

Application of Imperialist Competitive Algorithm for Feature Selection: A Case Study on Bulk Rice Classification

S. J. Mousavi Rad
Department of Computer
Engineering
University of Kurdistan
Sanandaj, Iran

F. Akhlaghian Tab
Department of Computer
Engineering
University of Kurdistan
Sanandaj, Iran

K. Mollazade
Department of Mech. Eng. of
Agri. Machinery
University of Kurdistan
Sanandaj, Iran

ABSTRACT

Feature selection plays an important role in pattern recognition. The better selection of a feature set usually results the better performance in a classification problem. This work tries to select the best feature set for classification of rice varieties based on image of bulk samples using imperialist competition algorithm. Imperialist competition algorithm is a new evolutionary optimization method that is inspired by imperialist competition. Results showed the feature set selected by the imperialist competition algorithm provide the better classification performance compared to that obtained by genetic algorithm technique

General Terms

Pattern recognition, data mining.

Keywords

Imperialist competition algorithm, feature selection, bulk rice classification, support vector machine

1. INTRODUCTION

Feature selection is the problem of selection a subset of features without reducing the accuracy of representing the original set of features. Algorithms for feature selection are divided into two categories: wrapper that uses the learning algorithm itself to evaluate the usefulness of features and filters that evaluates features according to statistical characteristics of the data. Among too many methods, which are proposed for feature selection, population based optimization algorithms such as particle swarm optimization(SOM)[1], genetic algorithm[2], and ant colonies optimization(ACO) [3] have attracted a lot of attention. These methods attempt to achieve better solution by application of knowledge from previous iteration. In this paper, a new wrapper feature selection technique is presented using imperialist competition algorithm to classify bulk samples of rice varieties. Imperialist competition is a new evolutionary optimization method that is inspired by imperialist competition[4]. All countries are divided into two types: imperialist states and colonies. Imperialist competition is the main part of this algorithm.

In the current grain-handling systems, grain type and quality are assessed by visual inspection. This evaluation process is, however, tedious and time consuming. The decision-making capabilities of a grain inspector can be seriously affected by his/her physical condition such as fatigue and eyesight, mental state caused by biases and work pressure, and working

condition such as improper lighting condition, etc. Hence, this needs to the automation of the process by developing an imaging system that should acquire, rectify, and analyze the rice grain images.

During the last decades, several studies have been carried out related to the application of machine vision for quality evolution of agro-food produce, Liu et al., [5] proposed a method of rice identification based on neural network and color and shape features with accuracy of 88.3%. In [6] researchers presented an algorithm for classifying five different varieties of rice using the color and texture features from individual rice kernels. Verma [7] developed a neural network with accuracy ranging from 90 to 95% to classify three varieties of Indian rice based on a feature set containing six morphological features.

The current paper is the first attempt to apply an imperialist competition based algorithm for feature selection. The result of feature selection process is then used for classification of rice varieties.

2. MATERIAL AND METHODS

2.1. Grain Samples

Seven rice varieties namely, Fajr, Hashemi, Daneboland, Gerde, Basmati, Domsiah and Tailandi were taken up for classification. The bulk samples were prepared by pouring a kg of rice kernels into a large plastic bag and shaking it to mix the grain thoroughly. The rice kernel was then slowly poured into a dish until it was completely filled. The excess rice was gently shaken off the dish so that the top level of grain was almost horizontal and matched up to the rim of the dish. This process was repeated 300 times for each rice variety. Thus, a total of 2100 images of bulk samples were acquired. Figure 1 shows a sample of images acquired from the seven different varieties of rice.

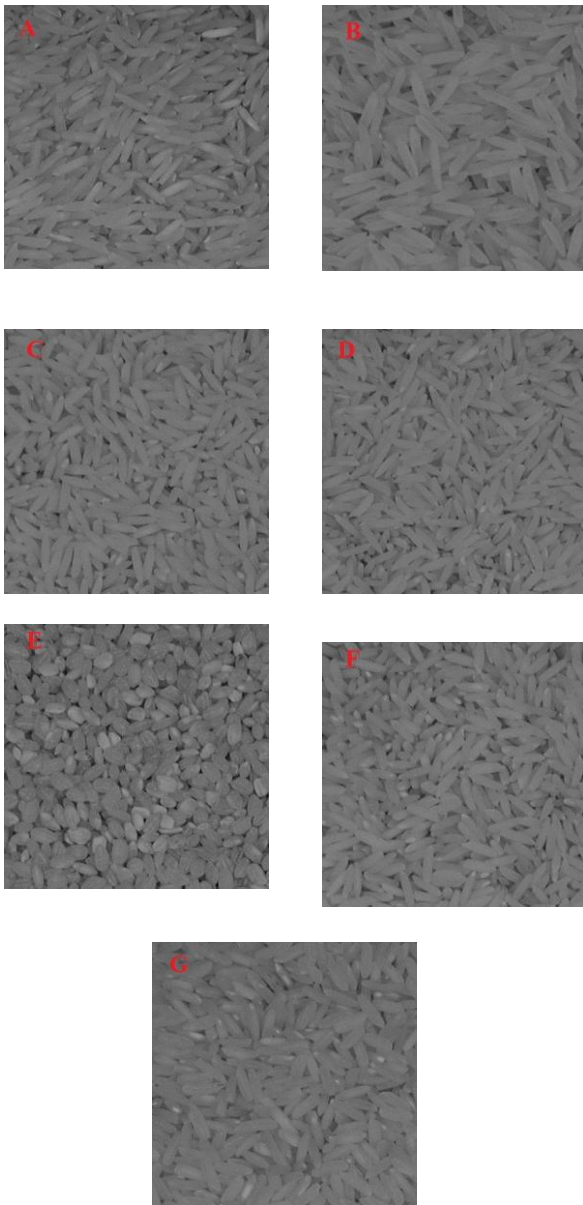


Figure 1: Images of seven different rice varieties (A: Fajr, B: Hashemi, C: Tailandi, D: Gerde, E: Domsiah, F: Daneboland, and G: Basmati)

2.2. Imaging Device

A CCD color camera (model No.C-2000 Olympus) was used to acquire the images. Camera was mounted on a black box which was equipped with a uniform lighting system. The distance between the camera and bulk samples was 20cm.

2.3. Texture Feature Extraction

Texture is one of the most important defining characteristics of images. It is characterized by the spatial distribution of gray levels in a neighborhood. Geometrical, statistic, model-based and signal processing are the common texture based methods for feature extraction. Statistical methods analyze the spatial distribution of gray values, by computing local features

at each point in the image and deriving a set of statistics from the distributions of local features. Geometric methods consider texture to be composed of texture primitives and rule governing their spatial organization by considering texture to be composed of texture primitives. Model based methods hypothesize the underlying texture process, constructing a parametric generative model, which could have created the observed intensity distribution. Signal processing methods analyze the frequency content of image. For food processing; the most widely used approach is statistical method like co-occurrence matrix methods[8].

2.3.1 Gray Level Co Occurrence Matrix (GLCM) Method

Gray level co-occurrence matrix is one of the mostly used statistical texture analysis methods in which texture feature is extracted by some statistical approaches from co-occurrence matrix , $p(k,l)$ [9]. The use of the co-occurrence matrix is based on the hypotheses that the same gray level configuration is repeated in texture. This pattern will vary more by fine textures than by coarse textures. The co-occurrence matrix $p_{\theta,d}(i, j)$ counts the co-occurrence of pixels with gray values i and j at a given distance d and a given direction θ . (Figure. 2) The direction θ can be selected from four different values of 0, 45, 90 and 135 deg while distance depends on the resolution of texture.

After the construction of the co-occurrence matrix, the matrix is normalized function:

$$p(k,l) = \frac{p(k,l)}{R}$$

where R is the normalized function, which is set as the sum of the matrix.

Using the normalized GLCM, the following textural features were extracted

- Maximum Probability

$$\max(p_{ij})$$

- Correlation: this shows how a pixel is correlated to its neighbor over the entire image.

$$\sum_{i=1}^k \sum_{j=1}^k (i-j)^2 p_{ij}$$

- Contrast: this measures contrast between a pixel and its neighbor

$$\sum_{i=1}^k \sum_{j=1}^k (i-j)^2 p_{ij}$$

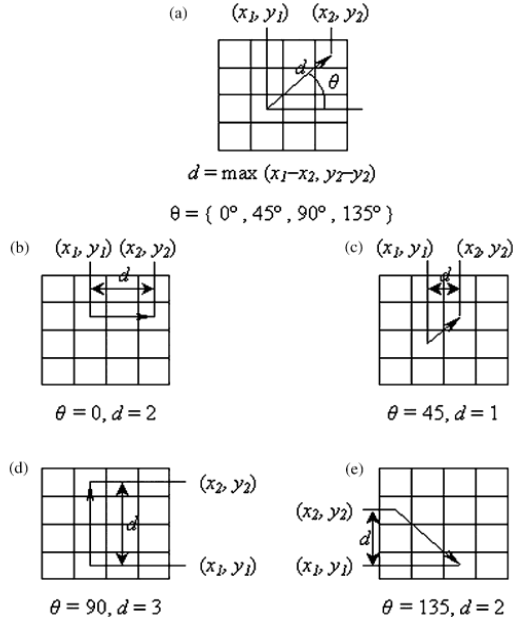


Figure 2. The direction θ of pixel pairs and the distance d between the pixel pairs used to construct the gray level co-occurrence matrix. (a) illustration of direction θ and distance d in image of pixel pairs (x_1, y_1) and (x_2, y_2) ; (b)-(e) four examples at direction $0, 45, 90$, and 135° , respectively[8].

- Uniformity(Energy)

$$\sum_{i=1}^k \sum_{j=1}^k P_{ij}^2$$

- Entropy: this measures randomness of a GLCM element

$$-\sum_{i=1}^k \sum_{j=1}^k P_{ij} \log_2 P_{ij}$$

- Homogeneity: this measures the closeness of the distribution of elements in the GLCM

$$\sum_{i=1}^k \sum_{j=1}^k \frac{P_{ij}}{1+|i-j|}$$

- Dissimilarity

$$\sum_{i=0}^k \sum_{j=0}^k P_{ij} |i-j|$$

- Mean

$$imean = \mu_i \sum_{i=0}^k \sum_{j=0}^k i(P_{ij})$$

$$jmean = \mu_j \sum_{i=0}^k \sum_{j=0}^k j(P_{ij})$$

- Cluster shade (CS) and cluster prominence (CP): CS and CP are measure the skewness of the matrix, in other words the lack of symmetry. When CS and CP are high, the image is not symmetric. In addition, when CP is low, there is a peak in the co-occurrence matrix around the mean value. For the image, this means that there is a little variation in gray scales.

$$CS = \sum_{i=0}^k \sum_{j=0}^k ((i - \mu_i) + (j - \mu_j))^3 P_{ij}$$

$$CP = \sum_{i=0}^k \sum_{j=0}^k ((i - \mu_i) + (j - \mu_j))^4 P_{ij}$$

So, 44 features (11 features \times 4 orientations) were extracted. After completing the feature extraction step, outputs were normalized to the range of 0 to 1.

2.4. Feature Selection

As stated in the previous section, 44 features were totally extracted from the bulk samples of rice images. Of course, this set of features contains redundant, noisy or even irrelevant information for classification purposes. To optimize the number of features that contributed significantly to the classification, a new feature selection algorithm was selected using imperialist competition algorithm.

2.4.1. Brief Description of Imperialist Competition Algorithm

Imperialist competition algorithm is a new evolutionary optimization method that is inspired by imperialist competition[4]. Similar other evolutionary algorithms, it starts with an initial random population called countries. Some of the best countries are selected to be the imperialist states and the rest from the colonies of these imperialist. Imperialist competition among these empires forms the basis of the ICA. During this competition, weak empires collapse and powerful ones take possession of their colonies imperialist competition converge to a state in which there exists only one empire and colonies have the same cost function value as the imperialist.

After forming initial empires, colonies in each of imperialist start moving toward their relevant imperialist country. This movement is a simple model of assimilation policy which was pursued by some of the imperialist states. The total power of an empire depends on both the power of the imperialist country and the power of its colonies. This fact is modeled by defining the total power of an empire as the power of

imperialist country plus a percentage of mean power of its colonies. In imperialist competition, all empires try to take possession of the colonies of other empires and control them. Figure 3 shows a picture of the modeled imperialist competition.

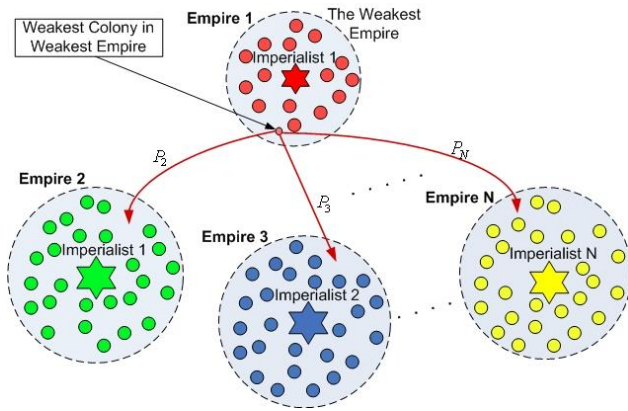


Figure 3. Imperialist competition: the more powerful an empire is, the more likely it will possess the weakest colony of weakest empire

Based on their total power, in this competition, each of empires will have a likelihood of taking possession of the mentioned colonies. In other words these colonies will not be possessed by the most powerful empires, but these empires will be more likely to possess them. When an empire loses all of its colonies, it is assumed to be collapsed.

After a while all the empires expect the most powerful one will collapse and all the colonies will be under the control of this unique empire. In this ideal new world colonies have the same position and same costs and they are controlled by an imperialist with the same position and cost as themselves, which means the algorithm converge to the best solution.

The pseudo code of imperialist competition algorithm is as follows:

- 1) Initialize the empires and their colonies position randomly.
- 2) Move the colonies toward their relevant imperialist (assimilation).
- 3) Randomly change the position of some colonies (revolution).
- 4) If there is a colony in an empire which has lower cost than the imperialist, exchange the position of that colony and the imperialist.
- 5) Compute the total cost of all empires (related to the power of both imperialist and its colonies).
- 6) Pick the weakest colony from the weakest empire and give it to the empire that has the most likelihood to possess it(imperialist competition).
- 7) Eliminate the powerless empires.
- 8) If stop conditions satisfied, if not go to 2.

2.4.2. Feature Selection using ICA Algorithm

In this section, a new feature selection method using ICA algorithm was presented. The steps of the proposed algorithm are described in details in the following subsections.

2.4.2.1. Initialization of Empires

In ICA Algorithm, each solution is shown by an array. In GA terminology, this array is called “chromosome” while in ICA, “country” plays the same role. In an N dimensional optimization problem, a country is a $1 \times N$ array. This array is defined by

$$Country = [p_1, p_2, \dots, p_j]$$

That p_i is the variable that to be optimized.

In proposed ICA, each country is a string of binary numbers. When value of variable in country is 1, then the feature is selected and when it is 0, the feature is not selected. Figure 4 demonstrates of the feature representation as a country from initial empires.

The cost value of a country is defined as the classification accuracy of the country. Many classifiers can be used for classification. For example, the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are two popular classifiers. Although the SVM is a powerful classification method, it takes too long to build a classifier. Since the KNN is simpler and quicker than the SVM, KNN was chosen to compute the cost value of a country. The cost function is determined as follows:

$$Cost = \frac{\text{Number of correctly classified samples}}{\text{Total Number of samples}}$$

The algorithm starts with N_{pop} initial countries randomly in the population size, and the best of them, N_{imp} , is chosen as the imperialist. The remaining population N_{col} will be the colonies each of which belongs to an empire. The colonies are randomly distributed among imperialists based on the imperialist’s power.

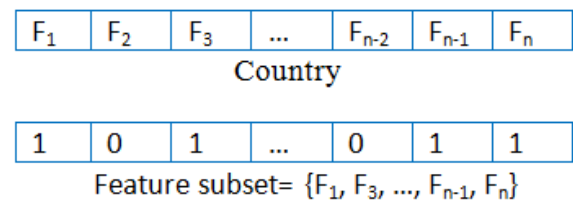


Figure 4. Example of feature encoding using ICA

To calculate the power of imperialists, first, the normalized cost of an imperialist is applied based on the following Equation

$$C_n = c_n - \max(c_i)$$

That c_n is the cost of n th imperialist and C_n is its normalized cost. Then, the power of each imperialist as follows:

$$P_n = \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i}$$

By attention to the imperialist's power, the colonies are distributed among the imperialists. In addition, the initial number of colonies of the n th empire will be:

$$N.C_n = \text{round}\{p_n \cdot N_{col}\}$$

where $N.C_n$ is the initial number of colonies of the n th empire. As shown in Figure 5, empires with the bigger power have a large number of colonies while, those with weaker power has less.

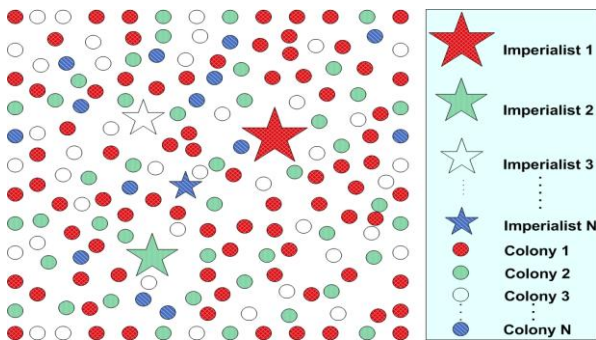


Figure 5. Generating of initial empires

2.4.2.2. Assimilation

The original version of imperialist competition algorithm operates on continuous problems. Imperialist countries start by improving their colonies. This fact has been modeled by moving all of the colonies toward the imperialist. Since the selection problems are discrete in nature, in the current research, a new assimilating, which is suitable for the discrete problems, is presented.

This operator is as below:

For each imperialist and their colonies

- 1) Create a binary string and assign a random generated binary (0-1) to each cell
- 2) Copy the cells from the imperialist correspond to location of the '1's in the binary string to the same positions in the colonies.

2.4.2.3. Revolution

This mechanism is similar to mutation process in genetic algorithm. The revolution increases the exploration of the algorithm and prevents the early convergence of countries to local minimum. In each iteration, for every colony a random number which is varying between 0 and one is generated. Then this value is compared with probability of revolution rate. If random number is lower than probability revolution rate, the revolution is performed. The new colony will replace with previous colony while its cost is improved.

2.4.2.4. Exchanging Positions of the imperialists and Their Colonies

After assimilation for all colonies and revolution for a percentage of them, a colony might reach a position with lower cost than the imperialist. In this case, the imperialist and the colony change their positions. Algorithm will continue by the imperialist in a new position and colonies start moving toward this position.

2.4.2.5. Total Power of an Empire

The total power of an empire is mainly contributed by the power of imperialist country. It is clear that the power of an empire includes the imperialist power and their colonies. Moreover, the power of imperialist has main effect on the total power of an empire while colonies power has lower impact. This fact is modeled by defining the total cost of an empire:

$$TC_n = \text{cost}\{\text{imperialist}_n\}C + \zeta \text{mean}\{\text{cost}(\text{colonies of empire}_n)\}$$

where TC_n is the total cost of the n th empire and ζ is a positive number which is considered to be less than 1. A small value for ζ causes the total power of the empire to be determined by just the imperialist and increasing it will increase the role of the colonies in determining the total power of an empire.

2.4.2.6. Imperialist Competition

All empires try to take possession of colonies of other empires and control them. To apply this concept, the imperialist competition is modeled by picking the weakest colony of weakest empires and making a competition among all empires to possess this colony.

To start the competition, after selecting a colony, the possession probability of each empire is found. The normalized total cost of an empire is simply obtained by

$$N.T.C_n = T.C_n - \max\{T.C_i\}$$

Where $T.C_n$ and $N.T.C_n$ are the total cost and normalized total cost of n th empires respectively.

By having the normalized total cost, the possession probability of each empire, P_{emp_n} is calculated as below:

$$P_{emp_n} = \frac{|N.T.C_n|}{\sum_{i=1}^{imp} N.C.T_i}$$

Roulette wheel method was used for assigning the mentioned colony to empires.

2.4.2.7. Eliminating the Powerless Empires

An empire will collapse when this empire loses all of its colonies. In this case, this imperialist is considered as a colony and it assigned to the other empires.

2.4.2.8. Stopping Criteria

Stopping criterion in proposed algorithm is to get the maximum decades.

2.5. Support Vector Machine

Support vector machines (SVM) are binary classifiers able to classify data samples in two disjoint classes by a hyperplane defined in a suitable space. The basic idea behind this technique is that two classes are linearly separable. Actually, there are more than one hyperplane for satisfying this condition, and one of them is chosen as classifier on the basis of the margin it creates between the two classes. Support vector machine uses support vectors for classification and discards other points. Support vectors are points that are on maximum margin hyperplane. SVMs can be used for classifying data that is not linearly separable. The data space is transformed using kernel function into a higher dimensional space where classes become linearly separable.

The classifier for the problem of binary classification is

$$f(x) = \text{sign}[w^T \cdot \phi(x) + b]$$

where the input vector x is mapped into a feature space by a kernel function $\phi(x)$, and w and b are the classifier parameters. Determining the classifier from the SVM theory is equivalent to solving the following optimization problem

$$\min_{w,b,\xi} \frac{1}{2} W^T \cdot W + C \sum_{i=1}^l \xi_i$$

Subject to

$$y_i [w^T \cdot \phi(x_i) + b] \geq 1 - \xi_i \quad \xi_i \geq 0, \quad i = 1, \dots, l$$

where ξ_i is a positive slack variable and $C > 0$ is a penalty parameter. This problem is a quadratic optimization problem that can be solved using Lagrange multipliers. Therefore, the hyper plane function can be written as:

$$f(x) = \text{sgn} \left(\sum_{i=1}^m y_i \alpha_i \cdot K(x, x_i) + b \right)$$

that

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Support vectors have non-zero Lagrange multipliers and m is the number of support vectors.

SVM has been developed to solve binary classification problems. But, real problems have more than two classes. For constructing a multi class SVM, two approaches are introduced. One approach is by combining several binary SVMs, and the other approach is implemented by direct multi class classification. The first approach includes “one against one” and “one against all” methodologies.[10]

In this research, different kernel functions were tested for classifying the rice bulk samples. Based on the results obtained, SVM with the RBF kernel and “one against one” methodology showed the best classification accuracy.

2.6. Performance Criteria

Three criteria were used to evaluate the performance of the proposed system, which are defined as follows:

$$\text{Precision} = \frac{fp}{tp + fp}$$

$$\text{Recall} = \frac{fp}{tp + fn}$$

$$F\text{measure} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

where tp is the number of true positive, fp is the number of the false positive, and fn is the number of false negative. $F\text{measure}$ is an evenly balanced precision and recall.

3. EXPERIMENTS

To show the performance of proposed ICA based feature selection algorithm, proposed algorithm was compared with genetic based feature selection. Various values were tested for the parameters of proposed algorithm. The results showed that the best performance is achieved when parameters are tuned as shown in Table 1.

An example of the process of ICA based feature selection to find the optimal solution is given in Figure 6. This Figure shows the average and best cost for each iteration. If exhaustive search is used to find the optimal feature subset in this dataset, there will be tens of millions of candidate subsets, which is impossible to execute. While the optimal solution is obtained after 200 iteration using ICA.

Table 1: Parameter setting for feature selection by ICA

Parameters	value
Number of population	200
Number of imperialist	20
Max Decades	400
Revolution rate	0.3
Zeta	0.1

Totally, 21 features were selected as the superior ones, The selected features are:

- Maximum probability at direction 0, 45, 90, and 135 deg.
- Homogeneity at direction 90 deg.
- Correlation at direction 0, 45, and 135 deg.
- Uniformity at direction 0, and 135 deg.

- Dissimilarity at direction 0, and 135 deg.
- Correlation at direction 0, 45, and 135 deg.
- Entropy at direction 45 deg.
- Imean at direction 0, 45, and 135 def.
- jmean at direction 135 deg.
- Cluster shade at direction 0, and 90 deg.
- Cluster performance at direction 0, and 135 deg.

Furthermore, the best features were selected by genetic algorithm and its classification performance for rice variety identification was compared to that of ICA. Results obtained for precision, recall, and F-measure shown in Table 2 indicate that ICA based algorithm creates higher F-measure value than that of genetic algorithm technique. Also, feature selection by both ICA and genetic algorithm techniques showed a considerable increase in classification accuracy compared to when all feature set is used for rice classification.

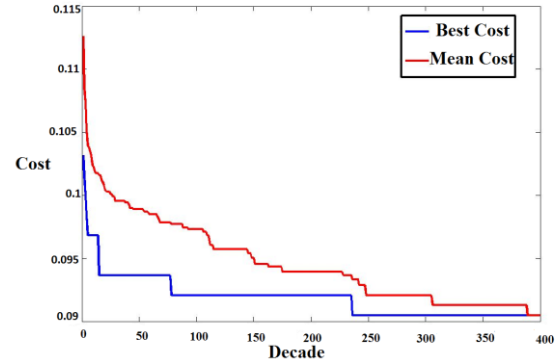


Figure 6: ICA based feature selection searching for optimal solution

4. CONCLUSION AND FUTURE WORK

In this paper, a new optimal feature selection technique based on imperialist competition algorithm was proposed. The performance of the proposed algorithm was tested for classification of bulk rice samples. The obtained results indicate the high performance of the proposed algorithm to find superior features leads to high classification accuracy.

The proposed feature selection algorithm can be combined with other population based feature selection techniques in future works. Furthermore, the robustness of proposed algorithm should be tested on different datasets.

Table 2. The performance of rice classification by ICA and genetic based feature selection methods

Variety of Rice	SVM prediction using All features			SVM prediction using genetic based feature selection			SVM prediction using ICA based feature selection		
	Precision	Recall	Fmeasure	Precision	Recall	Fmeasure	Precision	Recall	Fmeasure
Basmati	95.56	98.86	97.19	100	95.75	97.83	98.89	98.89	98.89
Daneboland	83.34	90.37	86.72	91.12	100	95.36	92.23	100	95.96
Domsiah	95.56	92.48	94	95.56	100	97.73	98.89	98.89	98.89
Fajr	98.89	74.79	85.17	100	80.36	89.12	100	81.82	90.01
Gerde	61.12	98.22	75.36	73.34	100	84.62	72.23	100	83.88
Hashemi	94.45	95.51	94.98	98.89	98.89	98.89	98.89	98.89	98.89
Tailandi	97.78	95.66	96.71	100	98.91	99.46	100	98.91	99.46
Total	89.53	92.27	90.02	94.13	96.28	94.72	94.45	96.78	95.14

5. REFERENCES

- [1] X. Wang,j. Yang, X.Teng, W.Xia, R.Jensen, 2007, Feature selection based on rough sets and particle swarm optimization, *Pattern Recognition Letters*, vol. 28, pp. 459-471.
- [2] H. Uguz, 2011, A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm, *Knowledge-Based Systems*, vol. 24, pp. 1024-1032
- [3] A. A. Ahmed, 2005, Feature subset selection using ant colony optimization, *International Journal of Computational Intelligence*, vol. 2, pp. 53-58.
- [4] E. Atashpaz-Gargari and C. Lucas, 2007, Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition, pp. 4661-4667.
- [5] Z.Liu,F. Cheng., Y.Ying, X.Rao, 2005, Identification of rice seed varieties using neural network, *Journal of Zhejiang University Science B*, vol. 6, pp. 1095-1100.
- [6] S.J Mousavi Rad, F. Akhlaghian, K.Mollazade, 2011, classification of rice varieties using optimal color and texture features and BP Neural Network, *The 7th Iranian Conference on Machine Vision and Image Processing*, Iran University of Science and Technology.
- [7] B. Verma, 2010, Image processing techniques for grading & classification of rice, *International conferences on computer and communication technology*, pp. 220-223.
- [8] C. Zheng, D.W. Sun, L. Zheng, 2006, Recent applications of image texture for evaluation of food qualities-a review, *Trends in Food Science & Technology*, vol. 17, pp. 113-128.
- [9] R. M. Haralick, k.shanmugam, I.H. Dinstein , 1973, Textural features for image classification, *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 3, pp. 610-621.
- [10] A. Mucherino,P.J. papajorgji, P.M. Pardalos, 2009, *Data mining in agriculture* vol. 34, Springer Verlag.