AI Course

Capstone Project   
Final Report

For students (instructor review required)

ⓒ 2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

| Detection and Classification of Brain Tumors from Magnetic Resonance Imaging using Convolutional Neural Networks |
| --- |

23/12/2025

Group 1

Alejandro Gallego Domínguez

Alejandro Lobillo Becerra

Almudena Neva Alejo

Content

1. Introduction

1.1. Background Information

1.2. Motivation and Objective

1.3. Members and Role Assignments

1.4. Schedule and Milestones

2. Project Execution

2.1. Data Acquisition

2.2. Training Methodology

2.2.1. CNN

2.2.2. Duplicate Removal

2.2.3. Outliers Removal

2.2.4. Background Suppression

2.2.5. Data Augmentation

2.2.6 Training Callbacks

2.2.7. Grad-CAM

2.3. Workflow

2.4. System Design

2.5. Supporting tools and resources

3. Results

3.1. Data Preprocessing

3.2. Exploratory Data Analysis (EDA)

3.3. Modeling

3.3.1. Model Architecture and Metrics

3.3.2. Explaining model predictions using Grad-CAM

3.4. Testing and Improvements

3.4.1. Classical algorithm-based approximation

3.4.2. Pretrained Models

3.4.3. Reducing Background influence in Model Training

3.4.4. Dealing with misclassifications with an "inconclusive” category

4. Projected Impact

4.1. Accomplishments and Benefits

4.2. Future Improvements

5. Team Member Review and Comment

6. Instructor Review and Comment

Figures and Tables Index

Figure 1. Example MRI of glioma, meningioma, pituitary tumor and non-tumor from the dataset used in this project (Images: *Tr-gl\_0051.jpg* , *Tr-me\_0161.jpg* , *Te-pi\_0025.jpg* and *Te-no\_0389.jpg* ), p. 6

Figure 2. Schematic representation of the project's motivation and objective. Original figure created with Biorender, p. 7

Figure 3. Operation of a CNN. Image obtained from 4, p. 9

Figure 4. Comparison of the image Te-gl\_0053.jpg before and after background removal, p.10

Figure 5. The data's journey, visualized from end to end. Figure generated with NotebookLM. p. 12

Figure 6. Example of duplicate images found in our dataset in the notumor class. Images: *Tr-noTr\_0006.jpg* and *Te-no\_0171.jpg* p.12

Figure 7. Bar chart of image distribution across classes prior to the elimination of duplicates and outliers, p.13

Figure 8. Bar chart of images distribution across classes after the removal of duplicates and outliers in the training, validation and test sets., p.13

Figure 9. Architecture of the Convolutional neural network used, p.14

Figure 10. Evolution of accuracy and loss in the training and validation sets during the different training epochs, p.15

Figure 11. Confusion matrix of the classification model, where the rows represent the actual labels and the columns the predicted ones, with the color intensity being proportional to the number of observations of each actual-predicted combination, p.15

Table 1. Model metrics by class, global and average, p.16

Figure 12. Grad-CAM visualization. Original image, activation map and overlay of glioma*.Te-gl\_0167.jpg:* Pred=glioma, Conf=0.472, Conf=0.472, p.16

Figure 13. Grad-CAM visualization. Original image, activation map, and overlay of glioma predicted as meningioma. *Tr-gl\_1211.jpg:* Pred=meningioma, Conf=0.373, p.17

Figure 14. Grad-CAM visualization. Original image, activation map, and overlay of meningioma. *Tr-me\_1169.jpg* : Pred=meningioma, Conf=0.468, p. 17

Figure 15. Grad-CAM visualization. Original image, activation map, and overlay of meningioma. *Tr-me\_0669.jpg* : Pred=meningioma, Conf=0.471, p. 17

Figure 16. Grad-CAM visualization. Original image, activation map, and overlay of NoTumor. *Te-no\_0027.jpg:* Pred=notumor, Conf=0.475, p. 18

Figure 17. Grad-CAM visualization. Original image, activation map and overlay of NoTumor. Te-no\_0238.jpg: Pred=notumor, Conf=0.475, p. 18

Figure 18. Grad-CAM visualization. Original image, activation map, and overlay of Pituitary. *Te-piTr\_0001.jpg* : Pred=pituitary, Conf=0.475, p.18

Figure 19. Grad-CAM visualization. Original image, activation map, and overlay of Pituitary. *Tr-pi\_1432.jpg:* Pred=pituitary, Conf=0.465, p.19

Figure 20. Grad-CAM visualization. Original image, activation map, and overlay of notumor *Te-noTr\_0002.jpg,* in which the old model uses the background to make the decision instead of brain tissue. P.20

Figure 21. Example of prediction confidence distribution in a misclassification, p.20

Figure 22. Example of prediction confidence distribution in a correct prediction, p.21

Figure 23. Comparison of results between the base model and the modified model with a threshold of 97.5%, p.21

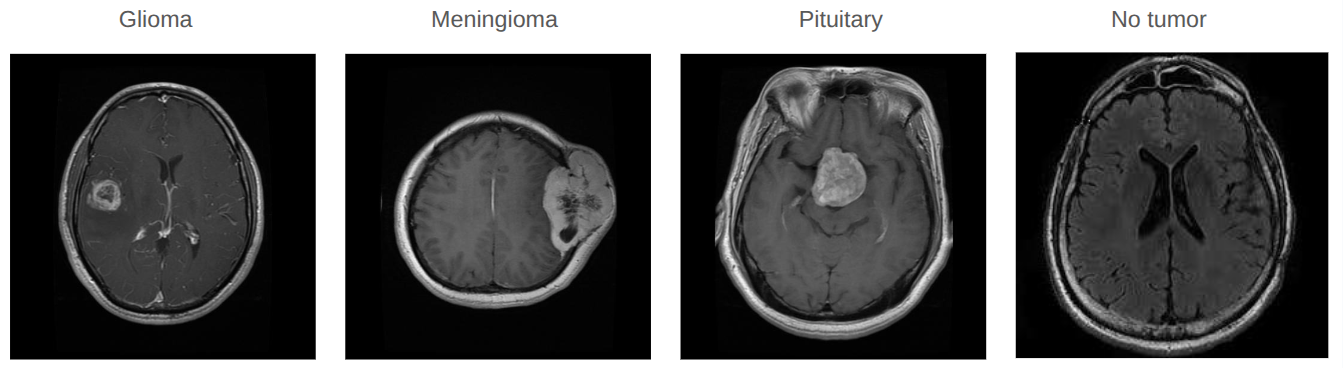
1. Introduction

1.1. Background Information

**Brain tumors** are associated with low survival rates; only 20.8% of men and 24.2% of women survive more than 5 years after **diagnosis** 1.Furthermore, these cancers are difficult to treat because they are protected by the skull, cerebrospinal fluid, and the blood-brain barrier, a defense mechanism that prevents toxins and pathogens from reaching the brain, but which also hinders the passage of some treatments. These same factors make detection difficult, the permeability of the blood-brain barrier affects the entry of contrast agents used in magnetic resonance imaging, especially in the earliest stages of the disease when it is usually less permeable 1.

**Magnetic resonance imaging (MRI)** is the standard method for detecting, classifying, and monitoring brain tumors. Classifying brain tumors is complex because they are so diverse. Many types exist depending on the location and cell type from which they originate, yet they can have a very similar appearance. There are primary tumors, originating directly in the brain, ranging from benign tumors to highly aggressive tumors such as glioblastomas, and brain metastases, which have spread to the brain from the original tumor, such as breast cancer and melanoma. 1,2

Despite the difficulty in distinguishing between gliomas, meningiomas, and pituitary tumors using MRI, we can consider the following characteristics. Gliomas are usually diffuse and located within brain tissue, as they originate from glial cells. Meningiomas, on the other hand, originate in the meninges, are located outside the brain, and are usually well-defined and delimited masses. Finally, pituitary tumors originate in the pituitary gland at the base of the skull (Figure 1)2 .

Figure 1. Example MRI of glioma, meningioma, pituitary tumor and non-tumor from the dataset used in this project (Images: *Tr-gl\_0051.jpg* , *Tr-me\_0161.jpg* , *Te-pi\_0025.jpg* and *Te-no\_0389.jpg*).

**Artificial intelligence**, specifically **Deep Learning (DL)** algorithms, is revolutionizing the world of medical imaging. Numerous studies have demonstrated the potential of Machine Learning (ML) and DL algorithms in diagnostic imaging, including for early disease detection. However, some limitations to their clinical application remain. One limitation is the high heterogeneity in MRI acquisition protocols, as there is no standardization across different hospitals. Another is the "black box" that represents the characteristics of the models for clinicians, thus limiting their acceptance. Furthermore, legal and data protection regulations hinder their implementation in hospitals, as cloud-based frameworks cannot be used. Instead, federated learning frameworks must be employed to train AI models across multiple devices and servers without transferring sensitive data, demonstrating responsible use while respecting privacy. These limitations can be overcome through collaboration between different professionals such as clinicians, data scientists, biomedical engineers, and bioinformaticians, forming **multidisciplinary teams** to address these challenges 2.

1.2. Motivation and Objective

There are many types of brain tumors with different biological behaviors and prognoses, ranging from meningiomas or pituitary tumors, which are mostly benign, to more aggressive tumors such as high-grade gliomas. Accurately classifying them using a non-invasive technique like magnetic resonance imaging (MRI) is essential for determining the appropriate treatment strategy for each patient. Artificial intelligence classification algorithms could contribute by assisting radiologists in the diagnostic process (Figure 2).

Therefore, we propose to develop an artificial intelligence model capable of detecting the presence of a brain tumor, as well as the type of tumor (glioma, meningioma, and pituitary tumor), from a magnetic resonance imaging (MRI) scan. To achieve this, we will use convolutional neural networks and explain the predictions using activation maps (Grad-CAM).

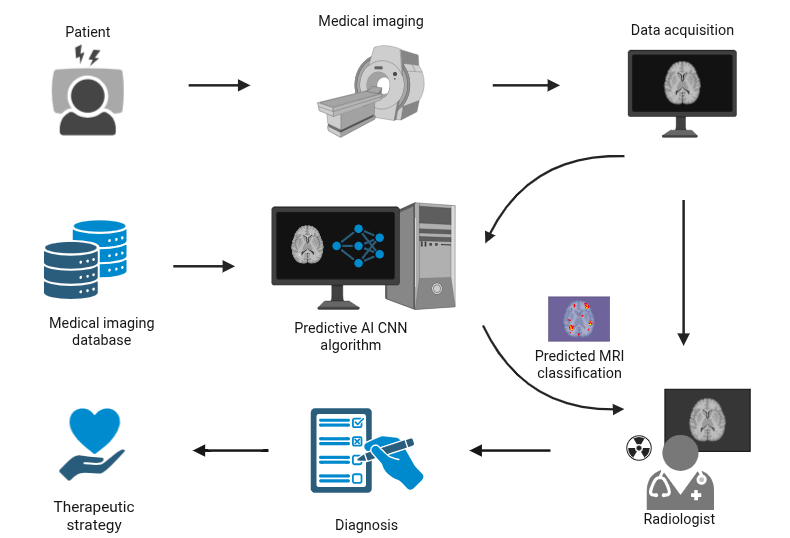


Figure 2. Schematic representation of the project's motivation and objective. Original figure created with Biorender.

1.3. Members and Role Assignments

No roles were defined; instead, everyone participated in different aspects of this project as a way to reinforce and apply everything they had learned in the course. A task list was created and updated as the project needed, and any member could select a task to work on. All team members contributed equally to the project's development, integrating their specific expertise.

1.4. Schedule and Milestones

* At the end of October, the working group was created and the first contact was made.
* In mid-November, brainstorming and discussion of the project topic began.
* The search and evaluation of datasets was carried out
* November 25th, Discord meeting and topic selection. Collaborative work tools were established.
* Literature review and evaluation of the algorithms to be used, as well as the type of preprocessing for the images.
* We started implementing the code in early December. Although we worked on the code and the report in parallel, the first few weeks we focused more on image preprocessing, algorithm development and optimization, and the later weeks on report writing and finding optimal ways to visualize the data.
* December 23, delivery of the work.

2. Project Execution

2.1. Data Acquisition

After evaluating different datasets from Kaggle and other public repositories, the “Brain Tumor MRI Dataset” was chosen3. Exclusion criteria included synthetic datasets, small sample size, low quality, low applicability, poorly documented datasets, or data with low scientific rigour..

The selected dataset consists of 7023 brain magnetic resonance imaging (MRI) scans, classified into four fairly balanced categories: glioma, meningioma, pituitary, and notumor. These are .jpg images, indicating that they have already been partially preprocessed, as the original format for this type of image is usually DICOM (Digital Imaging and Communications in Medicine).

This dataset is a combination of three datasets, allowing an increase in the sample size of the dataset and its variability (coming from different hospitals, different machines, different parameters when taking the image, etc.), thus helping our model to generalize.

When exploring the dataset, we defined a series of considerations to take into account for its preprocessing:

* Not all images are the same size.
* There are duplicate or very similar images in which the brightness, zoom, or other characteristics vary.
* It may contain low-quality images.
* MRI images can be taken from various planes (axial, sagittal, coronal) and using different sequences, which are the parameters that are adjusted on the machine according to the information that you want to highlight in the image (T1W, T2W, FLAIR).

2.2. Training Methodology

2.2.1. CNN

Solving a problem of this nature requires using an advanced image classification algorithm. For this purpose, we have chosen a convolutional neural network, as it represents one of the best algorithms currently available for this type of task 2,4,5 .

The pre-trained 'EfficientNetV2B3' model was chosen as the base model. This architecture stands out for its excellent balance between accuracy and computational efficiency. Its design uses optimized convolutions (such as separable depth convolutions) and composite scaling that uniformly adjusts the depth, width, and resolution of the network, enabling high performance in resource-constrained environments or with large data volumes.

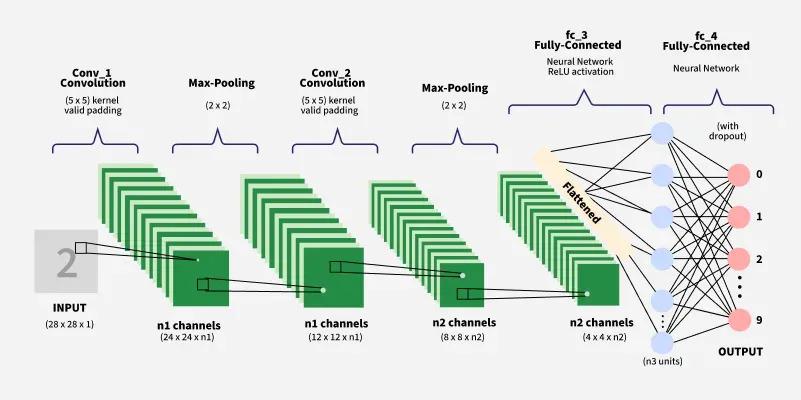
[](https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/)

Figure 3. Operation of a CNN. Image obtained from 4

2.2.2. Duplicate Removal

The perpetual hashing (pHash) algorithm was used to identify and remove images that were identical or highly correlated. This technique prevents overfitting and ensures that the test set contains genuinely new information for the algorithm.

2.2.3. Outliers Removal

To mitigate the impact of extreme outliers on pixel intensities, a percentile-based method was used. Values outside a predefined range (95%) were adjusted to the limits of these percentiles.

2.2.4. Background Suppression

Background suppression is a technique we have implemented to systematically remove the black and irrelevant background from MRI images. This forces the neural network to focus its attention on the brain structure and any potential tumor lesions, improving learning efficiency. We achieved this by applying two techniques. One is known as ROI cropping, where the background is reduced and the important part of the image is enlarged. Next, a radial mask is applied, where we apply a radius to cover the insignificant values ​​of the image, so that the final image is the structure of the skull along with the brain, but without the corners or background space occupying a large size that ends up hindering the AI ​​(Figure 4).

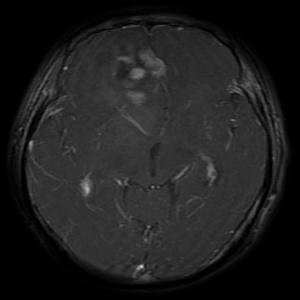
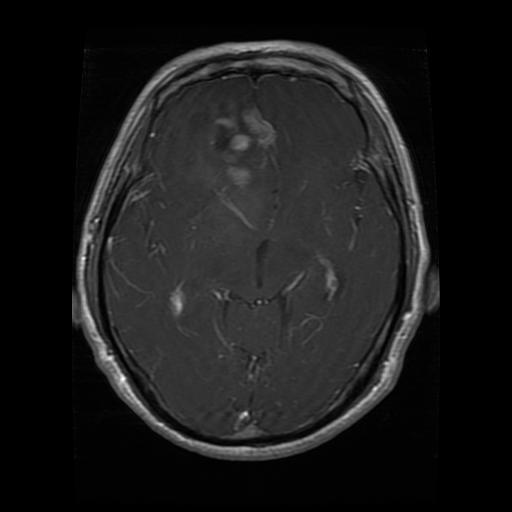


Figure 4. Comparison of the image *Te-gl\_0053.jpg* before and after background removal.

2.2.5. Data Augmentation

To increase the variability of the training set and improve the model's generalizability, an aggressive data augmentation strategy was applied. The transformations applied were random rotation up to 30 degrees, 15% horizontal and vertical shifting, angular distortion, random zoom, random horizontal inversion, random brightness adjustment, and random color channel shifting.

2.2.6. Training Callbacks

During neural network training, it is crucial to apply control mechanisms that ensure the selection of the best version of the model and prevent overfitting, while simultaneously optimizing computation time. To achieve this, we have implemented a set of callback functions that automatically act at specific points in the training process.

* **ModelCheckpoint**: This function automatically saves the model whenever an improvement is detected in a key metric. In our case, it uses the 'val\_accuracy' metric, which is the standard for obtaining the best version of a model, as it represents the accuracy of the validation set. It is configured in 'maximum' mode, meaning its task is to maximize the 'val\_accuracy' metric.
* **Early Stopping**: This is a regularization strategy that prevents overfitting. It stops training once the model's performance stops improving. It is configured to monitor 'val\_loss', which is the loss in the validation set. When this metric stops decreasing, it means overfitting is occurring, which triggers the training stop.
* **ReduceLROnPlateau**: This setting is the learning rate, a critical hyperparameter that determines the size of the steps the model takes to adjust its weights during gradient descent. If it is too high, the model can "skip" the minimum optimum; if it is too low, training is slow. It is configured to monitor the 'val\_loss'. If this metric stagnates, it means the model is stuck in a flat region of the optimization space. A patience of 4 was applied; that is, if it does not improve in 4 epochs, its learning rate is reduced, specifically by 20%, until reaching a lower limit defined at 10-6.

2.2.7. Grad-CAM

To translate confidence in the model to a clinical setting, an artificial intelligence explainability (XAI) technique was incorporated. Grad-CAM activation maps were used. This technique generates a heatmap that overlays the original image, highlighting the image regions that most contributed to the model's classification decision. This not only validates the prediction logic (ensuring the model focuses on the tumor region and not artifacts) but also provides a crucial visual tool for validation by clinical experts.

2.3. Workflow

As work tools, we primarily used Google Drive to share scripts and to write the document using Google Docs, which allows everyone to have real-time, up-to-date access. We also used WhatsApp for team communication and progress reporting.

Once we decided on the means by which we would collaborate, from a set of possible projects with datasets that we found on the Kaggle platform, we voted using a priority list to determine which project to undertake, resulting in this brain tumor analysis project.

With that, we began to decide which algorithms to apply to solve this problem, because although it can be achieved by applying classical Machine Learning, our research into alternatives led us to discover that there are more powerful options, such as the use of convolutional neural networks.

With this in mind, we considered running this project in two parallel versions: one applying traditional machine learning algorithms (PCA + SVM) and the other deep learning (CNN). This way, we could see if the faster, classic machine learning approach could match the accuracy of deep learning, making it a better alternative. However, we found that the deep learning model solved the problem significantly better. Therefore, the other project was discarded, and we focused on the deep learning model.

Once we had the project's framework, we applied data cleaning techniques (pHash for duplicates or highly correlated data, percentiles for outlier removal, etc.), resulting in an improvement. We also observed that applying fine-tuning to the model worsened its accuracy, so after several tests with different parameters, we decided it was better to eliminate fine-tuning, as the training model was consistently better.

Next, we optimized and improved our script, as described in detail in the "Testing and Improvements" section. We assessed the model's performance using a confusion matrix, applying metrics such as recall, F1 score, and especially accuracy.

Finally, we evaluated the model's decisions using Grad-CAM, which also allowed us to improve the model. In this way, we identified, for example, the need to apply background suppression techniques.

2.4. System Design

The following image illustrates the processes the data undergoes in a simplified manner, from its input into the AI model to the final heatmap. You can see how each image is processed by a pHash algorithm and percentile analysis to remove outliers. Then, ROI cropping and radial masking are applied. Finally, data augmentation is applied to the training set, which is then fed into the CNN artificial intelligence algorithm. After training, for each image in the test set or any new image we want to evaluate, this model will provide an answer as to whether it is a tumor and, if so, the type of tumor. Finally, each image receives a heatmap (GradCAM) showing the area the AI focused on to determine its decision, thus adding to the model's explainability.

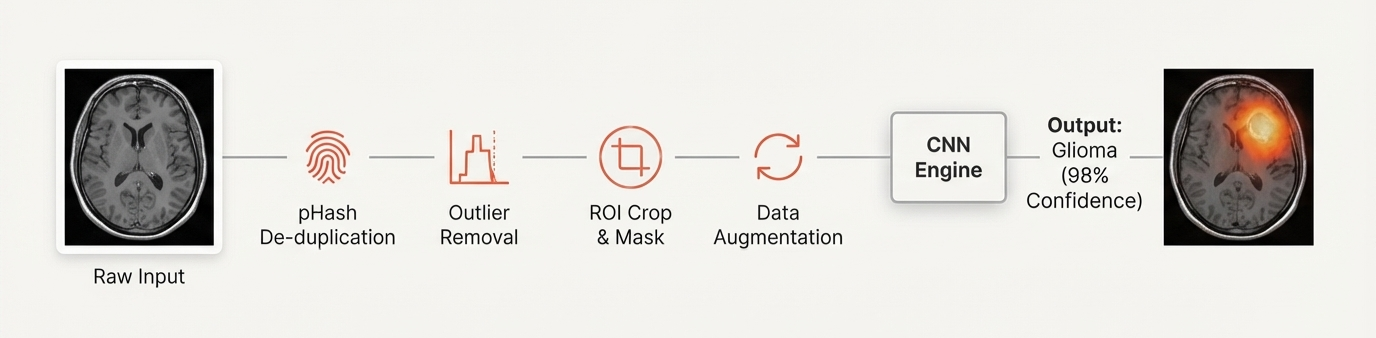


Figure 5. The data's journey, visualized from end to end. Figure generated with NotebookLM.

2.5. Supporting tools and resources

For the development of this project, various support tools were used, such as Discord, to facilitate communication among team members. Development environments like Visual Studio Code, Google Colab, and Google Antigravity were used, along with AI assistants like GitHub Copilot to support development tasks such as detecting and correcting code errors. The Google Drive platform was also used for collaborative work through shared documents. We used NotebookLM to generate slides and incorporate relevant images into this document.

The code to reproduce this analysis is in the GitHub repository “ <https://github.com/ImAle/IA_CNN_BrainTumor>”.

3. Results

3.1. Data Preprocessing

Before training the model, it is essential to process the data we are working with. This data comes from Kaggle, from a single common source; however, it was not clean. To improve data quality, we first had to run it through algorithms to remove highly correlated images (pHash), such as those shown in Figure 6, and to eliminate outliers (percentiles). Applying these strategies significantly improved the model, as it reduced noise and, with significantly fewer images, allowed for faster model training.

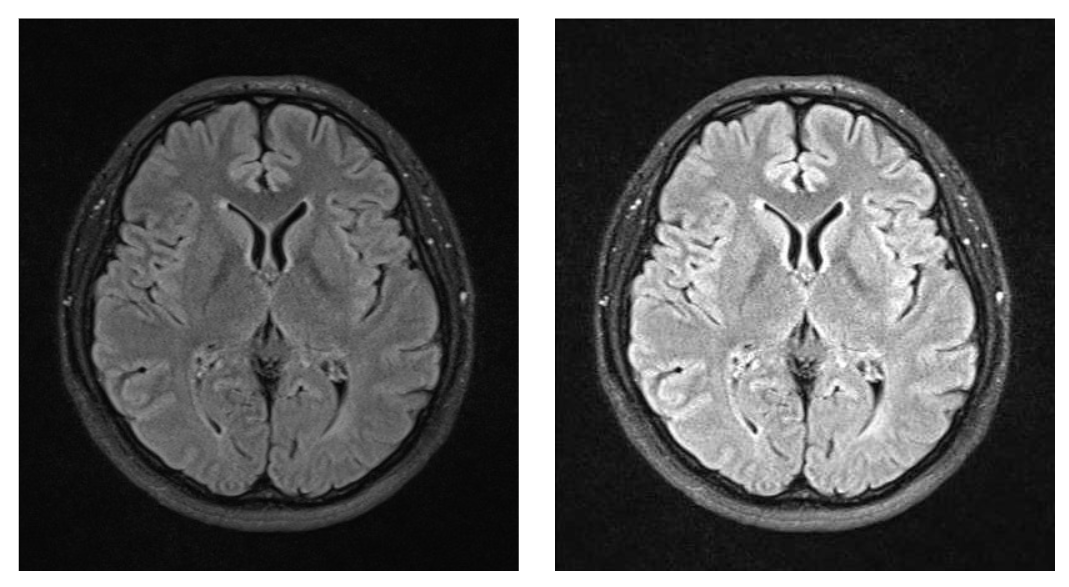


Figure 6. Example of duplicate images found in our dataset in the notumor class. Images: *Tr-noTr\_0006.jpg* and *Te-no\_0171.jpg*

With the application of Grad-CAM, we observed that the model focused on the background when making predictions. Upon discovering this, we had to add an extra step to the data preprocessing, which involved removing as much of the background as possible within our limitations. Applying ROI cropping (to enlarge the relevant part of the image and reduce the irrelevant part) and a radial mask (circle-shaped removal from the outside in) allowed the model to focus more on the brain to determine whether a tumor was present, and if so, the type of tumor, thus improving its generalizability.

3.2. Exploratory Data Analysis (EDA)

As we can see in Figure 7, the distribution of images across the different classes before preprocessing is fairly balanced, although the 'notumor' class has slightly more samples than the other classes. After preprocessing, we obtain the data distribution shown in Figure 8. A significant change is observed in the number of 'noTumor' images, as there was a high number with a high correlation (>=95).

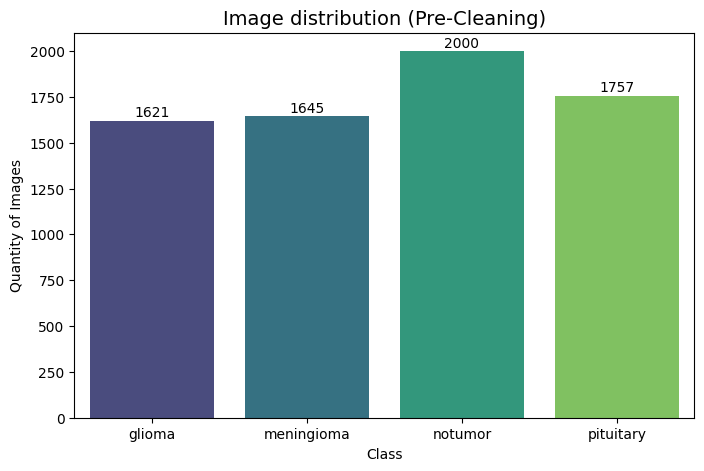


Figure 7. Bar chart of image distribution across classes prior to the elimination of duplicates and outliers.

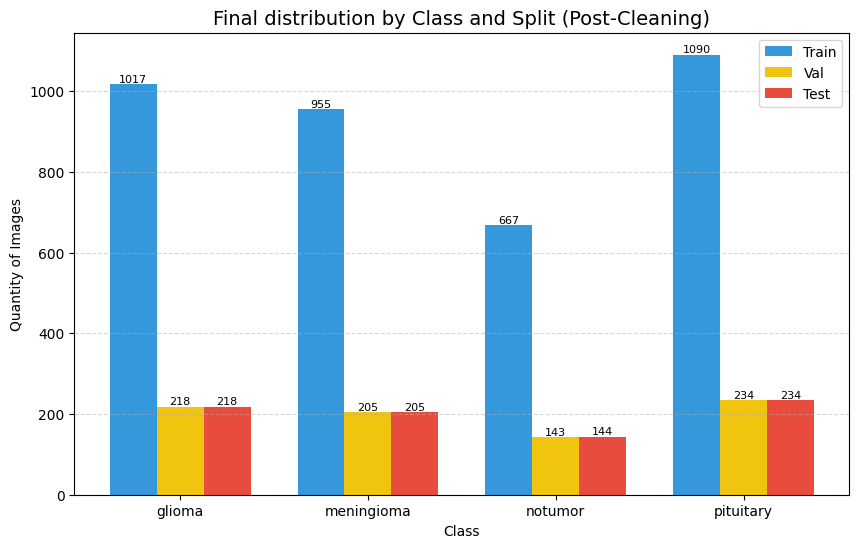


Figure 8. Bar chart of images distribution across classes after the removal of duplicates and outliers in the training, validation and test sets.

3.3. Modeling

3.3.1. Model Architecture and Metrics

Our model has the following defined parameters (Figure 9):

* **Base layer**: The pre-trained EfficientNetV2B3 model acts as a feature extractor.
* **Global Pooling Layer (GlobalAveragePooling2D)**: Reduces the dimensionality of the feature tensor to a flat vector, while maintaining key spatial information.
* **Batch Normalization**: Stabilizes and accelerates the training process.
* **Dense Layer**: Fully connected layers for high-level learning. A layer with ReLu activation was used to introduce nonlinearity.
* **Output Layer**: A Dense layer with softmax activation that generates the probability that the image belongs to each of the classes (Glioma, Meningioma, Pituitary or No Tumor).

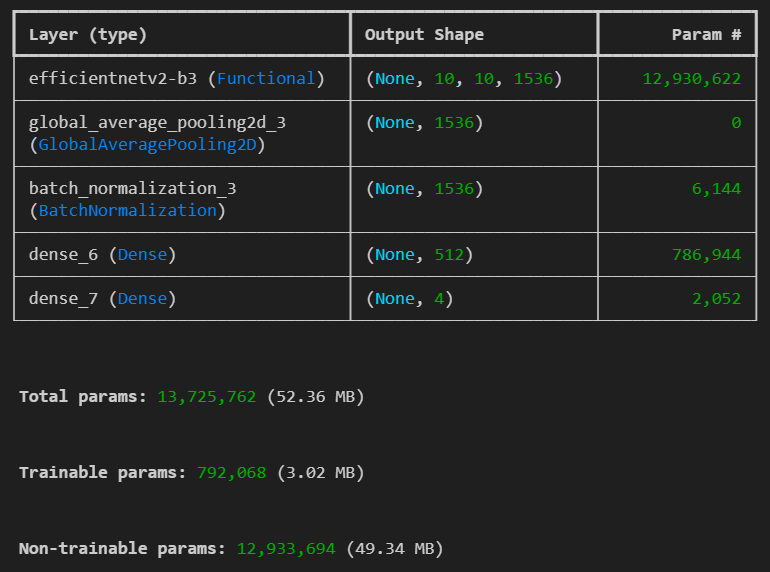


Figure 9. Architecture of the convolutional neural network used.

The training ended with EarlyStopping using a patience of 12, which stopped the training at an optimal time, as can be seen in Figure 10, since in both the Training and Validation graphs they no longer improve and both converge.

The results obtained can be observed both in Confusion Matrix format (Figure 11) and in comparative table format (Table 1), where the precision, recall, f1-score and support of both the overall mean and each specific class separately can be observed.

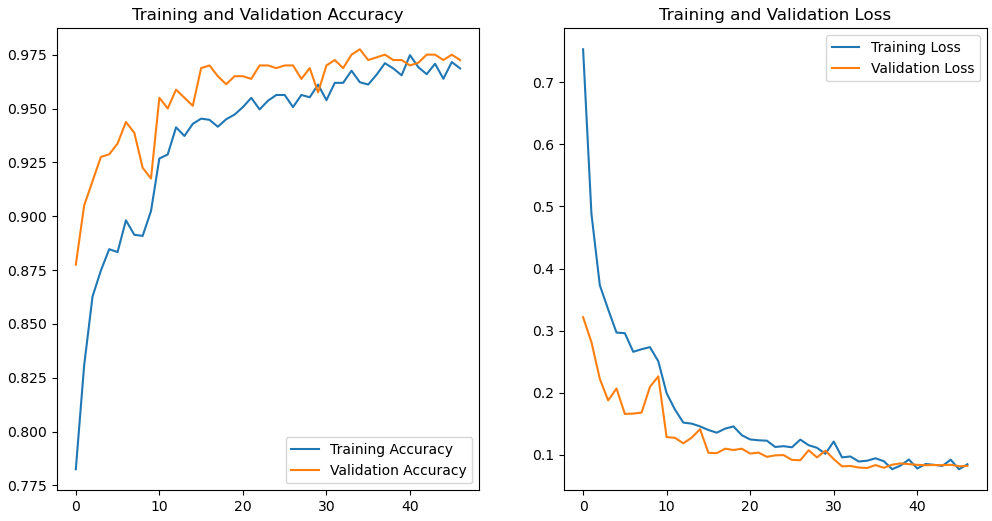


Figure 10. Evolution of accuracy and loss in the training and validation sets during the different training epochs.

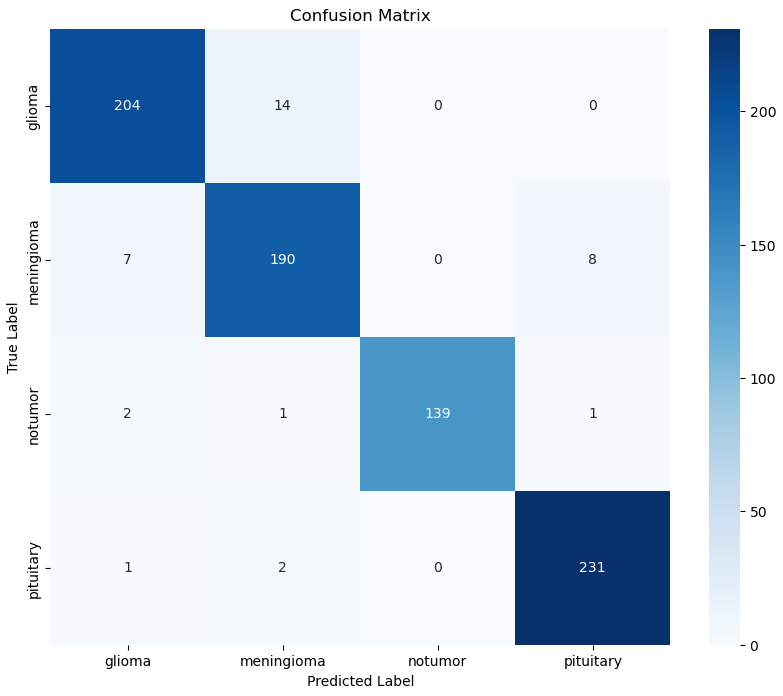


Figure 11. Confusion matrix of the classification model, where the rows represent the actual labels and the columns the predicted ones, with the color intensity being proportional to the number of observations of each actual-predicted combination.

|  | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| **Glioma** | 0.95 | 0.94 | 0.94 | 218 |
| **Meningioma** | 0.92 | 0.93 | 0.92 | 205 |
| **Notumor** | 1.00 | 0.97 | 0.99 | 143 |
| **Pituitary** | 0.96 | 0.99 | 0.97 | 234 |
|  |  |  |  |  |
| **Accuracy** |  |  | 0.95 | 800 |
| **Macro avg** | 0.96 | 0.96 | 0.96 | 800 |
| **Weighted avg** | 0.96 | 0.95 | 0.96 | 800 |

Table 1. Model metrics by class, global and average.

3.3.2. Explaining model predictions using Grad-CAM

As we can see in Figures 12-19, our model primarily uses the tumor area to classify the images, even in mispredicted images such as Figure 13, where a glioma is predicted as a meningioma. Improvements to the model, including background suppression, are reflected in these images. Even so, we can still see how the background interferes in some images, as shown in Figure 18, although a significant improvement can be observed compared to the heatmaps obtained before applying background suppression (Figure 20).

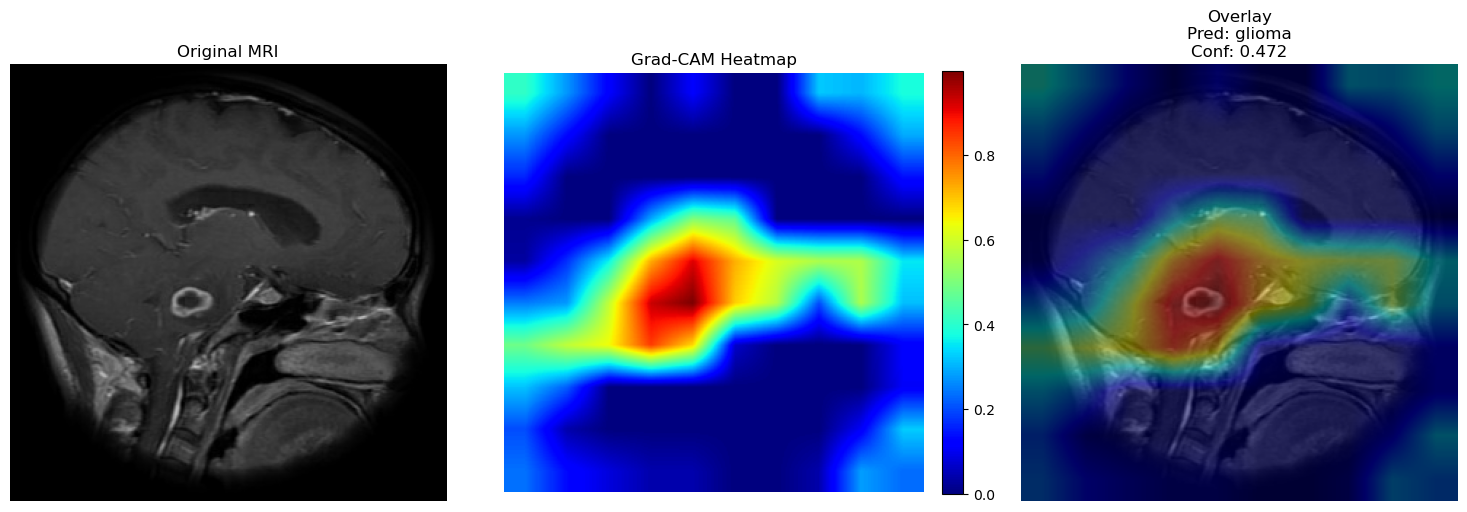


Figure 12. Grad-CAM visualization. Original image, activation map and overlay of glioma*.Te-gl\_0167.jpg:* Pred=glioma, Conf=0.472.

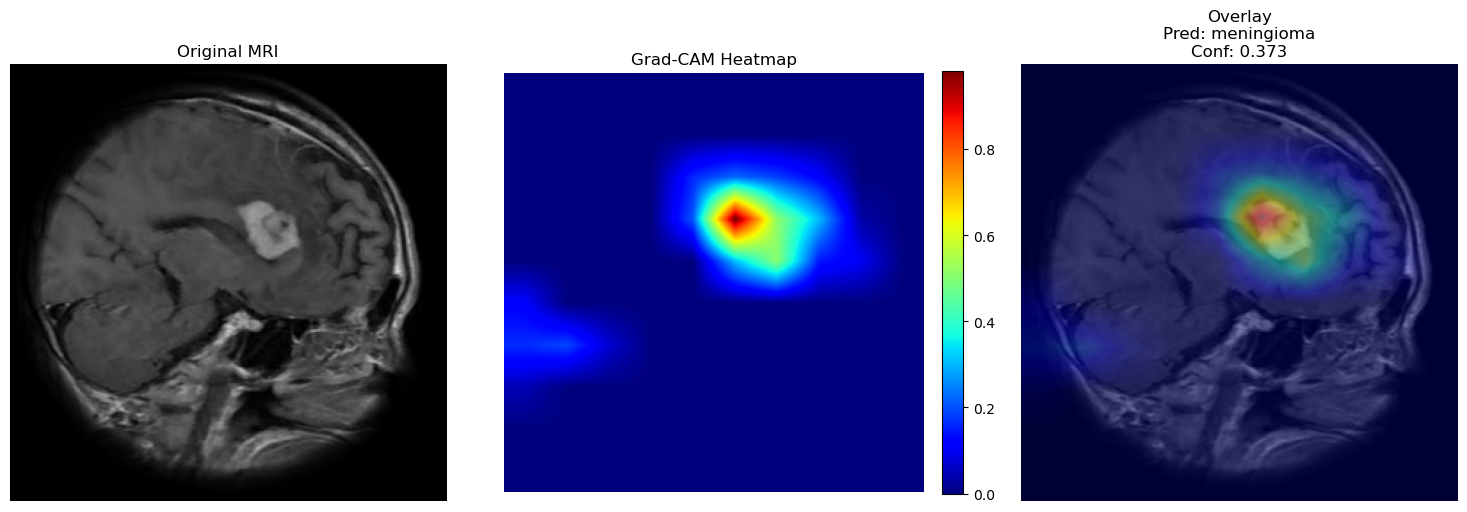


Figure 13. Grad-CAM visualization. Original image, activation map, and overlay of glioma predicted as meningioma. *Tr-gl\_1211.jpg:* Pred=meningioma, Conf=0.373.

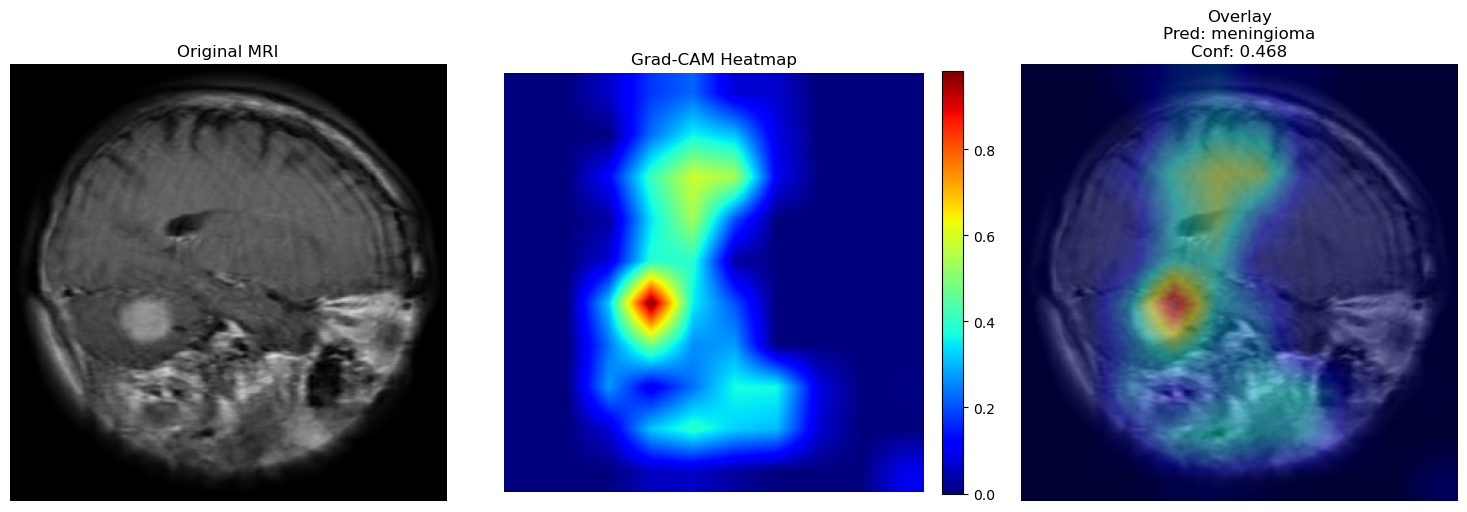


Figure 14. Grad-CAM visualization. Original image, activation map, and overlay of meningioma. *Tr-me\_1169.jpg* : Pred=meningioma, Conf=0.468.

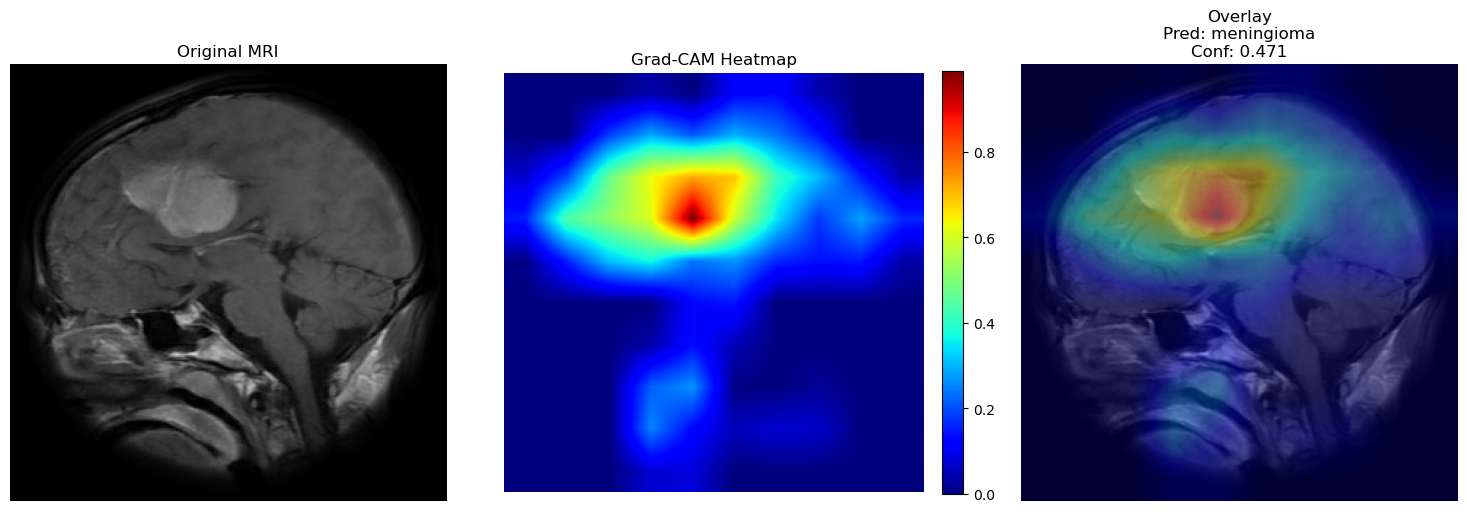


Figure 15. Grad-CAM visualization. Original image, activation map, and overlay of meningioma. *Tr-me\_0669.jpg* : Pred=meningioma, Conf=0.471.

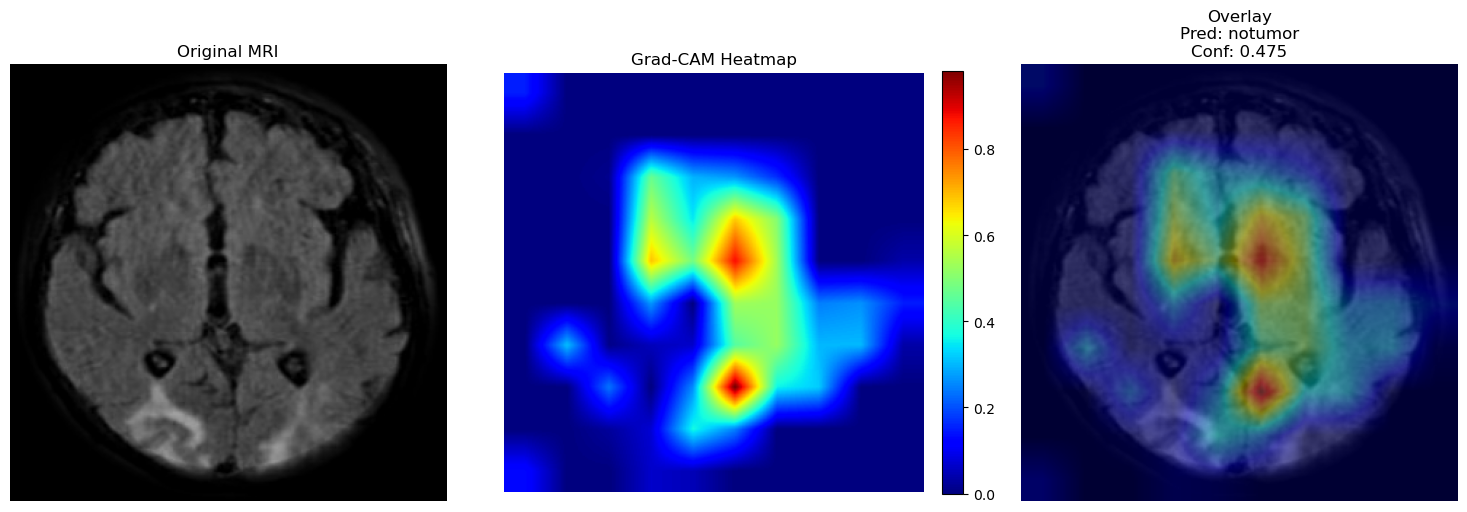


Figure 16. Grad-CAM visualization. Original image, activation map, and overlay of NoTumor. *Te-no\_0027.jpg:* Pred=notumor, Conf=0.475.

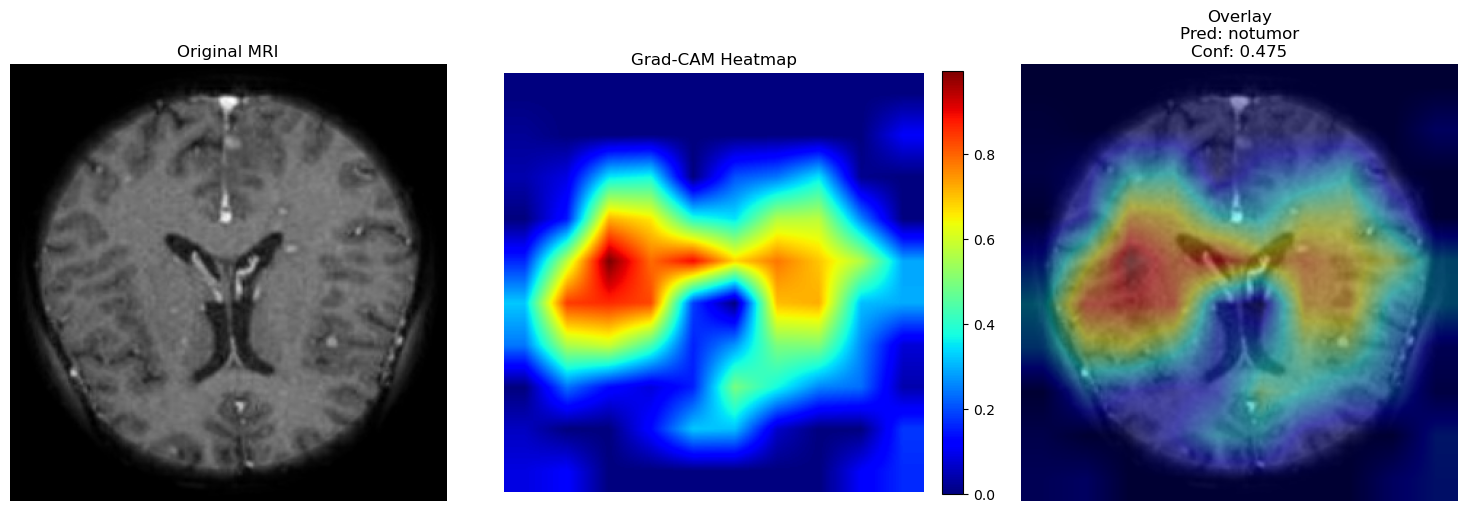


Figure 17. Grad-CAM visualization. Original image, activation map and overlay of NoTumor.Te-no\_0238.jpg: Pred=notumor, Conf=0.475.

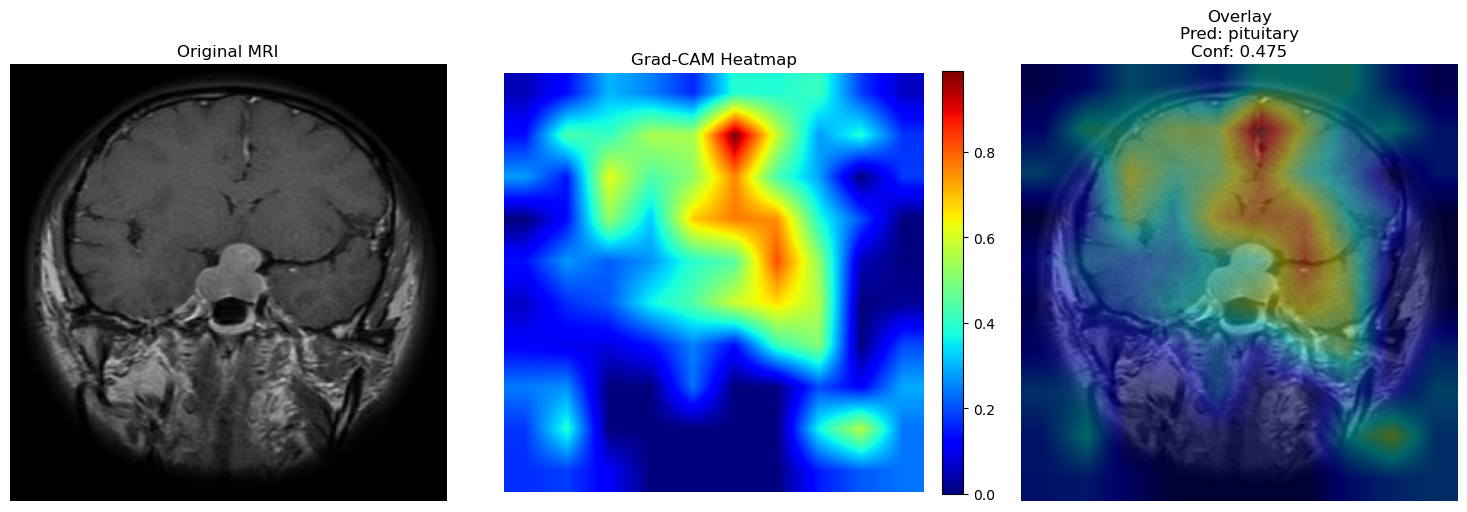


Figure 18. Grad-CAM visualization. Original image, activation map and overlay of Pituitary. *Te-piTr\_0001.jpg* : Pred=pituitary, Conf=0.475.

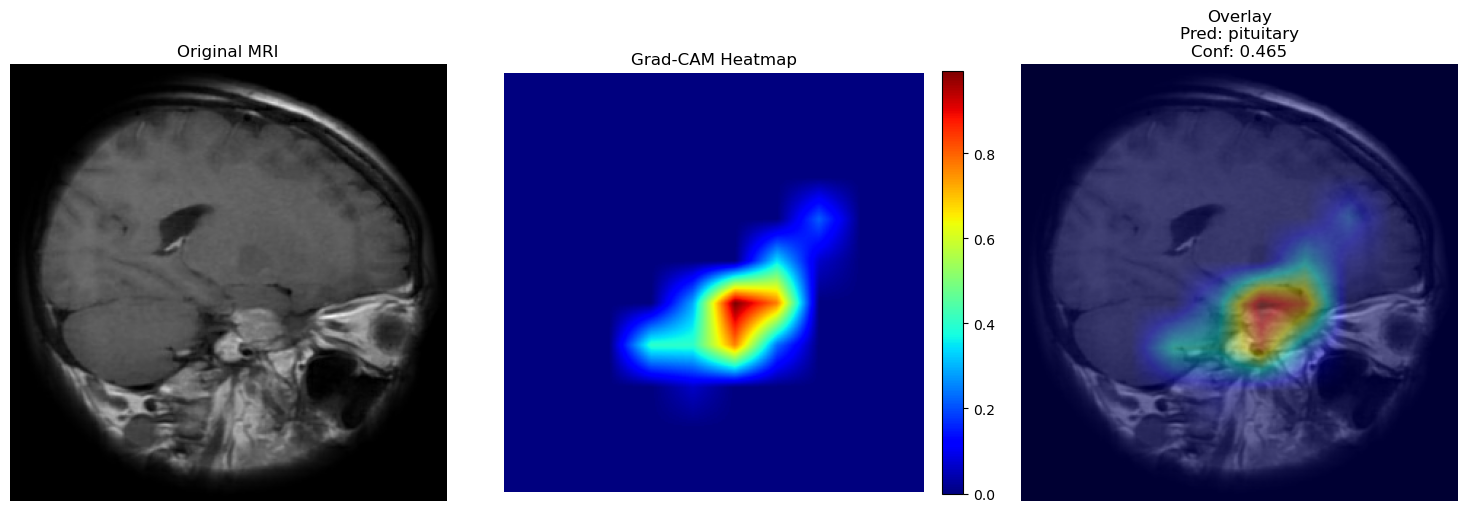


Figure 19. Grad-CAM visualization. Original image, activation map, and overlay of Pituitary. *Tr-pi\_1432.jpg:* Pred=pituitary, Conf=0.465

3. 4. Testing and Improvements

3.4.1. Classical algorithm-based approximation

A classical algorithm was also tested to compare results and assess whether similar results could be obtained with simpler, more interpretable algorithms that required less computation time. A pre-trained neural network was chosen to extract image *embeddings*, followed by a dimensionality reduction algorithm (PCA). Finally, a supervised learning algorithm, Support Vector Machines (SVM), was applied to generate the classification model. This last algorithm was chosen because it was the standard for image classification before the advent of convolutional neural networks and because it has yielded the best results in various MRI classification studies among traditional machine learning algorithms 5,6,7. Since the results obtained were worse than those obtained with the convolutional neural network, we decided to focus on convolutional neural networks and discontinue this approach.

3.4.2. Pretrained Models

Tests were also carried out with different pre-trained models such as EfficientNetB3, Resnet50V2 and DenseNet121, but none showed an improvement over the performance of the EfficientNetV2B3 model, so no changes were made, and it was kept until the end.

3.4.3. Reducing Background Influence in Model Training

Upon observing the Grad-CAM results, we realized that the model was classifying some images based on parts of the image background. Figure 20 shows one of the most pronounced examples, highlighting the importance of addressing this issue. It's worth noting that this dataset is a combination of three datasets, and all the non-tumor images originate from one of them. Since we wanted our model to be able to generalize to images from any hospital rather than learn from the background, we decided to use a preprocessing step, commonly used in neuroimaging, called skull stripping. This involves applying a filter to remove the background from the MRI using tools such as HD-BET.

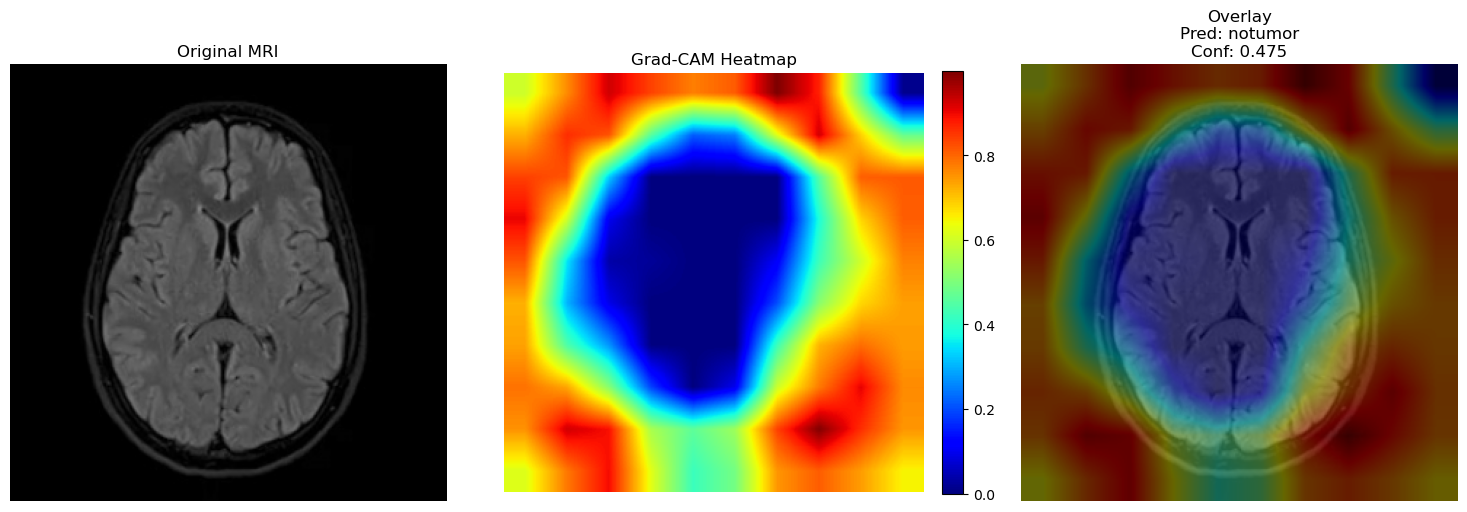


Figure 20. Grad-CAM visualization. Original image, activation map and overlay of notumor *Te-noTr\_0002.jpg* in which the old model uses the background to make the decision instead of brain tissue.

The problem with applying this method is that HD-BET uses a convolutional neural network trained on images obtained from MRI machines, which are usually in DICOM format and include a lot of metadata such as orientation and the information needed to reconstruct the 3D image. However, our dataset is only partially preprocessed, and we started with 2D images in .jpg format, so it wasn't possible to implement this solution. Finally, we used ROI cropping along with a radial mask as an alternative, which is explained in detail in the " Training Methodology" section.

3.4.4. Dealing with misclassifications with an “Inconclusive” category

After reviewing the *Prediction Confidence* of the incorrect results and comparing them with that of the correct ones, it was observed that the model tended to be much more confident in its decisions regarding correct results than incorrect ones (Figures 21 and 22). Consequently, the possibility of implementing a modification to the model was explored, creating a new class called 'Inconclusive' in which all results with a *Prediction Confidence* lower than a specified threshold would be added. After testing different thresholds, it was concluded that this was a very feasible option if the goal was to reduce the number of incorrect results at the expense of a lower number of correct ones. As shown in Figure 23, using a threshold of 97.5%, the error rate can be reduced by 86% proportionally to the overall rate, sacrificing only 16% of the correct results and leaving 19% of the results inconclusive.

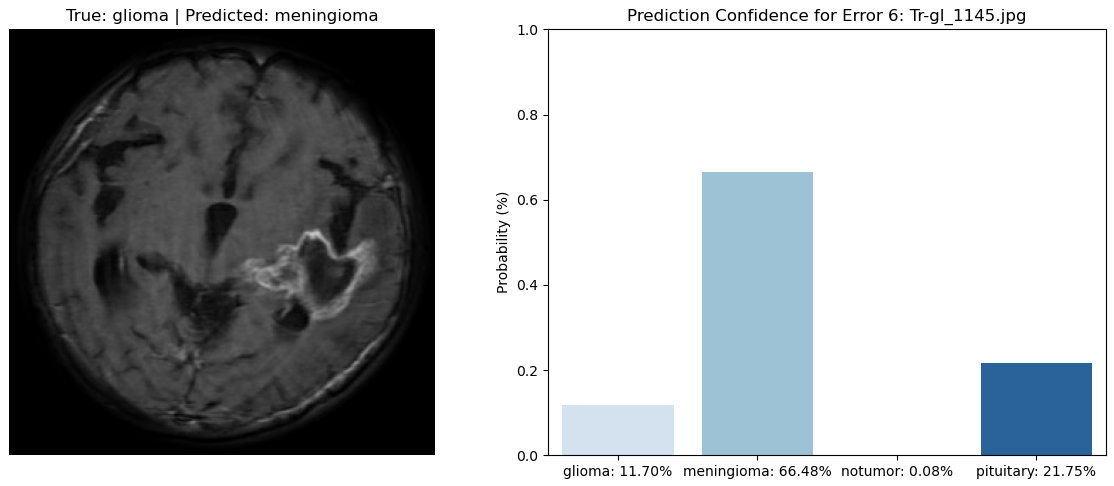


Figure 21. Example of prediction confidence distribution in a misclassification

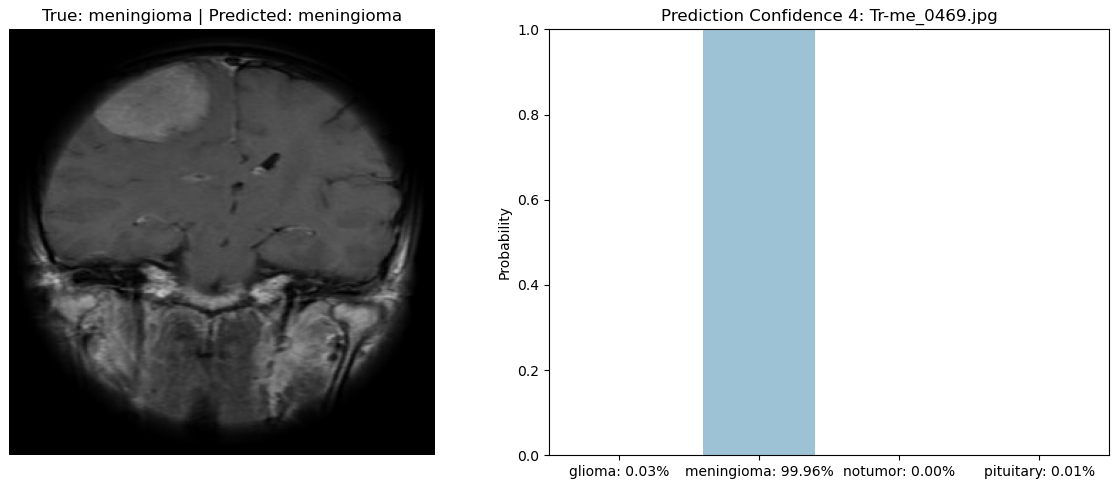


Figure 22. Example of prediction confidence distribution in a correct prediction.

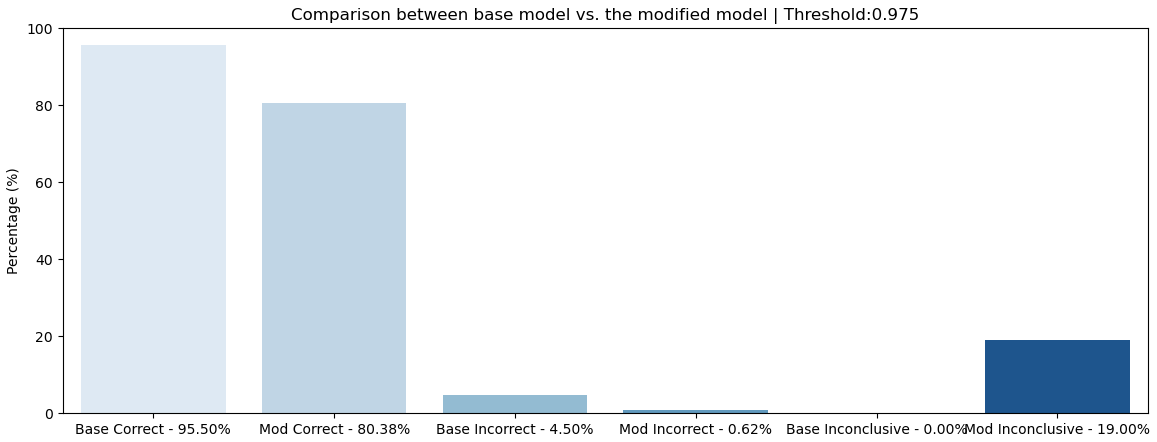


Figure 23. Comparison of results between the base model and the modified model with a threshold of 97.5%.



4. Projected Impact

4.1. Accomplishments and Benefits

In this project, we have developed a multiclass classification model using a convolutional neural network that allows us to classify future brain MRIs into four classes: “glioma,” “meningioma,” “pituitary,” and “notumor,” with 95.5% accuracy. This model has great potential for clinical application in AI-assisted diagnosis. It can generate valuable information, such as prediction and confidence levels, and Grad-CAM activation maps, enabling radiologists to determine the most appropriate strategy for each patient.

Recognizing that even a small percentage of error in a clinically applicable model can have significant consequences, we evaluated in detail the types of errors our model makes and how to mitigate them. Our model has the advantage that all cases it predicts as non-tumor are indeed non-tumor, thus avoiding the problem of sending home a patient with a tumor who received a negative diagnosis. Most confusion arises between gliomas and meningiomas, specifically gliomas that are predicted to be meningiomas. However, a review of the literature reveals that differentiating between these two tumor types, even with the more traditional, manual methods used by radiologists, remains a clinical challenge due to the high degree of similarity in their radiological characteristics 2.

There is also a considerable number of meningiomas predicted as pituitary tumors. To minimize the number of erroneous predictions, we have incorporated the 'Inconclusive' class to include images that the model cannot clearly differentiate and therefore predicts with low confidence.

4.2. Future Improvements

In the short term, an attempt would be made to improve the accuracy of the model by testing other algorithms described in scientific articles for solving similar problems, such as the **Vision Transformers (ViTs) deep learning models**, which use the “Transformer” architecture, originally designed for natural language processing, for image analysis problems 2,8.

We would also explore the possibility of obtaining raw image data to apply **skull-stripping** or searching for alternative datasets that include this type of image. Although our dataset is already a combination of three different datasets, we would consider increasing the sample size by including more datasets, both for model training and evaluation.

In the long term, future improvements would focus primarily on transforming the developed algorithm into a **clinically applicable tool** for use in hospitals for **AI-assisted diagnosis**. It would serve as a support tool for radiologists, assisting and streamlining image classification, always under the radiologist's supervision and final decision.

To further improve our model, it would be essential to work closely with radiologists so they can evaluate the tool, suggest improvements, and help us interpret the results. This would allow us to jointly analyze misclassified images, Grad-CAM results to see which parts of the image the model is using to make decisions, the initial image labels, etc. In this way, we could refine the tool together, following **the radiologists' criteria**. For example, we could identify mislabeled or ambiguous images from which the model is learning, and then re-label them to increase its accuracy.

Furthermore, we could train a model to obtain a more complex multiclass classifier that, in addition to classifying tumors into the four initial types, would also be able to classify them according to disease stage and tumor grade, information that is very useful for deciding on the therapeutic strategy. To do this, we would need to re-annotate the initial dataset with the help of radiologists or find more complex datasets.

Finally, to make this tool more accessible to radiologists or researchers, a **graphical interface would be created** in which new resonances could be uploaded and, in addition to the prediction, some images, such as those in figures 12 to 22, could be obtained.

**References**

1. “Tumor cerebral e investigación” | Contra el Cáncer. <https://blog.contraelcancer.es/investigacion-tumores-cerebrales/>.
2. Mounika, G., Kollem, S. & Samala, S. Investigating brain tumor classification using MRI: a scientometric analysis of selected articles from 2015 to 2024. Neuroradiol. 2025 6710 67, 2635–2672 (2025).
3. Msoud Nickparvar. (2021). Brain Tumor MRI Dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/2645886>.
4. “Convolutional Neural Network (CNN) in Machine Learning” <https://www.geeksforgeeks.org/deep-learning/convolutional-neural-network-cnn-in-machine-learning/>(09/08/2025).
5. Slides and notes from the Artificial Intelligence Course 2nd Ed. Samsung and Escula de Organización Industrial.
6. García, F. “*Unidad 7. Máquinas de Vectores de Soporte (SVM*)”<https://fmgarcia.github.io/aprendizaje_supervisado.github.io/7.%20Algoritmo%20SVM/> (19/11/2025).
7. Kang, J., Ullah, Z. & Gwak, J. MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. Sensors (Basel). 21, 1–21 (2021).
8. Sankari, C., Jamuna, V. & Kavitha, AR Hierarchical multi-scale vision transformer model for accurate detection and classification of brain tumors in MRI-based medical imaging. Sci. Rep. 15, (2025).

5. Team Member Review and Comment

|  |
| --- |

| YAM | REVIEW and COMMENT |
| --- | --- |
| Alejandro Gallego Domínguez | This project has been a key experience in consolidating the knowledge acquired throughout the course and giving it real-world application. It allowed me to tackle complex problems, make decisions independently, and explore ways to implement improvements on my own. As the project progressed, it became clear that learning artificial intelligence is a continuous process of testing, analysis, and refinement. I am not only satisfied with the project itself, but also grateful for the team I worked with, as my teammates were hardworking, friendly, and consistently focused on continuous improvement. |
| Alejandro Lobillo Becerra | Working on this project not only allowed me to put the knowledge acquired during the course into practice, but also provided an opportunity to go a step further and conduct independent research in order to resolve the difficulties that arose during its development. I am very satisfied with the final result and with the way the project was carried out, as the working environment was always excellent thanks to the combined effort of all group members. |
| Almudena Neva Alejo | Participating in this project with my colleagues has been a very enriching experience, in which we applied the skills and concepts learned during the Artificial Intelligence course to a real-world problem: medical image classification. Furthermore, it allowed me to deepen my understanding of how to carry out an artificial intelligence project, both through independent learning and collaboration with my colleagues, working as a team to overcome the various challenges encountered.  This experience has shown me the potential of artificial intelligence in the field of biomedicine, and I am highly motivated to apply everything I have learned to future professional challenges. |

6. Instructor Review and Comment

| CATEGORY | SCORE | REVIEW and COMMENT |
| --- | --- | --- |
| IDEA | \_\_/10 |  |
| APPLICATION | \_\_/30 |  |
| RESULT | \_\_/30 |  |
| PROJECT  MANAGEMENT | \_\_/10 |  |
| PRESENTATION & REPORT | \_\_/20 |  |
| TOTAL | \_\_/100 |  |