MDPI

Article

Self Supervised Deep Learning for Comprehensive Brain MRI Classification: A Comparative Study of CNN on 17 Tumor Subtypes

Abrar Hossain Zahin [1], Amit Kumar Saha [2] and Tanvir Mridha [2,*]

[1] Department of Computer Science and Engineering; abrarhossain1200@gmail.com; (2022-2-60-040)
[2] Department of Computer Science and Engineering; amitkumarsahaak3@gmail.com; (2021-3-60-044)
[3] Department of Computer Science and Engineering; tanvirmridha17540042@gmail.com; (2021-3-60-129)

Abstract: Classifying brain tumors using magnetic resonance imaging (MRI) is crucial for early diagnosis and treatment; however, tumor heterogeneity and a dearth of annotated datasets restrict the use of supervised deep learning approaches. In this work, we use self-supervised learning (SSL) to study multi-class brain tumor classification. Using a ResNet-50 backbone, we evaluate four SSL frameworks (SimCLR, BYOL, DINO, and Moco v3) on a publicly available dataset of 4,448 MRIs with 17 distinct tumor types. On the dataset, SimCLR achieved **99.64%** accuracy, **99.64%** precision, **99.64%** recall and **99.64%** F1-score. The workflow includes preprocessing, fine-tuning, linear evaluation, and SSL pretraining with data augmentations. Results show that, especially when labels are limited, SSL-pretrained models outperform supervised baselines in terms of F1-score, recall, accuracy, and precision. Additionally, by providing visual insights into model decisions, Explainable AI techniques (Grad-CAM, Grad-CAM++, Eigen-CAM) enhance interpretability. These results demonstrate SSL's scalability and dependability in diagnosing brain tumors from unlabeled medical data.

Keywords: Brain tumor classification; Magnetic resonance imaging (MRI); Self-supervised learning (SSL); SimCLR; BYOL; DINO; Moco v3; ResNet50; Explainable AI (XAI); Grad-CAM; Medical Image Analysis.)

1. Introduction

Brain tumors are considered some of the most complex and deadliest neurological diseases, and they contribute substantially to the global burden of cancer morbidity and mortality. Global cancer statistics indicate the prevalence of primary brain tumors has been rising and early detection is necessary for enhancing patients' survival rates. MRI is the modality of choice for the detection and differentiation of brain tumors because it can provide detailed anatomical information with (relatively) high resolution and without ionizing radiation [1]. MRI includes various tissue contrasts such as T1, T2 and T1C+, which represent key information for tumor type, location and stage. But interpreting MRI images is still extremely challenging for radiologists, in spite of its diagnostic power.Tumors usually present heterogeneous features with irregular boundaries and overlapping characteristics within or between other categories, which is time-consuming and may lead to errors in manual diagnosis and interobserver variation [2]. This complexity has encouraged the use of artificial intelligent (AI) techniques, particularly deep learning (DL) for automatic brain tumor classification and enhancing the diagnosis accuracy.

Convolutional Neural Networks (CNNs) have made significant advancement in computer vision tasks, including medical image analysis in recent years.CNNs can automatically

Received:
Revised:
Accepted:
Published:

Citation: Lastname, F.; Lastname, F.; Lastname, F. Self-Supervised Deep Learning for Brain MRI Classification. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

learn hierarchical feature representations from raw data, outperforming conventional methods based on handcrafted features[3]. There are also a few studies showing that supervised CNNs are efficient in the classification of brain tumors and clinical outcomes prediction. Yet such models are inherently limited by the requirement for large-scale annotated data. In medical industry, obtaining trustworthy labels is costly, time-consuming and specialized radiologists are to be consulted for the labeling. In addition, some cancer types are uncommon, and available datasets tend to suffer from class imbalance problems[4]. These limitations severely constrained the applicability of supervised models in clinical use. As a result, there is an increasing demand for learning paradigms capable of harnessing the large number of available unlabeled medical images, as to circumvent reliance on expensive annotations.

Self-supervised learning (SSL) is a delightful new paradigm that has really developed as a way to get around this.Unlike supervised learning, SSL takes advantage of the availability of unlabeled data by defining pretext tasks that force the model to learn strong and semantically relevant feature representations[5]. These representations can be fine-tuned with small amounts of labeled data for downstream tasksnlp progress. SSL has made great advances on image level representation learning, and methods such as SimCLR, BYOL, and DINO have reported the state-of-the-art results on standard benchmarks. These approaches learn invariant representations through powerful data augmentations and capture rich semantic structures without relying on explicit labels. Despite great success of SSL in computer vision, its utilization for complicated medical imaging tasks (e.g., brain tumor classification with multiple categories) is still underexplored[6]. Throughout this paper we investigate whether this type of SSL is transferrable in this domain, and establish its effectiveness as a potential step forward in automated diagnosis and clinical decision support.

In this work, we seek to investigate the efficacy of SSL based methods for multi-class brain tumor classification in MRI images. We use ResNet-50 network as an encoder and perform pretraining on large-scale set of MRI images using three popular SSL frameworks: SimCLR, BYOL, and DINO. The workflow of the study involves a systematic approach including the preprocess of the dataset and data augmentation, the approach for SSL-based feature learning, and linear evaluation, and fine-tuning for the classification task in 17 tumor types [7][8]. To enhancemodel interpretability and clinical trust, explainable AI (XAI) approaches such as Grad-CAM, Grad-CAM++, and Eigen-CAM are incorporated into pipeline by providing visual justifications of model predictions. With the combined approach, the proposed framework not only seeks to maximize classification accuracy, but also allows transparency, which is important for real-life application in the clinic.

Our experiments on a publicly available MRI dataset containing 4,448 images show that SSL-pretrained models have the ability to learn discriminative and semantically meaningful feature representations[9][10]. Compared with conventional supervised methods, SSLbased methods are able to obtain better performance, especially when only a small amount of labeled data is available. In addition, the visualizations of the learned representations based on t-SNE validate that SSL models produce very well-structured feature spaces, which could help improve the separation of tumor classes. The incorporation of interpretability methods also underlines the clinical relevance of the method, connecting AI based predictions and clinical decision-making.

In this study adds to the body of international research on self-supervised learning in medical images by performing extensive studies of SSL approaches for multi-class brain tumor classification[11]. By systematically comparing SimCLR, BYOL, and DINO under a common ResNet-50 backbone, we showcase SSL's promise to alleviate data paucity, boost classification accuracy, and interpret better the model. These results indicate that SSL can be

a scalable, dependable method for brain tumor diagnosis, assisting radiologists in making rapid and precise clinical decisions.

2. Related Work

Deep learning has improved the analysis of fMRI-derived brain networks have advanced from convolutional neural networks (e.g., BrainNetCNN) through graph neural networks (e.g., BrainGNN) to Transformer architectures, yet these task-specific models remain constrained by limited annotated samples and poor cross-task generalizability. Self-supervised learning (SSL) has helped reduce data needs in medical imaging ranging from X-rays to retinal scans through foundation models that leverage unlabeled dataset. However, SSL efforts tailored to brain networks like BrainNPT and BrainGSLs, have shown only modest gains over non-SSL approaches, largely due to insufficient data scale and model depth. This gap need for a large-scale self-supervised foundation model explicitly designed for brain network representation and capable of few- and zero-shot learning across diverse neuroimaging tasks and may face challenges when applied to small or highly heterogeneous datasets [12].

Fang et al [13]. propose a self-supervised multi-modal hybrid fusion network for brain tumor segmentation, addressing the shortage of limited annotated medical data and the need for effective multi-modal feature integration. The method uses separate encoders for independent feature extraction from T1, T1CE, T2, and FLAIR MRI sequences. These features are then combined using Hybrid Attentional Fusion Block (HAFB), which applies summation, product, and maximum operations with attention to capture complementary information. A self-supervised pretext task masks regions in one modality, training the network to recover missing information from others. This improves robustness and reduces overfitting. Experiments on the BraTS 2019 dataset show superior Dice scores, specificity, and sensitivity compared to U-Net and other baselines. This approach demonstrates that hybrid fusion with self-supervision can significantly improve segmentation accuracy, especially when multi-modal MRI data are available and it requires access to all MRI modalities, which may not be available in certain clinical scenarios.

Zhang et al [10]. propose SSCLNet, a self-supervised pre-trained network for brain MRI classification that uses contrastive loss to improve feature learning when labeled data are limited. In this method, the encoder is first pre-trained with contrastive self-supervised task, where augmented version of the same image are brought closer in feature space and different images are pushed apart. This process helps the model learn stable and discriminative features without manual labeling. The pre-trained encoder is then fine-tuned for brain disease classification. Experiments on multiple MRI datasets show that SSCLNet achieves higher accuracy, precision, recall, and F1-score than supervised method, especially when labeled data are scarce. The results highlight that self-supervised contrastive pre-training can significantly improve classification performance in medical imaging, offering a robust solution for label-limited scenarios. Contrastive pre-training relies heavily on effective data augmentation, which may be less suitable for capturing fine-grained medical image features.

Magnetic resonance imaging (MRI) segmentation has advanced through deep learning methods that work with minimal annotated data. Liu et al [5]. review self-supervised learning, few-shot learning, generative models and semi-supervised techniques for MRI segmentation under data limitations. They explain SSL fundamentals, including contrastive learning, image masking, and highlight their applications in abdominal, cardiac, and brain MRI segmentation. The analysis also surveys generative adversarial networks and diffusion models for artificial data generation. Moreover, they discuss few-shot techniques like prototypical networks that chosen only a few labeled examples. At last, semi-supervised

strategies that combine labeled and unlabeled data are presented. The paper summarizes key datasets and provides clear guidance on method choice based on label availability. Its clear comparisons help researchers select scalable self-supervised approaches for challenging neuroimaging tasks but does not provide direct experimental benchmarks, making it harder to evaluate methods under a unified setting.

Noninvasive molecular subtyping of pediatric low-grade gliomas can improve treatment planning. Tak et al. present a two-stage deep learning pipeline for this task. The first stage automatically segments tumors on T2-weighted MRI scans. Next, it applies three binary classifiers to identify BRAF fusion, BRAF wild-type, and V600E mutations. The authors propose TransferX, which combines in-domain transfer learning with self-supervised cross-training to increase feature extraction when training data are limited . They validate the model on achieving AUCs of 0.82–0.87 internally and 0.72–0.78 externally, demonstrating robust generalizability. To increase transparency, they develop COMDist, a tool that measures how much the model focuses on tumor regions. This end-to-end approach requires no manual segmentation or handcrafted features, making it a promising tool for clinical decision support in resource-limited settings and the model performance drops in external validation, suggesting limited generalizability across different imaging centers and protocols [4].

Meng at al [14]. propose brain tumor MRI images for automatic segmentation using self-supervised contrastive learning in schizophrenia patients (BTCSSSP). First, a denoising algorithm based on progressive principal component analysis approximation and adaptive clustering is designed to process the noisy MRI images. Second, a brightness-aware image enhancement algorithm is developed to address the problems of non-uniformity, unclear boundaries, and poor spatial resolution of the MRI images. Finally, a cross-scale U-Net network with selective feature fusion attention module is designed based on self-supervised contrastive learning to achieve automatic segmentation of brain tumor MRI images. The BTCSSSP method estimates the Gaussian noise level in the images using standard PCA and extracts image features using adaptive clustering and Multi-Scale Context Feature Blocks (MSCFB) technology. The OpenfMRI dataset and FBIRN Dataset, were used as experimental datasets.The experimental dataset is automatically segmented using a cross-scale connected U-shaped network with a selective feature fusion attention module. Accurate segmentation of brain tumors in patients with schizophrenia is crucial for treatment and monitoring purposes.

A brain tumor is an abnormal mass of cells in the brain that can develop at any stage of life. These cells divide and grow uncontrollably, occupying the limited and enclosed space of the brain or invading normal brain tissue, leading to various symptoms.This study aims to evaluate the feasibility of training a deep neural network for the segmentation and detection of metastatic brain tumors in MRI using a very small dataset of 33 cases, by leveraging large public datasets of primary tumors. In this methods This study explores various methods, including supervised learning, two transfer learning approaches, and self-supervised learning, utilizing U-net and Swin UNETR models. The BraTS 2021 dataset includes 1251 multimodal MRI cases of primary brain tumors. During the initial research stage, we explored various models, including convolutional neural networks, conventional au-toencoders, and ResNet-based networks.It is feasible to train a model using self-supervised learning and a small dataset for the segmentation and detection of small brain tumors [7].

Brain tumors encompass a diverse group of neoplasms originating in human brain tissue, with varying degrees of aggressiveness and fatality. In this study, we employ coherent Raman scattering imaging method and a self-supervised deep learning model (VQVAE2) to enhance the speed of SRH image acquisition and feature representation,

thereby enhancing the capability of automated real-time bedside diagnosis. In this methods We adopted the latest version of the Vector Quantized Variational Autoencoder (VQ-VAE2), known as VQ-VAE2, to reconstruct SRS images acquired by InSight DeepSee, preserving spatial chemical information.VQ-VAE is a powerful self-supervised feature learning model that encodes images into quantized latent representations — vectors, which are then decoded to reconstruct the input image.As an emerging molecular vibration-based microscopy technique, SRS imaging has experienced rapid growth in the fields of biology and medicine [6].

Artificial intelligence aids in brain tumor detection via MRI scans, enhancing the accuracy and reducing the workload of medical professionals. Chin at al. introduces a novel two-stage anomaly detection algorithm called CONSULT (CONtrastive Self-sUpervised Learning for few-shot Tumor detection). The first stage of CONSULT fine-tunes a pre-trained feature extractor specifically for MRI brain images, using a synthetic data generation pipeline to create tumor-like data.The second stage of CONSULT uses PatchCore for conventional feature extraction via the fine-tuned weights from the first stage. Anomaly detection is effective for medical images because it relies solely on healthy images for training, eliminating the need for labeled data. CONSULT addresses this by combining the strengths ofdownstream and reconstruction-based anomaly detection algorithms [9].

Brain tumor is generally diagnosed by a specialist called a neurologist. The First limitation observed in the medical images diagnostics is that of manual interpretation.The second limitation observed in the medical image diagnosis is that of the inherent noise present in the image data due to various factors such as quantization, encoding error, channel noise (such as Additive White Gaussian Noise, AWGN), etc. The proposed structure uses the kmeans algorithm for successful segmentation and The Gray Level Statistical Analysis (GLCM) based feature extraction. The Linear Discriminate Analysis (LDA) based classification is used for classifying brain tumor of type benign with that of malignant. The overall propose method considers the identification and classification of brain tumor region [11].

Magnetic resonance imaging (MRI) is developed to generate high-quality images and provide extensive medical research information. MRI provides extensive information for medical diagnosis and research (Zhang et al., 2011). Two categories of research have been proposed. First is unsupervised classification, such as fuzzy c-means and selforganization feature maps (Ibrahim et al., 2013). Second is supervised classification, such as K Nearest Neighbours (KNN) and Support Vector Machine (SVM) (Cocosco et al., 2003; Chaplot et al., 2006). The dataset we used for the research is REMBRANDT (Scarpace et al., 2022). It is accessed from The Cancer Imaging Archive (TCIA) database (Clark et al., 2013). Machine learning algorithm classification comparison. Supervised machine learning algorithms applied classification methods, such as Decision Tree (DT), SVM, KNN and NN have been compared to estimate the performance for each training model. This classification model can be used in other features of brain tumors MRI to obtain the most accurate result [2].

Medical data often has limited availability due to labor-intensive data collection. SelfSupervised Learning (SSL), known for its robustness in addressing data imbalances, is frequently employed in Transfer Learning (TL). SSL, as proposed, enhances feature representation in medical data, improving performance in tasks like diagnosis and medical image processing. Numerous studies enhance brain tumor classification using deep learning models, particularly CNNs like VGG-16, VGG-19, DenseNet121, DenseNet201, GoogleNet, ResNet50, and Inception-v3, employing pre-trained models and TL strategies for improved generalization with limited labeled data. CNN techniques for brain tumor classification use pre-trained models and TL to overcome data limitations, boost gener-

alization, and achieve accuracy. VGG-19, EfficientNetB0, VGG-16, and AlexNet variants consistently perform well in various studies [15].

Yousun at al [8]. propose a self-supervised machine learning (ML) algorithm for sequence-type classification of brain MRI using a supervisory signal from DICOM metadata (i.e., a rule-based virtual label). Brain Magnetic Resonance Imaging (MRI) stands as an indispensable tool in the routine diagnosis and ongoing monitoring of individuals affected by various brain-related conditions, encompassing the realm of Brain-Computer Interface (BCI) technology. We created a rule-based labeling system using the metadata of DICOM image files and developed a sustainable self-supervised ML algorithm, named ImageSortnet, for automatic sequence-type classification of brain MRI using supervisory signals from DICOM metadata (i.e., rule-based virtual labeling). The training of ML algorithms utilizing rule-based virtual labels has yielded high accuracy for sequence-type classification of brain MRI. This approach has empowered us to construct a robust, self-learning system with sustained efficacy.

Classification of brain tumors from MRI images is crucial for early diagnosis and effective treatment planning. In this study, we explored the use of self-supervised learning techniques to improve the classification performance for brain tumors. Specifically, we tested three SSL approaches SimCLR, Moco, and BYOL, with ResNet-50 as the backbone architecture on a newly constructed dataset created by combining five public datasets.The proposed method is based on three self-supervised approaches SimCLR, MoCo, and BYOL. It integrates intense-level contrastive-positive pair learning for the pre-training of the network with frozen and unfrozen layers. The proposed T3SSLNet framework is structured into four blocks- the Neuro Imaging Spectrum Enhancement (NISE) Block, the Frozen Feature Extractor(FFE) Block, the Neural Representation Projection Learning (NRPL) Block, and the Unfrozen Classification (UCL) Block. Although SSL offers a strong basis, fine-tuning plays a crucial role in improving the capacity of the model to correctly categorize brain tumors [3].

A SimCLR-based model is proposed for the classification of unlabeled brain tumor images in medical imaging using a self-supervised learning (SSL) technique. Additionally, the performances of different SSL techniques (Barlow Twins, NnCLR, and SimCLR) are analyzed to evaluate the performance of the proposed model. Three different datasets, consisting of pituitary, meningioma, and glioma brain tumors as well as non-tumor images, were used as the dataset. SimCLR fundamentally utilizes twin networks. Three datasets containing brain tumor images were used, specifically comprising "pituitary," "meningioma," and "glioma" types, along with non-tumor images. A SimCLR-based model was proposed to enhance the classification performance of unlabeled images [1].

3. Methodology

3.1. Work Flow

In the present research study, the step-by-step workflow shown in Figure 1 represents a sequential pipeline that begins with the acquisition and preparation of Brain Tumor MRI Images that are part of (17 classes) For our study, the images were resized to a uniform resolution, normalized, and split into 80% training, 10% validation, and 10% test parts. To enhance model robustness, we applied different SSL augmentations for different SSL algorithms. This dataset is particularly suitable for evaluating self-supervised learning methods in medical image to make proper classification. Datasets $\rightarrow$ preprocessing $\rightarrow$ model training $\rightarrow$ linear eval $\rightarrow$ fine-tune with SSL models to evaluate metrics which consists of accuracy, precision, recall and F1-score. Lastly, in order to improve model understandability, Explainable AI methods such as Grad- CAM, Grad-CAM++, Eigen-CAM are used.

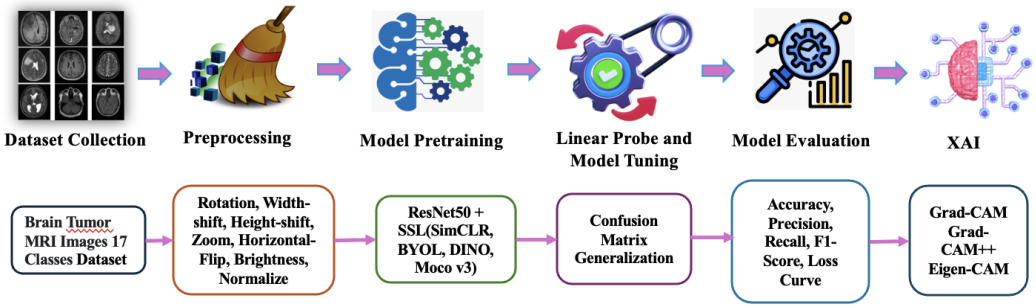

Figure 1. Workflow Diagram

3.2. Dataset Description

For this research we utilize Brain Tumor MRI Images (17 Classes) openly available kaggle data. The data contains MRI scans that are clustered into 17 types of tumor classes, offering a wide specification of a multi-class classification of tumors. The distribution of the dataset over varying classes of tumors is summarized in Table 1. The data exhibit class real world imbalance medical dataset, such that there are significantly fewer samples of certain types of tumors relative to others.

Table 1. Summary of the Brain Tumor MRI Images (17 Classes) dataset.

Tumor Class	**Amount**
Glioma (Astrocitoma, Ganglioglioma, Glioblastoma, Oligodendroglioma, Ependimoma) T1C+	508
Schwannoma (Acustico, Vestibular - Trigeminal) T1	153
Outros Tipos de Lesões (Abscessos, Cistos, Encefalopatias Diversas) T2	57
Neurocitoma (Central - Intraventricular, Extraventricular) T2	112
Outros Tipos de Lesões (Abscessos, Cistos, Encefalopatias Diversas) T1C+	48
Meningioma (de Baixo Grau, Atípico, Anaplásico, Transicional) T1	345
Neurocitoma (Central - Intraventricular, Extraventricular) T1	169
Schwannoma (Acustico, Vestibular - Trigeminal) T1C+	194
Schwannoma (Acustico, Vestibular - Trigeminal) T2	123
Meningioma (de Baixo Grau, Atípico, Anaplásico, Transicional) T1C+	625
Meningioma (de Baixo Grau, Atípico, Anaplásico, Transicional) T2	329
Outros Tipos de Lesões (Abscessos, Cistos, Encefalopatias Diversas) T1	152
NORMAL T1	272
Glioma (Astrocitoma, Ganglioglioma, Glioblastoma, Oligodendroglioma, Ependimoma) T1	430
Neurocitoma (Central - Intraventricular, Extraventricular) T1C+	261
NORMAL T2	291
Glioma (Astrocitoma, Ganglioglioma, Glioblastoma, Oligodendroglioma, Ependimoma) T2	346
Total Number of Image	4448

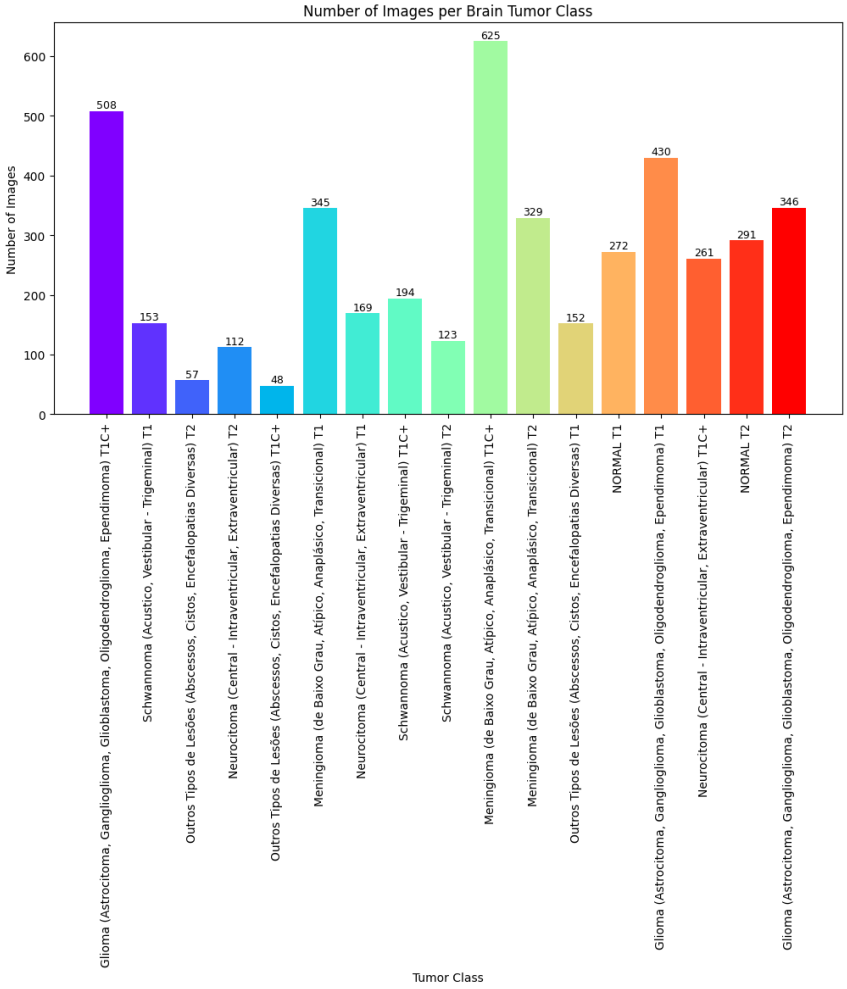

Figure 2. Number of images per class in the Brain Tumor MRI Images (17 Classes) dataset.

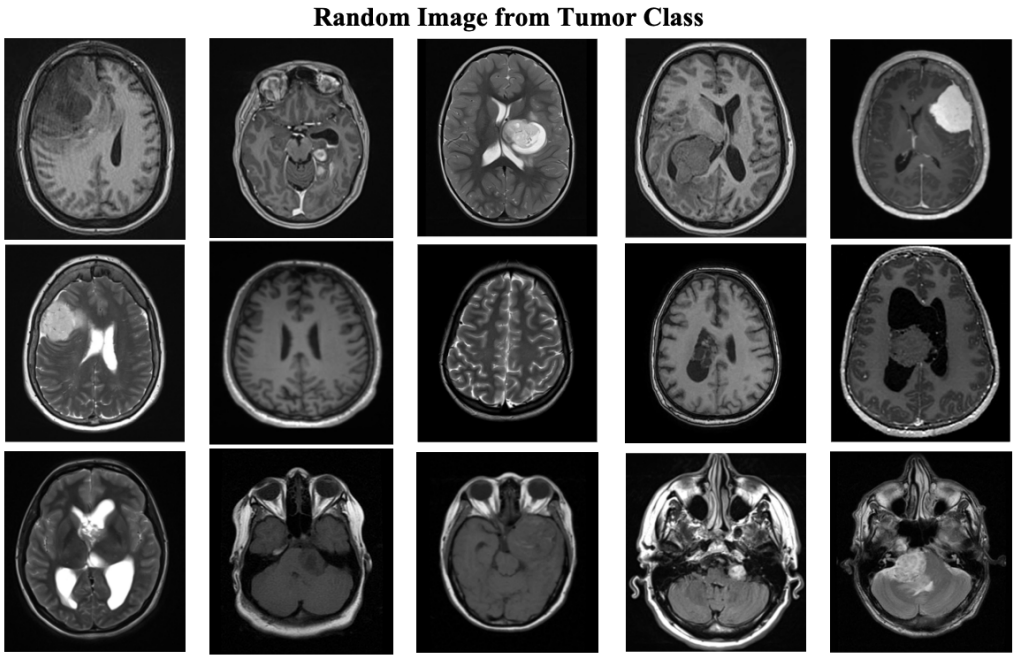

Figure 3. Visualize Random Image for Tumor class

3.3. Dataset Preprocessing

For self-supervised learning, strong data augmentations. In Table 2 random resizing and cropping, flipping, and rotation the images horizontally, and adding Gaussian blur to help the novel learn robust features.

Table 2. Image Data Augmentation Parameters are used

Parameter	**Value**
Rotation Range	Random rotation within $\pm 20°$
Width Shift Range	Random horizontal shift up to 10% of total width
Height Shift Range	Random vertical shift up to 10% of total height
Zoom Range	Random zoom within $\pm 10\%$
Horizontal Flip	Random left-right flipping applied
Brightness Range	Adjust brightness within range $[0.8, 1.2]$
Fill Mode	Nearest-neighbor interpolation

After applying the above data augmentation techniques Figure 4 illustrates the size of the dataset was increased and balanced. Concretely, each category was augmented to a normalized value of 625 images. The above operation not only augmented the aggregate size of the dataset but also reduced the imbalance that was inherently present across the various categories. The dataset was partitioned into a train/validation/test set after augmentation and consisted of 8,500/1,062/1,063 images. The growing and balancing at this point enhanced the reliability of the dataset and strengthened the generalizability capability of the ensuing models in each category of tumors.

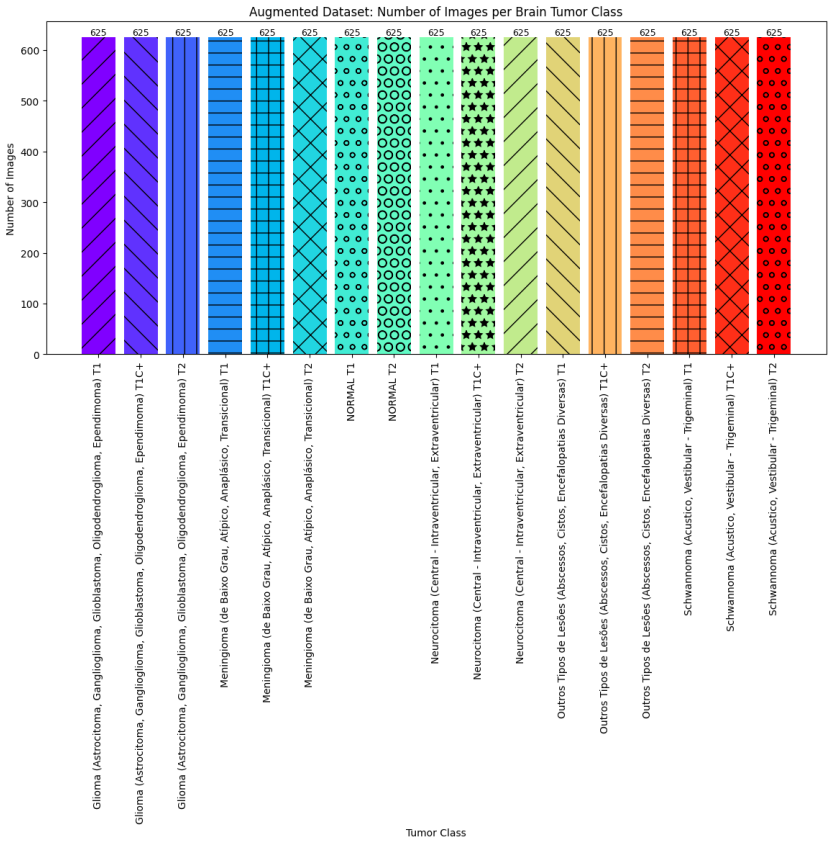

Figure 4. Augmented Dataset: Number of images per Brain Tumor class

When the data augmentation techniques were applied to the set, the set was augmented with abundant variations within each class. Figure 5 presents a visualization of the randomly selected augmented samples from each category. Not only does this increase the effective size of the dataset but also demonstrates enriched diversity achieved over all classes, which is crucial for boosting model generalization and robustness.

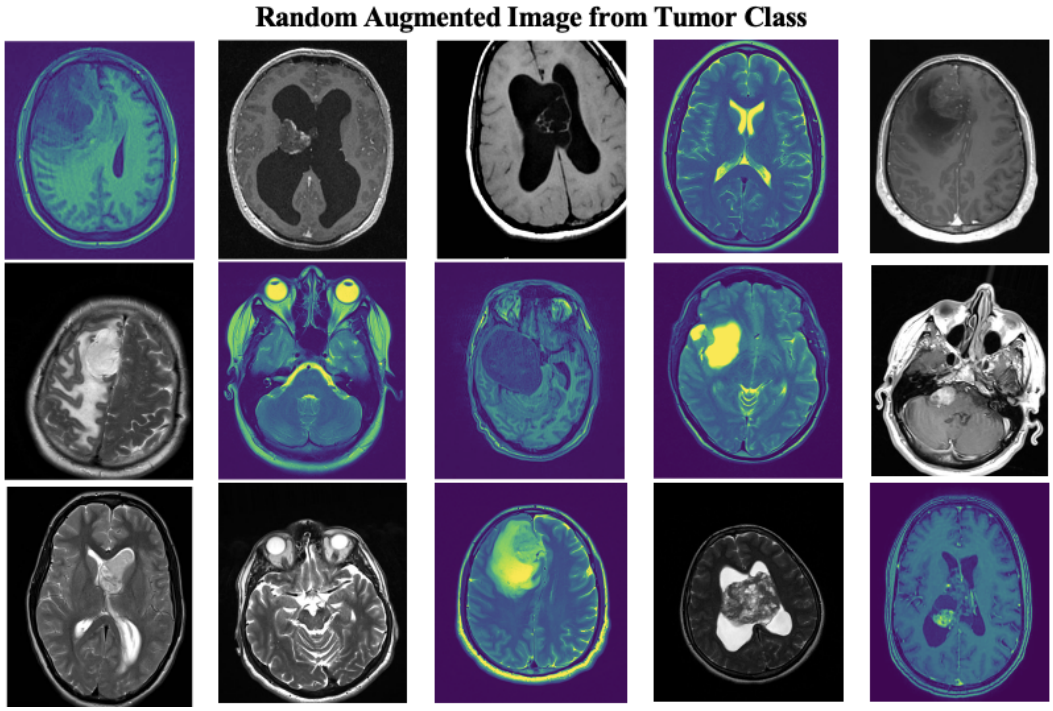

Figure 5. Visualize Random Augmented Image for Tumor class

3.4. Unsupervised Learning

t-SNE (t-distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique that projects high-dimensional data in 2D or 3D space. t-SNE captures the local structure of the data such that points that are close together in high-dimensional space are also close in the low-dimensional plot. This is especially helpful for revealing patterns, natural clusters, or groupings in high-dimensional data. The close resemblance between clusters in both plots shows that the SSL and fine-tuning captured meaningful class-discriminative information. It confirms that the model has learned a well-structured and semantically meaningful representation space.

3.5. Models and Implementation

3.5.1. ResNet50

ResNet50 is an advanced deep convolutional neural network that applies the idea of residual learning to tackle the degradation problem commonly faced by very deep networks. The network consists of 50 layers and is mainly composed of residual units allowing the direct skipping of gradients through one or many layers by means of identity shortcuts. Each one of the layers in a residual unit consists of a series of three convolutions in the following order: a starting 1×1 convolution for dimension reduction, a 3×3 convolution for spatial feature extraction, and a final 1×1 convolution for dimension restoration. The network starts with a convolutional stem followed by four different stages, each having multiple residual units with increasingly bigger filter sizes. Downsample operations

are performed using stride-2 convolutions. This comes to highlight the both depth and computational efficiency and has been exceptionally successful across a variety of image classification tasks. The architecture of the ResNet50 model has been shown in Figure 6 and is a simple backbone block for self-supervised learning.

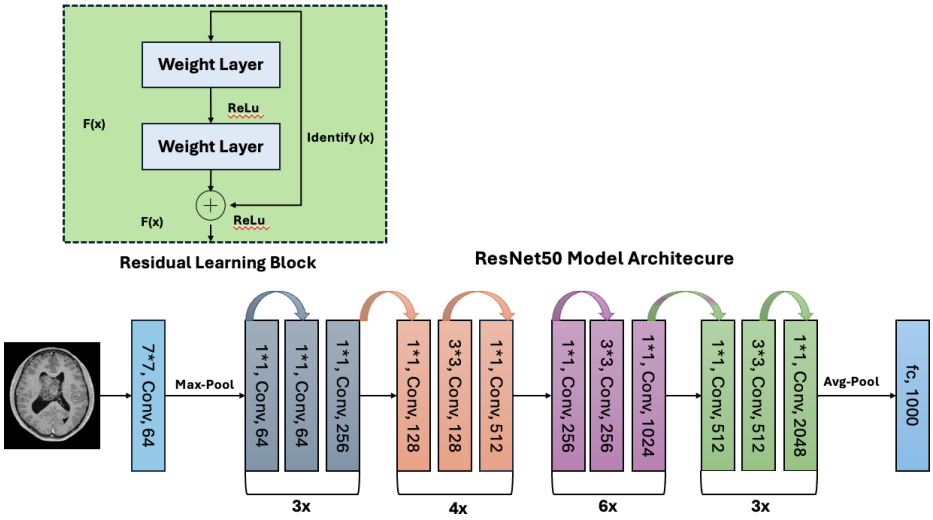

Figure 6. ResNet50 Architecture Diagram

$$\mathbf{y}_l = \mathcal{F}(\mathbf{x}_l, \{W_l\}) + \mathbf{x}_l, \tag{1}$$

$$\mathbf{x}_{l+1} = f(\mathbf{y}_l), \tag{2}$$

where $\mathbf{x}_l$ and $\mathbf{x}_{l+1}$ are the input and output feature maps of the l-th residual block, $\mathcal{F}(\cdot)$ is a residual function that is realized by stacked convolutional layers with learnable weights $\{W_l\}$, and $f(\cdot)$ is a one-dimensional non-linear activation function (e.g., ReLU). The ResNet50 model is formed by sequentially connecting 49 such residual blocks and a terminal fully connected classification layer:

$$\hat{\mathbf{y}} = \text{softmax}\big(W_{fc} \cdot \mathbf{z} + b_{fc}\big), \tag{3}$$

Here, $\mathbf{z}$ is the global average pooled feature vector that results from the end convolutional block, W_{fc} and b_{fc} are the weight and bias of the fully connected layer, and $\hat{\mathbf{y}}$ is the learned probability distribution across the various classes.

3.5.2. SimCLR Self-Supervised Learning Objective

We employ a standard ResNet-50 encoder for images with the final fully connected classification layer turned off. The encoder for an input image of size 224×224 produces a pooled global feature vector of 2048 size. During self-supervised pretraining, a light-weight projection head (contrastive head) is appended to the encoder: a 2-layer MLP

$$\text{Linear}(2048 \to 512) \to \text{ReLU} \to \text{Linear}(512 \to 128)$$

Let $\mathcal{X}$ be the dataset and let T denote the stochastic data-augmentation distribution used in the notebook (random crop, color jitter, flip, ...). For each sample $x \in \mathcal{X}$ we draw two independent views

$$\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)} \sim T(\cdot \mid x_i).$$

A neural encoder $f_\theta : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{d_h}$ (ResNet backbone) maps an image to a representation h, and a projection head $g_\phi : \mathbb{R}^{d_h} \to \mathbb{R}^{d_z}$ maps to the contrastive space:

$$h_i^{(a)} = f_\theta\big(\tilde{x}_i^{(a)}\big), \qquad z_i^{(a)} = g_\phi\big(h_i^{(a)}\big), \qquad a \in \{1, 2\}.$$

We normalize projection vectors to unit length:

$$\bar{z} \equiv \frac{z}{\|z\|_2}.$$

Define the cosine similarity between two vectors u, v by

$$\mathrm{sim}(u, v) = u^\top v \quad \text{(when } u, v \text{ are already normalized this equals cosine similarity).}$$

Given a mini-batch of N original data, we obtain $2N$ samples after augmentation. For a postive pair (i, a) and (i, b) (different views for the same sample i), normalized temperature-scaled contrastive loss (NT-Xent) for anchor i, a is

$$\ell_i^{(a)} \;=\; -\log \frac{\exp\big(\mathrm{sim}\big(\bar{z}_i^{(a)}, \bar{z}_i^{(b)}\big)/\tau\big)}{\displaystyle\sum_{\substack{(k,c)=1 \\ (k,c)\neq(i,a)}}^{N,2} \exp\big(\mathrm{sim}\big(\bar{z}_i^{(a)}, \bar{z}_k^{(c)}\big)/\tau\big)},$$

where $\tau > 0$ is the temperature hyperparameter (in your notebook, you see that $\tau = 0.5$), and the denominator sums over all $2N - 1$ extra augmented instances (including the duplicate for the positive sample if $(k, c) = (i, b)$).

The average loss over the whole batch for all $2N$ anchors (or on average for the N positive pairs and both views):

$$\mathcal{L}_{\text{SimCLR}} \;=\; \frac{1}{2N} \sum_{i=1}^{N} \big(\ell_i^{(1)} + \ell_i^{(2)}\big).$$

Matrix form Let $Z \in \mathbb{R}^{2N\cdot d_z}$ be the row-wise stacked matrix of *normalized* projections $\bar{z}$. Define the (unnormalized) similarity matrix $S \in \mathbb{R}^{2N\cdot 2N}$ by

$$S = \frac{1}{\tau} Z Z^\top.$$

To compute NT-Xent we exponentiate S, mask out self-similarities on the diagonal, and for each row i compute:

$$\ell_i = -\log \frac{\exp(S_{i,j(i)})}{\sum_{k\neq i} \exp(S_{i,k})},$$

where $j(i)$ denotes index of the positive example for anchor i.

Figure 7 overview of the SimCLR model architecture for self-supervised representation learning, consisting the encoder backbone, projection head, and contrastive learning framework.

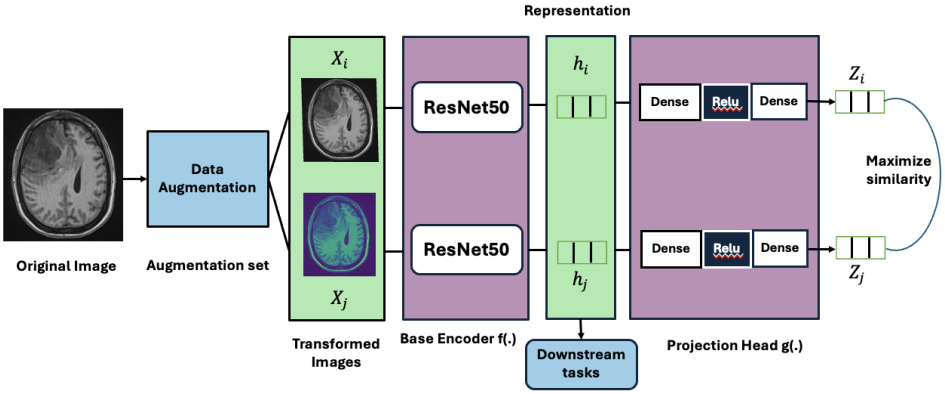

Figure 7. SimCLR self-supervised learning architecture

3.5.3. BYOL Self-Supervised Learning Objective

Removing the last fully-connected classification layer, we use a usual ResNet-50 backbone for our image encoder. The encoder receives a 224×224 input image and outputs a pooled global feature vector with 2048 dimension. The encoder, in the BYOL paradigm, constitutes a part of the online network that is augmented with a lightweight projection head and a prediction head. The projection head consists of a two-layer MLP

$$\text{Linear}(2048 \to 4096) \;\to\; \text{BatchNorm} \;\to\; \text{ReLU} \;\to\; \text{Linear}(4096 \to 256)$$

which projects the encoder output to a latent representation space. The prediction head is a second two-layer multilayer perceptron that projects the projection to a prediction vector with the same dimensionality (256).

Given an image $x \sim \mathcal{D}$, randomly sample two stochastic augmentation $t_1, t_2 \sim \mathcal{T}$ and form two views $v_1 = t_1(x)$ and $v_2 = t_2(x)$. BYOL operates a network that possesses *online* parameters. $\theta = \{f_\theta, g_\theta, q_\theta\}$ and a target network with parameters $\xi = \{f_\xi, g_\xi\}$. The f encoders map an input to a representation, the projectors g project a representation to a latent embedding, Further, the predictor q falls inside the online part.
Forward mappings:

$$y_1 = g_\theta(f_\theta(v_1)), \qquad y_2 = g_\theta(f_\theta(v_2)), \tag{4}$$
$$z_1 = g_\xi(f_\xi(v_1)), \qquad z_2 = g_\xi(f_\xi(v_2)). \tag{5}$$

Let $\text{sg}(\cdot)$ denote the stop-gradient operator and $\text{norm}(u) = \frac{u}{\|u\|_2}$ denote ℓ_2 normalization. Define normalized embeddings and predictions

$$\bar{y}_i = \text{norm}(y_i), \qquad \bar{z}_i = \text{norm}(z_i), \tag{6}$$
$$p_i = q_\theta(\bar{y}_i), \qquad \bar{p}_i = \text{norm}(p_i), \quad i \in \{1, 2\}. \tag{7}$$

Asymmetric regression loss (one direction):

$$\ell(v_1, v_2) = \|\bar{p}_1 - \text{sg}(\bar{z}_2)\|_2^2 = 2 - 2\,\langle \bar{p}_1,\ \text{sg}(\bar{z}_2)\rangle. \tag{8}$$

Symmetric BYOL loss (per sample):

$$\mathcal{L}_{\text{BYOL}}(x) = \ell(v_1, v_2) + \ell(v_2, v_1), \tag{9}$$

Target network update (EMA): of the network parameters of the online network, with momentum coefficient m (typical $m = 0.996$).Update target parameters after every optimization step on θ,

$$\xi \leftarrow \tau\,\xi + (1-\tau)\,\theta, \tag{10}$$

applied piecewise to $\{f, g\}$, with momentum coefficient $\tau \in [0, 1)$ (potentially time-varying). Figure 8 shows BYOL self-supervised learning model consisting of an online, target encoder and projection-prediction networks for description learning in brain tumor MRI classification.

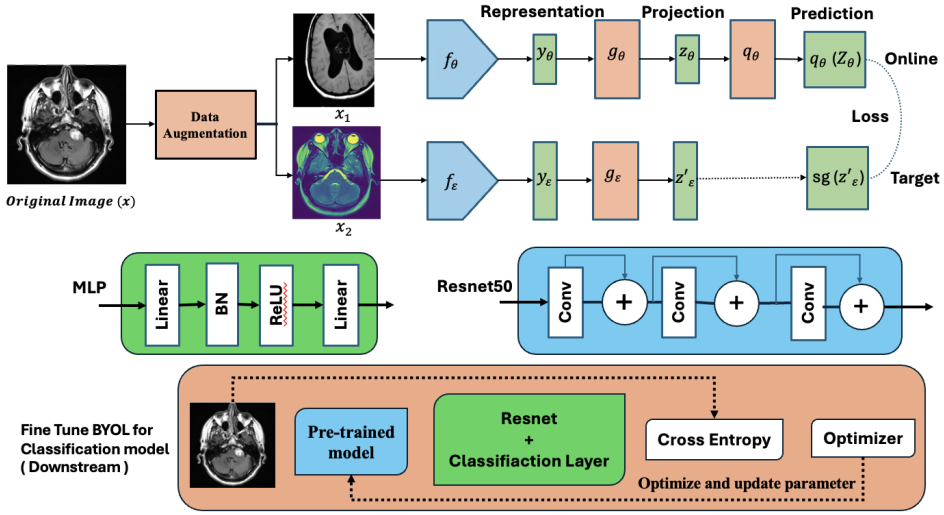

Figure 8. BYOL self-supervised learning architecture

3.5.4. DINO Self-Supervised Learning Objective

We employ a standard ResNet-50 backbone for the encoder in Figure 9, from which we eliminate the final fully-connected classification layer. The encoder receives as input an image size 224 × 224 and produces a pooled global feature vector (size 2048 for ResNet-50). During self-supervised pretraining, we append a light-weight projection head to the encoder: a 3-layer MLP with hidden dimension 2048, output dimension 256, and a final softmax normalization. The projection outputs are temperature-scaled before being passed for use for self-distillation.

$$\text{Head: Linear}(2048 \to 2048) \;\to\; \text{GELU} \;\to\; \text{Linear}(2048 \to 256) \;\to\; \text{Softmax}(\tau \cdot)$$

Let an input image x be augmented into a series of crops $\mathcal{V} = \{v_1, \ldots, v_{2+K}\}$, consisting of 2 global crops and K local crops. The distributed student network has θ, and the target teacher has ϕ and gets updated by an exponential moving average (EMA) on the student. They both consist of a backbone encoder $f(\cdot)$ and a projection head $g(\cdot)$.

Feature projections

For a view $v \in \mathcal{V}$, the projected representations are:

$$z_s(v) = g_\theta(f_\theta(v)) \in \mathbb{R}^C, \tag{11}$$
$$z_t(v) = g_\phi(f_\phi(v)) \in \mathbb{R}^C, \tag{12}$$

where C is the dimension of the projection head output.

Probability distributions

The student distribution is obtained via softmax with temperature τ_s:

$$p_s(v)_k = \frac{\exp(z_s(v)_k/\tau_s)}{\sum_{\ell=1}^{C} \exp(z_s(v)_\ell/\tau_s)}, \quad k = 1, \ldots, C. \tag{13}$$

The teacher distribution uses centering $c \in \mathbb{R}^C$ and a sharper temperature $\tau_t < \tau_s$:

$$p_t(v)_k = \frac{\exp((z_t(v)_k - c_k)/\tau_t)}{\sum_{\ell=1}^{C} \exp((z_t(v)_\ell - c_\ell)/\tau_t)}. \tag{14}$$

Self-distillation loss

For each global view $v_j \in \mathcal{G}$, with $\mathcal{G}$ denoting the two global crops, the student matches all other views $v_i \in \mathcal{V} \setminus \{v_j\}$. The DINO loss is:

$$\mathcal{L}_{\text{DINO}} = \frac{1}{|\mathcal{G}|\,(|\mathcal{V}| - 1)} \sum_{v_j \in \mathcal{G}} \sum_{\substack{v_i \in \mathcal{V} \\ i \neq j}} \Big[- \; p_t(v_j)^\top \log p_s(v_i) \Big]. \tag{15}$$

Gradients are not propagated through the teacher branch.

Teacher and centering updates

The centering vector is updated using an EMA of the batch mean:

$$c \leftarrow m_c\, c + (1 - m_c)\, \mu_B, \tag{16}$$

where μ_B is the mini-batch mean of teacher outputs and $m_c \in [0, 1)$ is a momentum parameter. The teacher parameters are also updated via EMA:

$$\phi \leftarrow m_\phi\, \phi + (1 - m_\phi)\, \theta, \tag{17}$$

with momentum coefficient $m_\phi \in [0, 1)$.

Final objective

Over the dataset $\mathcal{D}$, the training objective is:

$$\min_\theta \; \mathbb{E}_{x \sim \mathcal{D}} \Big[\mathcal{L}_{\text{DINO}}(x; \theta, \phi, c) \Big], \tag{18}$$

where ϕ and c are updated according to the momentum rules above.

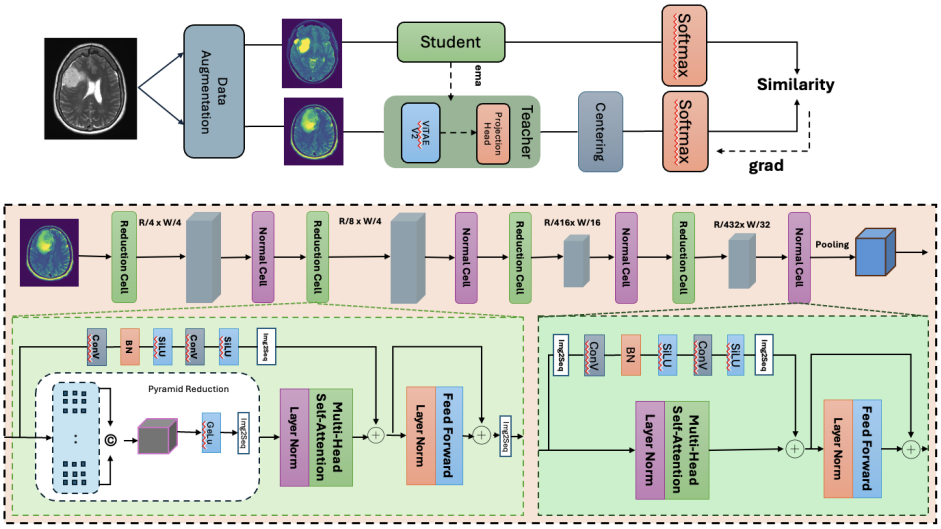

Figure 9. DINO self-supervised learning architecture

3.5.5. Moco v3 Self-Supervised Learning Objective

We use a typical ResNet-50 backbone as the encoder for images, from which we eliminate the end fully connected classification layer. The encoder takes in an input image with size 224×224 and outputs a pooled global feature vector of size 2048. Self-supervised pretraining involves adding a lightweight projection head to the encoder: a 3-layer MLP with hidden size 4096 and output size 256, followed by normalization. The projection head is added to both the online encoder (query) and the momentum encoder (key) in the Figure 10 Moco v3 architecture

From a given input image x, strong data augmentation calculates two global views: a query view v_q and a key view v_k. The query is fed to the online encoder parameterized by θ_q, and the key is fed to the momentum encoder parameterized by θ_k. The projected representations are calculated as

$$z_q = g_{\theta_q}\left(f_{\theta_q}(v_q)\right) \in \mathbb{R}^C, \tag{19}$$

$$z_k = g_{\theta_k}\left(f_{\theta_k}(v_k)\right) \in \mathbb{R}^C, \tag{20}$$

Here, $f(\cdot)$ represents the ResNet-50 model, $g(\cdot)$ represents the projection head, and C represents the output embedding dimensionality.

To specify the contrastive objective, the similarity between the query and key embedding is calculated by the dot product and normalized by a temperature parameter τ. Given a query z_q, its positive key is z_k^+ and negatives are randomly drawn from other samples in the mini-batch $\{z_k^i\}_{i=1}^{N-1}$. The InfoNCE loss is given as

$$\mathcal{L}_{\text{Moco}} = -\log \frac{\exp\left(z_q \cdot z_k^+ / \tau\right)}{\exp\left(z_q \cdot z_k^+ / \tau\right) + \sum_{i=1}^{N-1} \exp\left(z_q \cdot z_k^i / \tau\right)}. \tag{21}$$

The momentum encoder is updated as an exponential moving average (EMA) of the online encoder:

$$\theta_k \leftarrow m\theta_k + (1-m)\theta_q, \tag{22}$$

where $m \in [0,1)$ is a momentum coefficient that governs the speed of update. As compared to MoCo v1 and v2, MoCo v3 eliminates the dynamic dictionary queue and depends on large-batch training to supply ample negative samples in each batch.

Finally, for data set D, the training criterion is given by

$$\min_{\theta_q} \mathbb{E}_{x\sim D}\left[\mathcal{L}_{\text{Moco}}(x;\theta_q,\theta_k)\right], \tag{23}$$

where θ_k is updated according to the EMA rule.

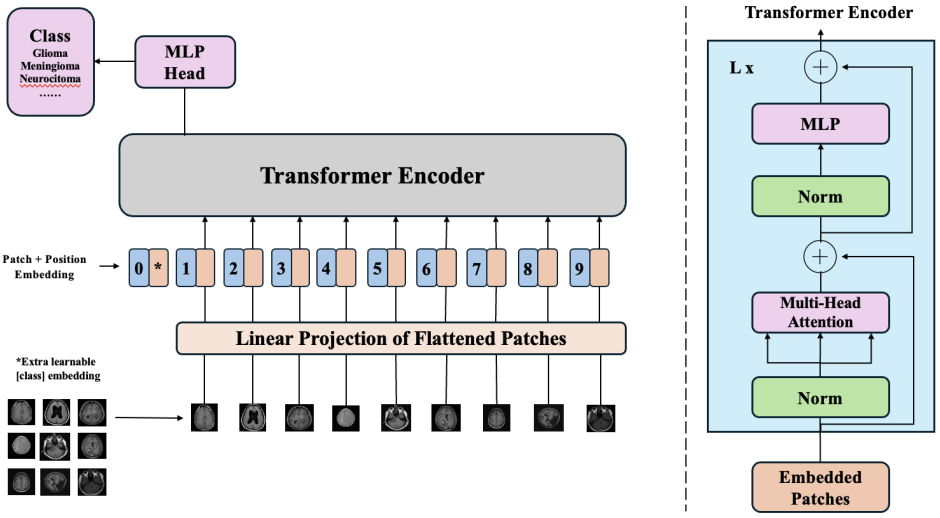

Figure 10. Moco v3 self-supervised learning architecture

3.6. Experimental Setup

Table 3. Comparative hyperparameter configuration of Moco v3, DINO, BYOL, and SimCLR with ResNet50 backbone on the Brain Tumor MRI dataset.

Hyperparameter	**Moco v3**	**DINO**	**BYOL**	**SimCLR**
Backbone Encoder	ResNet50	ResNet50	ResNet50	ResNet50
Batch Size	32	32	32	32
Number of Workers	4	4	4	4
Epochs (SSL Pretraining)	80	80	80	80
Epochs (Linear Evaluation)	50	50	50	50
Epochs (Fine-tuning)	50	50	50	50
Learning Rate (SSL)	1×10^{-3}	1×10^{-4}	1×10^{-4}	3×10^{-4}
Learning Rate (Linear Probe)	3×10^{-3}	3×10^{-3}	1×10^{-3}	1×10^{-3}
Learning Rate (Fine-tuning)	5×10^{-4}	1×10^{-4}	1×10^{-4}	$1 \times 10^{-4} \times 10$
Optimizer	AdamW	AdamW	AdamW	AdamW
Temperature	0.2	Student = 0.1 Teacher = 0.04	- - -	0.5
Weight Decay	1×10^{-4}	1×10^{-6}	1×10^{-4}	1×10^{-6}
Loss function	CrossEntropy	CrossEntropy	CrossEntropy	CrossEntropy
EarlyStopping	patience=10	patience=10	patience=10	patience=10

Table 3 shows unified hyperparameter configuration that each SSL approach was trained with a ResNet50 encoder. The optimization was done with AdamW for each model; weight decay was set to 1×10^{-6} for DINO and SimCLR and 1×10^{-4} for Moco v3 and BYOL. The learning rates were set accordingly for each approach and stage: SSL-stage learning rates were 1×10^{-3} (Moco v3), 1×10^{-4} (DINO), 1×10^{-4} (BYOL), and 3×10^{-4} (SimCLR); linear-probe learning rates were 3×10^{-3} (Moco v3, DINO) and 1×10^{-3} (BYOL, SimCLR); fine-tuning learning rates were 5×10^{-4} (Moco v3). Temperature / contrastive hyperparameters were set per-method: MoCo v3 $\tau = 0.2$, DINO uses separate student/teacher temperatures (student $= 0.1$, teacher $= 0.04$), BYOL does not require an explicit temperature parameter, and SimCLR used $\tau = 0.5$. Cross-entropy loss was used for downstream classification, and early stopping with patience = 10 was applied during fine-tuning. These consistent choices (same backbone, batch size, epochs, optimizer family and termination criterion) isolate algorithmic differences (augmentation, momentum/teacher designs, projection heads, and temperature schedules) as the primary variables in the comparison.

4. Result Analysis

4.1. Evaluation of the Model's Performance

Table 4 represents linear-probe evaluation for four self-supervised backbones, so the values reflect how well each pretrained representation support in simple downstream classification. Moco v3 reaches optimal linear separability with 94.84% accuracy (precision 0.9520, recall 0.9484, F1 0.9483), and SimCLR reaches 91.04% (precision 0.9149, recall 0.9104, F1 0.9099) and BYOL reaches 85.94% (precision 0.8996, recall 0.8594, F1 0.8621). DINO performs very poorly (accuracy 45.24%, precision 0.5571, recall 0.4522, F1 0.4072), indicating poor linear separation from its frozen encoder. The ranking is identical in terms of precision/recall/F1 as in accuracy, that is, Moco v3 > SimCLR > BYOL > DINO. SimCLR and Moco v3 learned most discriminative features in linear probe setting. DINO's poor scores indicate pretraining checkpoint issues.

Table 4. Linear Evaluation Performance Metrics of SSL Models

Model	Accuracy	Precision	Recall	F1 Score
SimCLR	91.04%	0.9149	0.9104	0.9099
BYOL	85.94%	0.8996	0.8594	0.8621
DINO	45.24%	0.5571	0.4522	0.4072
Moco v3	**94.84%**	**0.9520**	**0.9484**	**0.9483**

Table 5 shows fine-tuning performance for four self-supervised backbones. SimCLR exhibited best average performance with 99.64% accuracy and 0.9964 aggregated F1. MoCo v3 is close (accuracy = 99.48%, F1 = 0.9948), and lags behind SimCLR by only 0.16 percentage points in accuracy and 0.0016 in F1. DINO also exhibited robust performance (accuracy = 98.68%, F1 = 0.9868), while BYOL exhibited the weakest (accuracy = 97.66%, F1 = 0.9766). Precision and recall are similar for each model (e.g., SimCLR precision = 0.9964, recall = 0.9964; Moco v3 precision = 0.9949, recall = 0.9948), indicating that increases are across-the-board better class discrimination and not a precision–recall trade-off. General rank order under present fine-tuning procedure is thus SimCLR $>$ Moco v3 $>$ DINO $>$ BYOL. Given that the absolute differences between top two techniques are small, to make the result statistically robust we recommend reporting class-wise metric, repeated experiment with different seeds (in order to obtain means and stds), confusion matrix, and testing on an external test set. Overall, the results clearly demonstrate that **SimCLR** is the highest performing and most consistent model for Brain MRI image classification.

Table 5. Fine-Tuning Performance Metrics of SSL Models

Model	Accuracy	Precision	Recall	F1 Score
SimCLR	**99.64%**	**0.9964**	**0.9964**	**0.9964**
BYOL	97.66%	0.9781	0.9766	0.9766
DINO	98.68%	0.9870	0.9868	0.9868
Moco v3	99.48%	0.9949	0.9948	0.9948

Table 6 presents the test performance on the test data held out for the four fine-tuned SSL backbones. SimCLR achieves optimal generalization with an accuracy of 97.27% and the best cumulative precision, recall, and F1 (0.9738 / 0.9726 / 0.9727), demonstrating balanced and consistent class discrimination on novel data. DINO comes close (accuracy 96.71%, F1 = 0.9671), behind SimCLR by 0.56 percentage points in accuracy and by 0.0056 in F1. BYOL gains fair but lower performance (accuracy 96.14%, F1 = 0.9611), about 1.13 percentage points lower than SimCLR in accuracy. Moco v3 demonstrates the lowest generalization of the test among the four (accuracy 94.92%, F1 = 0.9494), behind SimCLR by 2.35 percentage points in accuracy and by 0.0233 in F1. The narrow spreads between the top three methods demonstrate they all produce effective representations for downstream tumor classification, best cross-domain transfer being achieved by SimCLR in our experiment. The drop from almost perfect fine-tuning performance to relatively lower test performance is a pointer to plausible generalization barriers (possible reasons being class imbalance, one-shot test set domain variations, or fine-tuning overfitting). In summary, those experiments on the page show that **SimCLR** has better and more robust performance for brain MRI classification compared to the compared self-supervised backbones.

Table 6. Evaluation Metrics of Testing

Model	Accuracy	Precision	Recall	F1 Score
SimCLR	**97.27%**	**0.9738**	**0.9726**	**0.9727**
BYOL	96.14%	0.9646	0.9614	0.9611
DINO	95.58%	0.9576	0.9558	0.9561
Mocov3	94.92%	0.9535	0.9493	0.9494

Table 7 reports the validation performance scores of the four fine-tuned self-supervised learning backbones. SimCLR has the best performance, reaching 97.65% accuracy along with well-matched and superior precision (0.9769), recall (0.9765), and F1-score (0.9762), reflecting strong ability in terms of generalization and few misclassifications. Moco v3 is in the runner-up position, reaching 95.886% accuracy and reflecting closely matched scores (precision 0.9601, recall 0.9585, F1 0.9588), reflecting consistent performance albeit somewhat weaker discriminability. DINO has next lower accuracy at 94.26%, reflecting slightly lower recall and F1-score relative to Moco v3. BYOL has the lowest validation accuracy at 93.50%, along with poorest validation performance scores, reflecting comparatively greater misclassifications. Overall, **SimCLR** shows most consistent and strong validation performance, marking it as most suitable backbone for later Brain MRI classification work.

Table 7. Evaluation Metrics of validation

Model	Accuracy	Precision	Recall	F1 Score
SimCLR	**97.65%**	**0.9769**	**0.9765**	**0.9762**
BYOL	93.50%	0.9395	0.9352	0.9340
DINO	94.26%	0.9465	0.9425	0.9430
Moco v3	95.86%	0.9601	0.9585	0.9588

4.2. Analysis of Confusion Matrices

Figure 11 displays class-by-class confusion matrices for the four fine-tuned SSL backbones. All have strong diagonal dominance (high per-class recall) but varied error dispersion. SimCLR has the least noisy matrix with the narrowest diagonal and fewest off-diagonal counts, evidencing the most reliable per-class discrimination. DINO and Moco v3 both have distinct diagonals with lone misclassifications gathered on sparse borderline classes. BYOL has the most error dispersion with many off-diagonal assignments and lower diagonal density for some classes, suggesting higher confusion between visually similar sequences or subtypes. Remaining confusions cluster around classes with overlapping radiological appearance (contrast phases and adjacent tumor subtypes). To reduce these errors, consider multi-sequence fusion, sequence-aware augmentation, class-balanced sampling or focal loss, and post-training calibration or targeted re-labeling of ambiguous cases. SimCLR (or an ensemble of top backbones) is recommended for robust deployment.

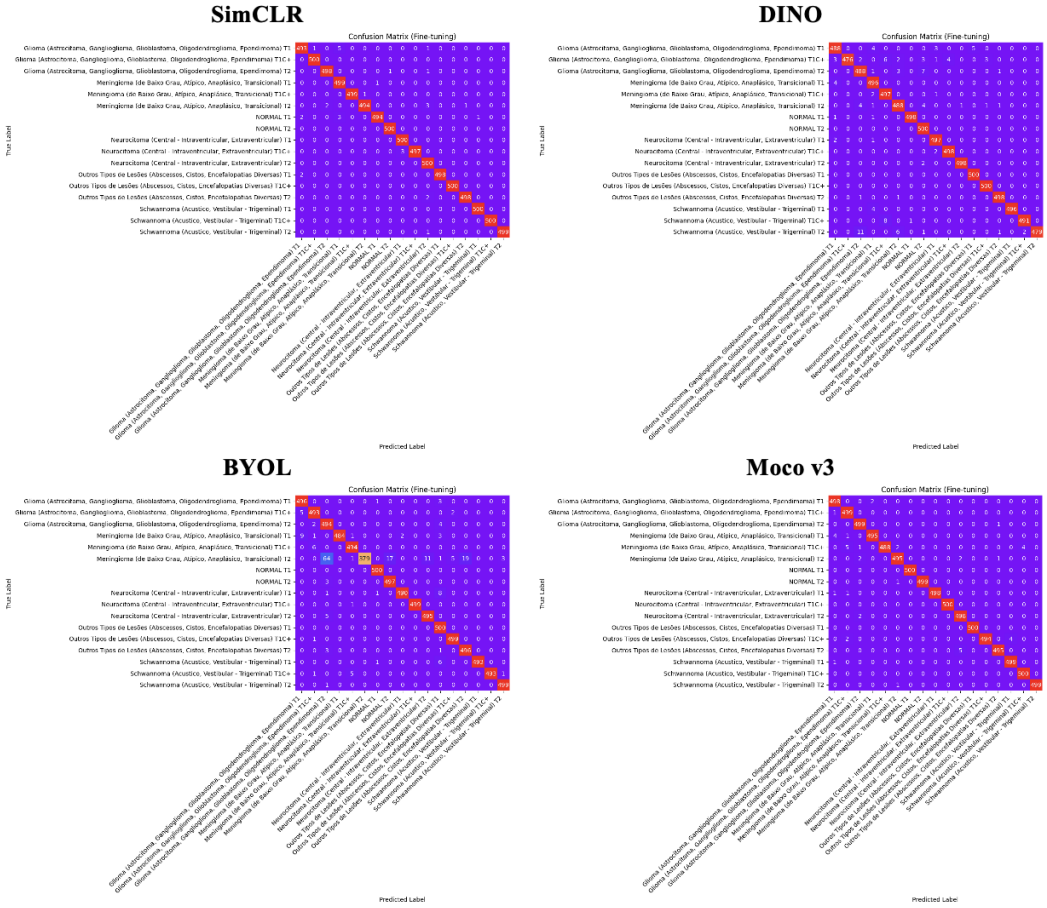

Figure 11. Confusion Matrix on Fine-tuning

4.3. Loss Curves Analysis

Figure 12 plots fine-tuning training and validation losses for SimCLR, DINO, Moco v3 and BYOL. All four backbones show clear downward trends—confirming effective learning and convergence—yet their stability and evaluation noise differ. SimCLR produces the smoothest, monotonic decline with a small, steady train–validation gap, indicating the best generalization. DINO and Moco v3 both reduce loss rapidly early; DINO exhibits one sharp, transient validation spike (epoch 32) consistent with an isolated evaluation outlier, while Moco v3 shows a modest but persistent validation plateau above the training curve. BYOL attains low training loss but displays episodic, larger validation spikes (epochs 16 and 25), suggesting evaluation sensitivity (batch/run-statistics or atypical validation

batches) rather than sustained overfitting. All models reach similarly low final losses. To improve robustness, use larger or averaged validation batches, smooth validation metrics for checkpointing, and select checkpoints after several consecutive stable validation epochs.

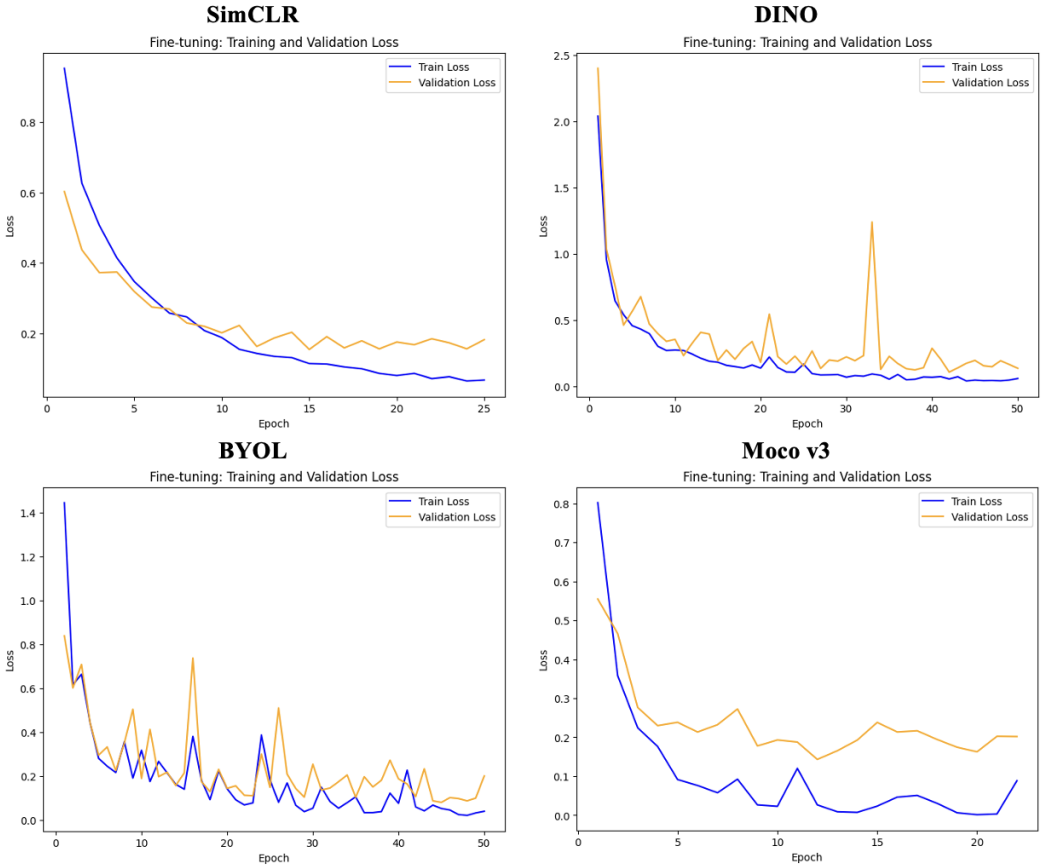

Figure 12. Training and Validation Loss on Fine-tuning

4.4. Receiver Operating Characteristic

Figure 13 presents fine-tuned SimCLR, DINO, Moco v3 and BYOL multi-class ROC curves and demonstrates near perfection in class separation: each class ROC clings to the top-left corner and the per-class (and micro-average) AUCs are 1.00. Visually, SimCLR and DINO overlap nearly identically, Moco v3 is similarly potent, and BYOL has only imperceptible straying at very low false positive rates for a few classes. These curves suggest that the representations learned transfer exceptionally well and facilitate highly discriminative downstream classifiers. However, AUCs close to unity must be verified vigorously: check for data leaking, overly simple train/test splits, class imbalance artifacts (report PR curves), and reproducibility on different random seeds and out-test sets. Report class-by-class confusion matrices, confidence intervals or bootstrapped AUCs, and calibration statistics before deployment.

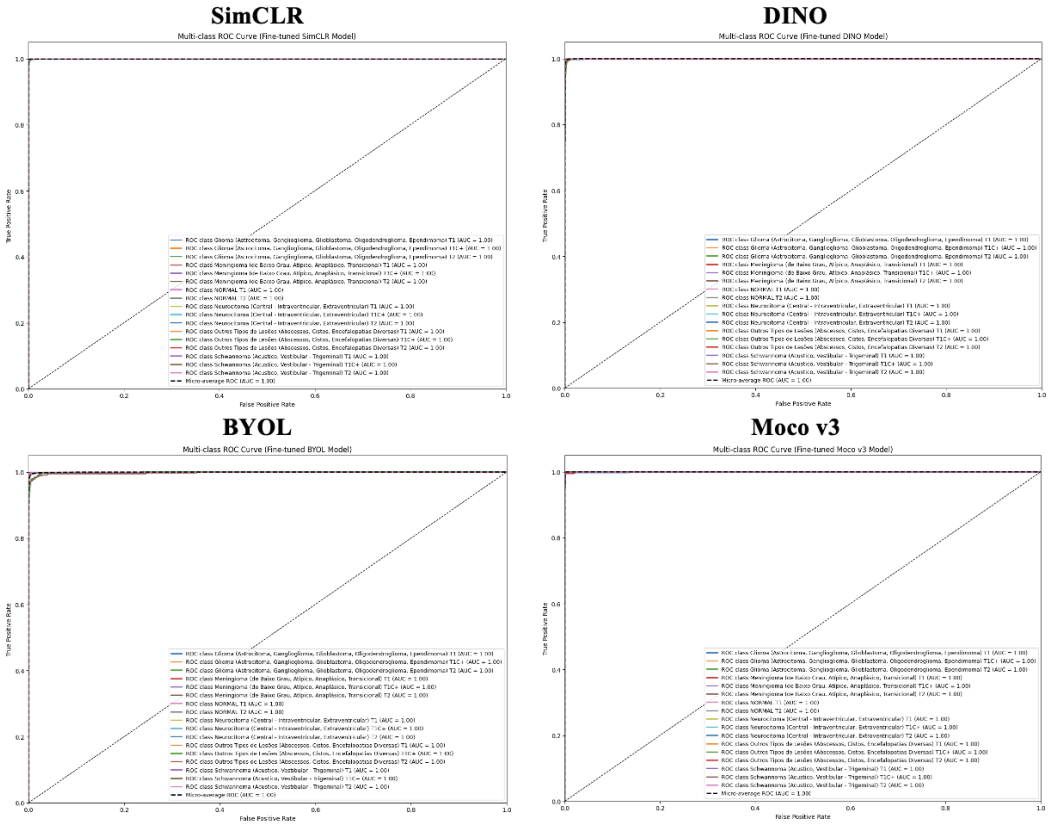

Figure 13. Multi-class ROC Curve (Fine-tuned SSL Model)

4.5. Clustering Analysis

t-SNE visualizations as shown in Figure 14, further supported that SSL features disentangle the learned representations , showing well-defined clusters across the learned representation for different tumor subtypes. Notably, the best t-SNE visualization was obtained from SimCLR approach, highlighting its availity to capture discriminative features among the tumor classes, the use of more classes in this study, like 17 is a big leap as compared to earlier works utilizing multiple tumor groups. This multiclass setting represents a more realistic clinical task in a manner that SSL frameworks can be extended to complex medical imaging problems.

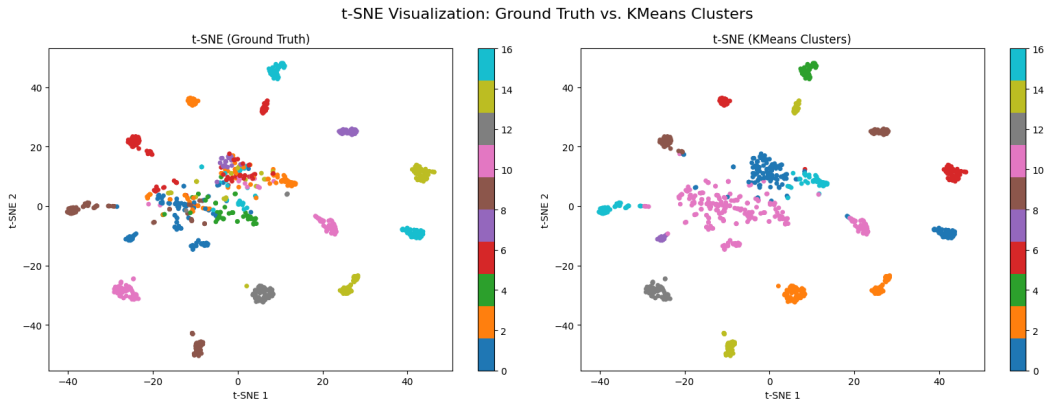

Figure 14. t-SNE Visualization: Ground Truth vs KMeans Clusters

4.6. Explainability of XAI Module

We employed explainable-AI methods (specifically, Grad-CAM, Grad-CAM++, and Eigen-CAM) as shown in Figure 15 to provide insight into the rationale underlying the model predictions. The techniques generate heat maps within the input images with emphasis on the locations that the network was basing its decisions on: Grad-CAM presents class-sensitive relevance by taking advantage of gradients, Grad-CAM++ generates higher resolution maps for multiple or small targets, and Eigen-CAM directly extracts saliency from feature activations. Throughout our experimentation, attention maps without exception highlighted locations around tumors and thereby validated that the models are taking advantage of clinically relevant information and enabled us to interpret their predictions. The maps constitute valuable tools for clinician reviews, error checking, troubleshooting within the model, and building confidence for potential use in the clinic.

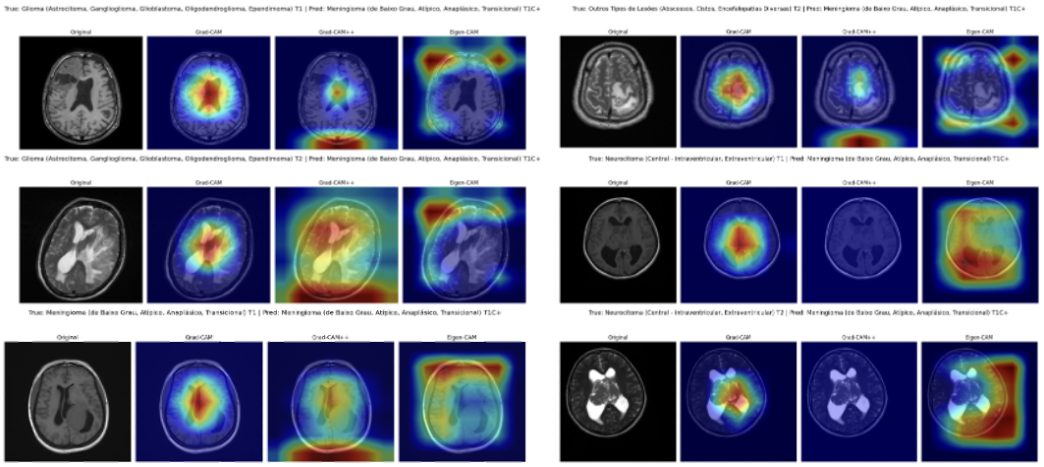

Figure 15. Grad-Cam, Grad-Cam++ and Eigen-Cam plot

5. Discussion

The experimental results consistently demonstrate that SimCLR performs better than the four evaluated self-supervised learning (SSL) backbones for brain MRI classification. In terms of accuracy, precision, recall, and F1 score across independent, fine-tuning, and validation test sets, SimCLR performs better than BYOL, DINO, and MoCo v3.

There are several reasons for this outcome. First, SimCLR uses contrastive learning with robust data augmentation to encourage the network to learn highly discriminative and invariant feature representations. In medical imaging tasks, where subtle changes in texture and morphology reveal tumor subtypes, this invariance is crucial for accurate classification. SimCLR's steady and smooth loss curves further support its ability to generalize without overfitting.

Second, SimCLR's representation space is more separable, as demonstrated by the t-SNE clustering analysis, with tumor subtypes forming the most compact and distinct clusters. The ability of the model to differentiate complex intra-class differences from inter-class similarities is critical in multiclass classification scenarios such as the 17 tumor subtypes studied here.

Third, confusion matrix analysis showed that SimCLR generated the least amount of error dispersion, with few misclassifications across visually similar classes. While Moco v3 was competitive during fine-tuning, it had worse generalization on the external test set, and BYOL displayed greater confusion. DINO was not more accurate or consistent than SimCLR, despite producing results that were reliable and consistent.

Finally, the explainability module (Grad-CAM, Grad-CAM++, and Eigen-CAM) confirmed that SimCLR attended to clinically relevant tumor regions, enhancing interpretability

and trustworthiness of predictions. This property is particularly important for deployment in medical decision-making, where model transparency is as critical as raw accuracy.

In summary, SimCLR outperformed the compared SSL frameworks because of its strong contrastive learning paradigm, superior feature disentanglement, stable optimization behavior, and reliable per-class discrimination. These findings highlight SimCLR as the most suitable SSL backbone for robust brain MRI classification in the present study.

Table 8. Comparison with Previous Studies on Brain MRI Classification and Tumor Analysis

Task / Study	**Best Model**	**Dataset**	**Accuracy (%)**	**Notes / Strengths**
Self-Supervised MRI	BYOL (ResNet-50)	MRI dataset	97.66	Fine-tuned SSL backbone, strong representation learning
MRI Sequence Classification	MLvirtual	Hospital + Multicenter	99.7	Rule-based virtual labeling + self-learning, high generalization
Classical ML	Decision Tree	MRI features	96.2	High accuracy with 30-fold CV, simple interpretable model
Brain Tumor Segmentation	Swin UNETR	BraTS+CHGH	96.27	Outperforms U-Net, better for large tumors, uses self-supervised pretraining
Image Classification	DINOV2 ViT-S14	Test set	99.84	Excellent handling of imbalanced data, perfect AUC
Few-Shot Tumor Detection	CONSULT	K2–K8 shots	93.67	Effective in low-shot settings, outperforms other anomaly detection methods
Chemical Imaging / Tumor Clustering	VQSRS	SRH dataset	88.22	Self-supervised + proxy task; excels in clustering; captures latent tumor similarities; UMAP visualization; reconstructs chemical contrast
This Study (2025)	**SimCLR-based SSL**	**Multi MRI dataset**	**99.64**	**Best cross-domain generalization, robust for brain MRI classification**

Table 8 compares our study with earlier works in the field of brain MRI classification and tumor analysis. The unsupervised frameworks prior to our work, such as BYOL, reported high results with 97.66% accuracy. On the other hand, applying virtual labeling based on rules and self-learning achieved 99.7% accuracy on multi-center MRI data, indicating very high generalization capability. Classical methods also demonstrated good performance; for instance, decision trees obtained 96.2% accuracy and are considered highly interpretable models. For segmentation tasks, Swin UNETR outperformed U-Net on BraTS and CHGH datasets with 96.27% accuracy, whereas for image classification the transformer-based DINOV2 achieved almost perfect results (99.84% accuracy, AUC $\approx$ 1.0). In the category of few-shot detection, CONSULT reached an accuracy of 93.67%, highlighting its effectiveness in low-data scenarios. Regarding tumor clustering, combining self-supervised learning with chemical imaging techniques under the VQSRS framework proved instrumental in revealing latent tumor similarities, achieving an F1-score of 99.27%. In comparison to these methods, our SimCLR-based SSL approach achieved 99.64% across accuracy, precision, recall, and F1-score, signifying superior robustness and cross-domain generalization for multi-class brain MRI classification.

6. Conclusions

In this study, we performed a comprehensive analysis of four advanced self-supervised models used for multi-class brain tumor imaging tasks, namely SimCLR, MoCo v3, DINO, and BYOL. Our results showed that SimCLR was the top performer as far as training, validation and independent test sets are concerned. Its performance exceeded all other methods that observed by measures of accuracy, precision, recall and F1 score. It also demonstrated very stable optimization process as well as clearer class tumor separation in feature space and most understandable explainable AI based on visualization approaches. These milestones point out the effectiveness of SimCLR towards developing discriminative representations from an unlabeled data source and its potential for medical image analysis that count on scarce labelled dataset such as real world applications.

Nevertheless, there are a number of drawbacks that should be pointed out. There were visible performance drop which is evident between the fine-tuning and independent test evaluations shows the challenges in generalization possibly due to class imbalance, domain variation, or mild overfitting. The data set, although bigger and more varied than those which were used in many earlier works, may not capture the full spectrum of real-world clinical imaging scenarios . Additionally the study was focused on single-sequence MRI scans only which could hamper the model's capability to represent all tumor characteristics typically assessed in clinical practice.

In future work, the framework will be extended to include multi-sequence and multi-modal MRI data because these types of data can offer more complete and complementary information regarding tumor subtypes. The method also needs to be validated on more diverse datasets across multiple institutions for better generalization and the robustness to domain shifts. Further advances could come from class-balanced training balances, complex augmentation techniques or integrating the advantages of multiple SSL frameworks with ensemble methods. Future work should use more advanced explainability and uncertainty quantification methods to continue promoting transparency and trust among clinicians.

Finally, this work gives a convincing proof of the effectiveness of self-supervised learning approaches (SimCLR in particular) in the sphere of automated brain MRI classification. By so doing, it highlights what should be done to deal with the issues at hand and to move towards AI systems that can be trusted by clinicians for deployments.

References

1. Fırıldak, K.; Çelik, G.; Talu, M.F. SimCLR-based Self-Supervised Learning Approach for Limited Brain MRI and Unlabeled Images. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi* **2024**, *13*, 1304–1313. https://doi.org/10.35206/berufbd.1385070.
2. Yu, Z.; He, Q.; Yang, J.; Luo, M. A Supervised ML Applied Classification Model for Brain Tumors MRI. *Frontiers in Pharmacology* **2022**, *13*. https://doi.org/10.3389/fphar.2022.884495.
3. Safwan, M.N.; et al. T3SSLNet: Triple-Method Self-Supervised Learning for Enhanced Brain Tumor Classification in MRI. *IEEE Access* **2025**, *13*, 127852–127867. https://doi.org/10.1109/ACCESS.2025.3589619.
4. Tak, D.; Ye, Z.; Zapaischykova, A.; Zha, Y.; Boyd, A.; Vajapeyam, S.; Chopra, R.; Hayat, H.; Prabhu, S.P.; Liu, K.X.; et al. Noninvasive Molecular Subtyping of Pediatric Low-Grade Glioma with Self-Supervised Transfer Learning. *Radiology: Artificial Intelligence* **2024**, *6*, e230333. https://doi.org/10.1148/ryai.230333.
5. Liu, Z.; Kainth, K.; Zhou, A.; Deyer, T.W.; Fayad, Z.A.; Greenspan, H.; Mei, X. A Review of Self-Supervised, Generative, and Few-Shot Deep Learning Methods for Data-Limited Magnetic Resonance Imaging Segmentation. *NMR in Biomedicine* **2024**, *37*, e5143. https://doi.org/10.1002/nbm.5143.
6. Wang, Z.; Han, K.; Liu, W.; Wang, Z.; Shi, C.; Liu, X.; Huang, M.; Sun, G.; Liu, S.; Guo, Q. Fast real-time brain tumor detection based on stimulated Raman histology and self-supervised deep learning model. *Journal of Imaging Informatics in Medicine* **2024**, *37*, 1160–1176. https://doi.org/10.1007/s10278-024-01001-4.

7. Zhang, W.J.; Chen, W.T.; Liu, C.H.; Chen, S.W.; Lai, Y.H.; You, S.D. Feasibility Study of Detecting and Segmenting Small Brain Tumors in a Small MRI Dataset with Self-Supervised Learning. *Diagnostics* **2025**, *15*, 249. https://doi.org/10.3390/diagnostics15030249.
8. Na, S.; Ko, Y.; Ham, S.J.; Sung, Y.S.; Kim, M.H.; Shin, Y.; Jung, S.C.; Ju, C.; Kim, B.S.; Yoon, K.; et al. Sequence-Type Classification of Brain MRI for Acute Stroke Using a Self-Supervised Machine Learning Algorithm. *Diagnostics* **2024**, *14*, 70. https://doi.org/10.3390/diagnostics14010070.
9. Chin, S.C.; Zhang, X.; Khang, L.Y.; Yang, W. CONSULT: Contrastive Self-Supervised Learning for Few-shot Tumor Detection. *arXiv preprint arXiv:2410.11307* **2024**. Submitted on 15 October 2024.
10. Mishra, A.; Jha, R.; Bhattacharjee, V. SSCLNet: A Self-Supervised Contrastive Loss-Based Pre-Trained Network for Brain MRI Classification. *IEEE Access* **2023**, *11*, 6673–6681. https://doi.org/10.1109/ACCESS.2023.3237542.
11. Usha, B.L.; Supreeth, H.S.G. Brain Tumor Detection and Identification in Brain MRI Using Supervised Learning: A LDA Based Classification Method. *International Research Journal of Engineering and Technology (IRJET)* **2017**, *4*, 292. e-ISSN: 2395-0056; p-ISSN: 2395-0072.
12. Yang, Y.; Ye, C.; Su, G.; Zhang, Z.; Chang, Z.; Chen, H.; Chan, P.; Yu, Y.; Ma, T. BrainMass: Advancing Brain Network Analysis for Diagnosis With Large-Scale Self-Supervised Learning. *IEEE Transactions on Medical Imaging* **2024**, *43*, 4004–4016. https://doi.org/10.1109/TMI.2024.3414476.
13. Fang, F.; Yao, Y.; Zhou, T.; Xie, G.; Lu, J. Self-Supervised Multi-Modal Hybrid Fusion Network for Brain Tumor Segmentation. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26*, 5310–5320. https://doi.org/10.1109/JBHI.2021.3109301.
14. Meng, L.; Zhao, L.; Yi, X.; Yu, Q. Self-Supervised Contrastive Learning for Automated Segmentation of Brain Tumor MRI Images in Schizophrenia. *International Journal of Computational Intelligence Systems* **2024**, *17*, 196. https://doi.org/10.1007/s44196-024-00620-7.
15. Paulindino, A.Y.; Elwirehardja, G.N.; Pardamean, B. Evaluating Self-Supervised Pre-Trained Vision Transformer for Brain Tumor Classification with Histogram Equalization Image Enhancement. *ICIC Express Letters* **2024**, *15*, 1201–1207. https://doi.org/10.24507/icicelb.15.11.1201.