JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Pseudocolor - Assisted Segmentation of Brain Tumors, Validation of Interpretability, and Examination of Radiologist Decision-Making.

[1]Devalla Hitesh Sri Sai, [2] Chirag M, [3] Anoop Soudri, [4] Abhimanyu Hiremath, [5] Ms. Dhivya V

[1,2,3,4] UG Students, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India,

[5] Assistant Professor of Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka, India,

ABSTRACT

Deep learning models are proficient at segmenting brain tumors; however, they face challenges in clinical environments due to the "black box" problem and a lack of adequate testing involving real users. This research investigates a pseudocolor-guided U-Net segmentation approach by assessing its technical effectiveness and interpretability for users. The model, which was trained on the BraTS 2021 dataset and utilized confidence-based post-processing, recorded Dice scores of 0.87 for the necrotic core, 0.82 for edema, and 0.85 for the enhancing tumor. A study conducted with five neuroradiologists demonstrated that pseudocolor visualization enhanced boundary identification accuracy by 23% ($p<0.001$). Additionally, it reduced diagnostic time by 29% ($p<0.01$) and increased confidence scores from 3.1 out of 5 to 4.5 out of 5 ($p<0.001$). Interrater agreement improved from $\kappa=0.68$ to $\kappa=0.84$. These results indicate that pseudocolor visualization significantly enhances comprehension and decision-making. This emphasizes the importance of validating AI technologies in medical imaging with an emphasis on human usability.

Keywords—Brain tumor segmentation; Explainable AI; Pseudocolor visualization; User study validation; Deep learning interpretability; Medical imaging; Neuroradiology

1. Introduction

Artificial intelligence has achieved remarkable outcomes in the analysis of medical images, notably in segmenting brain tumors, where convolutional neural networks are performing comparably to expert radiologists. AI-driven diagnostic systems can facilitate quicker diagnoses, decrease interobserver variability, and aid in personalized treatment planning. The primary barrier to clinical implementation remains the inherent lack of transparency in deep learning models. Clearly, interpretability and explainability are critical safety components; healthcare professionals must have confidence in these models, and regulatory agencies must endorse them. Although recent advancements in explainable AI techniques for medical imaging, such as attention maps, GradCAM, and saliency maps, have been made, validation involving clinicians is still limited. In this study, for the first time, a segmentation system utilizing pseudocolor guidance will be proposed and evaluated through a human-centered protocol co-developed with medical professionals.

2. Related Work

Deep learning models such as U-Net and its variations have significantly speed up brain tumor segmentation processes. Benchmark datasets like BraTS facilitate reproducible assessments but also lead to generalization issues across different clinical environments.Various visualization methods have emerged from explainable AI studies, yet validation with clinician involvement remains frequently absent. The perceptual advantages of pseudocolor mapping have been demonstrated in techniques like multi parametric MRI, providing a basis for its application to brain tumor segmentation.

3. Materials and Methods

3.1 System Overview

The system is made up of several components:

(1) preprocessing, (2) a two-dimensional U-Net segmentation model developed using MONAI/PyTorch, (3) post processing through a confidence model and morphological refinement, and (4) an embedding engine for pseudocolor visualization accompanied by a web interface for clinical assessment.

3.2 Data and Preprocessing

We utilized the BraTS 2021 dataset containing multiparametric MRI images, which include T1, T1ce, T2, and FLAIR scans, along with expert-annotated masks marking the necrotic core, edema, and enhancing tumor. The images were resized to 256 × 256, and their intensity values were normalized to the range [0,1]. The channels for T1, T2, and FLAIR were combined to create three-channel inputs. Labels were mapped onto discrete classes:

3.3 Segmentation Model

A 2D U-Net architecture was employed, consisting of an encoder with four blocks (with channels of 64, 128, 256, and 512) and a bottleneck comprising 1024 channels. The decoder mirrored this structure. The final layer generated 4-channel logits (including background and three tumor classes). The model encompasses around 31 million parameters. Training utilized a combined Dice and Cross-Entropy loss (DiceCELoss) with the AdamW optimizer, a ReduceLROnPlateau scheduler, batch size of 4, and early stopping protocols. Training lasted for a maximum of 30 epochs with an 85/15 ratio for the train/validation split.

3.4 Confidence-Based Post-Processing

Logits were transformed into class probabilities using softmax. Class-specific confidence thresholds were established: p1=0.45 for necrotic, p2=0.55 for edema, and p3=0.60 for enhancing tumors. Small connected components below area thresholds were eliminated (50 px for necrotic, 350 px for edema, and 350 px for enhancing) to minimize false positives while maintaining true tumor regions.

3.5 Pseudocolor Mapping

A pseudocolor scheme was developed that is perceptually distinct: the background is set to (128,128,128); the necrotic core is represented by (0,255,0—green); edema is denoted as (0,255,255—cyan); while the enhancing tumor appears in either blue or red based on interpretation. Colors were chosen to ensure they are perceptually separable in the CIE L*a*b* color space and align with established clinical standards. Overlays were created by merging pseudocolor masks with grayscale images using an alpha transparency of 0.4 for optimal visibility.

3.6 Novel Interpretability Metrics

In addition to Dice and Hausdorff distance, the following metrics were introduced:

- Color Distinctiveness Score (CDS): calculated as the average ΔE00 between the pixels at the boundaries of regions in the CIE L*a*b* color space. A higher CDS value indicates improved visual separability.
- Region Boundary Clarity (RBC): computed as the mean gradient magnitude at the boundaries of regions in the pseudocolor image.
- Tumor Detection Confidence: represented by the average softmax probability across correctly identified tumor pixels.

3.7 User Study Design

A total of five observers took part: three board-certified neuroradiologists (8 to 15 years of experience) and two final-year residents in radiology. Using a within-subject design, each observer assessed 20 cases in both grayscale and pseudocolor to evaluate boundary delineation, classify subregions (low vs. high-grade), estimate volume, and gauge confidence on a scale of 1 - 5. Following a training session, the evaluation sequence was arranged as: grayscale assessment, a break, pseudocolor assessment, and finally a comparative survey. The time taken for decisions was recorded automatically.

3.8 Comparative Explainability Methods

Two other interpretability methods were incorporated for benchmarking purposes: self-attention visualization (extracted attention weights from a bottleneck self-attention module) and GradCAM. The radiologists provided ratings for intelligibility, clinical relevance, trust enhancement, and actionability for each of the three methods using a 5-point Likert scale, while metrics such as Dice, decision time, and confidence were gathered as objective measurements.

4. Results

4.1 Segmentation Performance

In the evaluation set containing 45 cases, the metrics per class were:

Tumor Class	**Dice**	**Hausdorff (mm)**	**Precision**	**Recall**	**Specificity**
Necrotic Core	0.87	5.2	0.89	0.85	0.98
Edema	0.82	7.8	0.84	0.80	0.95
Enhancing Tumor	0.85	6.1	0.88	0.83	0.97
Mean	**0.85**	**6.4**	**0.87**	**0.83**	**0.97**

4.2 Effects of Post-processing

Post-processing led to a reduction in false positives across classes: for raw predictions - Precision 0.79, Recall 0.89, FPR 0.14, Dice 0.82; following confidence filtering - Precision 0.87, Recall 0.83, FPR 0.06, Dice 0.85; and after morphological refinement - Precision 0.87, Recall 0.83, FPR 0.03, Dice 0.85.

4.3 Novel Interpretability Metrics

The CDS values for pairs of regions were: Necrotic versus Edema, 72.3 (indicating high distinctness); Necrotic versus Enhancing, 68.5 (indicating high distinctness); Edema versus Enhancing, 45.8 (indicating distinctness). The RBC was found to be 187.4 (in normalized units). The average tumor detection confidence scores were: necrotic, 0.82; edema, 0.76; enhancing, 0.79.

4.4 User Study Findings

The diagnostic accuracy, as measured by participant-defined Dice, significantly increased from 0.71 ± 0.08 to 0.87 ± 0.05 when using pseudocolor ($p<0.001$). The volume estimation error was reduced from 18.3% ± 6.2 to 11.7% ± 4.1 (a decrease of 36%, $p<0.01$). The accuracy of region classification improved from 78.5% to 91.2% (an increase of 12.7%, $p<0.01$). Decision times decreased from 94.8 ± 14.2 seconds to 67.5 ± 9.8 seconds (a reduction of 29%, $p<0.01$), with the most significant savings occurring in the initial tumor detection phase (a reduction of 39%) and in region delineation (a reduction of 28%). Self-reported confidence levels rose from 3.1 to 4.5 ($p<0.001$). The interrater agreement, as measured by Fleiss' κ, improved from 0.68 to 0.84 with the pseudocolor technique ($p<0.05$).

4.5 Comparative Explainability Performance

In general, pseudocolor excelled compared to attention maps and GradCAM across subjective evaluations of comprehensibility, clinical relevance, trust, and actionability. Objectively, it produced better metrics: Dice of 0.87 (pseudocolor) versus 0.74 (attention) and 0.78 (GradCAM); decision times of 67.5 seconds compared to 89.2 seconds and 82.3 seconds; and confidence levels of 4.5 versus 3.1 and 3.4.

4.6 Failure Modes

Identified limitations involve very small tumors (less than 500 pixels), postoperative cases exhibiting bleeding, and low-grade diffuse gliomas. For these particular cases, the Dice score dropped to 0.63, highlighting the need for specialized training data and adjustments specific to the domain.

5. Discussion

The segmentation process utilizing pseudocolor is a semantic visualization that is aligned with tasks and significantly improves radiologist performance when compared to model-focused explanations like attention and GradCAM. This pseudocolor method provides pixel-level class labeling, adheres to clinical conventions, and takes advantage of preattentive color processing for quicker understanding. Future clinical AI research should integrate evaluations centered around human factors and metrics regarding interpretability of reports into their studies, in addition to current assessment of segmentation accuracy. Some obstacles to clinical adoption include integration with PACS/DICOM systems, real-time GPU processing, quality assurance for model stability, and legal considerations. This study does have limitations related to the sample size, reliance on BraTS-specific data, 2D slice segmentation, and issues related to color-blind testing.

6. Conclusion

The pseudocolor guidance significantly improves the interpretability and clinical decision-making in AI-supported brain tumor segmentation. This research presents a reproducible validation framework that connects algorithmic efficacy with human task performance and preferences, advancing the journey towards explainable medical imaging AI that is clinically valuable. Acknowledgments: We express our gratitude to the neuroradiologists who participated and recognize the BraTS organizers for providing the dataset. Funding details and institutional acknowledgments will be included.

References

1. Menze, B. H., et al. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).” IEEE Transactions on Medical Imaging, 2015.
2. Bakas, S., et al. “Advancing glioma MRI collections with expert segmentation labels.” Scientific Data, 2018.
3. Ronneberger, O., Fischer, P., & Brox, T. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” MICCAI, 2015.
4. Isensee, F., et al. “nnU-Net: A self-configuring method for deep learning based biomedical segmentation.” Nature Methods, 2021.
5. Baid, U., et al. “The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation.” arXiv preprint, 2021.
6. Kather, J. N., et al. “Color-coded visualization of multiparametric MRI maps.” Scientific Reports, 2017.
7. Li, T., et al. “Color-appearance-model-based fusion of medical images.” Information Fusion, 2014.
8. Ware, C. Information Visualization: Perception for Design. Morgan Kaufmann, 2019.
9. Lundberg, S. M., & Lee, S. I. “A unified approach to interpreting model predictions.” NeurIPS, 2017.
10. Selvaraju, R. R., et al. “Grad-CAM: Visual explanations from deep networks.” IJCV, 2020.
11. Chen, H., et al. “Explainable medical imaging AI needs human-centered design.” npj Digital Medicine, 2022.
12. Ghassemi, M., et al. “The false hope of current approaches to explainable AI in healthcare.” The Lancet Digital Health, 2021.
13. Ennab, M., & Mcheick, H. “Improving interpretability and accuracy of AI in medical imaging.” Frontiers in Medicine, 2024.
14. Ostrom, Q. T., et al. “CBTRUS Statistical Report: Brain and CNS Tumors in the US.” Neuro-Oncology, 2022.
15. Taha, A. A., & Hanbury, A. “Metrics for evaluating medical image segmentation.” BMC Medical Imaging, 2015.
16. Rogowitz, B. E., & Treinish, L. A. “Data visualization: The end of the rainbow.” IEEE Spectrum, 1998.
17. Badano, A., et al. “Consistency and standardization of color in medical imaging.” Journal of Digital Imaging, 2015.
18. Rudin, C. “Stop explaining black box models and use interpretable models instead.” Nature Machine Intelligence, 2019.