

Predictive immune biomarkers in renal cancer

Sean Klein

University Tübingen

2023-09-05

Introduction

Overview

- Aim: Detection of novel biomarkers in renal cell carcinoma (RCC) patients
- Multimodal neural networks
- Available data
 - Whole Slide Images (WSI)
 - Gene Expression counts (Subcohort)

Motivation

- RCC is a complex disease
 - No well-defined heterogeneity
 - 40% of patients treated with surgery experience recurrence
- Deep learning promises to better model this complexity
- Unimodal networks do not provide the “full picture”
- Multimodal networks promise better results

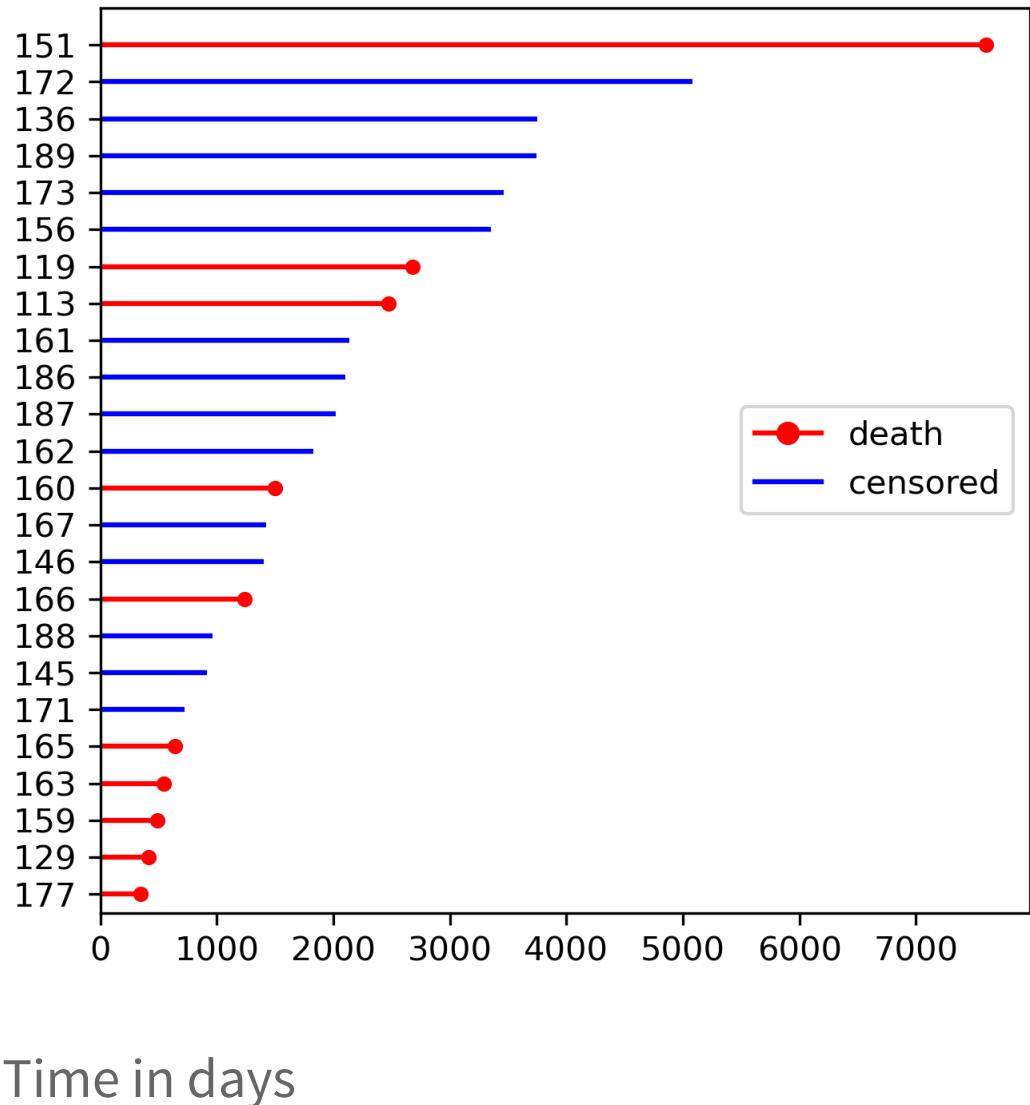
Background

Survival Analysis

- Expected duration until an event occurs
 - Time-to-Event data
 - Given: Event has not yet occurred
- Survival function:
 - Probability that event has not yet occurred at time t
- Hazard function:
 - Probability of event occurring at time t divided by δt
 - Trends to 0 with shrinking δt , hazard $h(t) \in [0, \infty]$

Censored Data

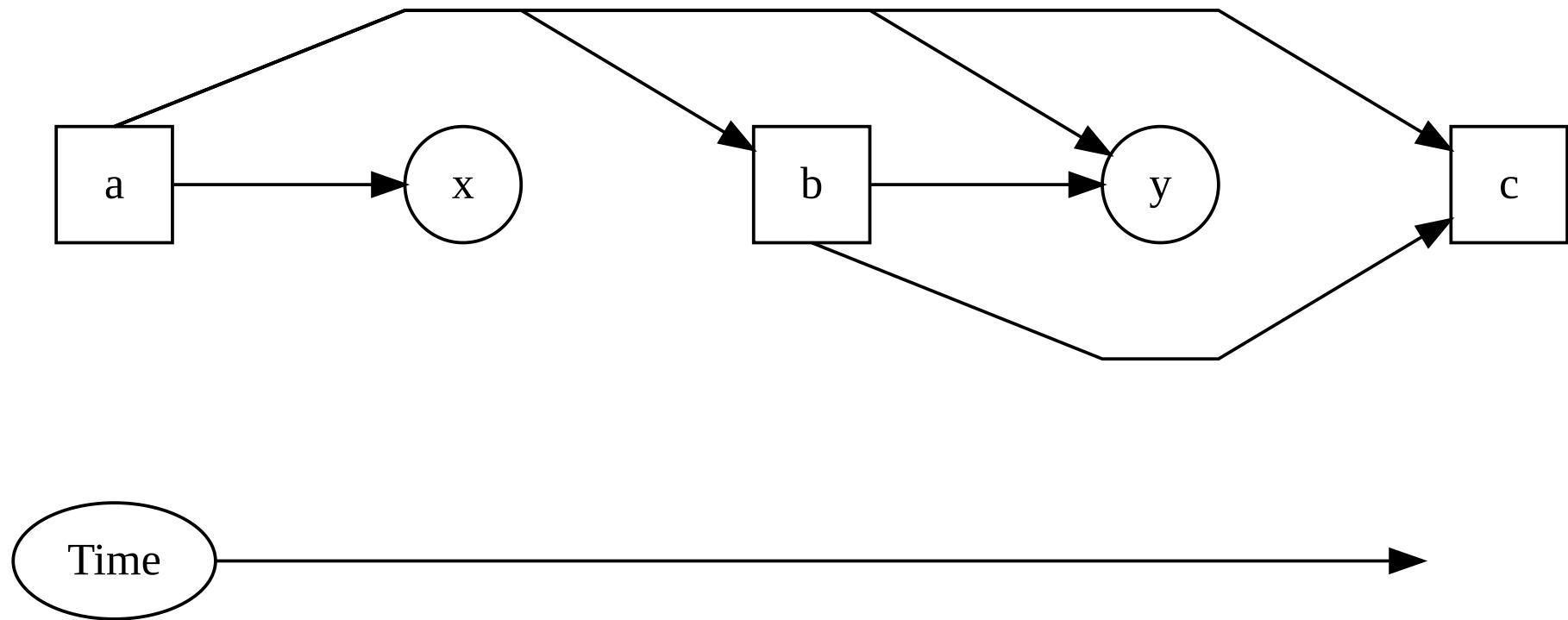
- Premature removal from study
- Event did not occurred yet
- Only right-censoring due to the study design
- All patients already affected at $t = 0$



Concordance Index (CI) I

- Main metric for evaluation of performance
- Accuracy of the *ranking* of predictions
- Calculation: Pairwise comparisons
 - Concordant if higher prediction & lower true survival
 - 1 if concordant, else 0
 - Average of binary results $\rightarrow c \in [0, 1]$

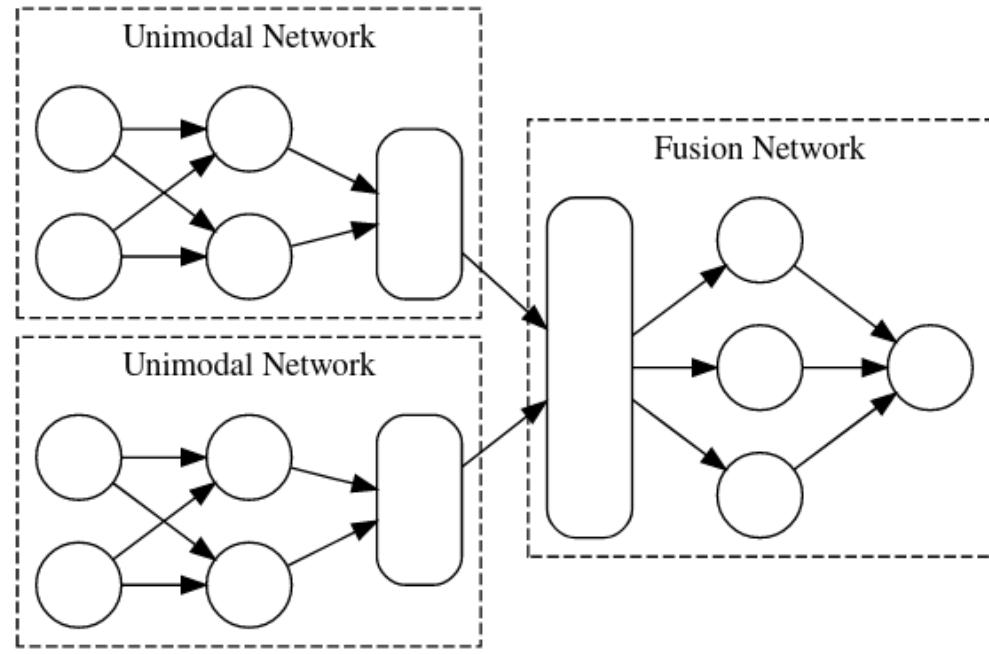
Concordance Index (CI) II



Uncensored samples are compared to all samples with later event occurrence.

Multimodal Networks

- Usage of multiple modalities
 - Different kinds of data
 - Domain-specific
- Multimodal fusion
 - Three general types
 - Varying time-point
 - Learned correlations vary



Intermediate Fusion

Interpretability methods

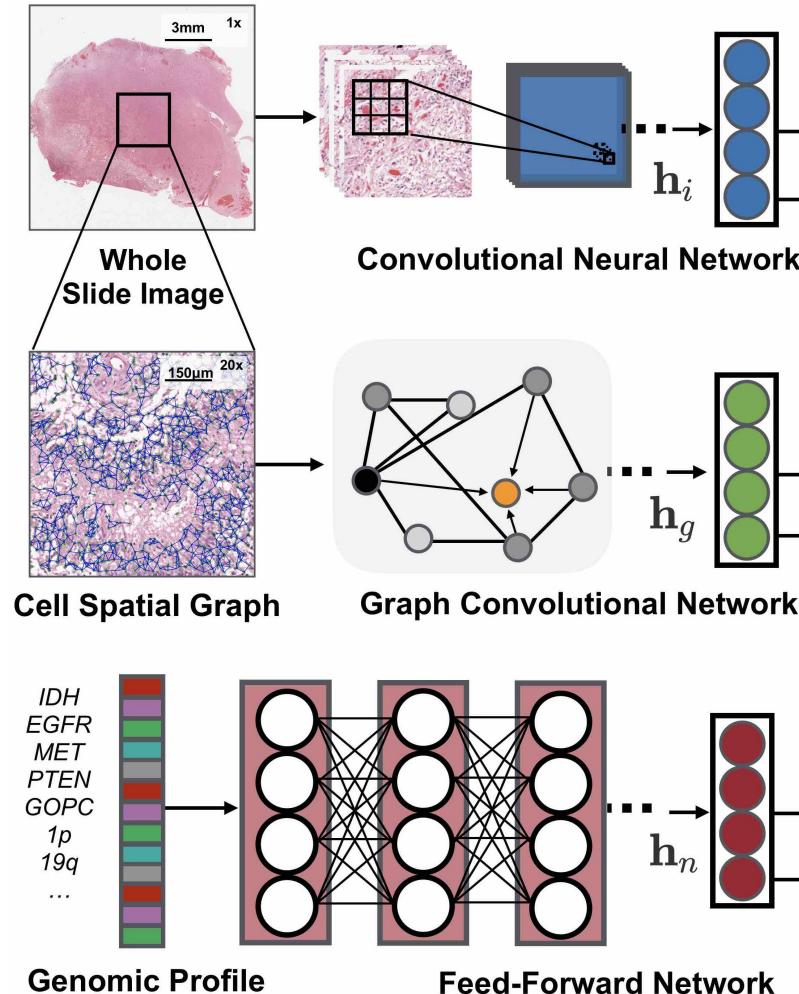
- Score attention of network to each feature
- Usage for biomarker detection
 - Compute attribution of pixels in WSI
 - Tissue regions of interest
- Post-Training procedure → No re-training needed

Related Work

General

- First adaption of neural networks to Cox proportional hazards model (CPH) by Faraggi and Simon ([1995](#))
 - CPH: linear regression model for survival analysis
- DeepSurv: first usage of “modern” deep learning techniques by Katzman et al. ([2018](#))
- Highest reported C-Index for RCC: 0.808 by Ning et al. ([2020](#))
 - Eigengenes (RNA-Seq)
 - Non-overlapping image patches

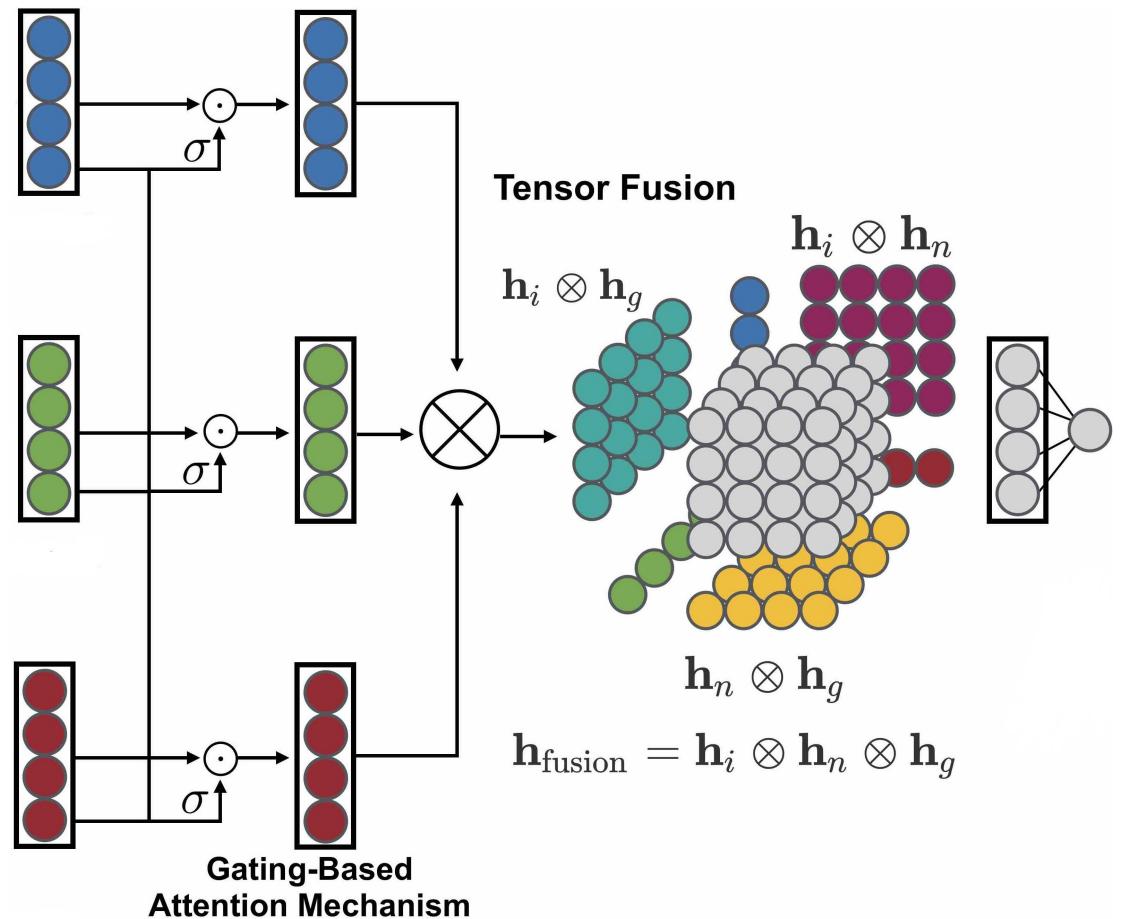
PathomicFusion I



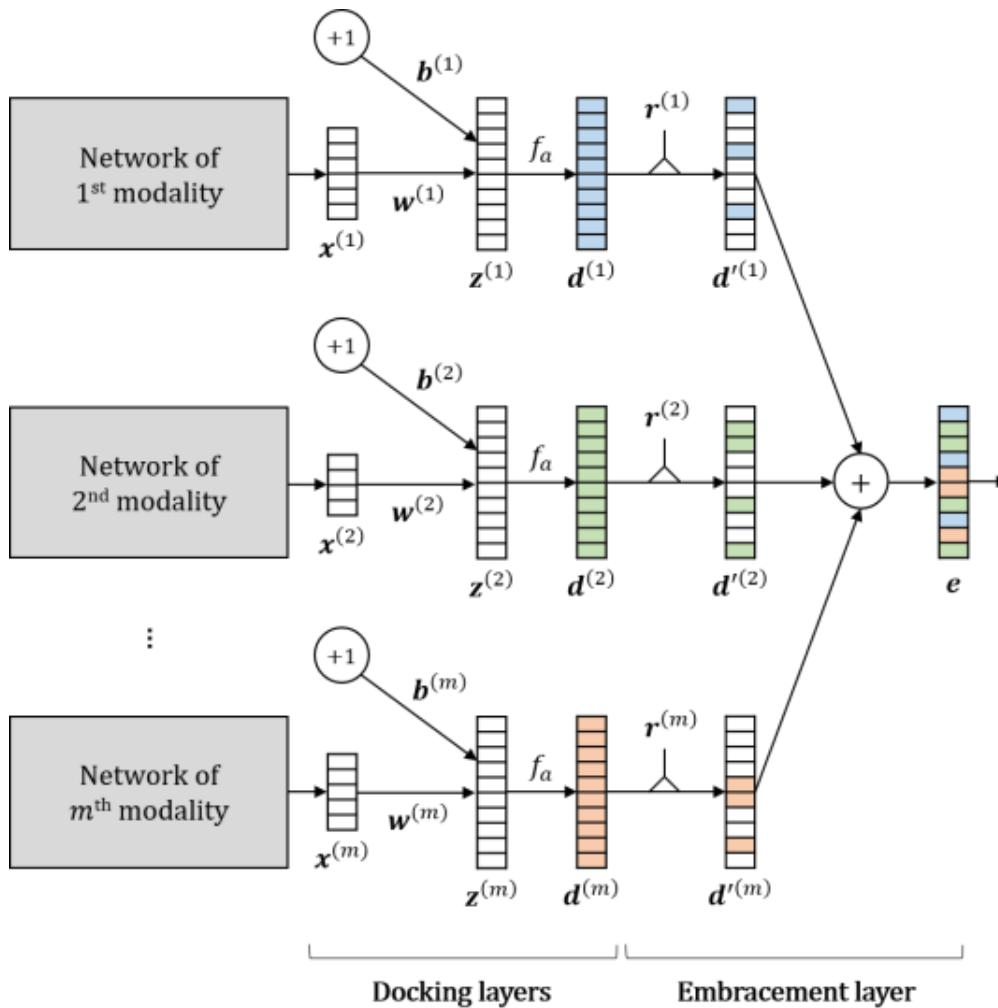
- Whole Slide image patches
- Graph convolutional network
 - Cell detection and localisation
 - K-nearest Neighbor
- Self-Normalizing Networks (SNN)
 - Zero mean and unit variance
 - Lower risk of Overfitting

PathomicFusion II

- Attention Mechanism
 - Controls expressiveness or vector size
 - Weaken noisy features
- Kronecker product
 - Same-time fusion
 - Pair-wise interactions



EmbraceNet



- Docking Layer
 - Independent unimodal networks
 - Create common-size outputs
- Embracement Layer
 - Stochastic sampling
 - Input-sized fusion vector

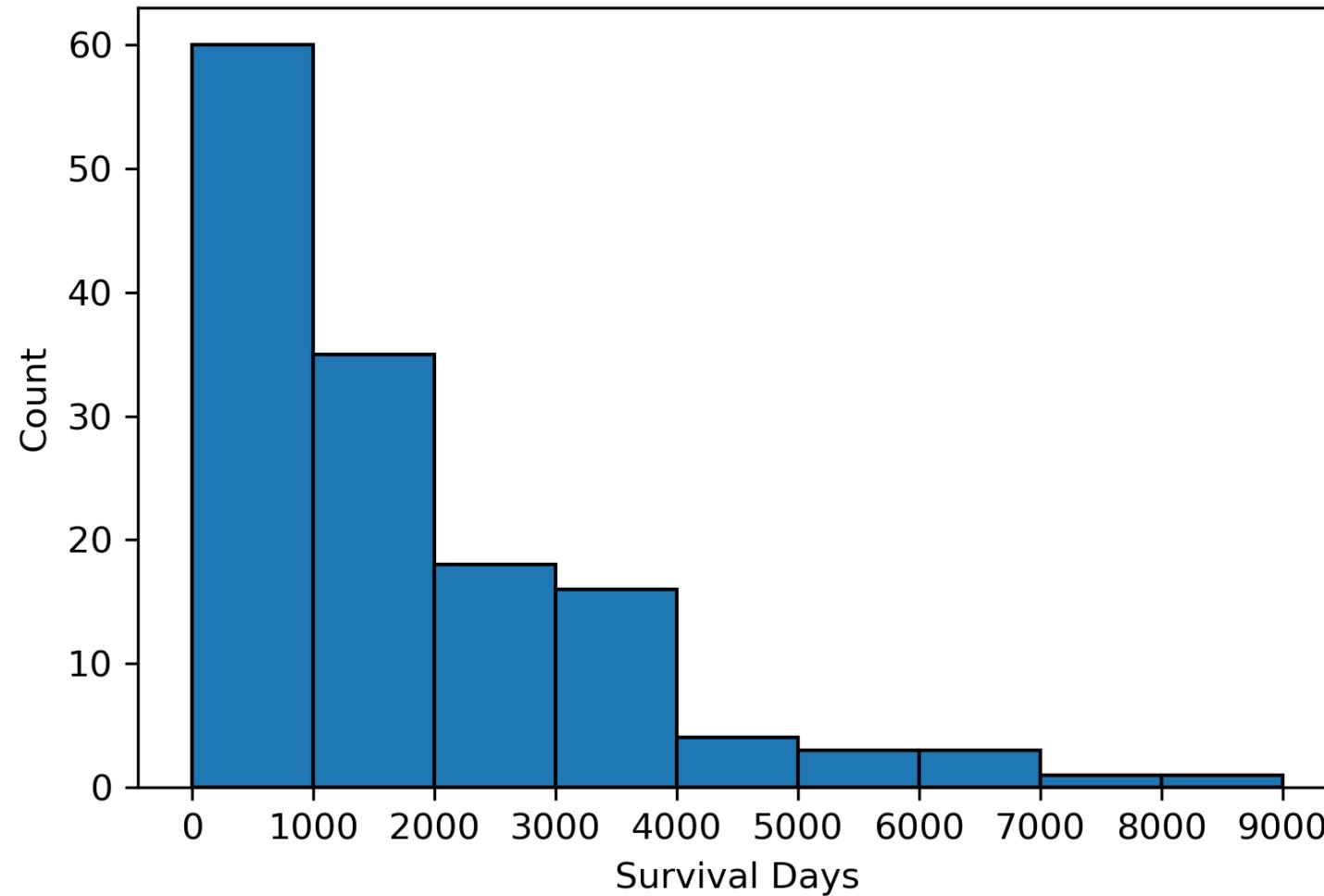
Material & Methods

Censored Samples

- Patients of the Hannover Medical School (MHH)
- From 171 to 141 patients due to filtering
- Filtering
 - Data issues (missing dates, images unavailable)
 - Subtype: clear cell renal cell carcinoma (ccRCC)
- Subcohort of 24 with gene expression data
- Total time span: 1992 and 2022

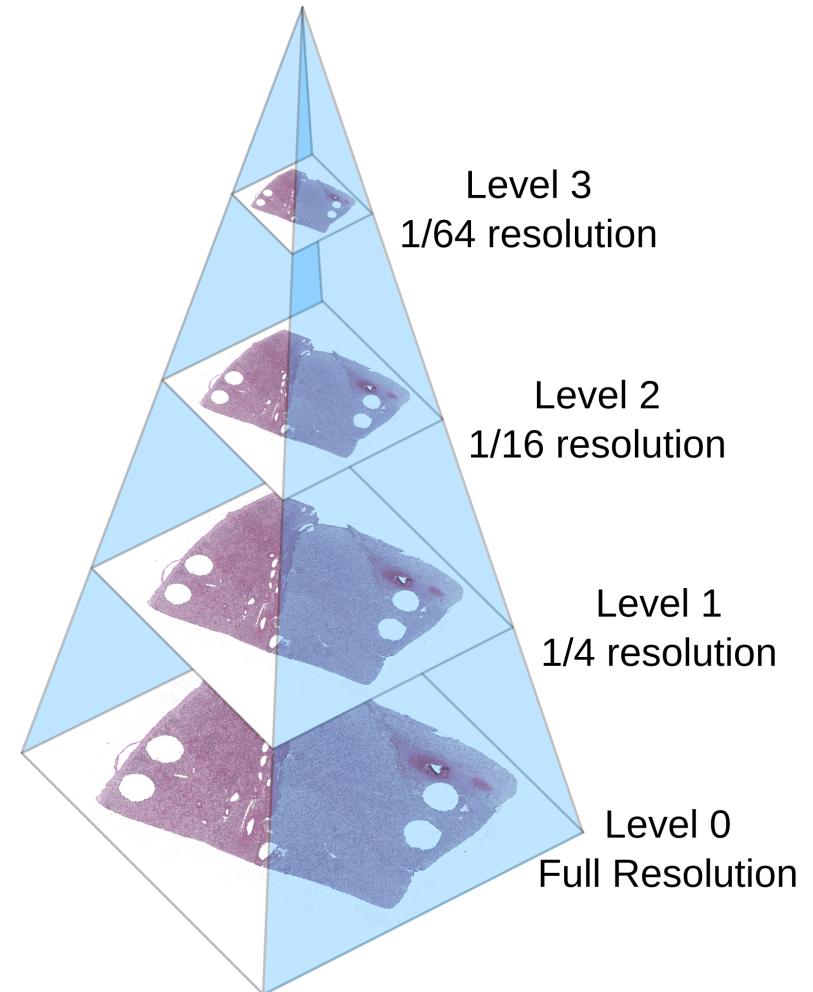
Survival Time

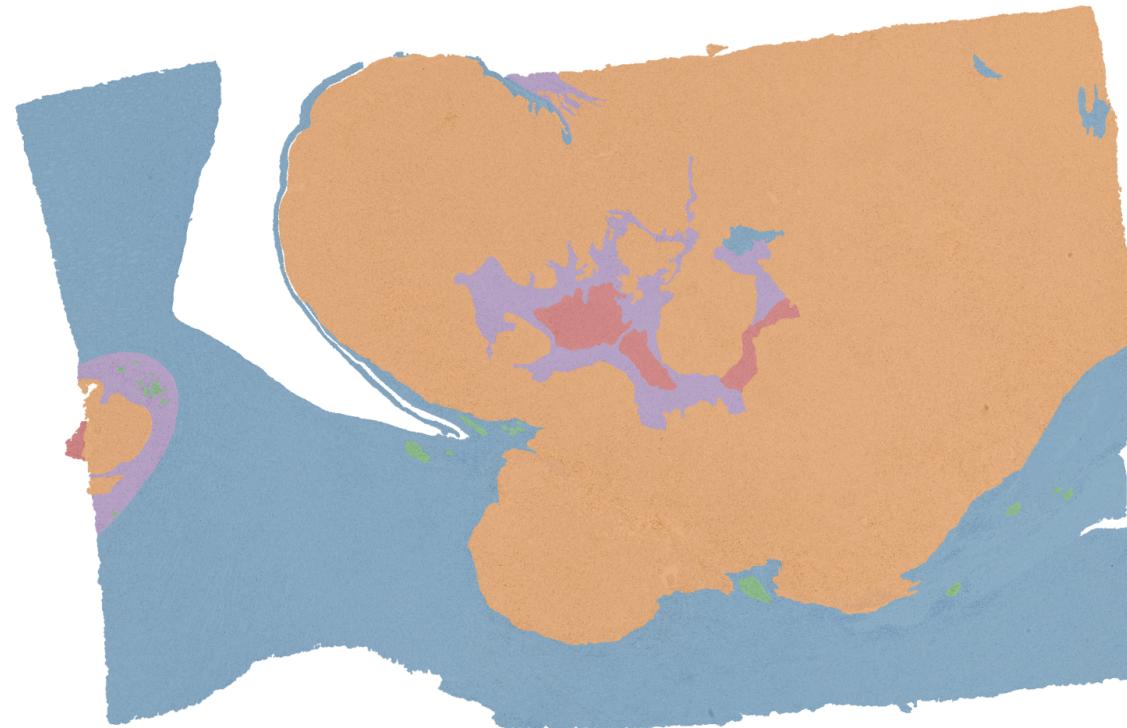
Date of nephrectomy to recorded date of death / last check-up



Whole Slide Images

- Around 10^9 pixels large
- Pre-computed lower resolutions
- Five images per patient
- Differently stained
 - Haematoxylin and Eosin (H&E)
 - Immunohistochemistry:
CD68; CD204; CD8 and CD20;
CD4 and FoxP3



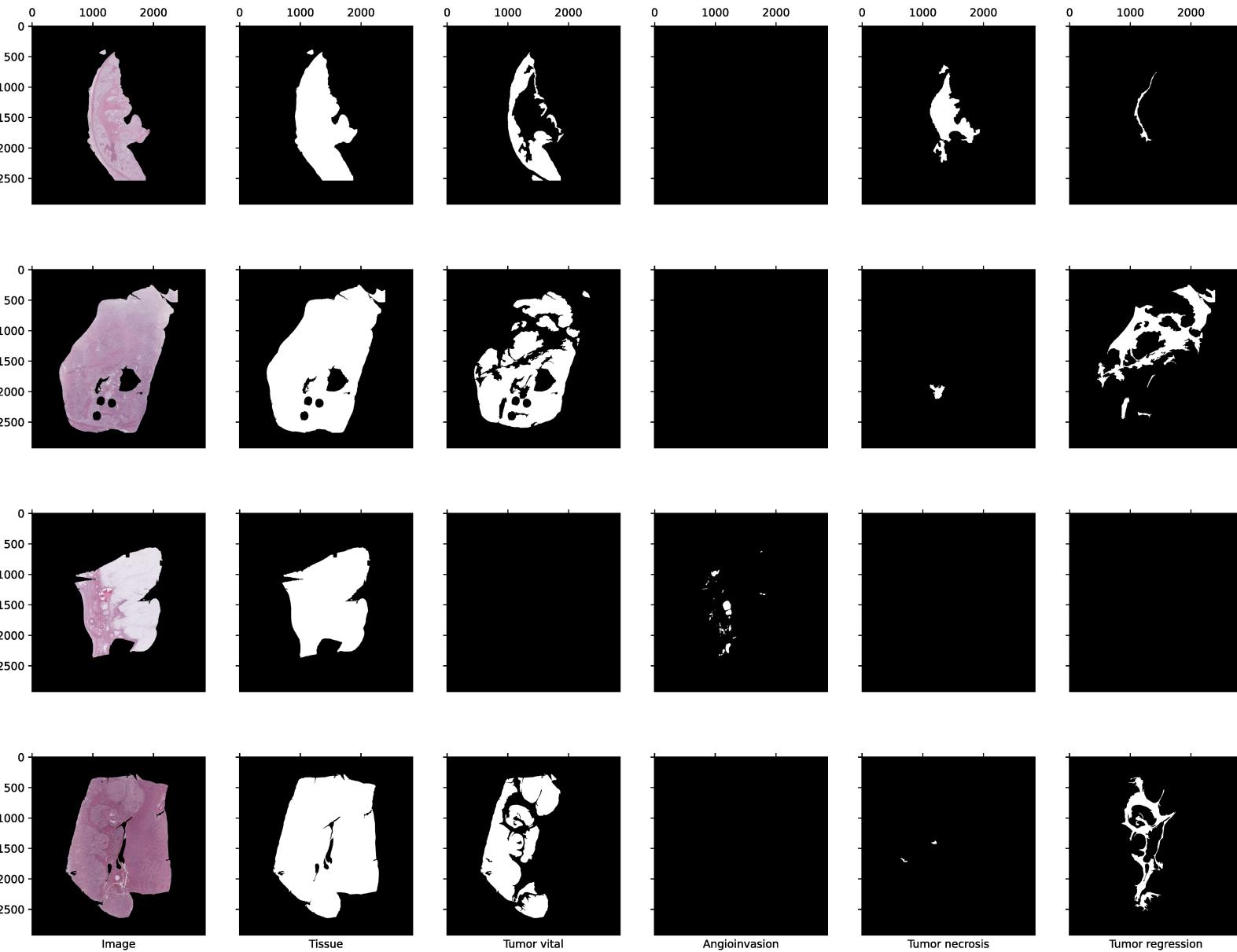


Area varies per image per annotation!

Data Cleanup - Scanning Artifacts

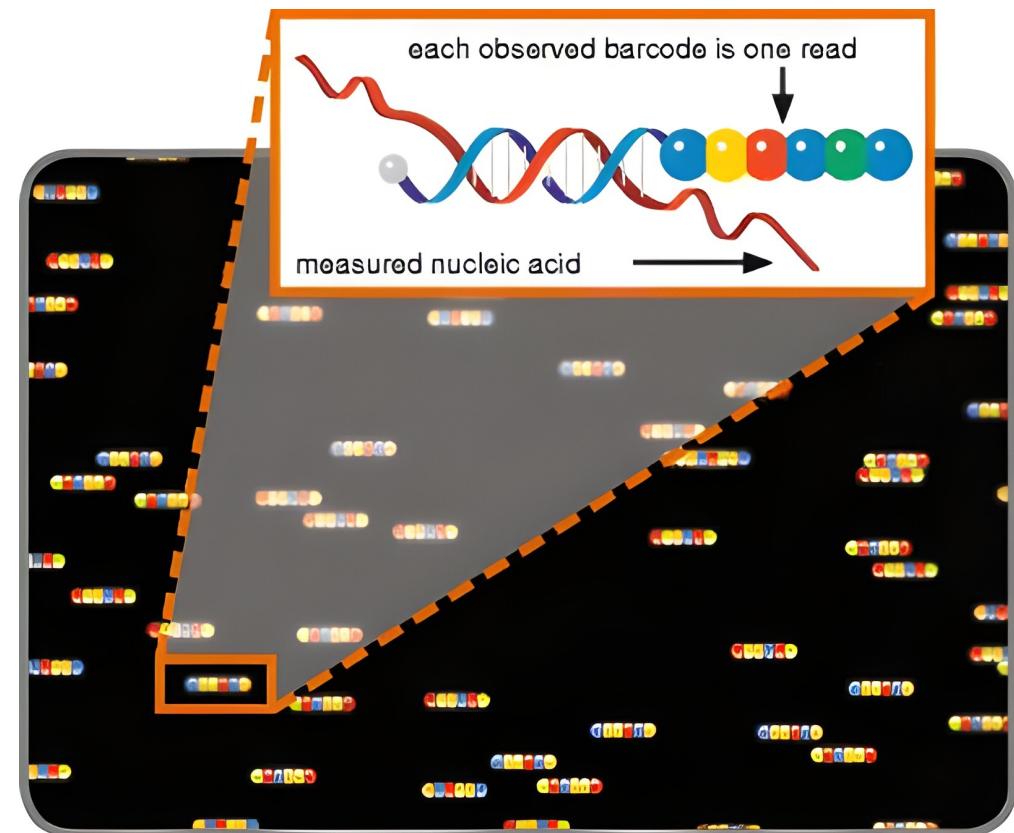
- Slide contaminations
- Scanning process
 - Filters empty tiles
 - Failed in some cases
- Usage of disk space
- Solution: Annotation-based Cropping





Gene Expression Data

- nCounter® PanCancer 10 360™ Panel by Nanostring
- Molecular barcodes
- 750 genes: tumour, microenvironment, and immune response
- 14 out of the 24 patients were censored



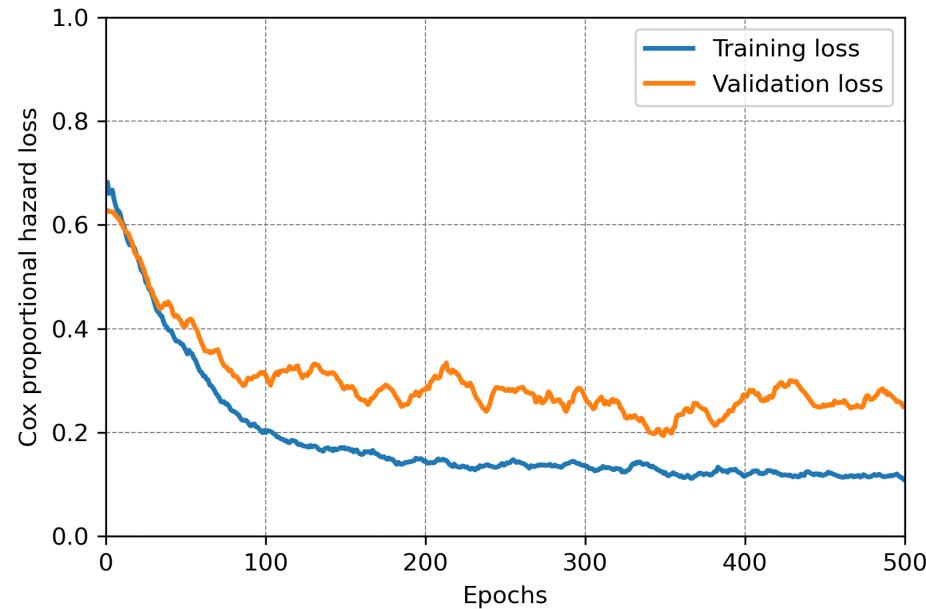
Results

Network Training

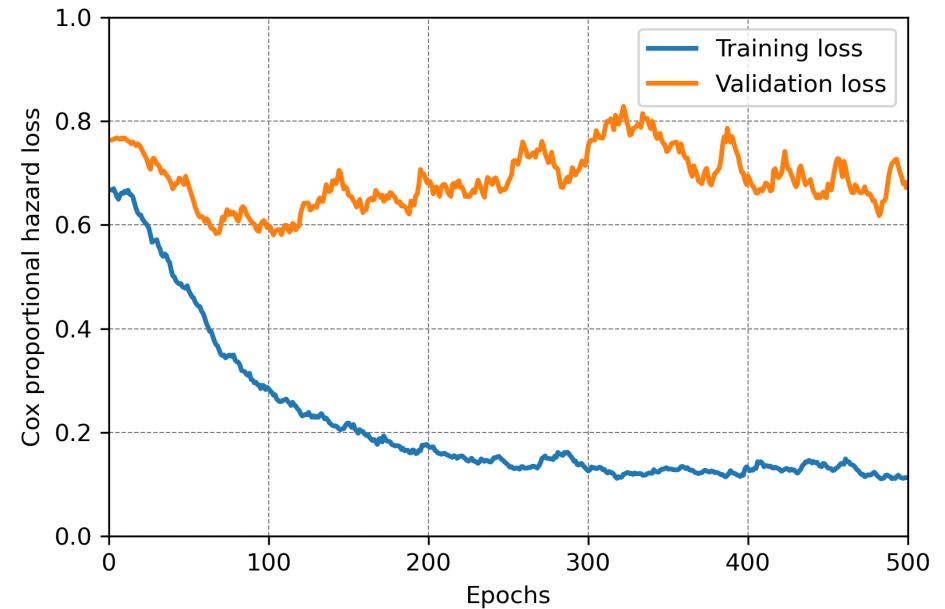
WSI - Network architecture

- Input - images at downsampled resolution
 - Memory limitations
- Output - Hazard / Input into CI
- Convolutional neural network
 - Modified ResNet architecture
 - Convolutional layer to reduce input dimensions
 - Fully connected head replaced with custom layers

WSI - Model Capability



Full Dataset



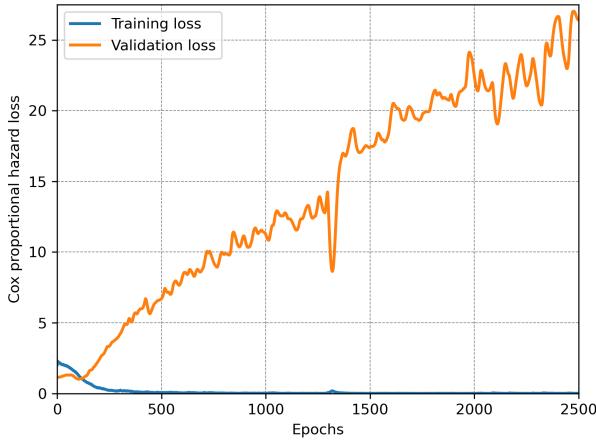
Subset of 5%

C-Index = 0.89

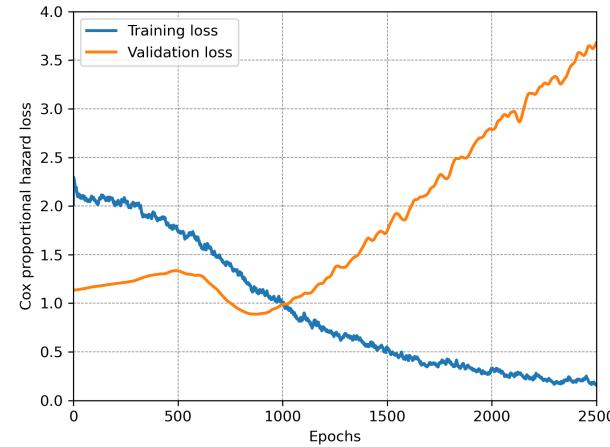
Genes - Self Normalizing networks

- Type of fully connected network
 - Scaled exponential linear unit (SeLU)
 - Alpha dropout layer
- Outputs → zero mean & unit variance
- Counteract vanishing/exploding gradient problem
- Applied to our data
 - Strong overfitting after certain amount of epochs
 - Multiples parameters explored - Unsuccessfull

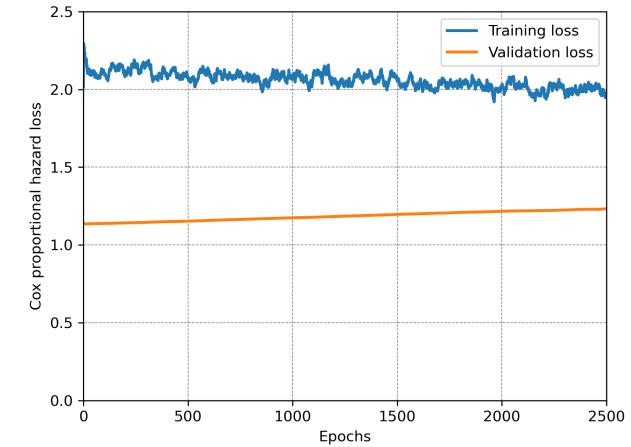
Genes - Learning rates



$1e - 4$



$1e - 5$



$1e - 6$

Lowering or raising learning rate worsens loss considerably

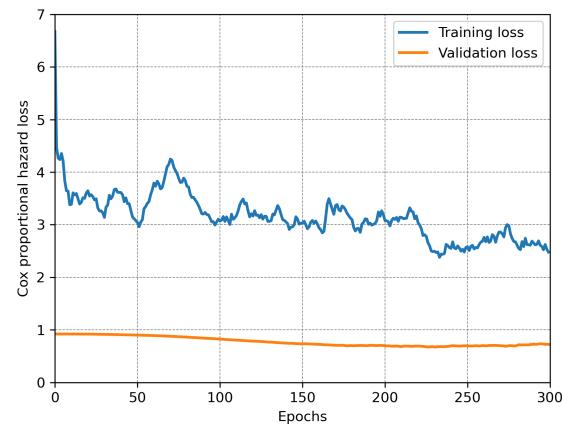
Network Fusion

Fusion experiments - 2 Issues

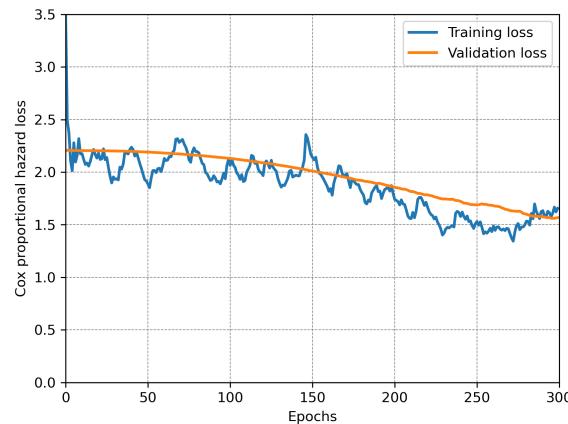
- Different Learning rates
 - Sub-networks with individual learning rates
 - Solution: Variable learning rates for sub-networks and $1e-3$ post fusion
- Number of epochs
 - Aim for 500 (same as WSI network)
 - Only 300 due to thesis deadline

Fusion experiments - Approaches

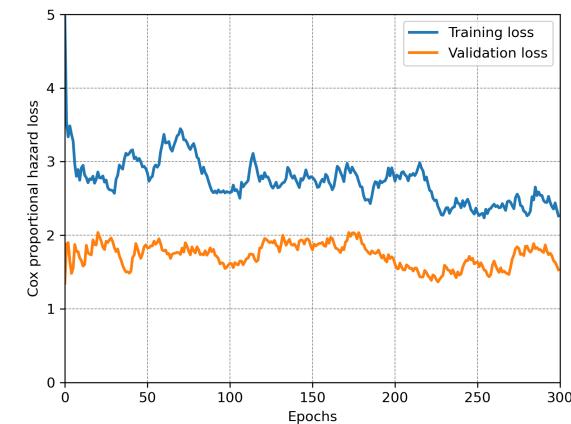
- Comparison of fusion methods
 - Element-wise summation and maximum
 - Vector concatenation
 - Probabilistic selection from EmbraceNet
 - Attention-based summation
 - Pathomic Fusion - Attention-Gating & Kronecker Product
- Only Difference: fusion method / fusion layers



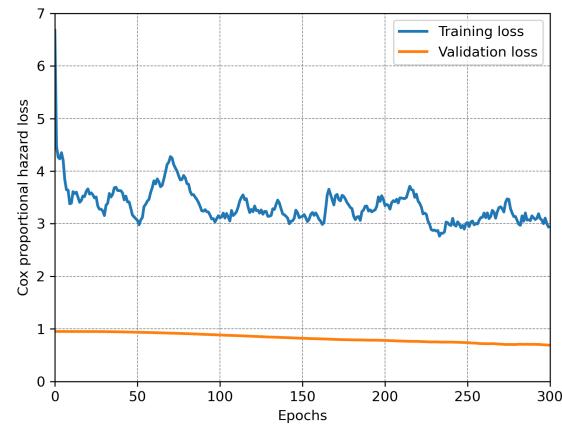
Elementwise summation



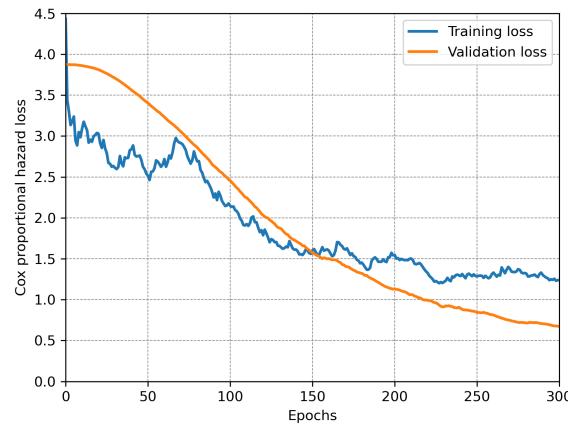
Vector Concatenation



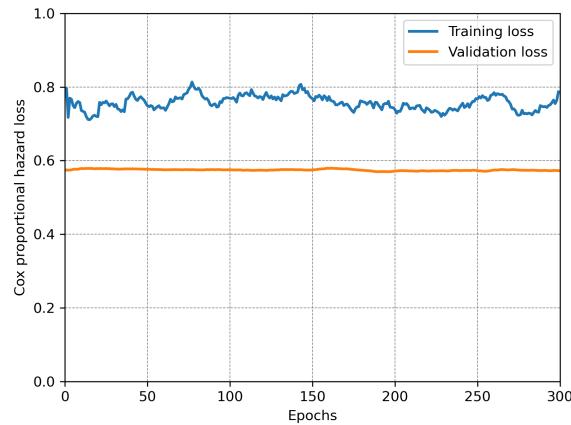
Probability-based Sampling



Elementwise maximum



Attention Mechanism



Pathomic Fusion

Fusion experiments - C-Indices

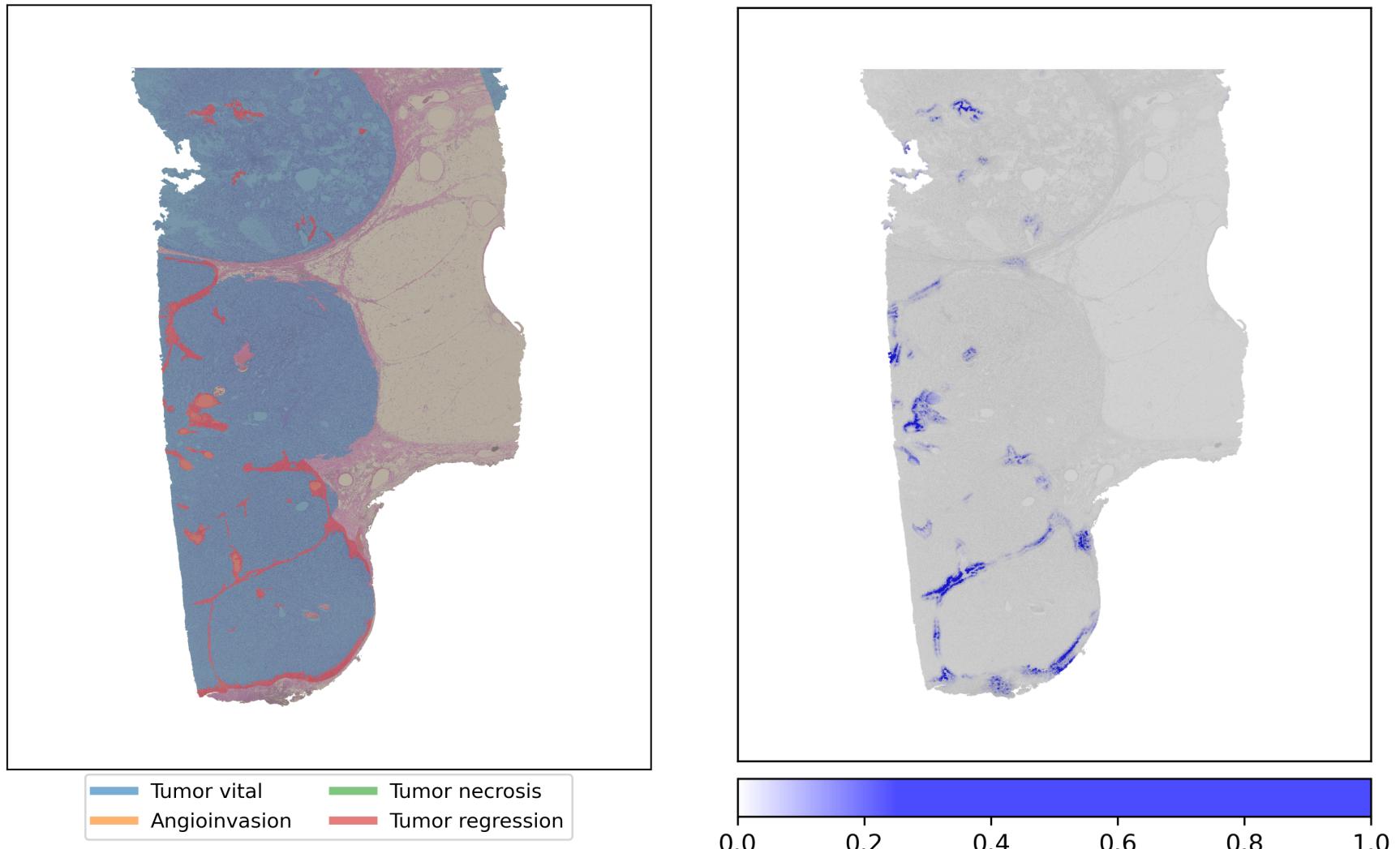
Experiment	Training	Validation
Maximum	0.594	0.590
Summation	0.637	0.651
Concatenated	0.709	0.522
EmbraceNet	0.596	0.644
Attention	0.492	0.500
Kronecker	0.580	0.594

Interpretability

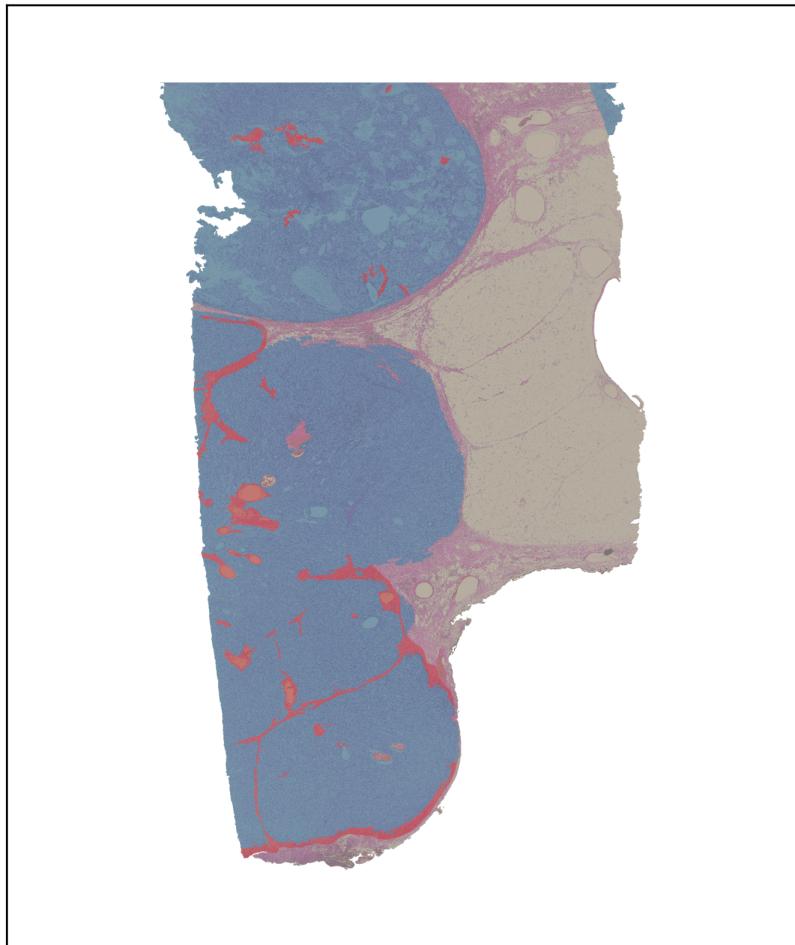
⚠ Disclaimer

The unimodal whole slide image network was used for the following!

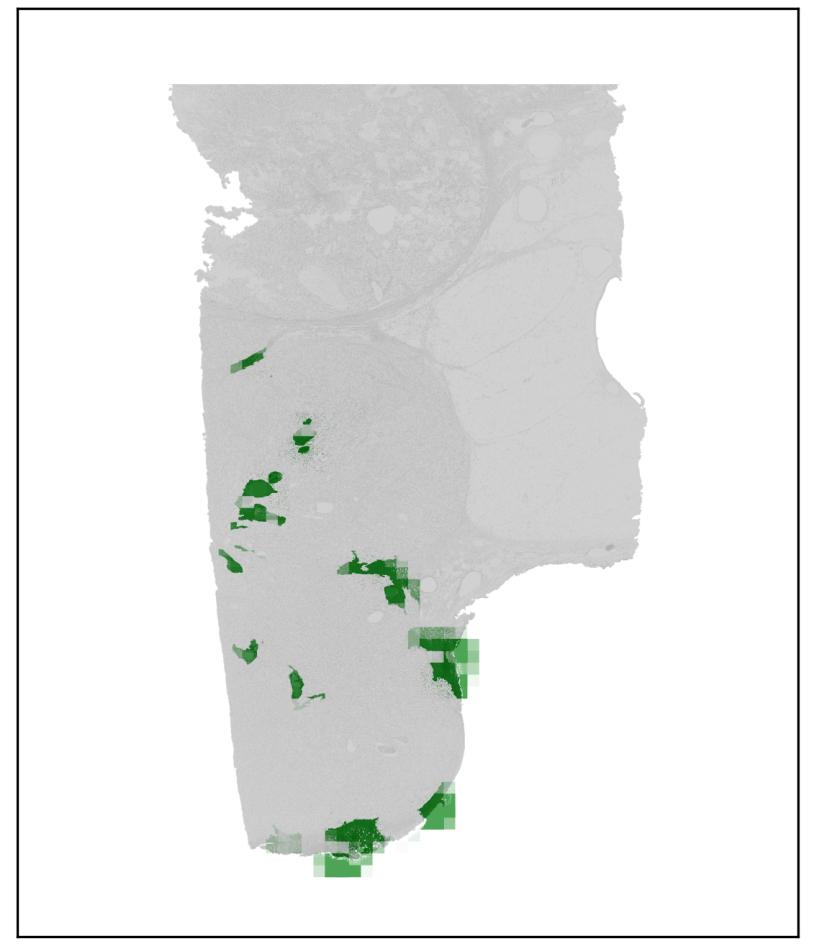
Integrated Gradients



Grad-CAM

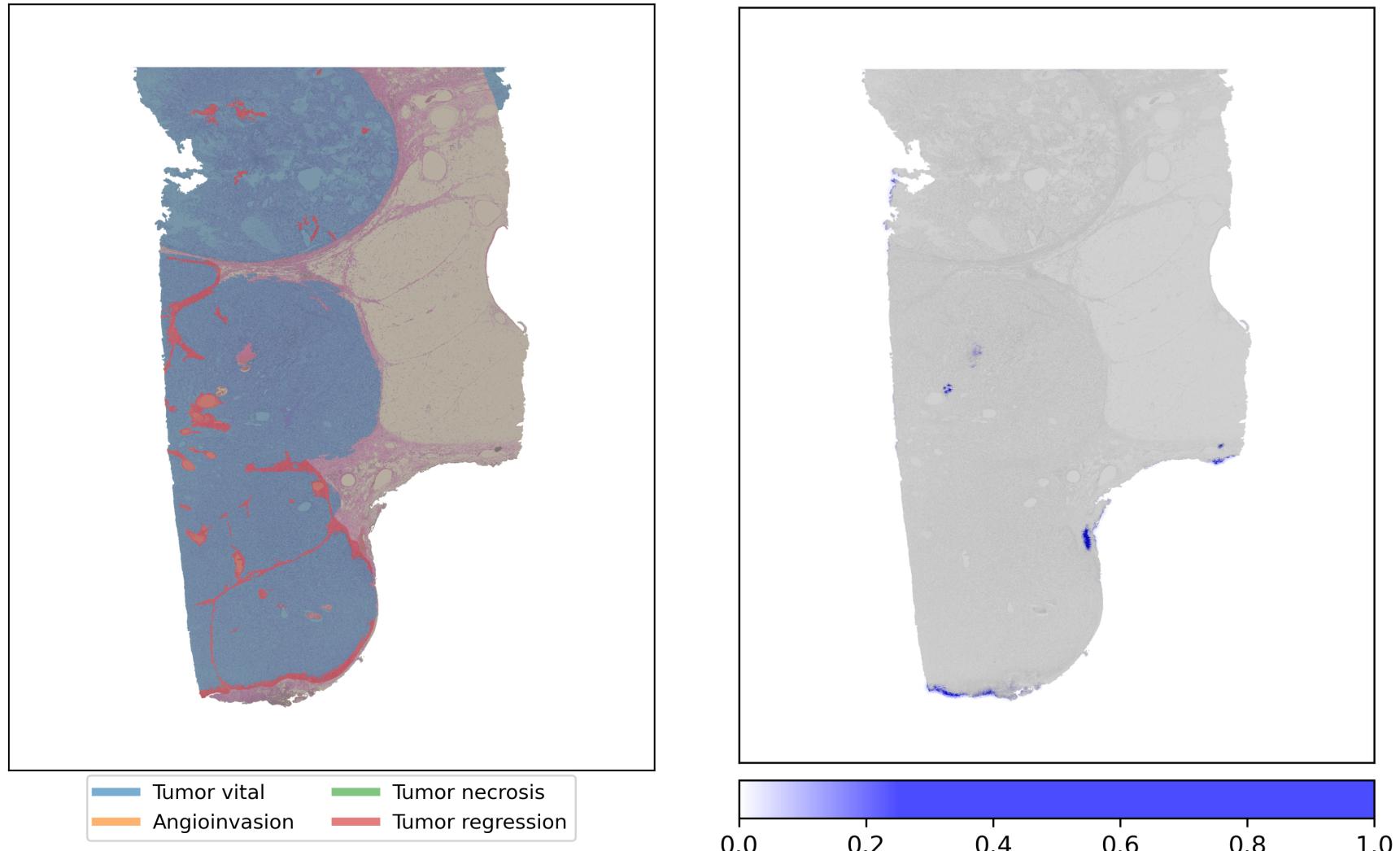


Tumor vital
Angioinvasion
Tumor necrosis
Tumor regression

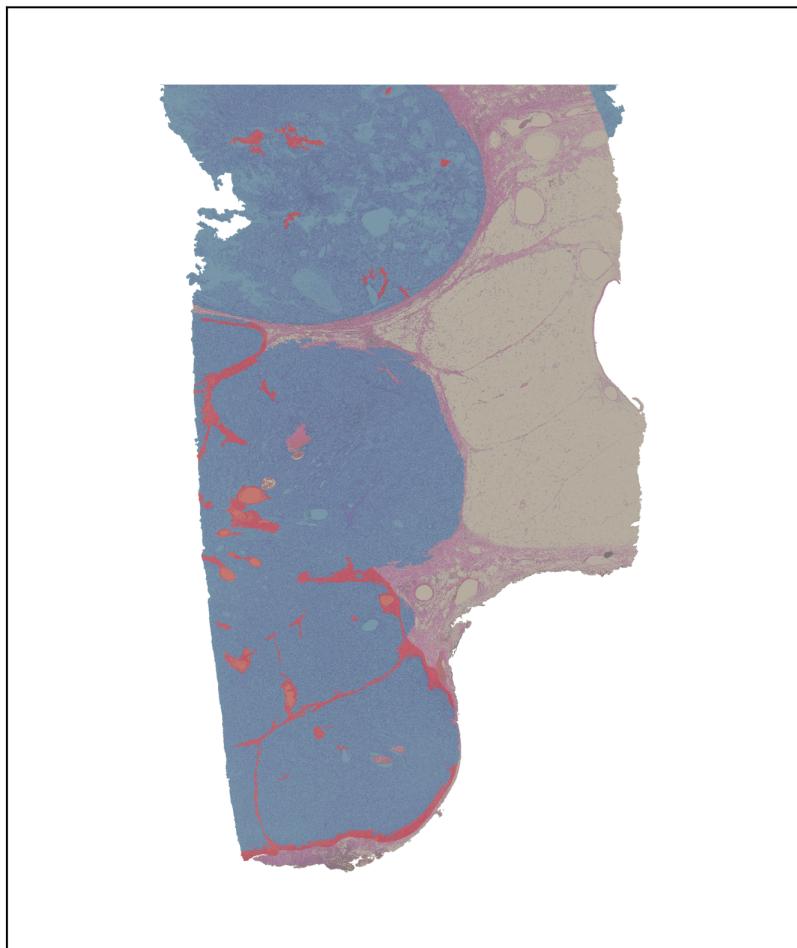


0.0 0.2 0.4 0.6 0.8 1.0

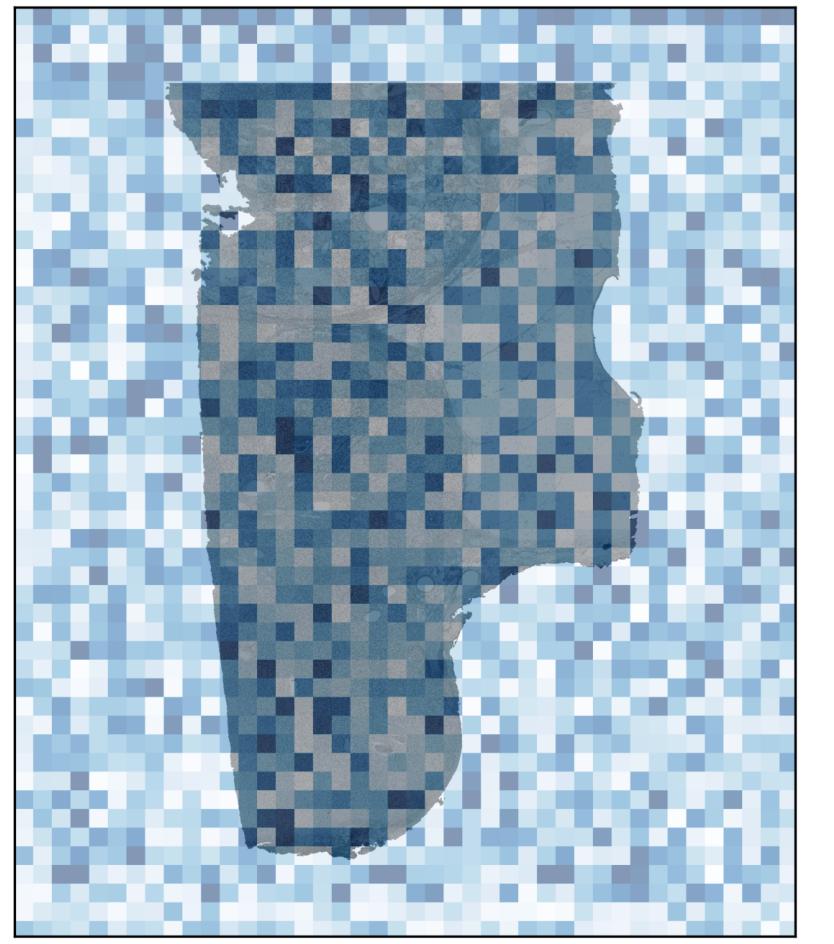
Saliency



Occlusion

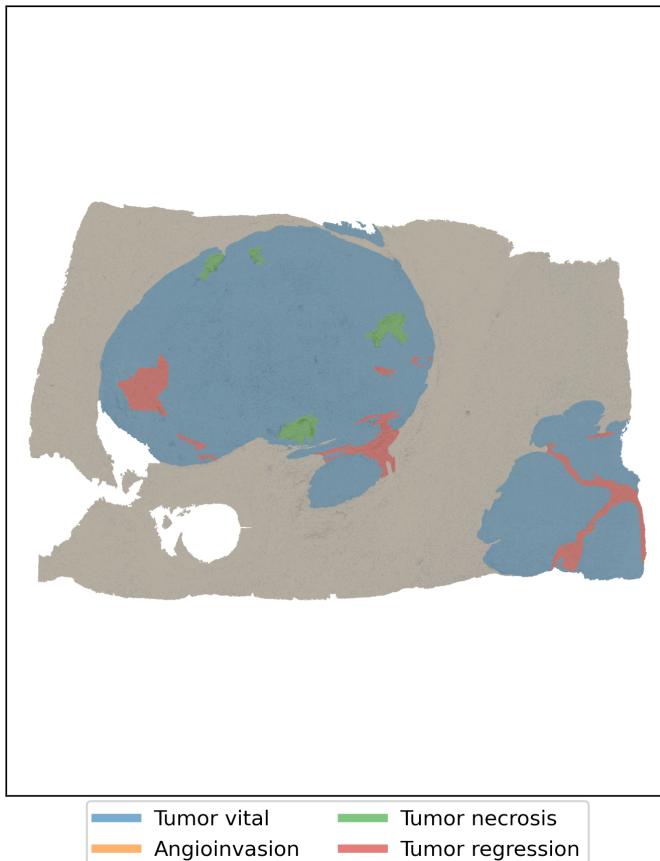


Tumor vital
Angioinvasion
Tumor necrosis
Tumor regression

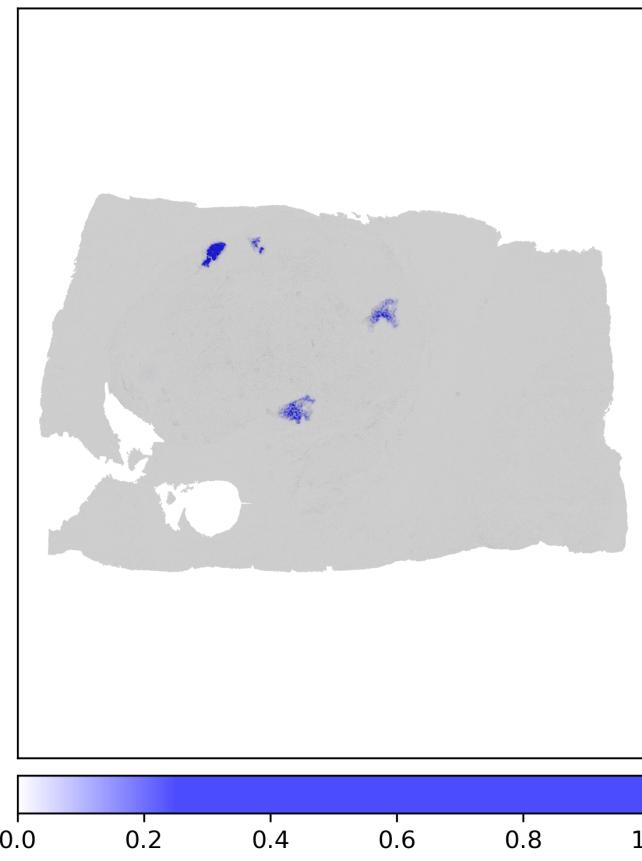


0.0 0.2 0.4 0.6 0.8 1.0

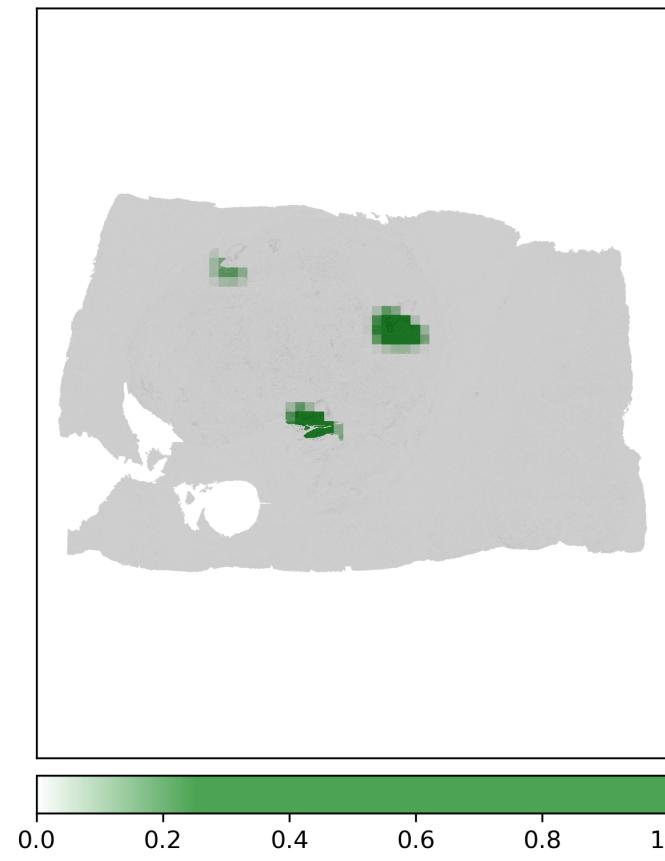
Importance of Annotations



Reference



Integrated Gradients



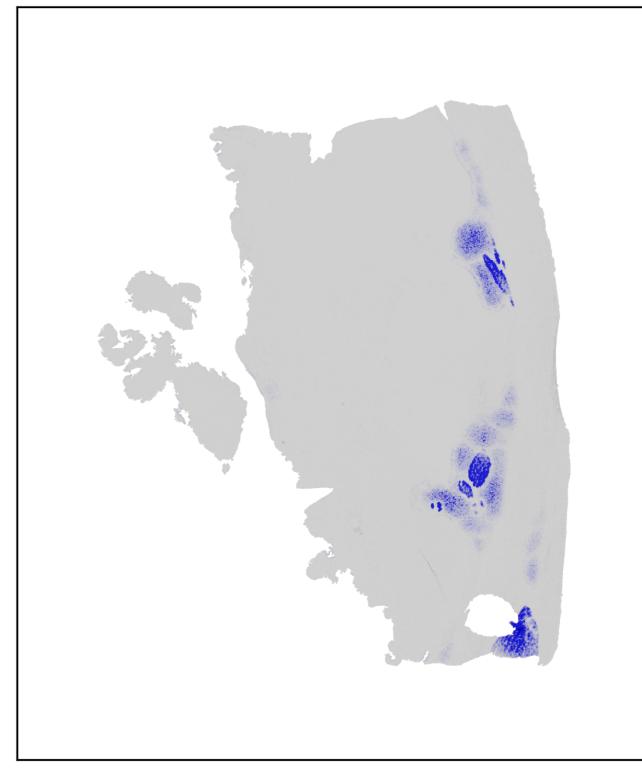
GradCAM

Other Regions

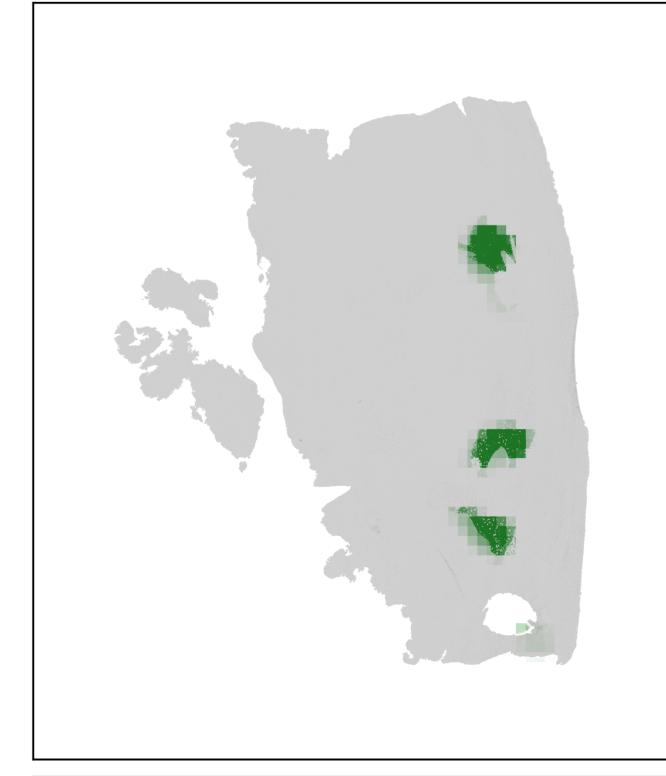


Tumor vital Tumor necrosis
Angioinvasion Tumor regression

Reference

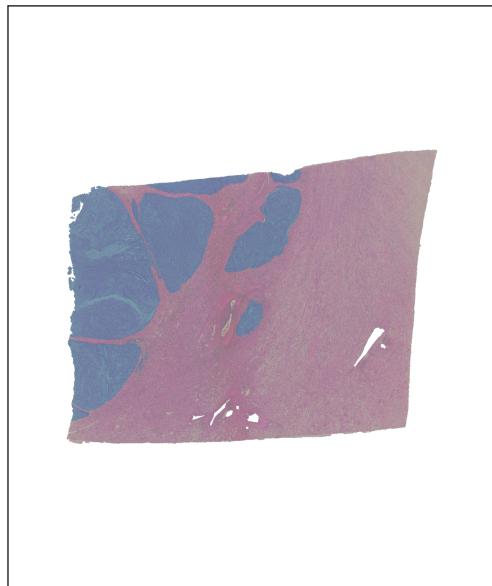


Integrated Gradients

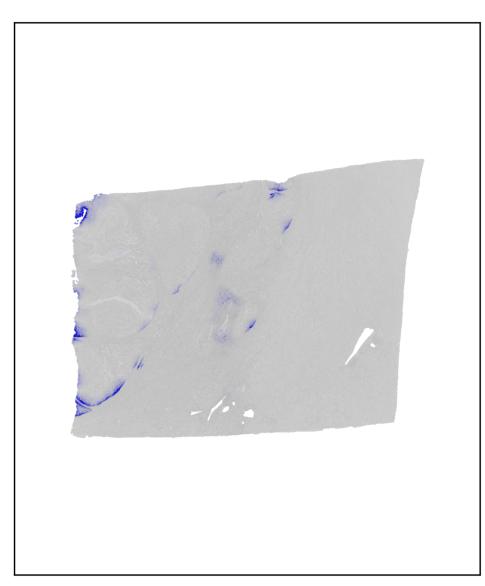


GradCAM

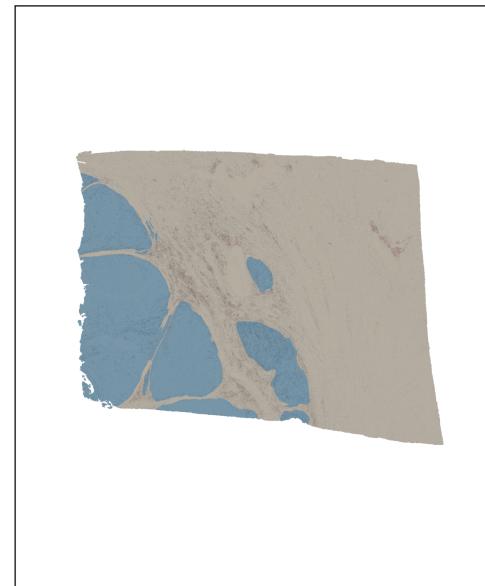
Same Patterns without Annotation



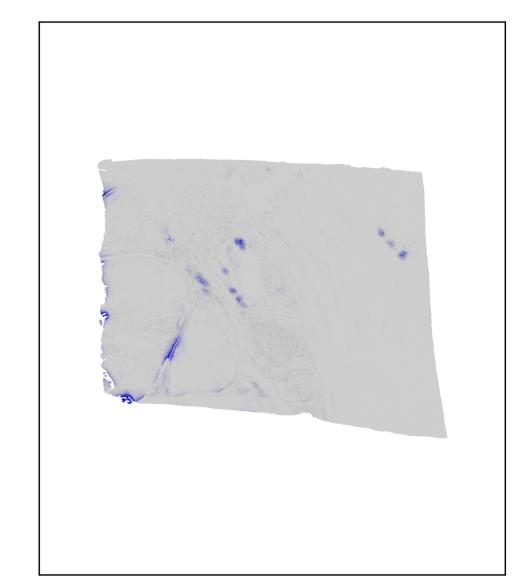
H&E Stain



Integrated Gradients



IHC - CD4 & FoxP3



Integrated Gradients

Discussion

Technical limitations

- Low batch size
 - Ideal: Process full data set at once
 - Else: Approximation of true survival function
- Training time
 - 10+ hours for WSI networks
 - Slow iteration

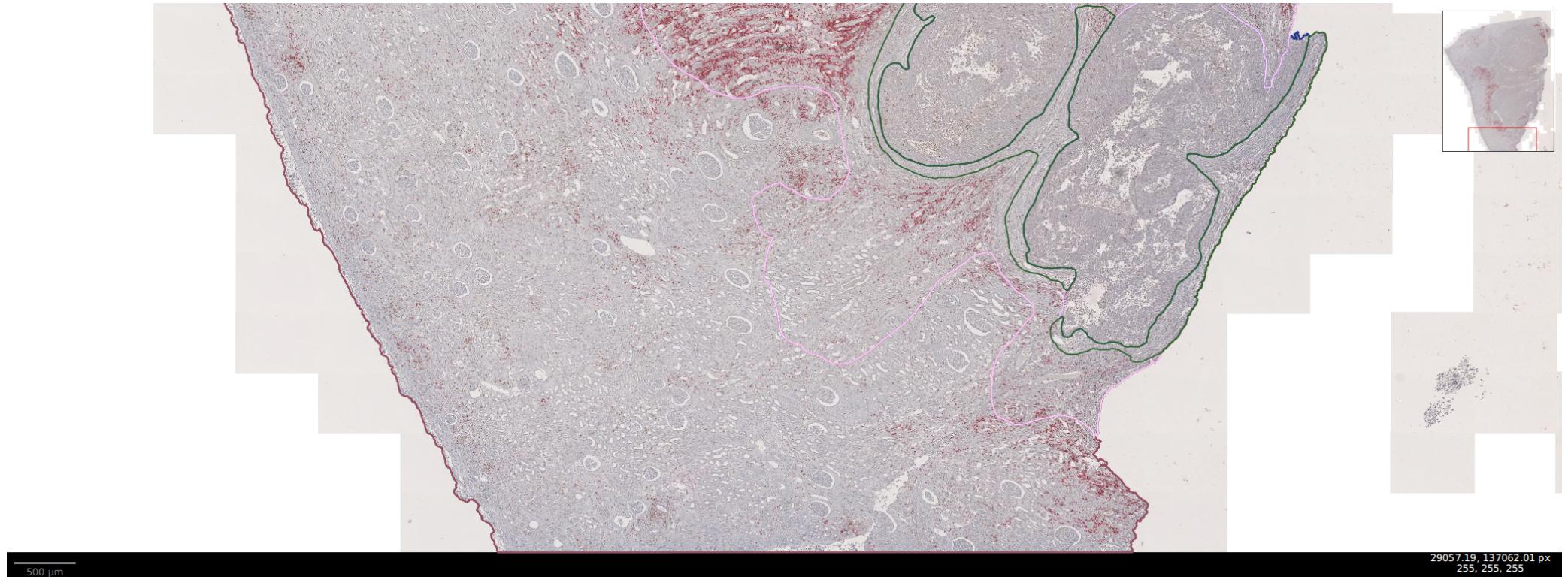
Concerns

- Data separation by image rather than patient
 - Training set and test set contain similar images
 - May lead to memorization of outer appearance
- Loss of Information due to downsampling
- Strong influence introduced by annotations

Data - Quality and Amount

- Too few gene expression samples
 - Late fusion might be more suitable
- Discarded: images with unannotated / mislabelled regions
- Irregularities impact results
 - Microarray sampling
 - Slide preparation

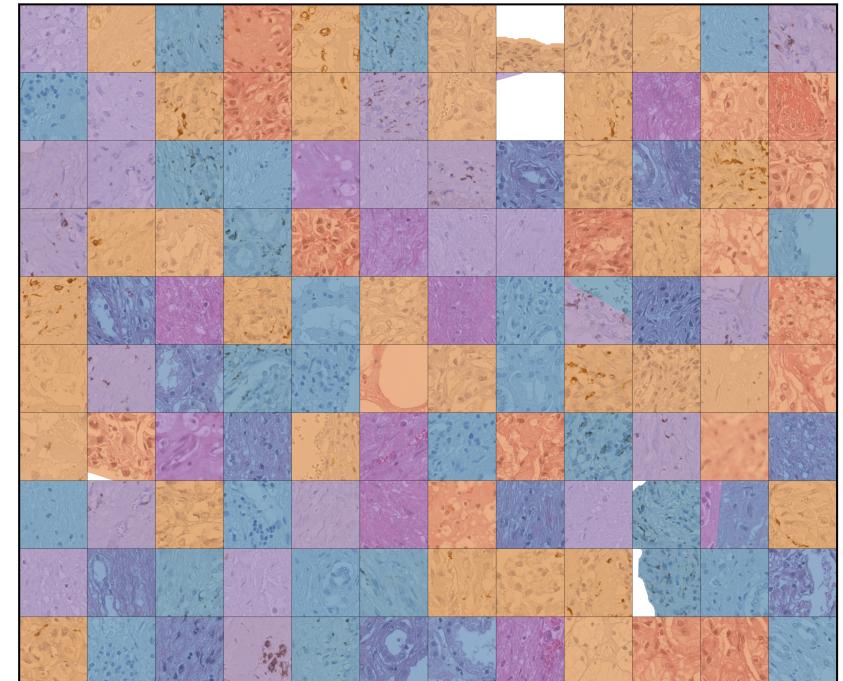
Cut off tissue



- Parts of tissue missing from scan
- Potentially valuable information

Original contribution

- Stacking annotations as pixel-masks on top of images
- Comparison of different fusion approaches
- Annotation-based patch extraction (not used)



Extracted Patches

Future Work

- Investigate interesting regions
- Acquire more gene expression data
- Aim to increase batch size
- Investigate relevance of annotations
- Perform proper hyperparameter search
- Investigate differences of interpretability methods

Thank you!

References

- Capitanio, Umberto, and Francesco Montorsi. 2016. “Renal cancer.” *Lancet* 387 (10021): 894–906. [https://doi.org/10.1016/S0140-6736\(15\)00046-X](https://doi.org/10.1016/S0140-6736(15)00046-X).
- Cesano, Alessandra. 2015. “nCounter PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA).” *Journal for ImmunoTherapy of Cancer* 3 (1). <https://doi.org/10.1186/s40425-015-0088-7>.
- Chen, Richard J., Ming Y. Lu, Jingwen Wang, Drew F. K. Williamson, Scott J. Rodig, Neal I. Lindeman, and Faisal Mahmood. 2022. “Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis.” *IEEE Transactions on Medical Imaging* 41 (4): 757–70. <https://doi.org/10.1109/tmi.2020.3021387>.
- Choi, Jun-Ho, and Jong-Seok Lee. 2019. “EmbraceNet: A Robust Deep Learning Architecture for Multimodal Classification.” *Information Fusion* 51 (2019) 259-270 51 (April): 259–70. <https://doi.org/10.1016/j.inffus.2019.02.010>.
- Cmglee. 2015. “Illustration of an Image Pyramid with 5 Levels.” August.
- Faraggi, David, and Richard Simon. 1995. “A Neural Network Model for Survival Data.” *Statistics in Medicine* 14 (1): 73–82. <https://doi.org/10.1002/sim.4780140108>.

- Hancock, S. Brandon, and Christos S. Georgiades. 2023. “Kidney Cancer.” *The Cancer Journal* 22 (6): 387–92. <https://doi.org/10.1097/ppo.0000000000000225>.
- Katzman, Jared L., Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. 2018. “DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network.” *BMC Med Res Methodol* 18 (1): 24. <https://doi.org/10.1186/s12874-018-0482-1>.
- Ning, Zhenyuan, Weihao Pan, Yuting Chen, Qing Xiao, Xinsen Zhang, Jiaxiu Luo, Jian Wang, and Yu Zhang. 2020. “Integrative analysis of cross-modal features for the prognosis prediction of clear cell renal cell carcinoma.” *Bioinformatics* 36 (9): 2888–95. <https://doi.org/10.1093/bioinformatics/btaa056>.

