Early Fusion of H&E and IHC Histology Images for Pediatric Brain Tumor Classification

Christoforos Spyretos[1,2] CHRISTOFOROS.SPYRETOS@LIU.SE
Iulian Emil Tampu[1,2] IULIAN.EMIL.TAMPU@LIU.SE
Nadieh Khalili[3] NADIEH.KHALILI@RADBOUDUMC.NL
Juan Manuel Pardo Ladino[1,4] JUAPA351@STUDENT.LIU.SE
Per Nyman[2,5] PER.NYMAN@REGIONOSTERGOTLAND.SE
Ida Blystad[2,6] IDA.BLYSTAD@REGIONOSTERGOTLAND.SE
Anders Eklund[1,2,4] ANDERS.EKLUND@LIU.SE
Neda Haj-Hosseini[1,2] NEDA.HAJ.HOSSEINI@LIU.SE

1 Department of Biomedical Engineering, Linköping University, Sweden
2 Center for Medical Image Science and Visualization, Linköping University, Sweden
3 Department of Pathology, Radboud University Medical Center, The Netherlands
4 Department of Computer and Information Science, Linköping University, Sweden
5 Department of Health, Medicine and Caring Sciences, Linköping University, Sweden
6 Department of Radiology and Department of Health, Medicine and Caring Sciences, Linköping University, Sweden

Abstract

This study explores the application of computational pathology to analyze pediatric brain tumors utilizing hematoxylin and eosin (H&E) and immunohistochemistry (IHC) histology images. Experiments were conducted on H&E images for predicting tumor diagnosis and fusing them with unregistered IHC images to investigate potential improvements. Patch features were extracted using UNI, a vision transformer (ViT) model trained on H&E data, and whole slide classification was achieved using the attention-based multiple instance learning CLAM framework. In the astrocytoma tumor classification, early fusion of the H&E and IHC significantly improved the differentiation between tumor grades (balanced accuracy: 0.82 ± 0.05 vs 0.84 ± 0.05). H&E only stain had a balanced accuracy of 0.79 ± 0.03 for the overall classes without any improvement obtained when fused with IHC. The findings highlight the potential of using multi-stain fusion to advance the diagnosis of pediatric brain tumors, however, further fusion methods should be investigated to explore the potentials.

Keywords: pediatric brain tumour, H&E, immunohistochemistry (IHC), Ki-67, GFAP, computational pathology, early fusion, UNI, CLAM, foundation model

1 Introduction

Central nervous system (CNS) tumors were the second leading cause of cancer incidences among children and adolescents aged 0-19 years old, with worldwide incidence and mortality rates of 1.2 and 0.6 (per 100,000 people) in 2022 (Ferlay J, 2024). Whole slide images (WSIs) are one of the primary tools for pathologists to diagnose brain tumors, providing a cost-effective illustration, sharing, and archiving for pathology information. The introduction of WSIs has led to an extensive volume of data generated, facilitating the implementation of deep learning in assisting pathologists with making faster and more consistent decisions.

The world health organization's (WHO) latest edition of guidelines for the CNS tumors classification published in 2021, emphasizes the integration of molecular information into diagnosing brain tumors (Louis et al., 2016). Despite the recent advancements in molecular diagnostics, the assessment of histology remains a crucial element in the evaluation of CNS tumors (Viaene,

2023). More precisely, since some molecular tests can take several days to weeks to provide results, the histologic diagnosis can guide initial treatment planning and patient counseling while awaiting the integrated diagnosis. In addition, histologic evaluation is frequently essential for determining the specific ancillary molecular tests needed for diagnosis. Pathologists often use various immunohistochemical (IHC) stains to identify specific molecular alterations, detect mutant proteins, and determine the molecular subgroups of brain tumors.

The increasing availability of biomedical data, such as medical imaging, electronic health records and genome sequences, has led to the development of multimodal artificial intelligence applications (Kline et al., 2022; Acosta et al., 2022). These applications mimic the multimodal nature of clinical expert decision-making and aim to enhance predictions and achieve more accurate diagnoses. enhance predictions and achieve more accurate diagnoses. Few studies in computational pathology have analyzed H&E-stained WSIs alongside IHC-stained WSIs, and those available are focused on breast cancer (Weitz et al., 2021; Liu et al., 2020). In this study, the classification of pediatric brain tumors is investigated using H&E stained images. In addition, the study examines whether the early fusion of unregistered H&E, Ki-67, and glial fibrillary acidic protein (GFAP) stained images improves diagnostic predictions compared to only using H&E slides. To the best of our knowledge, no published study has yet explored the potential and feasibility of fusing unregistered H&E and IHC images for predicting pediatric or adult brain tumor diagnoses. Potential uses for Ki-67 include diagnosis of medulloblastomas and astrocytomas and distinguishing between astrocytoma grades (Sengupta et al., 2012; Sharma et al., 2018). GFAP is a reliable marker for histological diagnosis between glial and non-glial tumors and the grading of astrocytomas (Varma et al., 2018; van Bodegraven et al., 2019).

2 Data

In this study, WSIs were utilized from the children's brain tumor network (CBTN) dataset, which consists of over 2,000 subjects and more than 8,000 slides (Lilly et al., 2023; Shapiro et al., 2023). However, the CBTN dataset is based on the 2016 WHO guidelines, which include brain tumor classifications no longer used in clinical practice. Therefore, subjects with outdated tumor classifications were excluded, and the brain tumor classifications listed in the 2021 WHO guidelines were used to conduct the analysis.

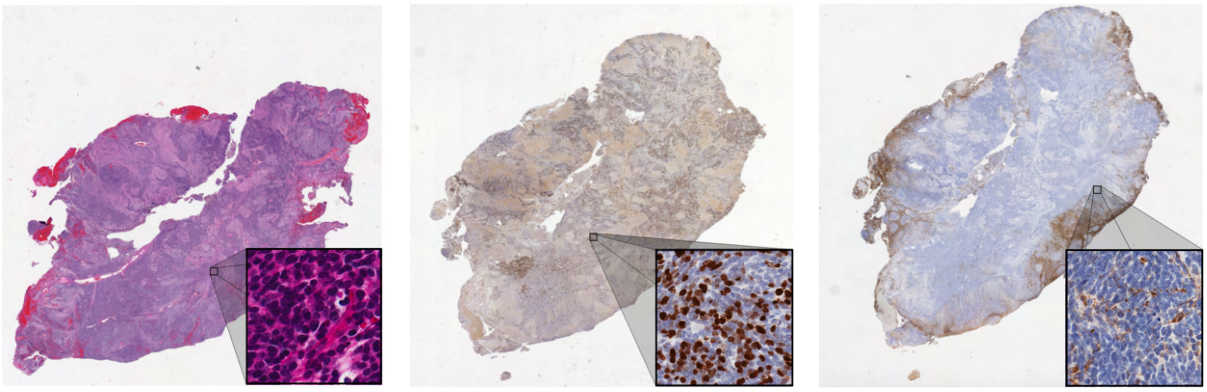
Most WSIs are H&E-stained, and the most represented IHC stains are Ki-67 and GFAP. Thus, only subjects with slides containing all three stains were chosen to conduct the experiments. It is important to note that the slides are unregistered, and subjects might contain several single-modality WSIs. In addition, a threshold of 10 subjects per class was set to ensure sufficient representation of each tumor family/type. Consequently, the study analyzes the cancer types of ependymoma (EP), medulloblastoma (MED), and ganglioglioma (GANG), and the tumor families of astrocytoma low-grade glioma (grades 1, 2) (ASTR-LGG) and astrocytoma high-grade glioma (grades 3, 4) (ASTR-HGG). ASTR-LGG, ASTR-HGG, GANG, and EP are glial tumors, while MED is a non-glial tumor. Additionally, images with artifacts such as pen marks and air bubbles were included in the experiments, as their effect was negligible in the model's performance (Pardo Ladino, 2024). Table 1 summarizes the number of subjects and WSIs included in the study, and figure 1 is an example from the dataset.

3 Methodology

End-to-end classification of WSIs using deep learning algorithms is challenging due to their immense size, typically in the range of gigapixels (Hosseini et al., 2024). This limitation is commonly circumvented by training patch-based networks using pixel-level annotations and then aggregating the patch-level results. Those approaches are known as supervised learning methods. However, obtaining a large amount of patches with fine-grained annotations is expensive, time-consuming, and requires the input of skilled and experienced pathologists. There-

Table 1: Tumor families/types and the number of subjects, H&E, KI-67, and GFAP slides used in the analysis. Only subjects with slides containing all three stain modalities were included in the experiments.

Tumor family/type	Number of subjects	Number of H&E slides	Number of Ki-67 slides	Number of GFAP slides
ASTR-LGG	173	342	195	196
ASTR-HGG	64	127	71	76
EP	47	106	54	57
MED	46	80	52	51
GANG	40	91	47	52
Totals	370	746	421	432



(a) H&E staining (b) Ki-67 staining (c) GFAP staining

Figure 1: Examples of H&E, Ki-67, and GFAP WSIs from the same subject diagnosed with medulloblastoma.

fore, weakly-supervised and self-supervised learning methods have been developed in computational pathology, addressing the aforementioned issues. Under the weakly-supervised learning paradigm, models are trained on partially or sparsely labeled data, such as one label for an entire WSI or per subject. With self-supervised learning, the algorithms learn feature representations through unlabeled data.

Multiple instance learning (MIL) is a weakly-supervised learning method, where each WSI is considered as a bag containing multiple patches, also called instances (Carbonneau et al., 2018). If a WSI (bag) is labeled class-positive, then at least one patch (instance) in that WSI is class-positive. Otherwise, if a WSI is class-negative, all patches in that WSI are negative. In this study, a MIL approach named clustering-constrained attention multiple instance learning (CLAM) was utilized to perform the classification tasks (Lu et al., 2021), aggregating patch-level features in slide-level representations for classification.

In the pre-processing phase of the images, the CLAM toolbox was employed to segment the tissue and extract patches of 256x256 pixels and their features. The ViT named UNI (UNI-ViT) was utilized as a feature extractor, pre-trained on a proprietary histology dataset. UNI is a recently introduced foundation model trained on more than 100 million patches from over 100,000 diagnostic H&E-stained WSIs across 20 major tissue types. Feature extractors pre-trained on in-domain histology datasets can capture a broad spectrum of patterns, such as different fixations, staining characteristics, scanning protocols, and tissue architecture across multiple centers. After extracting the features, those corresponding to the same stain modality for each subject were concatenated, representing subject-level features per stain modality.

Healthcare professionals utilize various information sources to make comprehensive decisions and a thorough understanding of the patient's condition. By fusing multiple modalities, the diverse tissue characteristics of each modality could be captured and obtain a more comprehensive and accurate analysis (Lipkova et al., 2022). An early fusion approach was utilized to explore the potential improvement in diagnostic prediction through the integration of IHC with H&E staining. Specifically, all possible fused combinations between the three stain modalities were investigated by concatenating features at the subject level. Figure 2 illustrates the workflow framework for the single modality and multi-stain approaches.

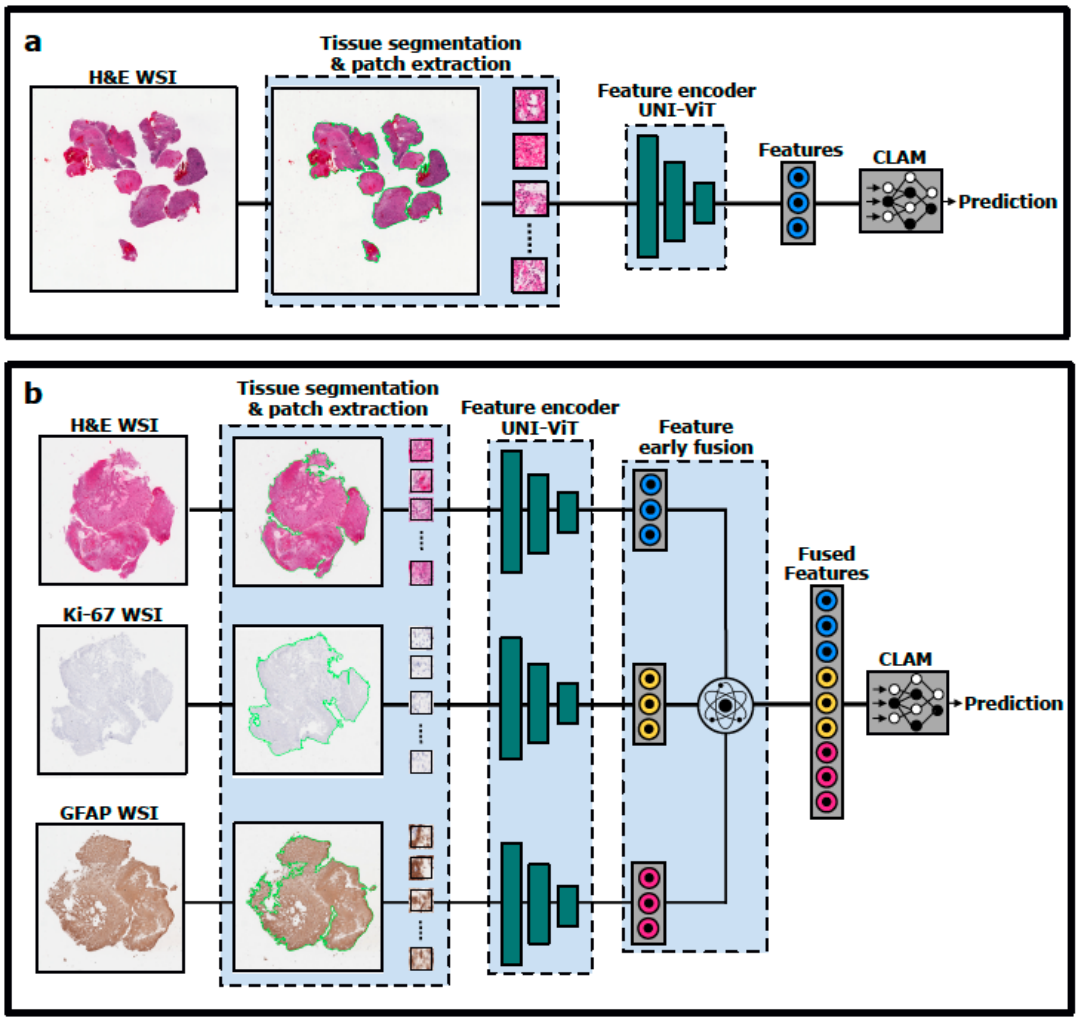


Figure 2: Overview of the **a)** single modality and **b)** multi-stain frameworks. **a)** Tissue segmentation and patch extraction from the whole slide images (WSIs) were accomplished using the CLAM toolbox. The extracted patches were encoded to feature representations using UNI, a ViT pre-trained on an in-domain proprietary H&E histology dataset. These features were then input into the CLAM model to perform the classification tasks. **b)** Tissue segmentation, patch and feature extraction were conducted in the same manner as in the single-modality procedure for each stain. An early fusion approach was employed between all possible combinations of the three stain modalities by concatenating features at the subject level. The fused features were then fed into the CLAM model to perform the classification tasks.

The classification tasks were accomplished using the small-sized single-branch CLAM model. In the training phase, the learning rate was 1e-4, and the maximum and minimum number of epochs was set at 20 and 10, respectively. Early stopping was utilized with patience set to 5 epochs. All other settings were set at the default values suggested by the authors. In addition

to the CLAM framework, class weight adjustment was used to mitigate the imbalanced data distribution, using the Scikit-learn library, with weights being inversely proportional to the number of samples in each class (Pedregosa et al., 2011).

The dataset was class-stratified and subject-wise split into 50% for training, 20% for validation, and 30% for testing to prevent data leakage between the sets. Since the class distribution of the dataset is imbalanced, the models were assessed using balanced accuracy, Matthew's correlation coefficient (MCC), area under the receiver operating characteristic curve (AUC-ROC), and weighted F1-score. Experiments were conducted on individual modalities and all combinations of the fused modalities using non-parametric bootstrapping with 50 replicates. Non-parametric bootstrapping involves repeatedly sampling from the dataset with replacement to generate multiple training, validation, and test sets (replicates). This approach allows for robust estimations of statistical measures without making assumptions about the data distribution. For the statistical comparison, 10000 permutations were conducted between the test sets to assess observed differences in the performance of the models, at a significance level of $\alpha = 0.05$. Statistical comparisons were conducted between the H&E single modality model and the fused modality models, as well as comparisons between the fused models. Bonferroni correction was utilized to adjust the significance level, and each hypothesis was performed at a statistical significance level $\alpha = 0.05/6 \approx 0.0083$ (Napierala, 2012).

Attention maps were used to visualize and interpret the importance of regions in the WSI, utilizing the CLAM toolbox. Furthermore, QuPath was employed to generate the positive and negative cell density maps of the Ki-67 and compared them to the attention maps (Bankhead et al., 2017; Pai et al., 2022). The Ki-67 immunostaining is interpreted as a labeling index (Ki-67 LI) and is defined as the percentage of the number of Ki-67 positive tumor nuclei (brown cells) divided by the number of all tumor nuclei, correlating with the histological tumor grade.

4 Results

Experiments were conducted using the three different staining modalities individually and in all possible combinations. Table 2 illustrates the classification performance between ASTR-LGG and ASTR-HGG and between the five tumor families/types. The tables show the mean values of the metrics along with their standard deviation across 50 repetitions. The models were assessed using unseen test sets.

In binary classification, the fusion of all three stains achieves the highest performance, with significant differences compared to using only the H&E WSIs in most metrics, although it is not always significantly better than fusing H&E WSIs with either Ki-67 or GFAP WSIs. The performance of each class has steadily improved by gradually combining different stain modalities across 50 repetitions. For ASTR-LGG, the balanced accuracy has increased from 0.91 ± 0.04 to 0.92 ± 0.05, and the f1-score has improved from 0.73 ± 0.07 to 0.77 ± 0.06 when fusing three stains. Similarly, for ASTR-HGG, the balanced accuracy has enhanced from 0.72 ± 0.1 to 0.75 ± 0.1, and the f1-score has increased from 0.91 ± 0.02 to 0.92 ± 0.02 when fusing the three stains. However, in the five-class classification, fusing Ki-67, GFAP, or both with H&E does not significantly improve performance compared to using H&E WSIs alone, and there is no improvement in each class's performance. Specifically, EP and MED are almost perfectly classified, and ASTR-LGG is correctly classified in most cases. However, ASTR-HGG is incorrectly classified in less than half of the cases, often classified as ASTR-HGG, and GANG is frequently misclassified as ASTR-LGG.

Figure 3 illustrates the attention maps for individual stain modalities and fused modalities. Red regions indicate higher attention from the model, while blue regions indicate lower attention in those tissue regions. Figures 5 and 4 illustrate the attention and cell density maps of an ASTR-LGG Ki-67 WSI and an ASTR-HGG Ki-67 WSI, respectively. The positive cell density map shows Ki-67 positive stained nuclei (brown cells); the greater the number of the positive

Table 2: Classification performances between ASTR-LGG and ASTR-HGG and five-class classification of tumor families/types using the three different staining modalities, individually and in all possible combinations, across 50 repetitions. The models with the best overall performance are highlighted in bold.

ASTR-LGG vs ASTR-HGG				
Stain Modalities	Balanced Accuracy	MCC	AUC-ROC	Weighted F1-Score
H&E	0.82 ± 0.05^{a}	0.65 ± 0.08^{a}	0.90 ± 0.04^{a}	0.86 ± 0.03^{a}
H&E + Ki-67	0.83 ± 0.05^{b}	0.68 ± 0.09^{b}	0.90 ± 0.03^{b}	0.87 ± 0.03^{b}
H&E + GFAP	0.82 ± 0.05^{c}	0.66 ± 0.09^{c}	0.91 ± 0.04^{c}	0.87 ± 0.04^{c}
H&E + Ki-67 + GFAP	$\mathbf{0.84 \pm 0.05}^{a}$	$\mathbf{0.69 \pm 0.08}^{a}$	$\mathbf{0.90 \pm 0.04}^{d}$	$\mathbf{0.88 \pm 0.03}^{a}$
Ki-67	0.81 ± 0.04	0.62 ± 0.07	0.88 ± 0.05	0.85 ± 0.03
GFAP	0.77 ± 0.05	0.55 ± 0.08	0.86 ± 0.04	0.82 ± 0.03
Five-class classification between tumour families/types				
H&E	$\mathbf{0.79 \pm 0.03}^{a}$	$\mathbf{0.72 \pm 0.04}^{a}$	$\mathbf{0.94 \pm 0.01}^{a}$	$\mathbf{0.80 \pm 0.03}^{a}$
H&E + Ki-67	0.78 ± 0.03^{b}	0.73 ± 0.04^{b}	$0.94 \pm 0.01^{a,b}$	0.80 ± 0.03^{b}
H&E + GFAP	0.78 ± 0.04^{c}	0.72 ± 0.05^{c}	0.94 ± 0.01^{b}	0.80 ± 0.03^{c}
H&E + Ki-67 + GFAP	0.78 ± 0.03^{d}	0.72 ± 0.05^{d}	0.94 ± 0.01^{c}	0.80 ± 0.03^{d}
Ki-67	0.66 ± 0.04	0.58 ± 0.05	0.88 ± 0.02	0.70 ± 0.04
GFAP	0.69 ± 0.05	0.61 ± 0.05	0.90 ± 0.02	0.72 ± 0.04

^{a,b,c,d} Within a column, two-sided p-value < 0.05 permutation test, significance level adjusted after Bonferroni correction to $\alpha = 0.05/6 \approx 0.0083$. Model performances with a common superscript differ significantly.

stained nuclei, the higher the tumor grade. Conversely, the negative cell density map shows Ki-67 negatively stained nuclei (blue cells); the more negatively stained nuclei present, the lower the tumor grade. The Ki-67 labeling index of the ASTR-HGG is approximately 3.67%, suggesting a high proliferation rate associated with high-grade pediatric brain tumors. By comparing the attention map to the positive cell density map, it could be inferred that the model's attention is localized in the WSI regions that have a high density of positive Ki-67 (brown) cells. In contrast, the Ki-67 labeling index of the ASTR-LGG is approximately 0.4%, indicating a low proliferation rate associated with low-grade pediatric brain tumors. When comparing the attention map to the negative cell density map, it could be interpreted that the model's attention is mainly localized in the upper left part of the tissue, which represents the negative cell (blue) region according to the cell density map. However, the model's attention is also concentrated on tissue regions with a low density of negative cells.

5 Discussion

The experiments conducted using unregistered WSIs with three different staining modalities (H&E, Ki-67, and GFAP), both individually and in fusion, produced satisfactory results in binary and five-class classifications. When distinguishing between ASTR-LGG and ASTR-HGG, fusing all three stains significantly outperformed H&E slides alone. This highlights the potential diagnostic value of Ki-67 and GFAP in differentiating between astrocytoma grades confirmed by the observation made using the Ki-67 labeling index. However, in the five-class classification task, although the performance of the model on the H&E WSIs alone was high,

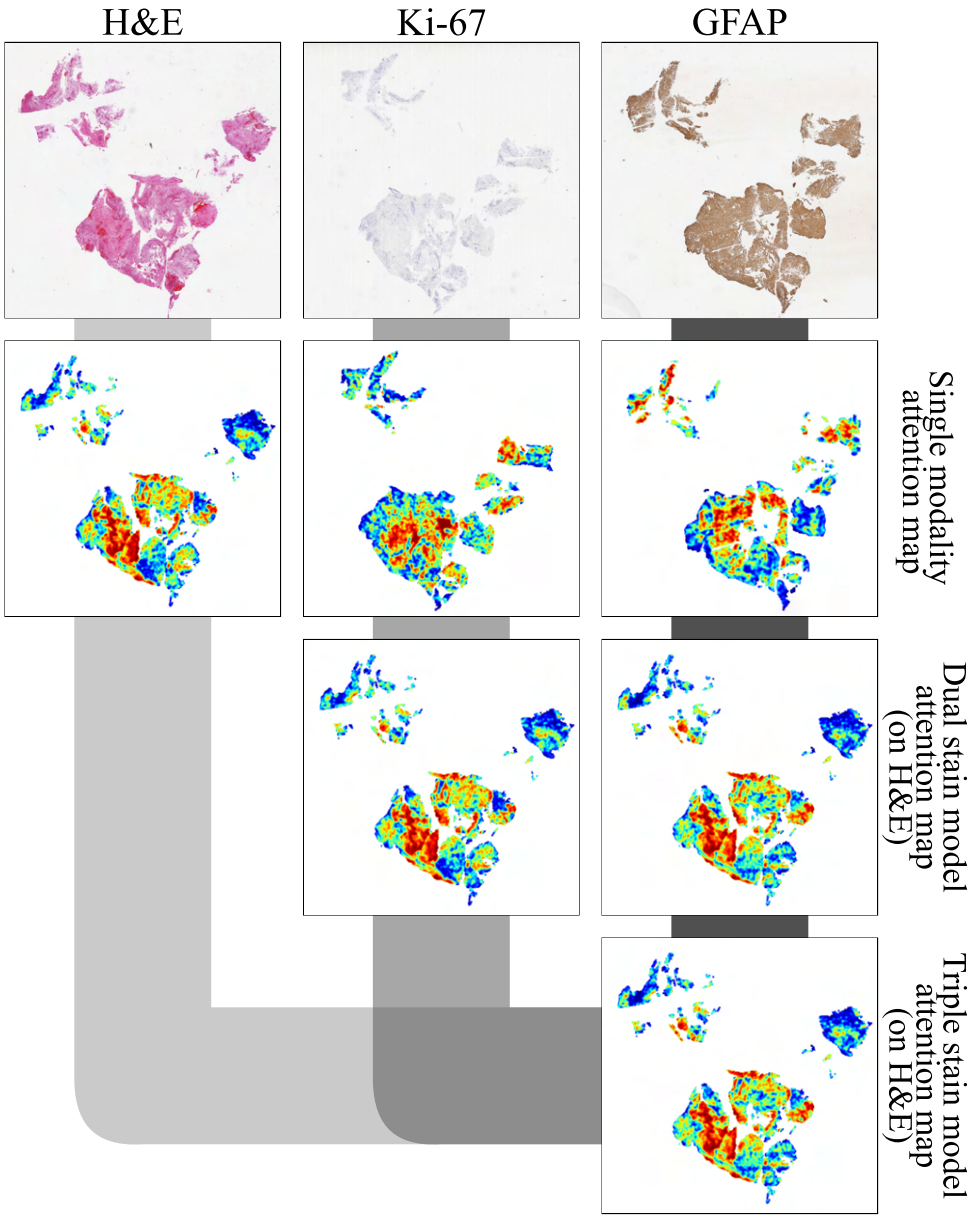
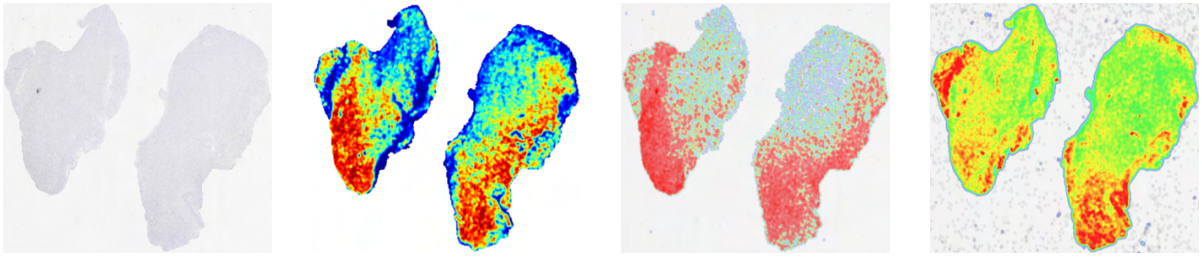
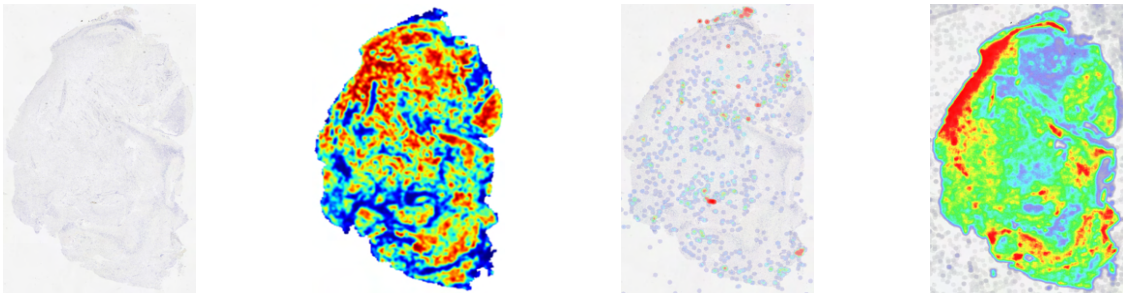


Figure 3: Attention maps of ASTR-LGG WSIs of the ASTR-LGG vs ASTR-HGG classification task between individual and fused modalities.



(a) Ki-67 WSI (b) Attention map (c) Positive cell density map (d) Negative cell density map

Figure 4: A Ki-67 WSI of an ASTR-HGG along with its corresponding attention and cell density maps.



(a) Ki-67 WSI (b) Attention map (c) Positive cell density map (d) Negative cell density map

Figure 5: A Ki-67 WSI of an ASTR-LGG along with its corresponding attention and cell density maps.

fusing Ki-67 WSIs, GFAP WSIs, or both with H&E WSIs did not improve the results, implying that the IHC information is not useful for multi-class tumor-type classification due to the multiple tumor classes or that a different fusion strategy is needed.

To further explore the potential of multi-stain fusion in computational pathology, alternative fusion strategies beyond early fusion should be investigated. Multimodal strategies also include late and intermediate fusions, with intermediate fusion further subdivided into single-level, gradual, and guided. In intermediate fusion, the prediction loss is backpropagated to the feature extraction layer of each modality to iteratively learn improved feature representations. Using the late fusion approach, individual classifiers are trained separately on each staining modality, and their predictions are combined at the decision level.

The relationship between the attention and the cell density maps provides useful understandings of how the model's attention is localized to the regions of the tissue based on cell proliferation activity. In low-grade tumors, the model emphasizes tissue areas with negatively stained nuclei, whereas in high-grade tumors, the model's attention is mainly concentrated on regions with high Ki-67 positively stained nuclei. This attention pattern underlines the model's potential for classifying tumor grades based on cellular proliferation markers. However, a more thorough analysis would require further interpretation of the maps and eventual statistical analysis.

6 Conclusion

The experiments showed promising results in classifying pediatric brain tumor families/types using H&E slides. Early fusion of unregistered Ki-67 and GFAP with H&E slides improved the distinction between ASTR-LGG and ASTR-HGG with statistical significance, suggesting the diagnostic potential of Ki-67 and GFAP stains. However, over multiple tumor classes, fusion of the H&E and IHC did not improve the results from what is obtained by the H&E images only. Alternative fusion strategies beyond early fusion, such as late and intermediate fusions, should be explored in the next step to maximize the potential of multi-stain fusion for this application.

Acknowledgements and Disclosure of Funding

The research was made possible in part due to The Children's Brain Tumor Tissue Consortium (CBTTC)/ The children's brain tumor network (CBTN). The study was financed by Swedish Childhood Cancer Foundation (MT2021-0011, MT2022-0013), Joanna Cocozza's Foundation (2023-2024), Linköping University's Cancer Strength Area (2022) and Vinnova project 2017-02447 via Medtech4Health and Analytic Imaging Diagnostics Arena (1908).

References

Julián N Acosta, Guido J Falcone, Pranav Rajpurkar, and Eric J Topol. Multimodal biomedical ai. *Nature Medicine*, 28(9):1773–1784, 2022.

Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.

Marc-André Carbonneau, Veronika Cheplygina, Eric Granger, and Ghyslain Gagnon. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77:329–353, 2018.

Lam F Laversanne M Colombet M Mery L Piñeros M Znaor A Soerjomataram I Bray F Ferlay J, Ervik M. Global cancer observatory: Cancer today. Lyon, France: International agency for research on cancer, 2024. URL `https://gco.iarc.who.int/today`. Accessed: 2024-06-19.

Mahdi S Hosseini, Babak Ehteshami Bejnordi, Vincent Quoc-Huy Trinh, Lyndon Chan, Danial Hasan, Xingwen Li, Stephen Yang, Taehyo Kim, Haochen Zhang, Theodore Wu, et al. Computational pathology: a survey review and the way forward. *Journal of Pathology Informatics*, page 100357, 2024.

Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1):171, 2022.

Jena V Lilly, Jo Lynne Rokita, Jennifer L Mason, Tatiana Patton, Stephanie Stefankiewiz, David Higgins, Gerri Trooskin, Carina A Larouci, Kamnaa Arya, Elizabeth Appert, et al. The children's brain tumor network (cbtn)-accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia*, 35:100846, 2023.

Jana Lipkova, Richard J Chen, Bowen Chen, Ming Y Lu, Matteo Barbieri, Daniel Shao, Anurag J Vaidya, Chengkuan Chen, Luoting Zhuang, Drew FK Williamson, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer cell*, 40(10):1095–1110, 2022.

Yiqing Liu, Xi Li, Aiping Zheng, Xihan Zhu, Shuting Liu, Mengying Hu, Qianjiang Luo, Huina Liao, Mubiao Liu, Yonghong He, et al. Predict ki-67 positive cells in h&e-stained images using deep learning independently from ihc-stained images. *Frontiers in Molecular Biosciences*, 7: 183, 2020.

David N Louis, Arie Perry, Guido Reifenberger, Andreas Von Deimling, Dominique Figarella-Branger, Webster K Cavenee, Hiroko Ohgaki, Otmar D Wiestler, Paul Kleihues, and David W Ellison. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta neuropathologica*, 131:803–820, 2016.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

Matthew A Napierala. What is the bonferroni correction? *Aaos Now*, pages 40–41, 2012.

Rima Pai, Susan Karki, Rakhee Agarwal, Steven Sieber, and Samuel Barasch. Optimal settings and clinical validation for automated ki67 calculation in neuroendocrine tumors with open source informatics (qupath). *Journal of Pathology Informatics*, 13, 2022.

Juan Manuel Pardo Ladino. Multiple instance attention-based learning for slide-level brain tumor histopathology classification. Master's thesis, Linköping University, The Division of Statistics and Machine Learning, 2024.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Subhalakshmi Sengupta, Uttara Chatterjee, Uma Banerjee, Samarendranath Ghosh, Sandip Chatterjee, and Ashit K Ghosh. A study of histopathological spectrum and expression of ki-67, tp53 in primary brain tumors of pediatric age group. *Indian Journal of Medical and Paediatric Oncology*, 33(01):25–31, 2012.

Joshua A Shapiro, Krutika S Gaonkar, Stephanie J Spielman, Candace L Savonen, Chante J Bethell, Run Jin, Komal S Rathi, Yuankun Zhu, Laura E Egolf, Bailey K Farrow, et al. Openpbta: The open pediatric brain tumor atlas. *Cell genomics*, 3(7), 2023.

Vikas Sharma, Yawar Shoaib, Laxmi Gupta, and Amit Dagar. P53 and ki-67 expression in primary pediatric brain tumors: Does it correlate with presentation, histological grade, and outcome? *Asian journal of neurosurgery*, 13(04):1026–1032, 2018.

Emma J van Bodegraven, Jessy V van Asperen, Pierre AJ Robe, and Elly M Hol. Importance of gfap isoform-specific analyses in astrocytoma. *Glia*, 67(8):1417–1433, 2019.

Amit V Varma, Garima Gupta, Jagdish Gupta, and Sapan Gupta. Gfap expression in neuroglial tumours–immunohistochemical confirmation for diagnosis and grading. *Journal of Evolution of Medical and Dental Sciences*, 7(46):5834–5839, 2018.

Angela N Viaene. Pediatric brain tumors: A neuropathologist's approach to the integrated diagnosis. *Frontiers in Pediatrics*, 11:1143363, 2023.

Philippe Weitz, Balazs Acs, Johan Hartman, and Mattias Rantalainen. Prediction of ki67 scores from h&e stained breast cancer sections using convolutional neural networks. In *Medical Imaging with Deep Learning*, 2021.