kim, J. *et al.* Pure transformers are powerful graph learners. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22 (Curran Associates Inc., Red Hook, NY, USA, 2022).

[42] Ying, C. *et al.* Do transformers really perform bad for graph representation? In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21 (Curran Associates Inc., Red Hook, NY, USA, 2021).

[43] Hu, W. *et al.* Open graph benchmark: datasets for machine learning on graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20 (Curran Associates Inc., Red Hook, NY, USA, 2020).

[44] Jin, W. *et al.* Amazon-m2: a multilingual multi-locale shopping session dataset for recommendation and text generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23 (Curran Associates Inc., Red Hook, NY, USA, 2023).

[45] Lubos Takac, M. Z. Data analysis in public social networks (2012). URL https://api.semanticscholar.org/CorpusID:18659578.

[46] Wu, Q., Zhao, W., Li, Z., Wipf, D. & Yan, J. Nodeformer: A scalable graph structure learning transformer for node classification. In *Advances in Neural Information Processing Systems (NeurIPS)* (2022).

[47] Wu, Q. *et al.* Difformer: Scalable (graph) transformers induced by energy constrained diffusion. In *International Conference on Learning Representations (ICLR)* (2023).

[48] Wu, Q. *et al.* Sgformer: Simplifying and empowering transformers for large-graph representations. In *Advances in Neural Information Processing Systems (NeurIPS)* (2023).

[49] Wu, Q. How to build graph transformers with o(n) complexity (2023-04-19) (accessed 2026-02-18). URL https://towardsdatascience.com/how-to-build-graph-transformers-with-o-n-complexity-d507e103d30a/.

[50] Rahimi, A. & Recht, B. Random features for large-scale kernel machines. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, NIPS'07, 1177–1184 (Curran Associates Inc., Red Hook, NY, USA, 2007).

[51] McMahan, H. B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. *arXiv preprint* (2023). URL https://arxiv.org/abs/1602.05629. 1602.05629.

[52] Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *International Conference on Machine Learning* (2017). URL https://api.semanticscholar.org/CorpusID:16747630.

[53] Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).

[54] Seo, M., Baek, J., Lee, S. & Hwang, S. J. Paper2code: Automating code generation from scientific papers in machine learning. In *The Fourteenth International Conference on Learning Representations* (2026). URL https://openreview.net/forum?id=3DcaUTjdKc.

[55] Trisovic, A., Lau, M. K., Pasquier, T. & Crosas, M. A large-scale study on research code quality and execution. *Sci. Data* **9**, 60 (2022).

[56] Semmelrock, H. *et al.* Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Mag.* **46** (2025).

[57] Hastings, J., Haug, K. & Steinbeck, C. Ten recommendations for software engineering in research. *Gigascience* **3**, 31 (2014).

[58] Wilson, G. *et al.* Good enough practices in scientific computing. *arXiv preprint* (2016). 1609.00037.

[59] Wu, Q. *et al.* Sgformer: Simplifying and empowering transformers for large-graph representations. https://github.com/qitianwu/SGFormer/tree/3578e101c701491ce068bf26a9b029d2134903be (2023). GitHub repository, commit 3578e10.

[60] Yadan, O. Hydra - a framework for elegantly configuring complex applications. Github (2019). URL https://github.com/facebookresearch/hydra.

[61] Ansel, J. *et al.* PyTorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, 929–947 (ACM, New York, NY, USA, 2024).

[62] Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).

[63] Solodskikh, K. *et al.* Integral neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16113–16122 (2023).

[64] Gu, A. & Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling* (2024). URL https://openreview.net/forum?id=tEYskw1VY2.

[65] Kuş, Z., Aydin, M., Kiraz, B. & Kiraz, A. Neural architecture search for biomedical image classification: A comparative study across data modalities. *Artif. Intell. Med.* **160**, 103064 (2025).

[66] Hu, Y. *et al.* Omnimedvqa: A new large-scale comprehensive evaluation benchmark for medical lvlm. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22170–22183 (2024).