Brain Tumor classification using Machine Learning and Deep Learning techniques

Rostyslav Dovganyuk
Department of Computer Science and Engineering - DISI
University of Bologna
Cesena Department
Email: rostyslav.dovganyuk@studio.unibo.it

***Abstract*—The paper explores recent advancements in brain tumor classification, delving into various methodologies employed within the realm of machine learning. Drawing upon a comparative analysis, this study evaluates the efficacy of different techniques, including decision trees, support vector machines (SVM), and deep learning models such as artificial neural networks (ANN), notably the Perceptron, alongside specialized image classification algorithms like Convolutional Neural Networks (CNNs). Through a rigorous examination of these approaches, this research sheds light on their respective strengths and limitations, offering insights crucial for advancing the field of medical imaging and diagnosis.**

1. Introduction

In the human body the brain is an enormous and complex organ that controls the whole nervous system, therefore any kind of abnormality may put the human in serious health danger. Among such abnormalities, brain tumors are the most severe ones. Brain tumors can be classified into two groups, primary tumors, that are present in the brain tissue, and secondary tumors that expand from other parts of the human body to the brain tissue.
Among the primary tumors *glioma* and *meningioma* are the most lethal ones, and may lead the patient to death if not diagnosed at an early stage [4]. According to WHO (World Health Organization), brain tumors can have huge variations and can be classified in different types and grades, but for our research we'll be focusing on three different tumors: glioma, meningioma and pituitary [5]. The following research goal is to compare different algorithms and approaches in the classification of tumors from MRI (Magnetic Resonance Imaging) brain images. In a more broad view, there will be a comparison between Machine Learning algorithms, with "manual" feature extraction from images, and Deep Learning algorithms, one using only ANN (Artificial Neural Networks) and "manual" feature extraction, and another one using CNN (Convolutional Neural Networks) to extract features automatically. The layout of this research is organized as follow: The related work is in section 2. The proposed methods are in section 3. The results are described in section 4.

2. Related Work

In the medical field the use of Artificial Intelligence has increased drastically within the last years, we have better machine learning algorithms, more access to data and a cheaper hardware [3], this permit to researchers and institutes to develop their own methodologies for specific problems, like the one that we're facing today, the classification of brain tumors based on the MRI images received.
A 2020 paper, from MDPI, present a work related to the classification of brain tumors using the BraTS datasets and reaching an accuracy of 97.8%, the classification goal was to classify 4 different classes of tumors (T1, T1CE, T2, Flair) [6]. The classification task require two steps, (1) features extraction and (2) classifier based classification. They used two different CNN models extract features and at the end perform a fusion of the features, to get a new feature vector with more information. Going deeply in the work, their method consist of five core steps: (1) linear contrast stretching, (2) deep learning features extraction, (3) corr-entropy-based joint learning approach, (4) PLS-based fusion of selected features and (5) ELM-based classification.

Those steps, respectively, take the MRI image,(1) increase the contrast of the image, to better visualize the region where tumor is present, (2) then extract deep learning features two pre-trained CNN models (VGG16 and VGG19), (3) (4) then modifying those pre-trained models accordingly to the problem, and (5) finally classifying them using the ELM (Extreme Learning Machine). Other classification algorithms are used to compare the results with the ELM, like MSVM and Ensemble Tree, but reaching the best results with the ELM, 98.16% of accuracy, second position for the Ensemble Tree with an accuracy of 95.67%.
Another interesting work was done by the Korean researchers, that preset their work on MDPI [4], where they actually are classifying MRI images, in 4 different classes, one is no tumor and the other three represent a tumor (Glioma, Meningioma, Pituitary), and also their work focus on extracting deep learning features and use several machine learning classifiers. Their work consist of, (1) crop the image using extreme point calculation, removing parts that can interfere with the classification process, like empty spaces around the MRI, then perform some *dilation* and *erosion* to

remove noise, and resizing it, the with Image Augmentation they create other images from the ones that they have, to increase the number of elements in the datasets. The model is composed by 13 pre-trained CNN models, to extract deep features, and the features extracted are then passed to nine different ML classifiers, then the top three deep features are selected based on the evaluation results, and concatenated together, and at the end those three feature vectors concatenated are given to the nine ML classifiers used in the step before. The model was trained and assessed using three different datasets, each one respectively with higher amount of images and with the first two datasets with only two classes, and the last one with 4 classes (the ones defined above). The accuracy obtained in the large dataset (the one that better represent our work), reach an accuracy, on average, of 90.19%, using a SVM classifier with RBF kernel.

3. Proposed Methods

In this research we're gonna compare our results using three different methods, the first one will be using "classic" machine learning classification algorithms and extracting manually some features from images, the second method will be by using deep learning techniques with a model created using ANN, and then the third one using some pre-trained models, in the same way as was done in section 2.

3.1. Dataset

The dataset used is taken by kaggle.com [2], and it's already subdivided in Training and Testing, each one respectively containing 2870 and 394 Magnetic Resonance images (MRI) in a gray scale format. Each dataset is divided in four folders, each one represent the class, that are: glioma, memingioma, no tumor and pituitary, and are graphically visible in Figure 1.

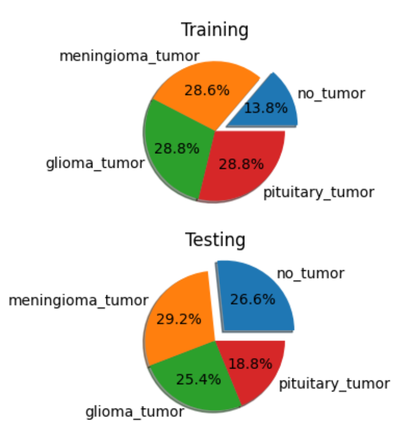

Figure 1: Dataset Subdivision

It is also visible from Figure 1 that in the Training set the no tumor elements are far lower that other classes (13.8%), this can impact the classification models, in this study we wouldn't balance the dataset.

3.2. Manual Feature Extraction

Feature extraction is a fundamental phase in the image classification field, in our first study we are going to extract features from images using the *scikit-image* packets, that contain a collection of algorithms for image processing. In our research will be used 2 different algorithms, *Histogram of Oriented Gradients* (HOG) and *Local Binary Pattern* (LBP). Below there is a little explanation of the feature descriptor, and in Figure 2 we can see the two algorithms put on work in comparison to the original image.

3.2.1. Histogram Oriented Gradients - HOG. This type of mechanism is used to describe the shape of an object, more precisely, HOG provides the edge direction. The image is divided in smaller regions, and for each regions the edge directions are calculated.

3.2.2. Local Binary Pattern - LBP. This feature describe the texture of the surfaces, where the texture is the visual surface appearance. The LBP use 2 elements, local spatial patterns, that is a 3x3 neighborhood of each pixel that is thresholded with the center value, and a gray scale contrast measure. That is computed on different cells of the image.

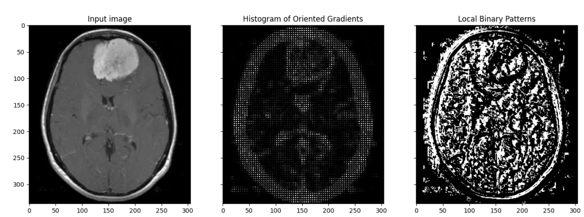

Figure 2: Feature Representation (normal, HOG, LBP)

3.3. Machine Learning classification algorithms

Machine Learning classification algorithms are very efficient and with the new technology they run everywhere, but have some limitations when talking about image classification, especially if the image is a raw image, without any feature extracted.

Before training or testing our models we need to perform some pre-processing steps, or in a simplest way, we can't pass to the model a raw image. The first thing the image should be viewed, read, as a sequence of values, where each value represent the pixel, so the representation is a matrix, that each cell represnt a pixel of the image, after that the first real pre-processing step will be feature extraction, and we will try two different methods, as we saw in section 3.2 plus just the raw image without any feature extraction, just a normalization of pixels, and after that the matrix will be flatten, so will become an individual array of values. In our case the images will be firstly resized to 64x64, this means that the matrix will have a shape of 64x64x1, that means that we have single matrix, where the cell represent the pixel, in this case it will be in a gray scale dimension, for that there is only one two-dimensional matrix and not three like we would have for an RGB image (RxGxB). After the flattening the model will receive an array of length 4096 (64x64). The representation of this concept can be better visualized in Figure 3.

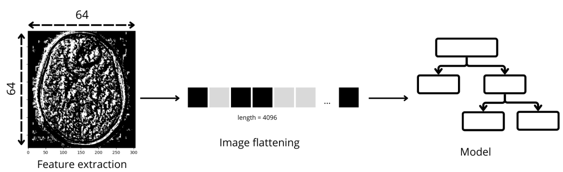

Figure 3: Pipeline of the image flow

The algorithms used are:

- SVC
- Random Forest
- AdaBoost
- Decision Tree

The testing of algorithms are done by training the model using the feature extracted, or the raw image, and then comparing the results for each algorithm. Resuming it, the images are first process, there is a feature extraction and are flatten, then we create the classifiers, and train each classifier on the data and assess the algorithm using accuracy and score, then we change the feature extraction method and start again, and at the end compare all the results. All the results of this part are in Section 4.

3.4. Deep Learning Classification

The deep learning classification is done in three different ways, the first one is a simple model created by the author, without the use of Convolutional Layers, the second one is also a model created by the author, but with the use of Convolutional Layers, and at the last model, was created by using the transfer learning method, by using as base the MobileNet model.
As we saw in Section 3.1 (Dataset), the dataset is already subdivided in training and testing, for a better training and for better results, a part of the training set (20%) is used as validation set, that is used during the training set to evaluate the model and make some adjustment in weights to be more ”general” for the real world. Below there are the descriptions of the three model, and all the results are in Section 4. All the three models were trained on 20 epochs.

3.4.1. Model No Convolutional Layer. The model created is a simple model that has as first layer the Normalization layer, that is something that in the classification with machine learning was done ”manually”, and after that has a Flatten layer that transform the matrix into an array, in the same way it was done in Machine Learning classification, and at the end some Dense layers that has as a principal goal to implement the activation function to the elements. The Model can be better seen in Figure 4

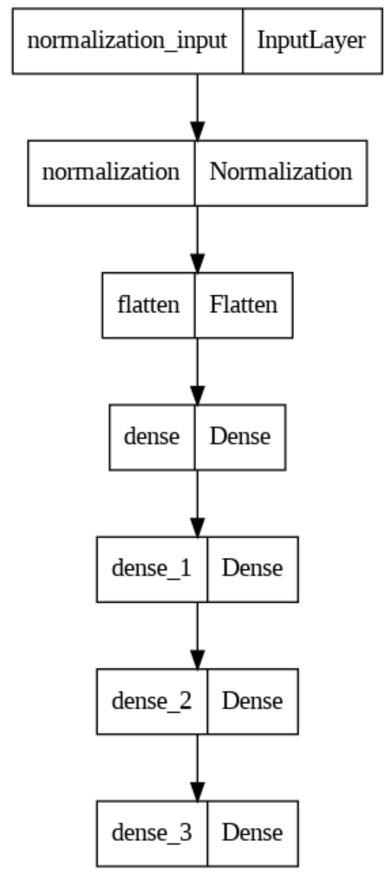

Figure 4: Visual representation model with no Convolutional layers

3.4.2. Model Convolutional Layer. The model is also a simple model but now implementing some Convolutional Layers, that act as feature extractors, and the model is like the one above but has two Concolutional layers after the Normalization layer, can be better visualized in the representation in Figure 5

3.4.3. Transfer Learning to MobileNet. MobileNet is an efficient Convolutional Neural Networks for Mobile Vision applications [1], and was fine tuned specifically for IMR tumor classification.
Fine tuning consist of training specifically, a model already trained, for something else, this by changing weights in the neurons, but not only that, also the final layer (Output

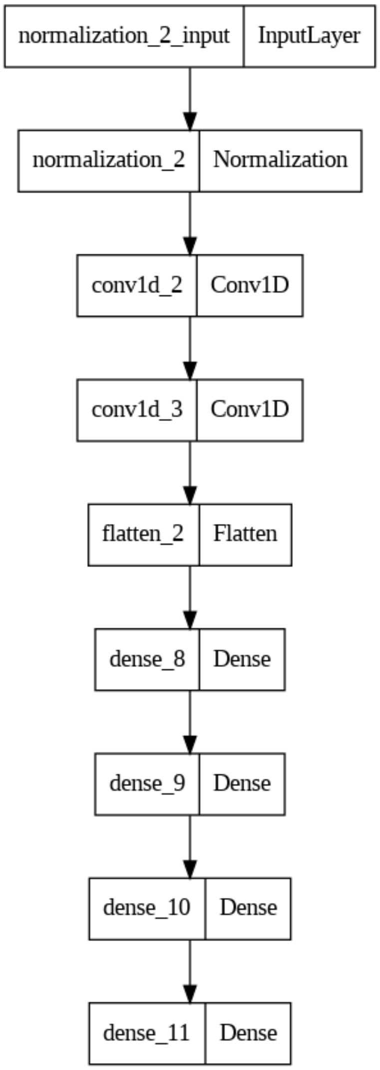

Figure 5: Visual representation model with Convolutional layers

Layer), need to be modified, the actual MobileNet model is capable of classifying 1.000 different classes, but for our case we need to classify only 4 classes, so the ”original” output layer is replaced with a Dense Layer. After replacing the last layer the model is trained on the training set and then ready to use.

4. Results

The results of the Machine Learning classification algorithms are visible in Table 1.

It's visible that SVC (Support Vector Classification) has obtained some great results especially when the *hog* algorithm for feature extraction is used, but also has great results when is used the raw image, but we can also see that DecisionTrees algorithm has performed very well. On the other hand the other two algorithms, RandomForest

	hog		lbp		img	
Classifier	score	accuracy	score	accuracy	score	accuracy
SVC	**0,8571**	**0,7386**	0,2962	0,7107	0,8362	0,7259
RandomForest	0,5645	0,2462	0,6655	0,2843	0,5749	0,2766
AdaBoost	0,7387	0,4188	0,5889	0,4112	0,6725	0,3629
DecisionTree	0,6899	0,7335	0,7108	0,6244	0,7805	0,6650

TABLE 1: Table with all the results, in bold the higher ones

and AdaBoost, have obtained some scarce results, with an accuracy that never goes over 0.5.
In Figure 6 it's graphically visible the pour performance of RandomForest and AdaBoost, but this is due to a poor hyper-parameters of the models, so the model is actually performing pretty well on the score, but this parameter is assessed using the validation model, on the other hand the accuracy is assessed on the training set, so this means that the two model are under-fitting, so maybe increasing the depth in RandomForest and changing some parameters in AdaBoost, also those model can perform well as the others.

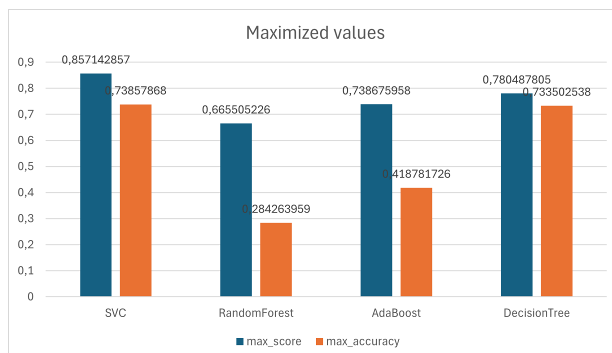

Figure 6: Visual representation of classifiers maximal results

To better assess the Deep Learning models, the results are shown based on the validation set, and it's visible in Table 2.

model	accuracy	loss
model 1 (no Conv)	0,7456	2,641
model 2 (w Conv)	0,8101	1,245
MobileNet (transfer l.)	**0,9199**	**0,2739**

TABLE 2: Results of Deep Learning models

It's clearly visible from Table 2 that the fine tuned model is clearly performing better than the other ones, it's also visible from it's graphical representation of loss (Figure 7) and accuracy for each epoch, that it's more linear and completely different from model 2, this is particularly visible in Figure 9, where the validation loss line is really far away from the training loss line, and at some points are visible some spikes, that in MobileNet are not visible. The results of the ”home made” models can be improved by adding some more Convolutional layers, increasing the number of epochs and increasing the number of the dataset.

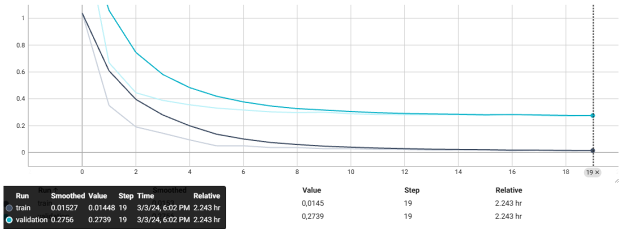

Figure 7: epoch Loss for MobileNet model

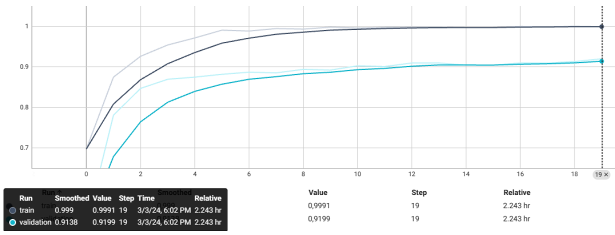

Figure 8: epoch Accuracy for MobileNet model

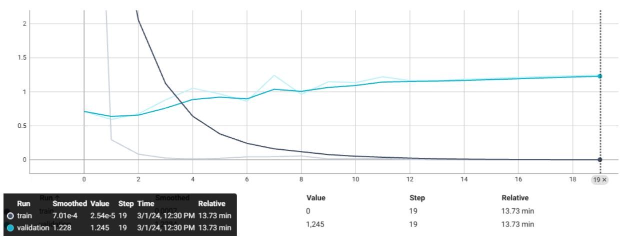

Figure 9: epoch Loss for model 2 (w Conv)

5. Conclusion

In conclusion the Deep Learning model performed better than the Machine learning ones, but also some of them reached some good results, even if the Feature extraction was made manually.
More work on the topic can be done in the future, by first trying some different combination of hyper parameters for the Machine Learning algorithms and also trying new algorithms.
Another possible solution will be to combine Deep Learning and Machine Learning, precisely Deep Learning to extract features, in a feature vector and use Machine learning algorithms to classify then the images using those feature vectors, like was done in the paper *MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers* [4].

References

[1] Bo Chen Andrew G. Howard, Menglong Zhu. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arxiv.org*, 2017.

[2] Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, and Swati Kanchan. Brain tumor classification (mri), 2020.

[3] IBM Education. The benefits of ai in healthcare. Technical report, IBM, 2023.

[4] Jeonghwan Gwak Jaeyong Kang, Zahid Ullah. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. *MDPI*, 2021.

[5] Wesseling P Louis DN, Perry A. The 2021 who classification of tumors of the central nervous system: a summary. *Neuro Oncol*, 2021.

[6] Majed A. Muhammad A., Imran A. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *MDPI*, 2020.