Glioma Diagnosis Aid through CNNs and Fuzzy-C Means for MRI

I. Amaya-Rodriguez, L. Duran-Lopez, F. Luna-Perejon [a], J. Civit-Masot [b],
J. P. Dominguez-Morales [c], S. Vicente [d], A. Civit [e], D. Cascado [f] and A. Linares-Barranco [g]
Robotics and Technology of Computers Lab., University of Seville, Seville, Spain

Abstract: Glioma is a type of brain tumor that causes mortality in many cases. Early diagnosis is an important factor. Typically, it is detected through MRI and then either a treatment is applied, or it is removed through surgery. Deep-learning techniques are becoming popular in medical applications and image-based diagnosis. Convolutional Neural Networks are the preferred architecture for object detection and classification in images. In this paper, we present a study to evaluate the efficiency of using CNNs for diagnosis aids in glioma detection and the improvement of the method when using a clustering method (Fuzzy C-means) for pre-processing the input MRI dataset. Results offered an accuracy improvement from 0.77 to 0.81 when using Fuzzy C-Means.

1 INTRODUCTION

A brain tumor is an abnormal growth of brain cells. Tumors can be malignant or benign. Malignant tumors are cancer cells that grow quickly and can cause metastasis. These brain tumors can be primary (they start in the brain) or metastatic, i.e., coming from another part of the body (Zülch, 2013).

Doctors diagnose these tumors after a neurological examination and other tests such as Magnetic Resonance Images (MRI), computerized tomography or biopsy. The treatment includes surgery, radiation, chemotherapy and directed therapy, with medicines and other substances to identify and attack specific cancer cells and save normal ones.

There exist several kinds of primary brain tumors. Their names depend on the type of affected cells (Gavrilovic, 2005):

a) Gliomas. These tumors start in the brain or the spinal cord, and they c beinclude astrocytomas, ependinomas, glioblastomas, oligoastrocytomas and oligodendrogliomas.

b) Meningiomas. This type of tumor arises from the membranes around the brain and the spinal cord (meninges). Most meningiomas are not cancerous.

c) Acoustic Neuromas (Schwannomas). These are benign tumors that develop in the nerves that control balance and audition from the inner ear to the brain.

d) Pituitary Adenomas. These are mainly benign tumors that grow in the hypophysis, in the base of the brain. They can affect the pituitary hormones and cause effects throughout the entire body.

e) Medulloblastomas. These are the most frequent cancerous tumors in children. They develop in the lower back of the brain and they extend through the cerebrospinal fluid.

f) Germ Cell Tumors. They can develop during childhood when the testicles or ovaries are formed. Nevertheless, sometimes these tumors affect other parts of the body, such as the brain.

[a] https://orcid.org/0000-0002-4352-8759
[b] https://orcid.org/0000-0003-3306-3537
[c] https://orcid.org/0000-0002-5474-107X
[d] https://orcid.org/0000-0001-9466-485X
[e] https://orcid.org/0000-0001-8733-1811
[f] https://orcid.org/0000-0002-6138-1009
[g] https://orcid.org/0000-0002-6056-740X

g) Craniopharyngiomas. These non-cancerous tumors are infrequent and they develop near the pituitary gland.

In this paper we focus on three tumors: glioma, meningiomas and pituitary tumors.

1.1 Pituitary Tumor

It is an abnormal growth of brain cells in the pituitary gland of the brain. This tumor can affect several hormones that are regulateing important functions of the body. Most pituitary tumors are benign, also called adenomas. They stay in the pituitary gland or surrounding tissues and do not disseminate to other parts of the body. There are several treatment options, such as surgery, to remove the tumor, or the use of medicines, to control its growth and the hormone levels. Figure 1 shows a coronal (left) and sagittal (right) MRIs.
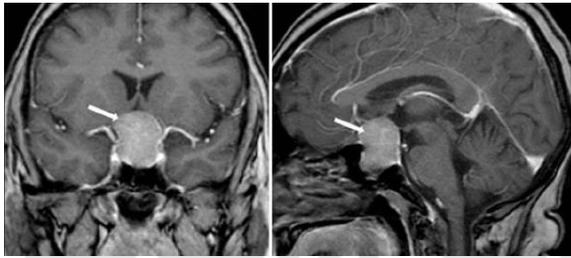

Figure 1: Pituitary tumor MRI.

1.2 Meningeal Tumor

It develops in the meninges, which is the tissue that surrounds the brain. Although technically it is not a brain tumor, it is included in this category because it compresses and presses the brain, nerves and adjacent vessels. Meningiomas are benign tumors and they rarely become malignant. The most frequent localization is in the convexity or base of the skull. They usually grow slowly, although sometimes they cause severe disabilities. Figure 2 shows axial (left) and coronal (right) MRIs with this tumor.

1.3 Glial Tumor

Gliomas are the most common type of primary brain tumor. They develop in the brain and the spinal cord, starting in glial cells present around the nerve cells (DeAngelis, 2001).

---

[1] American Brain tumor association. "Glioblastoma and Malignant Astrocytoma. ABTA-2017.
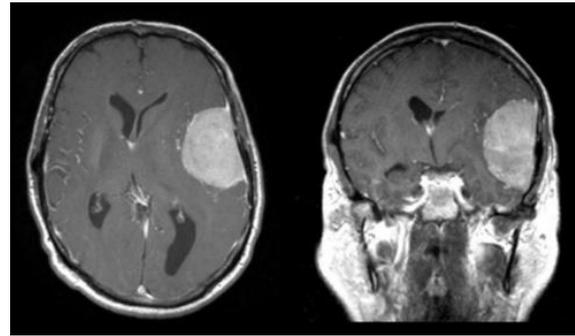

Figure 2: Meningioma tumor MRI.

Glial tumors are classified depending on the glial cell they start and on the genetic characteristics of the tumor. These types are astrocytomas, ependiomomas and oligodendrogliomas.

Glioma tumors are divided into 4 grades depending on their malignancy and aggressiveness. Grades I and II are benign and grades III and IV are malignant (DeAngelis, 2001). Glial tumors affect the brain function and they can cause mortality depending on the location and growth speed.
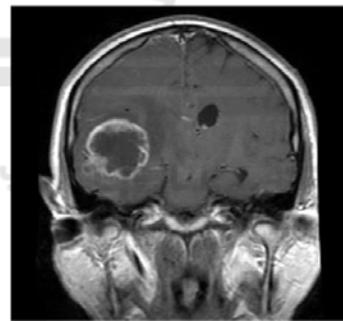
Figure 3 shows a coronal MRI of the glial tumor.

Figure 3: Glial tumor MRI.

For adults with the more aggressive glioblastoma (astrocytoma), treated with concurrent temozolomide and radiation therapy, the median survival is about 14.6 months with a two-year median survival rate of 27%; five-year survival is 10%. However, there are case reports of patients surviving for 10-20 years.[1]

Magnetic Resonance Imaging (MRI) is the diagnosis test used to detect these tumors. MRI is a technolgy that generates detailed 3D anatomic images without the need of hazardous radiation. It is frequently used for disease detection, diagnosis and treatement monitoring. Its technology stimulates and detects proton direction changes in rotation axes present in the water of the tissues. MRI uses powerful

magnets that change these protons orientations. MRI sensors detect these protons energy changes and produce gray-level images taking into account the latency and amount of energy that protons, from different tissues, need to return to their initial states (Ogawa, 1990). The lower the latency, the brighter the tissue appearas in the image.

In this paper, we present a viability study for diagnosis aid in glioma tumor detection using deep-learning based on two convolutional neural networks. We used a public available MRI dataset together with a pre-processing technique, called Fuzzy C-means, in order to compare the classification accuracies of the CNNs with or without the pre-processing technique.

The rest of the paper is structured as follows: Ssection 2 presents the used dataset and the pre-processing technique selected for this study, section 3 presents two CNN architectures, Fuzzy C-Means and their results, and section 4 presents the conclusions.

2 MRI DATASET CLUSTERING WITH FAST AND ROBUST FCM

The used dataset (Cheng, 2016) is available at Figshare. It consists of a collection of 3064 MRI images from 233 different patients. These images are labeled with three types of brain cancer: meningioma, glioma and pituitary tumor. Images are T1-CE MRI and include coronal, sagittal and axial planes. From these images, there are 1426 that include a glial tumor (from 89 patients), 708 that include meningioma (from 82 patients) and 930 that include a pituitary tumor (62 patients). Some images are 512x512 pixels and others are 256x256. Therefore, all of them were resized to 256x256. Figure 4 shows sample images.

Clustering is a data association technique when the data share particular characteristics. Mathematically, it consists in finding a point in an N-dimensional space that could represent the characteristics of a set of data defined in that space. Clustering algorithms are widely used in data mining. The most popular ones are K-NN, K-means, and Fuzzy C-means (FCM) (Bezdek, 1984). The latter allows associating each data to all the clusters defined in the same N-dimensional space. Therefore, each data could be associated to several clusters. This is known as a fuzzy association of clusters. These algorithms define a grouping criterion based on a fuzzy partition used to express the objective function.
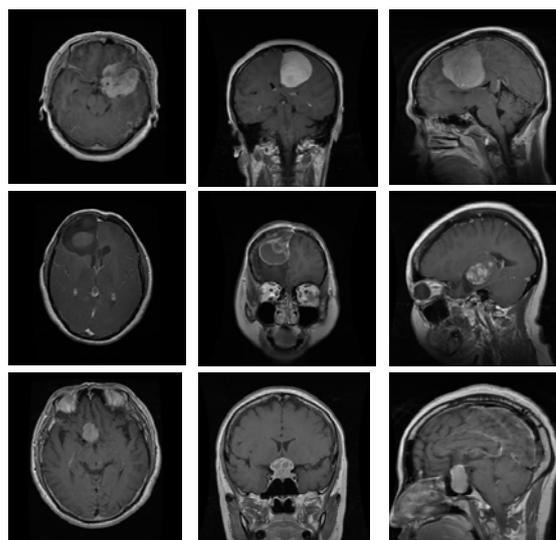

Figure 4: From left to right: axial, coronal & sagittal MRI. Top: Meningeal tumor. Middle: Glial tumor. Bottom: Pituitary tumor.

The Fast and Robust Fuzzy C-Means (FRFCM) improves in time and results the Fuzzy C-Means (Bhide, 2012). In this method, the membership relationship is improved by applying a morphologic reconstruction and a filter from the membership partition matrix.
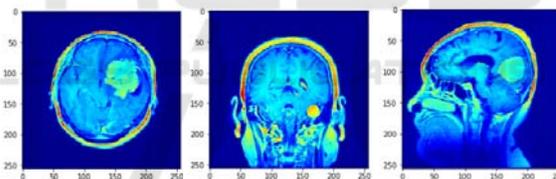

Figure 5: From left to right: axial, coronal and sagittal colored MRI images before Fuzzy C-means.

The most complex part of any clustering algorithm is to determine the number of clusters beforehand. The Fuzzy Partition Coefficient (FPC) is a measure used to determine the optimal number of clusters for a fixed set of data. This FPC is defined in the range from 0 to 1. Although this coefficient is typically used to minimize the number of clusters, in our case, we aim to find an adequate number of clusters that will serve to improve the discrimination of the tumor by selected CNNs. In this study, this FPC was used to determine the optimal number of clusters between 2 and 10 for our set of data. Several MRIs were processed with this clustering measure in order to determine the best number of clusters. After that, the MRI dataset was processed with the Fuzzy algorithm to obtain clustered images, which will

represent the training, validation and test dataset for our deep-learning architecture.
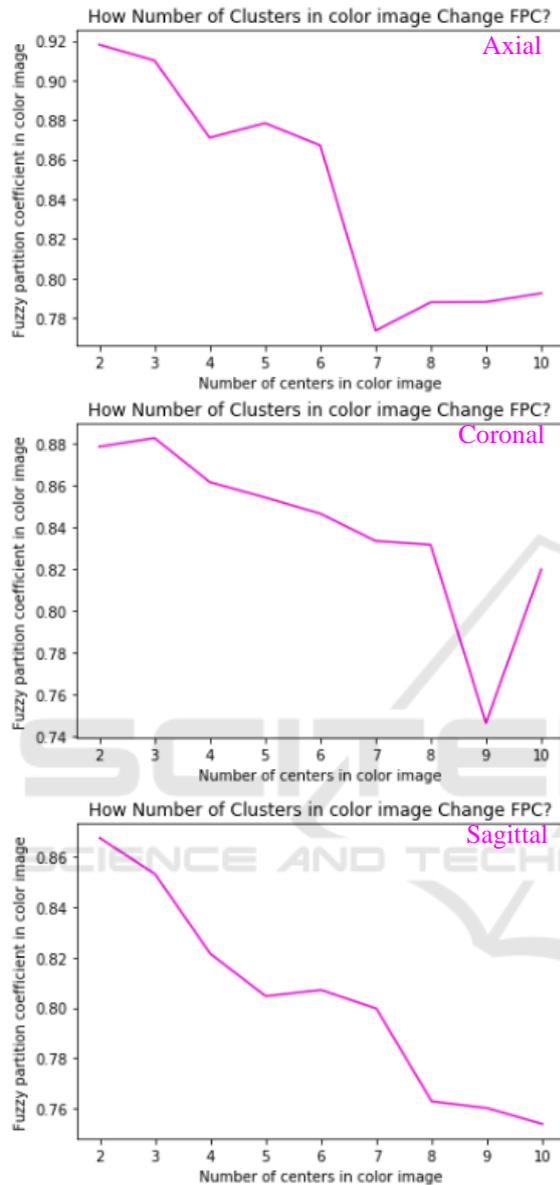

Figure 6: FPC for axial (top), coronal (mid) & sagittal (bottom) planes for 2-10 clusters.

The aim of obtaining this modified dataset was to evaluate whether the accuracy of the Deep-learning architecture can be improved by including the segmentation process. This pre-processing implementation was performed with the use of Jupiter Notebooks and Python.

In order to determine the best number of clusters, the FPC was calculated for each kind of plane from the MRI dataset. In our case, the lower the index, the better the number of clusters for the Fuzzy C-Means

colormap images conversion. Figure 5 shows colored MRIs before applying the Fuzzy C-means.

Figure 6 shows the results of calculating the FPC for the three-plane images shown in Figure 5. As can be seen, 7 clusters is the best choice for the axial plane, 9 for coronal, and 10 for sagittal. In this study, 9 clusters were selected, since, on average, the use of these 9 clusters has the lowest FPC value. The whole MRI dataset was processed with Fuzzy C-means to be used to train the selected CNN architecture.

3 CONVOLUTIONAL NEURAL NETWORKS

CNNs were successfully used for the first time in 1998 (LeCun, 1989), although they gained interest after the impressive results obtained in the ImageNet competition in 2012 (Russakovsky, 2015). The name of the CNN architecture used in that competition was AlexNet (Krizhevsky, 2012), where a set of around 1 million images that represented around 1 thousand different classes, was reduced by this CNN to almost half the error rate with respect to the previous winner algorithm. A typical CNN architecture is composed of several layers of convolution operations, pooling, activation and classification (fully connected). The convolutional layers produce characteristic maps by applying a convolution function to the input image. The pooling layer is used to reduce the size of the output of the convolution layer by using MaxPooling or AveragePooling. This output represents the input of the next convolutional layer. Typically, a convolutional layer can be finalized through an activation function. The most common ones are the Rectified Linear Units (ReLU) (Nair, 2010) and Leaky ReLU (Xu, 2015). A ReLU transforms any negative data into zero, and it keeps the positive values with no change. The classification layer produces series of predictions over the output classes. These predictions are connected to a Loss function, which is used during back-propagation (LeCun, 1989) training to determine how good the selected weights are for the architecture. These weights are also known as architecture parameters and they are fixed after the training phase. The Loss function measures how close is the prediction of the CNN with current modified weights with respect to the ground truth used for training.

Before the training procedure, the dataset must be divided into three parts, and each part must correspond to completely different patients to ensure that there will be no relationships between data of

these three parts, which are called *training*, *validation* and *test*. The training part is used by the back-propagation training algorithm, as explained in the next section. The validation part is also used during training after each iteration, once the weights are fixed to check the goodness of the calculated weights. The last part of the dataset (test) is used once the weights calculation is finished and the training procedure is finished and validated. This test is used to calculate and offer the final accuracy values of the CNN.

The training process starts with a set of random weights for all the convolutional layers and the fully connected layers. After each training iteration, the architecture is validated with validation MRI samples from the dataset. Then, the loss function is calculated. This process is automatic and shows iteratively the accuracy (prediction classes output) and loss function values. The number of iterations depends on the complexity of the architecture, the number of output classes and the size of the dataset.

In this study, we used two different CNN architectures for Glial tumor MRI classification: LeNet-5 and GoogleNet.

3.1 LeNet-5

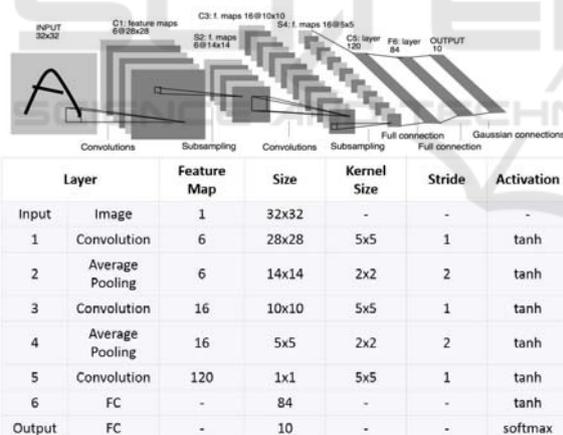| Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	32x32	-	-	-
1	Convolution	6	28x28	5x5	1	tanh
2	Average Pooling	6	14x14	2x2	2	tanh
3	Convolution	16	10x10	5x5	1	tanh
4	Average Pooling	16	5x5	2x2	2	tanh
5	Convolution	120	1x1	5x5	1	tanh
6	FC	-	84	-	-	tanh
Output	FC	-	10	-	-	softmax

Figure 7: LeNet-5 CNN architecture diagram and layer characteristics: 2 convolutional layers and 2 fully connected layers, with 10 output classes.

Figure 7 shows the architecture of the LeNet-5 architecture proposed by (LeCun, 1998). It was developed and trained for handwritten digits prediction, and consisted of 7 levels, which comprised Conv layers, Pooling and fully connected classifiers with SoftMax functions. The training dataset used (called MNIST) is composed of 60k images of handwritten digits, of which 10k were used

to test the architecture. The images size was 32x32 pixels in gray level with digits between 0 and 9.

This CNN is a basic start point widely used in many applications. It is available as a tutorial in deep-learning tools such as Caffe (Jia, 2014). The architecture can be adapted and modified to other datasets by modifying the input image sizes, kernel sizes, number of convolutions per layer, adding new convolutional layers or changing the number of output classes. Table 1 presents the CNN architecture characteristics. LeNet-5 was adapted for our MRI dataset sizes to 256x256 and output classes.

Table 1: LeNet modified architecture for MRI Glial tumor.

Parameter	Value
Input image size	256x256
C1 (K size, K #)	3x3 and 6 units
S2 (pooling size)	2
C3 (K size, K#)	3x3 and 16 units
S4 (pooling size)	2
F5 (inner-product ReLU size)	120
F6 (inner-product ReLU size)	84
Output classes	3 (meninge, glioma, pituitary)

After training, for the testing phase of the CNN, the accuracy of this network for the dataset without Fuzzy C-means was 76.81% with a loss of 3.42%. Figure 8 shows the evolution of the accuracy and loss values for both train and validation during 1000 iterations of the training process.

3.2 GoogleNet / Inception V1

This net was successfully used in similar medical applications previously, e.g., in (Havaei, 2015) for brain tumor segmentation and classification (Deepak, 2019). GoogleNet (Szegedy, 2015) won in 2014 the ILSVRC ImageNet contest (Russakovsky, 2015). It was the first time an architecture offered the optimal convolution sizes by using different sizes and combining them in a block. Several blocks are connected along the architecture to provide the deep to the net. It achieved a top-5 error rate of 6.67%, which is very close to human level performance. The network used a CNN inspired by LeNet, although it implemented a novel element called the "inception module". It used batch normalization, image distortions and RMSprop. This module is based on several very small convolutions in order to drastically reduce the number of parameters. Their architecture consisted of a 22-layer deep CNN, but with a reduced number of parameters: from 60 million (AlexNet) to 4 million. The architecture diagram is shown in Figure 9.

each dimension. The learning rate is adapted by dividing it by that sum of squares.
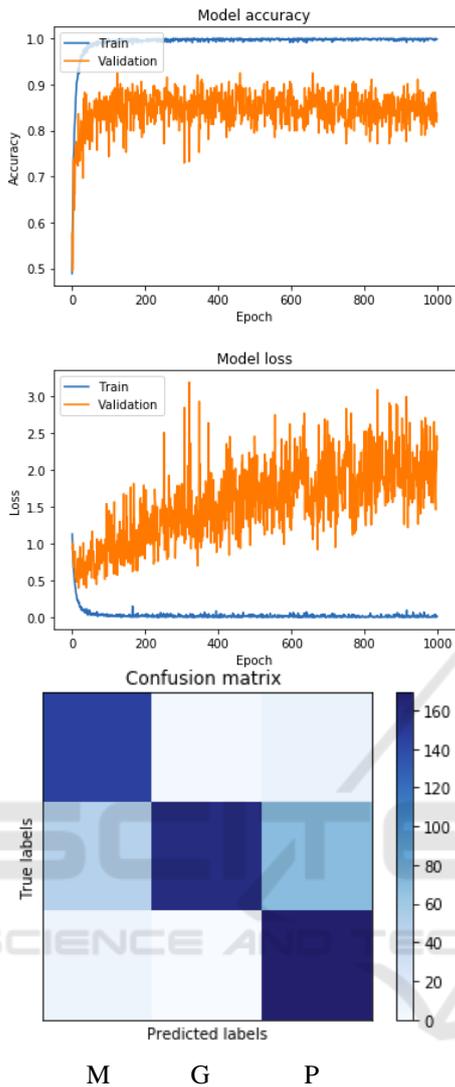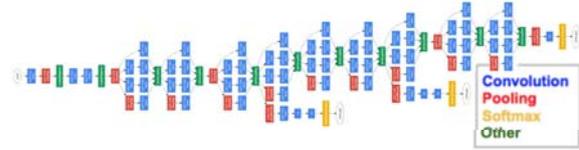

Figure 9: GoogleNet architecture.

Figure 8: LeNet accuracy, loss and confusion matrix for the three output classes: M (meninge), G (glial), P (pituitary).

RMSprop (Ruder, 2017) is an adaptive learning rate method whose popularity has increased in recent years, while also getting some criticism (Wilson, 2017). There are two ways to introduce RMSprop. On the one hand, it can be considered as the adaptation of the Rprop (Igel, 2000) algorithm for mini-batch learning. It was the initial motivation for developing this algorithm. Another way is to look at its similarities with Adagrad (Duchi, 2011) and view RMSprop as a way to deal with its radically diminishing learning rates. Rprop uses the sign of the gradient and adapts the step size individually for each weight. Adagrad adds element-wise scaling of the gradient based on the historical sum of squares in
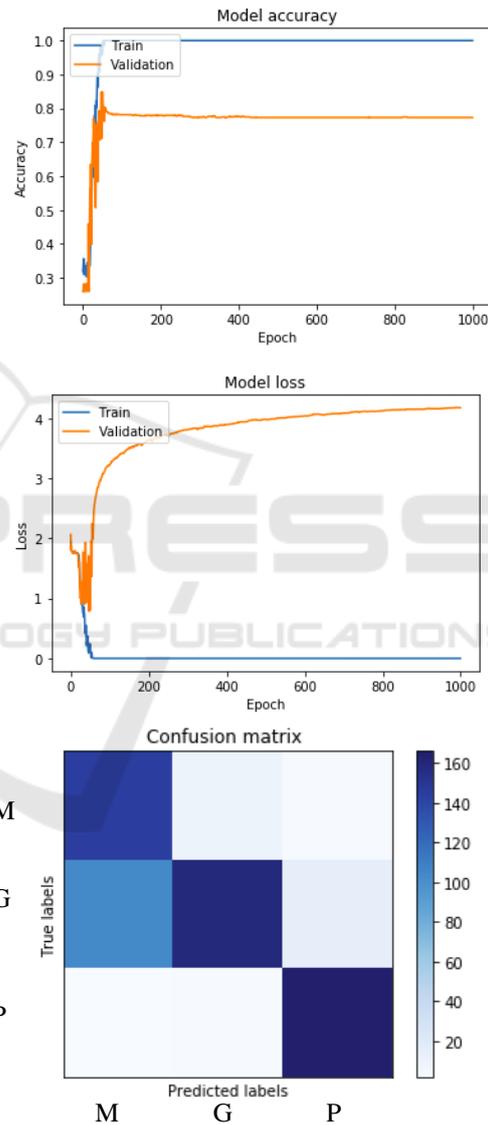
Figure 10: GoogleNet accuracy, loss and confusion matrix for the three output classes: M (meninge), G (glial), P (pituitary).

Each block of the net is composed of 4 parallel routes. From left to right, the first three paths are convolutions of 1x1, 3x3 and 5x5 kernels. These

extract information from the input features at different sizes. The two paths in the middle implement 1x1 convolutions to reduce the number of input channels for the next convolutions, which reduces the complexity of the model. The fourth path uses a MaxPooling followed by a 1x1 convolution. Then, each of these paths apply lately transformation to obtain coincidences int the outputs. Finally, the four outputs are concatenated to offer the output of the block. After training, for the test phase of the CNN, the accuracy of this network for the dataset without Fuzzy C-means was 77.3% with a loss of 2.65%. Figure 10 shows the evolution of the accuracy and loss values for both training and validation during 1000 iterations of the training process.

3.3 Introducing Fuzzy C-Means

In order to test whether the use of Fuzzy C-means improves the accuracy of glial tumor detection from MRI datasets, we repeated the LeNet-5 training with the pre-processed dataset using Fuzzy C-means with 9 clusters.

The accuracy results obtained were 80.87% with a loss of 2.71%. Figure 11 shows the confusion matrix, which demonstrates the efficiency of the proposed architecture.
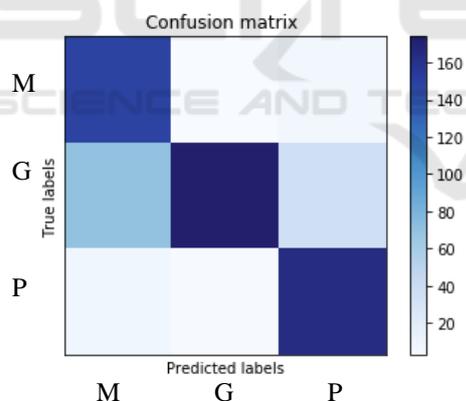

Figure 11: LeNet-5 with Fuzzy C-means pre-processed MRI dataset Confusion matrix for the three output classes: M (meninge), G (glial), P (pituitary).

4 CONCLUSIONS

This work presents a deep-learning study applied to glial tumor detection over MRI for diagnosis aid. It presents the analysis of two convolutional neural network architectures: LeNet-5 and GoogleNet. In order to improve the classification accuracy, the dataset was pre-processed using the Fuzzy C-means

(FCM) clustering method to apply a colormap over a fixed number of clusters. The best number of clusters was obtained over the Fuzzy Partition Coefficient (FPC) for the three different planes of images in the MRI original dataset: axial, coronal and sagittal. The obtained results justify the use of FCM for this application, improving the accuracy of the classification from 76.81% to 80.87%.

ACKNOWLEDGEMENTS

REFERENCES

Bezdek, J.C, Ehrlich, R, Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences Vol. 10, Issues 2–3, Pages 191-203.

Bhide, P.A.S, Patil, P, Dhande, S. (2012). Brain Segmentation using Fuzzy C means clustering to detect tumour Region, International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 1, n. 2, p. pp:85-90-90.

Cheng J, Yang W, Huang M, Huang W, Jiang J, Zhou Y, Yang, R, Zhao, J, Feng, Y, Feng, Q, Chen, W. (2016) Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation. PLoS ONE 11(6): e0157112. https://doi.org/10.1371/journal.pone.0157112.

DeAngelis, L.M. (2001). M.D. Brain Tumors. N Engl J Med; 344:114-123.

Deepak, S, Ameer, P.M. (2019). Brain tumor classification using deep CNN features via transfer learning, Computers in Biology and Medicine, vol. 111, p. 103345.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adagrad: Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research, 12, 2121–2159.

Gavrilovic, I.T. & Posner, J.B. J Neurooncol (2005) 75: 5. https://doi.org/10.1007/s11060-004-8093-6

Havaei, M, Davy, A, Warde-Farley, D, Biard, A, Courville, A, Bengio, Y, Pal, C, Jodoin, P.M, Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks, Medical Image Analysis, vol. 35, pp. 18-31.

Igel, C, Hüsken, M. (2000). Improving the Rprop Learning Algorithm, Proceedings of the second International Symposium on Neural Computation.

Jia, Y, Shelhamer, E, Donahue, J, Karayev, S, Long, J, Girshick, R, Guadarrama, S, Darrell, T. (2014). Caffe:

Convolutional Architecture for Fast Feature Embedding. arXiv:1408.5093.

Krizhevsky, A., Sutskever, I., Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25, 1097-1105, NIPS - 2012.

LeCun, Y, Boser, B, Denker J.S, Henderson, D, Howard, R.E, Hubbard, W, Jackel, L.D.(1989). Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, vol. 1, n. 4, pp. 541-551.

Lecun, Y, Bottou, L, Bengio, Y, Haffner, P. Gradient-based learning applied to document recognition (1998). Proceedings of the IEEE. Pages 2278-2324.

Nair, V, Hinton, G. (2010). Rectified linear units improve restricted Boltzmann machines. Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814.

Ogawa, S, Lee, T.M., Kay, A.R., Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. PNAS, 87 (24), 9868-9872.

Ruder, S. (2017). An overview of gradient descent optimization algorithms. arXiv:1609.04747v2.

Russakovsky, O, Deng, J, Su, H, Krause, J, Satheesh, S, Ma, s, Huang, Z. Karpathy, A, Khosla, A, Bernstein, M, Berg, A.C, Fei-Fei, L. (2012). ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis (2015) 115: 211. https://doi.org/10.1007/s11263-015-0816-y.

Szegedy, C, Liu, W, Jia, Y, Sermanet, P, Reed, S, Anguelov, D, Erhan, D, Vanhoucke, V, Rabinovich, A. (2015). Going Deeper with Convolutions. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9.

Wilson, A.C, Roelofs, R, Stern, M, Srebro, N, Recht, B. (2017). The Marginal Value of Adaptive Gradient Methods in Machine Learning. arXiv:1705.08292v2

Xu B, Wang N, Chen T, Li, M. (2015) Empirical evaluation of rectified activations in convolutional network. arXiv:1505.00853

Zülch, K.J. (2013). Brain Tumors: Their Biology and Pathology. Springer-Verlag ISBN 978-1-4899-6264-5.