Enhancing Glioma Detection through Multimodal Classifiers: Integrating MRI and WSI

Tomé Albuquerque [1,2]
tome.m.albuquerque@inesctec.pt

Beatriz Coutinho [2], João Rodrigo [2], Miguel Almeida [2]
{up201906333,up201705110,up201907088}@edu.fe.up.pt

Benedikt Wiestler[3,4]
b.wiestler@tum.de

Claire Delbridge[5,6]
c.delbridge@tum.de

Maria João M. Vasconcelos[7]
maria.vasconcelos@fraunhofer.pt

Peter Schüffler[6]
peter.schueffler@tum.de

Jaime S. Cardoso[1,2]
jaime.cardoso@inesctec.pt

[1] INESC TEC
Porto, Portugal

[2] Faculty of Engineering of the University of Porto
Porto, Portugal

[3] Department of Neuroradiology, MRI, TUM
Munich, Germany

[4] TranslaTUM, TU Munich
Munich, Germany

[5] Department of Neuropathology, MRI, TUM
Munich, Germany

[6] Institute of General and Surgical Pathology, TUM
Munich, Germany

[7] Fraunhofer Portugal AICOS
Porto, Portugal

Abstract

Adult-type diffuse gliomas represent the predominant malignant neoplasms within the central nervous system. The advent of targeted therapeutic options has amplified the allure of molecular biomarkers, directly impacting on selection of appropriate interventions. Nevertheless, the manual assessment process within pathological laboratories is burdened by time intensiveness and error susceptibility. In order to surmount this constraint, multimodal fusion models have been explored. These models aim to identify the two pivotal molecular biomarkers (IDH1 mutation and 1p/19q codeletion) within gliomas by harnessing the synergy of MRI and digital pathology examinations.

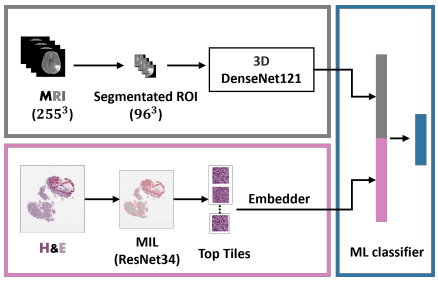
1 Introduction

Within the realm of glioma molecular markers, the mutation status of IDH1 and the codeletion of 1p/19q serve as crucial elements for precise diagnosis. The automation of biomarker detection holds the potential to spare patients from unnecessary examinations and alleviate the burden on both patients and medical institutions. Furthermore, leveraging predictions generated from the fusion of diverse medical data sources can significantly aid in the decision-making process for treatment strategies. This study introduces a novel multimodal learning framework designed to classify biomarkers. Our approach integrates brain Magnetic Resonance Imaging (MRI) and whole slide images (WSI) from hematoxylin and eosin staining (H&E) slides. Since the update of the WHO protocol, intensive research has been conducted on biomarker classification and glioma grading. In recent years, several researchers tried to detect IDH1 and 1p/19q biomarkers from MRI or WSI [1, 3]. Nevertheless, most of the works only make a binary classification between IDH1-mutant and IDH1-wildtype or 1p/19q codeleted and non-codeleted.

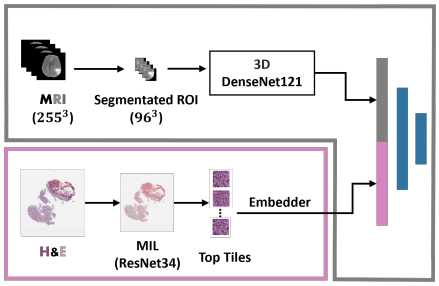
This research is structured into two distinct phases. First, we utilize two separate Convolutional Neural Network (CNN) models, each pre-trained on MRI and WSI scans. These CNN models function as feature encoders, supplying inputs to seven distinct machine learning (ML) classification algorithms. We then compare and analyze the results of these methods. In the second phase, we develop an end-to-end model based on a CNN architecture. This model processes MRI scans and extracts relevant patterns and features. In the final linear layers of this CNN model, we incorporate features extracted from WSI data to enhance the model training process. By integrating multimodal data and exploring the best aggregation strategies, including an end-to-end CNN approach, our study aims to advance the classification of critical glioma biomarkers, thereby enhancing diagnostic precision and decision-making in the field of glioma research.

2 Proposed Methodology

As mentioned in the previous section, this study includes two distinct phases: using conventional ML classifiers with pre-trained CNN encoders for both modalities (Figure 1 (a)) and multimodal deep learning (DL) models using an end-to-end training strategy (Figure 1 (b)).



(a) Phase I - Multimodal aggregation ML models.



(b) Phase II - End-to-end CNN model

Figure 1: Representation of different proposed methods: (a) Phase I - Aggregation ML models; (b) Phase II - End-to-end CNN model.

All models underwent training using a comprehensive approach that included both MRI and WSI modalities. For the MRI branch, segmented ROI blocks of size $96 \times 96 \times 96$ pixels (px) were extracted from the original MRI exams, which were of size $255 \times 255 \times 255$ px. These ROI blocks were used to train both the encoder and the end-to-end models. Two different MRI exam types, T1ce and FLAIR, were tested both individually and in combination to determine the optimal method. DenseNet121 was employed as the backbone for biomarker classification. In the case of the WSI encoder, a weakly supervised DL model based on multiple instance learning (MIL) was utilized. This model relied solely on the reported diagnosis as the label per slide. The embedder/encoder was trained for binary classification (presence vs. absence of biomarkers) using ResNet34. The top 10 tiles per slide were selected during training based on their output scores. The first phase of the study involved employing seven distinct machine learning algorithms: CatBoost, Xgboost, K-Nearest-Neighbor (KNN), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). Feature vectors generated by the pre-trained models were concatenated into a unified feature

vector, which was then used as input for subsequent aggregation models to arrive at a final consensus prediction. During the second phase, the WSI encoder remained unchanged or "fixed," maintaining its previously learned features. In contrast, the MRI encoder was fine-tuned in an end-to-end manner. This fine-tuning involved training the MRI encoder using concatenated features derived from the WSI data, allowing for the integration of information from both modalities.

3 Experimental Details

3.1 Dataset

In this work, a public dataset from The Cancer Genome Atlas (TCGA). [1] was used. It includes 3D MRI images ($255 \times 255 \times 255$ px) from four modalities: T1, T1ce, T2, and FLAIR with segmentation blocks ($96 \times 96 \times 96$ px), digital pathology slides with hematoxylin and eosin staining (H&E), and clinical data (age and gender) from 187 patients. Figure 2 shows three examples of both exams for three different patients. The dataset consists of three classes (without any biomarker; with IDH1 mutation; and with IDH1 mutation and 1p19q codeletion).

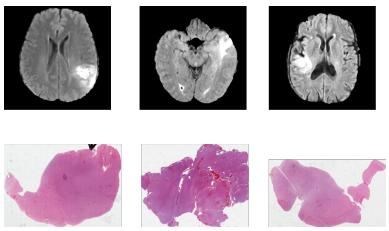


Figure 2: Examples of MRI and WSI exams for three different patients.

3.2 Data pre-processing

The dataset was partitioned into five distinct folds to preserve the class ratios using stratified cross-validation. The results are the average and standard deviation of these five folds. For the Pathology deep learning embedded, it was necessary to extract tiles with 512×512 px and 2048×2048 px dimensions from all WSI. Only tiles with more than 60% of tissue were used. Normalization was also performed to scale the pixel values to 0-1. Each pathology image is subjected to a series of random transformations during training, including HEDJitter, random elastic transformations, random color affine, random Gaussian blur, and random image rotations. MR images were preprocessed using the publicly available BraTS Toolkit [2]. For tumor segmentation, the SIMPLE fusion approach implemented in BraTS Toolkit was applied. MR images were subjected to random image rotations and intensity changes during training.

3.3 Training

A grid search was also performed for the seven ML models of phase I to find the best combination of hyperparameters. For the WSI deep learning encoder, a grid search was first performed to find the best combination of hyperparameters. The best parameters found for the fine-tuned model were an image size of 552×552px, a batch size of 384, and 50 training epochs. ADAM was used in both modalities as the optimizer and started with a learning rate of 10^{-4} for all models. Cross-entropy (CE) was used as loss function.

4 Results and Discussion

The methods proposed in this study underwent training using a dataset consisting of 149 digital pathology WSI and MRI scans encompassing FLAIR and T1ce modalities. Subsequently, validation was conducted on 38 WSI and MRI examinations, with the T1ce modality yielding the most promising results. The following tables are based on the T1ce MRI exams. Table 1 illustrates the performance of phase I, employing seven distinct ML algorithms. The best-performing results are highlighted in bold. The MLP emerged as the top-performing algorithm, achieving an accuracy of 78.09%, an MAE of 0.28, and an F1 score of 77.46. In contrast, CatBoost demonstrated remarkable competitiveness, boasting an AUC of 0.90, surpassing the MLP by 0.30 in this regard.

Table 1: Evaluation metrics for biomarkers detection using seven different models for decision aggregation.

Model	ACC (%)	MAE	F1(%)	AUC
CatBoost	77.01 ± 5.67	0.29 ± 0.14	73.77 ± 2.72	$\mathbf{0.90 \pm 0.05}$
Xgboost	75.40 ± 6.61	0.33 ± 0.10	72.83 ± 8.56	0.87 ± 0.06
KNN	70.64 ± 8.50	0.34 ± 0.11	66.16 ± 10.03	0.85 ± 0.03
LR	75.39 ± 7.49	0.31 ± 0.10	74.13 ± 7.44	0.86 ± 0.04
RF	77.01 ± 4.99	0.29 ± 0.08	73.43 ± 6.37	0.88 ± 0.05
SVM	73.79 ± 6.77	0.31 ± 0.09	72.47 ± 7.33	0.86 ± 0.03
MLP	$\mathbf{78.09 \pm 5.13}$	$\mathbf{0.28 \pm 0.08}$	$\mathbf{77.46 \pm 4.80}$	0.87 ± 0.03

Taking a deeper dive and evaluating the performance of the end-to-end MRI model, which integrates WSI features, we can observe the results presented in Table 2. These results have exceeded the performance metrics of most of the previous ML models that relied on pre-trained encoders. This suggests that integrating multimodal features into the training process of the models may have a beneficial effect on the detection of glioma biomarkers.

Table 2: Evaluation metrics for biomarkers detection using an end-to-end CNN model.

Model	ACC (%)	MAE	F1(%)	AUC
End-to-end	81.28 ± 3.94	0.24 ± 0.06	80.84 ± 3.52	0.87 ± 0.04

5 Conclusions and Future Work

In summary, when comparing the MLP to the other six machine learning methods, MLP demonstrated superior efficiency in the classification of glioma biomarkers. The end-to-end model yielded promising results across various metrics, underscoring its potential for identifying biologically significant biomarker features and effectively categorizing distinct glioma subtypes by integrating multimodal features in deep learning systems. As we look ahead to future research, there is an opportunity to extend and test the capabilities of the end-to-end model to include WSI exams and validate these methods using external datasets.

Acknowledgments

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project HfPT, with reference 41. Tomé Albuquerque was supported by Ph.D. grant 2021.05102.BD provided by FCT.

References

[1] Pranathi Chunduru, Joanna J. Phillips, and Annette M. Molinaro. Prognostic risk stratification of gliomas using deep learning in digital pathology images. *Neuro-Oncology Advances*, 4, 2022.

[2] Florian Kofler, Christoph Berger, Diana Waldmannstetter, Jana Lipkova, Ivan Ezhov, Giles Tetteh, Jan Kirschke, Claus Zimmer, Benedikt Wiestler, and Bjoern H. Menze. Brats toolkit: Translating brats brain tumor segmentation algorithms into clinical and scientific practice. *Frontiers in Neuroscience*, 14, 2020. ISSN 1662-453X. doi: 10.3389/fnins.2020.00125. URL https://www.frontiersin.org/articles/10.3389/fnins.2020.00125.

[3] Yingping Li, Samy Ammari, Littisha Lawrance, Arnaud Quillent, Tarek Assi, Nathalie Lassau, and Émilie Chouzenoux. Radiomics-based method for predicting the glioma subtype as defined by tumor grade, idh mutation, and 1p/19q codeletion. *Cancers*, 14, 2022.

[1] https://www.cancerimagingarchive.net/