Expert Systems With Applications 114 (2018) 313–321

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Nested cross-validation based adaptive sparse representation algorithm and its application to pathological brain classification

Lingraj Dora [a], Sanjay Agrawal [b,*], Rutuparna Panda [b], Ajith Abraham [c]

[a] Department of Electrical and Electronics Engineering, VSSUT, Burla 768018, India
[b] Department of Electronics and Telecommunication Engineering, VSSUT, Burla 768018, India
[c] Machine Intelligence Research (MIR) Labs, Scientific Network for Innovation and Research Excellence, Washington 98071-2259, USA

ARTICLE INFO

Article history:
Received 21 February 2018
Revised 16 July 2018
Accepted 17 July 2018
Available online 25 July 2018

Keywords:
Pathological brain classification
Gray level co-occurrence matrix
Nested cross-validation based adaptive sparse representation algorithm
Nested cross-validation technique

ABSTRACT

Brain disease such as brain tumor, Alzheimer's disease, etc. is a major public health problem, and the main cause of death worldwide. Expert systems are gaining much attention in the medical image analysis field for the clinical treatment and follow up study. Traditional sparse representation based classifiers use a random subset in a limited range. It suffers from the problem of repetition of the training samples which may prevent obtaining optimal subset having all variations. To overcome this problem, nested cross-validation based adaptive sparse representation algorithm is newly proposed. The novelty of the work are: (i) a novel strategy for optimal subset selection, (ii) adaptively selects an optimal subset, (iii) ability to overcome the problems like overfitting, underfitting and bias results, (iv) better accuracy in all variations of training samples, and (v) newly applied to pathological brain classification problem. The proposed system is based on a hybrid methodology of feature selection followed by classification. The gray level co-occurrence matrix is used to extract the spatial texture feature vectors of the brain MRI samples. The nested cross-validation based adaptive sparse representation algorithm is used for classification. It uses a nested cross-validation technique to obtain the optimal value of the subset size (N) based on maximum classification accuracy. The results demonstrate the superiority of the proposed algorithm over the state-of-the-art methods.

1. Introduction

Pathological brain classification is an increasingly important task in medical image analysis. It classifies a given brain image as normal or diseased. It helps us in the early detection of the various brain diseases like brain tumors, Alzheimer disease, Parkinson disease, etc. Magnetic resonance imaging (MRI) is a standard neuroimaging technique used in the clinical assessment of the pathological brain (Kharrat, Benamrane, Messaoud, & Abid, 2009). However, the domain experts make the final decision. Over the last few decades, expert systems play an important role in disease diagnosis and classification (Dora, Agrawal, Panda, & Abraham, 2017b; El-Dahshan, Hosny, & Salem, 2010; El-Dahshan, Mohsen, Revett, & Salem, 2014). The main objectives behind the development of the expert systems are manifold. They could assist domain experts in analyzing a disease. Additionally, false diagnosis due to fatigue as well as inter and intra reader variability could be avoided. Further, they help to reduce the workload. Nevertheless, the complex anatomical structure of the brain makes the design of an expert system for pathological brain classification a challenging task. In recent years, there has been an increasing interest in the design of classifiers using machine learning techniques (Chakraborty, Midya, & Rabidas, 2018; Katzir & Elovici, 2018; Shihabudheen, Mahesh, & Pillai, 2018; Zhang et al., 2018). Sparse representation based methods have gained much appreciation in the field of pattern recognition for the applications like face recognition (Wright, Yang, Ganesh, Sastry, & Ma, 2009). So far, however, there has been little discussion about their use in medical image analysis (Xu, Wu, Chen, & Yao, 2015).

A Gauss-Newton representation based algorithm (GNRBA) was proposed in Dora, Agrawal, and Panda (2017a) and Dora et al. (2017b). It is a sparse representation based method used for breast cancer classification and pathological brain classification. The experiments were conducted using different training-testing partition ratios (50/50, 60/40 and 70/30) to investigate the performance of the GNRBA. However, performance evaluation using these types of ratios may involve class imbalance

* Corresponding author.
E-mail addresses: lingraj02uce157ster@gmail.com (L. Dora), agrawals_72@yahoo.com (S. Agrawal), r_ppanda@yahoo.co.in (R. Panda), ajith.abraham@ieee.org (A. Abraham).

https://doi.org/10.1016/j.eswa.2018.07.039
0957-4174/

problem. To overcome this problem, a 10-fold cross-validation technique was used to evaluate the performance measures. It partitioned the entire dataset into 10 equal sized blocks. The 90% of the blocks were used for training and the remaining 10% utilized for testing. The testing block cover the entire database (i.e. 10 repetitions of the 10% of testing block, exchanged in all runs). However, it still suffers from the problem of repetition of the training samples during its evaluation procedure. This may give rise to problems like overfitting, underfitting and biased results. In view of this, nested cross-validation technique is a good alternative. This has inspired us to propose nested cross-validation based adaptive sparse representation (NCVASR) method. For the first time, it is applied to the problem on hand. The experimental results show the superiority of the system.

A large and growing body of literature has investigated the pathological brain classification problem (Ain, Jaffar, & Choi, 2014; El-Dahshan et al., 2010; El-Dahshan et al., 2014; Jothi et al., 2016). Mostly, the researchers adopted a hybrid scheme, i.e. feature extraction method with classification technique, to solve this problem. For instance, a study by El-Dahshan et al. (2010) uses the discrete wavelet transform (DWT) to extract the features from the T2-weighted brain magnetic resonance (MR) images. To avoid the dimensionality problem, principal component analysis (PCA) is used to transform the DWT features to a low dimensional space. Finally, classification is carried out by using two types of classifiers: k- nearest neighbour (k-NN) and artificial neural network (ANN). The method works well for the classification problem, i.e. normal brain vs. abnormal brain. However, the authors conclude that if the number of images in the database changes, the method needs to be trained again. Ain et al. (2014) reported a method for classification of brain MR image as normal or pathological. In the study, the authors used the first and second order texture feature methods to extract features. Subsequently, ensemble based support vector machine (SVM) is used to carry out the classification task. After the successful identification of the pathological brain, fuzzy c-means (FCM) is used to segment out the tumor region. However, the method is silent about the further classification of the tumor as benign or malignant.

A recent study by Jothi et al. (2016) investigated a hybrid intelligent system to classify normal brain from abnormal brain. A variety of methods are used to extract features from the MRI, such as gray level co-occurrence matrix (GLCM), gray level different matrix, etc. After feature extraction, a tolerance rough set firefly based quick reduct (TRSFFQR) technique is employed for feature selection. The authors identify three types of classifiers: Naive Bayes, J48 and IBK, for classification. However, the overall classification accuracy achieved by the method is low, i.e. 93.5%. Additionally, its performance to different types of medical images are not reported. A recent survey conducted by El-Dahshan et al. (2014) presents the state-of-the-art methods for brain MRI tumor detection and classification. In the survey, the authors also included their proposed method. The method uses the feedback pulse coupled neural network (FPCNN) to segment the tumor from the whole brain MRI. DWT is used to extract the features. Afterwards, PCA is used to avoid the dimensionality problem. Finally, a feedforward backpropagation neural network (BPNN) is used for classification. The state-of-the-art methods in brain tumor classification are used for comparison. However, the method needs training whenever the number of images in the database changes.

Recently, a wrapper approach for data sampling executed via nested cross-validation is found in More (2016). It is used to select an optimal feature subset for an induction learning algorithm. In this approach, the database is partitioned into disconnected training set and test set. The training set is divided into cross-validation sets. The first feature is chosen based on the highest cross-validation accuracy. The process is repeated continuously to add features one-by-one based on maximizing the cross-validation accuracy. Once the termination criterion is reached, the feature selection is stopped. A classifier is trained using the selected features and the full training set. The classifier is tested on the held-out test set, which are not used during the feature selection. However, the performance of this approach depends on the induction algorithm or predictive model used to score the feature subset. Literature study evident that there are some learning algorithms which performs feature selection as part of their overall operation such as sparse regression, regularized trees, decision tree, memetic algorithms, auto-encoding network, etc. The proposed NCVASR algorithm is a sparse regression based method. Unlike feature selection, the training feature vectors selection is a part of its overall operation. In the context of sparse representation based classifier, a new scheme for training feature vectors selection via nested cross-validation is proposed. The suggested method avoids the limitations faced in wrapper approaches such as increasing overfitting risk when the number of observations is insufficient and dependency of performance on induction algorithms.

In view of all that has been mentioned so far, it is evident that feature extraction followed by the classification technique is one of the most suitable ways adopted by the researchers, for the problem on hand. In this work, we have developed a hybrid methodology based on the novel NCVASR method for pathological brain MR image classification. It is hybrid in the sense that the feature extraction scheme is utilized followed by classification technique. GLCM is used to extract the second order texture feature vectors from the input brain MR images (Nabizadeh & Kubat, 2015; Valckx & Thijssen, 1997; Zacharaki et al., 2009). The novel NCVASR is implemented for classification purpose. It sparsely represents the unknown test feature vector as the linear weighted summation of all the training feature vectors. All the training feature vectors are from different classes. Thus, all the training feature vectors in the representation are not essential to classify a test feature vector, which belongs to a particular class only. These training feature vectors may have an adverse effect of increasing the misclassification rate. These types of training feature vectors need to be discarded from the representation for accurate classification. The suggested NCVASR method selects a subset having N relevant training feature vectors. A maximum class contribution criteria is employed to classify the test feature vector. We have proposed to use the nested cross-validation technique to adaptively optimize the value of N (Reunanen, 2003; Statnikov, Aliferis, Tsamardinos, Hardin, & Levy, 2004). It uses an inner cross-validation loop to obtain the optimal value of N using the complete classification procedure subjected to maximum classification accuracy, as discussed above. An outer cross-validation loop is used to classify a set of new unseen test feature vectors using the optimal value of N. The use of nested cross-validation technique facilitates us to adaptively optimize the value of N. Further, the problems like overfitting as well as bias in the result are also avoided.

Traditionally, the sparse representation based classifiers use a subset of training samples during classification. The size of the subset is randomly selected from the range $[0.05 \times T,\ 0.2 \times T]$, where T is the total number of training samples. This is a trial and error scheme. However, it suffers from the problem of repetition of the training samples during its evaluation procedure. This repetition may prevent us to obtain an optimal subset of training samples having all variations. In this context, we have proposed a classifier based on nested cross-validation technique to select the optimal subset for accurate classification, which is a novel idea. Whenever, a new test sample arrives, the proposed algorithm adaptively selects an optimal subset that best represents the new test sample. The novelty of this work are: (a) a novel strategy for optimal subset selection is proposed for the sparse representation based methods, (b) adaptively selects an optimal subset whenever a new test

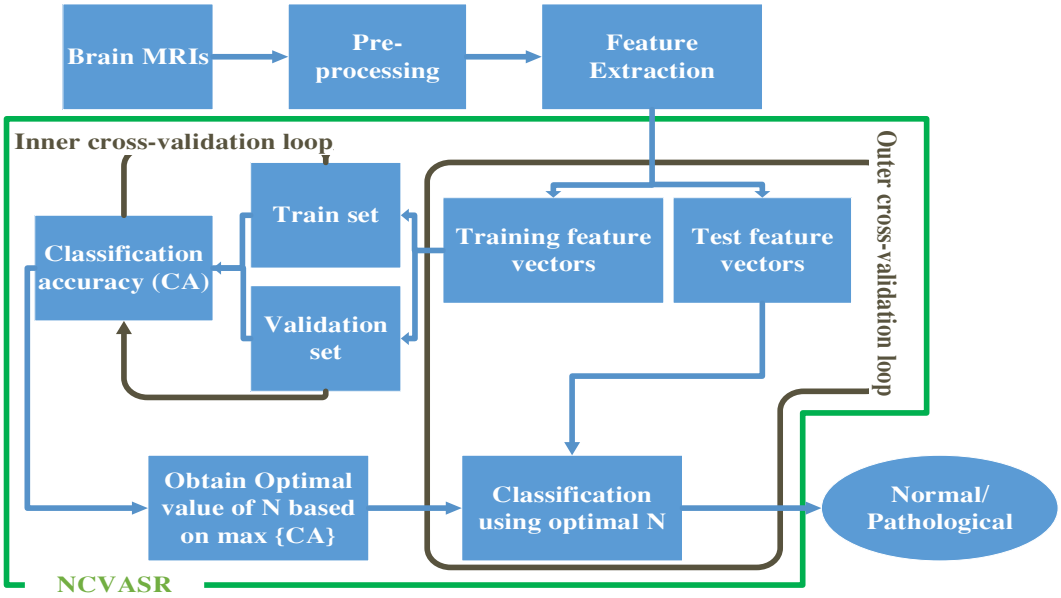

Fig. 1. Block diagram of the suggested hybrid methodology.

sample is applied for classification, (c) this idea provides ability to overcome the problems like overfitting, underfitting and bias results, which are common in conventional classification tasks, (d) it allows us to evaluate the classification accuracy in all variations of training samples by partitioning the samples into multiple training and validation sets, and (e) this idea is newly applied to pathological brain classification problem.

The overall arrangement of this paper is composed of four different sections, including this introduction section. Section 2 is concerned with the suggested method. Section 3 presents the experimental results as well as includes a discussion of the findings. Finally, Section 4 outlines the conclusion.

2. Proposed method

The suggested scheme is a hybrid methodology based on, feature extraction combined with classification, for pathological brain classification. At first, the second order texture features are extracted from the pre-processed brain MR images. NCVASR algorithm makes use of these features to classify an unknown brain MRI as normal or pathological. The block diagram of the suggested method is shown in Fig. 1. Initially, the pre-processing technique is applied to all the images in the database. The texture feature vectors from the pre-processed images are extracted using the GLCM. These feature vectors are partitioned into two different sets, i.e. training feature vectors and test feature vectors. The training feature vectors are used by the NCVASR for classification of the test feature vectors. It intends to determine a subset having N significant training feature vectors nearest to the test feature vector using nested cross-validation technique. The inner loop exploits these training feature vectors by partitioning them into a train set and a validation set. An optimal value of N is obtained from the inner cross-validation loop to select the subset, based on maximum classification accuracy. Finally, the subset is exploited for classification of the unseen test feature vectors. The following sub-sections explain the proposed method in detail.

2.1. Pre-processing

In this stage, skull stripping is employed as a pre-processing step. We have used Brainsuit to remove the skull from the brain MRI, which is a popular freely available toolbox (Brainsuit, 2017). After skull stripping, intensity normalization technique is used to make the suggested approach insensitive to contrast variation (Jen & Yu, 2015). For a given brain MR image $I(i, j)$, the normalized image is obtained as:

$$I_N(i, j) = 255\left(\frac{I(i, j) - I_{min}}{I_{max} - I_{min}}\right) \quad (1)$$

Where, $I(i, j)$ is the intensity of the pixels at location (i, j), $I_{\min}$ is the minimum intensity value and $I_{\max}$ is the maximum intensity value. An example of pre-processed image is shown in Fig. 2.

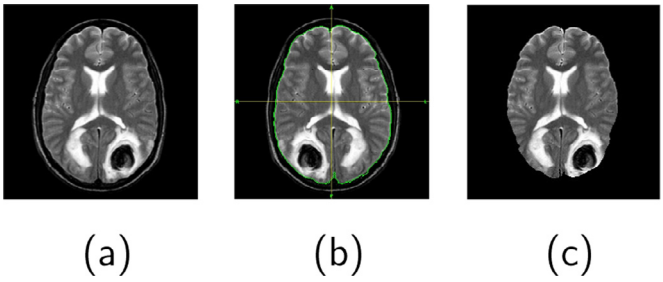

Fig. 2. Example of pre-processing. (a) Original brain MRI. (b) Brain boundary detection using Brainsuit. (c) Skull stripped MRI.

2.2. Feature extraction

A considerable amount of research has been published on the feature extraction methods. These studies include methods like wavelet transform, Gabor filter, GLCM, etc. (Nabizadeh & Kubat, 2015; Valckx & Thijssen, 1997; Zacharaki et al., 2009). A good set of features provides us an opportunity to perform accurate classification. However, the medical images like MRI has some inherent artifacts such as complex shape, size, texture, etc. making feature extraction a challenging task. Additionally, anatomical structure of the MRI differs remarkably with person to person. In recent years, much of the literature available on medical image analysis used GLCM for feature extraction (Ain et al., 2014; El-Dahshan et al., 2014; Jothi et al., 2016). Collectively, these studies, evidence that it is still used as one of the texture modelling technique. This is

the reason, we have used GLCM for feature extraction from the brain MRI. It is a technique which considers the spatial information of an image to capture the second order statistical features. They characterize the joint probability distribution between the intensity levels. This type of distribution informs us about the occurrence of two intensity levels, apart from each other by a distance d and angle θ. A detailed discussion of the GLCM is found in Valckx and Thijssen (1997). In the spatial domain, each intensity level is surrounded by eight neighbouring intensity levels. Thus, there are eight possible angles such as $\theta = 0°$, 45°, 90°, 135°, 180°, 225°, 270° and 315°. It is suggested that the GLCM possesses symmetric behavior, i.e. co-occurrence of the intensity levels calculated at $\theta = 0°$ is same with the pair evaluated using $\theta = 180°$. This is well carried out for other angles like $\theta = 45°$, 90° and 135°. In this work, we have used $d = 1$ and only four values of $\theta = 0°$, 45°, 90° and 135° to extract the texture features. The final feature vector is constructed by averaging the four GLCMs for each θ.

2.3. Nested cross-validation based adaptive sparse representation

This section describes the novel NCVASR. It utilizes the feature vectors extracted from GLCM for classification. It is based on the principle of the sparse representation (Wright et al., 2009). It represents an unknown test feature vector as the linear weighted summation of all the training feature vectors, given as:

$$t = \phi_1 y_1 + \phi_2 y_2 + \cdots + \phi_T y_T \tag{2}$$

where, t is the unknown test feature vector, $y = [y_1, y_2, \ldots, y_T]$ is the set of the T training feature vectors and $\phi = [\phi_1, \phi_2, \ldots, \phi_T]$ is the random weight assigned to all the training feature vectors. The expression in (2) uses all the training feature vectors to represent an unknown test feature vector. It is to be noted that, all the training feature vectors are from different classes. This study deals with a binary classification problem. It is evident that, if all the training feature vectors in the representation are utilized, this may possibly increase the misclassification rate (Xu, Zhang, Yang, & Yang, 2011). A way to decrease it is to select a subset of significant training feature vectors that are nearest to the test feature vector. The Euclidean distance is viewed as a measurement to calculate the similarity between a test feature vector (t) and a training feature vector y_i, for $i = 1, 2, \ldots, T$, as given below:

$$d_i = \|t - y_i\|_2 \tag{3}$$

From (3), a small distance d_i indicates that the ith training feature vector (y_i) is nearest to the test feature vector t. We have used (3) to extract a subset having N nearest training feature vectors, such that $N \ll T$. Additionally, the remaining training feature vectors are discarded for accurate classification. Further, we preserved the class labels of the N training feature vectors as the candidates for the class label of the unknown test feature vector. However, it is not clear, for the different values of N which subset will help us to achieve good classification accuracy from the available alternatives. Moreover, the selection of the value of N on trial and error basis may result in overfitting problem. If the subset with this value of N is used, it may not perform well to an unseen test feature vector despite its excellent performance on the known feature vectors. In Xu et al. (2011), it is suggested that, the value of N is selected from the range of [0.05 × T, 0.2 × T]. However, it still suffers from the problem of repetition of the training feature vectors during its evaluation procedure. This repetition may prevent us to obtain a good value of N to select a subset having all variations that would be observed, if each subset containing the training feature vector is completely independent of the previous subset.

In view of this, the nested cross-validation technique is a good alternative. In this paper, we have used this technique to optimize the value of the parameter N (Reunanen, 2003; Statnikov et al., 2004). Literature study suggests that it is a well-known technique used to estimate the performance of the method (Statnikov et al., 2004). The inner loop is used to optimize the hyper-parameters using the complete classification procedure of the method. The outer loop is used to estimate the classification rate using the optimal hyper-parameters value of an unseen test data. This test data is not used in the inner loop during hyper-parameters optimization. We have used the inner cross-validation loop to optimize the value of N based on maximum classification accuracy. This optimal value of N is used to select a subset to classify the unseen test feature vectors in the outer cross-validation loop.

Fig. 3 illustrates the procedure of optimizing the parameter N for the subset selection using a $m \times n$ nested cross-validation technique. In the outer loop, it divides all the feature vectors (training plus test) into m different blocks. Out of the m blocks, $m-1$ blocks are used as the training feature vectors. The remaining one block is used as the test feature vectors (test block). The inner loop performs n-fold cross validation using the feature vectors of the $m-1$ blocks by partitioning it into a training set and a validation set. The classification accuracy, i.e. $CA_i = \frac{TP+TN}{TP+TN+FP+FN}$ for $i = 1, 2, \ldots, n$, resulting from each fold in the inner loop is calculated, where $TP\rightarrow$ number of true positives, $TN\rightarrow$ number of true negatives, $FP\rightarrow$ number of false positives and $FN\rightarrow$ number of false negatives. On completion of the inner loop, an optimal value of N is obtained based on the classification accuracy, as given below:

$$N_{adap(x)} = \arg\max_{N}\{CA_i\} \tag{4}$$

where, $N_{adap(x)}$ for $x = 1, 2, \ldots, m$ is the optimal value of N used in the outer loop to classify the new unseen test block. The outer loop is repeated for m times by exchanging the test block in all runs. In this way, the nested cross-validation procedure facilitates us to adaptively select the optimal value of N whenever a new unseen test feature vector arrives. Additionally, it allows us to optimize the performance of the suggested NCVASR method while avoiding overfitting problem and possible bias in the result.

As depicted from Fig. 3, the inner cross-validation loop is used to find the optimal value of N to select a subset. The test feature vector is now represented as the linear weighted summation of the training feature vectors from the subset, given as:

$$t = \phi_1 \widehat{y}_1 + \phi_2 \widehat{y}_2 + \cdots + \phi_N \widehat{y}_N \tag{5}$$

where, $\widehat{Y} = [\widehat{y}_1, \widehat{y}_2, \ldots, \widehat{y}_N]$ is the optimal subset of the N training feature vectors. From (5), it is observed that the contribution of the jth training feature vector towards t is $\phi_j \widehat{y}_j$, for $j = 1, 2, \ldots, N$. However, the random value of ϕ in the representation results in inconsistency in indicating the contribution of the training feature vectors. The optimal value of ϕ is used to maintain the consistency in the contribution that each of the training feature vectors make towards t. A way to obtain the optimal ϕ is by minimizing the sum of square error (Dora et al., 2017b; Gill, Murray, & Wright, 1981; Jang, Sun, & Mizutani, 1997). We have employed this scheme to update the value of ϕ, as given below:

$$\phi_{next} = \phi_{now} + \Delta\phi \tag{6}$$

where, $\Delta\phi = (\widehat{Y}^T\widehat{Y} + \mu I)^{-1}\widehat{Y}^T \Delta f$, μ is a regularization parameter, I is the identity matrix and Δf is the difference between the actual output and the desired output when the input is $\widehat{Y}$. In this paper, $\mu = 0.1$ is used as suggested in Jang et al. (1997) and Gill et al. (1981). After the termination criterion is reached, the optimal ϕ_{opt} equals ϕ_{next}.

After obtaining the optimal ϕ_{opt}, a class contribution criterion is used to classify the test feature vector. The criteria calculates the weighted sum of the training feature vectors from the same class,

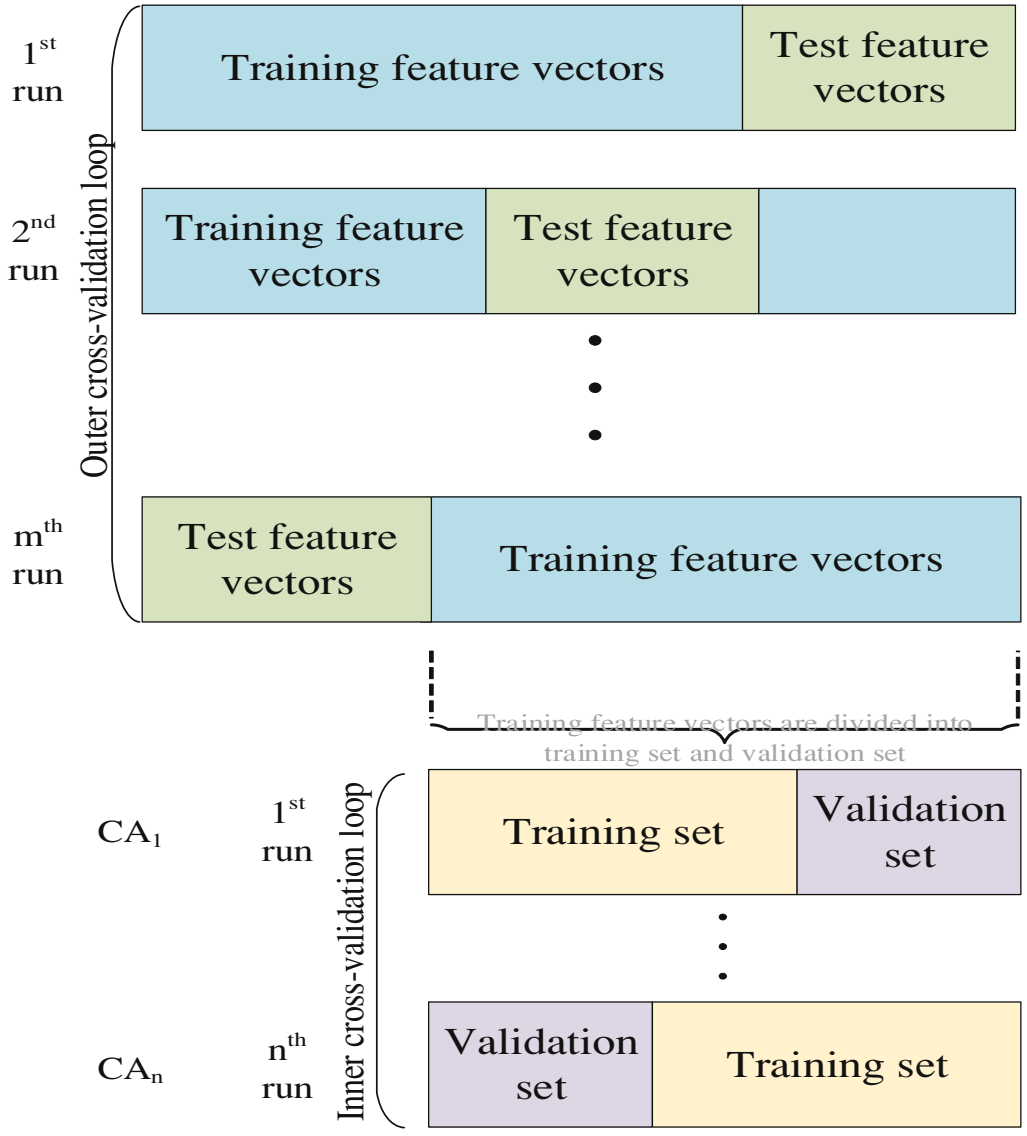

Fig. 3. $m \times n$ nested cross-validation technique for the subset selection.

given as:

$$C_L = \sum_{l=1}^{n_l} \phi_{(opt)l} \widehat{y}_l, \ L = 1, 2, \ldots, C \tag{7}$$

where, n_l is the number of training feature vectors from the Lth class and C is the total number of classes. The class label of the test feature vector is identified as:

$$D_L = \|t - C_L\|_2 \tag{8}$$

From (8), a smaller distance between t and C_L indicates the test feature vector is identified to L^{th} class. The algorithm of the proposed NCVASR for pathological brain classification is given as follows:

3. Results and discussions

The experiments are conducted using MATLAB software on a MAC with core *i*5. To show the robustness of the proposed method, real patients brain MRI samples are collected from the Harvard Whole Brain Atlas (HWBA) database (Atlas, 2017). Four different datasets (Dataset 1, Dataset 2, Dataset 3 and Dataset 4) are constructed from the above collected brain MR images. Dataset 1 and Dataset 2 contains seven types of pathological brains. The Dataset 3 consists of eleven types of pathological brains. The Dataset 4 includes eighteen types of pathological brain MRIs. The detail of training and testing partition used in the three datasets is depicted in Table 1. Detailed description about the database is found in the website.

Table 1
Detail of training and testing partition.

Dataset	Training		Testing		Total	
	H	P	H	P	H	P
Dataset 1 (D1)	15	40	3	8	18	48
Dataset 2 (D2)	16	112	4	28	20	140
Dataset 3 (D3)	16	144	4	36	20	180
Dataset 4 (D4)	28	176	7	44	35	220

H → Normal brain MRI; P → Pathological brain MRI

Example images from the HWBA database are shown in Fig. 4.

The proposed method uses the $m \times n$ nested cross-validation technique to obtain the best performance estimates as well as to avoid the overfitting problem. It involves two loops, i.e. an outer loop and an inner loop. For the experiment, we have used the

Algorithm 1 NCVASR.

Require: Feature vectors of all the brain MRIs from the database.
$Z \leftarrow$ Input feature vectors are partitioned into m equal blocks.
for $a = 1 : m$ **do**
  $B \leftarrow a^{th}$ block from Z (outer loop test feature vectors).
  $A \leftarrow$ remaining blocks from Z (outer loop training feature vectors).
  $W \leftarrow A$ is partitioned into n equal blocks.
  for $b = 1 : n$ **do**
    $F \leftarrow b^{th}$ block from W (validation set in the inner loop).
    $E \leftarrow$ remaining blocks from W (training set in the inner loop).
    for $i = 2 : n_2$ (number of vectors in E) **do**
      for $j = 1 : n_1$ (number of vectors in F) **do**
        $f \leftarrow F(j)$, j^{th} validation feature vector
        $d_x \leftarrow \|f - E_x\|_2$; $x = 1, 2, \ldots, n_2$.
        $d_{x(s)} \leftarrow \min\{d_x\}$; $s = 1, 2, \ldots, N$.
        find optimal ϕ_{opt} using (6).
        Calculate class contribution using (7).
        Calculate class label of f using (8).
      end for
    end for
    $[CA_b, N_b] \leftarrow \max\{CA_i\}$; CA = classification accuracy
  end for
  $N_{adap(a)} \leftarrow \arg\max_{N_b}\{CA_b\}$ using (4)
  for $i = 1 : n_3$ (number of vectors in B) **do**
    $g \leftarrow B(i)$, i^{th} test feature vector
    $d_h \leftarrow \|g - A_h\|_2$; $h = 1, 2, \ldots, n_4$ (number of vectors in A).
    $d_{h(r)} \leftarrow \min\{d_h\}$; $r = 1, 2, \ldots, N_{adap(a)}$.
    find optimal ϕ_{opt} using (6).
    Calculate class contribution using (7).
    Calculate class label of f using (8).
  end for
  Return: Class of the unknown test feature vectors in B.
end for

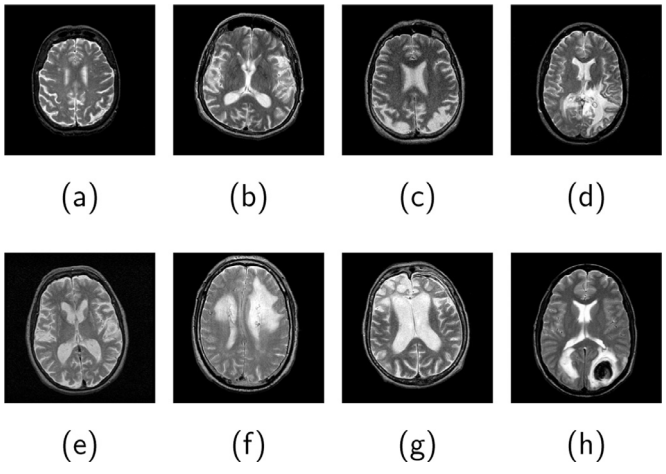

Fig. 4. Sample images from the HWBA database. (a) Normal brain MRI. (b) - (h) Pathological brain MRI.

value of $m = 5$ and $n = 2$. For instance, Table 2 shows the statistics of the number of images employed in both the loops using D3. The outer loop uses a stratified m-fold cross-validation technique. From Table 2, it is observed that for the value of $m = 5$ the database is divided into 5 equal blocks, i.e. $200/5 = 40$ images in each block. Out of the 5 blocks, 4 blocks are used in the outer loop as the training images (training block i.e. $4 \times 40 = 160$ images). The remaining one block is used as the unseen test images

Table 2
Details of training-testing ratio used in the experiment based on D3.

Outer loop iteration	Total images	Training-testing partition ratio			
		Outer loop		Inner loop	
		Training	Testing	Training	Validation
1	200	160	40	80	80
2	200	160	40	80	80
3	200	160	40	80	80
4	200	160	40	80	80
5	200	160	40	80	80

Table 3
CA, *SEN*, *SPE*, and *AUC* of the NCVASR.

Performance index	Value	D1	D2	D3	D4
CA	Mean	100%	99.38%	98.83%	99.5%
	Max	100%	100%	100%	100%
	Std. dev.	0	1.9220	1.1524	0.1250
SEN	Mean	1.0000	1.0000	0.9867	0.9973
	Max	1.0000	1.0000	1.0000	1.0000
	Std. dev.	0	0	0.0014	0.0280
SPE	Mean	1.0000	0.9600	1.0000	0.9826
	Max	1.0000	1.0000	1.0000	1.0000
	Std. dev.	0	0.0080	0	0.0135
AUC	Mean	1.0000	0.9978	0.9714	0.9935
	Max	1.0000	1.0000	1.0000	1.0000
	Std. dev.	0	0.0192	0.0061	0.0223

std. dev. → standard deviation

in the outer loop (test block). The outer loop is repeated 5 times by changing the test block in each iteration. The training block from the outer loop is exploited in the inner loop for parameter optimization. The inner loop uses a stratified n-fold cross validation technique. For the value of $n = 2$, the training block is partitioned into 2 equal blocks, i.e. $160/2 = 80$ images in each block. These blocks are employed as training and validation blocks in the inner cross-validation loop.

The performance indices such as classification accuracy (*CA*), sensitivity (*SEN*), specificity (*SPE*) and area under receiver operating characteristics (*AUC*) of the suggested technique, calculated using the four datasets is presented in Table 3 (Fawcett, 2006; Sokolova & Lapalme, 2009). From the table it is observed that the suggested method achieved 100% accuracy in terms of *CA*, *SEN*, *SPE* and *AUC* in D1. The reason may be the use of less instances for testing as compared to the training instances. Additionally, for the other three datasets it also shows a good performance with accuracy more than 97%. Moreover, the suggested NCVASR method is compared with the different methods to show its superiority.

The comparison is based on the performance indices such as *CA*, *SEN*, *SPE*, confusion matrix (*CM*) and *AUC* (Fawcett, 2006; Sokolova & Lapalme, 2009). Recent literature on pathological brain classification is used to select the methods for the comparison. All these methods used images from HWBA database. Thus, they are included for a fair comparison. Table 4 shows the comparison based on *CA* using D1, D2 and D3.

It is depicted from the table that our approach outperforms all other methods using the three datasets. The reason may be the optimal selection of the subset.

Table 5 presents the comparison of the proposed NCVASR method with different methods based on *CA*, *SEN* and *SPE* using D4. It is apparent from the table that our method outperforms all other methods. The reason for this superior performance is the ability to select the subset having the most significant training feature vectors from the optimal value of *N*. In addition, the value of *N* is adaptive to the new unseen test feature vector.

The following analysis provides deeper insights about the proposed method. The CA obtained with earlier methods using D1, D2

Table 4
Comparison based on *CA* using D1, D2, and D3.

Name of the method	D1	D2	D3
DWT+ self-organizing map (Chaplot, Patnaik, & Jagannathan, 2006)	94%	93.17%	91.65%
DWT+SVM (Chaplot et al., 2006)	96.15%	95.38%	94.05%
DWT+SVM+radial basis function (Chaplot et al., 2006)	98%	97.33%	96.18%
DWT+PCA+k-NN (El-Dahshan et al., 2010)	98%	97.54%	96.79%
DWT+PCA+ANN (El-Dahshan et al., 2010)	97%	96.98%	95.29%
DWT+PCA+SVM (Zhang & Wu, 2012)	96.01%	95%	94.29%
Wavelet entropy+naive Bayes classifier (NBC) (Ortuño & Rojas, 2015)	92.58%	91.87%	90.51%
Wavelet packet Shannon entropy+SVM (Zhang et al., 2015)	98.64%	97.12%	97.02%
Wavelet packet Tsallis entropy+SVM (Zhang et al., 2015)	99.09%	98.94%	98.39%
Fractional fourier entropy (FRFE)+ Welchs *t*-test+NBC (Wang et al., 2015)	97.12%	95.94%	95.69%
FRFE+multi-layer perceptron (Zhang et al., 2016)	99.85%	98.38%	97.02%
NCVASR	100%	99.38%	98.83%

Table 5
Comparison of *CA*, *SEN*, and *SPE* using D4.

Name of the method	***CA*(%)**	***SEN***	***SPE***
Independent component analysis + SVM (Wang & Fei, 2009)	79	0.8700	0.7500
Pearsons correlation coefficients + SVM (Wang & Fei, 2009)	82	0.8900	0.7700
PCA + SVM (Wang & Fei, 2009)	85	0.8900	0.8400
FPCNN+DWT+PCA +BPNN (El-Dahshan et al., 2014)	99	1.0000	0.9280
NCVASR	99.5	0.9973	0.9826

and D3 datasets is found above 90%, whereas with D4 dataset their performances fall well below 80%. The reason is that the number of pathological brain types and the number of testing instances are more in D4. On the other hand, our proposed method performs well above 99% with respect to *CA* in case of D4 dataset. Because the role of *N* (subset size) is crucial in evaluating the *CA*. The optimal subset selection affects the values of TP ↑, TN ↑, FP ↓ and FN ↓. It increases the numerator values (TP + TN) and decreases the denominator value (TP + TN + FP + FN). As a result, the *CA* increases. If FP and FN are zero, then CA becomes 100%.

For instance, in D1 dataset, the number of pathological brain and the testing instances are less (P = 7 and H + P = 11) during one outer loop of cross-validation. Our method yields 100% *CA* with unity value of *SEN, SPE* and *AUC*, which is interesting. The reason is the FP and FN values become zero, which makes the *CA* 100%. The optimal subset selection idea incorporated in this work helps in identifying a pathological brain as pathological and a normal brain as healthy. Further, the idea of nested cross-validation avoids inclusion of images from a certain class into a class having different characteristics. This controls the values of FP and FN and keeps them to a minimum.

It is also observed that the proposed work performs well as compared to the existing methods while considering D4 dataset. Note that number of pathological brain types and the number of testing instances are more (P = 18 and H + P = 51) during one outer loop of cross-validation. For example, the method using ICA + SVM could obtain a *CA* of 79% only whereas our method yields 99.5%. Though there is a slight decrease in the value from 100%, yet the method is able to score a much better value even with an increased number of testing instances. It is quite obvious from the literature that the increase in the number of instances will never make the values of FP and FN zero. Therefore, the *CA* is less than 100%. A similar trend is also observed in the values of *SPE* and *SEN*.

Table 6
Comparison based on *CM* using D4.

Name of the method	Actual value	Predicted value	
		H	P
FPCNN+DWT+PCA+BPNN (El-Dahshan et al., 2014)	H	87	0
	P	1	13
NCVASR	H	35	0
	P	1	219

Table 7
Comparison of the *AUC* value using D4.

Name of the method	***AUC* value**
FPCNN+DWT+PCA+BPNN (El-Dahshan et al., 2014)	0.9800
NCVASR	0.9935

It is to be noted that the *SEN* value obtained with our method is slightly less than the FPCNN+DWT+PCA+BPNN method. The reason is that the number of test samples used is very less and the authors have presented the maximum value of *SEN*. The number of test samples used in our method is more and we have presented the average value of SEN. In fact, we have also achieved a maximum value of 1 for *SEN* as shown in Table 3.

A comparison based on *CM* using D4 is shown in Table 6. *CM* is a quantitative measure used to evaluate the predicted classification result with the actual classification result. The FPCNN + DWT + PCA + BPNN method is a state-of-the-art method for brain tumor classification. The authors compared their method with the other methods for brain tumor classification using HWBA database.

As seen from Table 6, the *CM* value indicates the predicted classification result with the actual classification result. The sum of predicted values from each class should be close to the actual values in that class. For instance, the total number of test samples used in FPCNN + DWT + PCA + BPNN method is 101, out of which the actual number of normal sample (H) is 87 and pathological sample (P) is 14. It is observed that the predicted value of H using the above method is 87, whereas the predicted value of P is 13 and one sample is misclassified as H. On the other hand, in our proposed method, the total number of samples taken is 255 (H = 35 and P = 220). It is observed that the predicted values are H = 35, P = 219 and one sample is misclassified. The reason is that the number of test samples (H + P = 101) and the pathological samples (P = 14) used are very less and the authors have presented the maximum value of *CM*. Note that the number of pathological instances (P = 220) and the number of test samples (H + P = 255) used in our method is more (more than double). We have presented the sum of all *CM*s obtained from all the cross-validations. In fact, we have also obtained a maximum *CA* of 100%, as shown in Table 3, which indicates that all the predicted values are equal to the actual values with a zero misclassification, using D4 dataset. This is due to the fact that the incorporation of optimal subset selection via nested cross-validation helps us to decrease the misclassification rate. It is observed that the suggested NCVASR method performs better as compared to the state-of-the-art method.

Table 7 presents the results obtained from the *AUC* analysis of the FPCNN + DWT + PCA + BPNN and the proposed method using D4.

The area under receiver operating characteristic curves based on the trapezoidal rule is used to determine the *AUC* values. The *AUC* value varies between 0 (worst performance) and 1 (best performance). The optimal subset selection strategy helps us to control the values of TP, TN, FP and FN in such a way that *SEN* is increasing rapidly. At the same time, $1 - SPE$ is hardly increasing until *SEN* reaches a high value. Due to this characteristics, a larger

Table 8
p-value comparison of the proposed NCVASR method with other method using Friedman test.

Name of the method	***p*-value**
DWT + self-organizing map	0.0285
DWT + SVM	0.0285
DWT + SVM + radial basis function	0.0314
DWT + PCA + k-NN	0.0209
DWT + PCA + ANN	0.0209
DWT + PCA + SVM	0.0285
Wavelet entropy + NBC	0.0219
Wavelet packet Shannon entropy + SVM	0.0314
Wavelet packet Tsallis entropy + SVM	0.0432
FRFE + Welchs *t*-test + NBC	0.0314
FRFE + multi-layer perceptron	0.0432
Independent component analysis + SVM	0.0012
Pearsons correlation coefficients + SVM	0.0012
PCA + SVM	0.0012
FPCNN + DWT + PCA + BPNN	0.0455

area is occupied under the curve and thereby a large *AUC* value (close to 1) is achieved by our suggested approach. The results, as shown in Table 7, indicate that our method outperforms the FPCNN + DWT + PCA + BPNN method in terms of the *AUC* value. The *AUC* value of other methods, as discussed above, are not reported in the literature. Thus, they are not included in Table 7 for comparison.

Moreover, for a thorough comparison, we have conducted the Friedman statistical analysis test with a significance level of 0.05 (Demšar, 2006). Table 8, presents the *p*-value obtained by performing the Friedman test between the suggested NCVASR method and all other methods.

The results indicate that the proposed method outperforms (highly significant) all other methods except the Wavelet packet Shannon entropy + SVM, FRFE + multi-layer perceptron and FPCNN + DWT + PCA + BPNN methods. Although it is slightly significant as compared to the Wavelet packet Shannon entropy + SVM, FRFE + multi-layer perceptron and FPCNN + DWT + PCA + BPNN methods, it provides an improved performance in terms of other performance indices, as confirmed from above tables. Additionally, it is prone to the problems like overfitting and bias results due to the inclusion of the nested cross-validation technique.

4. Conclusion

In this investigation, the aim is to propose a novel NCVASR method for pathological brain classification. The GLCM is used to extract the spatial texture features from the brain MR images (training and test). The training feature vectors are exploited by the proposed NCVASR method to classify an unknown test brain MRI sample as normal or pathological. This study has shown that a subset having *N* significant training feature vectors is selected for accurate classification. The value of *N* is optimized by using the $m \times n$ nested cross-validation technique. The present study makes several noteworthy contributions: (a) avoids overfitting as well as underfitting problems, (b) avoids possible bias in the result and c) adaptively finds the optimal value of *N* whenever a new test image arrive. The relevance of the subset selection using nested cross-validation technique is clearly supported by the current findings. We have also compared the proposed NCVASR with the state-of-the-art methods for pathological brain classification in terms of *CA, SEN, SPE, CM* and *AUC*. Taken together, these results indicate that it is superior over all other methods. In addition, the statistical analysis suggests that there is a significant difference between the suggested NCVASR method and the other methods. The study has gone some way towards enhancing our understanding of parameter selection while avoiding some serious problems such as overfitting. The authors would like to reiterate the fact that the idea of cross-validation in sparse representation based classifier is very encouraging. A future study investigating its performance to different medical image databases would be very interesting.

References

Ain, Q., Jaffar, M. A., & Choi, T.-S. (2014). Fuzzy anisotropic diffusion based segmentation and texture based ensemble classification of brain tumor. *Applied Soft Computing, 21*, 330–340.

Atlas, B. (2017). Harvard whole brain Atlas. http://www.med.harvard.edu/aanlib/home.html.

Brainsuit (2017). Brainsuit toolbox. http://brainsuite.org.

Chakraborty, J., Midya, A., & Rabidas, R. (2018). Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. *Expert Systems with Applications, 99*, 168–179.

Chaplot, S., Patnaik, L., & Jagannathan, N. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control, 1*(1), 86–92.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*(Jan), 1–30.

Dora, L., Agrawal, S., & Panda, R. (2017a). Gauss-newton representation based algorithm for magnetic resonance brain image classification. In *Proceedings of the international conference on intelligent systems design and applications* (pp. 294–304). Springer.

Dora, L., Agrawal, S., Panda, R., & Abraham, A. (2017b). Optimal breast cancer classification using gauss–newton representation based algorithm. *Expert Systems with Applications, 85*, 134–145.

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.

El-Dahshan, E.-S. A., Mohsen, H. M., Revett, K., & Salem, A.-B. M. (2014). Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Systems with Applications, 41*(11), 5526–5545.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. Academic Press.

Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence*. Prentice-Hall.

Jen, C.-C., & Yu, S.-S. (2015). Automatic detection of abnormal mammograms in mammographic images. *Expert Systems with Applications, 42*(6), 3048–3055.

Jothi, G., et al. (2016). Hybrid tolerance rough set–firefly based supervised feature selection for mri brain tumor image classification. *Applied Soft Computing, 46*, 639–651.

Katzir, Z., & Elovici, Y. (2018). Quantifying the resilience of machine learning classifiers used for cyber security. *Expert Systems with Applications, 92*, 419–429.

Kharrat, A., Benamrane, N., Messaoud, M. B., & Abid, M. (2009). Detection of brain tumor in medical images. In *Proceedings of the 3rd international conference on signals, circuits and systems (SCS), 2009* (pp. 1–6). IEEE.

More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv:1608.06048

Nabizadeh, N., & Kubat, M. (2015). Brain tumors detection and segmentation in mr images: Gabor wavelet vs. statistical features. *Computers & Electrical Engineering, 45*, 286–301.

Ortuño, F., & Rojas, I. (2015). Bioinformatics and biomedical engineeringProceedings of the third international conference, IWBBIO, Granada, Spain, April 15–17 (vol. 9044). Springer.

Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research, 3*(Mar), 1371–1382.

Shihabudheen, K., Mahesh, M., & Pillai, G. (2018). Particle swarm optimization based extreme learning Neuro-Fuzzy system for regression and classification. *Expert Systems with Applications, 92*, 474–484.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., & Levy, S. (2004). A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics, 21*(5), 631–643.

Valckx, F. M., & Thijssen, J. M. (1997). Characterization of echographic image texture by cooccurrence matrix parameters. *Ultrasound in Medicine & Biology, 23*(4), 559–571.

Wang, H., & Fei, B. (2009). A modified fuzzy c-means classification method using a multiscale diffusion filtering scheme. *Medical Image Analysis, 13*(2), 193–202.

Wang, S., Zhang, Y., Yang, X., Sun, P., Dong, Z., Liu, A., & Yuan, T.-F. (2015). Pathological brain detection by a novel image feature fractional fourier entropy. *Entropy, 17*(12), 8278–8296.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(2), 210–227.

Xu, L., Wu, X., Chen, K., & Yao, L. (2015). Multi-modality sparse representation-based classification for alzheimer's disease and mild cognitive impairment. *Computer Methods and Programs in Biomedicine, 122*(2), 182–190.

Xu, Y., Zhang, D., Yang, J., & Yang, J.-Y. (2011). A two-phase test sample sparse representation method for use with face recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 21*(9), 1255–1262.

Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI tex-

ture and shape in a machine learning scheme. *Magnetic Resonance in Medicine, 62*(6), 1609–1618.
Zhang, Y., Sun, Y., Phillips, P., Liu, G., Zhou, X., & Wang, S. (2016). A multilayer perceptron based smart pathological brain detection system by fractional fourier entropy. *Journal of Medical Systems, 40*(7), 173.
Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., & Cichocki, A. (2018). Multi-kernel extreme learning machine for eeg classification in brain-computer interfaces. *Expert Systems with Applications, 96*, 302–310.
Zhang, Y., & Wu, L. (2012). An mr brain images classifier via principal component analysis and kernel support vector machine. *Progress In Electromagnetics Research, 130*, 369–388.
Zhang, Y.-D., Wang, S.-H., Yang, X.-J., Dong, Z.-C., Liu, G., Phillips, P., & Yuan, T.-F. (2015). Pathological brain detection in MRI scanning by wavelet packet tsallis entropy and fuzzy support vector machine. *SpringerPlus, 4*(1), 716.