Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 10 (16): 1755-1761, 2010
ISSN 1812-5654

Support Vector Machines for Brain Tumours Cells Classification

[1,2]C.M. Bentaouza and [1]M. Benyettou
[1]Modeling and Optimization of Industrial Systems Laboratory,
Department of Computer Science, Faculty of Science,
University of Science and Technology of Oran, P.O. Box 1505, El Menaour, Oran, Algeria
[2]Department of Computer Science, Faculty of Science and Technology,
University of Mostaganem, B.P. 188, Mostaganem, Algeria

Abstract: This research is a study applied to the supervised classification of brain tumours by a method resulting from the artificial intelligence which is the Support Vector Machines. The artificial intelligence quickly moved these last decades, with the evolution of the cerebral imagery to diagnose certain diseases such as the brain tumours by techniques like magnetic resonance imagery in order to treat this disease by the surgery and microscopy to detect the type and the rank of the tumour. The results obtained by the Support Vector Machines are satisfactory from the point of view of time of learning and convergence, which have in particular tendency to learn data too much, thus providing good performances in generalization. On the other hand the Support Vector Machines give automatically a reliable result.

Key words: Medical image, microscopy, multiclass, one versus rest, supervised learning

INTRODUCTION

Since, the beginning of the years 1990, various techniques of cerebral imagery revolutionized this search while making it possible to see the brain. If these techniques show us what occurs in the brain during a task without having to open the brain-pan, it is especially thanks to progress of computing. So, with the progress of the computing, the artificial intelligence (Cornuéjols *et al*., 2002; Bishop, 2006) experienced a great development. It forms a whole of techniques of representation and decision (Mitchell, 1997; De Beauville and Kettaf, 2005) making it possible the machines to simulate a behavior similar to that operated by the human one. The medical imagery (Ding and Wilkins, 2006) became an essential tool for the decision-making aid (Sharif and Amira, 2009). From where, need for classification (Alphonso, 2001) as process which assigns an entity with a class starting from a whole of measurements, by the use of new mathematical methods for the treatment and interpretation of data (Huang and Laia, 2010). The sector of the algorithms of training (Cristianini and Shawe-Taylor, 2000) remains a very active field of research; for that reason in this study, the authors propose the Support Vector Machines or Separators with Vast Margin (SVM) (Abe, 2005; Wang, 2005), exactly, the multiclass: One Versus Rest (OVR) for their performances in the supervised learning and the classification which are applied to the brain tumours cellules (Lukas *et al*., 2004), diagnosed by Magnetic Resonance Imagery (MRI) and on sure analysis under the microscope of a sample, to identify the type and the rank of the tumour. The study developed a machine learning capable to classify the type of the brain tumour by the use of SVM and presented the strong points of the method. Precisely, we have to show success of the SVM applied to brain tumour cells.

SUPPORT VECTOR MACHINES

The support vector machine is a new method for classification of data. This technique of supervised classification suggested in 1992 by Vladimir VAPNIK (Vapnik, 1998) results from the statistical theory of the training. Now, it is an active zone of search for machine learning. The fundamental idea is to build an optimal learning machine for better generalization, having:

- A definite surface in an optimal way
- In general non linear in the space of entry
- Linear in a dimensional space more raised
- Implicitly defined by a kernel function

Binary SVM: The principal idea of the binary SVM is to implicitly trace data with a dimensional space more raised by the intermediary of a kernel function and to solve a problem of optimization to identify the hyperplane whose

Corresponding Author: C.M. Bentaouza, Modeling and Optimization of Industrial Systems Laboratory, Department of Computer Science, Faculty of Science, University of Science and Technology of Oran, P.O. Box 1505, El Menaour, Oran, Algeria

margin is maximum, which separates the examples from training (Vapnik, 1998). The hyperplane is based on the support vector i.e., a whole of examples of training being on the margin. New examples are classified according to the side of the hyperplane. The problem of optimization generally is formulating in a manner and takes account of the inseparable data while penalizing distort them classifications.

Separable linear case: The solution of the problem of optimization (Burges, 1998) is as follows:

- **Result of the stage of training:** The optimal hyperplane is:

$$w.x+b = 0 \quad (1)$$

$$w = \sum_{i=1}^{n} \alpha_i x_i y_i \quad (2)$$

$$b = \frac{1}{2n}\sum_{i=1}^{n}(w.x_i^{+} + w.x_j^{-}) \quad (3)$$

Correspond to the closest points, i.e., the support vectors that are no null. They are found by solving the following equation:

$$\begin{cases} \max_{\alpha_i} W(\alpha_i) = -\sum_{i=1}^{n}\alpha_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i.x_j) \\ \forall i\, \alpha_i \geq 0 \\ \sum_{i=1}^{n}\alpha_i y_i = 0 \end{cases} \quad (4)$$

- **Result of the stage of test:** The function of decision is obtained by:

$$g(x) = \text{sgn}(\sum_{i=1}^{n}\alpha_i y_i (x_i.x) + b) \quad (5)$$

Non separable linear case: The algorithm established for the separable linear case cannot be used in many real problems. In general, of the noisy data will make separation linear impossible. For this there, the approach above can be extended to find a hyperplane that minimizes the error on the training example (Training error). This approach also indicated under the name of soft margin hyperplane (Vapnik, 1995). We introduce the slack variable ξ_i the error of the example x_i with ξ_i>0, I = 1,..., n. The value of ξ_i indicates, where, are x_i before a hyperplane of separation:

- If $\xi_i \geq 1$ then x_i is badly classified
- If $0<\xi_i<1$ then x_i is correctly classified, but is inside margin
- If $\xi_i = 0$ then x_i is correctly classified and is outside the margin or with the band of the margin

Alternatively, the expression of the function changes from

$$\frac{1}{2}\|w\|^2 \text{ to } \frac{1}{2}\|w\|^2 + C\sum_{i=0}^{n}\xi_i^{\,k}$$

where, C is the weight of error, it is a parameter to be chosen by the user, larger C corresponding to assign a higher penalty with the errors, is generally taken 1/n.

Non linear case: How the methods above can be generalized if the function of decision is not a linear function of the data?, in the linear case space used is Euclidean space. Separation by hyperplanes is a relatively poor framework of classification, bus for much of case; the examples which one wants to separate are not separable by hyperplanes. To continue to profit from the theory which justifies the search for hyperplanes of maximum margin while using more complex surfaces of separation (nonlinear), VAPNIK (Boser *et al*., 1992) suggests separating not items x_i, but of representatives this points Xi = $\varphi(x_i)$ in a space F of dimension much larger, than VAPNIK calls: feature space. One gives oneself a nonlinear function noted φ, such as φ: $R^m \rightarrow F$. The separation nonlinear of maximum margin consists in separating the points $X_i = \varphi(x_i)$, as the function of decision is expressed exclusively using scalar products is thus, one in practice never uses the function directly, but only the kernel K is defined by: K(x, x') = φ(x).(x') it is a nonlinear scalar product. Therefore the problem is solved; it is enough to make a transformation: $x_i.x_j \rightarrow \varphi(x_j).\varphi(x_j)$ in all the formulas of hard margin by using a type of following kernel (Cornuejols, 2002) function:

- Polynomial function:

$$k(x,y) = (x.y+1)^d \quad (6)$$

- Radial function:

$$k(x,y) = \exp(\frac{(x\ \ y)^2}{\sigma^2}) \quad (7)$$

- Sigmoid function:

$$k(x,y) = \tan(s.xy+c) \quad (8)$$

The solution of Problem of optimization in the non-linear case is:

- **Result of the stage of training:** The optimal hyperplane is:

$$w.\varphi(x)+b=0 \tag{9}$$

$$w=\sum_{i=1}^{n}\alpha_i\varphi(x_i)y_i \tag{10}$$

$$b=\frac{1}{2n}\sum_{i=1}^{n}(w.\varphi(x_i^+)+w.\varphi(x_j^-)) \tag{11}$$

α_i is found by solving the following equation:

$$\begin{cases} \max_{\alpha_i} W(\alpha_i)=-\sum_{i=1}^{n}\alpha_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j(\varphi(x_i).\varphi(x_j)) \\ \forall i\, 0\le\alpha_i\le C \\ \sum_{i=1}^{n}\alpha_i y_i=0 \end{cases} \tag{12}$$

- **Result of the stage of test:** The function of decision is obtained by:

$$g(x)=\text{sgn}(\sum_{i=1}^{n}\alpha_i y_i(\varphi(x_i).\varphi(x))+b) \tag{13}$$

Multiclass SVM: There are several methods for the multiclass (Platt, 1999) but in this article, we tried out One Versus Rest (OVR) which is implemented in multiclass SVM. OVR is conceptually the method of simplest SVM multiclass. Here we build of the binary classifiers of K SVM: classify 1 positive against all other negative classes, classifies 2 against all other classes..., class K against all other classes. The combined function of decision of OVR chooses the class of a sample, which corresponds to the maximum value of the binary functions of decision of K indicated by the other positive hyperplane (Statnikov *et al*., 2005). While thus making, hyperplanes of decision calculated by K SVM, which call into question the optimality of multiclass classification.

Implementation: The data base is extracted from classification of OMS (Brucher *et al*., 1999); the ranks are like degrees of malignity, indicating the biological behavior. The following diagram summarizes the essential characters of the tumours gathered according to their rank. In the atlas used, the various tumour entities are also grouped according to their rank, which corresponds to the images taken to make our classification by the two artificial methods. The machine used to carry out these tests is a PC (Pentium 4, 2.80 GHz with 256 MO of RAM). And concerning the programming three essential tools were used:

- C++ Builder 6 under Windows to create a graphic interface summer developed on the one hand for the pretreatment of the classification of the medical images, which consists in extracting significant information in RGB (Red, Green and Blue), which will make the object of the supervised training and the test, so, the input of the binary SVM and Multi-classes SVM
- Binary SVM under Linux with SVM Light (Joachims, 1999) for classification binary classification
- SVM Multiclass (OVR) under Linux is a free tool (free) (Joachims, 1999) which was developed for the scientific research whose problem of optimization was included

EXPERIMENTATION

Considering for all this part, following points:

- **CPU:** The time of training measured in seconds
- **Tr error:** The error on the training set
- **Te error:** The error on the test set
- The classification rate is the average of the well classified of each label

Case 1: binary classification (detection): We began with a binary classification, which consists in detecting the infected part (tumour) of the not infected part (healthy) as shown in Table 1. This data was a cerebral IRM, posterior fossa, whose type of tumour is Ependymoma (Brucher *et al*., 1999):

Figure 1-3 show the variation of the error rate of training and test according to the kernel chosen and to the parameter of regularization.

We note that in the Fig. 1, the error on training set and the error on test set will decrease with increasing regularization parameter to stabilize 1E-04.

We note that in Fig. 2, the error on training set and the error on test set are null despite the change of the regularization parameter.

We notice in the fig that the error on training set and the error on test set are stable and high compared to Fig. 1 and 2.

Table 1: No. of images for each type of tumour in case 1

Label	Class	No. of images
1	Tumour	3
2	Healthy	5
Total		8

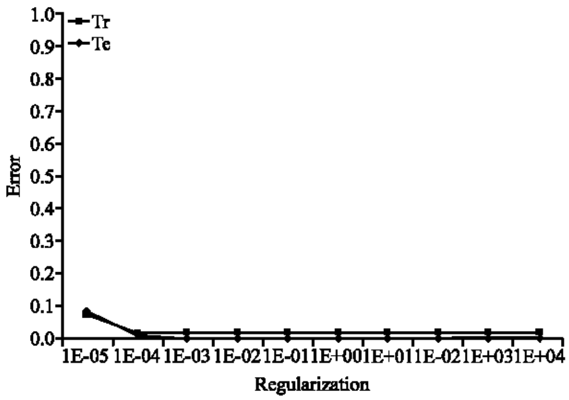

Fig. 1: Error rate of the polynomial kernel in case 1

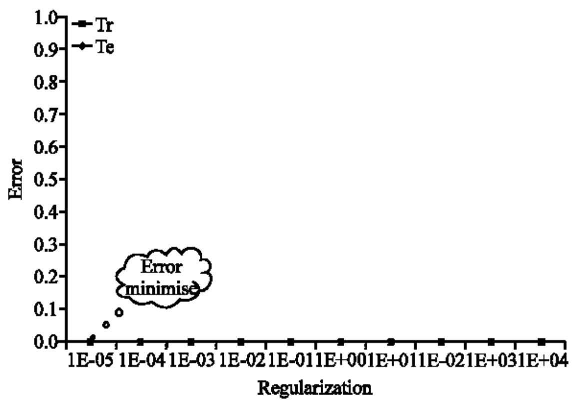

Fig. 2: Error rate of the radial kernel in case 1

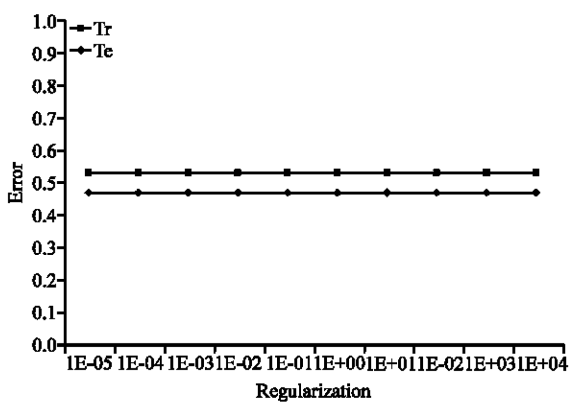

Fig. 3: Error rate of the tangential kernel in case 1

After several experiments and according to the comparison made between different Kernels, of which the goal is to minimize the error. We deduce that in the case of detection, the radial function as defined in Table 2 makes it possible to minimize the error on training set and the error on test set with any badly classified element in less time (i.e., 0,1 sec).

Table 2: Comparison between Kernels in case 1

Kernel	Regularization	Tr	Te	CPU
Polynomial	R≥1E-03	0.0	0.01891	1.0
Radial	R≥1E-05	0.0	0.0	0.1
Tangentiel	R≥1E-05	0.4673	0.5283	16.21

Table 3: Rate of classification per class in case 1

Label	Well classified (%)	Badly classified (%)
1	100.00	0.00
2	85.72	14.28
Classification rate	92.86	

Table 4: No. of images for each type of tumour in case 2

Label	Pathology	Rank	No. of images
1	Ependymoma	I	3
2	Astrocytoma	II	8
3	Glioblastoma	IV	16
4	Medulloblastoma	IV	8
Total			35

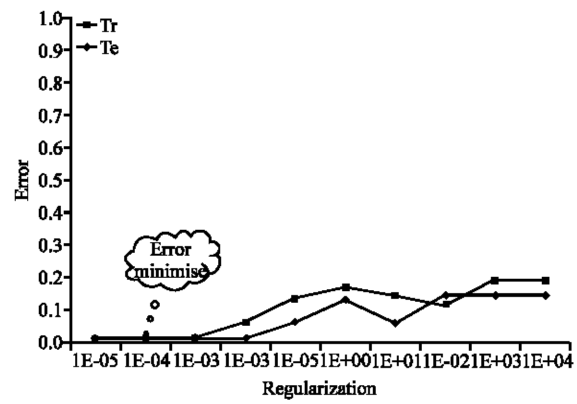

Fig. 4: Error rate of the polynomial kernel in case 2

One passes to the validation results by radial Kernel and regularization parameter higher than 1E-05 which gave a satisfactory rate of learning (Table 2). Table 3 represents the rate of classification of the examples well classified and badly classified for each class.

In this case, we note that the classification rate is good even all the examples of the validation set of label 1 have been well learned.

Case 2: classification with 4 classes: In this case, we will use four various types of classes of frequent tumour cells in the child as can be seen from Table 4 to explode the number of classes and test the performance of SVM on machine learning. This table represents the affected number of images for each class, but also the rank of the tumour (Brucher *et al*., 1999).

We also present in this case, three graphs as illustrated that corresponding to the error of training and test according to the parameter of regularization and the selected type of kernel.

We note that in Fig. 4, the error on training set and the error on test set are minimized when the regularization

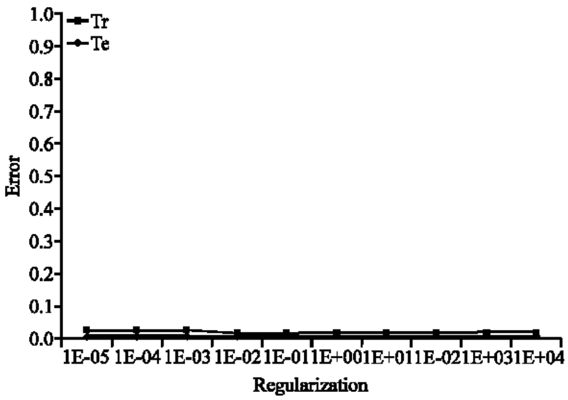

Fig. 5: Error rate of the radial kernel in case 2

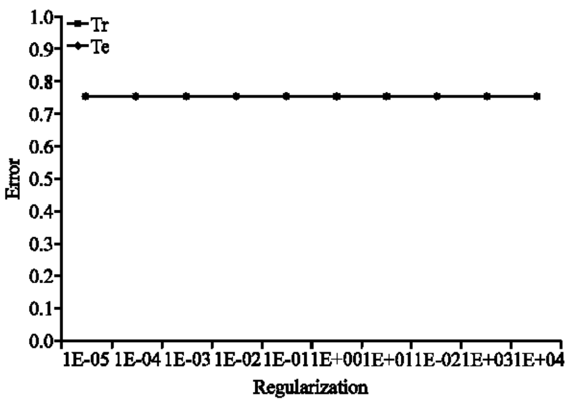

Fig. 6: Error rate of the tangential kernel in case 2

parameter is less than or equal to 1E-03. Then the value of the error begins to increase until stabilization.

We can say that in Fig. 5, the error on training set and the error on test set are minimized and stable after a regularization parameter greater than or equal to 1E-02.

In the case of Fig. 6, with a tangential function, the error on training set and the error on test set are stable and equal to 0,75 despite the change of regularization parameter.

According to the tests carried out for each Kernel (Polynomial, radial and tangential), we retain the Polynomial function as shown in Table 5 which minimizes the error on the basis of test and the base of training in less time.

One validates presented results with the parameters obtained by the tests made before (Kernel Polynomial) and one obtains a table which represents the rate of classification of the examples well classified and badly classified in each class depicted in Table 6.

Case 3: classification with 9 classes: In this case, we will burst our experimentation in nine classes of cerebral tumour in adults as shown in Table 7. This part was added in this study due to scientific curiosity to confirm the ability of SVM in machine learning. The following table represents the affected number of images for each class as well as the rank of the tumour (Brucher *et al*., 1999).

Table 5: Comparison between Kernels in case 2

Kernel	Regularization	Tr	Te	CPU
Polynomial	1E-05≤R≤1E-03	0.0091	0.0093	1.13
Radial	R≥1E-02	0.0091	0.0185	2.44
Tangentiel	R≥1E-05	0.75	0.75	190.47

Table 6: Rate of classification per class in case 2

Label	Well classified (%)	Badly classified (%)
1	96.43	3.57
2	96.43	3.57
3	96.43	3.57
4	100.00	0.00
Classification rate	97.32	

Table 7: No. of images for each type of tumour in case 3

Label	Pathology	Rank	No. of images
1	Pilocytic Astrocytoma	I	10
2	Astrocytoma Cells	I	2
3	Ependymoma	I	3
4	Astrocytoma	II	8
5	Anaplasic Astrocytoma	III	2
6	Anaplasic Ependymoma	III	1
7	Anaplasic Astrocytoma	III	1
8	Glioblastoma	IV	16
9	Medulloblastoma	IV	8
Total			51

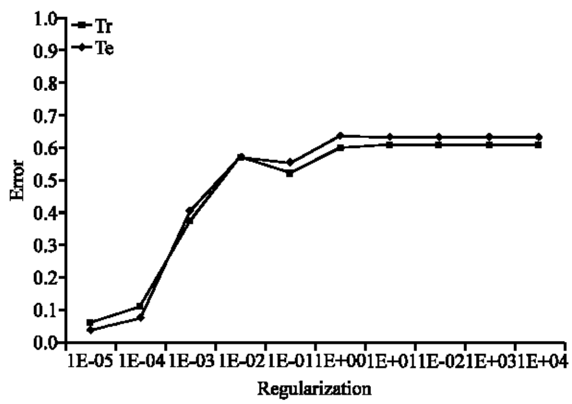

Fig. 7: Error rate of the polynomial kernel in case 3

There we take again the same type of experimentation on this case, by changing the type of the kernel, as well as the parameter of regularization, we obtain the following Fig. 7-9.

In Fig. 7, the error on training set and the error on test set are minimized when the regularization parameter is equal to 1E-05 and it starts to increase with increasing regularization parameter to stabilize after a regularization parameter equal to 1E+01.

We note that in Fig. 8 that the error on training set and the error on test set are stable even with the change

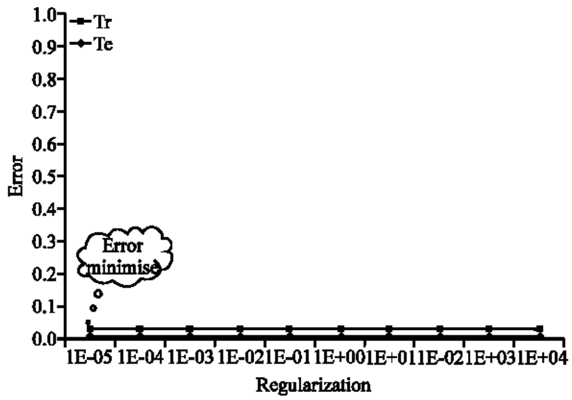

Fig. 8: Error rate of the radial kernel in case 3

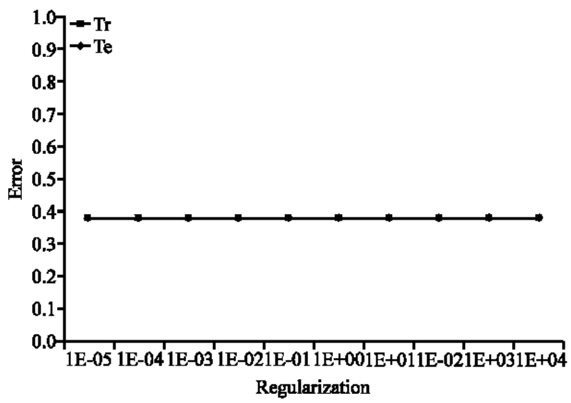

Fig. 9: Error rate of the tangential kernel in case 3

Table 8: Comparison between Kernels in case 3

Kernel	Regularization	Tr	Te	CPU
Polynomial	R = 1E-05	0.0242	0.037	678.89
Radial	R≥1E-05	0.0121	0.0329	9.19
Tangentiel	R≥1E-05	0.8889	0.8889	7.25

Table 9: Rate of classification per class in case 3

Label	Well classified (%)	Badly classified (%)
1	100.00	0.00
2	100.00	0.00
3	100.00	0.00
4	100.00	0.00
5	95.24	4.76
6	100.00	0.00
7	95.24	4.76
8	95.24	4.76
9	100.00	0.00
Classification rate	98.41	

of the regularization parameter and lower values obtained in the case of using a polynomial function (Fig. 7).

The Fig. 9 does not differ from other figures in the use of a tangential function in case 1 and case 2, who does not minimize the error on training set and the error on test set.

Us observe in this case, that it is not any more the Polynomial function which minimized the error in the case 2, on the other hand it is the same function which minimized the error in case 1, so, the radial function as defined in Table 8.

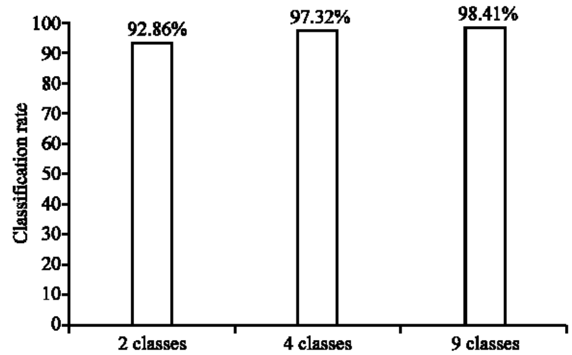

Fig. 10: Comparison of the rates of classification in the three cases

Table 10: Parameters which minimize the error in the three cases

	Kernel	Regularization	Tr	Te	CPU
Case 1 (2 classes)	Radial	R≥1E-05	0.0	0.0	0.1
Case 2 (4 classes)	Polynomial	1E-05≤R≤1E-03	0.0091	0.0093	1.13
Case 3 (9 classes)	Radial	R≥1E-05	0.0121	0.0329	9.19

We introduce the examples of validation and we obtain a rate of classification as shown in Table 9 higher than that found in the case of four classes and that of two classes.

DISCUSSION

One recapitulated the parameters which minimize the error in three cases of the SVM as illustrated in Table 10. We notice that:

- The kernel which minimizes the error changes in three cases between radial and polynomial, thus we cannot define it exactly
- The parameter of regularization to be chosen as the classification of the tumour cells, has to be amounts to 10^{-5} whether it is for the kernel radial or polynomial
- The time of learning is satisfactory

As a consequence, we do say that the SVM has a capacity to learn, in month of time with a convergence guaranteed according to our experiment. The following picture presents the classification rate of the tree cases.

By comparing the rate of classification in three cases by SVM as illustrated in Fig. 10, we discover that the SVM is not really influenced by the number of classes and more we increase the number of class, more we have a better rate of classification.

CONCLUSION

The aim of this research ensues from the need to compare the performances of the various kernels of SVM in classification. In particular, the necessity of being able to choose in an automatic way the parameters of the kernel of SVM. The results noticed in the experimentation, are all obtained by alternation between several parameters for SVM. This last, is one of the models of classification to have marked the forms recognition by supplying a statistical theoretical frame and either connexionnist in the learning theory. View Point of learning time and guarantee of convergence, we can say that SVM is the method the most adapted to the classification of the brain tumours. We showed examples of the success of the SVM and suggest some tracks to go farther: voluminous data base and comparison with others methods as Neural Networks or hybridization with Neuro-SVM (Mitra *et al*., 2005).

REFERENCES

Abe, S., 2005. Support Vector Machines for Pattern Classification. Springer-Verlag, London.

Alphonso, I., 2001. Support Vector Machines for Vowel Classification. Neural Computing, USA.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. 1st Edn., Springer, Heidelberg, ISBN: 0-387-31073-8.

Boser, B.E., I.M. Guyon and V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual ACM Workshop, COLT, (WCOLT'92), ACM, New York, USA., pp: 144-152.

Brucher, J.M., M. Chatel, F. Darcel and R. Theaux, 1999. Association of french speaking neuro-oncologists (ANOCEF's). Atlas of Neuro-Oncologie. http://www.anocef.org/atlas/index.html.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowledge Discovery, 2: 121-167.

Cornuejols, A., 2002. Une nouvelle methode d'apprentissage: Les SVM, separateurs a vaste marge. Bull. L'AFIA, 51: 14-23.

Cornuéjols, A., L. Miclet and Y. Kodratoff, 2002. Apprentissage Artificiel: Méthodes et Algorithmes. 1st Edn., Eyrolles, USA., pp: 277-309.

Cristianini, N. and J. Shawe-Taylor, 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. 1st Edn., Cambridge University Press, Cambridge.

De Beauville, J.A. and F.Z. Kettaf, 2005. Bases Théoriques pour l'Apprentissage et la Décision en Reconnaissances des Formes. 1st Edn., Cépaduès, USA., pp: 233-237.

Ding, Y. and D. Wilkins, 2006. Improving the performance of SVM-RFE to select genes in microarray data. Proceedings from the 3rd Annual Conference of the Midsouth Computational Biology and Bioinformatics Society, Sept. 26, The University of Mississippi, University, MS, USA., pp: S12-S12.

Huang, P.W. and Y.H. Laia, 2010. Effective segmentation and classification for HCC biopsy images. Pattern Recognition, 43: 1550-1563.

Joachims, T., 1999. Making Large-Scale SVM Learning Practical. In: Advances in Kernel Methods-Support Vector Learning, Scholkopf, B., C.J.C. Burges and A.J. Smola,(Eds.). MIT Press, Cambridge, MA., ISBN-10: 0-262-19416-3, pp: 169-184.

Lukas, L., A. Devos, J. Suykens, L. Vanhamme and F. Howe *et al*., 2004. Brain tumour classification based on long echo time proton MRS signals. Artificial Intel. Med., 31: 73-89.

Mitchell, T.M., 1997. Machine Learning. 1st Edn., McGraw-Hill, New York, ISBN: 978-0070428072.

Mitra, V., C. Wang and S. Banerjee, 2005. A neuro-SVM model for text classification using latent semantic indexing. Proceedings of the International Joint Conference on Neural Networks, (IJCNN'05), Canada. Technical Communications Ltd., pp: 564-569.

Platt, J., 1999. Fast Training of Support Vector Machines Using Sequential Minimal Optimization Advances in Kernel Methods, Support Vector Learning. 1st Edn., MIT Press, Cambridge, ISBN: 0-262-19416-3.

Sharif, M.S. and A. Amira, 2009. An Intelligent System for Pet Tumour Detection and Quantification. Processing of IEEE International Conference on Image, Nov. 7, Cairo, Egypt, pp: 2625-2628.

Statnikov, A., F. Aliferis, I. Tsamardinos, D. Hardin and S. Levy, 2005. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics, 21: 631-643.

Vapnik, V.N., 1998. Statistical Learning Theory. Wiley, New York.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Wang, L., 2005. Support Vector Machines: Theory and Application. 1st Edn., Springer, New York, ISBN-10: 3540243887.