TISSUE REGION SEGMENTATION IN H&E-STAINED AND IHC-STAINED PATHOLOGY SLIDES OF SPECIMENS FROM DIFFERENT ORIGINS

A PREPRINT

Sepideh Naghshineh Kani
Department of Computer Science
Bahcesehir University
Besiktas, Istanbul 34353, Turkey
sepidehnaghshinehk@gmail.com

Burak Can Soyak
Department of Computer Science
Bahcesehir University
Besiktas, Istanbul 34353, Turkey

Melih Gokce
Faculty of Medicine
Bahcesehir University
Kadikoy, Istanbul 34734, Turkey

Zeynep Duyar
Faculty of Medicine
Bahcesehir University
Kadikoy, Istanbul 34734, Turkey

Hasan Alicikus
Faculty of Medicine
Bahcesehir University
Kadikoy, Istanbul 34734, Turkey

Özlem Yapıcıer
Faculty of Medicine
Bahcesehir University
Kadikoy, Istanbul 34734, Turkey
ozlem.yapicier@med.bau.edu.tr

Mustafa Umit Oner *
Department of Computer Science
Bahcesehir University
Besiktas, Istanbul 34353, Turkey
onermustafaumit@gmail.com

January 16, 2025

ABSTRACT

In the era of advancing digital pathology, integrating digital slides and deep learning-based decision support systems has become increasingly prevalent in clinical practice. The effectiveness of these systems heavily relies on tissue region segmentation, a process essential for slide scanning and success of deep learning (DL) models. While thresholding-based methods are fast, they usually do not accurately detect tissue regions, especially in diverse staining scenarios or debris-laden images. Hence, this study develops a fast, lightweight, accurate, and robust deep-learning framework for segmenting tissue regions in digital pathology slides of various tissue types with different staining. The model is trained and tested on datasets of seven tissue types collected from seven different institutes. Internal and external validation studies have shown promising results, with area under receiver operating characteristics curve values greater than 0.98. The model has a memory footprint of 19.8 MBs, and it takes approximately 22 seconds to segment out a digital pathology slide using the CPU. The model could be used in digital slide scanners to improve the scanning process and in the pre-processing stages of DL pipelines to prepare high-quality datasets.

Keywords Digital Pathology · Tissue Segmentation · Deep learning · Convolutional Neural Networks

1 Introduction

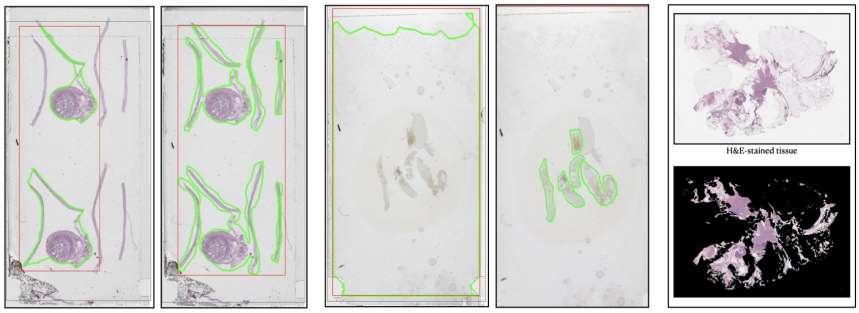
Cancer is a leading cause of death worldwide, with one in five people developing it(Bray et al. [2024]). Histopathology is an essential field in cancer studies, analyzing tissues at cell level. The samples obtained from patients, after undergoing

*https://onermustafaumit.github.io

several steps, are stained using various dyes and techniques to highlight specific structures or cell components within the tissue. Hematoxylin and Eosin (H&E) staining, for example, is commonly used to visualize cell nuclei (with Hematoxylin) and cytoplasm (with Eosin).

Traditional pathology in the clinic involves manual examination under a microscope, while digital pathology uses digital slide scanners to convert glass slides into high-resolution digital images called whole slide images (WSIs). These images are stored in a pyramid structure where the base level contains the highest resolution image. Subsequent levels in the pyramid are down-sampled versions of the base image, representing the slide at progressively lower resolutions. WSIs enable researchers to view and analyze the sample on a computer screen rather than using a traditional microscope. Additionally, WSIs can be processed by powerful deep learning models, which can support human pathologist and decrease their workload while providing more objective results.

Tissue segmentation, an essential step in processing WSIs, divides the images at the pixel level to detect and label specific tissue areas, a critical task in digital pathology. Traditional methods like thresholding (Otsu [1979]), region-growing (Nock and Nielsen [2004]), watershed segmentation (Vincent and Soille [1991]), and edge-based methods (Canny [1983]) have been widely used for this purpose, each with strengths and limitations. These traditional techniques, however, often struggle with issues like omitting details, being time-consuming, or being sensitive to noise, particularly in diverse and complex staining conditions, while in digital pathology, accurate tissue segmentation in WSIs is essential for diagnostics, as missing tissue regions can lead to misdiagnosis, and false detection can increase processing times, as illustrated in Figure 1 These traditional techniques often fail to detect all tissue regions. Recently, deep learning (DL) models have improved segmentation accuracy by leveraging labeled data.



(a) Tissue regions detected by a commercial scanner's automated software where

(b) manual annotation of tissue regions.

(c) Tissue regions detected by a commercial scanner's automated software for an IHC-stained slide.

(d) manual annotation of tissue regions for an IHC-stained slide.

(e) RGB tissue mask created by OTSU method for the H&E stained tissue above.

Figure 1: Problems with the OTSU method as a semi-automated classical segmentation method. (a) shows the tissue areas detected by a commercial scanner's detection algorithm in an H&E-stained slide, where green contours highlight the tissue regions and red contours the region of interest (ROI), which are inaccurate. (b) shows the manually annotated tissue regions which are accurate. (c) shows that the scanner's detection algorithm detected all pixels and regions as tissue areas causing unnecessary processing and using more storage in an IHC-stained slide. (d) shows the manually annotated tissue regions. (e) Original RGB image and masked tissue regions for machine learning dataset preparation.

Large deep learning models, such as deep convolutional neural networks (CNNs) like ResNet, VGG, and specialized architectures like U-Net (Ronneberger et al. [2015])and Fully Convolutional Neural Networks (FCNNs)(Long et al. [2015]), have demonstrated remarkable success in various image analysis tasks, including tissue region segmentation. These models leverage complex architectures to capture intricate patterns in high-dimensional data, leading to high accuracy and robustness in diverse staining conditions. For instance,Hossain et al. [2023] applied DoubleU-Net, ResU-Net++ to tasks like artifact segmentation, and Bándi et al. [2017] applied FCNN and U-net CNN (UCNN) method to tasks like foreground extraction in HE and IHC stained tissues, comparing them with the Foreground Extraction from Structure Information (FESI) method. However, despite the advantages of large DL models in generalizing across diverse datasets (Szegedy et al. [2015]), they require significant computational resources and have long inference times, making them impractical for institutions with limited infrastructure and real-time clinical applications (Chen et al. [2016]). Their complexity also increases the risk of overfitting, especially with smaller datasets (Zhang et al. [2021]).

To address these limitations, our study introduces a lightweight CNN model, LeNet-5, designed for efficient, real-time tissue segmentation in resource-constrained healthcare centers (Fukushima [1980]). With a small memory footprint

and quick inference time, our model offers a cost-effective alternative for digital pathology. Additionally, it improves pre-processing by creating accurate tissue masks, outperforming traditional methods and enhancing downstream DL model performance.

2 Materials and Methods

2.1 Data Source

The success of this system heavily relies on training it on diverse and representative datasets, ensuring its accuracy and generalizability. Therefore, we have carefully chosen multiple datasets, including:

TCGA The Cancer Genome Atlas Program (TCGA) from the GDC Data Portal (GDC), which is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. TCGA data are in SVS file format and are suitable-quality breast cancer pathology tissue slides stained with haematoxylin and eosin (H&E) stain.

Camelyon17(H&E stained lymph node) CAMELYON 17 dataset (cam), contains whole-slide images (WSI) of hematoxylin and eosin (H&E) stained lymph node sections and is collected from 5 medical centers in the Netherlands. WSIs are provided as TIFF images and are organized by the patient. Patients consist of 5 lymph nodes, and every slide holds sections of just one lymph node.

Her2 scoring contest @ Nottingham Pathology 2016 Since IHC-stained slides are one of the most problematic in digital histopathology studies, especially in making tissue masks using image processing techniques, training our model with this dataset was essential. This contest provided a dataset of nearly 100 whole-slide images (100,000 x 80,000 pixels). By including the Her2 scoring contest dataset (her), which provides for both H&E and IHC stained slides, in our research, we aim to train our system on a more extensive dataset, allowing it to effectively predict and segment a broader range of tissue regions.

HEROHE EDCP2020 The WSIs for this dataset (ecd), were scanned in a 3D Histech Pannoramic 1000 digital scanner and saved in the MIRAX format. Each WSI is stored divided into one ".mrxs" and a folder (with the same base name) containing many .dat files and is a collection of digitized H&E-stained slides of breast cancer.

Bahcesehir Medical School This study has received Institutional Review Board approval (document no: 2022-10/03) from Bahcesehir University Clinical Research Institutional Review Board. By gathering the private data from Bahcesehir Medical School from both staining types(The H&E and IHC stained) of breast cancer tissues, we ensured that our model works for unseen and real-world clinical data. All images have been deposited at Zenodo under `https://zenodo.org/records/14131968` and are publicly available Naghshineh Kani et al. [2024a].

Prostate Cancer Grade Assessment (PANDA) Challenge It was important that our model could perform on other cancer types. Since with more than 1 million new diagnoses reported every year, prostate cancer (PCa) is the second most common cancer among males worldwide, resulting in more than 350,000 deaths annually, we improved our dataset by adding the PANDA (pan) prostate tissue slides which are H&E-stained biopsies from two centers.

Digital pathology dataset for prostate cancer diagnosis (TTS Singapore) Digitized hematoxylin and eosin (H&E)-stained whole-slide-images of 40 prostatectomy and 59 core needle biopsy specimens were collected from 99 prostate cancer patients at Tan Tock Seng Hospital, Singapore (Oner et al. [2022]). There were 99 WSIs in total, such that each specimen had one WSI. H&E-stained slides were scanned at 40× magnification (specimen-level pixel size0.25 μm $\times$ 0.25 μm) using Aperio AT2 Slide Scanner (Leica Biosystems).

2.2 Dataset Preparation

All the steps of dataset preparation are shown briefly in Figure 2 and are discussed below:

Data segregation Data segregation is an important step in dataset preparation, where patients are divided into training, validation, and test sets. The training set trains machine learning models, while the validation set refines hyperparameters and chooses the best model. The model's performance is then tested on unseen patients to simulate a

Tissue	Stain	Center	Count
Breast	H&E	TCGA	81
Lymph node	H&E	CAMELYON17	66
Breast	H&E	HEROHE	148
Breast	H&E	Her2-HE	75
Breast	IHC	Her2-IHC	86
Prostate	H&E	PANDA	50
Prostate	H&E	TTS Singapore	89
Breast	H&E	BAU-HE	72
Breast	IHC	BAU-IHC	107

Table 1: The 9 different tissue-staining categories of the development dataset were used. The dataset was randomly divided into training, validation, and testing sets with a 5:2:3 distribution, respectively.

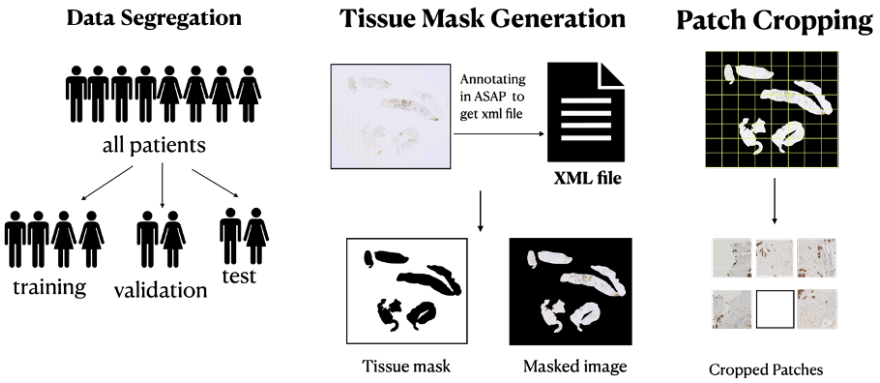


Figure 2: Dataset preparation

real-world scenario. Data segregation at the patient level is crucial to avoid data leakage problems (Oner et al. [2020]), such as the correlation of tumor slides. In this study, slides from different cohorts were segregated with a 70% training, 10% validation and 20% test distribution.

Slide annotation We first annotated the WSIs individually in an annotation tool called ASAP (ASA) which is an annotation tool used primarily for image annotation to mark regions of interest within an image and associate them with specific labels or metadata (Figure 3).

Tissue mask generating The annotations made in ASAP are typically stored in an XML file format, including information about the coordinates of the annotated regions on the image. These coordinates define the shape and location of the marked and annotated areas within the image. Using the obtained XML files, we created the tissue mask over a down-sampled version of a WSI where black areas of the binary tissue mask show tissue regions within the slide. This way, we make labeled datasets for the model we are training,as shown in Figure 4.

Cropping In this study cropping patches across tissue areas in WISs was the last stage in creating datasets.The reason for cropping the WSIs is their huge size, whereas most DL models work with small-sized images and cropping the WSIs significantly reduces the computational burden during training or analysis. Using a sliding window over the mask, patches are cropped from the desired resolution level of the slide using binary tissue masks as guides. Depending on the purpose, overlapping or non-overlapping sliding window methods may be used for cropping.
During cropping, we label the tissue crops based on the hand-labeled mask. For each cropped region, if there is at least 10% tissue area on the mask of the same region, we check the intensity values of the RGB channels. If the average intensity is lower than 240 out of 255, we label the region as tissue.
In this study we aimed to build a model that had minimum FLOPs and inference time; To achieve this, we cropped the WSIs at level 4 with patch sizes of 128*128 (L4-128).
FLOPS stands for Floating Point Operations Per Second. In the context of DL models, it refers to the computational complexity of the model, specifically the total number of arithmetic operations required for a single forward pass.

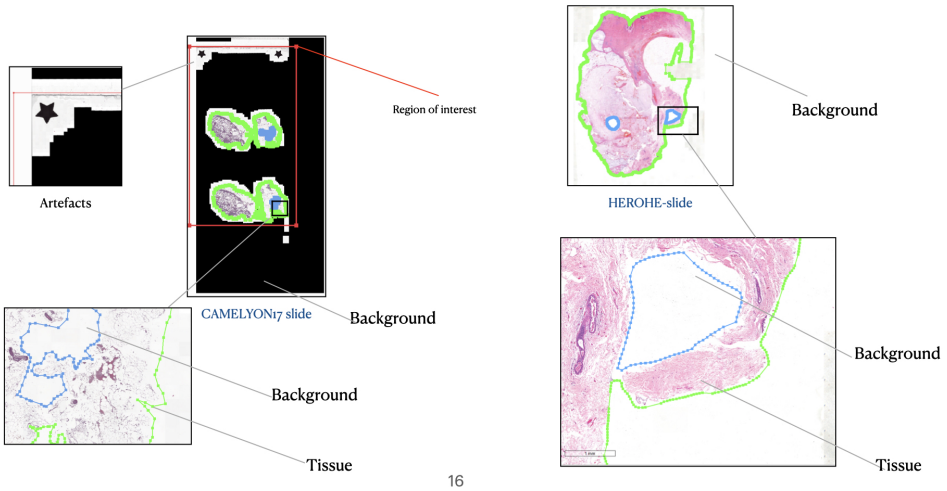


Figure 3: Tissue annotation is green, and background annotation is blue. Anything outside the tissue region is considered background. If the size of the tissue compared to the background is very small, to balance the classes, the tissue region and background are placed within a region of interest (ROI) box, which is red. Some WSIs have both black and white backgrounds.

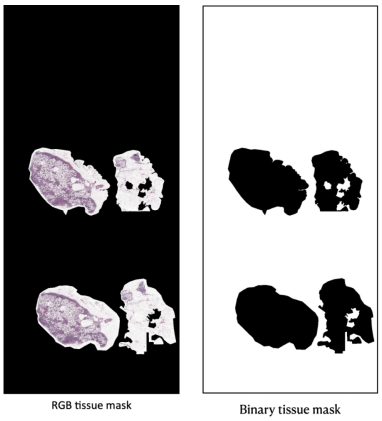


Figure 4: Created ground truth tissue masks on Level 5 with a Python code to parse the XML files for a WSI.

Higher FLOPS implies more complex calculations and, therefore, higher resources (like time and energy) needed to run the model.
To calculate the models' FLOPs and their computational efficiency to estimate the model performance, we used a Facebook library named fvcore for this task. fvcore is the first tool that can provide both operator-level and module-level flop counts together and contains a flop-counting tool for PyTorch models.

2.3 OTSU

Over time, various image segmentation algorithms and techniques have been developed using domain-specific knowledge to address specific applications like medical imaging, automated driving, video surveillance, and machine vision. Each method has advantages and disadvantages. The threshold method Otsu [1979]), which works best with high-contrast pictures, separates an image into foreground and background areas depending on the threshold. Still, many details can get omitted during thresholding, and threshold errors are common. The process of generating a binary mask involves basic image processing operations over a down-sampled version of a WSI, including an OTSU thresholding, morphological operations, and median filtering to eliminate noise.

2.4 Deep learning model

Another method employed in our research is the integration of LeNet5 (LeCun et al. [1998]), a very efficient CNN for handwritten character recognition. What makes LeNet5 so unique for this research is its small network to achieve our aim of having small FLOPs and inference time so the model can be efficiently run and used without GPU and easily applied and used in microchips. LeNet5 contains the basic modules of DL: convolutional layer, pooling layer, and full link layer, which are the basis of other DL models. By leveraging LeNet5, conducting various experiments, and performing several internal and external validations, our system can identify tissue regions, enabling more precise and targeted segmentation of tissue regions in WSIs. we modified LeNet5 for RGB images in different sizes. As expected, after calculating the models' FLOPs and size using the Facebook library fvcore, LeNet5-16 had the minimum FLOPs, with a total FLOPs of 0.309768 MB, and LeNet5-128 had the maximum FLOPs of 19.18 MB. All original code has been deposited at Zenodo under `https://zenodo.org/records/14294426` and is publicly available Naghshineh Kani et al. [2024b]. The repository provides a detailed step-by-step explanation, from preparing the ML dataset and training of tissue segmentation models to inference with the trained model.

Training and Evaluation Before loading the data to CNN model for training, The cropped patches were changed into tensors and be normalized based on the mean and standard of the training set patches and were loaded underwent a sequence of augmentation procedures to enhance the resilience of our networks to the diverse range of data fluctuations seen in real-world scenarios.
Our objective is to accurately replicate the distinctions among scanners, stains, and staining processes.By implementing data augmentation techniques like rotation and flipping, the number of patches in a dataset can be increased, and it may help mitigate over-fitting issues. Also by using ColorJitter transformation from the transforms module in the torchvision library, the hue of the patches were adjusted by a random amount within the range [-0.4, 0.4].to adjust hue and doing color adjustments, we made every color possible on the image patches to make model to be able to recognize tissue with with unseen stains that were not included in this study.
We pushed the data into our CNN model with a batch size of 1024 for training with 1000 epochs with the learning rate of LR = 0.0001 and used Cross-Entropy loss or log loss as a loss/cost function to optimize the model during training in our classification model. We used Stochastic gradient descent (SGD) optimizer to minimize the loss function and used momentum to accelerate the gradient vectors in the right direction in SGD to lead to a faster converging. To prevent overfitting, we used weight decay (also known as L2 regularization) in the SGD optimizer, which works by adding a penalty term to the loss function during training which aims not only to minimize the error between the predicted output and the actual target but also to minimize the magnitude of the weights.We didn't do any batch normalization or dropouts because we wanted to stick to the original architecture of LeNet5 as much as possible. Training and validation in each epoch took approximately 58 seconds for all the datasets in all our experiments.

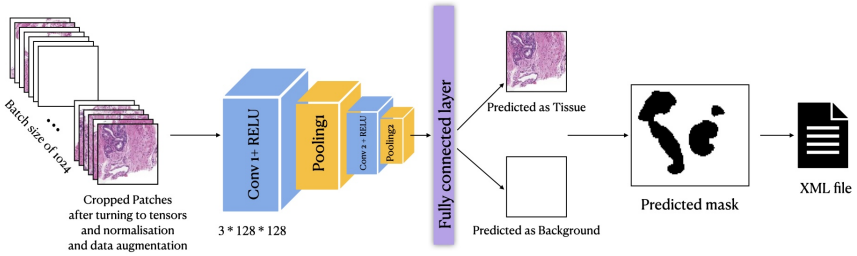


Figure 5: LeNet5 for RGB image with a patch size of 128*128.

Inference To prove how the segmentation model outperform OTSU method, we quantitatively compared the algorithms using the Dice score (Dice [1945]) and Jacard Index(Gilbert [1884]), comparing the output of the algorithms and generated masks by OTSU against the binary reference masks of the 70 different types of tissue images in the testing part of the development dataset. By giving a WSI to our inference segmentation model, we obtain a binary tissue mask and XML file of coordination of annotations for the tissue regions 6.

3 Results

Overall 10 LeNet5 model were trained on the prepared datasets. The last model was tained using a combination of H&E and IHC-stained breast cancer and H&E-stained prostate and lymph node biopsies which we named this model CNN-Final-L4-128. The results for the trained models on L4-128 with different data sources are shown in table 2

Figure 6: Tissue region Segmentation inference.

Data Source	train loss	train accuracy	Val loss	Val accuracy	test loss	test accuracy	Val AUC	Test AUC
TCGA_L4_128	0.0800	97.6037	0.0695	97.8335	0.0896	97.2937	0.9947	0.9914
CAMELYON17_L4_128	0.0507	98.3818	0.0653	97.4784	0.0645	97.9061	0.9979	0.9955
HER2_HE_L4_128	0.0773	97.6447	0.0524	98.4615	0.0716	97.7764	0.9944	0.9939
HER2_IHC_L4_128	0.0513	98.6427	0.0515	98.7623	0.1327	96.7706	0.9895	0.9830
PANDA_Karolinska_L4_128	0.0086	99.7070	0.0058	99.7851	0.0414	98.5288	0.9999	0.9995
BAU_HE_L4_128	0.1036	96.4383	0.0510	98.5595	0.0926	96.5633	0.9981	0.9926
BAU_IHC_L4_128	0.0469	98.3711	0.0463	98.4536	0.0420	98.5250	0.9976	0.9970
HEROHE_L4_128	0.0523	98.2668	0.0073	99.7835	0.0077	99.7748	0.9995	0.9995
TTS_Singapore_L4_128	0.0295	99.1058	0.0223	99.2887	0.0448	97.9653	0.9995	0.9994
CNN-Final_L4_128	0.0531	98.3798	0.0152	99.5415	0.0281	99.0849	0.9989	0.9984

Table 2: Training, validation, and testing's loss and accuracy and validation and testing's AUC results on trained models on L4-128 with different data sources.

Results on dissimilar dataset It is crucial for DL models trained for histopathology tasks to be validate by completely unseen data from other cohorts to make sure the model will show the same performance on real-world data. For thus, we did external validation on the trained models on previous experiments, with the dataset explained in Table3 as well as the results of AUC value of external validation on the mentioned datasets compared to their unseen similar test dataset. the combined data for CNN-Final-l4-128 includes data from TTS-Singapore, HEROHE, HER2-HE,HER2_IHC and Camelyon17. unfortunatly there was not another data source to do external validation on Camelyon17.

Model	External test set Including:	Test set AUC	External validation AUC
TCGA_L4_128	HER2-HE, HEROHE, BAU-HE	0.9914	0.9778
HER2_HE_L4_128	TCGA, HEROHE, BAU-HE	0.9939	0.9398
HEROHE_L4_128	TCGA, HER2-HE, BAU-HE	0.9995	0.9724
BAU_HE_L4_128	TCGA, HEROHE, HER2-HE	0.9926	0.9568
HER2_IHC_L4_128	BAU-IHC	0.9830	0.9988
BAU_IHC_L4_128	HER2_IHC	0.9970	0.9988
TTS-Singapore_L4_128	PANDAS_Karolinska	0.9994	0.9988
PANDAS_Karolinska_L4_128	TTS-Singapore	0.9995	0.9988
CNN_Final_L4_128	Combined-data	0.9984	0.9988

Table 3: External test dataset summary with the comparison of the AUC value on their unseen similar and dissimilar test datasets.

Results of segmentation Inference The Confidence Intervals(CIs) of Jaccard index was (0.66-0.93) for The segemntation model while this range changed to (0.02-0.94) for OTSU. in the same manner, the CIs were (0.81-0.97) for the Segmentation model while this changes to (0.05-0.98) for OTSU model.
It should be mentioned that OTSU method highly owes its reults to the prostate biopsies which were perfectly scanned and the pixels of background and tissue region were in high contrast and they did not have fatty regions but as shown in figure 7, the OTSU method not only fails in detecting all of the tissue regions but also detects an artifact as tissue region; This is while our segmentation model detects the all of the tissue regions correctly while it can detect the artifact as background and discriminate it from tissue. In figure 8, our CNN model can detect the fatty areas of tissue regions easily with high probability while the OTSU fails. The heat maps of the predictions using the probabilities of our model's predictions show that it has predicted the tissue regions with high probabilities for the example slides9.

Figure 7: Binary masks for an IHC-stained WSI using OTSU and our model, compared to the mask generated from manual annotation.

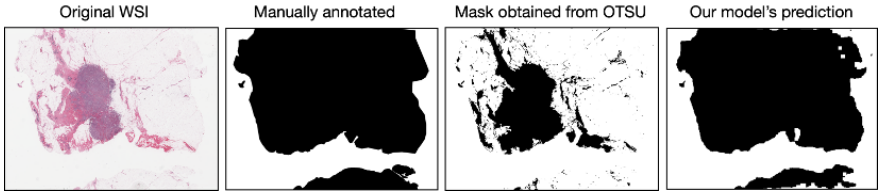


Figure 8: Binary masks for a HE-stained breast tissue WSI using OTSU and our model, compared to the mask generated from manual annotation.

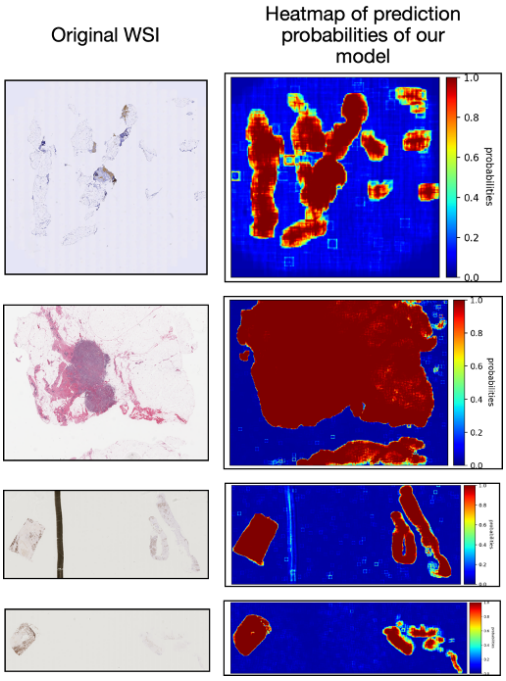


Figure 9: Heatmap of the model's predictions for the generating mask for example slides.

4 Discussion

Different tissue types, such as breast cancer tissues, lymph nodes, and prostate biopsies, affected the model's performance. IHC-stained breast cancer tissues had poor performance, with AUCs under 0.97 and F1 scores under 0.94 for HER2-IHC and 0.67 for BAU-IHC. One reason for the poor performance of a model just trained by IHC-stained data was that the model considered noises and artifacts as tissue. After all, the IHC-stained tissues in the training set looked like noise and artifacts. The appearance of IHC-stained slides looks so noisy that, sometimes, as a human, even by zooming in, it was hard to distinguish between the tissue and the noise during annotating. This difficulty increases when the models are trained on down sampled levels rather than in the most magnificent.
Lymph node biopsies showed better performance, as they contain fatty areas and are a constant source of difficulty for tissue segmentation techniques. The CAMELYON17 dataset had white and black backgrounds and large and small tissue areas with markers, signs, and artifacts, allowing the LeNet5 to see and learn as many variant features as possible. Comparing results based on institutes demonstrates the significance of data origin and quality in model performances.

Comparing the results of segmentation inference with OTSU,the tissue mask generated by the OTSU method does not cover the whole tissue region. Instead, it only generated a tissue mask for a tiny area, missing crucial data. This might lead to severe issues in labs if the scanner operator cannot identify the issue and scan the material correctly. This might result in significant problems in the lab. This means that the pathologist can only see the whole tissue in WSI if they rescan or use an antiquated method that involves a microscope; also, when using DL models like the ones whose job is to grade HER2 scores and identify abnormalities or do prognosis. Additionally, the OTSU method detected the artifact as tissue and incorrectly masked it in the slide. In some other cases, the OTSU approach creates a tissue mask for the whole area of WSI, which raises storage expenses.

Although our upper boundaries could not reach 0.99 for the Jaccard index metric like UCNN and FCNN in Bándi et al. [2017], but still the LeNet5 CNN solutions were proven superior to traditional methods in IHC-stained tissues where the mean Jaccard index for HER2-IHC slides using OTSU was 0.35 and for our model was 0.84 and for BAU-IHC data this value for OTSU was 0.65 while it was 0.82 for our model. Comparing our model to OTSU, using Dice coefficient metrics demonstrated the significant difference and outperformance, where the Dice coefficient is a more accurate metric for this task.

Because of the diversity of the training dataset, the model with integrated data(CNN-Final-L4-128) performed better results with an average value of 0.94 of the Dice coefficients in all the tissues. Ultimately, although not all of the models that were just trained with data of only one origin did not function as predicted, all of them were able to detect and segment the fatty tissues, while most of the time, the OTSU method failed.

5 Conclusion

Addressing multifaceted challenges is imperative for advancing the accuracy and robustness of these models in real-world applications.The success of tissue segmentation models in digital pathology imaging is contingent upon many factors.
Paramount among these are the quality and origin of the data, as evidenced by this study's findings. Moreover, the nuanced characteristics of the samples, such as tissue types, composition, and the inherent complexities of the training process, significantly influence model performance. Specifically, mitigating the impact of artifacts and noise in images, comprehending the intricate variations within tissues, and continually optimizing training methodologies are crucial. These tasks enhance the models' accuracy and fortify their adaptability to diverse scenarios encountered in clinical or research settings.

Our model's ability to handle tissue segmentation intricacies, especially in distinguishing fatty tissues from IHC-stained ones, showcases its potential utility in practical clinical applications. As digital pathology imaging continues to evolve, refining tissue segmentation models remains an ongoing pursuit.

In future work, we plan to repeat the experiments where our models with more sophisticated CNN architecture and use knowledge distillation (Gou et al. [2021]) to keep the same small architecture for the inference while our models could generalize more with having higher performance. Also, we want to add more classes to the annotations, such as artifact or tissue border, to make the segmentation more precise. This continuous pursuit of improving tissue

segmentation models enhances their efficacy in clinical diagnostics and research and holds promise for revolutionizing biomedical imaging practices, ultimately benefiting patient care and scientific advancements.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The dataset is publicly accessible at `https://zenodo.org/records/14131968` Naghshineh Kani et al. [2024a], and the accompanying code is available on `https://github.com/sepidehnaghshineh/Tissue-region-segmentation-in-HE-and-IHC-stained-pathology-slides` Naghshineh Kani et al. [2024b].

Acknowledgment

This work was supported by BAU BAP (Grant No. 2022-02.13). Ethical approval for this study was obtained from the Bahçeşehir University Clinical Research Institutional Review Board (IRB approval number: 2022-10/03).

References

Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

Richard Nock and Frank Nielsen. Statistical region merging. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11):1452–1458, 2004.

Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598, 1991.

John F Canny. A variational approach to edge detection. In *AAAI*, volume 1983, pages 54–58, 1983.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

Md Shakhawat Hossain, Galib Muhammad Shahriar, MM Mahbubul Syeed, Mohammad Faisal Uddin, Mahady Hasan, Md Sakir Hossain, and Rubina Bari. Tissue artifact segmentation and severity assessment for automatic analysis using wsi. *IEEE Access*, 11:21977–21991, 2023.

Péter Bándi, Rob van de Loo, Milad Intezar, Daan Geijs, Francesco Ciompi, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. Comparison of different methods for tissue segmentation in histopathological whole-slide images. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 591–595. IEEE, 2017.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2487–2496, 2016.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

National Cancer Institute. https://portal.gdc.cancer.gov. Accessed: 2022-12-01.

camelyon17.grand-challenge. https://camelyon17.grand-challenge.org.

her2contest. https://warwick.ac.uk/fac/cross_fac/tia/data/her2contest/.

ecdp2020.grand-challenge. https://ecdp2020.grand-challenge.org/Dataset/.

Sepideh Naghshineh Kani, Burak Can Soyak, Mert Gokce, Zeynep Duyar, Hacer Alicikus, Özlem Yapıcıer, and Murat U. Oner. Digital pathology dataset for breast cancer diagnosis [data set], 2024a. URL https://doi.org/10.5281/zenodo.14131968.

panda.grand-challenge. https://panda.grand-challenge.org.

Mustafa Umit Oner, Mei Ying Ng, Danilo Medina Giron, Cecilia Ee Chen Xi, Louis Ang Yuan Xiang, Malay Singh, Weimiao Yu, Wing-Kin Sung, Chin Fong Wong, and Hwee Kuan Lee. An ai-assisted tool for efficient prostate cancer diagnosis in low-grade and low-volume cases. *Patterns*, 3(12), 2022.

Mustafa Umit Oner, Yi-Chih Cheng, Hwee Kuan Lee, and Wing-Kin Sung. Training machine learning models on patient level data segregation is crucial in practical clinical applications. *medRxiv*, pages 2020–04, 2020.

ASAP. https://computationalpathologygroup.github.io/ASAP/.

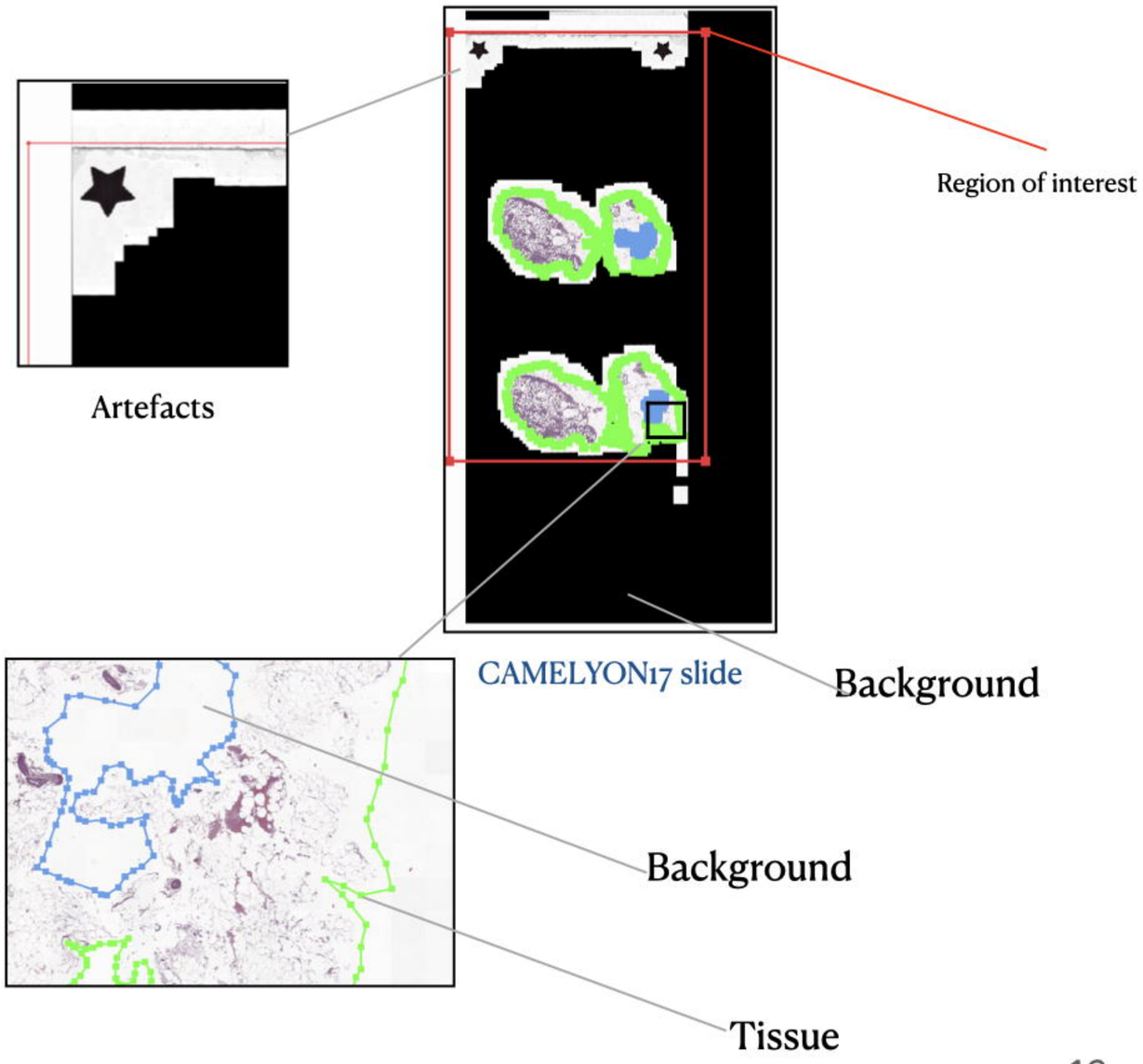
Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

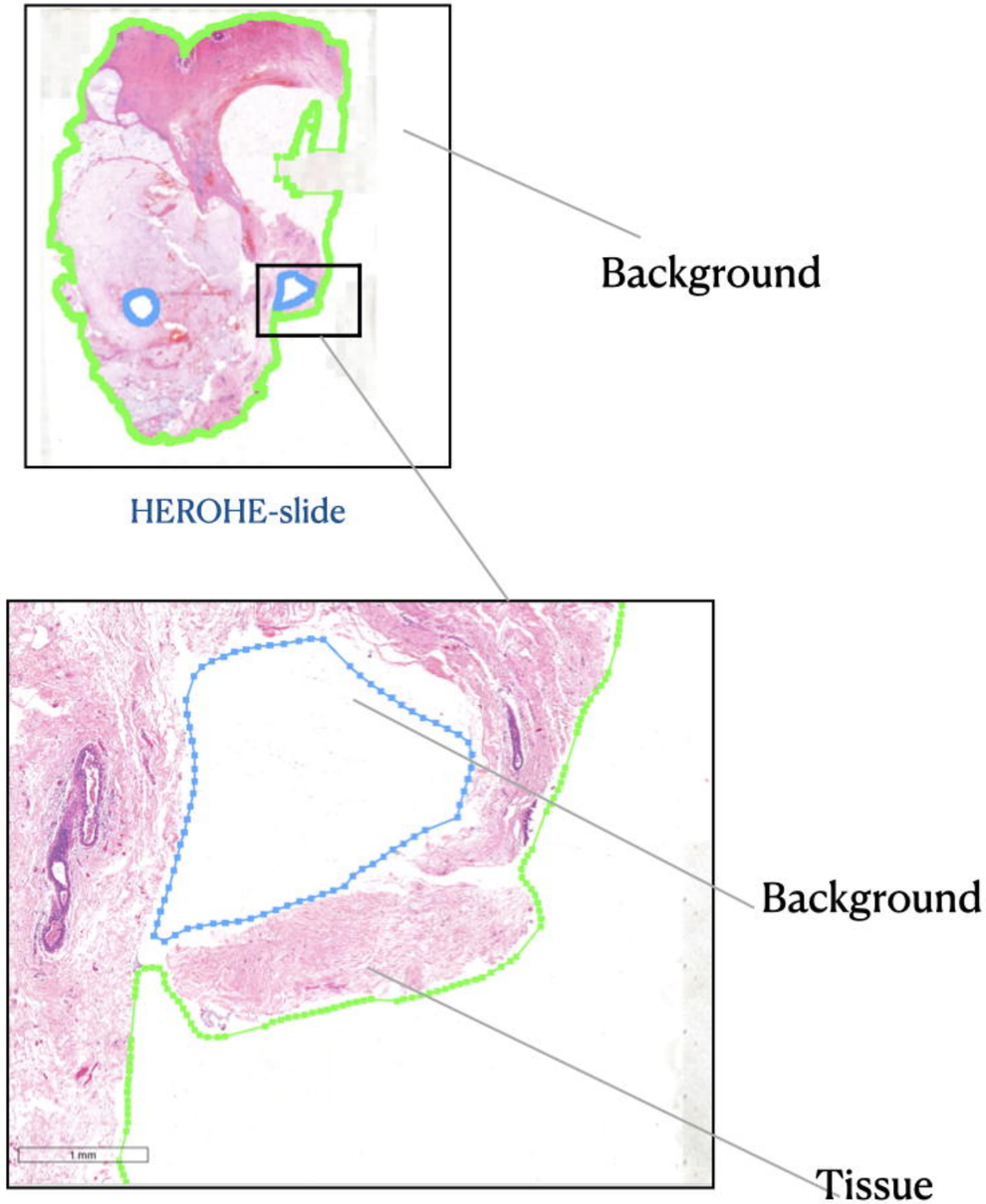
Sepideh Naghshineh Kani, Burak Can Soyak, Mert Gokce, Zeynep Duyar, Hacer Alicikus, Özlem Yapıcıer, and Murat U. Oner. Tissue region segmentation in he-stained and ihc-stained pathology slides of specimens from different origins. *Zenodo*, 2024b. doi:10.5281/zenodo.14294426. URL https://doi.org/10.5281/zenodo.14294426.

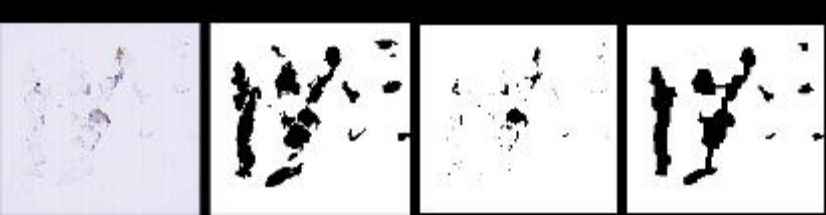
Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

Grove Karl Gilbert. Finley's tornado predictions. *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, 1(5):166, 1884.

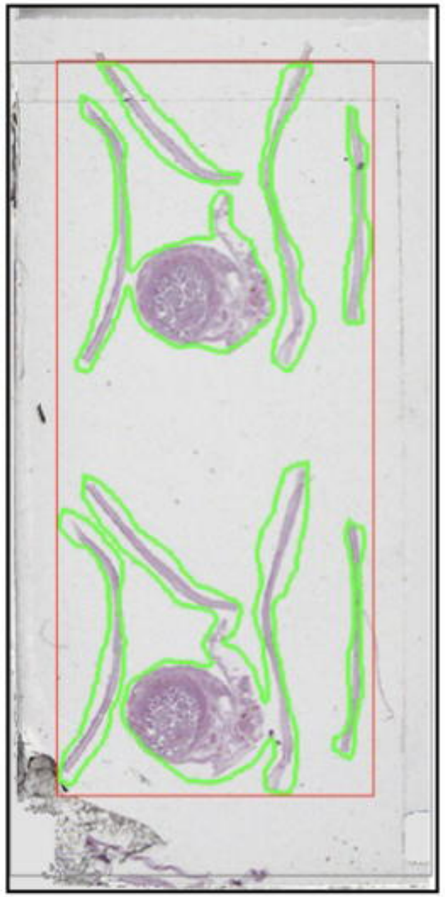
Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819, 2021.


Artefacts
Region of interest
CAMELYON17 slide
Background
Background
Tissue


Background
HEROHE-slide
Background
Tissue



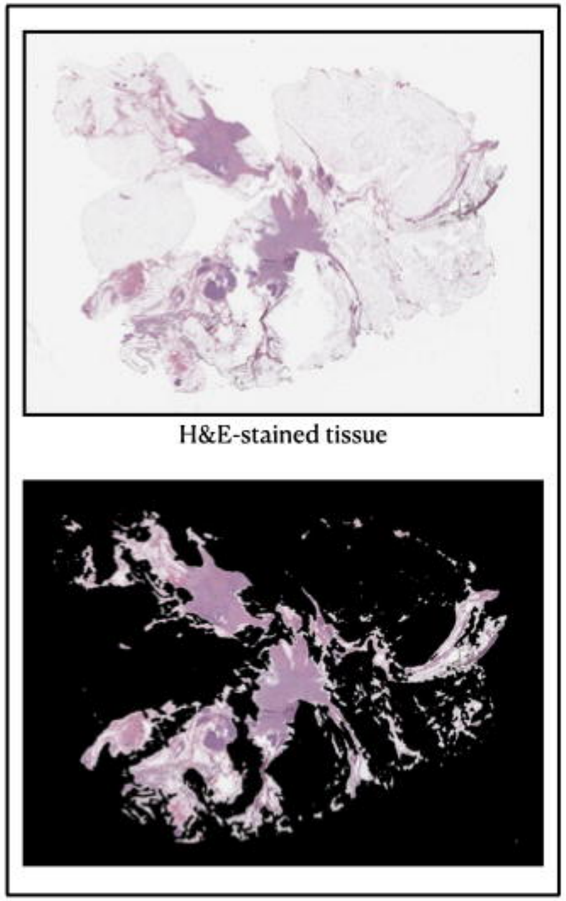
(a) Tissue regions detected by a commercial scanner's automated software where



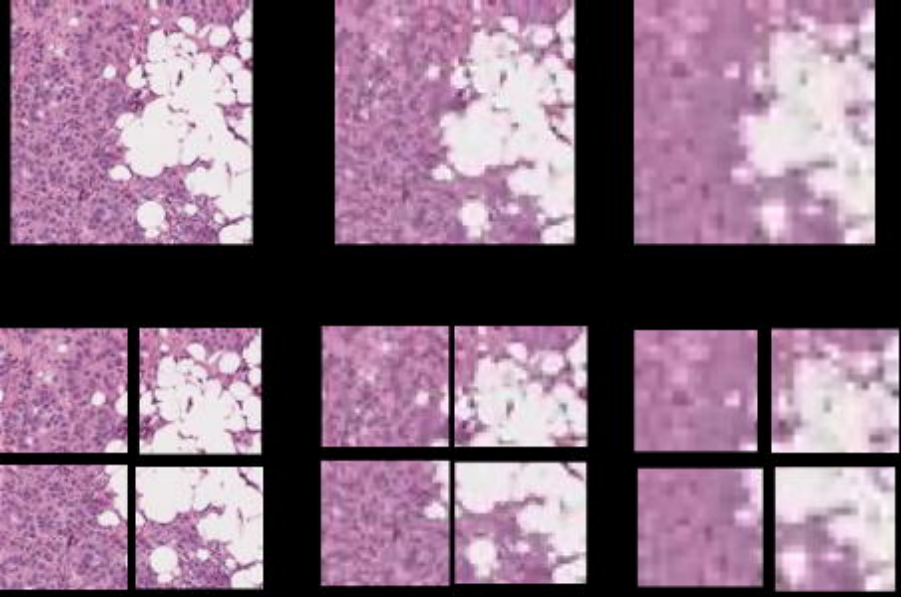
(b) manual annotation of tissue regions.

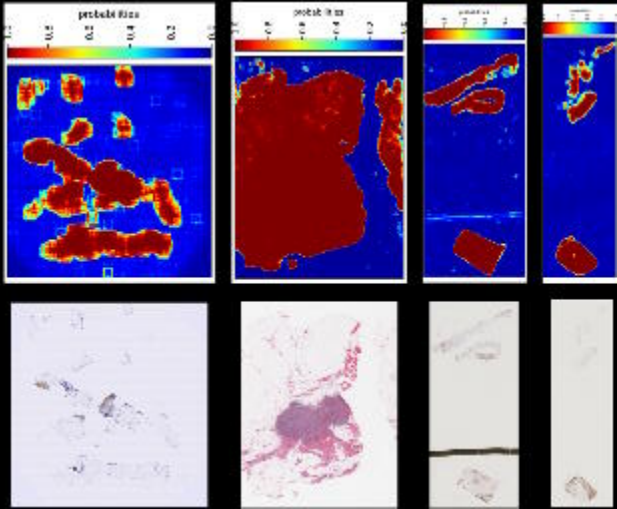
(c) Tissue regions detected by a commercial scanner's automated software for an IHC-stained slide.

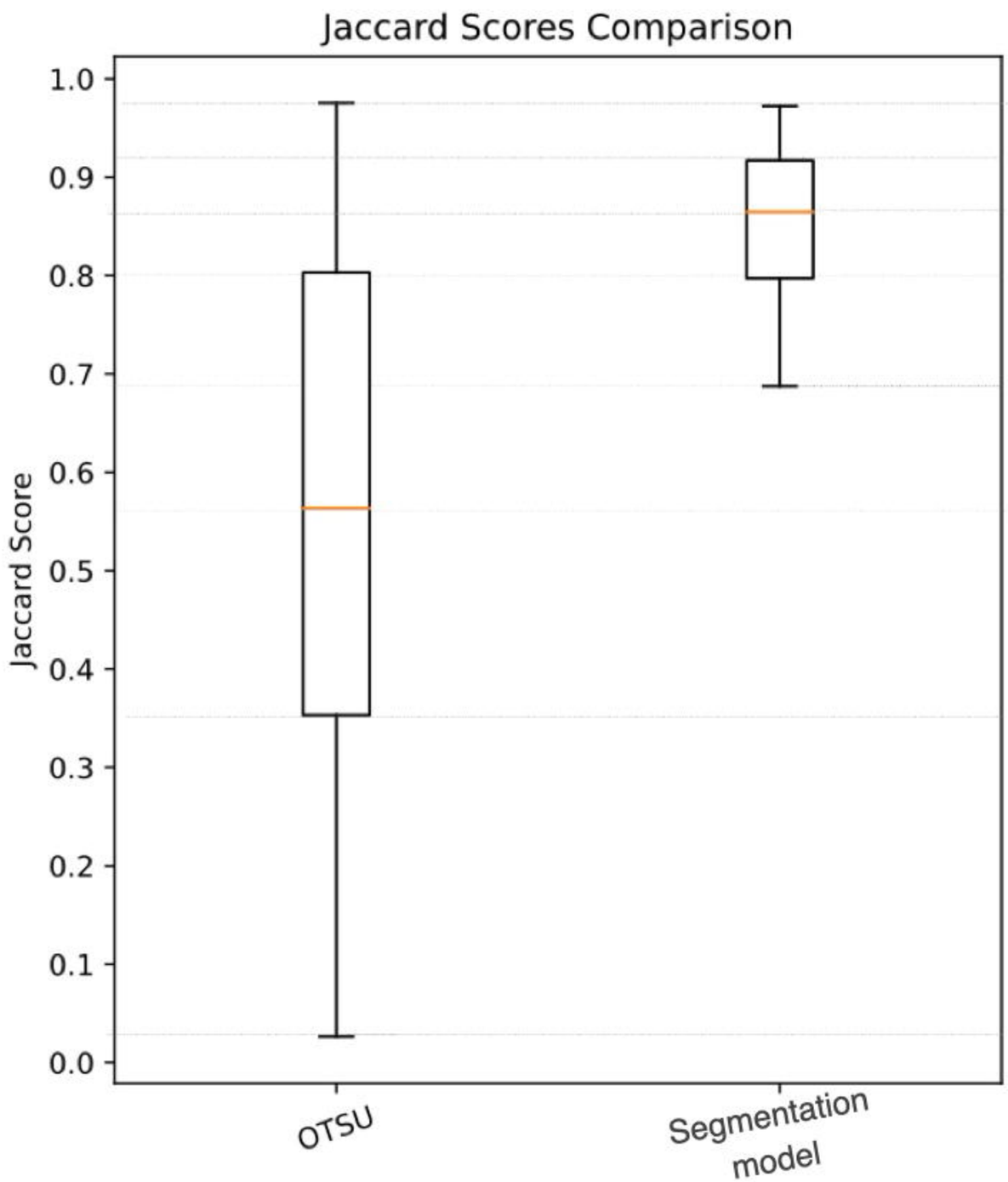
(d) manual annotation of tissue regions for an IHC-stained slide.

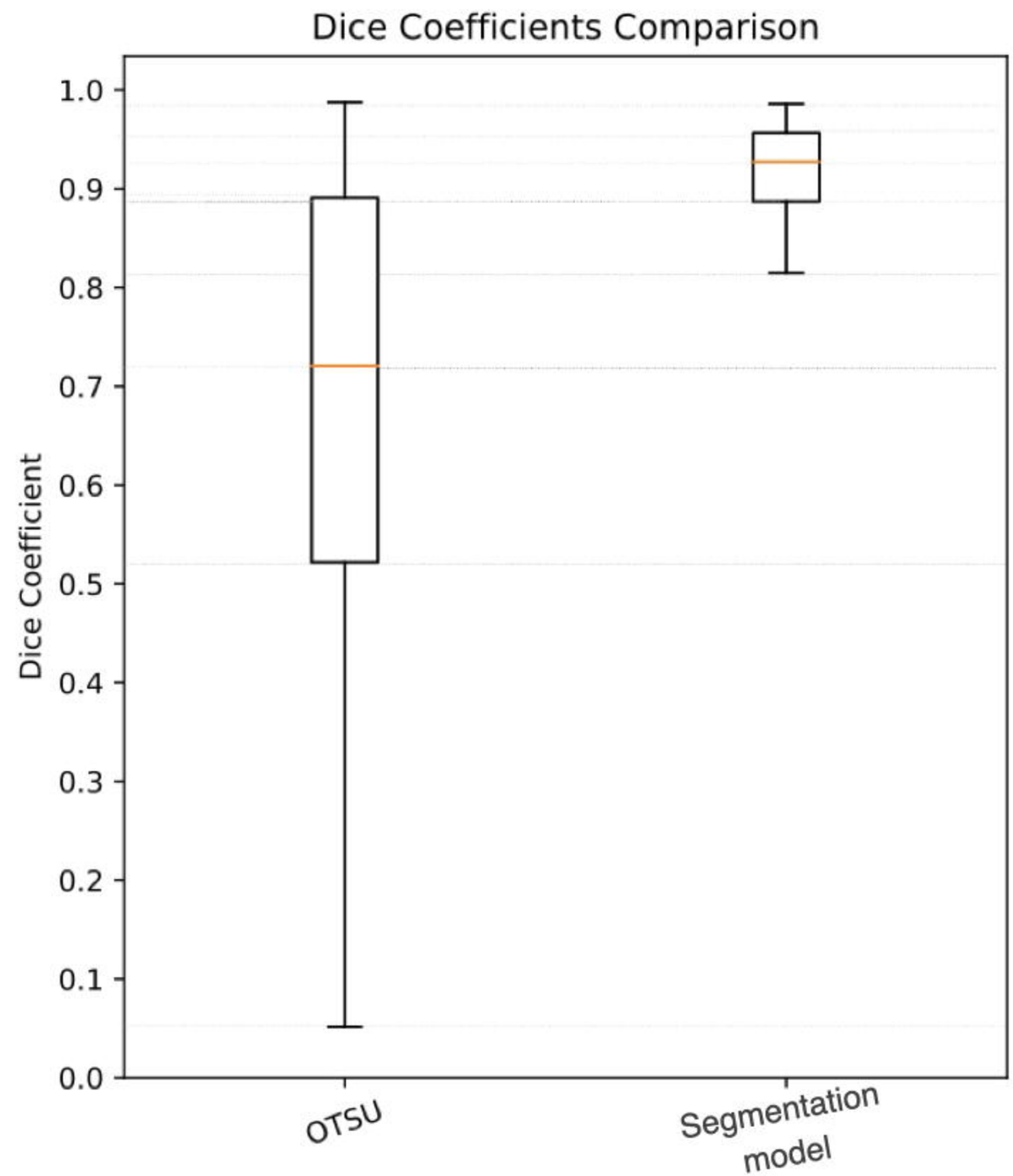


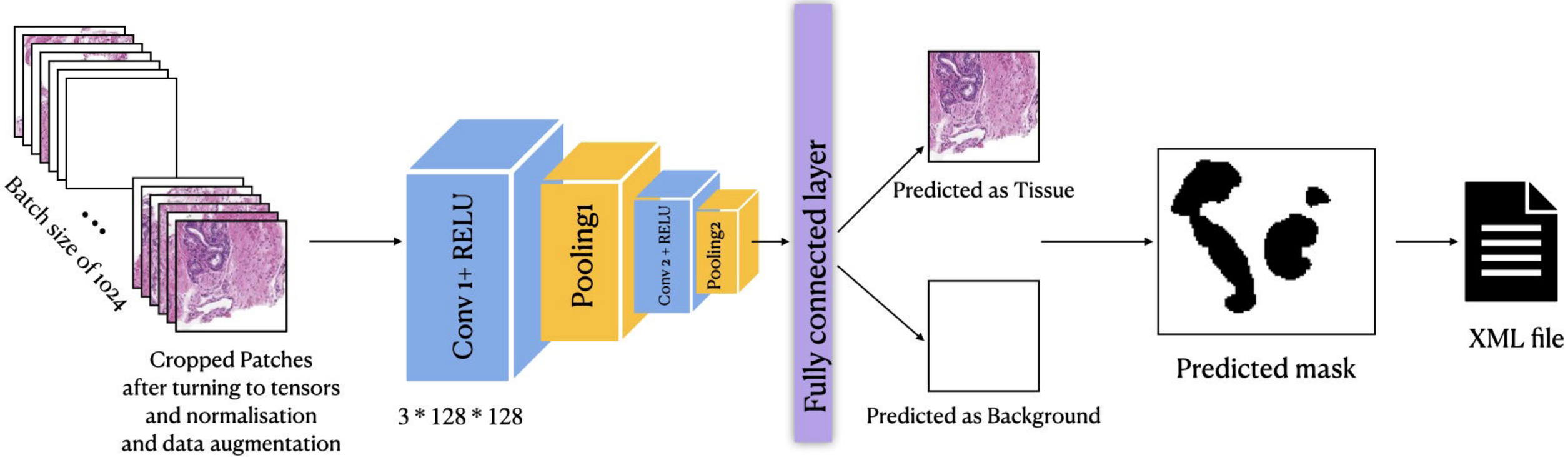
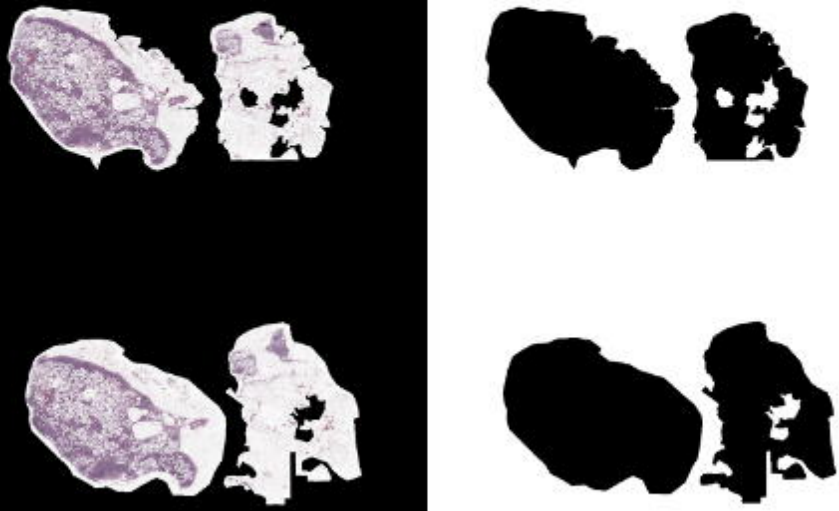
(e) RGB tissue mask created by OTSU method for the H&E stained tissue above.






Jaccard Scores Comparison
Jaccard Score
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OTSU
Segmentation model


Dice Coefficients Comparison
Dice Coefficient
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
OTSU
Segmentation model


Batch size of 1024
Cropped Patches
after turning to tensors
and normalisation
and data augmentation
Conv 1+ RELU
Poolingl
Conv 2 + RELU
Pooling2
3 * 128 * 128
Fully connected layer
Predicted as Tissue
Predicted as Background
Predicted mask
XML file