Mathematical Sciences Letters
An International Journal

http://dx.doi.org/10.12785/msl/040215

Proposing a Features Extraction based on Classifier Selection to Face Recognition and Image Processing

*Sajad Parvin and Zahra Rezaei**

Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad, Iran

Received: 17 Jul. 2014, Revised: 24 Dec. 2014, Accepted: 8 Jan. 2015
Published online: 1 May 2015

Abstract: feature is a Gabor response of image with a different tuple (x, k, θ). We use a pre-processing whereby we can use a fixed point x for all images without missing of the generality. Eight orientation frequency values are selected for θ parameter. Five spatial frequency values are also selected for domain of k parameter. So we reach a k×θ Gabor-wavelet based feature space. Also to get rid of the curse of dimensionality problem again without loss of the generality we omit the versatility of values in the k parameter. Indeed we compute the similarity of a pair faces in two images by averaging their similarity defined for all possible values of k parameter for a given θ parameter. Then considering the similarities of as a matrix we produce eight matrices for eight different θ parameters. By considering each of these matrices as a classifier we finally use an ensmble mechanism to aggregate them into final classification. We turn to a weighted majority average voting classifier ensemble to handle the problem. We show that the proposed mechanism works well in an employees' repository of our laboratory.

Keywords: Classifier Ensemble, Gabor Wavelet Features, Face Recognition, Image Processing.

1. Introduction

Feature extraction for object representation performs an important role in automatic object detection systems. Previous methods have used many representations for object feature extraction, such as raw pixel intensities [16, 24, 25, 26], [17] and [21], rectangle features [18], [19] and [20], and local binary pattern [22]. Gabor-wavelet based feature extraction methods have been successfully employed in many computer-vision problems, such as fingerprint enhancement and texture segmentation [10, 11]. Also similar to the human visual system, Gabor-wavelet features represent the characteristics of the spatial localities and the orientation selectivity, and are locally optimal in the space and frequency domains [12]. Therefore, Gabor-wavelet features are the proper choice for image decomposition and representation when the goal is to derive local and discriminating features [13, 27, 28].

Gabor filter can capture salient visual properties such as the spatial localization, the orientation selectivity, and the spatial frequency characteristics. The Gabor responses describe a small patch of gray values in an image around a given pixel. It is obtained based on a wavelet transformation. To obtain a Gabor response form a typical image 3 inputs must be chosen: (a) the pixel around that the Gabor response is to be extracted denoted by x, (b) spatial frequency value denoted by k and (c) orientation frequency value denoted by θ. We can call each Gabor response to a tuple (x, k, θ) in a typical image a Gabor wavelet-feature.

Usage of recognition systems has found many applications in almost all fields. However, Most of classification algorithms have obtained good performance for specific problems; they have not enough robustness and generality for other problems. Ensemble of multiple classifiers can be considered as a general solution method for pattern recognition problems. It has been shown that combination of classifiers can usually operate better than a single classifier provided that its components are independent or they have diverse outputs. It is shown that the necessary diversity of an ensemble can be achieved by manipulating of dataset features. Parvin et

* Corresponding author e-mail: zamio.rez@gmail.com

al. have proposed some methods of creating the necessary diversity for an ensemble success [14] and [15].

As it is said combinational classifiers are so versatile in the fields of artificial intelligence. It has been proved that a single classifier is not able to learn all the problems because of three reasons:

Z. Rezaei: Using Different Extracted Gabor-Wavelet Features to Identify Images

1. Problem may inherently be multifunctional.

2. From other side, it is possible that a problem is well-defined for a base classifier which its recognition is very hard problem.

3. And finally, because of the instability of some base classifiers like Artificial Neural Networks, Decision Trees, and Bayesian Classifier and so on, the usage of combinational classifiers can be inevitable.

Applications of combinational classifiers to improve the performance of classification have had significant interest in image processing recently. Singh and Singh [8, 29, 30] have proposed a new knowledge-based predictive approach based on estimating the Mahalanobis distance between the test sample and the corresponding probability distribution function from training data that selectively triggers classifiers. They also have shown the superior performance of their method over the traditional challenging methods empirically.

There are several methods to combine a number of classifiers in the field of image processing. Some of the most important are sum/mean and product methods, ordering (like max or min) methods and voting methods. There is a good coverage over their comparisons and evaluations in the [1], [2], [3] and [4]. In [5] and [6] it is shown that the product method can be considered as the best approach when the classifiers have correlation in their outputs. Also it is proved that in the case of outliers, the rank methods are the best choice [4, 35]. For a more detailed study of combining classifiers, the reader is referred to [7, 31, 32].

This paper aims at producing an ensemble-based classification of face recognition by use of Gabor-wavelet features with different orientation and spatial frequencies. The face images are first gave to the Gabor feature extractor with different orientation and spatial frequencies, and then the features of all trainset union with the test data are compared with each other in each orientation frequency. This results in a similarity matrix per each orientation frequency. The similarity matrices are finally combined to vote to which training image the test data belongs.

2. Weighted Voting Classifier Ensemble

An ensemble learns classification better than a single classifier because different single classifiers with the different characteristics and methodologies can complement each other and cover their internal weaknesses. If a number of different classifiers vote as an ensemble, the overall error rate will decrease significantly rather using each of them individually.

One of the oldest and the most common policy in classifier ensembles is majority voting. In this approach as it is obvious, each classifier of the ensemble is tested for an input instance and the output of each classifier is considered as its vote. The class is the winner which the most of the classifiers vote for it. The correct class is the one most which is often chosen by different classifiers. If all the classifiers indicate different classes, then the one with the highest overall outputs is selected to be the correct class.

Let us assume that E is the ensemble of n classifiers $\{e_1, e_2, e_3 \dots e_n\}$. Also assume that there are m classes in the case. Next, assume applying the ensemble over data sample d results in a binary D matrix like equation 1.

$$D = \begin{bmatrix} d_{1\ 1} & d_{1\ 2} & . & d_{1\ n} \\ . & . & . & . \\ d_{m-1\ 1} & d_{m-1\ 2} & . & d_{m-1\ n} \\ d_{m\ 1} & d_{m\ 2} & . & d_{m\ n} \end{bmatrix} \qquad 1$$

where $d_{i,j}$ is equal to one if the classifier j votes that data sample belongs to class i. Otherwise it is equal to zero. Now the ensemble decides the data sample to belong class b according to equation 2.

$$b = \arg\max_i \left| \sum_{j=1}^{n} d_{i\ j} \right| \qquad 2$$

Another method to combine a number of classifiers which employs d_{ij} as confidence of classifier j for belonging the test data sample to class i is called majority average voting. The majority average voting uses equation 2 as majority voting.

Weighted majority vote is another approach of voting; in this method members' votes have different worths. Unlike the previous versions of voting this is not like democracy. For example if a classifier has 99% recognition ratio, it is more worthy to use its vote with a more effect than the vote of another classifier with 80% accuracy rate. Therefore in weighted majority vote approach, every vote is multiplied by its worth. Kuncheva [7, 33, 34, 35] has shown that this worth can optimally be a function of accuracy.

To sum up assume that the classifiers existing in the ensemble E have accuracies $\{p_1, p_2, p_3 \ldots p_n\}$ respectively. According to Kuncheva [7] the worth of them are $\{w_1, w_2, w_3 \ldots w_n\}$ respectively where

$$w_i = \log \frac{p_i}{1-p_i} \qquad 3$$

Weighted majority vote mechanism decides the data sample to belong class b according to equation 2.

$$b = \arg\max_i \left| \sum_{j=1}^{n} w_j * d_{i\ \ j} \right| \qquad 4$$

Similarly another method of combining which again employs d_{ij} as confidence of classifier j for belonging the test data sample to class i is called weighted majority average voting. Weighted majority average voting method uses equation 4 as weighted majority voting.

3. Overview of Proposed Method

Gabor filter can capture salient visual properties such as the spatial localization, the orientation selectivity, and the spatial frequency characteristics. The Gabor responses describe a small patch of gray values in an image around a given pixel. It is obtained based on a wavelet transformation. To obtain a Gabor response form a typical image 3 inputs must be chosen: (a) the pixel around that the Gabor response is to be extracted denoted by x, (b) spatial frequency value denoted by k and (c) orientation frequency value denoted by θ. We can call each Gabor response to a tuple (x, k, θ) in a typical image a Gabor wavelet-feature.

It has been proven that Gabor wavelet-feature based recognition methods are useful in many problems including face detection. It has been shown that these features can tackle the image recognition problem well. In image identification, while there is a number of human faces in a repository of employees, it is aimed to identify the face of an arrived employee is which one? So due to the ability of Gabor-wavelet feature in well-encoding and well-representing the characteristics of an image the application of Gabor-wavelet based features in the case of face identification is reasonable. Each image is an employee face in our benchmark. So we can use a specific spatial localization for all images without lacking generality of the problem. So a pre-processing phase is necessary to get rid of the high possible domain for the pixel x around that the Gabor response is to be extracted. As it is presented in Fig. 1, we first cut the images so as to all marginal non-facial pixel be removed in pre-processing phase. Then we rescale all modified images in a fixed size. After that we can select the middle point of all rescaled images as the pixel x around that the Gabor response is to be extracted.

4. Feature Extraction

Gabor filter can capture salient visual properties such as the spatial localization, the orientation selectivity, and the spatial frequency characteristics. The Gabor responses describe a small patch of gray values in an image $I(x)$ around a given pixel $x=(x,y)^T$. It is based on a wavelet transformation, given by the equation 5.

$$R_i(x) = \int I(x')\psi_i(x-x')dx' \qquad 5$$

which $\psi_i(x)$ is a convolution of image with a family of Gabor kernels like equation 6.

$$\psi_i(x) = \frac{\|k_i\|^2}{\sigma^2} e^{\frac{\|k_i\|^2 \|x^2\|}{2\sigma^2}} \left[e^{jk_i x} - e^{-\frac{\sigma^2}{2}} \right] \qquad 6$$

Where

$$k_i = \begin{pmatrix} k_{ix} \\ k_{iy} \end{pmatrix} = \begin{pmatrix} k_v \cos\theta_\mu \\ k_v \sin\theta_\mu \end{pmatrix} \qquad 7$$

Each $\Psi_i(x)$ is a plane wave characterized by the vector k_i enveloped by a Gaussian function, where σ is the standard deviation of this Gaussian. The center frequency of i^{th} filter is given by the characteristic wave vector k_i having a scale and orientation given by (k_v, θ_μ). Convolving the input image with a number of complex Gabor filters with 5 spatial frequencies ($v = 0,...4$) and 8 orientations ($\mu = 0,...7$) will capture the whole frequency spectrums, both amplitude and phase as illustrated in [9].

According to equation 5, each image I^q of face train dataset is mapped to *40* features $I^{q\prime}_{v,\mu}$, where $v\in\{0,...4\}$ and $\mu\in\{0,...7\}$. Test image H is also mapped to $H'_{v,\mu}$. Now we define the similarity vector $\boldsymbol{sim^f}$ whose *i*th element indicates the similarity between *i*th train image and the test image, H. Similarity between the train images I^q and the test image H in a fixed orientation frequency f is defined according equation 8.

$$sim_q^f(H)=\frac{1}{5}\sum_{v=0}^{4}\underset{x\in C}{mean}\left(\left|I_{v,f}^q(x)-H_{v,f}(x)\right|\right) \quad 8$$

where C is a 9×9 square in the middle of the image, e.g. for image with size 80×40, it is $\{36,...44\}\times\{16,...24\}$. Indeed we try to compute for a very fewer points in the middle of the images rather than all of them. So we have really 3240 ($9\times9\times8\times5$) features.

5. Employed Classification

Let assume that there exists n training images and one test image. Also assume that the training images are indexed as number one to n respectively and the test image indexed as number $n+1$. The goal is to understand to which training image the test image is similar. The Gabor-wavelet features of r_1 orientation frequency and five orientation frequency are first extracted from images number one to $n+1$. Then the similarities between each of train images and test image are evaluated according to equation 8, as discussed in the previous section.

It is obvious that in order to become these similarities comparable they must be normalized in such a way that the sum of the similarity vector of test image becomes unit. So they are normalized in range [0,1]. After calculating each of these similarities between each two training and test images, a similarity vector named $\boldsymbol{sim^{r1}}$ which is a vector with n elements, is obtained. It is important to note that the sim^{r1}_i means the similarity between images number i and test image.

As the reader can guess, the problem mentioned here, is an n class problem. $\boldsymbol{sim^{r1}}$ can be also served as a simple classifier C_{r1} which uses images number 1 to n as its train dataset. It acts very similar to 1-NN classifier where it assigns the index of the maximum value in the vector to class label of test image.

Considering $\boldsymbol{sim^{r1}}$, $r1\in\{0,...7\}$ there are eight classifiers to classify the test image. Now the majority-votes ensemble is employed to classify the test image. Assume that the accuracy of classifier C_{r1} is denoted by p_{r1}, the weight vector $\boldsymbol{w}$ can straightforwardly be calculated in the weighted-majority-votes ensemble.

6. Parameters of Classification

Parameter k_v is set to one of the values {0.2, 0.4, 0.6, 0.8, 1}. In the experiments, there exist 2×300 training images. Here there are 300 real classes, 2 images per each class denoted by TI_i and VI_i where $i\in\{1,..300\}$. Indeed one image of class i is denoted by TI_i and the other by VI_i. 300 fixed images i.e. TI_i, are selected as training dataset. Running the algorithm 300 times, each time one of VI_i is considered as test image and the other 299 images as validation dataset. In zth running of algorithm image VI_z is selected as test image and images VI_j where $j\in\{1,...300\}$-$\{VI_z\}$ are considered as validation dataset. Now we obtain 8 classifiers C_{r1}, $r1\in\{0,...7\}$ based on $\boldsymbol{sim^{r1}}$. To calculate the accuracies of C_{r1} the mentioned validation dataset is used as following. The similarities between each pairs of images denoted by TI_i and VI_i, where $i\in\{1,..300\}$ and $j\in\{1,...300\}$-$\{VI_z\}$ are evaluated employing equation 9.

$$SIMILARITY_{i,j}^{r1}=sim_i^{r1}(I^j) \quad 9$$

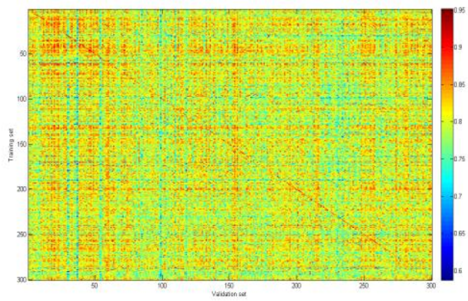


Figure1. Similarity matrix in frequency 5

To show the effectiveness of the similarity matrix, the Figure 2 shows the matrix SIMILARITY^5.

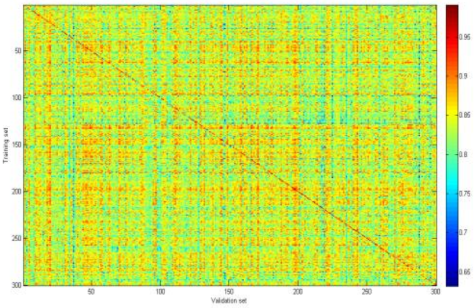


Figure2. Similarity matrix in frequency 10

Figure 3, Figure 4, Figure 5, Figure 6 also show the matrix $SIMILARITY^{10}$, $SIMILARITY^{15}$, $SIMILARITY^{20}$, $SIMILARITY^{27}$, $SIMILARITY^{36}$ and $SIMILARITY^{50}$ respectively.

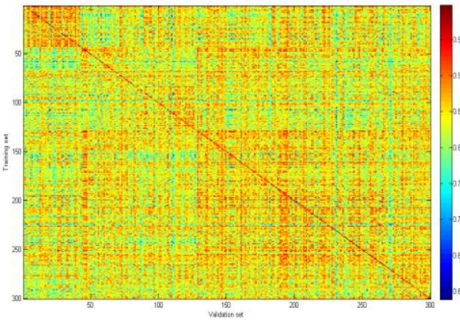


Figure3. Similarity matrix in frequency 20

As is obvious the best frequency is 27. As we increase the frequency, after the 27 the quality of classification decreases.

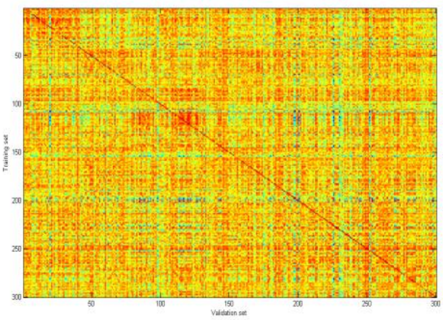
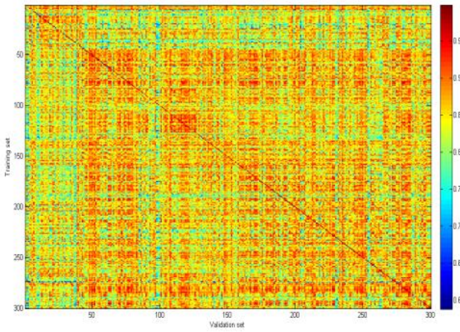


Figure4. Similarity matrix in frequency 27



Figur5. Similarity matrix in frequency 36

It is obvious that these similarities must also again be normalized in order to become them comparable. So they are again normalized in range [0,1] as mentioned before. After calculating each of these similarities between each two of training datasets, a similarity matrix named $SIMILARITY^{r1}$ which is an $n \times n$ matrix, is obtained. It is important to note that the $SIMILARITY^{r1}_{i,j}$ means the similarity between image number i of training dataset and image number j of validation dataset and the VI_zth column of that matrix is invalid.

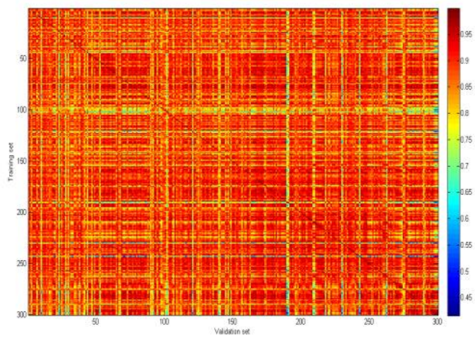


Figure6. Similarity matrix in frequency 50

Now the accuracy of classifier C_{r1}, on the training data, is the number of training data that correctly assigned to its correct class, divided to n. In other words, the number of the columns which its maximum value is over matrix diagonal, divided to n can be considered as the accuracy of this classifier as stated in equation 10. Although it is obvious that diagonal elements of this matrix must be the largest in their columns, it is not true in many cases.

$$p_{r1} = \frac{1}{299} \sum_{j \in \{1,\dots 300\} - \{VI_z\}} isequal(\arg\max_{i}(SIMILARITY^{r1}_{i,j}), j)$$

where $isequal(x,y)$ is defined as equation 11.

$$isequal(x,y)=\begin{cases}1 & x==y\\ 0 & otherwise\end{cases} \quad (11)$$

7. Experimental Study

Experimental results are reported over 300 pairs of images. Each pair of images belongs to an employee (personnel) of our laboratory. All the images have the same resolution. All of them are first equalized using equalizing their histograms.

Live-one-out technique is used to test ensemble classifier over these images. Also features of 5 different scales and 8 orientations are extracted. So, there are forty similarity matrices. 599 images, except in weighted majority voting, are used as training set because there is no longer need to validation set. It is worthy to mention that the best classifier using only one of the similarity matrix, has just 76.63% recognition ratio. While recognition ratio of classifier mentioned above has 90.17% recognition ratio with majority voting, by use of the average voting as final results the 89.32% recognition ratio is achieved. But the combinational proposed approach has 92.67% recognition ratio. The Table 1 summarizes the results.

Table1. Face recognition ratios of different methods.

Best C^f	MV(C^f)	MAV(C^f)	WMAV
76.63	89.32	90.17	92.67

8. Conclusion and Discussion

In this paper, new face identification algorithm is proposed. We first extract a large number of different features from each employee face image. Each feature is a Gabor response of image with a different tuple *(*x, k, θ*)*. We use a pre-processing whereby we can use a fixed point x for all images without missing of the generality. Eight orientation frequency values are selected for θ parameter. Five spatial frequency values are also selected for domain of k parameter. So we reach a k×θ Gabor-wavelet based feature space. We compute a similarity matrix per different values θ parameter. By considering each of these matrices as a classifier we finally use an ensmble mechanism to aggregate them into final classification.

To validate the employed face identification algorithm we use live-one-out technique. We turn to a weighted majority average voting classifier ensemble to handle the problem. It is shown that the proposed mechanism works well in an employees' repository of laboratory containing 600 face images from 300 different individuals.

References

1. B. Minaei-Bidgoli, G. Kortemeyer and W.F. Punch, Optimizing Classification Ensembles via a Genetic Algorithm for a Web-based Educational System, (SSPR /SPR 2004), Lecture Notes in Computer Science (LNCS), Volume 3138, Springer-Verlag, ISBN: 3-540-22570-6, pp. 397-406, 2004.

2. Saberi., M. Vahidi , B. Minaei-Bidgoli, Learn to Detect Phishing Scams Using Learning and Ensemble Methods, IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Workshops (IAT 07), pp. 311-314, Silicon Valley, USA, November 2-5, 2007.

3. T.G. Dietterich, Ensemble learning, in The Handbook of Brain Theory and Neural Networks, 2nd edition, M.A. Arbib, Ed. Cambridge, MA: MIT Press, 2002.

4. H. Parvin, H. Alinejad-Rokny and S. Parvin, Divide and Conquer Classification, Australian Journal of Basic & Applied Sciences, 5(12), 2446-2452 (2011).

5. H. Parvin, B. Minaei-Bidgoli and H. Alinejad-Rokny, A New Imbalanced Learning and Dictions Tree Method for Breast Cancer Diagnosis, Journal of Bionanoscience, 7(6), 673-678 (2013).

6. H. Parvin, H. Alinejad-Rokny and M. Asadi, An Ensemble Based Approach for Feature Selection, Journal of Applied Sciences Research, 7(9), 33-43 (2011).

7. S. Gunter and H. Bunke, Creation of classifier ensembles for handwritten word recognition using feature selection algorithms, IWFHR 2002 on January 15, 2002.

8. B. Minaei-Bidgoli, G. Kortemeyer, W. F. Punch, Mining Feature Importance: Applying Evolutionary Algorithms within a Web-Based Educational System, Proc. of the Int. Conf. on Cybernetics and Information Technologies, Systems and Applications, CITSA 2004.

9. L. I. Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, New York: Wiley, 2005.
10. L. Shapley and B. Grofman, Optimizing group judgmental accuracy in the presence of interdependencies, Public Choice, 43:329-343, 1984.
11. H. Parvin, H. Helmi, B. Minaie-Bidgoli, H. Alinejad-Rokny and H. Shirgahi, Linkage learning based on differences in local optimums of building blocks with one optima, International Journal of Physical Sciences, 6(14), 3419-3425 (2011).
12. H. Parvin, B. Minaei-Bidgoli, H. Alinejad-Rokny and S. Ghatei, An innovative combination of particle swarm optimization, learning automaton and great deluge algorithms for dynamic environments, International Journal of Physical Sciences, 6(22), 5121-5127 (2011).
13. H. Parvin, H. Alinejad-Rokny and S. Parvin, A Classifier Ensemble of Binary Classifier Ensembles, International Journal of Learning Management Systems, 1(2), 37-47 (2013).
14. F. Roli and J. Kittler, editors. Proc. 2nd Int. Workshop on Multiple Classifier Systems (MCS 2001), Vol. 2096 of Lecture Notes in Computer Science LNCS Springer-Verlag, Cambridge, UK, 2001.
15. F. Roli and J. Kittler, editors. Proc. 3rd Int. Workshop on Multiple Classifier Systems (MCS 2002), Vol. 2364 of Lecture Notes in Computer Science LNCS Springer-Verlag, Cagliari, Italy, 2002.
16. L. Lam. Classifier combinations: implementations and theoretical issues. In J. Kittler and F. Roli, editors, Multiple Classifier Systems, Vol. 1857 of Lecture Notes in Computer Science, Cagliari, Italy, 2000, Springer, pp. 78-86.
17. T.G. Dietrich, Machine-learning research: four current direction, AI Magazine, 18, 4, winter 1997, 97-135.
18. H. Parvin, B. Minaei-Bidgoli, H. Alinejad-Rokny and W.F. Punch, Data weighing mechanisms for clustering ensembles, Computers & Electrical Engineering, 39(5), 1433-1450 (2013).
19. H. Parvin, H. Alinejad-Rokny, B. Minaei-Bidgoli and S. Parvin, A new classifier ensemble methodology based on subspace learning, Journal of Experimental & Theoretical Artificial Intelligence, 25(2), 227-250 (2013).
20. H. Parvin, H. Alinejad-Rokny, N. Seyedaghaee and S. Parvin, A Heuristic Scalable Classifier Ensemble of Binary Classifier Ensembles, Journal of Bioinformatics and Intelligent Control, 1(2), 163-170 (2013).
21. A.K. Jain, R.P.W. Duin, J. Mao, Satanical pattern recognition: a review, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI-22, 1, January 2000, 4-37.
22. R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed. John Wiley & Sons, NY, 2001.
23. J. C. Sharkey, editor, Combining Artificial Neural Nets. Ensemble and Modular Multi-Net Systems, Springer-Verlag, London, 1999.
24. L. K. Hansen, P. Salamon, Neural network ensembles. IEEE Transaction on Pattern Analysis and Machine Intelligence, 12(10):993-1001, 1990.
25. Krogh, J. Vedelsdy, Neural Network Ensembles Cross Validation, and Active Learning, In: G. Tesauro, D. Touretzky, T. Leen (Eds.), Advances in Neural Information Processing Systems, Volume 7. MIT Press, Cambridge, MA, p.231-238, 1995.
26. L. Breiman, Bagging predictors. Machine Learning, 24(2):123-140, 1996.
27. M.H. Fouladgar, B. Minaei-Bidgoli, H. Parvin and H. Alinejad-Rokny, Extension in The Case of Arrays in Daikon like Tools, Advanced Engineering Technology and Application, 2(1), 5-10 (2013).
28. H. Parvin, M. MirnabiBaboli and H. Alinejad-Rokny, Proposing a Classifier Ensemble Framework Based on Classifier Selection and Decision Tree, Engineering Applications of Artificial Intelligence, 37, 34-42 (2015).
29. R.E. Schapire, The strength of weak learn ability, Machine Learning, 5(2):1971-227, 1990.

30. P. Melville, R. Mooney, Constructing Diverse Classifier Ensembles Using Artificial Training Examples, Proc. of the IJCAI-2003, Acapulco, Mexico, p.505-510, 2003.
31. B. E. Rosen, Ensemble learning using decorrelated neural network. Connection Science, 8(3-4):373-384, 1996.
32. Y. Liu, X. Yao, Evolutionary ensembles with negative correlation learning, IEEE Trans. Evolutionary Computation, 4(4):380-387, 2000.
33. D. Opitz, J. Shavlik, Actively searching for an effective neural network ensemble, Connection Science, 8(3-4):337-353, 1996.
34. I. Jamnejad, H. Heidarzadegan, H. Parvin and H. Alinejad-Rokny, Localizing Program Bugs Based on Program Invariant, International Journal of Computing and Digital Systems, 3(2), 141-150 (2014).
35. Lazarevic, Z. Obradovic, Effective pruning of neural network classifier ensembles. Proc. International Joint Conference on Neural Networks, 2:796-801, 2001.
36. H. D. Navone, P. F. Verdes, P. M. Granitto, H. A. Ceccatto, Selecting Diverse Members of Neural Network Ensembles, Proc. 16th Brazilian Symposium on Neural Networks, p.255-260, 2000.
37. Z. H. Zhou, J. X. Wu, Y. Jiang, S. F. Chen, Genetic algorithm based selective neural network ensemble, Proc. 17th International Joint Conference on Artificial Intelligence, 2:797-802, 2001.
38. M.H. Fouladgar, B. Minaei-Bidgoli, H. Parvin and H. Alinejad-Rokny, Extension in The Case of Arrays in Daikon like Tools, Advanced Engineering Technology and Application, 2(1), 5-10 (2013).
39. H. Parvin, M. MirnabiBaboli and H. Alinejad-Rokny, Proposing a Classifier Ensemble Framework Based on Classifier Selection and Decision Tree, Engineering Applications of Artificial Intelligence, 37, 34-42 (2015).
40. H. Parvin, H. Alinejad-Rokny, S. Parvin and H. Shirgahi, A new Conditional Invariant Detection Dramework (CIDF), Scientific Research and Essays, 8(6), 265-273 (2013).
41. Q. Fu, S. X. Hu, S. Y. Zhao, A PSO-based approach for neural network ensemble, Journal of Zhejiang University (Engineering Science), 38(12):1596-1600, 2004, (in Chinese).
42. Y. Freund, R.E. Schapire, A decision-theoretic generalization of online learning and an application to boosting, in Proceedings of the 2nd European Conference on Computational Learning Theory, Barcelona, Spain, pp.23–37, 1995.
43. B. Efron, R. Tibshirani, An Introduction to the Bootstrap, New York: Chapman & Hall, 1993.
44. V. Dobra, Scalable Classification And Regression Tree Construction, Ph.D. Dissertation, Cornell University, Ithaca, NY, 2003.
45. T. G. Dietterich, Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, Multiple Classifier Systems, volume 1857 of Lecture Notes in Computer Science, Springer, pp. 1–15, Cagliari, Italy, 2000.
46. F. Roli, G. Giacinto, G. Vernazza. Methods for designing multiple classifier systems. In J. Kittler and F. Roli, editors, Proc. 2nd International Workshop on Multiple Classifier Systems, Vol. 2096 of Lecture Notes in Computer Science, Springer- Verlag, pp. 78–87, Cambridge, UK, 2001.
47. S. Dudoit, J. Fridlyand, Bagging to improve the accuracy of a clustering procedure. Bioinformatics, 19 (9), pp. 1090-1099, 2003.
48. L.I. Kuncheva, L.C. Jain, Designing Classifier Fusion Systems by Genetic Algorithms. IEEE Transaction on Evolutionary Computation, Vol. 33, 351-373, 2000.
49. Strehl, J. Ghosh, Cluster ensembles a knowledge reuse framework for combining multiple partitions. Journal on Machine Learning Research, pp. 583-617, 2002.