Novel automatic tumor extraction method based on decision Tree Classifier

Sara SANDABAD [#1], Yassine SAYD TAHRI [#2], Achraf Benba[#3], Ahmed HAMMOUCH [#4]
[#]Laboratory LRGE, ENSET, Mohamed V University, Rabat, Morocco.
[1] sandabad.sandabad@gmail.com
[2] saydtahri@gmail.com
[3]achraf.benba@um5s.net.ma
[4] hammouch_a@yahoo.com

Abstract—Early detection of a brain tumor is much easier to cure than to find out at an advanced stage. In this article we present a tumor extraction method based on the decision tree classifier for early and rapid detection. This method can also provide the tumor characteristics, the exact location and it geometric properties. Our method will be applied to a huge number of MRI images to prove its accuracy, and speed. This method is carried out in three stages: the first step is filtering the image using the Nl-means filter. The second step is the brain extraction of the skull, to retain only the brain, in order to avoid the influence of another substance on the detection of the tumor, and at the last step the tumor area extraction of with the detection tree classifier.

Keyword- Brain tumor, MRI, decision tree classifier, Nl-means filter, Dice, Jaccard.

I. INTRODUCTION

Brain tumors make no distinctions they affect people of all ages, anywhere and anytime. we record the survival statistics for brain cancer is 25%, which means that a person diagnosed with brain cancer would, on average, 25% chance of living five years after diagnosis.

In 2006, the score of the criteria of tumors of the central nervous system is maximal for their connection with the environment. That is to say that the tumors of the central nervous system (head, parotid gland ...) are those that have the greatest increase and must be the most supervised, according to the National Institute for Public Health Surveillance.

Basis in MR images with contrast product injection, we will attempt to detect and extract the tumor area and its characteristics.

Extract brain tumors consist in removing the tumor area and give its geometric properties. The goal of our work is to detect in an automatic manner a tumor areas, this will free up radiologists of this tedious task of determining the tumor parameters from MRI.

In this axis several methods have been proposed, but all of them present some disadvantages as: some methods require the supervisor presence to change settings in each new tumor case (we can no longer talk about an automatic method), other methods require several acquisitions of MRI protocols, other take a very enormous simulation time…

We present a brief history of tumor detection : in 1994 Phillips presents the Fuzzy clustering method [1], 1996 Peck presents Eigenimage analysis [2], 1997, Zhu Yan and present a method based on Hopfield neural network and active contours [3], in 1998 Clark proposes Knowledge-based fuzzy clustering [4] with 70% accuracy but with a very important simulation time, in 1999 Kaus proposes Segmentation of meningiomas and Low Grade Gliomas in MRI [5] with 55% an accuracy and with a simulation time varying with the kind of image to be simulated, also in 1999 Vinitski presents k-Nearest neighbor [6], with an interesting simulation time "2 min", but poor accuracy, That same year Karayiannis presents Fuzzy clustering [7] with very low accuracy, in 2000 Fletcher Heath presents Knowledge-based fuzzy clustering [8] with 53% accuracy, in 2002 Ho presents the method based on 3D level [9] sets an accuracy that can grow 93%, but with a very important execution time, in 2003 Prastawa presents Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors [10], with an accuracy which can go up 71%, but a huge execution time, which is about 100 min, in 2004 Prastawa presents Knowledge-based / outlier detection [11] with an accuracy which can go up 80%, but with a high execution time is 90 min, in 2005 Lee presents the method based on discriminative Random Fields and SVM [12] with an accuracy of 40%, in 2008, Corso presents the method based on Multilevel Bayesian segmentation [13], with an accuracy that can grow 88% and an execution in 7 min, and most recently in 2014 Li and Guo present Fuzzy Clustering [14] with excellent accuracy “99%”, a modest simulation time “16 min”, but unfortunately the method is semi-automatic.

We present in this article our original approach tumor detection, with a reduced simulation time in comparaison with the methods already cited, and with an important accuracy that allows to present a very good result.

This method is based primarily on the extraction of the tumor by classification of MRI image into two classes tumor, and no tumor classes, the classification will be realized by using the decision tree classifier after brain extraction and MRI filtering.

II. PROPOSED METHOD

The MRI image quality depends on the following parameters:

- spatial resolution and contrast
- noise and artifacts

So to extract the tumor area, we must begin with the step of smoothing the image and noise reduction. This step will avoid us having bad result, which may be due to the influence of noise on the image.

A. Filtering MRI image by using the filter Nl-means:

The goal of denoising is to remove the noise and/or spurious details from a given possibly corrupted digital picture while maintaining essential features such as tumors.

This algorithm is based on the redundancy of information in the image. In fact, it looks for the pixel with similar neighborhood to be denoised, and then replace the value of the noisy pixel by the non-local average of other pixels that are similar in order to remove noise.

If compared with other well-known denoising techniques, such as the Gaussian smoothing model, the anisotropic diffusion model, the total variation denoising, the neighborhood filters and an elegant variant, the Wiener local empirical filter, the translation invariant wavelet thresholding, the non-local means method noise looks more like white noise. Recently non-local means has been extended to other image processing applications such as deinterlacing and view interpolation.

B. Brain extraction by using EM algorithm:

Method presented by our team " S. Sandabad and al. [15]". The goal of this step is to extract the brain from the brain MRI acquired. The region of interest now is the brain. To extract it, we use the AND operator between the original image and the binary mask obtained by the EM algorithm. This method is very simple, and requires very little time to execution of most of the methods presented earlier, it can be divided into four steps as follows:

1. Creation of the mask using the EM algorithm:

The expectation maximization algorithm was originally described in its general form by Dempster [16] it has been widely applied to estimate the hyper-parameters in statistical segmentation of MRI images. It is often used in evaluation problems where some missing data.

The principle is: it alternates evaluation expectancy (E) steps and and a maximization step (M), in the first step it calculates the expectation of the likelihood taking account the recent observed variables; and in the second step it estimates the maximum likelihood of the parameters by maximizing the likelihood found in (E) step; it still uses the settings found in M as the starting point of a new phase of evaluation expectancy, finally as it iterates phases E and M. The objective of the EM is then: find the maximum likelihood parameters of probabilistic models when the model depends on unobserved latent variables.

EM algorithm is defined as follows:

- Initialization randomly
- c = 0
- As long as the algorithm has not converged do
- Evaluation of the expectation (E step)
- Maximization (M-step)
- c = c + 1
- end

Log-likelihood is $L(x;\theta) = \sum_{i=1}^{n} logf(x_i\ ,\theta)$

2. Extraction of the largest connected component (brain):

Based on the result of step 'creation of the mask with the EM algorithm, we will need in the second step (which is the extraction of the largest connected component brain mask), so to achieve this we have uses statistical analysis of existing connected components in the image result of EM.

3. dilatation (softening of the mask outline):

After obtaining the mask with the expectation maximization algorithm, we need to soften the exterior boundaries of the component, for this we made use of mathematical morphology, the principle here is the expansion of adjacent white areas so we moved a circular structuring element surface (π * r2) on the resulting image, and applying a logical or (sum) on each of the ((π * r2) -1) neighboring pixels An empirical study and assigning different values to r; we went out with r = 1.5 as the radius that gives the best result of the brain mask.

4. Brain extraction:

Finally to extract our region of interest and that is the brain, we use the operator "and" between the original image and the final mask obtained.

The following figure illustrates these steps.

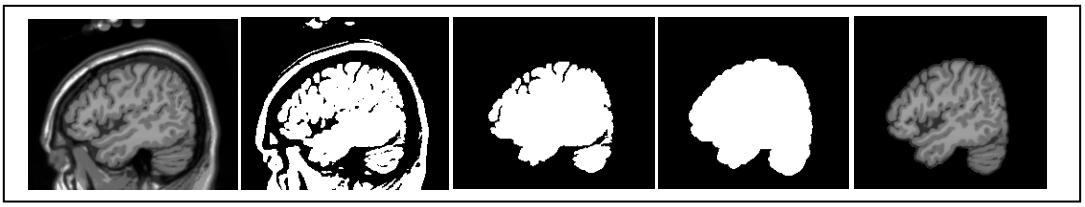

Fig. 1: a- MRI filtered by Nl-means; b- EM brain mask; c- Extraction of the largest component; d- dilatation; e- brain extraction.

Now after extraction of the MRI brain image, the third step is the extraction of the tumor area by the decision tree classifier.

C. Tumor extraction by using decision tree classifier

Initially to perform image classification MRI tumor and non tumor area, we started by filtering and extracting the brain. After collecting the resulting images of these two steps, we must apply the decision tree.

This classification algorithm is based on a decision tree. A decision tree is a set of simple rules. Decision trees [17] are also nonparametric because they do not require any assumptions about the distribution of the variables in each class. Every interior node contains a decision criterion depending only on one feature. For the first split into two parts, the feature with the highest relevance is used. This procedure is recursively repeated for each subset until no more splitting is possible. It followed from a root to a leaf node the decision tree corresponds to a rule-based classifier. An advantage of decision tree classifiers is their simple structure; the most important features are near the root node. A decision tree is built from a training set, which consists of objects, each of which is completely described by a set of attributes and a class label. The class that is associated with the leaf is the output of the tree. A tree misclassifies the image if the class label output by the tree does not match the class label.

The delicate part of the classification with association rule mining is the construction of the classifier. For this we always require that MRI should be taken treated with an injection of contrast product, the product allows contrast to the tumor area to take greater intensity value than non-tumor areas, which allow deciding a threshold after an empirical study, which is equal to 160, and then the general rule became:

Superior intensity at 160 involves tumor area;

Lower intensity than 160 involves no tumor area.

The association rules generated from the database in such a manner that they have as consequent a category (tumor aera) from the classification classes. The association rules could imply either normal or abnormal. When a new image has to be classified, the categorization system returns the association rules. The first intuition in building the classification system is to categorize the image in two classes (tumor and no tumor). This classification would work when the number of rules extracted for each class is balanced. In dealing with MRI images it is very important that the false negative rate be as low as possible. It is better to misclassify a normal image than an abnormal one. So in our tuning phase we take into consideration the recognition rate of abnormal images. It is not only important to recognize some images, but to be able to recognize those that are abnormal.

There we give the decision Tree algorithm:

Input : training dataset T.

Freatures F.

Output: Decision tree Tree.

Set Tree ={}

For f ϵ F do

Set Info (T) = 0

And split (T) = 0

Compute Entropy(f)

For v ϵ values (T) do

Set T subset of T where f = v

$\text{Info (T)} = \sum \frac{|\text{T vf}|}{\text{Tf}} \text{ Entropy (f)}$

$\text{Split (T)} = - \sum \frac{|\text{Tfv}|}{|\text{Tf}|} \log \frac{|\text{Tfv}|}{|\text{Tf}|}$

End for

Gain f (T) = Entropy(f) – Info (Tf)

$\text{GainRatio (T)} = \frac{\text{gainf (T)}}{\text{splitf(T)}}$

End for

Set best_f = max{GainRationf(T)}

Add best_f into Tree

Return Tree

We present in the following section the results and comparison of our method with another method often used in the literature.

III. RESULTS AND DISCUSSION

To prove the effectiveness of this method (based on the decision tree classifier), we apply it on a database of real MRI images and so on other database of simulated MRI images. We present the characteristics of the images used in the following paragraph

- Forty simulated weighted MRI images T1 size (181x217x181 pixels) are provided by the base Brain web data [http://brainweb.bic.mni.mcgill.ca/brainweb/]. We note that the manual segmentation of such images is also offered by Brain web.
- Twenty five real weighted MRI images T1 size (256x256x128) are retrieved from the database of MRI and manual segmentation of experts.(Internet brain segmentation repository IBSR V2.0) [www.cma.mgh.harvard.edu/ibsr] used by the center of morphomertry analysis of Massachusetts General Hospital. All MRI data and manual segmentations are provided by the experts' radiologists of the center.

The method of detection and automatic extraction of the tumor presented above can treat a variety of tumors since our process is not concerned with the location or the nature the tumor. But, it is necessary to inject to the patient a contrast produce prior to MRI session.

We will apply our method on authentic and simulated MRI images; the implementation of the program will be made with a scientific programming language "MATLAB 7.8". We'll execute the program on a personal computer processor Intel Core i3 and 2.3GHz frequency with 4GB memory and an Intel HD Graphics 3000. We present the results in the next section.

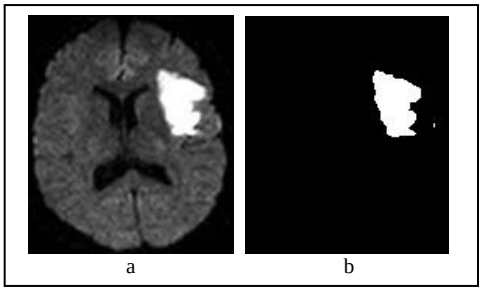

Fig. 2: a- MRI filtered by Nl-means and EM step; b- tumor extraction

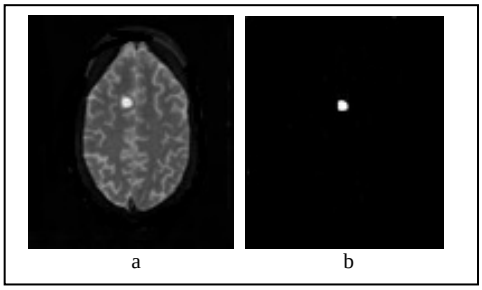

Fig. 3: a- MRI filtered by Nl-means and EM step; b- tumor extraction

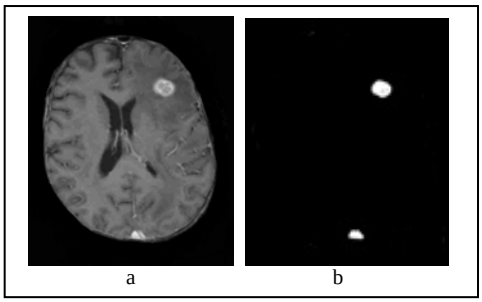

Fig. 4: a- MRI filtered by Nl-means and EM step; b- tumor extraction

We will identify the geometric properties of the tumor in Figure 2, its features are presented in Table 1.

TABLE 1: Characteristics of the tumor of figure 2

Area of the tumor:	4320
Real basis of the tumor region:	362
Width:	62
Height:	78
Orientation angle (deg):	-78.25
Eccentricity:	0.63
Compact:	0.92
Elongation:	0.77
Euler number:	1

Among the measures of the performances the most used in the literature to compare our method and the other tumor extraction methods most used in the literature, we have chosen Jaccard's similarity coefficient; Dice's similarity coefficient; the sensitivity; the specificity; precision and accuracy

1. *Jaccard's similarity coefficient: (*JSC)

Jaccard's similarity coefficient [18] is provided by the equation (1):

$$JSC = \frac{Card\ (R1 \cap R2)}{Card\ (R1\ U\ R2)} \quad (1)$$

- R1: The region automatically segmented.
- R2: The region manually segmented.
- Card (x): Indicates the number of pixels in region x.

We planned during our evaluation to normalize the results between 0 and. So that if Jsc value is 1, the result is perfect (this is what we desire). If the value is 0, we conclude that the result is far from what would be desired.

2. *Dice's similarity coefficient:*

The Dice similarity coefficient [19] is frequently used to calculate the degree of similarity either from two regions automatically segmented or between two regions: one of which is automatically segmented or the other is manually as in our case. Dice coefficient is provided by the equation (2):

$$DSC = 2 * \frac{Card\ (R1 \cap R2)}{Card\ (R1 + R2)} \quad (2)$$

- R1: The region automatically segmented.
- R2: The region manually segmented (reference).
- Card (x): pixels in region x.

Like the evaluation criteria of Jaccard, we have normalized the results in an interval [0, 1] in which (1) corresponds to a good result (desired) and (0) corresponds to a bad result (unwanted).

3. Sensitivity:

The possible outcomes of a two class prediction be represented as True positive (TP), True negative (TN), False Positive (FP) and False Negative (FN). The normal and abnormal images are correctly classified as True Positive and True Negative respectively. A False Positive is when the outcome is incorrectly classified as positive when it is a negative. False Positive is the False alarm in the classification process. A false negative is when the outcome is incorrectly predicted as negative when it should have been in fact positive.

We define:

- TP= True Positive as number of Abnormal images correctly classified;
- TN= True Negative as number of normal images correctly classified;
- FP= False Positive as Number of Normal images classified as Abnormal;
- FN= False negative as Number of Abnormal images classified as Normal.

The sensitivity is the percentage of pixels recognized by the algorithm. It is given by the equation (3):

$$\frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

4. Specificity:

Specificity is the percentage of pixels which is not recognized by the algorithm. It is given by the equation (4):

$$\frac{TN}{TN+FP} \tag{4}$$

5. Accuracy:

The fraction of test results those are correct.

$$\frac{TP+TN}{TP+FN+TN+FP} \tag{5}$$

6. Precision:

The fraction of abnormal images on correct results in the equation (6)

$$\frac{TP}{TP+FP} \tag{6}$$

We give in the Table2 the application results of all this evaluation criteria.

TABLE 2: Results for T1-weighted MRI volume 181*217*181 from brain web dataset

	Modification of Atlas Priors	**Our method**	**Multilevel Bayesian**	**Knowledge based/outlier detection**
Jaccard	0.991	0.900	0.820	0.911
Dice	0.976	0.913	0.833	0.901
Sensibility	0.903	0.899	0.842	0.904
Specificity	0.812	0.975	0.801	0.906
Accuracy	0.901	0.912	0.889	0.920
Precision	0.992	0.936	0.878	0.941
Time	100 min	1 min	7 min	90 min

Referring to Table 2, we can see that the most accurate and effective methods require a considerable length time during the simulation step. Generally, the interval time is set between 90 and 100 minutes. Our method requires only one minute to locate the tumor and to extract it. Time currently is very important front of the large number of MRI have treated daily. So our method has been able to satisfy time criteria

The results are normalized between 0 and 1; results close to the value 1 are the most accurate; as we can see; in addition to the reduced time advantage, our method has very good accuracy, accuracy of our method has been able be equal to the method “Modification of Atlas Priors”; while this method requires 100 minutes in times simulation and extraction.

IV. CONCLUSION

“Tumor Detection and Classification using Decision Tree in Brain MRI” is used to get accurate and efficient result. Using Decision tree classification technique tumor has been found as well as classified in Normal or Abnormal class. Here we used two algorithms, Naïve Bayesian and Decision Tree, to compare performance. After evaluating performance we can say that the proposed algorithm has been found to be performing well compared to the existing classifiers. The accuracy of 96% and sensitivity of 93% were found in classification of brain tumor using decision tree classifier. This will produce result into normal or abnormal in efficient way. The developed brain tumor classification system is expected to provide valuable diagnosis techniques for the physicians.

REFERENCES

[1] W. E. Phillips, R. P. Velthuizen, S. Phuphanich, L. O. Hall, L. P. Clarke, M. L. Silbiger, “application of fuzzy c-means segmentation technique of differentiation in MR images of a hemorrhagic glioblastoma multiforme”, Magn. Reson. Imaging, Vol. 13, No.2, July 1994, pp. 277-290.

[2] D. J. Peck, J. P. Windham, L. L. Emery, H. Soltanian-Zadeh, D. O. Hearshen, and T. Mikkelsen, “Cerebral Tumor Volume Calculations Using Planimetric and Eigenimage Analysis,” Medical Physics, vol. 23, no. 12, 1996, pp. 2035–2042.

[3] Y. Zhu and H. Yan, “Computerized Tumor Boundary Detection Using a Hopfield Neural Network,” IEEE Transactions on Medical Imaging, vol. 16, no. 1, 1997, pp. 55–67.

[4] M. C. Clark, L. O. Hall, D. B. Goldgolf, R. Velthuizen, F.R. Murtagh, M. S. Silbiger, ” Automatic tumor segmentation using knowledge-based techniques”, IEEE Transactions on Medical Imaging, Vol. 17, No. 2, April 1998, pp. 187-201.

[5] M. R. Kaus, S. K. Warfield, A. Nabavi, E. Chatzidakis, P. M. Black, F. A. Jolesz, R. Kikinis, “Segmentation of Meningiomas and Low Grade Gliomas in MRI”, Presented at the 1999 International Conference on Medical Image Computing and Computer-Assisted Intervention, Cambridge, UK, September 1999.

[6] S. Vinitski, C. F. Gonzalez, R. Knobler, D. Andrews, T. Iwanaga, and M. Curtis, “Fast tissue segmentation based on a 4d feature map in characterization of intracranial lesions Fast Tissue Segmentation Based on a 4D Feature Map in Characterization of Intracranial Lesions,” Journal of Magnetic Resonance Imaging, vol. 9, no. 6, 1999, pp. 768–776.

[7] N. B. Karayiannis , P. Pin, “Segmentation of magnetic resonance images using fuzzy algorithms for learning vector quantization”, IEEE Transactions on Medical Imaging, Vol. 13, 1999, pp. 172-180.

[8] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, F.R. Murtagh, “Automatic segmentation of non-enhancing brain tumors in magnetic resonance images”, Artificial Intelligence in Medicine, Vol. 21, No. 1-3 , pp. 43-63, August 2000.

[9] S. Ho, E. Bullitt, G. Gerig, “Level set evolution with region competition: automatic 3D segmentation of brain tumors”, Presented at the 2002 International Conference on Pattern Recognition, Quebec, Canada, August 11-15, 2002.

[10] M. Prastawa, E. Bullitt, N. Moon, K. V. Leemput, and G.Gerig “Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors”, Acad Radiol. Vol. 10, No. 12, pp. 1341–1348, December 2003.

[11] M. Prastawa, E. Bullitt, S. Ho, G. Gerig, “A brain tumor segmentation framework based on outlier detection”, Medical Image Analysis, Vol. 8, 2004, pp. 275–283.

[12] C. H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, “Segmenting brain tumor with conditional random fields and support vector machines,” in Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision, 2005.

[13] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, “Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification”, IEEE Transactions on Medical Imaging, 2008.

[14] K. Li, and Z. Guo, “Image Segmentation with Fuzzy Clustering Based on Generalized Entropy”, Journal of Computers, Vol. 9, No. 7, July 2014, pp. 1678-1683.

[15] S. Sandabad, A. Hammouch, YS. Tahri, A. Benba, B. Cherradi , “New brain extraction method using expectation maximization and mathematical morphology”. Journal of Theoretical and Applied Information Technology; Vol. 73, pp: 368-376, 2015.

[16] A. Dempster, N. Laird, D. Rubin, “Maximum likehool from incomplete data via the EM algorithm”. Journal of the Royal Statistical Society; vol. 39, pp: 1-38, 1977.

[17] R. Baskaran, M. Deivamani, A. Kannan, “A multi agent approach for texture based classification and retrieval (MATBCR) using binary decision tree." International journal of computing and information sciences, Vol. 2, No.1, pp: 13-22, 2004.

[18] P. Jaccard, “The distribution of the flora in the alpine zone”. New Phytol; vol. 11, pp: 37-50, 1912.

[19] L. Dice, “Measures of the amount of ecologic association between species”. Ecology; vol. 26, pp: 297-302, 1945.

AUTHOR PROFILE

Sara Sandabad PhD student, was born in 1988 in Morocco, she received her Master's degree in Electrical Engineering from Higher Normal School of Technical Education (Ecole Normale Supérieure de l’Enseignement Technique – ENSET), Rabat Mohammed V University, Morocco, in 2012. She is actually a PhD student since 2013 in Sciences and Technology of the Engineer in the National School of Computer Science and Systems Analysis (Ecole Nationale Supérieure d’Informatique et d’Analyse des Systèmes - ENSIAS), Research Laboratory in Electrical Engineering LRGE, Research Team in Informatics and Telecommunication ERIT located in the Higher Normal School of Technical Education ENSET, Mohammed V University, Rabat, Morocco. She is author or co-author of several papers in international conferences and journals. His research interests include the treatment of multimedia information, biomedical engineering, image processing.

Yassine Sayd-Tahri young Moroccan researcher, born in 1984 in Morocco, After obtaining his engineering degree in computer engineering from state to Morocco Mohammedia School of Engineers in 2007 , Yassine Sayd - Tahri joined the world of work since 2007 to 2014 during this period he gained extensive experience in the field of it development Currently he is part of the Research Laboratory in Electrical Engineering , he is co-author of several papers in international conferences and journals. His research interests include the treatment of multimedia information, biomedical engineering , image processing.

Achraf BENBA, PhD student, he received his Master's degree in Electrical Engineering from Higher Normal School of Technical Education (Ecole Normale Supérieure de l'Enseignement Technique – ENSET), Rabat Mohammed V University, Morocco, in 2013. He is actually a PhD student since 2014 in Sciences and Technology of the Engineer in the National School of Computer Science and Systems Analysis (Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes - ENSIAS), Research Laboratory in Electrical Engineering LRGE, Research Team in Informatics and Telecommunication ERIT located in the Higher Normal School of Technical Education ENSET, Mohammed V University, Rabat, Morocco. His interests are in Signal processing for detecting patients who suffer from neurological disorders.

Ahmed Hammouch received the master degree and the PhD in Automatic, Electrical, Electronic by the Haute Alsace University of Mulhouse (France) in 1993 and the PhD in Signal and Image Processing by the Mohammed V Agdal University of Rabat (Morocco) in 2004. From 1993 to 2013 he was professor in the Mohammed V-Souissi University in Morocco. Since 2009 he manages the Research Laboratory in Electronic Engineering. He is an author of several papers in international journals and conferences. His domains of interest include multimedia data processing and telecommunications. He is currently head of Department for Scientific and Technical Affairs in National Center for Scientific and Technical Research in Rabat (Morocco). During these two decades, Ahmed Hammouch was a member of several bodies and commissions ENSET and UM5S. He served as Head of the Electrical Engineering Department for several years. He is responsible for setting up and monitoring of several training courses, several research projects and several research facilities including Research Laboratory in Electrical Engineering. He is author or co-author of several papers in international conferences and journals. He is also a member of the reading committee of several international conferences and journals. He has supervised and co-supervised several doctoral theses. His research interests include the treatment of multimedia information, telecommunications and electronics.