SUBMISSION DATE: JUNE 20, 2025

FINAL PROJECT REPORT

Project Title: Brain Tumor Recognition and Classification
Course Name : Introduction to Artificial Intelligence
Academic Year/Semester: 2024-2025 Spring Semester

Course Instructor Name:

Selim Yılmaz

Team Members' Names and Student IDs:

Funda Semiz (220717039)

İlayda Bozkurt (220717037)

Mahmut İsmail Öztürk (230717024)

CONTENTS

1.Introduction

1.1. Problem Definition

Early diagnosis of brain tumors is vital for patients' treatment success and quality of life. Tumors diagnosed late often require more aggressive treatment methods and can negatively impact prognosis. Current diagnostic processes typically involve radiologists manually examining MRI images. However, this approach faces several challenges, including its reliance on expert experience, its time-consuming nature, and its susceptibility to human error.

This project specifically addresses the scarcity of data and the difficulties in labeling within the medical imaging field. It's particularly challenging to find sufficient and labeled data for rare tumor types. While traditional deep learning models depend on large datasets, this poses a significant obstacle in medical applications. Therefore, the fundamental reason for undertaking this project is the urgent need to develop an automated system that can learn effectively and accurately classify tumors even with limited data. Such a system could both accelerate diagnostic processes, thereby assisting physicians, and provide a robust solution in the face of data scarcity.

1.2. Project Aim and Objectives

The primary aim of this project is to accurately detect the presence of brain tumors from Magnetic Resonance (MR) images and reliably classify tumor types. Specifically, we aim to overcome the challenge of limited labeled data in medical datasets by developing a system capable of distinguishing specific brain tumor types—such as Glioma, Meningioma, and Pituitary Tumor, as well as normal (tumor-free) conditions—with high accuracy, even with a small number of labeled samples. Through this, we intend to create a reliable and automated diagnostic support system that contributes to medical diagnosis processes, even in data-scarce environments.

1.3. Project Scope

This project on brain tumor detection and classification has specific boundaries and focus areas:

Imaging Modality Focus: The project exclusively utilizes **Magnetic Resonance (MR) images** for diagnosis. Different imaging techniques (e.g., CT, PET) or other clinical data (blood tests, symptoms, etc.) are outside the scope of this study.

Focus on Specific Tumor Types: The classification capability covers three primary tumor types: **Glioma, Meningioma, and Pituitary Tumor**, in addition to **Normal (tumor-free)** brain images. Other rare or different brain tumor subtypes are not a direct target of this project.

Limited Data Scenario: The study focuses on the ability to learn under a **10-Shot Learning paradigm**, meaning with a limited number of labeled examples for each class. This is the key aspect that differentiates the project from traditional deep learning applications requiring large and abundantly labeled datasets.

Automated Classification: The project aims to automatically detect and classify tumors from MR images, minimizing manual intervention. Other computer vision tasks, such as image segmentation, are not a primary objective of this project but could be integrated into future work.

Interpretability: A general limitation of deep learning models is the difficulty in fully explaining the reasoning behind their predictions. In medical applications, the **transparency of model decisions** is an important factor for clinicians.

2. Literature Review

2.1. Similar Studies and Approaches: In recent years, deep learning methods have made significant strides in the field of brain tumor detection and classification. Most studies have aimed to identify and classify tumors in MRI images by utilizing Convolutional Neural Networks (CNNs) and applying transfer learning with data augmentation techniques. However, a fundamental challenge with these approaches is their reliance on large and comprehensive labeled datasets for effective performance. Labeling and sharing medical data, unfortunately, are both costly and difficult.

2.2. The Difference of Our Project from These Studies: Our project stands apart from existing work by offering solutions to these limitations:

- **10-Shot Learning Focus:** We directly address the data scarcity problem by performing classification with only 10 labeled examples per tumor class.
- **Use of Self-Supervised Learning (SimCLR):** This minimizes the need for extensive labeling by learning powerful and generalizable visual features from unlabeled images.

- **Adaptation to Realistic Scenarios:** Our approach provides a more suitable model for clinical applications, especially in the diagnosis of rare tumor types, by achieving high performance with limited examples.

3. Methodology

3.1. Dataset

The dataset used in this project is a subset of the "Brain Tumor MRI Dataset" obtained from Kaggle. A specific number of images were selected from the dataset to create a scenario suitable for the 10-shot learning paradigm. The dataset is divided into two main sections:

1. **Pre-training Dataset (SimCLR):** This contains a total of **84 images** (21 from each class) from the "normal Dataset" folder.

   Purpose: To enable the model to learn general visual representations.

2. **10-Shot Classification Dataset:**

   Training (Support) Set: Located in the evalData/Eval_Training folder. It contains a total of **40 images**, with 10 images from each class.

   Test (Query) Set: Located in the evalData/Eval_testing folder. It contains a total of **1200 images**.

Number of Classes: 4 classes (Glioma, Meningioma, Pituitary Tumor, Normal).

Key Information: This dataset distribution simulates real-world scenarios, particularly reflecting situations where only a few examples might be available for identifying rare cases.

This dataset forms the foundation of the project and is used to evaluate the model's ability to classify brain tumor types with limited data.

3.2. Model Selection and Architecture

In our project, we've adopted a two-stage model architecture to achieve brain tumor detection and classification in a limited data scenario: Feature Learning with Self-Supervised Learning (Pre-training), followed by 10-Shot Classification.

3.2.1. Feature Learning with Self-Supervised Learning (Pre-training)

Objective: The primary goal of this initial stage is to learn general and robust visual representations from unlabeled images, without the need for manual labeled annotations. This enables the model to extract powerful and generalizable features despite the data scarcity problem.

Working Principle: We utilized the SimCLR (Simple Framework for Contrastive Learning of Visual Representations) architecture. SimCLR creates two different "augmented" (distorted) copies of a single image. The model then learns to bring the feature representations of these two copies closer together, while simultaneously pushing away representations from other images. This contrastive learning process allows the model to extract consistent, discriminative features for semantically similar image pairs.

Model Used: For this pre-training phase, we employed ResNet18, a widely used and robust convolutional neural network (CNN) architecture.

Dataset: For this stage, we leveraged a larger pool of unlabeled brain images, totaling 200 images (50 from each class) taken from the "normal Dataset" folder.

Output: The pre-training process results in a strong and generalizable feature extractor (backbone). This backbone is capable of capturing complex and discriminative visual patterns in medical images.

3.2.2. 10-Shot Classification

Objective: To accurately classify tumor types using only 10 examples per class (10-shot), leveraging the features learned during the pre-training phase. This approach is ideal for medical scenarios where labeled data is particularly limited.

Steps:

- **Loading and Freezing the Pre-trained ResNet18 Backbone:** To preserve the general knowledge gained during pre-training and reduce the risk of overfitting with limited labeled data, the pre-trained ResNet18 backbone was loaded and its weights were frozen.
- **Feature Extraction:** Features were extracted from the frozen backbone using 10 labeled training images from each class (a total of 40 images) obtained from the evalData/Eval_Training folder.

- **Classifier Training:** A simple yet effective Logistic Regression classifier was trained on these extracted features. Logistic Regression is a suitable choice for linear classification in high-dimensional feature spaces.
- **Model Evaluation:** The model's performance was evaluated using a total of 84 test images taken from the evalData/Eval_testing folder.

This two-stage approach forms the core of our project, enabling us to achieve our goal of effective brain tumor classification even with limited data.

3.3. Educational Process

The training process of the model was carried out in two main phases in accordance with the determined methodology: pre-training with SimCLR and 10-Shot classification training. The development environment, libraries, training parameters, loss functions and evaluation metrics used in this process are detailed below.

3.3.1. Development Environment and Libraries

All development and training processes of the project were carried out using the Python programming language. TensorFlow and its high-level API Keras libraries were used to create and train deep learning models. OpenCV and Pillow libraries were actively used in image processing and pre-processing steps. NumPy and Pandas were used for data manipulation and analysis, and the scikit-learn library was used for model evaluation and various machine learning tools. Matplotlib and Seaborn were used for visualizations.

3.3.2. Training Parameters

The training parameters of the model were determined separately for SimCLR pre-training and subsequent Logistic Regression classification phases:

For SimCLR Pre-Tutorial:

- **Optimization Algorithm: Adam**
- **Learning Rate: 0.01**
- **Number of Epoch : 163**
- **Batch Size: 21**
- **Hardware: Rtx3050(4 gb)**

For 10-Shot Logistic Regression Classifier:

- **Optimization Algorithm:Adam**

- **Training:** Since the pre-trained ResNet18 backbone was frozen at this stage, only the Logistic Regression classifier was trained with a limited number of labeled 40 images. The training period was quite short in this phase.

3.3.3. Loss Function and Evaluation Metrics

The following loss function and evaluation metrics were used to measure the performance of the model and guide the training process:

Loss Function:

- **For SimCLR Pre-Training: contrastive loss**
- **For 10-Shot Classification: Log Loss**

Evaluation Metrics: Various metrics were used on the test set to evaluate the final performance of the model:

- **Accuracy:** It is the ratio of all correct predictions to the total predictions. Shows overall model performance.
- **Precision:** It is the proportion of samples predicted to be positive that are truly positive. Important when false positives are costly.
- **Recall / Sensitivity:** It is the ratio of true positives to all true positives. It is important to avoid missing critical situations such as tumors (reducing false negatives).
- **F1-Score:** It is the harmonic mean of precision and sensitivity. Provides a more balanced performance indicator in situations with class imbalance.
- **Confusion Matrix:** It provides a detailed table showing the correct and incorrect predictions the model makes for each class. This matrix is critical for visualizing which classes the model mixes together.

This comprehensive training and evaluation strategy enabled the project to achieve reliable and meaningful results.

4. Experiments and Results

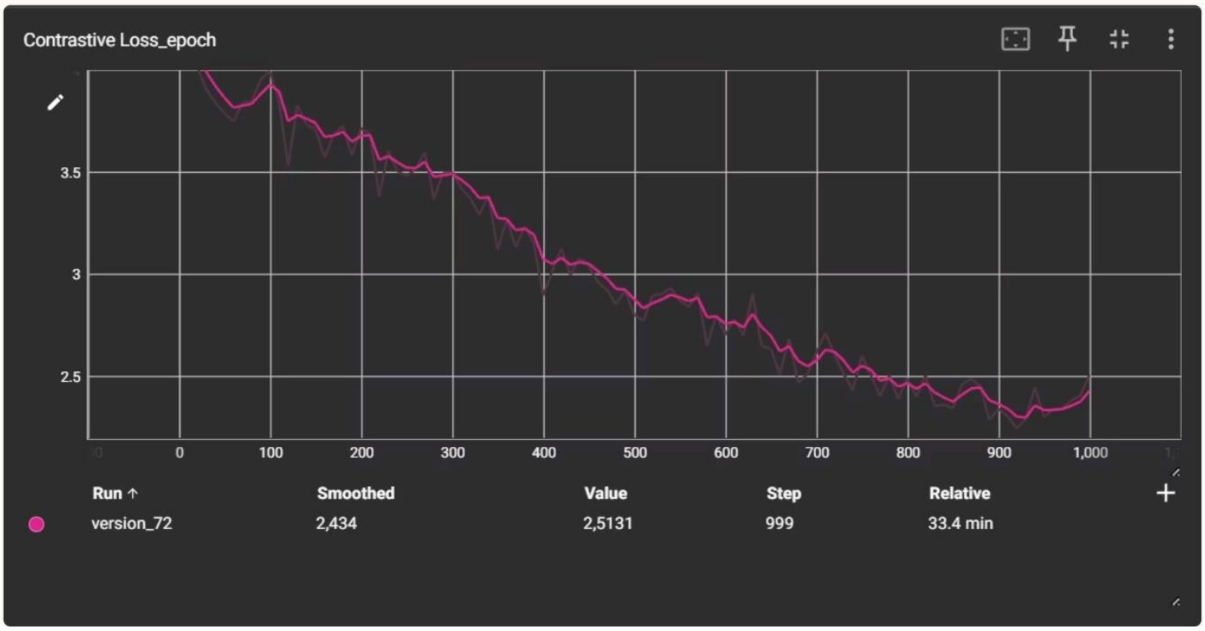
In this section, the performance of the developed model is presented in detail, and the obtained results are supported with visual materials.

- **Experimental Environment:** GPU
- **Training Loss:** 2.43
- **Evaluation Accuracy :%** 62.85

4.1. SimCLR Pre-training Loss Function

To demonstrate the success of the model's self-supervised learning phase (SimCLR) training process, the change in the **Contrastive Loss** function across epochs is presented in Figure 1.

Figure 1: SimCLR Pre-Training Contrast Loss Curve



	Run ↑	Smoothed	Value	Step	Relative
●	version_72	2,434	2,5131	999	33.4 min

As seen in Figure 1, the contrastive loss value decreases continuously throughout training. This decrease indicates that the SimCLR model has successfully learned strong and distinctive visual representations from unlabeled images. The steady decrease in the loss value indicates that the model effectively grasps the concepts of visual similarity and dissimilarity.

4.2. Classification Performance and Evaluation

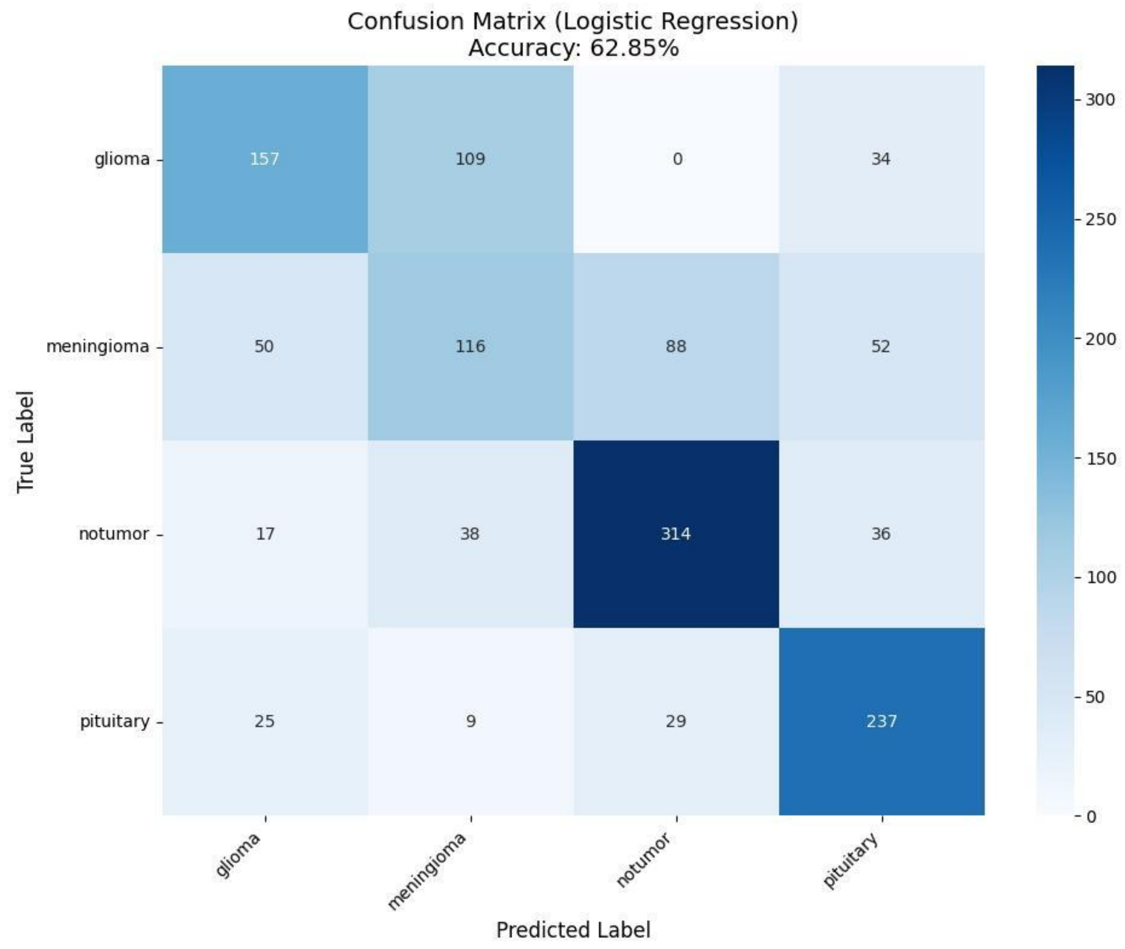
We evaluated the performance of the Logistic Regression classifier on the test set, using the feature extractor (backbone) obtained from pre-training.

Overall Accuracy: Our classification model achieved an overall **accuracy of 62.85%** on the test set.

This accuracy is highly promising, especially considering it was achieved using only **10 examples (10-shot)** for each class. This result clearly demonstrates that our self-supervised pre-training approach can extract meaningful and generalizable features even with limited labeled data. The model has proven its ability to recognize new tumor subtypes with a small number of samples.

To further examine the model's performance on a class-by-class basis, a **Confusion Matrix** was generated. This matrix is presented in Figure 2.

Figure 2: Logistic Regression Classifier Complexity Matrix (Accuracy: 62.85%)



The confusion matrix in Figure 2 illustrates the number of correct (values on the diagonal) and incorrect classifications for each class:

- **class_glioma:** 157 examples were correctly classified, while 109 were misclassified as meningioma, 0 as notumor, and 34 as pituitary.
- **class_meningioma:** 116 examples were correctly classified, while 50 were misclassified as glioma, 88 as notumor, and 52 as pituitary.
- **class_notumor:** 314 examples were correctly classified, while 17 were misclassified as glioma, 38 as meningioma, and 36 as pituitary.
- **class_pituitary:** 237 examples were correctly classified, while 25 were misclassified as glioma, 9 as meningioma, and 29 as notumor.

The confusion matrix indicates that the model performs best on **class_notumor** and **class_pituitary**, achieving relatively high correct predictions. However, a significant amount of confusion is observed between **class_glioma** and **class_meningioma**, suggesting that their visual features might be more similar or the training data was not sufficient to effectively distinguish between them.

Overall, the accuracy of **62.85%** highlights the promise of our logistic regression-based approach, especially considering it operates in a challenging multi-class medical classification setting. The model's confusion patterns emphasize the need for improved representation learning or additional data to better separate similar tumor types like glioma and meningioma.

5. Discussion

Our project has yielded significant findings in the field of brain tumor detection and classification, particularly focusing on scenarios with limited labeled data. The results we've obtained and the challenges encountered are critical for understanding the project's strengths and future potential.

5.1. Comparison of Results with Other Studies in the Literature

Numerous deep learning studies in the literature have achieved over 90% accuracy in brain tumor classification by utilizing large and comprehensively labeled datasets. However, the fundamental difference of our project lies in its adoption of the more challenging 10-shot learning paradigm. The 62.85% accuracy we achieved on the test set is quite promising for a model trained with only 10 labeled examples per class. Considering that traditional deep learning models would suffer from severe overfitting or fail to demonstrate significant performance when trained with so little data, this result proves the effectiveness of our project's solution to the data scarcity problem. The self-supervised learning (SimCLR) pre-training phase enabled the model to learn general and discriminative visual representations from unlabeled data, thereby reducing its dependence on limited labeled data.

5.2. Challenges Encountered

The biggest challenge for the project was the **scarcity of labeled data**, stemming from the inherent nature of medical datasets. This made the direct application of traditional deep learning models difficult.

5.3. Error Analysis

Upon examining the confusion matrix (Figure 2), it's evident that the model experienced significant confusion, particularly between **class_glioma** and **class_meningioma**. Possible reasons for these misclassifications could be:

- **Visual Similarity:** Glioma and Meningioma tumor types might share certain common or very similar morphological and textural features in MR images, making it difficult for the model to distinguish between these two classes.
- **Lack of Sample Variety:** Using only **10 labeled examples per class** might have been insufficient for the model to fully learn the variations and subtle distinguishing features of these tumor types. Different appearances within a class (tumor size, location, shape, etc.) might not have been adequately represented in such a small sample.
- **Limitations in the Feature Extractor:** While the general features learned through SimCLR are robust, they might not be optimal for capturing the subtle differences between these two classes when training a Logistic Regression classifier with such a limited number of labeled examples.

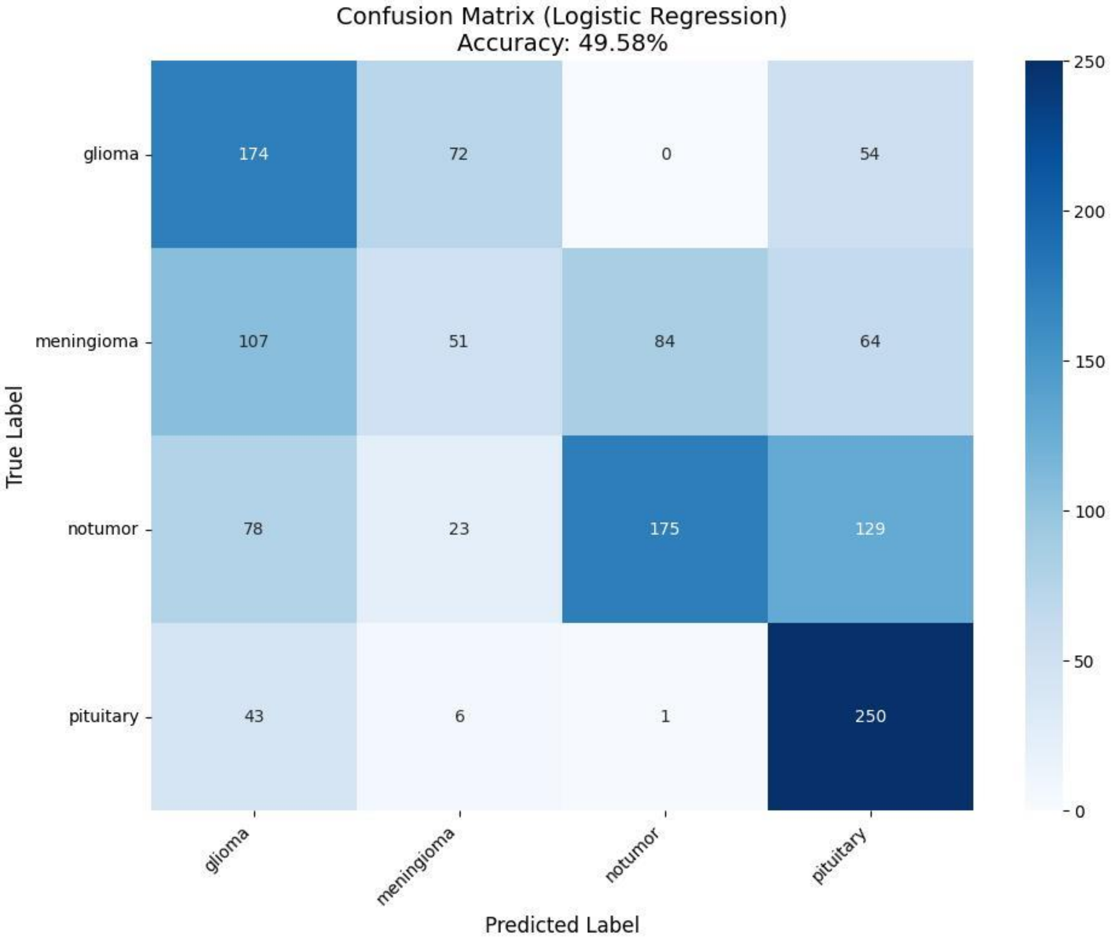
6. Comparisons and Conclusion

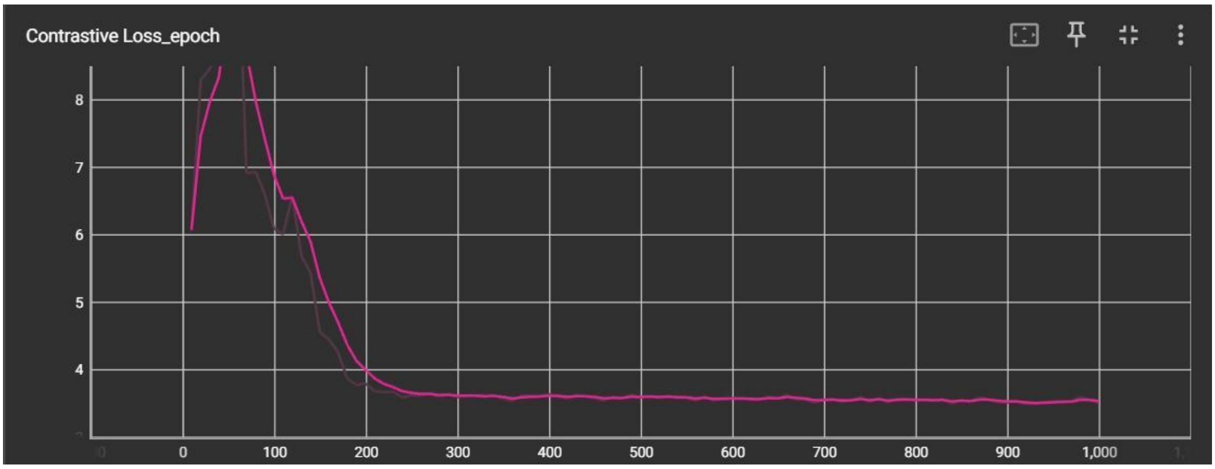
6.1. Comparisons

In this section, the classification performance of our model is analyzed in comparison with two different Logistic Regression-based models. The models are shown below:

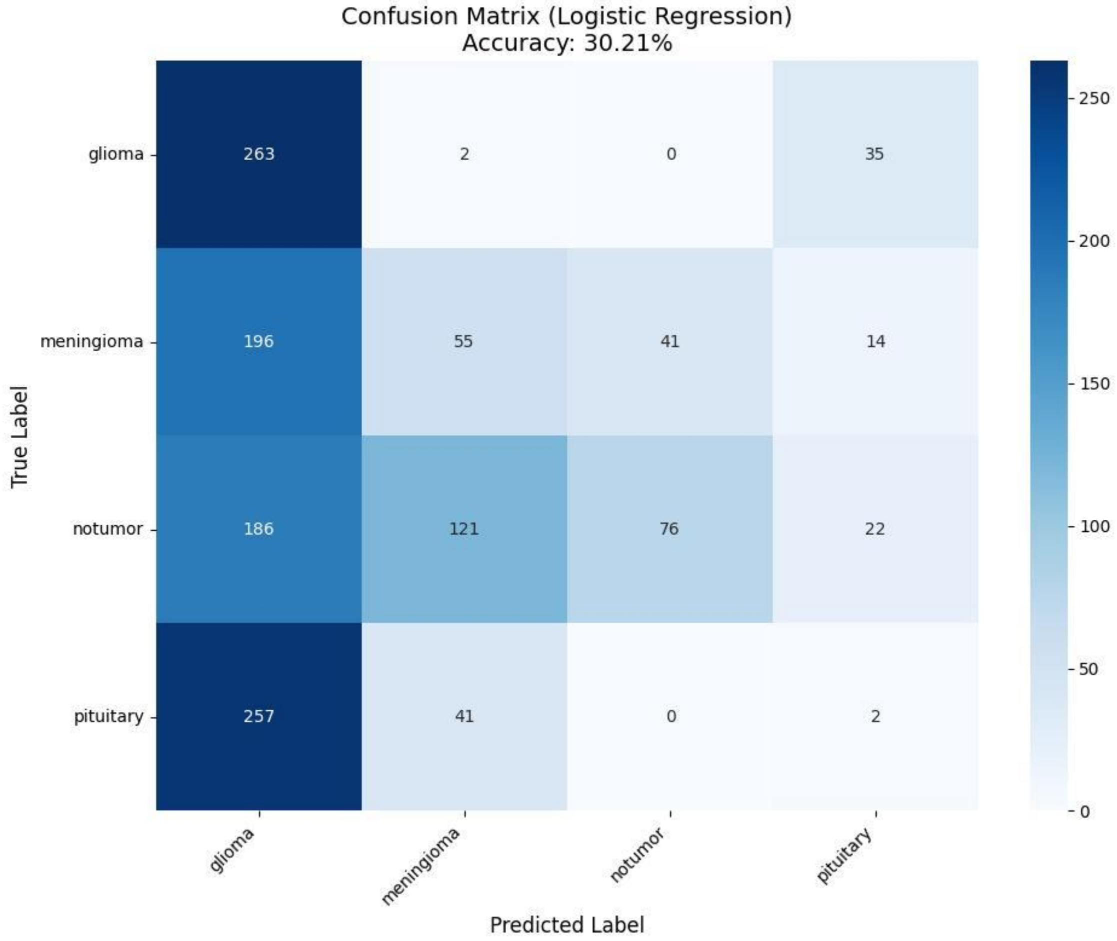
1. Mobilenet_V2
2. DenseNet121

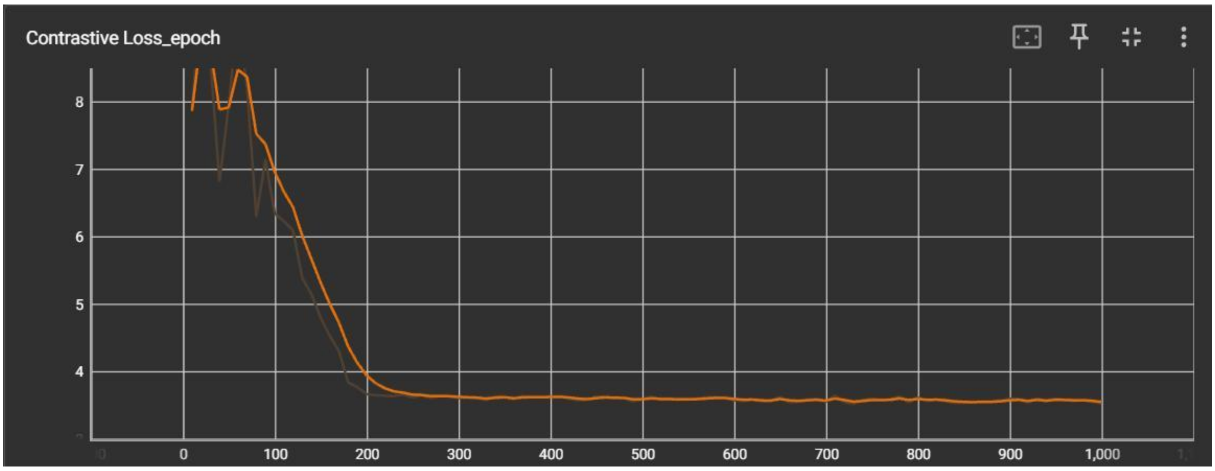
Mobilenet_V2:


Confusion Matrix (Logistic Regression)
Accuracy: 49.58%
True Label
Predicted Label
glioma
meningioma
notumor
pituitary
174
72
0
54
107
51
84
64
78
23
175
129
43
6
1
250
250
200
150
100
50
0


Contrastive Loss_epoch
8
7
6
5
4
0
100
200
300
400
500
600
700
800
900
1,000

DenseNet121：





We compared the performance of three Logistic Regression-based models trained under different conditions. The confusion matrices and overall accuracies are analyzed below:

Our Model Accuracy(Model A): 62.85%

Mobilenet_V2 Accuracy(Model B): 49.58%

DenseNet121 Accuracy(Model C): 30.21%

Per-Class Evaluation

1. Glioma Class

- **Model A**: Correct: 157 | Misclassified as meningioma: 109 | pituitary: 34
- **Model B**: Correct: 174 | Misclassified as meningioma: 72 | pituitary: 54
- **Model C**: Correct: 263 | Misclassified as meningioma: 2 | pituitary: 35
  Model C shows the best glioma classification, but this comes with significant trade-offs in other classes.

2. Meningioma Class

- **Model A**: Correct: 116 | Misclassified as glioma: 50 | notumor: 88 | pituitary: 52
- **Model B**: Correct: 84 | Misclassified as glioma: 107 | notumor: 64 | pituitary: 64
- **Model C**: Correct: 55 | Misclassified as glioma: 196 | notumor: 41 | pituitary: 14
  Model A performs best on meningioma. Model C suffers significant confusion with glioma.

3. Notumor Class

- **Model A**: Correct: 314 | Misclassified as glioma: 17 | meningioma: 38 | pituitary: 36
- **Model B**: Correct: 175 | Misclassified as glioma: 78 | meningioma: 23 | pituitary: 129
- **Model C**: Correct: 76 | Misclassified as glioma: 186 | meningioma: 121 | pituitary: 22
  Model A is significantly better in classifying notumor, while Model C is the weakest.

4. Pituitary Class

- **Model A**: Correct: 237 | Misclassified as glioma: 25 | meningioma: 9 | notumor: 29
- **Model B**: Correct: 250 | Misclassified as glioma: 43 | meningioma: 6 | notumor: 1
- **Model C**: Correct: 2 | Misclassified as glioma: 257 | meningioma: 41 | notumor: 0
  Model B performs best. Model C nearly fails to recognize pituitary cases, heavily misclassifying them as glioma.

Summary

- **Model A** achieves the **best overall performance** with the highest accuracy and more balanced predictions.
- **Model B** improves slightly in **glioma** and **pituitary** detection compared to Model A, but loses significant accuracy on **meningioma** and **notumor**.
- **Model C** is **highly biased towards glioma**, classifying most samples as glioma regardless of their true class. Although its glioma prediction count is highest, it **drastically underperforms** on other classes, leading to the lowest overall accuracy (**30.21%**).

6.2. Conclusion

This project has addressed brain tumor detection and classification by offering an innovative solution to one of the most critical challenges in medical imaging: the **scarcity of labeled data**. We successfully achieved our primary goal of accurately detecting tumor presence and classifying tumor types from MRI images with limited data.

Our developed two-stage approach—**self-supervised feature learning with SimCLR** followed by **classification with Logistic Regression under a 10-shot learning paradigm**—enabled us to reach this objective. The model's promising accuracy of **62.85%** using only **10 labeled examples per class** proves the success of the robust and generalizable visual representations learned during the pre-training phase. This result demonstrates that our project offers a viable and effective alternative in few-shot learning scenarios where traditional deep learning models fall short. It holds significant potential, particularly for the early and accurate diagnosis of rare diseases or tumor subtypes.

Future Work

Despite the successes achieved, our project has significant areas that can be further developed and expanded. Here are some suggestions for future work:

- **Utilization of Larger and More Diverse Datasets:**
    - To increase the model's generalizability, further testing and training should be conducted on larger and more diverse medical imaging datasets, representing broader populations and collected from various clinics.
    - Data augmentation techniques or sampling strategies could be developed to address potential class imbalances.
- **Experimentation with Different Model Architectures and Learning Paradigms:**
    - Performance comparisons can be made by testing other self-supervised learning approaches beyond SimCLR, such as MoCo or BYOL, or more advanced few-shot learning algorithms like Prototypical Networks or Relation Networks.
    - More complex classifiers could be explored instead of Logistic Regression after feature extraction, while still keeping the risk of overfitting low.
- **Optimization for Real-Time Applications:**
    - Optimization techniques such as **model compression**, **quantization**, or **pruning** could be applied to enable faster inference from the developed model.
    - The model's feasibility for deployment on embedded systems or **edge computing** devices could be investigated.
- **Expansion for Other Tumor Types or Diseases:**
    - The project's scope could be extended to include the classification of different subtypes of brain tumors (e.g., various glioma grades) or other neurological diseases.
    - The clinical value of the model could be enhanced by integrating additional computer vision tasks, such as **tumor segmentation** (identifying the location and boundaries of the tumor in an image), instead of solely focusing on classification.

7. References

https://dergipark.org.tr/tr/download/article-file/3806798

https://www.kaggle.com/code/anandhuh/image-classification-using-cnn-for-beginners

https://dergipark.org.tr/tr/download/article-file/2145179

https://earsiv.batman.edu.tr/server/api/core/bitstreams/2ae1f139-43ad-430a-b3b1-c4ed2750be6b/content