www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification

Amirreza Fateh[1], Yasin Rezvani[2], Sara Moayedi[2], Sadjad Rezvani[2], Fatemeh Fateh[3], Mansoor Fateh[2] ✉ & Vahid Abolghasemi[4] ✉

Accurate segmentation and classification of brain tumors from Magnetic Resonance Imaging (MRI) remain key challenges in medical image analysis, primarily due to the lack of high-quality, balanced, and diverse datasets with expert annotations. In this work, we address this gap by introducing BRISC, a dataset designed for brain tumor segmentation and classification tasks, featuring high-resolution segmentation masks. The dataset comprises 6,000 contrast-enhanced T1-weighted MRI scans, which were collated from multiple public datasets that lacked segmentation labels. Our primary contribution is the subsequent expert annotation of these images, performed by certified radiologists and physicians. It includes three major tumor types, namely glioma, meningioma, and pituitary, as well as non-tumorous cases. Each sample includes high-resolution labels and is categorized across axial, sagittal, and coronal imaging planes to facilitate robust model development and cross-view generalization. To demonstrate the utility of the dataset, we provide benchmark results for both tasks using standard deep learning models. The BRISC dataset is made publicly available.

Background & Summary

Brain tumors are among the most critical medical conditions, requiring accurate and timely diagnosis for effective treatment and management[1,2]. Magnetic Resonance Imaging (MRI) plays a critical tool in diagnosing and monitoring brain tumors, owing to its non-invasive imaging capabilities and ability to provide detailed visualization of brain structures[3–5]. Despite significant advancements in medical imaging technologies, developing automated systems for tumor detection and segmentation remains a major challenge[6–8]. This challenge arises primarily from the scarcity of high-quality labeled datasets designed for these tasks. Additionally, the complexity and variability of tumor appearances across patients further complicate accurate segmentation and classification[9–11].

Existing brain tumor segmentation datasets, such as the Brain Tumor Segmentation (BraTS)[12], Cheng[13] brain tumor dataset, and others, have significantly advanced the development of automated segmentation models. However, several limitations in these datasets drive the need for novel datasets to address emerging challenges in the field. For example, the BraTS dataset is widely used and comprehensive. However, it depends on pre-processed, standardized data that may not reflect the real-world variability in MRI acquisition protocols across institutions. Additionally, BraTS primarily focuses on gliomas and lacks representation of other tumor types, potentially limiting the generalizability of models trained on it[14,15]. The Cheng[13] brain tumor dataset, on the other hand, suffers from class imbalance and limited diversity in imaging conditions and patient demographics, which can restrict model robustness[16]. Many publicly available datasets face issues with inconsistent labeling. These inconsistencies can negatively affect the training and evaluation of segmentation models[17–20]. These limitations underscore the necessity of introducing a new dataset that offers balanced class distributions, multi-institutional diversity, and high-quality expert annotations to enhance the reliability and generalizability of automated brain tumor segmentation models. Additionally, including class labels for classification tasks, such as identifying glioma, meningioma, pituitary, and non-tumorous cases, enhances the dataset's utility and supports broader real-world applications in brain tumor analysis.

[1]School of Computer Engineering, Iran University of Science and Technology (IUST), Tehran, Iran. [2]Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran. [3]Northern Care Alliance NHS Foundation Trust (NCA), Manchester, UK. [4]School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK. ✉e-mail: mansoor_fateh@shahroodut.ac.ir; v.abolghasemi@essex.ac.uk

Class	Training Images	Testing Images	Total Images
Glioma	1,147	254	1,401
Meningioma	1,329	306	1,635
Pituitary	1,457	300	1,757
non-tumorous	1,067	140	1,207
Total	**5,000**	**1,000**	**6,000**

Table 1. Class distribution in the training and testing parts of BRISC.

	Train			Test		
Class\Plane	**Axial**	**Coronal**	**Sagittal**	**Axial**	**Coronal**	**Sagittal**
Glioma	347	428	372	85	81	88
Meningioma	423	426	480	134	89	83
Pituitary	428	496	533	116	98	86
non-tumorous	352	310	405	52	48	40
Total per plane	1550	1660	1790	387	316	297
Total	**5000**			**1000**		

Table 2. Class distribution based on MRI planes in the training and testing parts of BRISC.

To address these gaps, we present the BRISC dataset: a large-scale, balanced, and expert-annotated MRI dataset designed for both segmentation and classification of brain tumors. BRISC includes 6,000 contrast-enhanced T1-weighted MRI scans across four categories—glioma, meningioma, pituitary, and non-tumorous cases—covering multiple anatomical planes (axial, coronal, sagittal). The dataset emphasizes consistent quality, balanced distributions, and multi-institutional diversity, making it suitable for developing models that generalize across clinical settings.

Methods

The BRain tumor Image Segmentation and Classification (BRISC) dataset has been meticulously curated to address key challenges in brain tumor research, particularly in the domains of segmentation and classification tasks. It provides a balanced, high-quality collection of MRI data, annotated for both research and clinical applications. The dataset includes images with labels for four categories: Glioma, Meningioma, Pituitary tumors, and non-tumorous. By focusing on comprehensive data collection and rigorous annotation processes, the dataset aims to advance the development of robust machine learning models in medical imaging.

Motivation and Aims.

The primary goal of collecting and releasing this dataset is to overcome limitations observed in existing brain tumor datasets, such as class imbalance, lack of diversity, and annotation inconsistencies. While datasets like BraTS have driven significant advancements in glioma segmentation, their exclusive focus on gliomas and reliance on pre-processed data limit their generalizability to other tumor types and real-world scenarios. Our dataset expands the scope by incorporating multiple tumor types and includes a “non-tumorous” class to aid in broader diagnostic tasks. This addition makes the dataset highly versatile, enabling its use in applications ranging from multi-class tumor classification to binary tumor detection.

Dataset Composition and Planar Distributions.

The dataset comprises 6,000 MRI images, divided into training and testing sets, as detailed in Table 1. This structured division ensures robust evaluation metrics while providing ample data for training advanced machine learning models. For the training dataset, the total number of images across the planes is 5,000, and for testing, the total is 1,000.

In addition to the class-based distribution, we provide another form of distribution, which is dataset composition by MRI planes. This breakdown categorizes images into Coronal, Sagittal, and Axial planes, helping to analyze how different orientations are represented in the dataset. As shown in Table 2, the distribution of different MRI planes is nearly uniform, similar to the distribution of different classes in Table 1. This balanced distribution ensures that no particular class or plane is overrepresented, which is crucial for preventing model bias and improving generalization.

Data Preprocessing.

The original collection contains 7,023 brain MR images across four classes: glioma, meningioma, pituitary tumour, and non-tumorous. We then applied the following steps to ensure consistency and quality:

- **Sequence harmonization**: Only T1-weighted MRI sequences were retained to ensure consistency.
- **Label and mask verification**: A radiologist and a physician reviewed images to identify incorrect or inconsistent labels; images with tumour labels unsupported by the visible image, or with empty/misaligned masks, were removed.
- **Artefact/corruption screening**: Corrupted files and images with severe artefacts were excluded.

- **De-duplication and redundancy control**: Exact duplicates and near-duplicates (including consecutive images from very short series) were removed; de-duplication was completed before any train/test split to avoid leakage and over-representation.
- **Standardization**: Images were resized and borders/margins adjusted for consistent spatial dimensions.

Imaging Details.

All images in the dataset are T1-weighted contrast-enhanced MRI scans, selected specifically from the "Brain Tumor MRI Dataset" (Kaggle)[21]. Although the original dataset included some T2-weighted images, we exclusively selected T1-weighted scans for their superior ability to highlight tumor boundaries effectively. Another notable characteristic of this dataset is the length of MRI sequences. While typical brain MRI studies often consist of longer sequences, the majority of sequences in this dataset were notably short, ranging from 1 to 5 images per sequence. Sequences with only one image were excluded, as even experienced radiologists and physicians found it challenging to identify tumors accurately in these cases.

The original dataset provided only a train/test split and did not include any information about patients, sequences, or slices, making it impossible to definitively link images to the same subject. To minimize the risk of the same person appearing in both train and test sets, our annotators manually reviewed images as much as possible—through visual similarity checks and metadata cross-referencing—to separate likely same-subject cases. While complete subject-level independence cannot be guaranteed due to the source limitations, this conservative approach ensures no obvious same-patient images cross splits; multiple images from the same subject may therefore be present within a single split. This is documented in the repository metadata.

Annotation Process.

The dataset underwent a meticulous annotation and review process to ensure accuracy and reliability. Annotation was performed using the AnyLabeling tool[22], which facilitated precise delineation of tumorous and non-tumorous regions. Each image was reviewed and refined multiple times with input from a certified physician and radiologist. Key steps in the annotation process included:

- **Tumor Mask Refinement**: Using AnyLabeling[22], regions corresponding to tumorous lesions were iteratively refined to ensure accurate segmentation masks.
- **Class Verification**: The "non-tumorous" class was reviewed in detail. It includes both completely healthy brains and scans with non-neoplastic, space-occupying lesions (e.g., abscesses, cysts). Images misclassified in the original datasets were corrected or removed as needed.
- **Consensus Reviews**: Annotation was carried out by a trained team and verified under the supervision of a certified radiologist and physician. Discrepancies were detected through visual comparison and overlap inspection in AnyLabeling and resolved collaboratively during review sessions. A quality assessment on a representative subset showed a mean Dice coefficient of 0.924 between initial and expert-verified masks. Approximately 4.8% of images required correction, after which full consensus and final approval were achieved for all BRISC annotations.

Limitations and Intended Use.

BRISC comprises only contrast-enhanced T1-weighted MRI images collected from multiple public datasets that lack comprehensive acquisition metadata such as scanner type, field strength, or sequence parameters. Consequently, detailed harmonization across hardware and protocol variations was not achievable. The dataset is intended primarily for research in algorithm benchmarking, model comparison, and methodological development for brain-tumour segmentation and classification tasks. It is not designed or validated for direct clinical diagnostic use. Users should be aware that models trained on BRISC may experience domain-shift when applied to non-contrast T1-weighted, T2-weighted, or institution-specific datasets, and additional data normalization or fine-tuning is recommended to mitigate such effects.

Data Records

The BRISC (BRain tumor Image Segmentation and Classification) dataset is available in Kaggle[23]. The dataset release includes:

- Dataset images
- Manifest file (`manifest.csv`)
- JSON metadata (`manifest.json`)
- File checksums and per-file metadata

The public release follows the directory structure shown in Fig. 1.

File Naming Convention.

Filenames follow this pattern:

```
brisc2025_<split>_<index>_<tumor_code>_
<plane_code>_<sequence>.<ext>
```

Example: `brisc2025_test_00010_gl_ax_t1.jpg`
For segmentation pairs, images and masks share the same basename (image: `.jpg`, mask: `.png`).

Metadata Fields.

Each image row in the manifest includes:

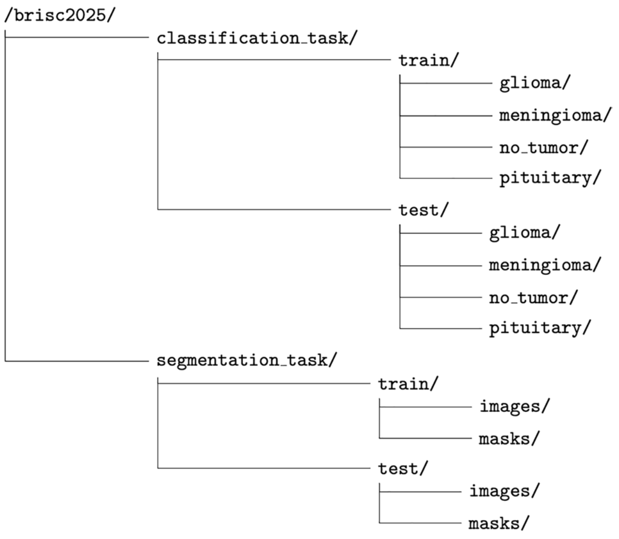

Fig. 1 Directory structure of the BRISC dataset.

- `relative_path`, `filename`, `task`, `split`
- `index`, `tumor_code`, `tumor_label`
- `plane_code`, `plane_label` (ax=axial, co=coronal, sa=sagittal)
- `sequence` (e.g., t1=T1-weighted)
- Spatial dimensions (`width`, `height`)
- `file_size_bytes`, `sha256` checksum

Mask rows have `is_mask = True` and include `linked_image` pointing to the matched image. BRISC is derived from Nickparvar "Brain Tumor MRI Dataset"[21], which aggregates:

- Cheng brain tumor dataset[13]
- SARTAJ[24]
- Br35H[25]

Selection and exclusion rules applied during curation are documented in the Methods section.

Data Overview

BRISC contains 2D single-slice T1-weighted brain MRI images stored as JPEG files, with corresponding pixel-wise segmentation masks stored as PNG files. The dataset is organized into two separate tasks:

1) **Classification task:** Balanced image-level classification dataset for four diagnostic classes: glioma, meningioma, pituitary tumor, and no_tumor. Contains 6,000 JPEG images (5,000 training, 1,000 test).
2) **Segmentation task:** Pixel-wise tumor annotation task. Contains 4,793 image files with exact paired masks. Figure 2 shows an example image from the dataset together with its mask.

Technical Validation

Establishing baseline performance is a critical step in evaluating any newly proposed dataset, as it sets a reference point for further research and model development[26]. To validate the effectiveness and versatility of the BRISC dataset, we conducted experimental evaluations on segmentation and classification tasks.

This section presents a comprehensive performance analysis of several standard baseline models. We report their performance metrics to establish clear benchmarks for the dataset. These results demonstrate the dataset's quality and utility for developing and testing new models in medical imaging.

Evaluation Metrics.

Segmentation Metric. In this part, we detail the evaluation metrics employed to assess the performance of segmentation models on the BRISC dataset. These metrics provide comprehensive insights into model efficacy for these two distinct tasks.

Intersection over Union (IoU). Intersection over Union (IoU), also known as the Jaccard Index, is a fundamental metric for evaluating binary segmentation tasks. It quantifies the overlap between the predicted tumor regions and the ground truth, normalized by their union[27]. For binary segmentation, IoU is computed as shown in Equation (1).

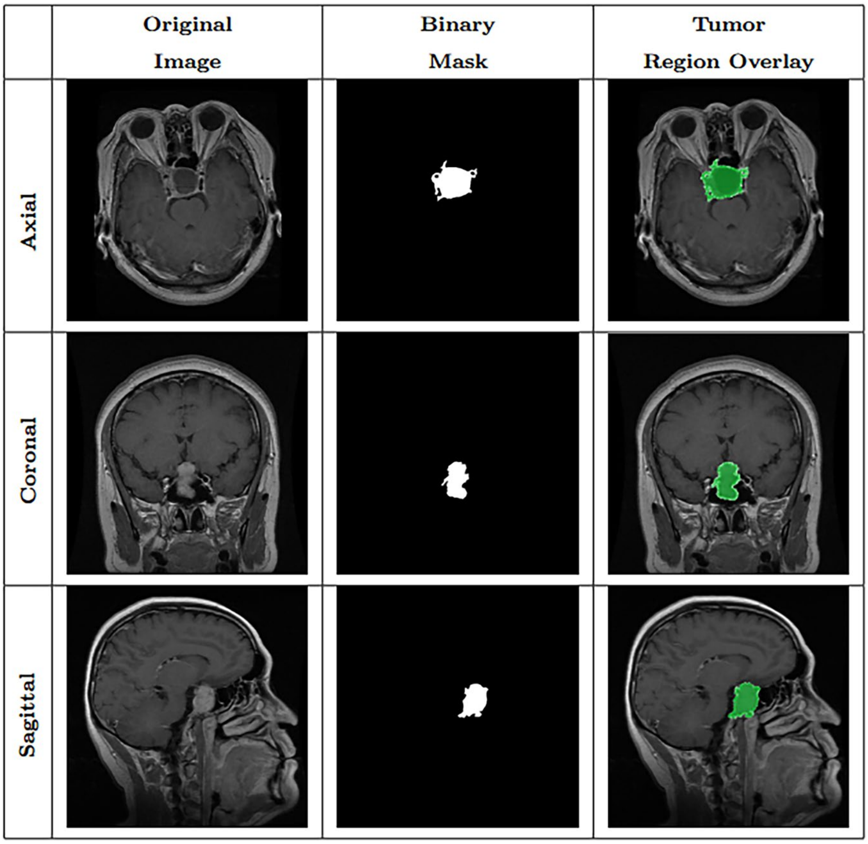

Fig. 2 Samples of Pituitary segmentation across different imaging planes.

Model	mIoU Glioma	mIoU Meningioma	mIoU Pituitary	Weighted mIoU
UNet[31]	69.7	77.1	79.3	75.7
UNet++[32]	71.7	74.2	79.7	75.3
MANet[34]	72.4	77.5	78.0	76.2
LinkNet[33]	71.7	74.8	79.0	75.3
DeepLabV3+[35]	72.0	77.5	78.7	76.3
PAN[36]	72.0	74.5	80.7	75.9
EINet[37]	73.6	78.4	80.3	77.7
EU-Net[38]	71.7	76.1	78.3	75.6
DAD[39]	75.2	80.4	82.3	79.5
BASNet[40]	74.0	77.5	81.7	77.9
SaberNet[41]	74.0	82.4	84.3	80.6
ABANet[42]	72.4	80.4	84.7	79.5

Table 3. IoU (%) for Brain Tumor Segmentation Models on Different Tumor Types. Weighted mIoU is calculated as a weighted average based on the number of samples per tumor type: Glioma, Meningioma, Pituitary.

$$\mathrm{IoU} = \frac{\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i - \sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}, \tag{1}$$

where $y_i \in \{0, 1\}$ denotes the ground truth label, $\hat{y}_i \in \{0, 1\}$ represents the predicted label (after thresholding), and ϵ is a small constant added for numerical stability.

As shown in Equation (1), this formulation captures the pixel-wise overlap between the predicted and actual tumor regions and is particularly effective for evaluating segmentation quality, especially along object boundaries.

Model	Class	Precision	Recall	F1-Score	Accuracy
ResNet50	glioma	0.9868 ± 0.0098	0.9751 ± 0.0164	0.9808 ± 0.0103	—
	meningioma	0.9815 ± 0.0150	0.9673 ± 0.0226	0.9741 ± 0.0087	—
	no_tumor	0.9906 ± 0.0107	0.9952 ± 0.0083	0.9929 ± 0.0035	—
	pituitary	0.9756 ± 0.0285	0.9967 ± 0.0000	0.9859 ± 0.0146	—
	Macro Avg	0.9836 ± 0.0064	0.9836 ± 0.0077	0.9834 ± 0.0072	—
	Weighted Avg	0.9823 ± 0.0076	0.9820 ± 0.0080	0.9820 ± 0.0080	0.9820 ± 0.0080
ResNet101	glioma	0.9726 ± 0.0347	0.9869 ± 0.0082	0.9794 ± 0.0138	—
	meningioma	0.9879 ± 0.0081	0.9575 ± 0.0279	0.9722 ± 0.0106	—
	no_tumor	0.9883 ± 0.0107	0.9905 ± 0.0165	0.9893 ± 0.0037	—
	pituitary	0.9793 ± 0.0113	0.9956 ± 0.0020	0.9874 ± 0.0066	—
	Macro Avg	0.9820 ± 0.0074	0.9826 ± 0.0093	0.9821 ± 0.0086	—
	Weighted Avg	0.9815 ± 0.0085	0.9810 ± 0.0092	0.9810 ± 0.0092	0.9809 ± 0.0092
DenseNet121	glioma	0.4838 ± 0.4753	0.5879 ± 0.5095	0.5197 ± 0.4731	—
	meningioma	0.6640 ± 0.5751	0.2800 ± 0.4569	0.3178 ± 0.4966	—
	no_tumor	0.5324 ± 0.4095	0.6976 ± 0.4871	0.4721 ± 0.4204	—
	pituitary	0.4502 ± 0.4128	0.6422 ± 0.5574	0.5259 ± 0.4678	—
	Macro Avg	0.5326 ± 0.4550	0.5519 ± 0.3376	0.4589 ± 0.4310	—
	Weighted Avg	0.5356 ± 0.4650	0.5253 ± 0.3850	0.5253 ± 0.3850	0.4531 ± 0.4390
DenseNet169	glioma	0.9543 ± 0.0690	0.3543 ± 0.5356	0.3840 ± 0.5173	—
	meningioma	0.7522 ± 0.2090	0.8007 ± 0.2607	0.7560 ± 0.2021	—
	no_tumor	0.4841 ± 0.4507	0.9738 ± 0.0393	0.5754 ± 0.3763	—
	pituitary	0.3333 ± 0.5774	0.3322 ± 0.5754	0.3328 ± 0.5764	—
	Macro Avg	0.6310 ± 0.3116	0.6152 ± 0.3305	0.5121 ± 0.4160	—
	Weighted Avg	0.6404 ± 0.3020	0.5710 ± 0.3689	0.5710 ± 0.3689	0.5093 ± 0.4169
MobileNetV2	glioma	0.3026 ± 0.0558	0.8517 ± 0.1229	0.4418 ± 0.0494	—
	meningioma	0.6667 ± 0.5774	0.0120 ± 0.0180	0.0233 ± 0.0348	—
	no_tumor	0.5343 ± 0.3066	0.6548 ± 0.1750	0.5249 ± 0.0899	—
	pituitary	0.1412 ± 0.2445	0.0400 ± 0.0693	0.0623 ± 0.1080	—
	Macro Avg	0.4112 ± 0.2078	0.3896 ± 0.0345	0.2631 ± 0.0307	—
	Weighted Avg	0.3980 ± 0.2379	0.3237 ± 0.0264	0.3237 ± 0.0264	0.2115 ± 0.0366
MobileNetV3	glioma	0.8912 ± 0.0353	0.9777 ± 0.0082	0.9321 ± 0.0154	—
	meningioma	0.9755 ± 0.0050	0.8639 ± 0.0334	0.9160 ± 0.0175	—
	no_tumor	0.9445 ± 0.0308	1.0000 ± 0.0000	0.9713 ± 0.0165	—
	pituitary	0.9679 ± 0.0073	0.9733 ± 0.0208	0.9706 ± 0.0138	—
	Macro Avg	0.9448 ± 0.0148	0.9537 ± 0.0113	0.9475 ± 0.0140	—
	Weighted Avg	0.9475 ± 0.0123	0.9447 ± 0.0140	0.9447 ± 0.0140	0.9442 ± 0.0142
EfficientNetB0	glioma	0.9960 ± 0.0000	0.9882 ± 0.0000	0.9921 ± 0.0000	—
	meningioma	0.9934 ± 0.0000	0.9869 ± 0.0000	0.9902 ± 0.0000	—
	no_tumor	0.9929 ± 0.0000	1.0000 ± 0.0000	0.9964 ± 0.0000	—
	pituitary	0.9868 ± 0.0000	0.9967 ± 0.0000	0.9917 ± 0.0000	—
	Macro Avg	0.9923 ± 0.0000	0.9929 ± 0.0000	0.9926 ± 0.0000	—
	Weighted Avg	0.9920 ± 0.0000	0.9920 ± 0.0000	0.9920 ± 0.0000	0.9920 ± 0.0000
EfficientNetB1	glioma	0.9987 ± 0.0023	0.9921 ± 0.0000	0.9954 ± 0.0011	—
	meningioma	0.9933 ± 0.0001	0.9750 ± 0.0136	0.9840 ± 0.0070	—
	no_tumor	0.9976 ± 0.0041	1.0000 ± 0.0000	0.9988 ± 0.0021	—
	pituitary	0.9773 ± 0.0116	1.0000 ± 0.0000	0.9885 ± 0.0059	—
	Macro Avg	0.9918 ± 0.0036	0.9918 ± 0.0034	0.9917 ± 0.0036	—
	Weighted Avg	0.9905 ± 0.0040	0.9903 ± 0.0042	0.9903 ± 0.0042	0.9903 ± 0.0042
EfficientNetB2	glioma	0.9919 ± 0.0040	0.9712 ± 0.0164	0.9814 ± 0.0091	—
	meningioma	0.9699 ± 0.0128	0.9782 ± 0.0105	0.9740 ± 0.0084	—
	no_tumor	0.9906 ± 0.0107	1.0000 ± 0.0000	0.9953 ± 0.0054	—
	pituitary	0.9879 ± 0.0082	0.9922 ± 0.0077	0.9900 ± 0.0017	—
	Macro Avg	0.9851 ± 0.0054	0.9854 ± 0.0047	0.9852 ± 0.0051	—
	Weighted Avg	0.9838 ± 0.0049	0.9837 ± 0.0049	0.9837 ± 0.0049	0.9837 ± 0.0049
Xception	glioma	0.0847 ± 0.1466	0.3333 ± 0.5774	0.1350 ± 0.2339	—
	meningioma	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	—
	no_tumor	0.0933 ± 0.0808	0.6667 ± 0.5774	0.1637 ± 0.1418	—

Continued

Model	Class	Precision	Recall	F1-Score	Accuracy
	pituitary	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	—
	Macro Avg	0.0445 ± 0.0165	0.2500 ± 0.0000	0.0747 ± 0.0230	—
	Weighted Avg	0.0346 ± 0.0259	0.1780 ± 0.0658	0.1780 ± 0.0658	0.0572 ± 0.0395
VGG16	glioma	0.9803 ± 0.0150	0.9396 ± 0.0741	0.9582 ± 0.0335	—
	meningioma	0.9427 ± 0.0509	0.9684 ± 0.0019	0.9549 ± 0.0257	—
	no_tumor	0.9790 ± 0.0068	0.9976 ± 0.0041	0.9882 ± 0.0020	—
	pituitary	0.9867 ± 0.0099	0.9822 ± 0.0117	0.9844 ± 0.0051	—
	Macro Avg	0.9722 ± 0.0132	0.9720 ± 0.0172	0.9714 ± 0.0162	—
	Weighted Avg	0.9706 ± 0.0157	0.9693 ± 0.0177	0.9693 ± 0.0177	0.9692 ± 0.0178
VGG19	glioma	0.9484 ± 0.0351	0.9541 ± 0.0421	0.9502 ± 0.0081	—
	meningioma	0.9624 ± 0.0130	0.9434 ± 0.0068	0.9527 ± 0.0030	—
	no_tumor	0.9725 ± 0.0231	0.9976 ± 0.0041	0.9848 ± 0.0111	—
	pituitary	0.9744 ± 0.0287	0.9745 ± 0.0267	0.9739 ± 0.0039	—
	Macro Avg	0.9645 ± 0.0061	0.9674 ± 0.0041	0.9654 ± 0.0047	—
	Weighted Avg	0.9639 ± 0.0039	0.9630 ± 0.0044	0.9630 ± 0.0044	0.9629 ± 0.0044
InceptionV3	glioma	0.6564 ± 0.5686	0.5315 ± 0.4983	0.5778 ± 0.5128	—
	meningioma	0.8972 ± 0.1694	0.6645 ± 0.5132	0.6401 ± 0.4458	—
	no_tumor	0.5887 ± 0.4180	0.9905 ± 0.0165	0.6719 ± 0.3794	—
	pituitary	0.6629 ± 0.5741	0.5722 ± 0.5136	0.6104 ± 0.5343	—
	Macro Avg	0.7013 ± 0.3670	0.6897 ± 0.3750	0.6250 ± 0.4676	—
	Weighted Avg	0.7225 ± 0.3490	0.6487 ± 0.4312	0.6487 ± 0.4312	0.6198 ± 0.4797

Table 4. Per-Class and Average Classification Performance (%) for Brain Tumor Classification Models. Metrics are reported as mean ± standard deviation over three runs.

Classification Metrics. As commonly employed in the evaluation of multi-class classification models, metrics such as Accuracy, Precision, Recall, and F1-Score are widely utilized due to their effectiveness in assessing performance across diverse tasks[28–30]. These metrics are defined below to ensure a comprehensive understanding of their applicability to our dataset.

Accuracy. Accuracy measures the overall correctness of predictions across all four classes. It is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} \text{Correct Predictions for Class } i}{\text{Total Samples}} \tag{2}$$

where C denotes the total number of classes, and "Correct Predictions for Class i" represents the samples correctly classified as class i.

Precision. Precision quantifies the proportion of correctly predicted positive instances for each class. For class i, Precision is defined as:

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{3}$$

In multi-class classification, Precision is averaged using either macro-averaging or weighted-averaging:

$$\text{Macro Precision} = \frac{1}{C}\sum_{i=1}^{C} \text{Precision}_i \tag{4}$$

$$\text{Weighted Precision} = \frac{\sum_{i=1}^{C} w_i \cdot \text{Precision}_i}{\sum_{i=1}^{C} w_i} \tag{5}$$

where w_i represents the proportion of samples in class i.

Recall. Recall, or Sensitivity, measures the proportion of actual positive instances correctly identified by the model. For class i, Recall is defined as:

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{6}$$

For multi-class classification, Recall is averaged similarly to Precision:

$$\text{Macro Recall} = \frac{1}{C}\sum_{i=1}^{C}\text{Recall}_i \tag{7}$$

$$\text{Weighted Recall} = \frac{\sum_{i=1}^{C} w_i \cdot \text{Recall}_i}{\sum_{i=1}^{C} w_i} \tag{8}$$

F1-Score. The F1-Score is the harmonic mean of Precision and Recall. For class i, it is defined as:

$$\text{F1} - \text{Score}_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{9}$$

For multi-class classification, F1-Score is averaged as follows:

$$\text{Macro F1} - \text{Score} = \frac{1}{C}\sum_{i=1}^{C}\text{F1} - \text{Score}_i \tag{10}$$

$$\text{Weighted F1} - \text{Score} = \frac{\sum_{i=1}^{C} w_i \cdot \text{F1} - \text{Score}_i}{\sum_{i=1}^{C} w_i} \tag{11}$$

By calculating these metrics per class and aggregating them through macro- or weighted-averaging, we ensure a detailed evaluation of model performance, particularly in datasets with imbalanced class distributions.

Comparison. *Segmentation results.* To establish segmentation benchmarks for the dataset, we conducted a comparative study against a diverse set of brain tumor segmentation models, including traditional convolutional architectures, attention-enhanced methods, and transformer-based approaches. The baselines include UNet[31], UNet++[32], LinkNet[33], MANet[34], DeepLabV3+[35], PAN[36], EINet[37], EU-Net[38], DAD[39], and BASNet[40], as well as two recent transformer-enhanced models, SaberNet[41] and ABANet[42].

Each model was evaluated using the mean Intersection over Union (mIoU) metric for three tumor types: *Glioma*, *Meningioma*, and *Pituitary*. Furthermore, we report a *weighted mIoU*, which is calculated based on the proportion of samples belonging to each tumor type, providing a more representative performance indicator across the dataset.

As summarized in Table 3, the results establish a strong set of baselines. Transformer-based models, such SaberNet, achieved the highest scores, with weighted mIoUs of 80.6% respectively. This suggests that architectures adept at capturing multi-scale contextual features perform well on this dataset. More traditional architectures like UNet provided a solid baseline with a weighted mIoU of 75.7%.

The reported weighted mIoU is calculated as a weighted average based on the number of samples in each tumor class to provide a more realistic assessment under dataset imbalance. Unlike simple arithmetic means, the weighted mean better reflects the overall segmentation performance in real-world clinical distributions.

It is important to emphasize that the primary goal of this work is to introduce and validate a new brain tumor segmentation dataset, which is designed to support the development of robust and generalizable medical segmentation models. These results serve as a foundational step, and future research is expected to build upon this dataset to explore a broader range of models, training protocols, and evaluation settings.

Classification Results. To evaluate the classification performance on our newly introduced brain tumor dataset, we conducted a comprehensive analysis of several baseline models for classifying brain tumor types: *Glioma*, *Meningioma*, *Pituitary*, and *non-tumorous*. The evaluated models include ResNet50, ResNet101, DenseNet121, DenseNet169, MobileNetV2, MobileNetV3, EfficientNetB0, EfficientNetB1, EfficientNetB2, Xception, VGG16, VGG19, and InceptionV3. Each model was trained and tested three times to ensure robust and reliable results, with performance reported as the mean and standard deviation of key metrics: Precision, Recall, F1-Score, and Accuracy.

The evaluation metrics were computed per class, alongside macro and weighted averages, to provide a comprehensive view of model performance across diverse tumor types. The macro average treats all classes equally, while the weighted average accounts for class imbalance by weighting each class's contribution based on the number of samples, offering a realistic assessment of performance in clinical scenarios where tumor type distributions may vary.

As presented in Table 4, the benchmarks show that high classification accuracy is achievable on this dataset. EfficientNetB0 performs strongly, with a weighted average F1-score of 0.9920 ± 0.0000 and an accuracy of 0.9920 ± 0.0000, achieving perfect recall (1.0000 ± 0.0000) for the *non-tumorous* class. EfficientNetB1 follows closely with a weighted F1-score of 0.9903 ± 0.0042, while ResNet50 and MobileNetV3 deliver competitive results (weighted F1-scores of 0.9820 ± 0.0080 and 0.9447 ± 0.0140, respectively).

In contrast, Xception exhibits the lowest performance, with a weighted F1-score of 0.1780 ± 0.0658, failing entirely on meningioma and pituitary (F1-score: 0.0000 ± 0.0000). Similarly, DenseNet121 and DenseNet169 show unstable performance, with high standard deviations, indicating limited generalizability. MobileNetV2 also struggles, particularly with meningioma (recall: 0.0120 ± 0.0180), likely due to insufficient model capacity.

The VGG variants (VGG16 and VGG19) achieve moderate performance, with weighted F1-scores of 0.9693 ± 0.0177 and 0.9630 ± 0.0044, respectively, while InceptionV3 shows inconsistent results (weighted F1-score: 0.6487 ± 0.4312), reflecting challenges in handling complex tumor morphology or class imbalances.

This evaluation underscores the strong performance of EfficientNet models, particularly EfficientNetB0, which combines high accuracy with remarkable stability across all tumor types. The results validate the utility of our dataset for developing reliable diagnostic tools, while the stark performance differences across architectures emphasize the importance of model selection in medical imaging tasks, where precision and consistency are critical. This work establishes a robust benchmark for brain tumor classification and provides a foundation for future research to explore diverse models and training protocols using this dataset.

Data availability

We introduce a new dataset, BRISC (Brain Tumor MRI Dataset for Segmentation and Classification), which is publicly available at Kaggle https://www.kaggle.com/datasets/briscdataset/brisc2025/.

Code availability

The custom code developed for the baseline models are publicly available at the BRISC dataset repository on Kaggle https://www.kaggle.com/datasets/briscdataset/brisc2025/.

Received: 1 September 2025; Accepted: 27 January 2026;

References

1. Ge, L.-K. *et al.* An open data for imaging acute aerobic exercise effects on brain and mind in emerging adulthood. *Scientific Data* **11**, 1422 (2024).
2. Usman Akbar, M., Larsson, M., Blystad, I. & Eklund, A. Brain tumor segmentation using synthetic MR images-A comparison of GANs and diffusion models. *Scientific Data* **11**, 259 (2024).
3. Sun, Z. *et al.* A low-field MRI dataset for spatiotemporal analysis of developing brain. *Scientific Data* **12**, 109 (2025).
4. Zhu, Q. *et al.* 7 Tesla multimodal MRI dataset of ex-vivo human brain. *Scientific Data* **12**, 845 (2025).
5. Gong, Z. *et al.* A multi-center, multi-parametric MRI dataset of primary and secondary brain tumors. *Scientific Data* **11**, 789 (2024).
6. Dorosti, S. *et al.* High-Resolution Ultrasound Data for AI-Based Segmentation in Mouse Brain Tumor. *Scientific Data* **12**, 1322 (2025).
7. Li, C. *et al.* DDEvENet: Evidence-based ensemble learning for uncertainty-aware brain parcellation using diffusion MRI. *Computerized Medical Imaging and Graphics* **120**, 102489 (2025).
8. Rezvani, S., Fateh, M., Jalali, Y. & Fateh, A. FusionLungNet: Multi-scale fusion convolution with refinement network for lung CT image segmentation. *Biomedical Signal Processing and Control* **107**, 107858 (2025).
9. Zhang, Q., Hang, Y., Qiu, J. & Chen, H. Application of u-net network utilizing multiattention gate for mri segmentation of brain tumors. *Journal of Computer Assisted Tomography* **48**, 991–997 (2024).
10. Askari, F., Fateh, A. & Mohammadi, M. R. Enhancing few-shot image classification through learnable multi-scale embedding and attention mechanisms. *Neural Networks* **187**, 107339 (2025).
11. Fateh, A., Mohammadi, M. R. & Motlagh, M. R. J. MSDNet: Multi-scale decoder for few-shot semantic segmentation via transformer-guided prototyping. *Image and Vision Computing* **162**, 105672, https://doi.org/10.1016/j.imavis.2025.105672 (2025).
12. Menze, B. H. *et al.* The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging* **34**, 1993–2024 (2014).
13. Jun, C. Brain tumor dataset. *figshare*, https://doi.org/10.6084/m9.figshare.1512427.v8 (2017).
14. Ghaffari, M., Sowmya, A. & Oliver, R. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 challenges. *IEEE reviews in biomedical engineering* **13**, 156–168 (2019).
15. LaBella, D. *et al.* Analysis of the BraTS 2023 Intracranial Meningioma Segmentation Challenge. *arXiv preprint arXiv:2405.09787*, (2024).
16. De Benedictis, S. G., Gargano, G. & Settembre, G. Enhanced MRI brain tumor detection and classification via topological data analysis and low-rank tensor decomposition. *Journal of Computational Mathematics and Data Science* **13**, 100103 (2024).
17. Sarkar, S., Singh, H. & Chawla, J. Evolution of Lung Tumor Segmentation: Comprehensive Analysis. *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 1–7 (2024).
18. Khajeha, H. R. *et al.* Advancing Glaucoma Diagnosis Through Multi-Scale Feature Extraction and Cross-Attention Mechanisms in Optical Coherence Tomography Images. *Engineering Reports* **7**, e70110 (2025).
19. Fakhima, M. S., Fateh, M., Fateh, A. & Jalalia, Y. DA-COVSGNet: Double Attentional Network for COVID Severity Grading. *International Journal of Engineering, Transactions A: Basics* **38**, 1568–82 (2025).
20. Rezvani, S., Siahkar, F. S., Rezvani, Y., Gharahbagh, A. A. & Abolghasemi, V. Single image denoising via a new lightweight learning-based model. *IEEE Access*, (2024).
21. Masoud, N. Brain Tumor MRI Dataset. *Kaggle Dataset*, https://doi.org/10.34740/KAGGLE/DSV/2645886 (2021).
22. Nguyen, V. A. AnyLabeling-Effortless data labeling with AI support. https://github.com/vietanhdev/anylabeling, (2024).
23. Fateh, A. *et al.* BRISC 2025: Brain Tumor MRI Dataset for Segmentation and Classification. *Kaggle*, https://doi.org/10.34740/kaggle/ds/7632487 (2025).
24. Sartaj, B., Kadam, A., Bhumkar, P. & Dedge, S. Brain Tumor Classification (MRI). *Kaggle Dataset*, https://doi.org/10.34740/KAGGLE/DSV/12745533 (2025).
25. Hamada, A. Br35H :: Brain Tumor Detection 2020. *Kaggle Dataset*, https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection (2020).
26. Cheng, J. *et al.* Interactive medical image segmentation: A benchmark dataset and baseline. *Proceedings of the Computer Vision and Pattern Recognition Conference.* 20841–20851 (2025).
27. Rezatofighi, H. *et al.* Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 658–666 (2019).
28. Goodfellow, I. Deep learning. *MIT Press*, (2016).
29. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
30. Dosovitskiy, A. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, (2020).

31. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241 (2015).
32. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. *Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4*, 3–11 (2018).
33. Chaurasia, A. & Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE visual communications and image processing (VCIP)*, 1–4 (2017).
34. Fan, T., Wang, G., Li, Y. & Wang, H. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access* **8**, 179656–179665 (2020).
35. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801–818 (2018).
36. Li, H., Xiong, P., An, J. & Wang, L. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180* (2018).
37. Li, C. & Jiao, G. EINet: camouflaged object detection with pyramid vision transformer. *Journal of Electronic Imaging* **31**, 053002–053002 (2022).
38. Patel, K., Bur, A. M. & Wang, G. Enhanced u-net: A feature enhancement network for polyp segmentation. *2021 18th conference on robots and vision (CRV)*, 181–188 (2021).
39. Li, J., He, W., Lu, F. & Zhang, H. Towards complex backgrounds: A unified difference-aware decoder for binary segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*. IEEE (2025).
40. Qin, X. *et al.* Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, (2021).
41. Saber, A. *et al.* Efficient and accurate pneumonia detection using a novel multi-scale transformer approach. *Sensors* **25**(23), 7233 (2025).
42. Rezvani, S., Fateh, M. & Khosravi, H. ABANet: Attention boundary-aware network for image segmentation. *Expert Systems* **41**, e13625 (2024).

Acknowledgements

We thank Fatemeh Gheisari, our radiologist, for her invaluable assistance with expert annotations and guidance throughout this study.

Author contributions

Amirreza Fateh wrote the original draft of the manuscript, supervised the research, and contributed to the methodology. Yasin Rezvani contributed to the dataset collection, curation, and methodology, and implemented the code. Sara Moayedi contributed to the dataset creation. Sadjad Rezvani contributed to the methodology and implementing the baselines. Fatemeh Fateh, as the consulting physician, supervised the dataset design and labeling and contributed to writing the manuscript. Mansoor Fateh supervised the research, reviewed the manuscript, and contributed to the methodology. Vahid Abolghasemi reviewed the manuscript and supervised the work. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.F. or V.A.