

Automatic segmentation of histopathological glioblastoma whole-slide images utilizing MONAI

Ellena Spiess, Dominik Müller, Moritz Dinser, Volker Herbort,
Friederike Liesche-Starnecker, Johannes Schobel, Daniel Hieber

Angaben zur Veröffentlichung / Publication details:

Spiess, Ellena, Dominik Müller, Moritz Dinser, Volker Herbort, Friederike Liesche-Starnecker, Johannes Schobel, and Daniel Hieber. 2025. "Automatic segmentation of histopathological glioblastoma whole-slide images utilizing MONAI." In Intelligent health systems - from technology to data and knowledge: proceedings of MIE 2025, edited by Elisavet Andrikopoulou, Paris Gallos, Theodoros N. Arvanitis, Rosalynn Austin, Arriel Benis, Ronald Cornet, Panagiotis Chatzistergos, et al., 88–92. Amsterdam: IOS Press.
<https://doi.org/10.3233/shti250279>.

Automatic Segmentation of Histopathological Glioblastoma Whole-Slide Images Utilizing MONAI

Ellena SPIESS^a, Dominik MÜLLER^{b,c}, Moritz DINSER^{d,e,f}, Volker HERBORT^a,
Friederike LIESCHE-STARNECKER^f, Johannes SCHOBEL^{d 1} and
Daniel HIEBER^{d,e,f 1 2}

^aFaculty of Computer Science, Ulm University of Applied Sciences

^bFaculty of Applied Computer Science, University of Augsburg

^cInstitute for Digital Medicine, University Hospital Augsburg

^dDigiHealth Institute, Neu-Ulm University of Applied Sciences

^eInstitute of Medical Data Science, University Hospital Würzburg

^fDepartment of Neuropathology, Pathology, Medical Faculty, University of Augsburg

ORCID ID: Dominik Müller <https://orcid.org/0000-0003-0838-9885>, Moritz Dinsler

<https://orcid.org/0000-0001-5700-4393>, Volker Herbort

<https://orcid.org/0009-0004-3733-0879>, Friederike Liesche-Starnecker

<https://orcid.org/0000-0003-1948-1580>, Johannes Schobel

<https://orcid.org/0000-0002-6874-9478>, Daniel Hieber

<https://orcid.org/0000-0002-6278-8759>

Abstract. Manual segmentation of histopathological images is both resource-intensive and prone to human error, particularly when dealing with challenging tumor types like Glioblastoma (GBM), an aggressive and highly heterogeneous brain tumor. The fuzzy borders of GBM make it especially difficult to segment, requiring models with strong generalization capabilities to achieve reliable results. In this study, we leverage the Medical Open Network for Artificial Intelligence (MONAI) framework to segment GBM tissue from hematoxylin and eosin-stained Whole-Slide Images. MONAI performed comparably well to state-of-the-art AutoML tools on our in-house dataset, achieving a Dice score of 79%. These promising results highlight the potential for future research on public datasets.

Keywords. Tumor Segmentation, Neuropathology, Computer Vision, Machine Learning, MONAI

1. Introduction

Glioblastoma (GBM) is among the most aggressive and deadly forms of brain tumors, presenting significant challenges in both diagnosis and treatment – especially due to their high heterogeneity [1]. Effective segmentation of GBM tissue in histopathological im-

¹Contributed equally & Shared supervision

²Corresponding Author: Daniel Hieber, e-mail: daniel.hieber@hnu.de.

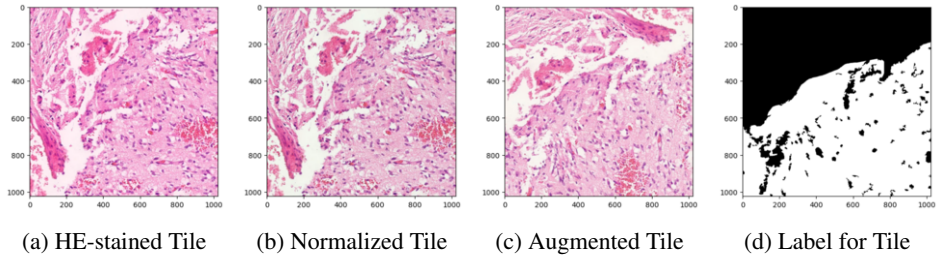


Figure 1. Visualization of an Example Tile (a) Prior and (c) After Preprocessing with its According Label (White is Tumor) (d) and the Effect of Stain-Normalization (b).

ages is crucial for precise treatment planning [2]. Despite being the standard in many clinical settings, manual segmentation is labor-intensive, costly and subject to inter-observer variance due to its reliance on expert interpretation [3]. With the advancement of machine learning, automated segmentation methods have become viable alternatives, offering improvements in both speed and accuracy [1,4].

The *Medical Open Network for Artificial Intelligence* (MONAI) is an open-source framework designed explicitly for analyzing medical images [4]. It provides comprehensive tools to develop, train, and evaluate models tailored to complex tasks, like the segmentation or classification of tumor tissue. The study presented in this manuscript evaluates whether the MONAI framework can be applied for GBM segmentation from hematoxylin and eosin (HE)-stained Whole-Slide Images (WSIs). Code and model are available at: https://github.com/hnu-digihealth/monai_gbm_segmentation

2. Methods

The model creation is split into three parts: (1) Data preparation and preprocessing, (2) model development, and (3) model evaluation. All steps are based on the MONAI framework, although other frameworks are used for sub-tasks.

2.1. Data Preparation and Preprocessing

An anonymized in-house dataset with 20,236 tiles from 103 HE-stained WSIs of 56 GBM patients is used in this study. Each tile has a dimension of 1,024 x 1,024 pixels. A label marking neoplastic tissue was assigned to each tile, created by a trained neuropathologist. The labels were refined using simple thresholding algorithms to reduce off-tissue labeling. Given the variability in histopathological images, the data augmentation techniques random rotations, random flips, random contrast adjustment, and gaussian noise were applied to increase the diversity of the training data and improve model robustness [5]. All of these augmentations and the tile preprocessing are handled online during the model training. Additionally, stain-normalization was applied to address variations in color intensity, which may negatively impact the model's generalizability [6]. Although HE-staining is standardized, little differences in the chemical composition, staining time, or the scanner used to digitize the WSI can introduce significant differences in hue and saturation [6]. The stain-normalization was realized using the *modified Reinhard method* [7] from the *torchstain* framework [8]. The normalization was imple-

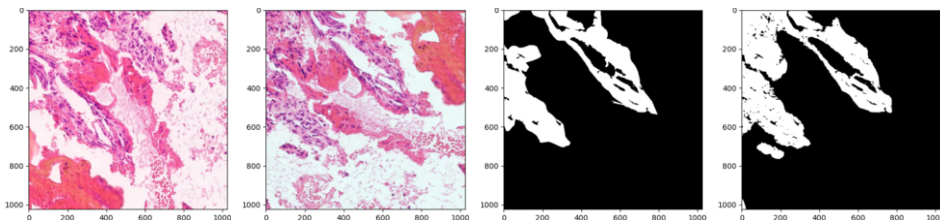


Figure 2. Results of the Machine Learning Model on Test Data. From Left to Right: Input Image, Augmented Image, Ground Truth, Segmentation Result (Tumor is white)

mented into the MONAI preprocessing composition by extending the `MapTransform` class. Figure 1 shows the preprocessing steps applied to an example image.

The dataset was sampled into three stratified subsets (80% – 10% – 10% split): training, validation, and test. The training set ($n = 16,186$) was used to develop the models, leveraging the labels to guide the learning process. The validation set ($n = 2,027$) served to prevent overfitting, while the test set ($n = 2,023$) was reserved for the final evaluation of the model’s performance on unseen data.

2.2. Model Development

The U-Net architecture was chosen for this study, as it currently is the gold-standard architecture for medical image segmentation [9,10].

The model was trained without pre-trained weights, allowing the model to learn feature representations directly from the training dataset, using the *Adam optimizer*. A learning rate scheduler was used to adjust the learning rate based on validation performance dynamically. A custom loss, combining *Dice Loss* and *Focal Loss*, was implemented to handle the class imbalance inherent in the dataset, ensuring that both small and large regions within the images were accurately segmented. The training was set to a maximum of 150 epochs with early stopping criteria to monitor the validation loss [11].

For the model implementation, *PyTorch Lightning* was used to streamline the training process, offering advanced features (e.g., model checkpointing, improved logging), thereby enhancing the overall efficiency of model development. MONAI directly supports *PyTorch Lightning* allowing its use without additional development effort.

The validation was evaluated using MONAI’s *F1/Dice* and *Intersection over Union* (IoU) implementation. Both metrics are standard in image segmentation tasks and measure the overlap between the predicted segmentation masks and the ground truth labels [12]. During validation, the metrics were calculated as a mean for each batch, and after an epoch, all batch-wise means were averaged in a total mean value (MONAI’s `mean_batch` reduction method). The `ignore_empty` parameter of the metrics was set to false. This was done to achieve the same results with the MONAI metrics as with scikit-learn’s `f1` and `Iou` metric [13], increasing the robustness and reproducibility of the evaluation.

2.3. Model Evaluation

After training, the model was evaluated on the test set, which consisted of unseen data. The evaluation metrics remain MONAI’s `Dice` and `IoU`. Instead of a batch-wise reduction, the mean value of all individual results in the test dataset was calculated. This

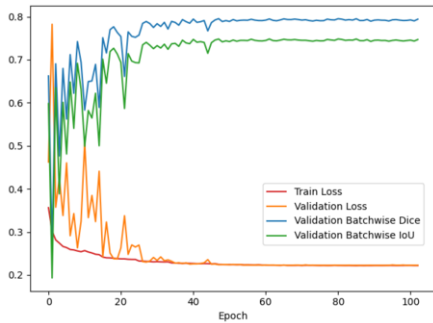


Figure 3.: Fitting Curve of Model Training

Metric	Model	Score
Dice	Our (MONAI)	0.790
	deepflash2	0.819
	nnU-Net v2	0.768
IoU	Our (MONAI)	0.743
	deepflash2	0.773
	nnU-Net v2	0.718

Table 1.: Performance on the Test Set of our Model (MONAI) in Comparison with other Established Frameworks like nnU-Net and deepflash2

provides more accurate results while at a higher resources cost and, therefore, was only done during testing. The results are further compared with metrics of current state-of-the-art AutoML tools *nnU-Net v2* [14] and *deepflash2* [15] on the same dataset.

3. Results

Figure 2 shows the segmentation of the final model compared to the label provided by a domain expert from neuropathology. Figure 3 and Tabel 1 depict the progress during training and the final results of the model. In total, the model trained for 102 epochs and was terminated by early stopping after reaching its best score in epoch 80. The final test scores were Dice: 79.00%, IoU: 74.33%, proving superior performance compared to nnU-Net v2 (Dice: 76.8%, IoU: 71.8%) while being inferior to deepflash2 (Dice: 81.9%, IoU: 77.3%) from an earlier study on a subset of the used dataset [16].

4. Discussion

This study demonstrates the efficacy of the MONAI framework for histopathological tumor segmentation. Applying the U-Net architecture within the MONAI framework proved effective, with the model achieving high scores in both Dice and IoU. The implementation was easy due to the available MONAI transformers and their great extendability. Additionally, the custom loss function, which combined Dice Loss and Focal Loss, contributed to the robustness of the models by effectively managing the class imbalance and ensuring accurate segmentation across large and small regions. Compared to the AutoML solutions, MONAI has proven itself to be equivalent in this setting. However, differences in training subsets, stain normalization, and hardware configurations, due to a different focus of the study [16], limit direct comparison with the AutoML frameworks. Future work should address these differences to ensure a robust and fair comparison.

5. Conclusion

This study demonstrates the potential of the MONAI framework for segmenting GBM tissue in histopathological images. The integration of a U-Net architecture, combined

with data augmentation, stain normalization, and a custom loss function, achieved competitive results in terms of Dice and IoU metrics. These findings highlight MONAI's ability to handle the variability and complexity of histopathological data and its competitiveness with state-of-the-art AutoML tools.

Future work will focus on evaluating MONAI against AutoML technologies on public datasets to enable a robust comparison and assess the generalizability of the models.

References

- [1] Chihati S, Gaceb D. A review of recent progress in deep learning-based methods for MRI brain tumor segmentation. In: 2020 11th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan: IEEE; 2020. p. 149-54.
- [2] Islam M, Wijethilake N, Ren H. Glioblastoma multiforme prognosis: MRI missing modality generation, segmentation and radiogenomic survival prediction. *Computerized Medical Imaging and Graphics*. 2021 Jul;91:101906.
- [3] Head and Neck Tumor Segmentation and Outcome Prediction: Second Challenge, HECKTOR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings. vol. 13209 of Lecture Notes in Computer Science. Cham: Springer International Publishing; 2022.
- [4] Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al.. MONAI: An open-source framework for deep learning in healthcare. arXiv; 2022. Version Number: 1.
- [5] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*. 2019 Dec;6(1):60.
- [6] Hoque MZ, Keskinarkaus A, Nyberg P, Seppänen T. Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion*. 2024 Feb;102:101997.
- [7] Roy S, Panda S, Jangid M. Modified Reinhard Algorithm for Color Normalization of Colorectal Cancer Histopathology Images. In: 2021 29th European Signal Processing Conference (EUSIPCO). Dublin, Ireland: IEEE; 2021. p. 1231-5.
- [8] Barbano CA, Pedersen A. EIDOSLAB/torchstain: v1.2.0-stable. Zenodo; 2022.
- [9] Jwaid WM, Al-Husseini ZSM, Sabry AH. Development of brain tumor segmentation of magnetic resonance imaging (MRI) using U-Net deep learning. *Eastern-European Journal of Enterprise Technologies*. 2021 Aug;4(9(112)):23-31.
- [10] Sharma R, Halarikar P, Choudhari K. Kidney and Tumor Segmentation using U-Net Deep Learning Model. *SSRN Electronic Journal*. 2020.
- [11] Lauriola I. On the Impact of Early Stopping in Multiple Kernel Learning. In: Proceedings of the Future Technologies Conference (FTC) 2020, Volume 1. vol. 1288. Cham: Springer International Publishing; 2021. p. 205-15. Series Title: Advances in Intelligent Systems and Computing.
- [12] Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. *IEEE Transactions on Medical Imaging*. 2020 Nov;39(11):3679-90.
- [13] Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al.. API design for machine learning software: experiences from the scikit-learn project. arXiv; 2013. ArXiv:1309.0238 [cs].
- [14] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. 2021 Feb;18(2):203-11. Publisher: Nature Publishing Group.
- [15] Griebel M, Segebarth D, Stein N, Schukraft N, Tovote P, Blum R, et al. Deep learning-enabled segmentation of ambiguous bioimages with deepflash2. *Nature Communications*. 2023 Mar;14(1):1679. Publisher: Nature Publishing Group.
- [16] Hieber D, Haisch N, Grambow G, Holl F, Liesche-Starnecker F, Pryss R, et al. Comparing nnU-Net and deepflash2 for Histopathological Tumor Segmentation. In: *Studies in Health Technology and Informatics*. IOS Press; 2024. .