How to Cite:
Singh, R., & Agarwal, B. B. (2022). Abnormality detection and classification from brain MRI using machine learning . *International Journal of Health Sciences*, *6*(S3), 9170–9180. https://doi.org/10.53730/ijhs.v6nS3.8242

Abnormality detection and classification from brain MRI using machine learning

Ravendra Singh
Department of Computer Science and Engineering, IFTM University, Moradabad, India
*Corresponding author email: ravendra85@gmail.com

Bharat Bhushan Agarwal
Department of Computer Science and Engineering, IFTM University, Moradabad, India
Email: bharatagarwal9@gmail.com

***Abstract*---**In our proposed method, wavelet transform is inked first to excerpt features from MR image, and then to lessen the dimensions of features the principle component analysis has been applied. The lessen features have been given as input to a Support Vector Machine (SVM) that uses kernel function for transforming the lower dimension feature space to upper dimension feature space. An approach of K-fold-stratified-cross validation has been taken care of to improve the generalization of kernel support vector machines. The proposed method has used different kernels and observed that the Gaussian Radial Basis function represents the accuracy result as 90.3% which is the highest.

***Keywords*---**principle component analysis, discrete wavelet transform, magnetic resonance image, gaussian radial basis function, support vector machine.

Introduction

A very popular imaging technique known as Magnetic Resonance Imaging plays a vital role in Clinical Diagnosis as well as Biomedical Researches. It is very useful in generating high-quality images that can act as an input for diverse medical processes mainly concerning with the central processing unit of the human body i.e. the brain [1-5]. During the process of MRI, various results are generated giving values corresponding to classifications performed with the help of MRI machine. The values can be enhanced by using the automated processes as well as diverse classification algorithms [6-8]. The Feature extraction can be done with

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 9 April 2022, Accepted for publication: 18 May 2022

the help of a very effective tool known as Wavelet Transform. It permits the user to analyze the images at a hierarchical level of resolutions as the wavelet transform has the resolution analytic property. This Transform has a prerequisite of the enormous capacity of storage as well as it is expensive in terms of computation [9]. For performing the dimensionality reduction and enhancing the discriminative power, a very powerful technique i.e. Principal Component Analysis (PCA) can be used [10]. With the above requirements of decreasing the cost (computation) as well as reducing the dimensions, this technique of PCA can be used [11]. This leads to a problem concern of classifying input data (raw facts and figures). From the last decade, many researchers have given their approaches for the solution required for the above problem that were broadly classified into two groups of classification. These belongs to Supervised and unsupervised form of classification. Under the Supervised form of classification, the techniques that are responsible for the solutions are Support Vector Machines [14], and K-Nearest Neighbour [7]. While the latter (unsupervised) [14] deals with fuzzy c means and Self-Organization Feature Maps [15].

These approaches are performing well but the need is to find a more appropriate model as when two major methods are compared that are supervised and unsupervised the former (supervised) is the winner with the limitation of 95%. The parameter that is considered here is Success Classification Rate (SCR). In terms of Supervised Learning, the support vector machines depict up to the mark classification scenarios that are typically based on a machine learning paradigm [16-18]. Among the diverse supervised techniques of ML such as Artificial Neural Network along with Bayesian Networks, Decision Tree, and SVM, SVM performs better in terms of accuracy and geometric interpretation. One key advantage of SVM is that it does not require numerous training samples as depicted in [14]. The primary phase of SVM dealt with a linear aspect of classifiers. Here, in this paper, we introduced the extension of the Support Vector Machine by using kernels. Here, we have compared the three major kernels that are Homogeneous Polynomial, Inhomogeneous Polynomial, GRB Kernel Function. Non-Linearity was introduced to exchange the dot product form in the actual linear support vector machine [14]. The kernelized version of SVM allows the user to adjust the maximum margin hyperplane. It could be non-linear and the transformed space that is generated can be of high dimension.

Materials and Methods

Our proposed approach has the following three stages:

- Step-1. Pre-processing (image segmentation, feature-extraction, feature-reduction).
- Step-2. Kernel SVM is trained.
- Step-3. As input, a new MRI is given to the trained kernel SVM, and the result is predicted by the classifier. In Figure 1, the flowchart shows a standard classification procedure.

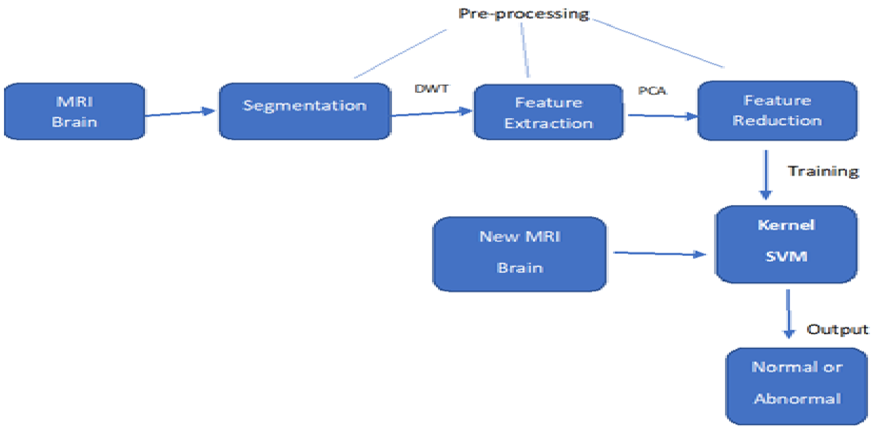

Figure 1. Proposed Method

Image Segmentation

In medical image processing, segmentation has a vital role. The segmentation is done by K-means. It is an unsupervised approach, in this unsupervised technique the outcome is not known in advance and it implements on the given data points by initialing labels. In the k-means method initially, k centroid points are given that are k=3 or k=5 and so on. The Euclidean distance has been used to compare the other data points and their clusters are formed that are closer to others. The mean of the clustered data is computed and the new centroid is discovered. The formation of new clusters in the curiosity of the new centroid and continue this procedure until no new centroids are formed.

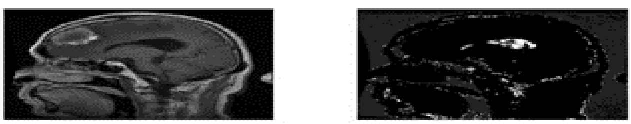

Figure 2. Segmented Image

Feature Extraction

For Analysis of Signals, a very popular method of Fourier Transform is used. It divides a broader domain of time into smaller modules of sinusoids that correspond to diverse frequencies. Fourier Transform has limitations in that it cannot reflect the abrupt changes in images. In the case of FT, the classification quality decreases as there is a loss in time information. In Gabor Fourier transform a small part of the signal is used to analyze in one time. This method is called short-time Fourier transform [22]. STFT insert a new window of a specific size to signal. It is an interface between frequency domain and time domain. STFT gives valuable information about frequency and time domain. Information precision is limited with the size of the window. Wavelet- transform presents a windowing approach of variable size. WT keeps both types of information regarding signal i.e. frequency and time. The process of the signal analysis is depicted in Figure 3.

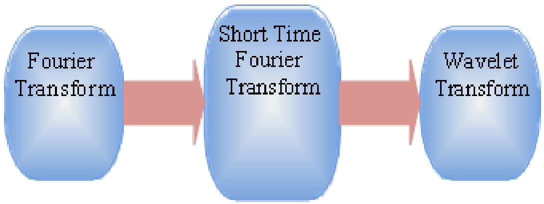

Figure 3. Process of the signal analysis.

One more advantage of Wavelet Transform (WT) is "scale" in place of traditional frequency, WT is not generating a frequency-time view since it produces a time scale view to see data, which is a more powerful and natural way.

Discrete Wavelet Transform

The DWT is a tool for denoising and compression images. It is also a powerful tool for the program of Wavelet transform [24]. The basics of discrete wavelet transform are described as supposing that x(t1) is an sqr-integrable function, then the continuous wavelet transform of x(t1) w.r.t to a given wavelet ψ(t1) is defined as

W ψ (a,b) = ψa,b(t1) dt ... (1)
where
ψa,b(t1)= ....(2)

where, the wavelet ψa,b(t1) is computed from the base wavelet ψ(t1) by dilation and translation: a is a dilation factor and b the translation parameter (both the values are positive and real numbers). In terms of wavelet analysis, diverse type of wavelets are present that has enhanced the concept as well as numerical interpretations of wavelet analysis. Harr Wavelet is the most popular of all types and can perform well in a variety of applications [25-27]. The decomposition operation that is depicted above represents iterations along with successive approximations and as a result here a bigger module of signals is broken off into many levels of resolutions. The entire procedure of the Decomposition tree of a Wavelet is shown in Figure 4.

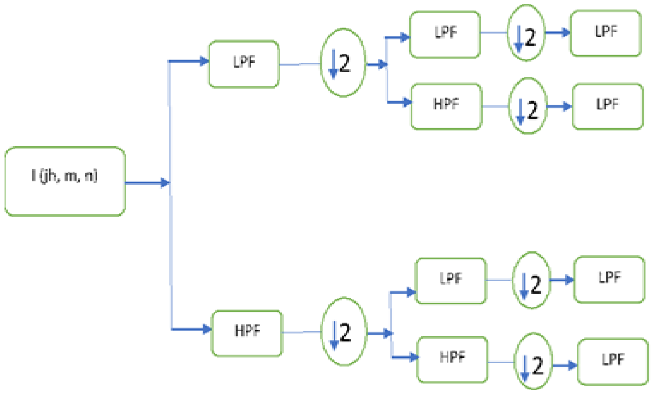

Figure 4. Level 3 Decomposition tree of a Wavelet.

In concern to a two-dimensional image, each dimension is applied with a Discrete Wavelet Transform. The systematic block diagram of 2-D DWT is shown below in Fig 5 for easy understanding. In the figure mentioned above, there are four sub-bands derived as a result of a denoising process. They are represented by the terms ILL, ILH, IHL, IHH. The (ILL) sub-band has been used for next 2Dimension Discrete Wavelet Transform (DWT). The (ILL) subband is represented as an approximation component of an image, the other subbands like (ILH) (IHL), and (IHH) represent the detailed components of the image. An approximation component was achieved as the level of decomposition growth and wavelets give a normal hierarchical framework for translating the image. In the proposed method, level-three decomposition is achieved to extract features by Harr wavelet. One technical issue occurs in the digital filter of Discrete wavelet transform (DWT) which is known as border distortion. When the filtering operation is applied to the image, the window of the filter will be stretched beyond the image at the edges, as the result is adding of the pixels beyond image. In the proposed method, the symmetric padding procedure [18] has been used to compute the boundary values.

Feature Reduction

Storage memory and CPU time increase due to large features and excessive features make the classification complicated, that is the curse of dimensionality. Now, this is the best time to reduce the no. of features. PCA is used here as a method to diminish the dimensions of a data set which consists of a huge no. of related variables dealing with variations. The reduction in features is a transformation phenomenon in which the set of data into ordered sequences according to their variances. PCA technique has the following impacts: The input vector components have been orthogonalized by PCA that are not correlated to each other, The resulting components have been ordered by PCA so that the large diversion comes first and it removes those which contribute less to variation in the data set. The input vector has been normalized to zero mean and unity variance before applying PCA.

Kernel Support Vector Machine

SVM plays a vital role in ML. Support vector machine provides high accuracy, mathematical tractability, and geometric interpretation [24]. SVM improved with the kernel in which the kernel Support vector machine is prominent and effective. The advantages of kernel SVMs are as follows: (1) noteworthy in the field of bioinformatics, natural language processing(NLP), and computer vision. (2) training involves convex quadratic optimization [29].

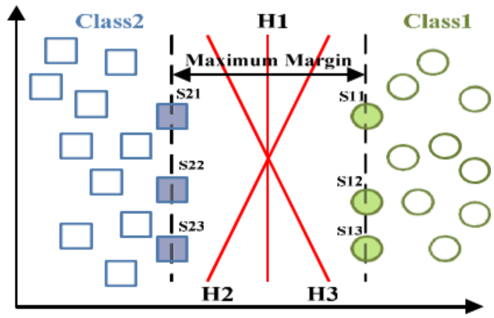

Figure 5. The geometric- interpolation of linear SVM

Conventional SVM draw a hyperplane to categorised data, so SVM can not be applied to categorise the problems since the different tpes of data present at different sides of hyperplane, to resolve a problem the kernel technique is introduced with a Support vector machine(SVM) [18]. In proposed method, nonkernel function is used in place of every dot. This kernel approach corresponds to the transformation φ (x_p) with the equation $k(x_p, x_q) = \varphi(x_p) \ \varphi(x_q)$.The value υ is also in the transformed space, with $\upsilon = \sum_i \alpha_i \gamma_i \varphi(xp)$. Dot products with υ can be computed for classification by v. φ (x)= $\sum_i \alpha_i \gamma_i k(xp, x)$. From another point of view, the kernel SVM(KSVM) permits fixing the large margin hyper-plane in a feature space. This transformation could be nonlinear and has transformed into high dimension even though the classifier is a hyperplane in higher dimension space-feature, it could be nonlinear in the orginal input space. In our proposed method, three kernels are used and their performance has been compared. For each kernel, one adjusting parameter is used to make the kernel tailor and flexible to practical data.

For given data, a classifier is trained so the classifier shows high accuracy for training dataset only, not for other independent datasets. To overcome the problem of overfitting, we have incorporated cross-validation in our proposed method. Since cross validation is integrated in our proposed method that makes the classifier more authentic and could be expanded to other independent datasets. This method can be categorised into three types: Random sub-sampling, leave-one-out validation and K-fold cross validation. In our proposed method, K-fold cross valdation is implemented since it is easy,simple and used for all data for validation and training. The mechanics is as follows: First, construct a k-fold division of the whole dataset, for training we have repeated k times to apply K1 folds, and for validation a left fold is used. Finally, we take the average of error-rates of K experiments. These folds are completely randomly partitioned, for few folds there might be a rather different distribution from other cross-folds. So, we have used stratified K-fold cross-validation, in which every cross fold has approximately same class distribution [29]. To get the no. of folds is another challenge. For choosing a large value of K, the bias value of error rate should be small. This is only possible when the estimator's variance value would be very large and the execution time would be very high. On another hand, if we choose a very small value of k, the computing time would decrease, the estimator of bias would be large and the variance of the estimator would small enough [30]. In our proposed method, The K has been determined as five through the error and trial method, that is, assume that the parameter K changes from three to ten and the step is one, and SVM is trained with each value. Finally, we have selected the best K-value for the highest accuracy of classification.

Result and Discussion

I have downloaded the open SVM toolbox and extended the SVM to kernel-SVM. The dataset contains T2- weighted MR images and axial planes with resolution of 256 x 256. The dataset has been downloaded from website of the Harvard Medical School (www.med.harvard.edu/AANLIB), ADNI dataset(adni.Ionic.ucla.edu) and OASIS dataset(www.oasis-brains.org). We have chosen the T2 weighted model for T2 images have a clear vision and higher contrast as compare to PET and T1 modality. To evaluate an optimal parameters of kernel functions in python, enormous computations were considered here depicting numerous simulation of the d order in the Linear and RBF kernel, and γ is the scaling factor in the GRB kernel. All classification methods were tested on 999 MRI scans of brains with tumors. There are three kinds of tumors (1. meningioma, 2. glioma, 3. pituitary tumors), with 333 images of each type of tumor. The performance details of each classifier are given below.

Table:1
Classification report for GRB classifier

Kind of Tumor	Precision	Recall	f1-score
1	0.89	0.68	0.77
2	0.86	0.90	0.89
3	0.73	1.00	0.85

Gaussian kernel is a general-purpose kernel and is used when there is no prior knowledge about the data. Its confusion matrix is given in Table 2.

Table 2
Confusion matrix for GRB classifier

	Predicted values		
Total values	17	4	4
	2	24	0
	0	0	11

Radial Basis Function (RBF) is a general-purpose kernel. Its performance matrix is shown in Table 3.

Table 3
Classification report for RBF classifier

Kind of Tumor	Precision	Recall	f1-score
1	0.82	0.69	0.75
2	0.84	0.88	0.86
3	0.73	0.90	0 .81

The confusion matrix of Radial Basis Function (RBF) kernel is given in Table 4.

Table 4
Confusion matrix

	Predicted values		
Total	18	4	4
	3	21	0
	1	0	11

We have compared our algorithm with common procedures SVM+DWT with a linear Kernel [12], SVM+DWT with a kernel based on RBF [12].

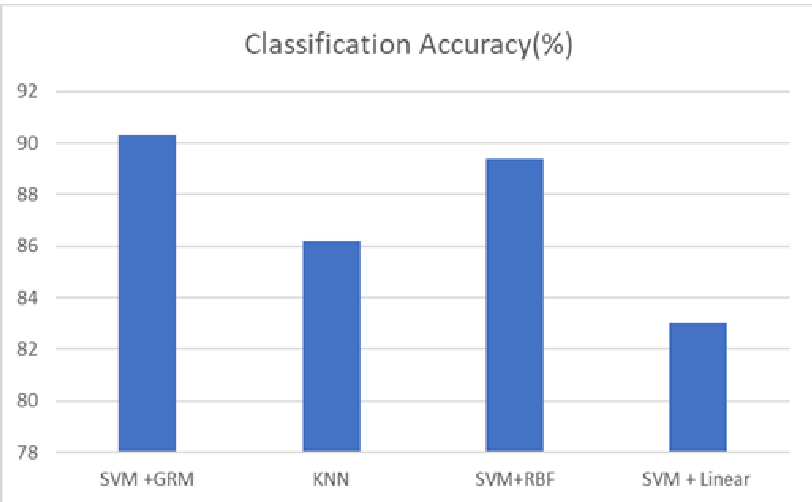

Figure 6. Comparison of Classification Accuracy

The experimental values of the proposed method KSVM+PCA+DWT with GRB kernel provide excellent results on both test images and training. The SVM with GRB-kernel clearly outperformed the two other core SVMs.

Conclusion

The proposed work has implemented a novella approach KSVM+PCA+DWT to categorize abnormal and normal brain MR images. In the proposed method, we have chosen three different kernels as GRB, Linear and RBF. The experimental result reveals that the GRB kernel support vector machine achieved an accuracy of 90.3% on one-sixty MR images which is higher than other methods in the literature in the last decade. Future work is focused on the following parameters: First, Multi-classification, that emphasis on particular disorder viewed using brain MRI, could be searched. Second, novel kernels can be tested for enhancing the accuracy of classification. We could effectively retrieve the information from MR images. In the proposed method, we have used DWT inplace of FT for the spatial resolution, viz.., Discrete Wavelet Transform captures both location and frequency information. Our proposed KSVM+ PCA+DWT+GRB kernel method has shown superiority to the Homogeneous Polynomial and Inhomogeneous Polynomial kernels SVM. An important contribution of this work is to propose a method for recognizing actual MR brain from abnormal brain MRI. Testing of three kernels, and find GRB kernel as the most successful one. Kernel SVM is a valuable tool that can be applied in assisted computer clinical diagnosis.

Acknowledgments

This paper and the research behind it would not have been possible without the exceptional support of my supervisor. His enthusiasm, knowledge, and exacting attention to detail have been an inspiration. I am grateful to all of those with whom I have had the pleasure to work during this paper.

References

1. Badža, M. M., & Barjaktarović, M. Č. (2020). Classification of brain tumors from MRI images using a convolutional neural network. Applied Sciences, 10(6), 1999.
2. Krishnaveni, P. R., & Kishore, G. N. (2020). Image Based Group Classifier for Brain Tumor Detection Using Machine Learning Technique. Traitement du Signal, 37(5).
3. Das, P., & Das, A. (2019, December). Automatic Detection and Classification of Enhanced Brain Tumor Using Machine Learning Algorithm. In International Conference on Computers and Devices for Communication (pp. 36-42). Springer, Singapore.
4. Shree, N. V., & Kumar, T. N. R. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. Brain informatics, 5(1), 23-30.
5. Krishnaveni, P. R., & Kishore, G. N. (2020). Image Based Group Classifier for Brain Tumor Detection Using Machine Learning Technique. Traitement du Signal, 37(5).
6. Tang, J., Gan, Z., & Yang, X. (2019, November). Cardiac Motion Tracking in Short-axis MRI using Siamese Convolution Network. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 865-870). IEEE.
7. Gunalan, K. (2018). Theoretical Predictions of Axonal Pathways Activated by Subthalamic Deep Brain Stimulation. Case Western Reserve University.
8. Kumar, S., Dabas, C., & Godara, S. (2017). Classification of brain MRI tumor images: a hybrid approach. Procedia computer science, 122, 510-517.
9. Zhang, Y., Lu, S., Zhou, X., Yang, M., Wu, L., Liu, B., ... & Wang, S. (2016). Comparison of machine learning methods for stationary wavelet entropy-

based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. Simulation, 92(9), 861-871.
10. Scapaticci, R., Di Donato, L., Catapano, I., & Crocco, L. (2012). A feasibility study on microwave imaging for brain stroke monitoring. Progress In Electromagnetics Research, 40, 305-324.
11. Gomes, T. A., Prudêncio, R. B., Soares, C., Rossi, A. L., & Carvalho, A. (2012). Combining meta-learning and search techniques to select parameters for support vector machines. Neurocomputing, 75(1), 3-13.
12. Iqbal, A., & Jeoti, V. (2012). A novel wavelet-Galerkin method for modeling radio wave propagation in tropospheric ducts. Progress in Electromagnetics Research, 36, 35-52.
13. [13] Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. Neuroimage, 62(2), 911-922.
14. Deris, A. M., Zain, A. M., & Sallehuddin, R. (2011). Overview of support vector machine in modeling machining performances. Procedia Engineering, 24, 308-312.
15. Ala, G., Francomano, E., & Viola, F. (2011). A wavelet operator on the interval in solving Maxwell's equations. Progress In Electromagnetics Research, 27, 133-140.
16. Mohsin, S. A. (2011). Concentration of the specific absorption rate around deep brain stimulation electrodes during MRI. Progress In Electromagnetics Research, 121, 469-484.
17. Chaturvedi, C. M., Singh, V. P., Singh, P., Basu, P., Singaravel, M., Shukla, R. K., ... & Singh, S. P. (2011). 2.45 GHz (CW) microwave irradiation alters circadian organization, spatial memory, DNA structure in the brain cells and blood cell counts of male mice, Mus musculus. Progress In Electromagnetics Research, 29, 23-42.
18. Wang, F. F., & Zhang, Y. R. (2011). The support vector machine for dielectric target detection through a wall. Progress In Electromagnetics Research, 23, 119-128.
19. Zhang, Y., & Wu, L. (2011). Crop classification by forward neural network with adaptive chaotic particle swarm optimization. Sensors, 11(5), 4721-4743.
20. Oikonomou, A., Karanasiou, I. S., & Uzunoglu, N. K. (2010). Phased-array near field radiometry for brain intracranial applications. Progress In Electromagnetics Research, 109, 345-360.
21. Asimakis, N. P., Karanasiou, I. S., Gkonis, P. K., & Uzunoglu, N. K. (2010). Theoretical analysis of a passive acoustic brain monitoring system. Progress In Electromagnetics Research, 23, 165-180.
22. Tagluk, M. E., Akin, M., & Sezgin, N. (2010). Classıfication of sleep apnea by using wavelet transform and artificial neural networks. Expert Systems with Applications, 37(2), 1600-1607.
23. Li, D., Yang, W., & Wang, S. (2010). Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. Computers and electronics in agriculture, 74(2), 274-279.
24. Zhang, Y. D., Wang, S., & Wu, L. (2010). A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO. Progress In Electromagnetics Research, 109, 325-343.
25. Muniz, A. M. S., Liu, H., Lyons, K. E., Pahwa, R., Liu, W., Nobre, F. F., & Nadal, J. (2010). Comparison among probabilistic neural network, support vector machine and logistic regression for evaluating the effect of subthalamic

stimulation in Parkinson disease on ground reaction force during gait. Journal of biomechanics, 43(4), 720-726.
26. El-Dahshan, E. S. A., Hosny, T., & Salem, A. B. M. (2010). Hybrid intelligent techniques for MRI brain images classification. Digital signal processing, 20(2), 433-441.
27. Ghosh, A., Shankar, B. U., & Meher, S. K. (2009). A novel approach to neuro-fuzzy classification. Neural Networks, 22(1), 100-109.
28. Martiskainen, P., Järvinen, M., Skön, J. P., Tiirikainen, J., Kolehmainen, M., & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. Applied animal behaviour science, 119(1-2), 32-38.
29. Zhang, Y. D., & Wu, L. (2008). Weights optimization of neural network via improved BCO approach. Progress In Electromagnetics Research, 83, 185-198.
30. , J. Y., & Fu, J. C. (2008). A hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI. Expert Systems with Applications, 34(2), 1285-1295.
31. Xu, Y., Guo, Y., Xia, L., & Wu, Y. (2008). An support vector regression based nonlinear modeling method for SiC MESFET. Progress in Electromagnetics Research, 2, 103-114.