Brain tumor classification based on magnetic resonance imaging using convolutional neural networks (CNN).

Milena Biernacka, Patrycja Lewczuk, Sara Świętek

21.7.2023

1 Research Objective

The objective of this project is to use Convolutional Neural Networks (CNN) for the automatic classification of brain tumors based on Magnetic Resonance Imaging (MRI) scans. We will explore two classification methods:

- A classifier that identifies 3 types of brain cancer (glioma, meningioma, pituitary adenoma) or their absence,
- Binary classification - identifying the presence or absence of cancer.

In this way, we aim to determine if precise classification is possible not only for the existence of cancer but also for identifying its specific type.

We will use a dataset of images from the Kaggle platform [4]. In addition to the main dataset, we also plan to use other similar datasets to test our model. This approach will allow us to investigate whether the accurate detection of brain tumors is feasible across images of varying qualities. Both the main dataset and the comparison datasets are thoroughly described in Chapter 2.

In the final phase of the project, we will present the results of binary classification using other machine learning algorithms: Support Vector Machine (SVM) and Decision Tree Classifier (DTC). This will enable us to evaluate which method performs best in tumor recognition.

2 Image Dataset

2.1 Main Dataset

Within the project, we will use the "Brain Tumor Classification (MRI)" dataset available on the Kaggle platform [1]. This dataset contains 3264 brain Magnetic Resonance Imaging (MRI) images, each categorized into four classes: glioma, meningioma, pituitary adenoma, and cases where no tumor is present. The creator of the dataset claims that the images have been curated and labeled under the supervision of a specialist doctor. The main dataset includes the following features:

- Images acquired from different sources (varied image qualities).

- Diverse perspectives of brain MRI scans (from the back, from the side, from the top) - see Figure [1].
- Images obtained using various MRI techniques (T1, T2, or FLAIR) - see Figure [2].

The aforementioned characteristics indicating the diversity of the dataset may pose challenges for the predictive model. As part of the project, we aim to determine whether it is still possible to create a model capable of classifying the images with high accuracy despite of mentioned challenges.

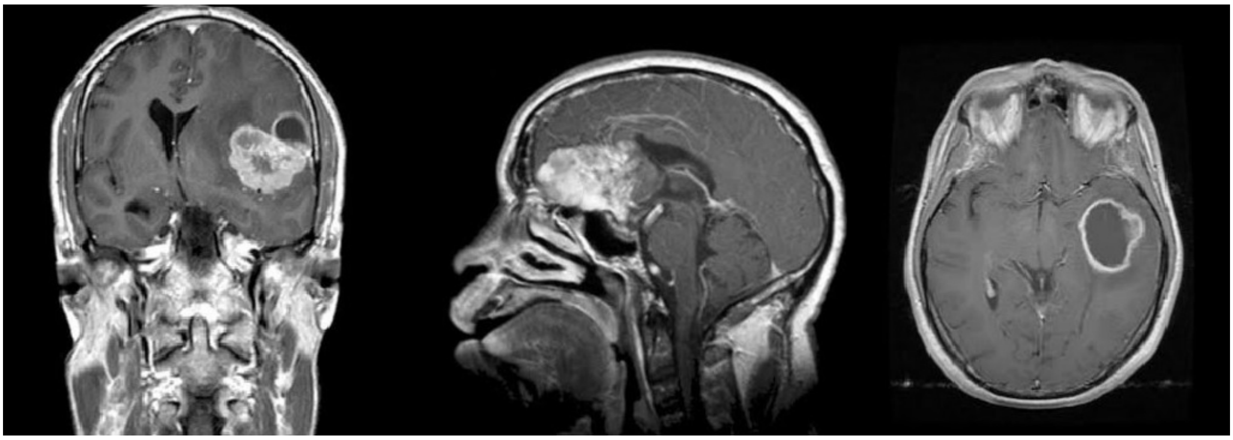

Figure 1: Diverse perspectives of brain MRI scans (from the back, from the side, from the top) [1]

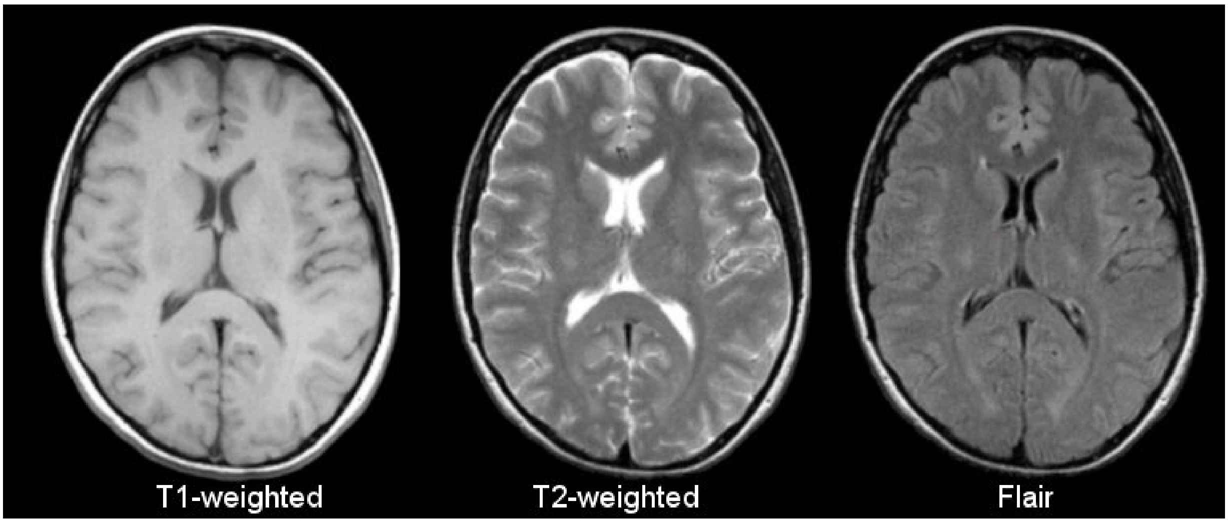

Figure 2: Images obtained using various MRI techniques (T1, T2, or FLAIR) [6]

2.2 Similar Image Datasets

To compare the accuracy of the model trained on the main dataset, we will use other datasets with similar data. We have found similar datasets on Kaggle: [2] (3000 images), [3] (253 images), and [5] (400 images). These datasets are divided into "cancer" and "no cancer" labels, so when comparing with other image datasets, we will use only binary models, without division into cancer categories.

Contrary to our main dataset, images from [2] and [3] are taken from a single perspective (from the top), making them less diverse. This leads to the question: If we train a predictive model on a less diverse dataset, will it yield better results? We will attempt to answer this question in this project as well.

Images from [5] are of lower quality compared to others. An example image from this dataset is shown in Figure 3. Using this dataset, we want to observe how the image quality affects the accuracy of the model's predictions.

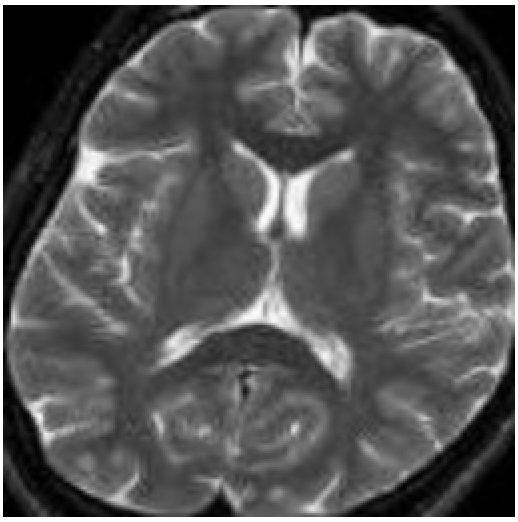

Figure 3: Example image from the dataset [5]

3 Model Training

3.1 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are a type of deep neural networks that are widely used in image analysis and visual signal processing. This method is particularly effective in tasks such as object recognition, image segmentation, and classification.

The fundamental components of CNN are convolutional layers, which scan the input image using convolutional filters (kernels). These filters perform convolution operations on the input data, extracting local patterns and characteristic features from the image. Convolutional layers are capable of automatically learning relevant features of the image, such as edges, textures, or shapes, based on which classification is performed.

Another essential element of CNN is the pooling layer (e.g., max-pooling layer), which serves to reduce the spatial size of the feature maps obtained from the convolutional layers. This allows convolutional operations to be performed on increasingly abstract features, and it also makes the model more robust to shifts and deformations in the image.

After the convolutional and pooling layers, fully connected layers and the output layer are typically added. The output layer's result corresponds to the classification of the input image into specific classes, such as benign tumor, malignant tumor, or healthy brain.

During the training of CNN, the network's weights are updated based on the backpropagation algorithm, which minimizes the difference between the predicted outputs and the true class labels.

Thanks to their ability to automatically extract features from images, CNNs have become highly effective in various fields, including medicine, particularly in classifying brain tumors based on magnetic resonance imaging (MRI) images.

3.2 Comparison of models with different parameters for classifying three types of brain cancer

The optimizer in the context of CNN (Convolutional Neural Networks) models is an optimization algorithm used to update the model's weights to minimize the cost function during the learning process. During the training of a CNN model, the optimizer utilizes gradients of the cost function with respect to the model's weights to make appropriate weight adjustments.

The investigated optimizers are:

- Adadelta: an adaptive optimization algorithm aimed at reducing the significance of manual hyperparameter tuning. It automatically adjusts the learning rate based on the history of weight updates. Adadelta also supports dynamically adapting the learning rate for each parameter.
- Adagrad: an optimizer that adjusts the learning rate for each parameter based on its previous updates. This means that parameters that rarely appear have a higher learning rate, while parameters that frequently appear have a lower learning rate.
- Adam (Adaptive Moment Estimation): combines the benefits of adaptive learning rate and momentum. It uses estimates of the first moment (mean of the gradient) and the second moment (mean of the squared gradient) to update the weights.
- RMSprop (Root Mean Square Propagation): also adjusts the learning rate for each parameter but only considers the mean of the squared gradient.
- SGD (Stochastic Gradient Descent): a simple and basic optimization algorithm. It updates the weights in the direction opposite to the gradient of the cost function.

Each of these optimizers has its own features and behaviors that may work well in different situations. The choice of optimizer depends on the nature of the problem, data, and preferences.

In the following chart, we compare the validation accuracy of models using different optimizers:

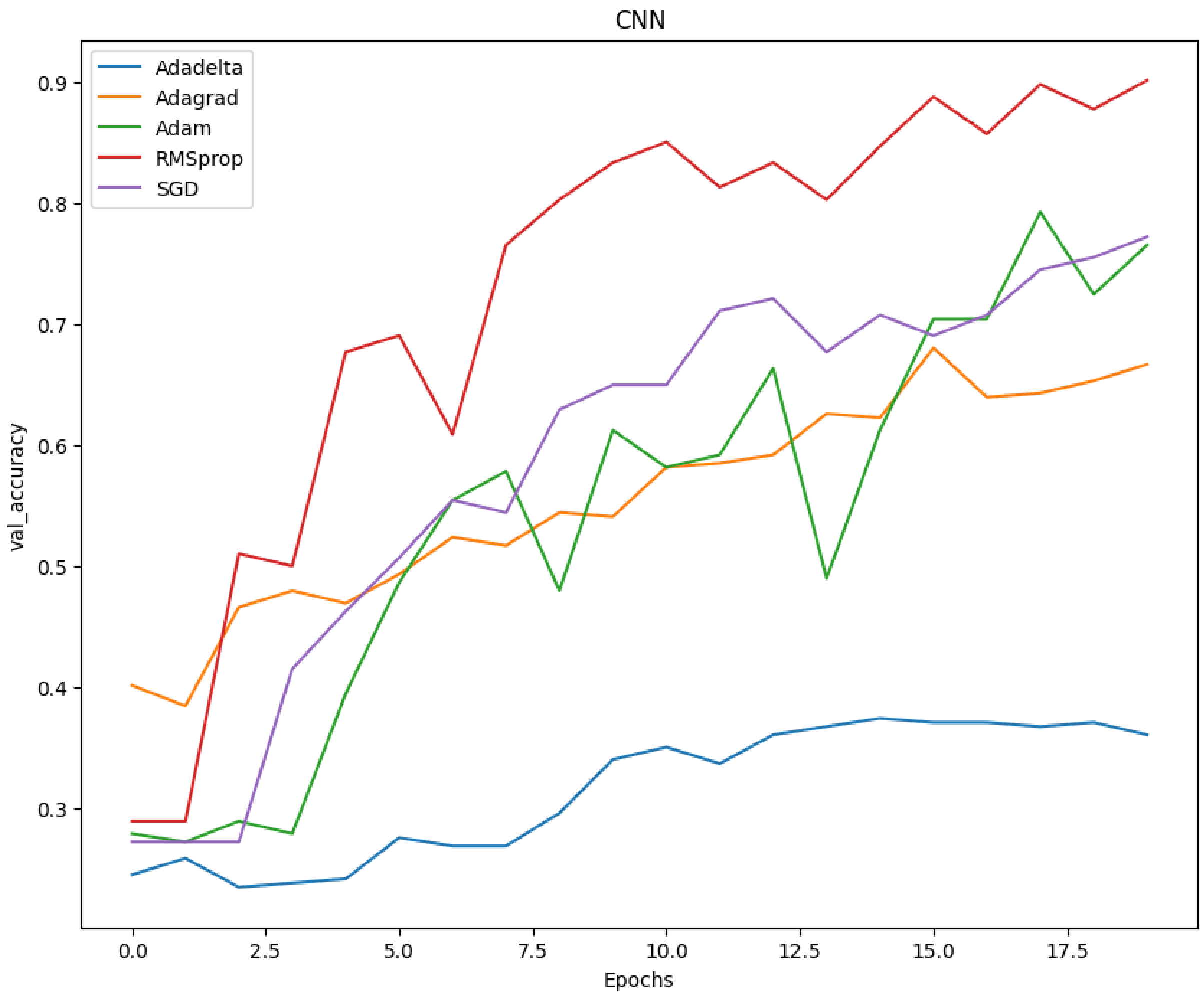

Figure 4: Validation accuracy chart for CNN models with different optimizer parameters

The model with the Adadelta optimizer consistently shows the lowest validation accuracy across all epochs. Let's further analyze this model in the context of comparing the accuracy between the training and test sets. For both sets, we observe low accuracy and significant discrepancies in trends, indicating the presence of overfitting.

We also plot the loss trend of the model for the train and test sets. We notice a significant difference between them - the trend for the train set is strongly decreasing, while for the test set, it remains nearly constant. This indicates a lack of generalization ability to new, previously unseen data.

Table 1: Final classification results for different optimizers

optimizer	accuracy
Adadelta	0.37
Adagrad	0.69
Adam	0.74
RMSprop	0.91
SGD	0.78

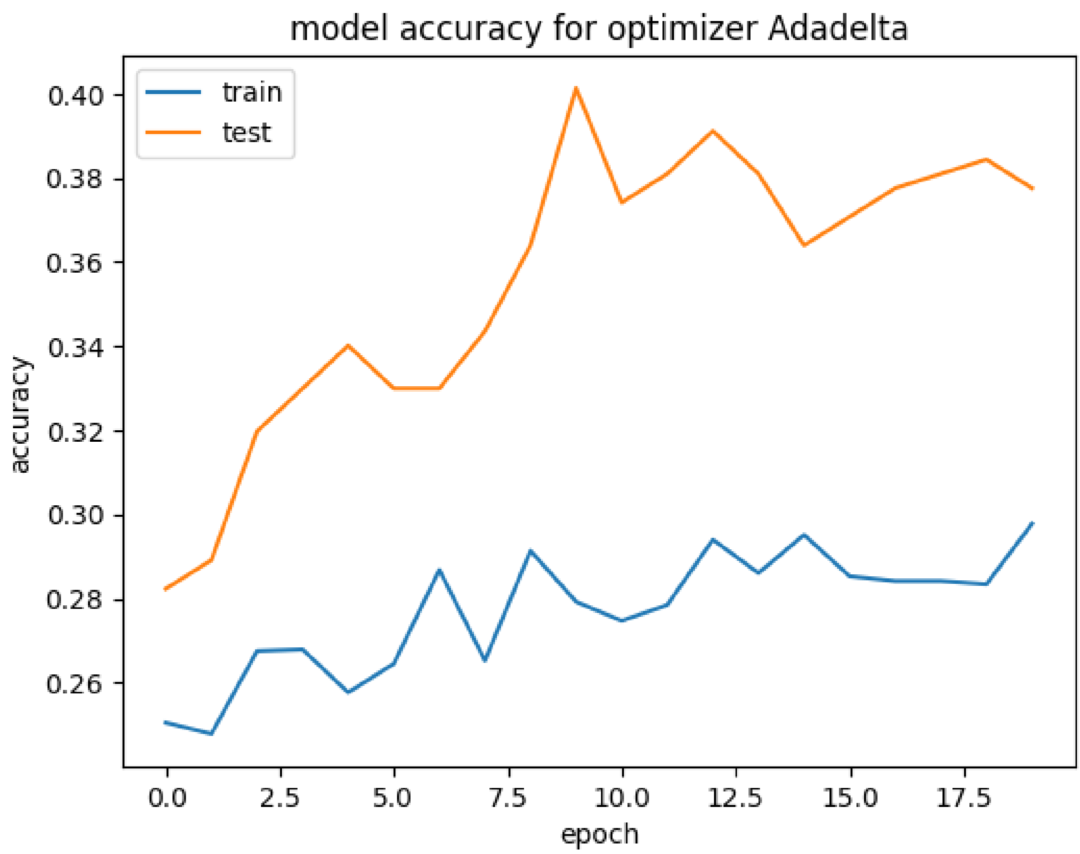

Figure 5: Accuracy chart for the model with Adadelta optimizer

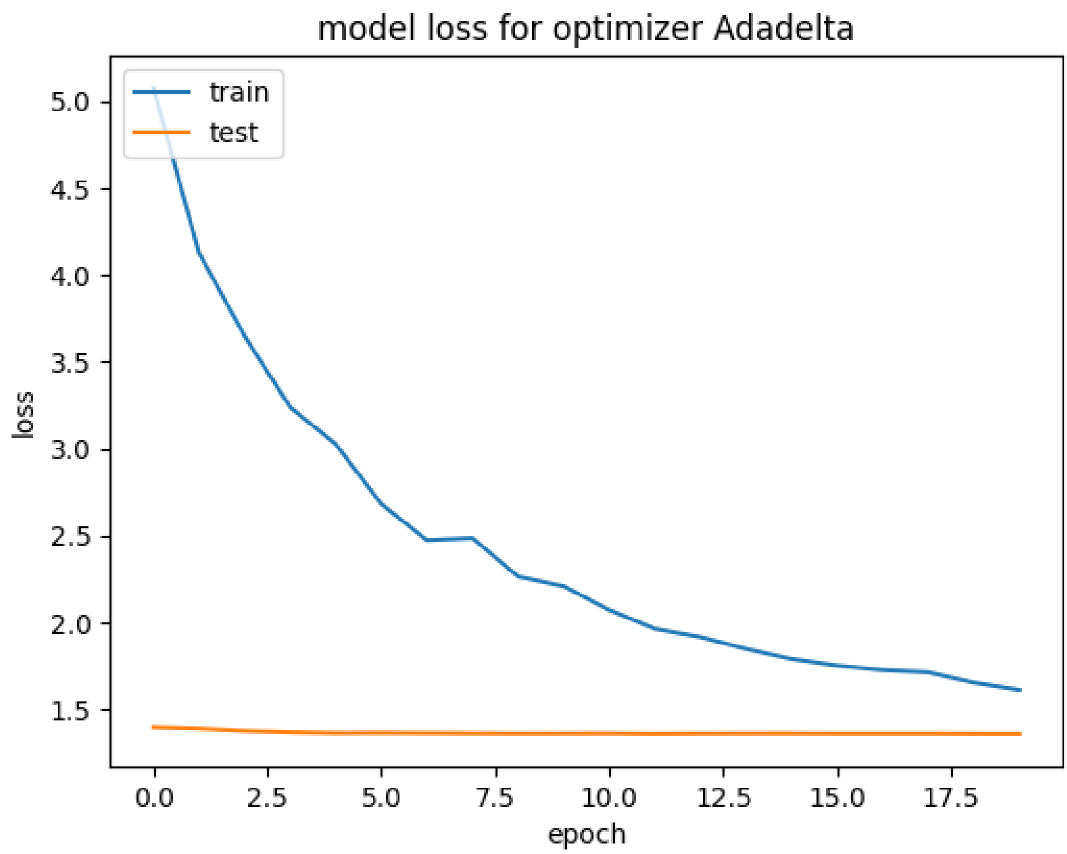

Figure 6: Loss chart for the model with Adadelta optimizer

Below, we provide a visualization comparison of predictions made on several MRI images for Adadelta optimizer and RMSprop, which achieved an accuracy of 91% after 20 epochs.

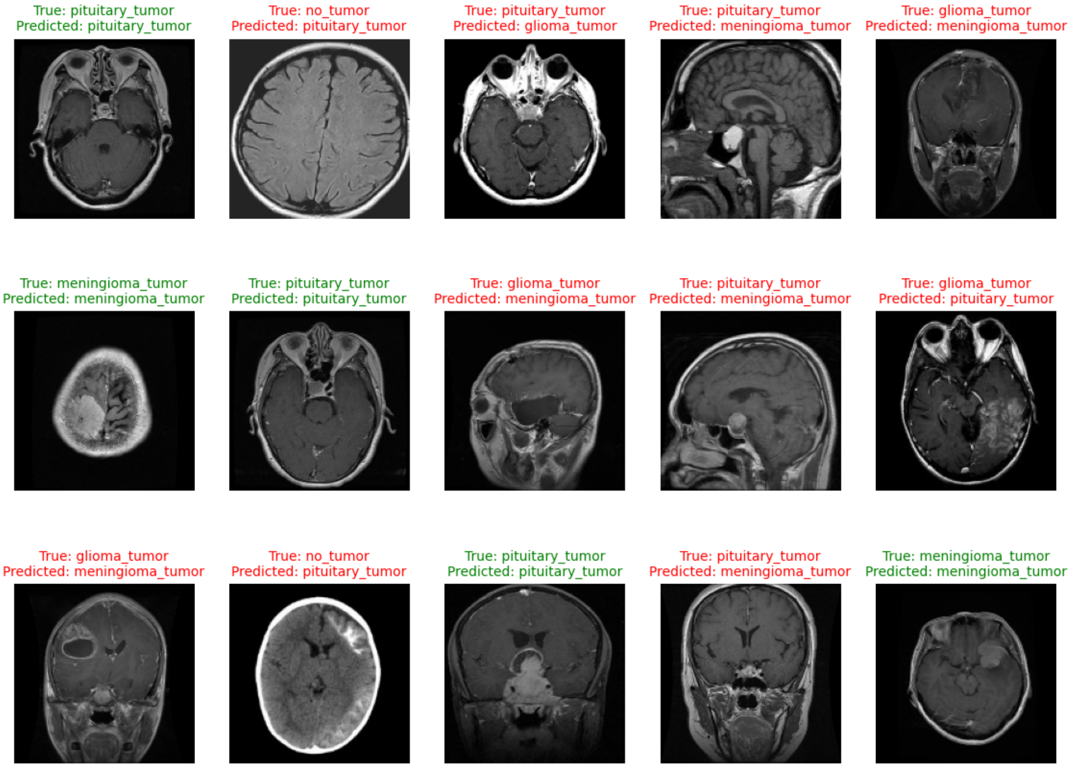

Figure 7: Predictions of the model with Adadelta optimizer

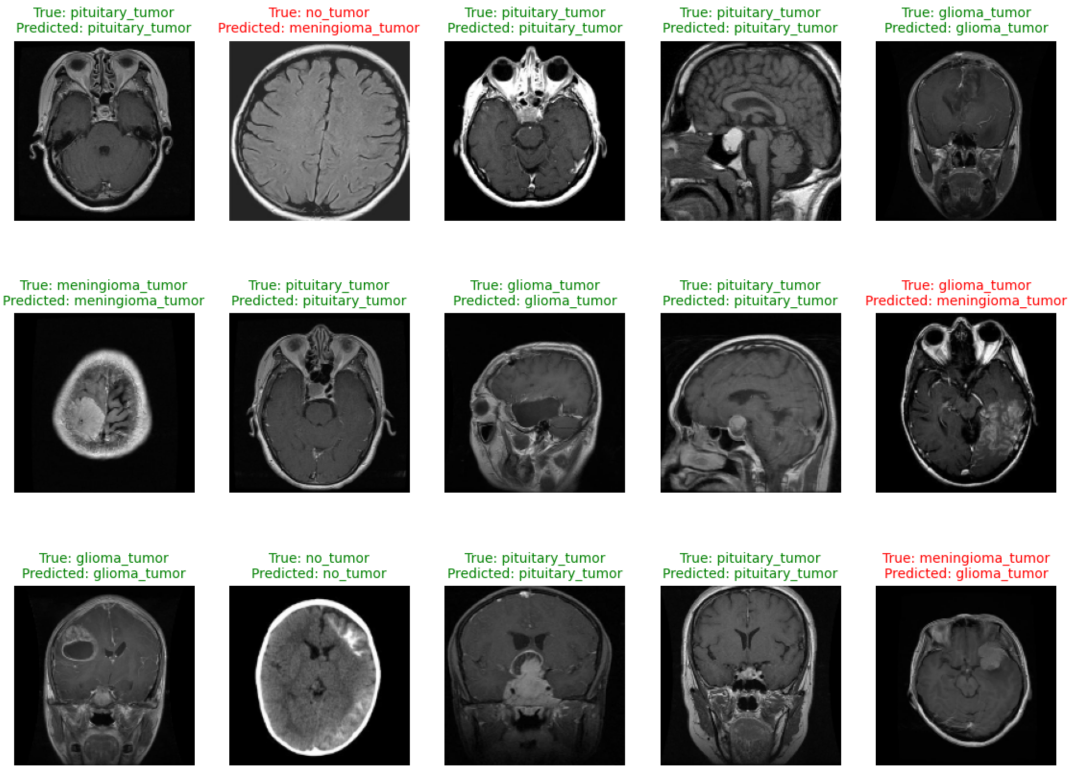

Figure 8: Predictions of the model with RMSprop optimizer

3.3 Binary Classification for the Main Dataset [1]

Binary classification is a type of machine learning task where the goal is to assign input data to one of two classes. In our dataset, we added an additional variable where images with three types of tumors are labeled as "tumor," and images of a healthy brain are labeled as "no tumor." This variable will be the target for binary classification.

Convolutional Neural Networks (CNN) with the "Adam" optimizer were used for binary classification. Comparing the results for binary classification (accuracy 0.93) and classification of brain tumor types (accuracy 0.74), a significant improvement in the predictive model can be noticed. This is because recognizing the type of brain tumor requires a more detailed analysis of the tumor's structure, which is more challenging than merely distinguishing between healthy and tumor images.

The table below presents the results of binary classification for the dataset [1]:

Table 2: Results of Binary Classification for the Dataset [1]

	precision	recall	f1-score	support
0	0.79	1.00	0.88	105
1	1.00	0.90	0.95	289
accuracy			0.93	394
macro avg	0.89	0.95	0.92	394
weighted avg	0.94	0.93	0.93	394

4 Testing the Model on Other Datasets

We will apply the binary classification model to two additional datasets [2] [3]. Both datasets contain brain MRI images captured solely from a top-down perspective, resulting in less variation compared to the previous dataset, which included scans from three different perspectives. Half of the images in these new datasets depict brains with tumors, while the other half shows healthy brains.

Below are the summary results of the model's predictions on these datasets:

Table 3: Classification Results for Dataset [2]

	precision	recall	f1-score	support
0	0.63	0.94	0.75	1500
1	0.89	0.45	0.59	1500
accuracy			0.69	3000
macro avg	0.76	0.69	0.67	3000
weighted avg	0.76	0.69	0.67	3000

Table 4: Classification Results for Dataset [3]

	precision	recall	f1-score	support
0	0.56	0.94	0.70	98
1	0.93	0.54	0.69	155
accuracy			0.70	253
macro avg	0.75	0.74	0.70	253
weighted avg	0.79	0.70	0.69	325

5 Model on a Uniform Dataset [2]

To address the question of whether training the model on images from the same perspective improves prediction performance, we utilized an appropriate dataset [3]. The model was tested on a smaller but also uniform dataset [2], which showed very high accuracy of 0.97.

The accuracy on the non-uniform dataset decreased, as expected, to 0.73.

We also tested the model on the [5] dataset, which contains images from the same perspective but with significantly lower quality. Here, the accuracy significantly decreased, showing a result of 0.79.

Table 5: Classification Results for Uniform Dataset [2]

	precision	recall	f1-score	support
0	1.00	0.92	0.96	98
1	0.95	1.00	0.97	155
accuracy			0.97	253
macro avg	0.98	0.96	0.97	253
weighted avg	0.97	0.97	0.97	325

Table 6: Classification Results for Non-Uniform Dataset [1]

	precision	recall	f1-score	support
0	0,29	0.67	0.41	395
1	0.93	0.74	0.83	2475
accuracy			0.73	2870
macro avg	0.61	0.70	0.62	2870
weighted avg	0.84	0.73	0.77	2870

Table 7: Classification Results for Low-Quality Dataset [5]

	precision	recall	f1-score	support
0	0.76	0.75	0.76	170
1	0.82	0.82	0.82	230
accuracy			0.79	400
macro avg	0.79	0.79	0.79	400
weighted avg	0.79	0.79	0.79	400

6 Comparison of CNN Quality with Other Models

The final step of our project will involve comparing the quality of CNN models with classical machine learning algorithms: Support Vector Machine (SVM) and Decision Tree Classifier (DTC). By conducting this comparison, we will be able to determine if using CNN was the best choice compared to other methods.

6.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm used for data classification and regression tasks. The main idea behind SVM is to find an optimal hyperplane that separates data points of different classes. In the context of brain tumor detection on MRI images, the input data is represented by image features such as textures, shapes, or pixel intensities. SVM analyzes these features and creates a model that can distinguish healthy brain areas from areas containing tumors.

6.2 Model Trained on Non-uniform Data [1]

Table 8: Classification results for the non-uniform dataset [1]

	precision	recall	f1-score	support
0	0.82	0.51	0.63	70
1	0.94	0.98	0.96	504
accuracy			0.93	574
macro avg	0.88	0.75	0.80	574
weighted avg	0.92	0.93	0.92	574

Table 9: Classification results for the uniform dataset [3]

	precision	recall	f1-score	support
0	0.47	0.81	0.60	98
1	0.78	0.43	0.56	155
accuracy			0.58	253
macro avg	0.63	0.62	0.58	253
weighted avg	0.66	0.58	0.57	253

6.3 Model Trained on Uniform Dataset [2]

Table 10: Classification Results for the Uniform Dataset [2]

	precision	recall	f1-score	support
0	0.97	0.98	0.98	287
1	0.98	0.97	0.98	313
accuracy			0.98	600
macro avg	0.98	0.98	0.98	600
weighted avg	0.98	0.98	0.98	600

Table 11: Classification Results for Additional Uniform Dataset [3]

	precision	recall	f1-score	support
0	1.00	0.97	0.98	98
1	0.98	1.00	0.99	155
accuracy			0.99	253
macro avg	0.99	0.98	0.99	253
weighted avg	0.99	0.99	0.99	253

Table 12: Classification Results for Non-Uniform Dataset [1]

	precision	recall	f1-score	support
0	0.15	0.70	0.25	395
1	0.89	0.37	0.53	2475
accuracy			0.42	2870
macro avg	0.52	0.54	0.39	2870
weighted avg	0.79	0.42	0.49	2870

6.4 Decision Tree Classifier (DTC)

Decision Tree Classifier is a machine learning algorithm that can also be applied for brain tumor detection on magnetic resonance imaging (MRI) scans. The classifier relies on a tree-like decision structure where each node represents a feature test, and each leaf corresponds to a class label.

6.5 Model Trained on Non-Uniform Dataset [1]

Table 13: Results of classification on non-uniform dataset [1]

	precision	recall	f1-score	support
0	0.84	0.70	0.77	70
1	0.96	0.98	0.97	504
accuracy			0.95	574
macro avg	0.90	0.84	0.87	574
weighted avg	0.95	0.95	0.95	574

Table 14: Results of classification on uniform dataset [2]

	precision	recall	f1-score	support
0	0.61	0.64	0.62	1500
1	0.62	0.60	0.61	1500
accuracy			0.62	3000
macro avg	0.62	0.62	0.62	3000
weighted avg	0.62	0.62	0.62	3000

6.6 Model trained on uniform dataset [2]

Table 15: Results of classification on uniform dataset [2]

	precision	recall	f1-score	support
0	0.73	0.70	0.71	287
1	0.73	0.76	0.75	313
accuracy			0.73	600
macro avg	0.73	0.73	0.73	600
weighted avg	0.73	0.73	0.73	600

Table 16: Results of classification on another uniform dataset [3]

	precision	recall	f1-score	support
0	0.76	0.81	0.78	98
1	0.87	0.84	0.86	155
accuracy			0.83	253
macro avg	0.82	0.82	0.82	253
weighted avg	0.83	0.83	0.83	253

Table 17: Results of classification on non-uniform dataset [1]

	precision	recall	f1-score	support
0	0.12	0.57	0.20	395
1	0.83	0.33	0.47	2475
accuracy			0.36	2870
macro avg	0.47	0.45	0.33	2870
weighted avg	0.73	0.36	0.43	2870

7 Conclusions

Based on the experiments conducted on the selected image datasets, we can draw the following conclusions:

For CNN:

- After testing various optimizers for the classification of three types of brain cancer, **RMSprop** proved to be the best. It achieved an accuracy of 91% on our main dataset [1].
- On the main dataset, **binary classification performed better (93% accuracy) than classifying three types of brain cancer (91% accuracy for RMSprop)**. However, it is worth noting that this difference is not significant, considering that classifying three types of brain cancer involves more labels, which increases the risk of misclassifying tumors into categories.
- When testing the binary model trained on the main dataset [1] on other datasets, the accuracy dropped to 69/70%.
- **The level of dataset diversity has a significant impact on the predictive model's quality.** In our case, after training the binary model on uniform images [2] (images from a single perspective), it achieved 97% accuracy on the test dataset, which is higher than in the diverse model.
- **The quality of images also has a significant impact on the prediction results.** After using the binary model trained on the uniform dataset [2] on the low-quality uniform images [5], we observed an accuracy of 79%.

Comparison of CNN with other methods:

- SVM performs similarly well on both uniform images (98% accuracy) and diverse images (93%), but only within a single dataset. When testing the accuracy of the model [1] on the dataset [2], SVM shows 58% accuracy, while CNN achieves 70%. This result indicates that CNN would be a better choice for training a model to be applied to different types of images.
- DTC performs well on our main dataset [1], achieving an accuracy of 95%. However, on uniform data [2], we observed a significant decrease in accuracy to 73%.

References

[1] S. Bhuvaji, A. Kadam, P. Bhumkar, and S. Dedge. *Brain Tumor Classification (MRI)*. 2020. URL: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri (visited on 05/09/2023).

[2] *Br35H :: Brain Tumor Detection 2020*. 2020. URL: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection (visited on 05/09/2023).

[3] *Brain Tumor MRI Images 17 Classes*. 2023. URL: https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-17-classes (visited on 05/09/2023).

[4] *Kaggle: Your Machine Learning and Data Science Community*. URL: https://www.kaggle.com/ (visited on 05/23/2023).

[5] *MRI Based Brain Tumor Images*. 2021. URL: https://www.kaggle.com/datasets/mhantor/mri-based-brain-tumor-images (visited on 06/14/2023).

[6] *MRI Basics*. URL: https://case.edu/med/neurology/NR/MRI%20Basics.htm (visited on 05/23/2023).