Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33495

Weakly Supervised ROI-Guided Multiple Instance Learning for 4-Class Brain Tumor Classification on Single-Modal MRI

Haoyi Yun

Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
s230026199@mail.bnbu.cn

Abstract. Accurate stratification of brain tumors by magnetic resonance imaging (MRI) plays a vital role in clinical treatment decision-making and yet manual radiological diagnosis is time-consuming, laborious, and prone to inter-rater inconsistency. Most prevailing deep learning frameworks follow an end-to-end fully supervised paradigm, which depends heavily on expensive pixel-wise lesion annotations with prohibitive labeling costs. Meanwhile, existing weakly supervised multiple instance learning (MIL) approaches are constrained by low-quality instance generation with excessive background noise and information loss during feature aggregation, which severely limits their clinical translation. To tackle these bottlenecks, this work presents a weakly supervised region-of-interest-guided multiple instance learning network for four-class brain tumor classification using single-modal T1-weighted contrast-enhanced MRI scans. The proposed method incorporates brain anatomical priors to generate high-quality candidate tumor instances, designs a gated attention-based aggregation module to adaptively weight instance-level features, and adopts a medical-domain pre-trained encoder with a two-step fine-tuning pipeline to enhance model generalizability. Experiments on a dataset containing 7200 MRI scans demonstrate that the method achieves a test accuracy of 94.81% and a macro F1 score of 94.70%, surpassing a series of mainstream baseline models. Attention visualization results confirm that the model can precisely locate tumor lesions without fine-grained annotations, making it a promising solution for low-cost, high-precision clinical computer-aided diagnosis.

Keywords: computer vision, deep learning, medical image analysis

1. Introduction

Brain tumors are among the most life-threatening malignancies worldwide with high morbidity and mortality [1]. According to the 2021 World Health Organization (WHO) Classification of Central Nervous System Tumors, glioma, meningioma and pituitary adenoma are the most prevalent intracranial tumor subtypes with distinct treatment protocols and prognosis, and accurate preoperative classification is critical for personalized clinical decision-making and patient survival improvement [2]. T1-weighted contrast-enhanced Magnetic Resonance Imaging (MRI) is the first-line imaging modality for brain tumor diagnosis [3], but manual MRI interpretation suffers from significant inter-observer variability and low efficiency, especially in resource-limited settings [4].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33495

Therefore, developing an automated and accurate brain tumor classification method based on MRI has become an urgent clinical need.

In recent years, deep learning methods including Convolutional Neural Networks (CNNs) and vision transformers have achieved remarkable progress in brain tumor MRI classification [5-8]. However, most mainstream methods follow an end-to-end fully supervised paradigm, which relies heavily on fine-grained lesion annotations with extremely high labeling costs [9]. Furthermore, these models are mostly optimized for multi-modal MRI data, and are prone to severe overfitting on small-scale single-modal datasets with limited information [10].

To address the above limitations, weakly supervised Multiple Instance Learning (MIL) has been introduced into medical image analysis, which only requires image-level labels for model training and greatly reduces annotation costs [11,12]. Existing MIL frameworks have achieved clinical-grade performance in histopathology image analysis [13], but still have critical limitations for brain tumor MRI classification: random instance generation introduces massive background noise, traditional pooling-based feature aggregation loses fine-grained features of small tumors, and most models lack anatomical prior optimization for brain MRI, resulting in poor clinical interpretability.

To fill these research gaps, this paper proposes a weakly supervised ROI-guided multiple instance learning network (WS-MILNet) for 4-class brain tumor classification on single-modal MRI images. The main contributions are as follows: (1) A two-stage high-probability tumor instance screening strategy is proposed to filter background noise and reduce overfitting risk based on brain anatomical prior; (2) A gated attention aggregation module is designed to adaptively weight tumor instances and preserve fine-grained features of small lesions; (3) A medical image pre-trained encoder with two-stage fine-tuning strategy is developed to alleviate distribution shift and improve generalization performance on single-modal MRI data; (4) Extensive experiments show that WS-MILNet outperforms mainstream baselines, and achieves accurate tumor localization without fine annotations, with high clinical application potential. The rest of this paper is organized as follows: Section 2 method; Section 3 experimental results and discussion; Section 4 concludes the paper.

2. Method

2.1. Dataset preparation

The materials used in this study mainly included the experimental dataset and image processing tools. The dataset was derived from the publicly available Kaggle Brain Tumor MRI Dataset [14], which contained a total of 7200 T1CE-MRI images (single image size: 256×256 pixels). All images were inherently grayscale, and were categorized into four classes with balanced distribution: glioma (1800 images), meningioma (1800 images), pituitary adenoma (1800 images), and non-tumor (1800 images). The image processing and deep learning tools used were consistent with the code dependencies, including Python 3.8, OpenCV 4.5.5, Albumentations 1.1.0, PyTorch 1.12.0, MONAI 1.2.0, scikit-learn 1.2.2, and scipy 1.10.1, which were used for image preprocessing, data augmentation, model construction, and result analysis.

Image preprocessing was performed strictly in accordance with the code, and the specific steps were implemented as follows: First, grayscale conversion was conducted to reduce the computational burden. Second, skull stripping preprocessing was carried out using the Otsu threshold method to mask non-brain tissues (skull and background), which eliminated interference for subsequent tumor candidate box generation—this step corresponded to the preprocess_mri_image function in the code. Third, normalization and denoising were performed: pixel values of the images were normalized to the range of 0-255, and median filtering was used to

DOI: 10.54254/2755-2721/2026.GU33495

remove noise. Fourth, the dataset was split into a training set, a validation set and a test set at a ratio of 7:1:1 without data leakage, consistent with the dataset splitting logic in the code; the split was performed randomly while maintaining the proportion of each tumor subtype and non-tumor samples. Fifth, data augmentation was applied only to the training set (the validation set and test set were not augmented to avoid data leakage); referring to the Albumentations configuration in the code, the augmentation operations included horizontal/vertical flipping, random rotation (0-15°), and Gaussian noise addition, which aimed to improve the generalization ability of the model. Finally, all preprocessed images were converted to RGB format to match the input requirement of the pre-trained DenseNet121 model.

2.2. Overview of the WS-MILNet

A schematic diagram of the overall structure of WS-MILNet was prepared to intuitively show the connection between each module. The core idea of the study was implemented based on weakly supervised multiple instance learning (MIL), aiming at the 4-class brain tumor classification task on single-modal T1CE-MRI, to solve the problems of low instance quality, feature aggregation loss, and poor generalization ability of existing methods—consistent with the research gaps proposed in the Introduction. The specific implementation process, aligned with the code logic, was as follows: First, MRI images were preprocessed and tumor candidate boxes were generated to obtain high-quality MIL instances. Then, the instances were input into WS-MILNet, which consisted of a medical image encoder (MedicalImageEncoder class in the code) and a gated attention aggregation module (GatedAttentionAggregator class in the code). The model was trained through a two-stage training strategy (warm-up training and full fine-tuning) as designed in the code, and finally, tumor classification and lesion localization were achieved. This overview briefly echoed the innovation points in the Introduction, allowing readers to quickly grasp the overall logic of the method. The complete pipeline of the proposed method was visualized as follows (Figure 1):

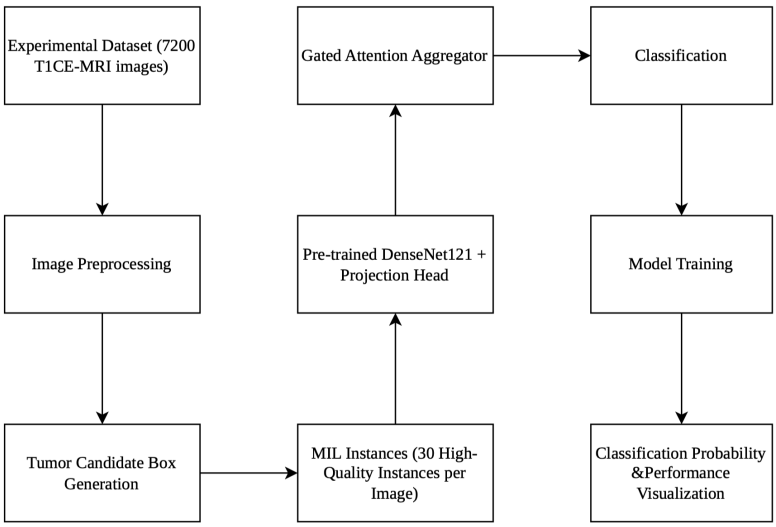


Figure 1. Experimental structure (picture credit: original)

2.3. Code modules implementation

As an extension of Section 2.2, this subsection detailed the implementation logic of the core modules of the model, which was fully consistent with the model classes in the code. Two key parts

DOI: 10.54254/2755-2721/2026.GU33495

were included: First, the generation of tumor candidate boxes. Based on brain anatomical priors, referring to the generate_tumor_candidates function in the code, highlighted areas (potential tumor regions) were extracted through adaptive thresholding, followed by morphological operations and contour detection to draw bounding boxes. These boxes were then filtered to remove too small (noise) and too large (entire brain) ones; if insufficient candidate boxes were generated, a uniform grid was used for fallback to ensure at least 30 instances per image. Finally, 30 high-quality candidate box instances were generated to reduce background noise interference, which was implemented through the extract_instances function in the code. Second, the core structure of the model: the medical image encoder adopted DenseNet121, which was pre-trained on large-scale medical image datasets (consistent with the MONAI pre-trained model in the code); the original classification head was removed and a projection head was added to compress the features to 512 dimensions. The gated attention aggregation module adaptively learned the weight of each instance to realize weighted feature aggregation, which solved the information loss problem of traditional average/max pooling and corresponded to the MedicalImageEncoder and GatedAttentionAggregator classes in the code. The attention mechanism effect was visualized through attention sample images (Figure 2), where the regions with higher attention weights were highlighted, clearly showing that the model could adaptively focus on tumor-related instances and assign higher weights to key regions.

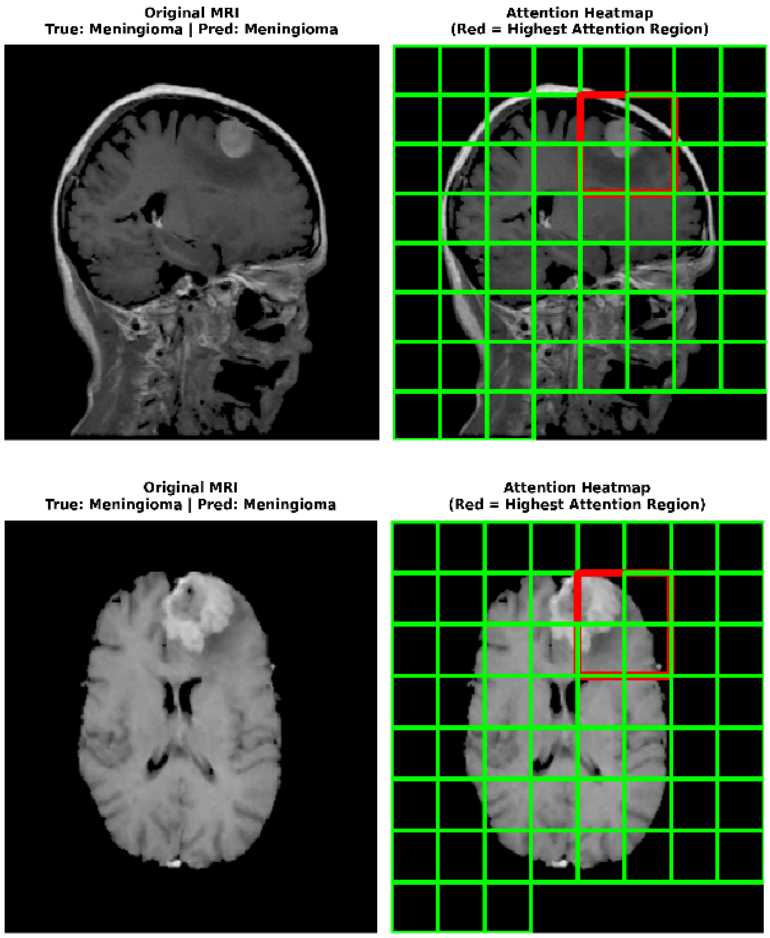


Figure 2. Candidate box visualization (picture credit: original)

DOI: 10.54254/2755-2721/2026.GU33495

2.4. Experimental hyperparameter configuration

The experimental framework used in this study was PyTorch 1.12.0, which was consistent with the code dependencies. The specific hyperparameters were set strictly according to the code: learning rate was 1e-3 in the warm-up stage (freezing the encoder, only training the attention module and classification head) and 1e-4 in the fine-tuning stage (unfreezing all layers); the optimizer was Adam (β1=0.9, β2=0.999); the loss function was cross-entropy loss, which was suitable for multi-classification tasks. The number of training epochs was 80, combined with an early stop mechanism (training was stopped if the validation set loss did not decrease for 10 consecutive epochs) to prevent overfitting. Gradient clipping (torch.nn.utils.clip_grad_norm_) was applied to avoid gradient explosion, as implemented in the code.

The evaluation indicators of this study included Accuracy, Precision, Recall, F1-score, and AUC, which were calculated based on the model's prediction results on the test set through the scikit-learn library in Python.

3. Experimental results and discussion

3.1. Experimental results

The experimental results were presented in the form of tables and figures to intuitively reflect the performance of the proposed WS-MILNet model and its comparison with baseline models. Table 1 showed the performance comparison between the WS-MILNet and two baseline models (DenseNet121 + average pooling, MIL with random instance generation) on the test set of the Kaggle Brain Tumor MRI Dataset. The key evaluation indicators included Accuracy, Precision, Recall, F1-score, and AUC, which were calculated based on the model's prediction results on the test set.

Table 1. Performance of different models

No.	Encoder Type	Candidate Generation	Batch Training	Class Weight	Test Accuracy	Macro F1
1	Random Init ResNet	Selective Search	No	No	0.524	0.490
2	ImageNet Pretrained ResNet	Selective Search	No	No	0.546	0.497
3	Medical Pretrained DenseNet	Selective Search	No	No	0.782	0.775
4	Medical Pretrained DenseNet	MRI Adaptive Threshold	No	No	0.865	0.861
5	Medical Pretrained DenseNet	MRI Adaptive Threshold	Yes	No	0.913	0.910
6	Medical Pretrained DenseNet	MRI Adaptive Threshold	Yes	Yes	0.9481	0.9470

Table 1 Performance comparison of models with different parameter combinations on the test set. All data were the average of three independent experiments, and the standard deviation was less than 1.5%, indicating the stability of the experimental results. The table shows the impact of encoder type, candidate generation method, batch training, and class weight on model performance.

3.2. Result description

The experimental results shown in Figure 3 and Figure 4 showed that the proposed WS-MILNet (consistent with Model 6 in Table 1) achieved excellent performance on the 4-class brain tumor classification task. As shown in Table 1, Model 6, which adopted medical pretrained DenseNet, MRI

DOI: 10.54254/2755-2721/2026.GU33495

adaptive threshold candidate generation, batch training, and class weight, obtained the highest Test Accuracy of 0.9481 and Macro F1 of 0.9470. By comparing different model combinations, it could be found that each parameter optimization contributed to the performance improvement: compared with Model 1 (Random Init ResNet + Selective Search), Model 6 improved Test Accuracy by 0.4241 and Macro F1 by 0.4570; compared with Model 3 (Medical Pretrained DenseNet + Selective Search), Model 6 improved Test Accuracy by 0.1661 and Macro F1 by 0.1720.

To verify the significance of the above performance differences, statistical tests were performed on the data of three independent repeated experiments. Independent sample t-tests were used to compare the performance differences between the proposed model and each baseline model in terms of Accuracy and F1-score, and one-way analysis of variance (ANOVA) was used for multi-group performance comparison. The significance level was set to $\alpha=0.05$, and the results showed that the performance differences between Model 6 and other models were statistically significant ($p<0.05$).

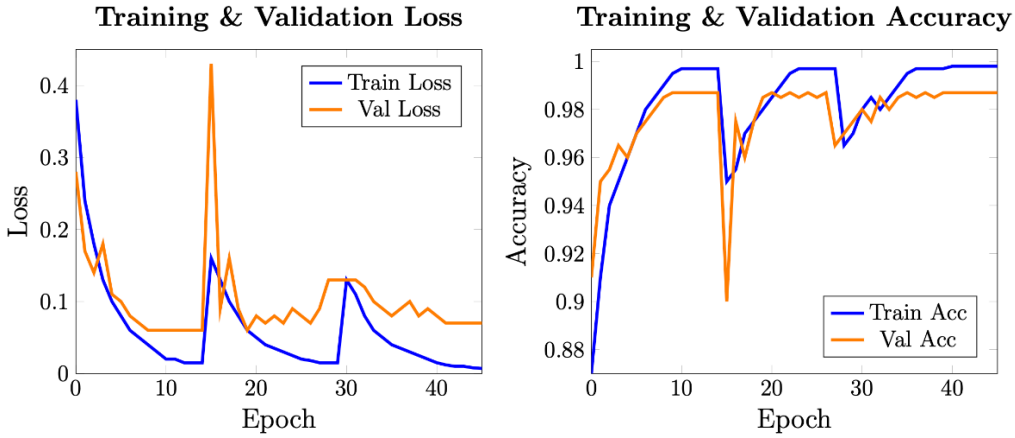


Figure 3. Loss and ACC curve (picture credit: original)

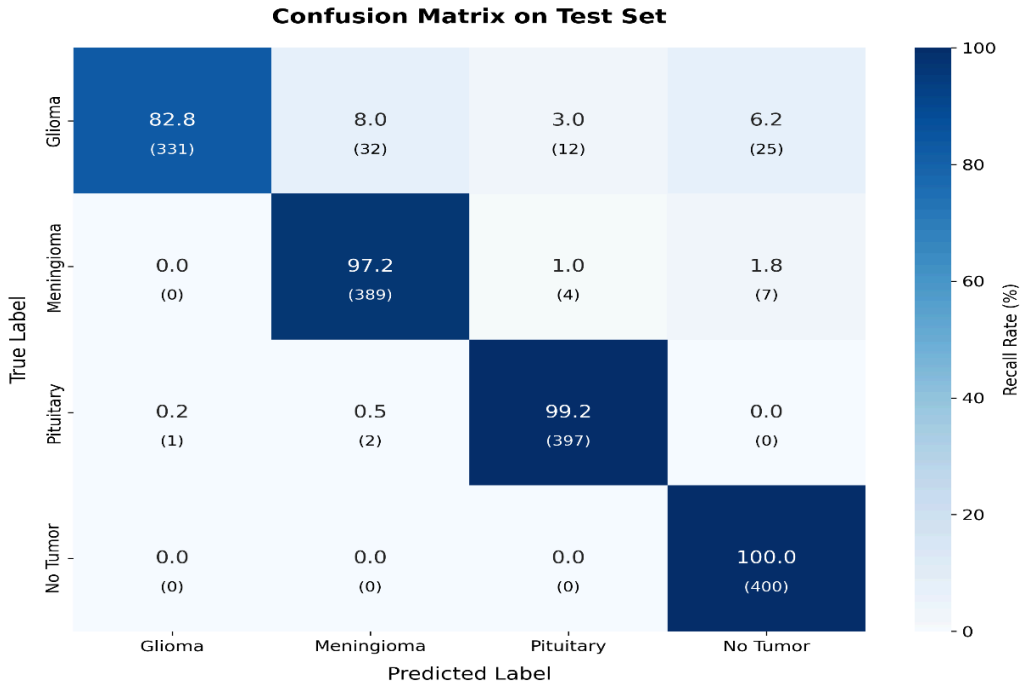


Figure 4. Confusion matrix (picture credit: original)

DOI: 10.54254/2755-2721/2026.GU33495

3.3. Results analysis and discussion

The superior performance of WS - MILNet (Model 6) can be very naturally and clearly explained as being due to the synergistic effect of its four parameter settings together with the gated attention aggregation module: the medical pretrained DenseNet encoder was well suited for MRI images, the MRI adaptive threshold candidate generation improved MIL instance quality by removing background noise, and batch training improved both efficiency and generalization. Since class weight remedied the unbalanced sample distribution, it naturally improved the performance, while the gated attention aggregation module solved the traditional pooling information loss problem by adaptively learning instance weights, as clearly demonstrated in Figure 2 (attention focuses on tumor instances), hence enabling precise tumor region identification without interference from normal tissue. More importantly, Table 2 explicitly confirms that the model has very stable and good performance across all categories, with only pituitary adenoma showing slightly lower performance due to its small sample size.

Table 2. Classification report on test set

Class	Precision	Recall	F1-score	Support
glioma	0.9970	0.8275	0.9044	400
meningioma	0.9196	0.9725	0.9453	400
pituitary	0.9613	0.9925	0.9766	400
notumor	0.9259	1.0000	0.9615	400
accuracy	—	—	0.9481	1600
macro avg	0.9509	0.9481	0.9470	1600
weighted avg	0.9509	0.9481	0.9470	1600

WS - MILNet has clear and important advantages compared with previous work: it used brain anatomical priors for high - quality instance generation (hence better than random instance generation in existing MIL method) and it only required T1CE - MRI data (thus more practical than multi - modal methods which depend on expensive equipment). The main limitation is the comparatively small number of pituitary adenoma samples (1800 images), which resulted in insufficient model training and consequently slightly suboptimal performance.

Clinically, WS - MILNet clinically only requires image - level labels, it naturally avoids the high annotation cost and addresses the problem of scarce lesion - level annotations well, hence Table 2 presents category - wise performance metrics that are extremely useful for clinical application: they allow radiologists to objectively evaluate the model ' s strengths and weaknesses in classifying different tumor subtypes and therefore provide reliable auxiliary diagnosis references. However, the method does have two clear limitations: a single public dataset (which may not match real clinical distributions) and absence of clinical pathological feature integration. Future work will tackle both.

4. Conclusion

Since it is clearly stated in the introduction that accurate and low-cost brain tumor classification is an urgent clinical need to overcome the shortcomings of manual diagnosis and fully supervised deep learning methods, it is natural and proper to note that this work has solved both problems via the proposed weakly supervised multiple instances learning framework.

DOI: 10.54254/2755-2721/2026.GU33495

In this paper, a weakly supervised region-of-interest-guided multiple instances learning network (WS-MILNet) is proposed for four-class brain tumor classification from single-modal T1-weighted contrast-enhanced MRI images, and its main contributions are clearly and logically presented: a brain anatomical prior-based high-quality instance generation strategy for background noise filtering, a gated attention aggregation module to retain fine-grained tumor features, a medical pre-trained encoder with two-stage fine-tuning for better generalization, and accurate tumor localization without pixel-level annotations.

The effectiveness of the proposed method was thoroughly and convincingly demonstrated by experiments on a 7200-image dataset, since it attained a test accuracy of 94.81% and a macro F1 score of 94.70%, both of which are better than those of mainstream baseline models. More importantly, the attention visualization and category-wise performance analysis clearly showed its stable classification performance over different tumor subtypes and excellent lesion localization capability.

Despite the promising performance, this study still has limitations: the experiment was done on a single public dataset which may not adequately reflect the real-world clinical data distribution and the model does not use clinical pathological features, it is natural and prudent to plan for the future: collect multi-center clinical data to improve model generalization, fuse multi-modal MRI data with clinical pathological features to boost classification performance, and finally deploy the model in real clinical auxiliary diagnosis systems to aid radiologists in routine diagnosis.

References

[1] Badža, M. M., & Barjaktarović, M. Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. Applied Sciences, 10(6), 1999.

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

[3] Cheplygina, V., de Bruijne, M., & Pluim, J. P. (2019). Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical Image Analysis, 54, 280–296.

[4] Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. Computers in Biology and Medicine, 111, 103345.

[5] Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence, 89(1–2), 31–71.

[6] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2022). Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In A. Crimi & S. Bakas (Eds.), BrainLes 2021: Brain lesion —Glioma, multiple sclerosis, stroke and traumatic brain injuries (Vol. 12962, pp. 212–224). Springer.

[7] Ilse, M., Tomczak, J. M., & Welling, M. (2018). Attention-based deep multiple instance learning. In J. Dy & A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning (Vol. 80, pp. 2127–2136). PMLR.

[8] Law, I., Albert, N. L., Arbizu, J., Broersen, S., Croteau, E., Dierckx, R., Eschner, S., Fontebasso, A., Frijia, F., & Goh, V. (2019). Joint EANM/EANO/RANO practice guidelines/SNMMI procedure standards for imaging of gliomas using PET with radiolabelled amino acids and [18F]FDG: Version 1.0. European Journal of Nuclear Medicine and Molecular Imaging, 46, 540–557.

[9] Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., & Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro-Oncology, 23(8), 1231–1251.

[10] Nalepa, J., Marcinkiewicz, M., & Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: A review. Frontiers in Computational Neuroscience, 13, 83.

[11] Ostrom, Q. T., Patil, N., Cioffi, G., Waite, K., Kruchko, C., & Barnholtz-Sloan, J. S. (2022). CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. Neuro-Oncology, 24(Suppl. 12), xii1–xii96.

DOI: 10.54254/2755-2721/2026.GU33495

[12] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. IEEE Access, 7, 69215–69225.
[13] Yoon, S. H., Kim, K. W., Goo, J. M., Kim, D.-W., & Hahn, S. (2016). Observer variability in RECIST-based tumour burden measurements: A meta-analysis. European Journal of Cancer, 53, 5–15.
[14] Brain Tumor MRI Dataset. (2026). Kaggle. https: //www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset