Journal of Mind and Medical Sciences

Volume 6 | Issue 1 Article 18

2019

Brain image clustering by wavelet energy and CBSSO optimization algorithm

Hasan Hosseinzadeh
Department of Mathematics, Ardebil Branch, Islamic Azad University, Ardebil, Iran, hasan_hz2003@yahoo.com

Mohammad Sedaghat
Ardebil Branch, Islamic Azad University, Ardebil, Iran

Recommended Citation
Hosseinzadeh, Hasan and Sedaghat, Mohammad (2019) "Brain image clustering by wavelet energy and CBSSO optimization algorithm," *Journal of Mind and Medical Sciences*: Vol. 6 : Iss. 1 , Article 18.
DOI: 10.22543/7674.61.P110120
Available at: https://scholar.valpo.edu/jmms/vol6/iss1/18

https://scholar.valpo.edu/jmms/
https://proscholar.org/jmms/
ISSN: 2392-7674

J Mind Med Sci. 2019; 6(1): 110-120
doi: 10.22543/7674.61.P110120

Received for publication: September 24, 2018
Accepted: December 12, 2018

Research article

Brain image clustering by wavelet energy and CBSSO optimization algorithm

Mohammad Sedaghat, Hasan Hosseinzadeh

[1]Ardebil Branch, Islamic Azad University, Ardebil, Iran

Abstract

Previously, the diagnosis of brain abnormality was significantly important in the saving of social and hospital resources. Wavelet energy is known as an effective feature detection which has great efficiency in different utilities. This paper suggests a new method based on wavelet energy to automatically classify magnetic resonance imaging (MRI) brain images into two groups (normal and abnormal), utilizing support vector machine (SVM) classification based on chaotic binary shark smell optimization (CBSSO) to optimize the SVM weights.

The results of the suggested CBSSO-based KSVM are compared favorably to several other methods in terms of better sensitivity and authenticity. The proposed CAD system can additionally be utilized to categorize the images with various pathological conditions, types, and illness modes.

Keywords : brain tumor, MRI, support vector machine, binary shark smell optimization

Highlights

- ✓ This paper presents a novel method to categorize MRI images into normal and abnormal groups by WE, support vector machin,e and CBSSO optimization method.
- ✓ The efficiency of this approach is confirmed through its comparison with several other methods (BP network, kernel SVM and PSO-KSVM).
- ✓ The WE is also authenticated as an efficient characteristic in the categorization of MRI images.

To cite this article: Sedaghat M, Hosseinzadeh H. Brain image clustering by wavelet energy and CBSSO optimization algorithm. *J Mind Med Sci*. 2019; 6(1): 110-120. DOI: 10.22543/7674.61.P110120

*Corresponding author: Hasan Hosseinzadeh, Department of Mathematics, Ardebil Branch, Islamic Azad University, Ardebil, Iran
E-mail: hasan_hz2003@yahoo.com

Introduction

Categorizing brains into normal and abnormal is highly important in medicine (1). Accordingly, MRI has received much attention and has produced an extensive database on each subject's brain situation. Many studies in the literature have been performed using automated detection of abnormal brains through MRI scans (2, 3).

Wavelet transformation (WT) is a highly efficient method for the analysis of images that shows the ability of feature selection in time as well as the frequency domain at the same time (4, 5). By implementing a one-level WT over images, 4 sub-bands are usually introduced, namely LL (low-low), LH (low-high), HL (high-low), and HH (high-high). The low-low sub-band provides an estimation of the image, and the remaining three give high frequency parameters for presenting the finer-scaled image details. WT has been efficiently implemented in many applications of computer-visualization thus far.

Recently, WT and other methods have been implemented for the diagnosis of abnormal brains through images. For instance, (6) the approximation parameters achieved through discrete WT (DWT), the SOM-NN (self-organization map neural network), and SVM have all been used. References (7) suggest that Slantlet transformation is the developed format of discrete WT. In exchange, the property vector of images is generated with regard to Slantlet transformation output levels related to 6 spatial locations, with the selection based on logic. Following that, back propagation NN (BPNN) has been used. In (8), parameters of three-level discrete WT are estimated. Also, this reference decreased the parameters using PCA (principal component analysis) and employed feed-forward BP network (FP-NN) as well as KNN (K-nearest neighbor) classification methods. Moreover, (9) suggested discrete WT as a feature selection method, principal component analysis for feature decrement, and the FNN along with the SCABC (scaled chaotic artificial bee colony) as a classification operator. In accordance with (9), reference (10) proposed a substitute for the SCABC method with the SCG (scaled conjugate gradient) approach. The authors in (11) utilized the expectation maximization Gaussian combined method equipped with fast Fourier transformation for the achievement of tissue categorization through MRI scans. The kernel support vector machine (KSVM) is suggested in (12). Also, 3 novel kernels including homogenous and non-homogenous polynomials as well as Gaussian radial basis are proposed in this reference. The authors in (13) suggested a new WE (wavelet entropy) characteristic, and used SWP (spider web plots) to additionally decrease the number of characteristics. Then, probabilistic NN (PNN) is utilized in this reference. Moreover, (14) suggested categorization precisions, and (15) performed the modeling of detailed parameters for a two-level discrete WT using the static model of GARCH (general autoregressive conditional heteroscedasticity), and its coefficients are regarded as initializing characteristic array. Also, KNN and SVM were selected as classification methods. Moreover, (16) proposed the PSO (particle swarm optimization) to learn the KSVM, whose results obtained 97.11% authenticity over an information set containing 90 images. Authors in (17) utilized synthetic wavelet statistic texture characteristics for the segmentation and categorization of AD benign and malignant tumor slices. Reference (18) utilized feedback-pulse-coupled NN to segment the images, discrete WT for feature exploitation, PCA method to decrease the count of wavelet parameters, and the FBP network to categorize the inputs into two classes (normal and abnormal). Furthermore, DWPT (discrete type of wavelet packet transformation) is proposed in (19), where harnessed Tsallis entropy is used to achieve characteristics from DWPT parameters. Following that, this reference employed GEP-SVM (the generalized Eigen value proximal SVM) alongside the RBF kernel.

All the above approaches have proper performance, but they faced 2 major challenges. First, they often utilized the discrete WT, whose parameters require a great store of memory and, second, their learning about classification operator does not have proper stability, hence, precision can be promoted.

The present work aimed at tackling these 2 challenges. This paper proposes the WE (wavelet energy) that is a newly-introduced feature exploitation method and computes the energy of each wavelet parameter. This method has also been implemented in many academic/industrial applications so far (20-22). Also, the CBSSO algorithm is proposed here for the learning process of classification operator in order to promote the clustering efficiency.

Materials and Methods

Feature extraction

1. Two dimensional discrete WT

The presented 2-D discrete WT is used in the decomposition of images of the many sub-sections based on a recursive procedure (23, 24). One-level decomposition gives 2 types of parameters. The first parameter includes LH1, HL1, and HH1 that provide details on the main images. The second one is the LL1,

which is equivalent to the estimation of the main images (25). Afterwards, the LL1 approximation will be decomposed to second level estimation and will detail the parameters. This procedure is continued until the favorable resolution level is achieved.

The achieved parameters of approximation as well as the detailed sub-sections can be seen as beneficial characteristics of texture classification. The images are decomposed through a two-dimensional discrete WT for spatial frequency elements, which explains the texture of the images. In fact, the parameters will be achieved by entwining the images with a set of filters. Then, the chosen characteristics will be elicited from the parameters for an additional process.

2. The WE method

Subsequent to the WT implementation on the images, the wavelet parameters of the detailed sub-sections of every decomposition level are utilized in naming the local WE characteristic (5). The WE of all four sub-bands are represented thus:

$$E(LL) = \sum_x \sum_y LL(x,y)^2 \qquad (1)$$

$$E(LH) = \sum_x \sum_y LH(x,y)^2 \qquad (2)$$

$$E(HL) = \sum_x \sum_y HL(x,y)^2 \qquad (3)$$

$$E(HH) = \sum_x \sum_y HH(x,y)^2 \qquad (4)$$

In these equations, E is the wavelet energy. The computed energies denote the ability of image details in various sub-bands. By considering a three-level decomposition, the final obtained characteristics will be the energies from the following ten sub-bands:

{LL3, LH3, HL3, HH3, LH2, HL2, HH2, LH1, HL1, HH1}.

Classification operator

The SVM is an efficient method in the field of machine learning. The high precision, the low mathematic complication, and the straight geometric description are the main benefits of this method. In recent years, several developed SVMs have been achieved with great speed. Among these developed methods, the kernel SVMs have the highest popularity and efficiency. Some of the merits of the kernel type methods are:

- The kernel SVMs operate properly over applications and are considerably effective in various fields such as natural-language classification, bioinformatics, and computer-vision.
- These methods have low adjustable variables.
- Learning them mostly includes convex quadratic optimization. In other words, the answers obtained are globally optimal and mostly unique. Therefore, they prevent local optimal trapping obtained using statistical training algorithms such as NNs.

1. Support vector machine

Assume a few predetermined data, each of which belongs to 1 or 2 categories and the aim is to determine in which category a new data point will be placed. We considered a datum in a p-D array form, and we had to generate a (p-1)-D hyper-plane. A large number of feasible hyper-planes could effectively categorize the data point (26). A reasonable option as to which is the best hyper-plane is the option that gives the greatest division (margin) between 2 categorizations, because a better treatment can be anticipated in responding to unforeseen information.

Such learning means better generalization efficiency (27). Hence, the hyper-plane is selected so that the space among that and the closest datum in each band is maximized. The geometric incorporation for linear support vector machines is depicted in Fig. 1, in which, H_i,i=1,2,3 denote the 3 hyper-planes that can categorize 2 categories efficiently. However, the second and the third ones (H2 and H3) do not have greater margins, hence they will not be effective for new trial information. The first one (H1) has the maximum margin to support arrays including: S11, S12, S13, S21, S22, and S23. Therefore, H1 is selected as the best categorization hyper-plane. A p-D with an N-dimension information set can be presented as follows:

$$\{(x_n, y_n) | x_n \in R^p, y_n \in \{-1, 1\}\}, n = 1, 2, \dots, N \qquad (5)$$

in which y_n will be -1 and 1 related to the first and second category, respectively. Moreover, x_n denotes a p-D array. The hyper-plane with the highest margin, which separates the first category from the second, is the favored support vector array. Regarding that, all hyper-planes can be presented by:

$$wx - b = 0 \qquad (6)$$

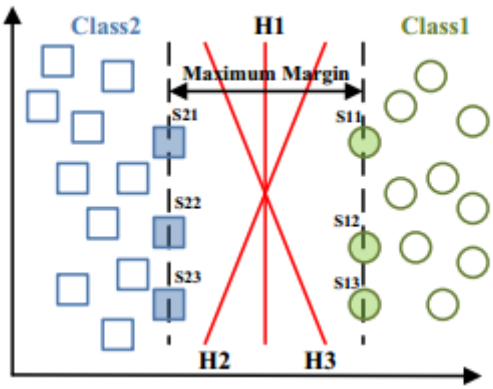

Figure 1. The geometric interpolation of linear support vector machines, where H and S indicate the hyper-plane and support vector, respectively.

The w and b are selected for the maximization of the margin between the 2 parallel hyper-planes as much as is plausible while keeping the division of the information. Therefore, these two parallel hyper-planes are introduced by:

$$wx - b = \pm 1 \tag{7}$$

Thus, we can present the problem under the form of optimization. This optimization aimed to maximize the space between the 2 mentioned hyper-planes and prevent the information from falling in the margin at the same time. Finally, this optimization is represented in the following form:

$$\min_{w,b} \|w\|$$

Subject to:

$$y_n(wx_n - b \geq 1), \quad n = 1, 2, \dots, N \tag{8}$$

In practice, the above optimization is mostly presented under the form of:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

Subject to:

$$y_n(wx_n - b \geq 1), \quad n = 1, 2, \dots, N \tag{9}$$

Therefore, the answer will not differ, but the optimization is transformed to a quadratic programming form, which is simple to solve by means of Lagrange multipliers and criterion quadratic programming (QP) methods.

2. Soft margin

In real-world problems, we do not have the hyper-planes splitting the samples perfectly. In this condition, the soft margin approach can select one of them, which divides the certain samples as plainly as possible, while still maximizing the space between the closest plainly-divided samples.

We presented the positive slack array $\zeta=\{\zeta_1,\dots,\zeta_n,\dots,\zeta_N\}$ for gauging the mis-categorization level of x_n, i.e. the space between the margin and the x_n arrays, which fall on the false side border. Afterwards, the best hyper-plane that divides the information is achievable by using the optimization below:

$$\min_{w,\zeta,b} \frac{1}{2} \|w\|^2 + ce^T\zeta$$

Subject to:

$$\begin{cases} y_n(w^T x_n - b) \geq 1 - \zeta_n \\ \zeta_n \geq 0 \end{cases}, \quad n = 1, 2, \dots, N \tag{10}$$

Here, c denotes the error punishment (it is known as the box constraint in a few related works); e represents an array of N-sized 1s. So, the problem compromises between a great margin and a low c. This limited optimization is solvable by a Lagrange multiplier under the form of:

$$\min_{w,\zeta,b} \max_{\alpha,\beta} \left\{ \frac{1}{2} \|w\|^2 + ce^T\zeta - \sum_{n=1}^{N} \alpha_n [y_n(w^T x_n - b) - 1 + \zeta_n] - \sum_{n=1}^{N} \beta_n \zeta_n \right\} \tag{11}$$

The optimization above which is a min-max type cannot be simply solved. Therefore, this dual method is mostly suggested for solving it:

$$\max_{\alpha} \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_m \alpha_n y_m y_n x_m^T x_n$$

Subject to:

$$\begin{cases} 0 \leq \alpha_n \leq c \\ \sum_{n=1}^{N} \alpha_n y_n = 0 \end{cases}, \quad n = 1, 2, \dots, N \tag{12}$$

The main merit of this scheme of optimization is the elimination of ζ_n, while the creation of a single parameter, i.e. c, is a further limitation over the Lagrange multipliers. Therefore, we have a QP problem that is introduced in the optimization form of a quadratic function in terms of multitude parameters under linear limitations. So, many

approaches are able to solve the equation (18) in only a few milliseconds including SMO (sequential minimal optimization), LS (least square), interior point approach, augmented Lagrange approach, conjugate gradient technique, and simplex method.

3. Kernel support vector machine

Linear forms of support vector machine have some drawbacks because of the linear hyper-planes that are not able to divide complex distributed practical data. For its development for non-linear ones, kernel is implemented on the support vector machines. The mechanism obtained is very similar with the difference that each dot-production is substituted by a non-linear kernel operator. In other words, the kernel SVMs permit fitting the max margin hyper-planes in a transformed characteristic domain. This transformation can be non-linear and the transformed domain can have greater dimensions, so, although the classification operator is a hyper-plane in a characteristic domain with greater dimensions, it can be non-linear in the basic input domain. In the case of each kernel, at least one tuned variable must exist to make the kernel flexible and fit for practical information. Because of its high efficiency, we selected the RBF type of kernel in the following form:

$$k(x_m, x_n) = e^{-\frac{\|x_m - x_n\|}{2\sigma^2}} \qquad (13)$$

Here, σ denotes the scaling parameter. By inserting the equation above in equation (12), the final SVM learning function will be obtained as follows:

$$\max_{\alpha} \sum_{n=1}^{N} \alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_m \alpha_n y_m y_n e^{-\frac{\|x_m - x_n\|}{2\sigma^2}}$$

Subject to:

$$\begin{cases} 0 \le \alpha_n \le C \\ \sum_{n=1}^{N} \alpha_n y_n = 0 \end{cases}, \quad n = 1, 2, \dots, N \qquad (14)$$

As it can be seen, the problem is still in the QP form. Furthermore, the interior point approach is selected to solve the optimization. Nevertheless, we still have a problem that specifies the amounts of C and σ in the final equation.

CBSSO Optimization approach

In order to achieve the optimal values of the two above-mentioned parameters, conventional approaches such as trial and error and grid exploration have been utilized in similar papers. Mostly, these conventional methods lead to high computational burdens, and they fail to ensure the convergence to optimal or at least sub-optimal answers. The present work proposed the CBSSO algorithm for this purpose.

The suggested CBSSO optimization method is presented in this part. First, the traditional type of this algorithm is provided, and subsequently it has been improved for the development of exploration ability.

1. SSO algorithm

This method was first presented in (28) and was inspired from the ability of sharks to find their prey. This procedure comprises the motion of sharks toward their prey due to their smell and the increase of the odor concentration. The stated motion is presented in an optimization form in order to achieve the optimal answer. More descriptions of the SSO method are available in (28). This optimization method has 2 main steps that are summarized below.

A. The initializing step

The initializing step of this method refers to the preparation of primary candidates specified according to the shark's location among the population as follows:

$$\mathrm{SX}_i = [\mathrm{sx}_1^1, \mathrm{sx}_2^1, \dots, \mathrm{sx}_{NP}^1], \quad NP = \text{Population Size} \qquad (15)$$

Here, sx_i^1 denotes the i-th initializing location array (initializing probable answer). The movement velocity of particles can be calculated by:

$$\mathrm{SP}_i = [\mathrm{sp}_{i,1}^1, \mathrm{sp}_{i,2}^1, \dots, \mathrm{sp}_{i,ND}^1], \quad i = 1, \dots NP \qquad (16)$$

in which, $sp_{i,j}^1$ indicates the j-th decision-variable that is produced randomly in a certain interval for particles. So, the cost function of each particle is introduced by OF($\mathrm{SX_{NP}}$) that will be recorded for the next computation of particles. The iterative step zero is the initializing step.

B. Movement procedure

In the second phase, the shark will move in order to attain the prey, which comprises two kinds of motions, i.e. forward and rotational motions. The orientation of the shark according to the great odor concentration is authenticated as:

$$sp_{i,j}^m = \mu_m . R1 . \frac{\partial(OF)}{\partial x_j}\Big|_{sx_{i,j}^m} + \alpha_m . R2 . sp_{i,j}^{m-1}$$
$$j = 1, \dots, ND, \; i = 1, \dots NP, \; m = 1, \dots, M \qquad (17)$$

in which, $\nabla(OF)$ denotes the gradient of cost function, μ_m indicates the gradient co-efficient, m and M are the steps count and max count of steps in forward motion, respectively, μ_m and α_m range between (0, 1]. Also, R1 and R2 range between [0, 1].

By bounding the velocity of the shark, the above relation can be re-written in the following form:

$$\left|sp_{i,j}^{m}\right| = Min\left[\left|\mu_m.R1.\frac{\partial(OF)}{\partial x_j}\right|sx_{i,j}^{m} + \alpha_m.R2.sp_{i,j}^{m-1}\right|, \left|\gamma_m.sp_{i,j}^{m-1}\right|\right]$$
$$j = 1,\ldots,ND,\ i = 1,\ldots,NP,\ m = 1,\ldots,M \quad (18)$$

In this formula, γ_m denotes the max allowable value of the current speed that depends on prior iteration speed. The components of SP_i^m array ($sp_{i,j}^m$) are indicated in equation (29). In case of forward motion (global exploration), the new location validation is provided as:

$$GY_i^{m+1} = SX_i^m + SP_i^m.\Delta t_m \quad i = 1,\ldots,NP \quad m = 1,\ldots,M \quad (19)$$

in which, Δt_m signifies the time range of step m. Thus, the local seeking is presented by:

$$NX_i^{m+1,l} = GY_i^{m+1} + R3.\Delta GY_i^{m+1}$$
$$l = 1,\ldots,L \quad i = 1,\ldots,NP \quad m = 1,\ldots,M \quad (20)$$

In this equation, l = 1, …,L. Also, R3 ranges randomly between (-1, +1), and L denotes the point count in the local seeking for each step.

The best obtained solutions in both shark motions can be written as follows:

$$SX_i^{k+1} = \text{argmax}\{OF(GY_i^{m+1}), OF(NX_i^{m+1,1}),\ldots,OF(NX_i^{m+1,L})\}$$
$$i = 1,\ldots NP \quad (21)$$

2. Enhanced version of the SSO algorithm

A. Chaotic development

Shark smell optimization is developed for its higher local/global exploration ability. The chaotic mechanism is furthered on based on a logistic mapping, which produced different random regions in the exploration domain (30). Hence, subsequent to computing equation (19) for each SSO particle production of GY_{iit+1}, the remaining probable answers are approximated by:

$$ChaoGY_{i,j}^{m+1,1} = GY_{i,j}^{m+1} + chm_{i,h}^{m}.(GY_{i,j}^{m+1} - SX_{i,j}^{m}) \quad (22)$$
$$j = 1,2,\ldots,ND, \quad i = 1,2,\ldots,NP$$

$$ChaoGY_{i,j}^{m+1,2} = GY_{i,j}^{m+1} + chm_{i,h}^{m}.(SX_{i,j}^{m} - GY_{i,j}^{m+1}) \quad (23)$$
$$j = 1,2,\ldots,ND, \quad i = 1,2,\ldots,NP$$

$$ChaoGY_{i,j}^{m+1,3} = SX_{i,j}^{m} + chm_{i,h}^{m}.(SX_{i,j}^{m} - GY_{i,j}^{m+1}) \quad (24)$$
$$j = 1,2,\ldots,ND, \quad i = 1,2,\ldots,NP$$

$$ChaoGY_{i,j}^{m+1,1} = SX_{i,j}^{m} + chm_{i,h}^{m}.(GY_{i,j}^{m+1} - SX_{i,j}^{m}) \quad (25)$$
$$j = 1,2,\ldots,ND, \quad i = 1,2,\ldots,NP$$

In these equations, j and i range between [1, ND] and [1, NP], respectively. Also, $chm^{it}_{i,j}$ denotes the chaotic generated number of j-th decision variable for i-th particle in iterative step m. The application of this operator follows the logistic map function as follows:

$$chm_{i,j}^{m} = 4.chm_{i,j}^{m-1}.(1 - chm_{i,j}^{m-1}) \quad 0 < chm_{i,j}^{m} < 1$$
$$chm_{i,j}^{0} = md_4, \quad chm_{i,j}^{0} \notin \{0.25, 0.5, 0.75\} \quad (26)$$

Here, the suggested function produces higher different values for $chm^{it}_{i,j}$ within the interval (0, 1). Thus, equation (21) can be re-written in the following form:

$$SX_i^{m+1} = \text{argmax}\{OF(GY_i^{m+1}), OF(ChaoGY_i^{m+1,1}), OF(ChaoGY_i^{m+1,2}),$$
$$OF(ChaoGY_i^{m+1,3}), OF(ChaoGY_i^{m+1,4}), OF(NX_i^{m+1,1}),\ldots,OF(NX_i^{m+1,L})\} \quad (27)$$

B. Binary development

According to the gradient computation in relation (17), shark smell optimization can be considered as a continuous-based method. In order to increase its capability, it is changed into a binary-based algorithm by the binary integer decision variables (as can be seen in Figure 2). Therefore, the suggested conversion of the binary space Bin(.) for binary parameters is presented as:

$$bsx_{i,j} = Bin(sx_{i,j}) = \begin{cases} 0.5.Rand(0,1) & if \quad sx_{i,j} = 0 \\ 1 - 0.5.Rand(0,1) & if \quad sx_{i,j} = 1 \end{cases} \quad (28)$$

Here, Rand (0, 1) is a randomly created number within the (0, 1) range. Based on the equation above, the particles can be produced in binary space and mapped to the bsxi,j within the (0, 1) range in real scheme. So, the binary particles are transformed to a favorable format in all the steps of both motions and chaotic development. Finally, all the evolved particles are reconverted to the proper format by sigmoid function as follows:

$$Sig(bsx_{i,j}) = \frac{1}{1 + \exp(-bsx_{i,j})} \quad (29)$$

in which, $Sig(bsx_{i,j}) \in (0,1)$. Also, the $Sig(bsx_{i,j})$ can be converted to the binary format by:

$$sx_{i,j} = \begin{cases} 1 & Rand(0,1) < Sig(bsx_{i,j}) \\ 0 & Rand(0,1) > Sig(bsx_{i,j}) \end{cases} \quad (30)$$

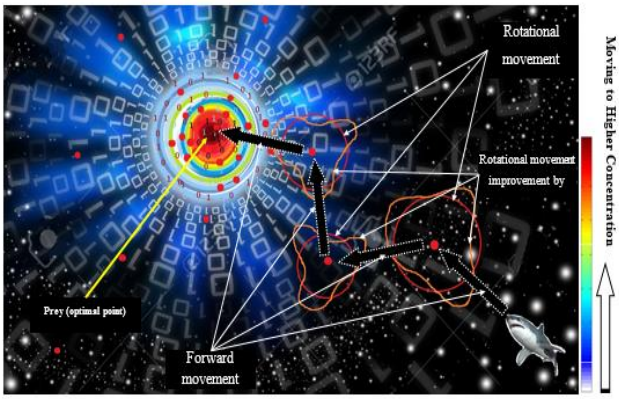

Figure 2. The suggested optimization process

Numerical results and analysis

The simulations are performed on a P4-IBM with a 3.2 GHz processor and eight GB of RAM, which operates under Windows7. The used algorithm is also an in-house improvement on the wavelet-toolbox of Matlab version 2014a. The utilized programs are runnable on any computer equipped with Matlab software.

Moreover, the utilized datasets are composed of ninety T2-weighted MRI brain images in the axial plane with a 256*256 resolution that are gathered from the Harvard Medical School website (URL: http://www.med.harvard.edu/aanlib/home.html). In this information set, images of abnormal brains are composed of some illnesses such as Glioma, Metastatic adenocarcinoma, etc., whose samples are virtualized in Figure 3. It's worth noting that the behavior of all the illnesses is specific to abnormal brains, and our aim was binary categorization into two sections (normal/ abnormal brain).

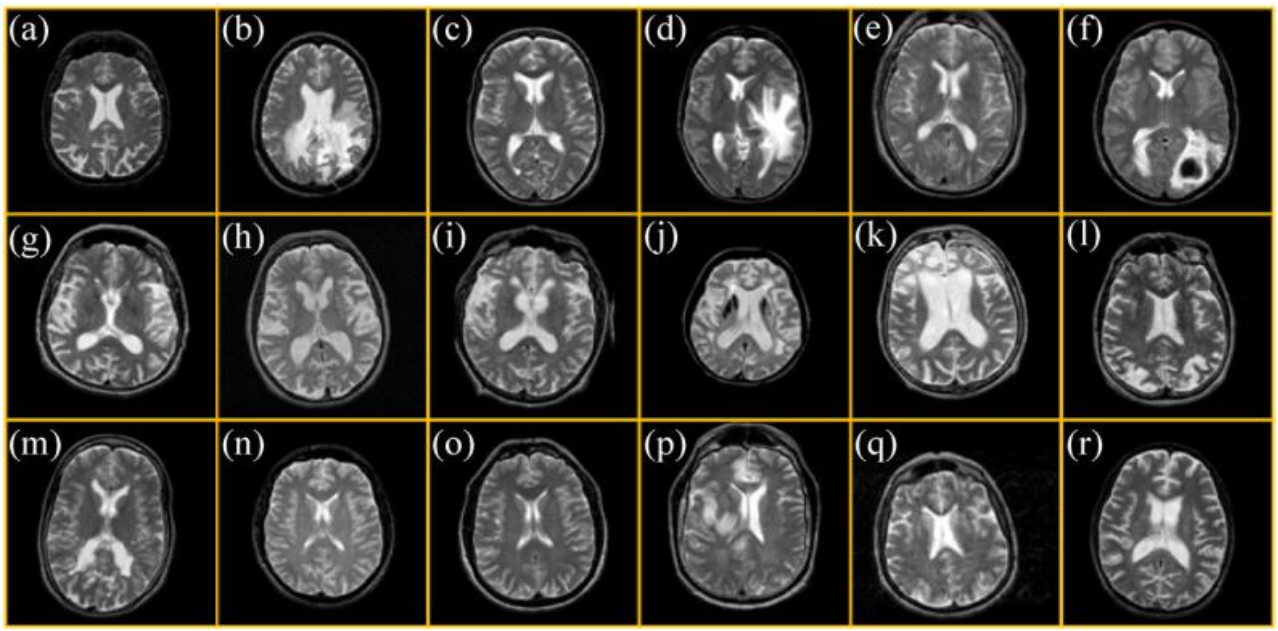

Figure 3. Samples of brain images: (a) normal brain; (b) Glioma; (C) Metastatic Adenocarcinoma; (D) Metastatic Bronchogenic Carcinoma; (E) Meningioma; (F) Sarcoma; (G) Alzheimer's; (H) Huntington; (I) Motor Neuron; (J) Cerebral Calcinosis; (K) Pick; (L) Alzheimer's Plus Visual Agnosia; (M) Multiple Sclerosis; (N) AIDS, Dementia; (O) Lyme Encephalopathy; (P) Herpes Encephalitis; (Q) Creutzfeld-Jakob; (R) Cerebral Toxoplasmosis.

Five images are chosen randomly from each case of brain disorders. Because we have seventeen kinds of abnormal brains and only one normal brain, ninety (5*(1+17)) images are chosen for gathering the information that include sets composed of five normal images and 85 abnormal ones.

Figure 4 depicts the adjustment of five-fold cross-validation. The present work divides the set into five equally diffused sub-classes. Each of the sub-classes includes one normal and seventeen abnormal brain images. Also, 5 experiments are made here, where 4 classes are utilized for each to learn, and the other class is utilized to test. Each class is utilized once for the test.

Additionally, the already mentioned five-fold cross-validation is iterated five times (5*5fold CV is applied in order to decrease haphazardness). The confusion matrices of the five executions are mixed for the organization of the ultimate confusion matrix, by the optimal results of 425 abnormal samples and 25 normal ones that are flawlessly formed.

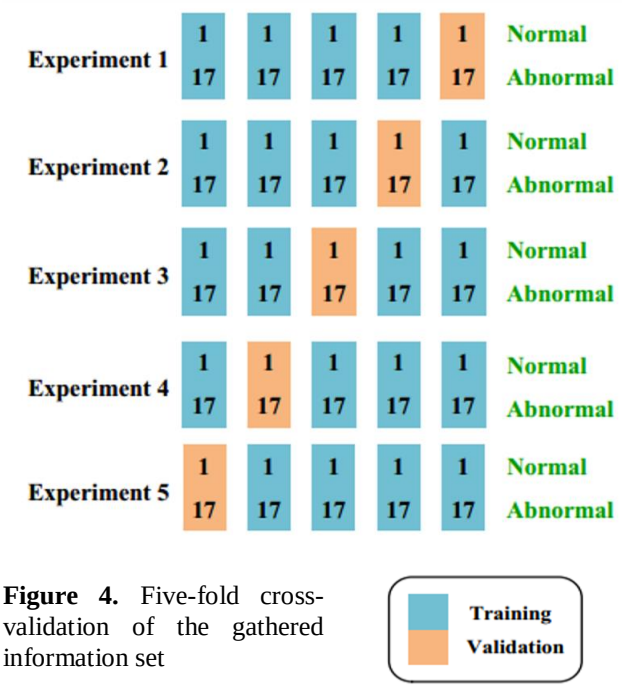

Figure 4. Five-fold cross-validation of the gathered information set

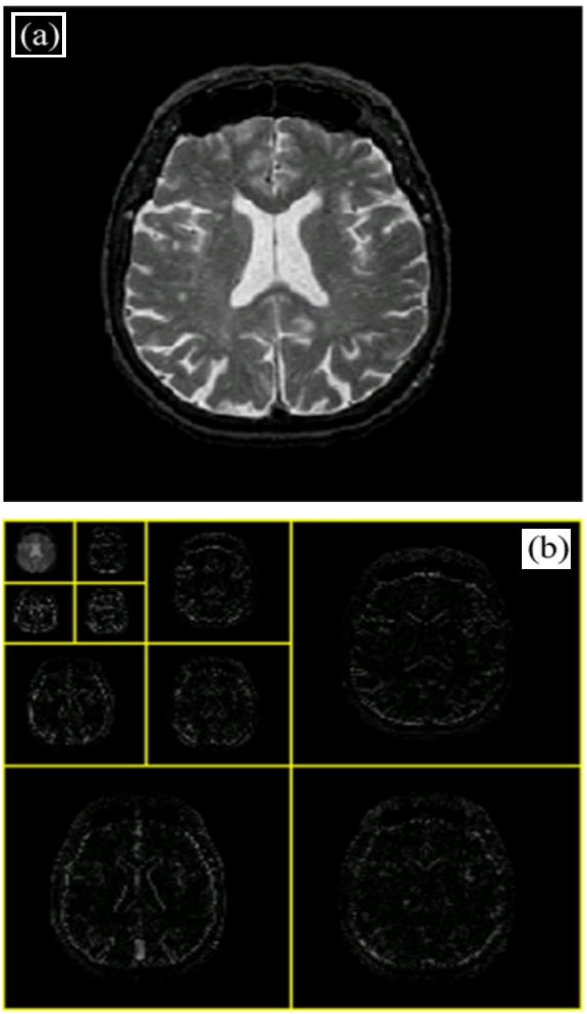

Figure 5. The results obtained for the three-level two dimensional discrete WT; (a) normal brain (b) the results of decomposition.

Feature extraction

The input images are decomposed by a three-level two-dimensional discrete WT into ten sub-bands that are depicted in Figure 5. We employed the symmetric padding approach (31) for the computation of the marginal amount to prevent margin distortion. Then, ten characteristics are obtained by computing the energy for ten sub-bands.

The WE is able to decrease the size of discrete WT parameters. In the 256*256 image used, the three-level discrete WT did not mitigate the size, while the three-level WE dramatically decreased the size of characteristics from 65.536 to ten.

We sent the WE characteristics to various classification operators. The proposed CBSSO-KSVM approach is contrasted with other methods including BP network (32), kernel SVM (29), and PSO-kernel SVM (16), and the results are listed in Table 1. The RBF kernel is employed.

Figure 6 illustrates 2 images of Lyme Encephalopathy, in which, the one on the left is wrongly categorized in normal subsets and the one on the right is properly categorized in abnormal sets. The reason behind this event is that the left one failed to depict the focus and deformed the area. This denotes that the suggested approach can be promoted by means of multi-slices.

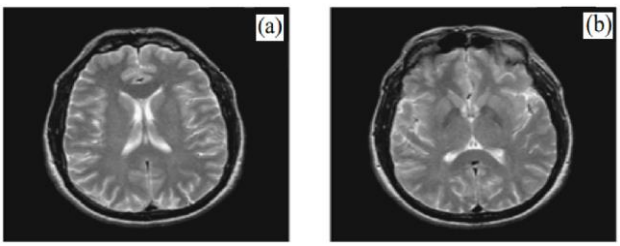

Figure 6. Images of Lyme Encephalopathy; (a) wrongly categorized (b) correctly categorized

The results presented in Table 1 can be considered as the key novelties of the present work. According to this table, the BP network (32), kernel SVM (29) and PSO-KSVM (16) matched 390, 413 and 437 cases correctly and have 86.67, 91.78 and 97.11 percent authenticity in categorization, respectively. Furthermore, the suggested CNSSO-KSVM matched 443 images with 97.94 percent authenticity.

Thus, the approach we suggest is the most efficient method. Furthermore, the sensitivity, specificity, and precision of the suggested method are 98.24, 94.00 and 99.61 percent, respectively. This fact confirmed the high performance of the suggested approach.

Table 1. The comparison of classifiers according to the wavelet energy characteristics

Approach	Confusion matrix	Correct cases	Sensitivity (percentage)	Specificity (percentage)	Precision (percentage)	Accuracy (percentage)
WE+BPNN	[376, 49, 1114]	390	88.47	56.00	97.16	86.67
WE+KSVM	[393, 32, 718]	413	93.18	68.00	98.02	91.78
WE+PSO+KSVM	[415, 10, 322]	437	97.65	88.00	99.28	97.11
Suggested method (WE+CBSSO+KSVM)	[417, 8, 223]	443	98.24	94.00	99.61	97.94

The proposed method consists of kernel SVM and CBSSO. The kernel SVM is a broadly utilized pattern diagnosis and yields proper results. Also, CBSSO is a new approach and an efficient optimization method, which is utilized for the tuning of the adjustable variables of kernel SVM (c and σ); otherwise obtaining the best value for them will be difficult. By unifying the CBSSO and kernel SVM, the categorization performance of kernel SVM is improved. According to the obtained results, this proposed optimization algorithm outperformed the PSO algorithm.

The main novelties of the present work can be summarized as follows:

1) This paper presented a novel method of categorizing MRI images into normal and abnormal groups by WE, support vector machine, and CBSSO optimization method.

2) The efficiency of the approach we suggest is confirmed by other methods (BP network, kernel SVM and PSO-KSVM).

3) The WE is also authenticated as an efficient characteristic in the categorization of MRI images.

Conclusions

The present work suggested a novel method of categorizing the MRI images automatically into two groups (normal and abnormal) by using WE, the support vector machine, and CBSSO. The high performance of the suggested approach was demonstrated by comparison with related works. The proposed CAD system can also be used to categorize the images with various pathological conditions, kinds, and illness modes. In addition, the following achievements can be considered in future works:

a) We can attempt to increase the decomposition level of two dimensional discrete WT in order to check whether a greater level will result in higher efficiency or not.

b) The WE can be substituted by another method with higher efficiency such as scale invariant characteristics.

c) Some efficient pattern diagnostic methods like deep-learning or RBF network can be utilized (30).

d) We can employ multiple-slices for promoting the efficiency of categorization.

Conflict of interest disclosure

There are no known conflicts of interest in the publication of this article. The manuscript was read and approved by all authors.

Compliance with ethical standards

Any aspect of the work covered in this manuscript has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

References

1. Yudong Zhang, Bradley S. Peterson, Genlin Ji, Zhengchao Dong. Energy preserved sampling for compressed sensing MRI. *Computational and mathematical methods in medicine* 2014; Article ID: 546814. DOI: 10.1155/2014/546814
2. Goh S, Dong Z, Zhang Y, DiMauro S, Peterson BS. Mitochondrial dysfunction as a neurobiological subtype of autism spectrum disorder: evidence from brain imaging. *JAMA Psychiatry* 2014; 71(6): 665-71. DOI: 10.1001/jamapsychiatry.2014.179
3. Yudong Zhang, Zhengchao Dong, Preetha Phillips, Shuihua Wang, Genlin Ji, JiquanYang. Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging. *Information Sciences* 2015; 322: 115-32. DOI: 10.1016/j.ins.2015.06.017
4. Razmjooy N, Sheykhahmad FR, Ghadimi N. A hybrid neural network–world cup optimization algorithm for melanoma detection. *Open Med (Wars)*. 2018; 13: 9-16. DOI: 10.1515/med-2018-0002.
5. Sin Ho Lee, Chun Ku Lee, Jin Bae Park, Yoon Ho Choi. Diagnostic method for insulated power cables based on wavelet energy. *IEICE Electronics Express* 2013; 10(12): 20130335. DOI: 10.1587/elex.10.20130335
6. Sandeep C, Patnaik LM, Jagannathan NR. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical signal*

processing and control 2006; 1(1): 86-92. DOI: 10.1016/j.bspc.2006.05.002

7. Madhubanti Maitra, Amitava Chatterjee. A Slantlet transform based intelligent system for magnetic resonance brain image classification. *Biomedical Signal Processing and Control* 2006; 1(4): 299-306.
8. El-Sayed Ahmed El-Dahshan, Tamer Hosny, Abdel-Badeeh M. Salem. Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Processing* 2010; 20(2): 433-41. DOI: 10.1016/j.dsp.2009.07.002
9. Zhang YD, Wu L, and Wang S. Magnetic resonance brain image classification by an improved artificial bee colony algorithm. *Progress In Electromagnetics Research* 2011; 116: 65-79.
10. Yudong Zhang, Zhengchao Dong, Lenan Wu, Shuihua Wang. A hybrid method for MRI brain image classification. *Expert Systems with Applications* 2011; 38(8): 10049-10053. DOI: 10.1016/j.eswa.2011.02.012
11. Ramasamy R, Anandhakumar P. Brain tissue classification of MR images using fast Fourier transform based expectation-maximization Gaussian mixture model. In: Wyld D.C., Wozniak M., Chaki N., Meghanathan N., Nagamalai D. (eds) Advances in Computing and Information Technology. ACITY 2011. Communications in Computer and Information Science, vol 198. Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-22555-0_40
12. Zhang YD, Wu L. An MR brain images classifier via principal component analysis and kernel support vector machine. *Progress In Electromagnetics Research* 2012; 130: 369-88.
13. Saritha M, K. Paul Joseph, Abraham T. Mathew. Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recognition Letters* 2013; 34(16): 2151-6.
14. Das S, Chowdhury M, Kundu MK. Brain MR image classification using multiscale geometric analysis of ripplet. *Progress In Electromagnetics Research* 2013; 137: 1-17.
15. Heidar Ali Shayanfar, Hossein Shayeghi, Aref Jalili. Design Rule-Base of Fuzzy Controller in Multimachine Power System Stabilizer Using Genetic Algorithm. *IC-AI*. 2010.
16. Yudong Zhang, Shuihua Wang, Genlin Ji, Zhengchao Dong. An MR brain images classifier system via particle swarm optimization and kernel support vector machine. *The Scientific World Journal* 2013; (2013). DOI: 10.1155/2013/130134
17. Oveis Abedinia, Morteza Dadash Naslian, Masoud Bekravi. A new stochastic search algorithm bundled honeybee mating for solving optimization problems. *Neural Computing and Applications* 2014; 25(7-8): 1921-39.
18. El-Sayed Ahmed El-Dahshan, Heba M. Mohsen, Kenneth Revett, Abdel-Badeeh M. Salem. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert systems with Applications* 2014; 41(11): 5526-45.
19. Yudong Zhang, Zhengchao Dong, Shuihua Wang, Genlin Ji, Jiquan Yang. Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* 2015; 17(4): 1795-1813. DOI:10.3390/e17041795
20. Bafroui HH, Ohadi A. Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions. *Neurocomputing* 2014; 133: 437-45. DOI: 10.1016/j.neucom.2013.12.018
21. Dai Y, Zhang J, Xue Y. Use of wavelet energy for spinal cord vibration analysis during spinal surgery. *Int J Med Robot.* 2013; 9(4): 433-40. DOI: 10.1002/rcs.1477.
22. Panchal MB, Mahapatra Roy D. Influence of crack parameters on wavelet energy correlated damage index. *Adv Vib Eng* 2013; 12(3): 311-8.
23. Nima Amjady, Oveis Abedinia. Short term wind power prediction based on improved Kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine. *Sustainability* 2017; 9(11): 2104. DOI: 10.3390/su9112104
24. Yudong Zhang, Shuihua Wang, Genlin Ji, Zhengchao Dong. Genetic pattern search and its application to brain image classification. *Mathematical Problems in Engineering* 2013; 2013, DOI: 10.1155/2013/580876
25. Abedinia Oveis, Bart Wyns, Ali Ghasemi. Robust fuzzy PSS design using ABC. Environment and Electrical Engineering (EEEIC), 2011 10th International Conference on. IEEE, 2011.
26. Cai M, Beiji Z, Gao H, Song J. Motion recognition for 3D human motion capture data using support vector machines with rejection determination. *Multimedia tools and applications* 2014; 70(2): 1333-62.
27. Yi-Leh Wu, Chun-Tsai Yeh, Wei-Chih Hung, Cheng-Yuan Tang. Gaze direction estimation using support vector machine with active appearance model. *Multimedia Tools and Applications* 2014; 70(3): 2037-62.

28. Oveis A, Amjady N, Ghasemi A. A new metaheuristic algorithm based on shark smell optimization. *Complexity* 2016; 21(5): 97-116.
29. Muthu Rama Krishnan Mookiah, U. Rajendra Acharya, Choo Min Lim, Andrea Petznick, Jasjit S. Suri. Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. *Knowledge-Based Systems* 2012; 33: 73-82.
30. Oveis Abedinia, Nima Amjady. Short-term wind power prediction based on Hybrid Neural Network and chaotic shark smell optimization. *International Journal of Precision Engineering and Manufacturing-Green Technology* 2015; 2(3): 245-54.
31. Messina A. Refinements of damage detection methods based on wavelet analysis of dynamical shapes. *International Journal of Solids and Structures* 2008; 45(14-15): 4068-97.
32. Navid R, Ramezani M, Ghadimi N. Imperialist competitive algorithm-based optimization of neuro-fuzzy system parameters for automatic red-eye removal. *International Journal of Fuzzy Systems* 2017; 19(4): 1144-56.