Weakly supervised Medulloblastoma tumor classification using domain specific patch-level feature extraction

Lennart Maack[a], Debayan Bhattacharya[a], Finn Behrendt[a], Michael Bockmayr[b c e], and Alexander Schlaefer[a]

[a]Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, 21073 Hamburg, Germany
[b]Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, Hamburg 20246, Germany
[c]Research Institute Children's Cancer Center Hamburg, Martinistraße 52, Hamburg 20251
[e]Mildred Scheel Cancer Career Center HaTriCS4, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

ABSTRACT

Medulloblastoma (MB) is the most common embryonal tumour of the brain. In order to decide on an optimal therapy, laborious inspection of histopathological tissue slides by neuropathologists is necessary. Digital pathology with the support of deep learning methods can help to improve the clinical workflow. Due to the high resolution of histopathological images, previous work on MB classification involved manual selection of patches, making it a time consuming task. In order to leverage only slide labels for histopathology image classification, weakly supervised approaches first encode small patches into feature vectors using an ImageNet pretrained encoder based on convolutional neural networks. The representations of patches are further utilized to train a data-efficient attention-based learning method. Due to the domain shift between natural images and histopathology images, the encoder is not optimal for feature extraction for MB classification. In this study, we adapt weakly supervised learning for MB classification and examine different histopathological specific encoder architectures and weights for the MB classification task. The results show that ResNet encoders pretrained with histopathology images lead to better MB classification results compared to encoders pretrained on ImageNet. The best performing method uses a ResNet50 architecture, pretrained on histopathology images and achieves an area under the receiver operating curve (AUROC) value of 71.89%, improving the baseline model by 2%.

Keywords: Histopathology, Weakly Supervised Learning, Attention, Medulloblastoma, Whole Slide Image Analysis, Transfer Learning, Deep Learning

1. INTRODUCTION

Medulloblastoma accounts for 69.5% of all embryonal tumours according to the most recent data from the Central Brain Tumour Registry of the United States (CBTRUS).[1] Due to the rapid growth of the tumour, patients with MB often show symptoms that develop over a period of weeks to months, such as cerebellar dysfunction and increased intracranial pressure.[2] In order to decide on an optimal therapy and thus increase patient survival rates, early and accurate diagnosis is crucial, e.g. by detecting and distinguishing between variants of the tumor.[3] There are four histological variants of MB, namely classic variant (CMB), desmoplastic/nodular variant (DN), MB with extensive nodularity (MBEN), and large cell anaplastic MB (LCA). CMB accounts for 72% of all MB and is associated with a poorer prognosis compared to DN and MBEN.[4]
To examine the type and stage of a possible tumour, the pathologist visually assesses an extracted biopsy sample from the suspected region of the brain using a digitized view of the sample slide that has an extremely high resolution of up to 150.000×150.000. The assessment of the Whole Slide Images (WSIs) is not only a laborious and

Further author information: (Send correspondence to Lennart Maack)
Lennart Maack: E-mail: lennart.maack@tuhh.de
Alexander Schlaefer: E-mail: schlaefer@tuhh.de

Medical Imaging 2024: Digital and Computational Pathology, edited by John E. Tomaszewski, Aaron D. Ward, Proc. of SPIE Vol. 12933, 129331C · © 2024
SPIE · 1605-7422 · doi: 10.1117/12.3006455

time consuming task, but additionally suffers from inter- and intra-observer variability.[5] Despite the advances of using neural networks for medical image analysis in recent years, the size of the WSI makes it computationally infeasible to feed the WSI with original size directly into a neural network. A mere downscaling would lose essential information about cells and cell-to-cell interactions.

Previous work on MB tumor classification focused on utilizing transfer learning, as well as different input tile sizes to classify the two major subtypes, CMB and DN[6,7]. However, image tiles of representative cancer areas in the WSI had to be extracted manually by neuropathologists, again requiring time and effort and thus reducing support for the clinical workflow. To avoid time-consuming pixel-wise or patch-wise manual annotations, weakly supervised approaches based on multiple-instance learning (MIL) achieved high classification performance using only slide labels for training.[8] To solve the problem of low data-efficiency in MIL based approaches, Lu et al.[9] proposed an attention-based learning method that achieves superior performance with scarce data on publicly available data sets for renal cell carcinoma, lung adenocarcinoma or breast lymph node metastasis detection. In addition, interpretable heatmaps of areas with high diagnostic value can be visualised in the WSI through the attention-based learning approach. Figure 1 depicts the architecture of this method. To reduce the dimensionality of the extracted patches of the WSI, a convolutional neural network (CNN) based encoder with pretrained weights is utilized. As the utilized encoder in their work is pretrained on ImageNet, it is not optimal for extracting representative features from histopathology images due to the domain shift.[10] Sharma et al.[11] showed that truncation of final blocks of a pretrained encoder, i.e. omitting the last block of convolutional layers to use the feature maps from the penultimate block, can yield to better results as representations become more domain-specific in the later blocks. Furthermore, they demonstrated that a domain-specific pretrained encoder can lead to more representative features, too.

Our contribution in this work is twofold: First, we adapt the weakly supervised, attention-based learning approach for MB classification. In particular, for a disease like MB, for which there is scarce data due to its rarity, a systematic evaluation of a training method using only slide labels is of interest. In addition, and to further assess the added value to the clinical workflow, we examine the generated heatmaps, which are designed to visually highlight valuable areas for the pathologist. Second, we improve the adapted weakly supervised method for MB classification by substituting the baseline encoder with a histopathological specific approach. To do so, we systematically adapt and compare different encoder architectures and pretraining strategies.

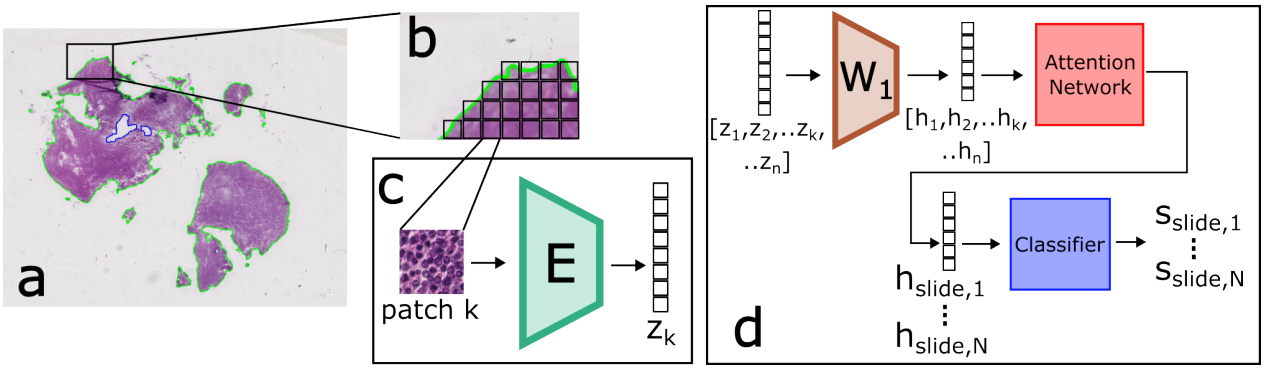

Figure 1. (a) The WSI is segmented into the region of interest indicated by the green borders using an adjusted threshold for background pixels. (b) Patches of size 224×224 are extracted from the region of interest. (c) An encoder E extracts the patch-level representation z_k for all n patches to reduce dimensionality. (d) A fully connected layer W_1 further compresses z_k into h_k. Afterwards, the attention network calculates the attention score for each patch k and aggregates all n attention scores to obtain the slide-level representation $h_{slide,i}$ for each of the N classes. Finally, a classifier calculates the slide-level prediction score $s_{slide,i}$ for each of the N classes.

2. METHODS AND MATERIALS

2.1 Data

The experiments are conducted with a data set of 161 WSI from individual patients collected at 12 clinical sites in Germany from 1989-2011. The haematoxylin and eosin (H&E) stained images are scanned at magnification

200× and labeled by trained neuropathologists as CMB or DN. Among the 161 samples, 103 cases belong to CMB and 58 cases belong DN. For comprehensive evaluation, we sample 18 CMB and 14 DN WSI for testing. The remaining 129 WSI are split into 10 training and validation sets for cross-validation.

2.2 Method

As mentioned in section 1, we build upon the weakly supervised, attention-based learning approach introduced by Lu et al.[9] The architecture is visualized and explained in depth in Figure 1. As our work focuses on improving the quality of the patch-level representations, different encoder architectures are adjusted for MB classification. Our utilized encoders are based on ResNet50 and ResNet18 architectures. To examine the influence of domain specific pretraining, we use publicly available ResNet50 and ResNet18 weights obtained by pretraining on histopathology data. The ResNet50 encoder weights are obtained by pretraining with a multi-task learning approach on a collection of publicly available data sets containing almost 900k images.[12] The data sets mostly contain H&E stained images of human breast cancer, but other organs, pathology and stains are represented in the training data as well. The ResNet18 encoder weights are obtained using the contrastive self-supervised learning method SimCLR[13] on a collection of 57 histopathology data sets.[14] The data sets contain different organs with different types of staining and resolution properties. In addition, the influence of the last block of the encoder is examined. For this, the last ResNet block of the encoder models is truncated for feature extraction. All models are trained for 180 epochs and hyperparameters are adapted from the original implementation.[9]

3. RESULTS

For evaluation, we report the area under the receiver operating curve (AUROC) over all ten folds with mean and standard deviation on the test data set. Table 1 displays the classification results for the different encoder architectures, the respective pretraining strategies and truncations of the last block. The results indicate that features extracted from a ResNet50 encoder pretrained with histopathology data lead to the best MB classification performance over all encoders. However, this is only the case if the last block of the encoder is not truncated. In contrast, truncating the last block of a ResNet50 encoder gives better results when the model is pretrained on ImageNet. The results for ResNet18 show the same trend. In all experiments, a stronger tendency of all models towards a CMB classification can be observed.

Table 1. Results for the experiments with different encoder architectures, pretraining based on natural images (ImageNet) or histopathology images (Histopath.) and whether the last block of the encoder is truncated. The baseline result is shown in *italics*, the best result is shown in **bold**.

Encoder Architecture	Pretraining	Truncated	AUC (↑)
ResNet50	ImageNet	Yes	*0.6989 ± 0.05*
		No	0.6619 ± 0.07
	Histopath.	Yes	0.6782 ±0.07
		No	**0.7134 ± 0.04**
ResNet18	ImageNet	Yes	0.6103 ±0.05
		No	0.5909 ±0.06
	Histopath.	Yes	0.6055 ±0.03
		No	0.6726 ±0.05

4. DISCUSSION AND CONCLUSION

This work addresses the task of MB tumor classification with weakly supervised learning using only slide level labels. Especially, the usage of histopathological specific encoders for better patch level representation extraction is evaluated.

Overall, our results indicate that non-truncated encoders pretrained on histopathology data sets lead to superior patch level representations for MB classification. The results for MB classification in Table 1 are also consistent with prior work,[11] i.e. the truncation of the last block of an encoder that is pretrained on ImageNet does not

lead to performance degradation. We hypothesise that the representations in the last layer of an ImageNet pretrained CNN become more domain-specific with respect to natural images. This effect can not be confirmed for encoders that are pretrained on histopathology images. They seem to continue to learn valuable medical domain-specific representations in the last layers. For all experiments, a stronger tendency of all models towards a CMB classification can be observed. This might be due to the imbalance in class distribution during training. In conclusion, this work introduces the task of MB classification when only slide labels are available for training. Additionally, our systematic experiments yield improved performance compared to the baseline method. The evaluation indicates an advantage of histopathological specific encoders for extracting patch representations for data scarce domains like MB. Future work could focus on extracting features not only on the cell level, but also consider cell-to-cell information for MB classification. In addition, problems that arise when training with an imbalanced data set can be addressed. As different encoders lead to different patch representation dimensions, an optimal architecture for dimensionality reduction can be tackled as well.

REFERENCES

[1] Ostrom, Q. T., Price, M., Neff, C., Cioffi, G., Waite, K. A., Kruchko, C., and Barnholtz-Sloan, J., "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2015–2019," *Neuro-Oncology* **24**, v1–v95 (10 2022).

[2] Millard, N. and Braganca, K. D., "Medulloblastoma," *Journal of Child Neurology* **31**(12), 1341–1353 (2016).

[3] Northcott, P. A., Robinson, G. W., Kratz, C. P., Mabbott, D. J., Pomeroy, S. L., Clifford, S. C., Rutkowski, S., Ellison, D. W., Malkin, D., Taylor, M. D., Gajjar, A., and Pfister, S. M., "Medulloblastoma," *Nature Reviews Disease Primers* **5** (Feb. 2019).

[4] Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., and Ellison, D. W., "The 2016 world health organization classification of tumors of the central nervous system: a summary," *Acta Neuropathologica* **131**, 803–820 (May 2016).

[5] Elmore, J. G., Barnhill, R. L., Elder, D. E., Longton, G. M., Pepe, M. S., Reisch, L. M., Carney, P. A., Titus, L. J., Nelson, H. D., Onega, T., Tosteson, A. N. A., Weinstock, M. A., Knezevich, S. R., and Piepkorn, M. W., "Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study," *BMJ* , j2813 (June 2017).

[6] Bengs, M., Bockmayr, M., Schüller, U., and Schlaefer, A., "Medulloblastoma tumor classification using deep transfer learning with multi-scale EfficientNets," in [*Medical Imaging 2021: Digital Pathology*], Tomaszewski, J. E. and Ward, A. D., eds., SPIE (Feb. 2021).

[7] Bengs, M., Pant, S., Bockmayr, M., Schüller, U., and Schlaefer, A., "Multi-scale input strategies for medulloblastoma tumor classification using deep transfer learning," *Current Directions in Biomedical Engineering* **7**, 63–66 (Aug. 2021).

[8] Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Silva, V. W. K., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., and Fuchs, T. J., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine* **25**, 1301–1309 (July 2019).

[9] Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., and Mahmood, F., "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering* **5**(6), 555–570 (2021).

[10] Ke, A., Ellsworth, W., Banerjee, O., Ng, A. Y., and Rajpurkar, P., "CheXtransfer," in [*Proceedings of the Conference on Health, Inference, and Learning*], ACM (Apr. 2021).

[11] Sharma, Y., Ehsany, L., Syed, S., and Brown, D., "Histotransfer: Understanding transfer learning for histopathology," in [*IEEE EMBS International Conference on Biomedical and Health Informatics*], IEEE (July 2021).

[12] Mormont, R., Geurts, P., and Maree, R., "Multi-task pre-training of deep neural networks for digital pathology," *IEEE Journal of Biomedical and Health Informatics* **25**, 412–421 (Feb. 2021).

[13] Chen, T., Kornblith, S., Norouzi, M., and Hinton, G., "A simple framework for contrastive learning of visual representations," in [*Proceedings of the 37th International Conference on Machine Learning*], III, H. D. and Singh, A., eds., *Proceedings of Machine Learning Research* **119**, 1597–1607, PMLR (13–18 Jul 2020).

[14] Ciga, O., Xu, T., and Martel, A. L., "Self supervised contrastive learning for digital histopathology," *Machine Learning with Applications* **7**, 100198 (2022).