Agricultural Plant Leaf Disease Detection and Diagnosis Using Image Processing Based on Morphological Feature Extraction

Mr. Sachin B. Jagtap[1], Mr. Shailesh M. Hambarde[2]

[1] *M.E.(Electronics-Digital Systems)* [2] *Assistant Professor(E&Tc Department)*
[1,2] *university Of Pune , Jspm's Jscoe, Pune, India.*

Abstract: *Leaf spots can be indicative of crop diseases, where leaf batches (spots) are usually examined and subjected to expert opinion. In our proposed system, we are going to develop an integrated image processing system to help automated inspection of these leaf batches and helps identify the disease type. Conventional Expert systems mainly those which used to diagnose the disease in agriculture domain depends only on textual input. Usually abnormalities for a given crop are manifested as symptoms on various plant parts. To enable an expert system to produce correct results, end user must be capable of mapping what they see in a form of abnormal symptoms to answer to questions asked by that expert system. This mapping may be inconsistent if a full understanding of the abnormalities does not exist. The proposed system consists of four stages; the first is the enhancement, which includes HIS transformation, histogram analysis, and intensity adjustment. The second stage is segmentation, which includes adaptation of fuzzy c-means algorithm. Feature extraction is the third stage, which deals with three features, namely color size and shape of spot. The fourth stage is classification, which comprises back propagation based neural networks.*

Keywords: *ANOVA, Classifier, FCM, Feature extraction, Image processing, Leaf disease*

I. Introduction

Expert systems are intelligent computer programs that are capable of offering solutions advices related to specific problems in a given domain, both in a way and at a level comparable to that of the human experts in the field. One of the advantages of employing expert system is its ability to reduce the information that human users need to process, reduce personnel cost and increase throughput. Another advantage of expert system is it performs tasks more consistently than human experts. Knowledge based expert system technology has been successfully applied to a variety of agricultural problems since the early 1980s. At CLAES (Central Lab of Agricultural Expert System in Egypt) several expert systems have been developed for different agricultural activities [1]. Many of these expert systems are used to diagnose disorders from observation symptoms. These diagnosing expert systems depend on the ability of an end user to understand abnormal symptoms of the plant and to convey these symptoms through a textual dialogue [2]. Depending on the user's level of understanding of the abnormal observations, expert system can reach the correct diagnosis. If end user interprets the abnormal observations in a wrong way and chooses a wrong textual answer to a presented question, then expert system will reach a wrong conclusion. Devising a method whereby abnormalities are automatically detected, would greatly reduce the risk of human error and would lead to a more accurate diagnosis. This could be achieved through the integration of an image processing component with a diagnostic problem solver. Image processing is a powerful tool that has been applied in many domains such as remote sensing via satellite, medical image analysis, radar, sonar robotics and automated inspection [3]. Image information can play a crucial role in the diagnosis of different diseases in the agricultural domain where the understanding of image symptoms is often essential to problem solving. Vegetable crops suffer from many leaf batches. Leaf batches differ in color, shape and size according to the cause. Leaf batches happen as a result of plant pathogens (Fungi, Bacteria and viruses), insect feeding (sucking insect pastes), and plant nutrition (lack of microelements). The importance of the fungal diseases came from its great distribution, short life cycle and propagation. In agricultural mass production, it is needed to discover the beginning of plant diseases batches early to be ready for appropriate timing control to reduce the damage production costs and increase the income. Plant leaves are considered first station for the rest and germination of bacterial, fungal capsules due to suitable macro environment. Leaf batch characteristic plays a crucial role in differentiating between different causes [4]. In the diagnosis of leaf batches there is some confusion due to the similarities between batch's shape size and color but only expert could identify it. The first step in fighting against these leaf batches is the adequate recognition of their presence i.e. correct diagnosis. An abnormal symptoms in an indication to the presence of the disease, and hence can be regarded as an aid in diagnosis. Spots are considered as the important units indicating the existence of diseases. Aim of our system is an efficient identification of these leaf spots. The development of such an intelligent system is justified by its economical relevance and by hard efforts necessary to perform a correct diagnosis. This includes the knowledge

and experience accumulated by the human experts. Other important aspects are the speed, safety and reliability of the response of the system [1].

II.   Review Of Prior Leaf Disease Detection Techniques

Detection of harmful viruses and bacteria in plant material, vectors or natural reservoirs is essential to ensure safe and sustainable agriculture [4]. The techniques available have evolved significantly in the last few years to achieve rapid and reliable detection of pathogens, extraction of the target from the sample being important for optimizing detection. Serological and molecular techniques are currently the most appropriate when high numbers of samples need to be analyzed. Specific monoclonal and/or recombinant antibodies are available for many plant pathogens and have contributed to the specificity of serological detection. Molecular detection can be optimized through the automatic purification of nucleic acids from pathogens by columns or robotics [5]. New variants of PCR, such as simple or multiplex nested PCR in a single closed tube, co-operative- PCR and real-time monitoring of amplicons or quantitative PCR, allow high sensitivity in the detection of one or several pathogens in a single assay. The latest development in the analysis of nucleic acids is microarray technology, but it requires generic DNA/RNA extraction and pre-amplification methods to increase detection sensitivity. Some of the limitations of the molecular techniques are that they are time-consuming and labor-intensive, and require an elaborate procedure, especially during sample preparation (collection and extraction) to obtain reliable The molecular techniques could be used as robust tool to ensure the presence of plant diseases, but cannot be used as a preliminary screening tool for processing large number of plant samples due to the time involved in the process [6].

III.   Features Of Leaf Image

When crops suffer from many diseases, batches (spots) often happen on leaves. Leaf spots are considered the important units indicating the existence disease and regarded as indicator of crops disease [7]. In order to classify disease leaf samples category, a set of spot features for classification and detection of the different disease leaves, as in Figure 1,

Fig1. Samples of diseased leaf images

are investigated in this paper. Spot features are extracted from image using the appropriate image processing method. These features are very important for the color and morphology of the leaf spots and they provide critical information about its visual representation [8]. The features correspond to color characteristics are the mean and variance of the gray level of the red, green and blue channel of the spots; and other features correspond to morphological and geometrical characteristics of the spots[8],[9],[10],[11],[12]. These features correspond to morphological characteristics of spots such as:

1.   Length and ratio of principal axes: Ratio of principal axes length is major axis length divided by Minor axis length.

2.   Center of Gravity: For a spot surface described by function $f(x, y)$ consisting of N pixels, the Center of Gravity coordinates$(\bar{x}, \bar{y})$ can be calculated as:

$$\bar{x} = \frac{1}{N}\Sigma\Sigma x \qquad \bar{y} = \frac{1}{N}\Sigma\Sigma y \qquad\qquad (1)$$

3.   Orientation: Orientation is defined as the angle between the major axis of spot and the horizontal axis. Because the major axis exhibiting the minimum moment of inertia, it can be calculated as:

$$\theta = \frac{1}{2}\arctan\left[\frac{2\times\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right] \qquad\qquad (2)$$

4.   Equivdiameter: The diameter of a disk with the same area as spot is computed as:

$$\sqrt{\frac{4\times spot\ area}{\pi}} \qquad\qquad (3)$$

5.   Eccentricity: The ratio of the distance between the foci and major axis length of the ellipse eccentricity ratio. Its value is between 0 and 1, the spot whose eccentricity ratio is 0 is actually a circle, while the spot whose eccentricity ratio is 1 is a line. The eccentricity ratio is calculated as:

$$2 * \sqrt{\left[\left(\left(\frac{major}{2}\right)\right)^2 - \left(\left(\frac{minor}{2}\right)\right)^2 \Big/ major\right]} \qquad (4)$$

6.   Solidity: also called compactness, has a value between 0 and 1, if the spot has a solidity value equal to 1, this means that it is fully compacted. It is the ratio the convex Area and the area of the spot, computed as:

$$\frac{Spot\ Area}{Convex\ Area} \qquad (5)$$

7.   Extent: also called rectangularity ratio, the proportion of the pixels in the bounding box that are also in the spot. It has a value between 0 and 1, when this ratio of spot has the value one then its shape is perfectly rectangle. Computed as the spot area divided by the area of the bounding box:

$$\frac{Spot\ Area}{Bounding\ Box\ Area} \qquad (6)$$

8.   Hydraulic Radius: It is calculated by dividing the spot area by the spot perimeter:

$$\frac{4 \times \pi \times Spot\ Area}{(Spot\ Perimeter\ )^2} \qquad (7)$$

9.   Complexity: also called scatter index, it indicates the perimeter of unit spot area. The bigger value, the spot shape is more complex, and vice versa. It can be computed as:

$$\frac{(Spot\ Perimeter\ )^2}{Spot\ Area} \qquad (8)$$

10.   Euler Number: It describes a simple, topologically invariant property of spot. It is computed as the number of objects in the region minus the number of holes in those objects.

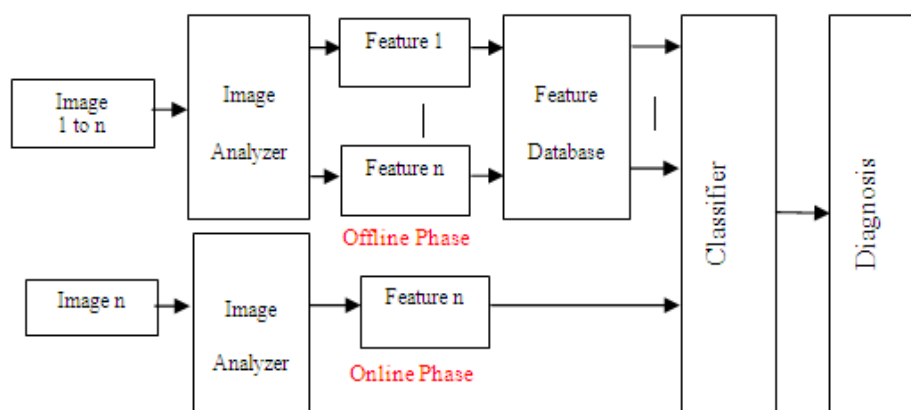## IV.   System Structure

Fig2. Architecture of proposed system

In order to diagnose the cause of the symptom by using an automated tool, so image processing system is proposed to develop to automate the identification and classification of the leaf batches into specific disorder. As shown in figure above the system consists of three main blocks: Image Analyzer, Feature Database and Classifier resp. [1]. The processing proposed to done by these blocks is divided into two phases as follows-

Offline Phase: A large set of defected images are processed by image analyzer for extracting abnormal features. Then these features are stored in feature database for later usage by the classifier.

Online Phase: Abnormal feature of a specific defected image is extracted by image Analyzer and then classified by the classifier into a specific disorder, Thus Diagnosis is facilitated.

Main aim of image analyzer is to extract the abnormal symptoms of the defected color image represented in spot size, spot color, and spot shape. As shown in figure below input to image analyzer is the acquired defected color image and the output of the image analyzer is the extracted features of the defected image [8].
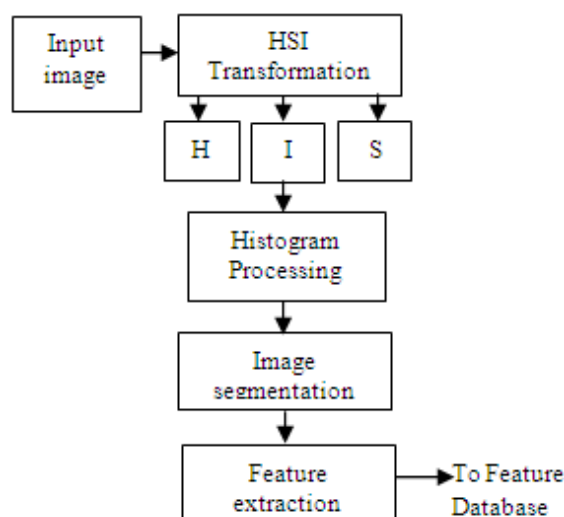
Fig3. Block diagram of Image analyzer

The camera is placed at about 60mm from top of the leaves. The image from camera is digitized into a 24 bit image with resolution 720x540 pixels. Image Enhancement is a sub field of image processing and consists of techniques to improve the appearance of an image, to highlight important features of an image, and to make this image more suitable for use in a particular application (e.g. make certain features easier to see by modifying the colors or intensities)[13][7][8].

The abnormality of the defected leaves is revealed by the appearance of spots. From the inspection of infected leaves it was found that spots have intensity values higher than other normal tissues. To extract those abnormal tissues, so our enhancement processing is as proposed in block diagram above consists of three steps:

STEP1: The first step is the transformation of defected image into HSI color space.

STEP2: The second step is analyzing the histogram of the intensity image to get the threshold by which we can increase the contrast of the image

STEP3: The final step is to adjust the intensity of the image by applying the thresholds [13][7][1].

Image segmentation is to be done to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. Image segmentation is the first step in image analysis and pattern recognition. It is a critical and essential step and is one of the most difficult tasks in image processing, as it determines the quality of the final result of analysis. The segmentation can be done using both classical and fuzzy techniques. We are going to use Fuzzy C-means (FCM) algorithm. The segmentation of defected plant images involves partitioning the image space into different cluster regions with similar intensity image values. The success of applying FCM to fit a segmentation problem depends mainly on adapting the input parameters such as No of clusters, Degree of fuzziness, feature of the data set etc. If any parameter is assigned improperly leads to a wrong decision i.e. partition results is not optimum [14][15]. We are going to use Fuzzy C-means (FCM) algorithm. The segmentation of defected plant images involves partitioning the image space into different cluster regions with similar intensity image values. The success of applying FCM to fit a segmentation problem depends mainly on adapting the input parameters such as No of clusters, Degree of fuzziness, feature of the data set etc. If any parameter is assigned improperly leads to a wrong decision i.e. partition results is not optimum [14][15][16].

In feature extraction phase our purpose is to reduce the image data by measuring certain features or properties of each segmented regions such as: color, shape, Texture etc. This phase consists of two steps namely spot isolation and spot feature extraction. Data consisting of those extracted features is stored in feature database for later usage by the classifier.

An Artificial Neural Network ANN is proposed to perform classification task in proposed system. Before online processing is done, the system needs to be manually trained using a set of training samples. There are many different types of ANNs. The most widely used is back propagation ANN, which is excellent for performing classification task [14][17]. The configuration of ANN needs to be determined accurately to give an optimal classification result. This configuration (also called Topology) includes no. of layers, the no. of neurons for each layer, and the minimal no. of training samples. There should be tradeoff between the selection of components of ANN e.g. Too many neurons may degrade the effectiveness of the model, leads to undesirable results, long training times etc. Large no. of connection weights in the ANN model may cause over fitting and loss of generalization capacity. On other hand, choosing too few neurons may not compute the full complexity

of the data. All techniques that existed, for finding the optimal no. of neurons in the hidden layer, employ trial and error method, in which the ANN training starts with a small no of neurons, and additional neurons are gradually added until some performance target is reached. Unfortunately, there is no theoretical consideration for determining the optimal network topology for a specific problem. We proposed to use statistical analysis technique called ANOVA, for determining the optimal configuration of the neural network [18].

V.   Results And Proposed Work

HSI transformation is used to transform defected image into HSI color space. Histogram analysis is used to increase contrast of image by adjusting its intensity for image enhancement purpose [19]. Following are the results of image enhancement:
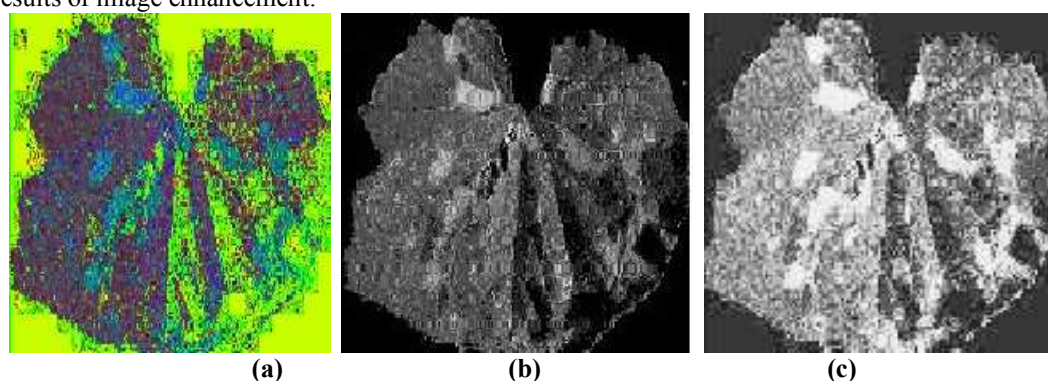

(a)   **(b)**   **(c)**
Fig 4.(a) HSI image **(b)** Intensity image **(c)** Enhanced image

Image segmentation is the first step in image analysis and pattern recognition it is a critical and essential step and is one of the most difficult tasks in image processing, as it determines quality of the final result of the analysis. The problem of segmentation is broadly investigated by scientists using both classical and fuzzy based techniques. Fuzzy set theory provides a mechanism to represent and manipulate the uncertainty and ambiguity. Therefore fuzzy clustering turns out to be particularly suitable for the segmentation of plant images. We are going to use fuzzy c-means (FCM) algorithm. Bezdek's FCM is most commonly used algorithm [16]. After the proper segmentation by setting all important parameters, such as feature of dataset, optimal no of clusters and degree of fuzziness, results should be as followed:
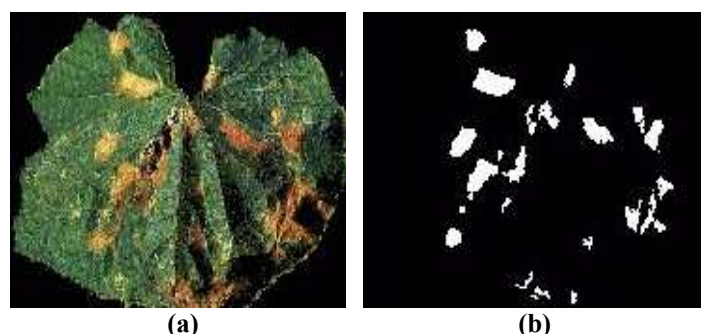

(a)   **(b)**
Fig 5. (a) Original image (b) Segmented image

Purpose of next feature extraction stage is to reduce image data by measuring certain features or properties of each segmented region such as color, shape or texture. This will be done in two steps, spot isolation and spot extraction. Component labeling algorithm will be used in the step of spot isolation. In feature extraction step values for various features of leaf images, as mentioned earlier in the paper, are extracted and summarized in a in a database table as a specific entries for each disease[20]. Table will be look like as shown below:

Table1. Features of Leaf Diseases

Features	Classes of leaf disease		
	Anthracnose	Black Spot	Red Leaf Blight
Area	76.03	13	56.73
Major axis	11.70	4.31	9.42
Minor axis	7.53	4.31	6.15
Orientation	24.88	27.44	5.40
Ratio of principle axes	1.53	1.00	1.53
Equivdiameter	9.04	4.07	7.233

Eccentricity	0.67	0.42	0.613
Solidity	0.96	1.00	0.943
Extent	0.65	0.52	0.65
Hydraulic radius	1.08	1.28	1.265
Complexity	12.95	9.85	12.00
Euler number	1.00	1.00	1.00
Moment of Inertia xx	6.53	4.08	8.98
Moment of Inertia yy	7.51	8.98	12.54

We are going to use Artificial Neural Network to perform classification task based on data summarize in feature database. For this purpose we are going to use most popular and widely used back propagation ANN which is excellent for performing classification task. While designing ANN we will decide best configuration of no. of layers, no. of neurons in each layer and minimum no. of training samples on the experimental basis. While performing such experiment identification factor for each disease will be the parameter under consideration.

Table2. Experimental design for neural network factors

ANN configuration			Identification factor		
Samples	Neurons	Layers	Anthracnose	Black Spot	Red Leaf Blight
50-250	5	One	0.44	0.26	0.52
		Two	0.68	0.70	0.68
	10	One	0.86	0.40	0.52
		Two	0.60	0.43	0.59
150-150	5	One	0.81	0.75	0.63
		Two	0.83	0.59	0.82
	10	One	0.76	0.52	0.75
		Two	0.79	0.72	0.62

This process mostly employ trial and error methods, in which ANN training starts with a small no. of neurons , and additional neurons are gradually added until some desired performance target is achieved[21]. There is no standard theoretical consideration for determining the optimum network configuration for specific problem. We will use a well known statistical analysis technique called ANOVA, for determining the optimal configuration of neural network. Good introduction to ANOVA is given by Norman R. [18].While performing ANOVA, we will determine the significance of each component of ANN on its classification performance.

VI.   Conclusion

Proposed System shows usefulness of integration of an image analyzer aided with pattern recognition within a diagnostic expert system model. In order to diagnose a disorder from leaf image four image processing phases have to be applied: Image enhancement, Image segmentation, Feature extraction, & classification. In order to employ proposed system we first have to train it with a set of images of disorders. Applying this model to any other crop disorder requires only spatial care to be taken in order to acquire a sufficient set of images for training purpose as representative of these disorders. Due to integration of this proposed system diagnosis accuracy will increase. Proposed system focuses on specific disorder's identification, it can be extended in order to include more disorders. Extension of system in such a way that it will be capable to detect and identify abnormalities on the other parts of plants also e.g. fruit, stem, & root. Potential future work will be development of a robotic expert system which may capable to see abnormalities of plant understand it and do treatment operations directly. Other important future work will be integration of diagnosis prior to disease in the proposed system with the help of extensive plant characteristics, behavioral study at micro level.

References

[1].   Mohammad El-Helly, Ahmed Rafea, Salwa El-Gammal, "An Integrated Image Expert Processing System for Leaf Disease Detection and Diagnosis", Central Lab. For Agricultural System(CLAES), Agricultural Research Center(ACR), Giza,Egypt

[2].   John Hartman, Brian Eshenaur, 'Plant Pathology fact sheet', University of Kentucky-college of agriculture,PPFS-GEN-02.

[3].   Report on plant disease, Dept. of crop sciences, University of Illinois at urbana-champaign, RPD NO.705 March 2004.

[4].   Agrios G.N., Plant Pathology (Academic Press, 4th Edition, 1997).

[5].   Sindhuja Sankaran, Ashish Mishra, Reza Ehsani, Cristina Davis, 'A review of Advanced techniques for detecting plant disease', Computer and electronics in agriculture 72(2010)1-13, 2010.

[6].   Review article, 'Innovative tools for detection of plant pathogenic viruses and bacteria' Int Microbiology ,6: 233-243, 2003.

[7].   EI-Helly M., Rafea. A., and EI-Gammal S."An integrated image processing system for leaf disease detection and diagnosis", 1st Indian international conference on AI (IICAI-03), Dec. 2003.

[8].   Rafael C. Gonzalez, Richard E. Woods, and Steven. L. eddins, Digital Image Processing (Publishing house of Electronics Industry, Beijing, 2nd Edition ,2003).

[9].   Rafael C. Gonzalez, Richard E. Woods, and Steven. L. eddins, Digital Image Processing Using Matlab (Publishing house of Electronics Industry, Beijing, 2005).

[10].   The Mathworks Inc, Image processing toolbox user's guide vision 5. 2004.

[11].   National Instruments, IMAQ Vision 6 Tutorial, 1998.

[12].   National Instruments, IMAQ Vision Builder Tutorial, 2002.
[13].   Ahmed, Salwa & Mohammad El-Helly, Reda, Integrated Diagnostic Expert system with image processing via Loosely coupled technique, Central Lab For Agricultural Expert System (CLAES) Agriculture Research Center(ARC), Cairo University.
[14].   Gonzalez R.C., Woods R.E., Digital Image Processing (Pearson Education.Inc. 3rd Edition, 2002).
[15].   Jang, Chuen & Eiji, Neuro-Fuzzy and Soft Computing (PHI Learning Private Limited,Original Edition, 2010).
[16].   Robert L. Cannon, Jitendra V. Dave, and James C. Bezdek," Efficient Implementation   of the Fuzzy c-Means Clustering Algorithms" IEEE transactions on pattern analysis and machine intelligence. vol. pami-8, no. 2, march 1986.
[17].   Image Processing Toolbox Users Guide Version 3. Mathworks Incorporation, Matlab Doc. 2002
[18].   Norman R., Introduction to social statistics (McGraw-Hill,Inc.,1983).
[19].   Jiaozuo, P. R. China," Local Adaptive Image Enhancement Based on HSI Space.",   Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT '10) , 14-15,August 2010, pp. 253-255,ISBN 978- 952-5726-10-7
[20].   Zhang, Mao, Hu and Li, "Feature Selection of Cotton Disease Leaves Image Based on Fuzzy Feature Selection Techniques", Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007
[21].   Boger Z. Weber R., Finding an Optimal Artificial Neural Network Topology in Real-Life Modeling, Proceedings of the ICSC Symposium on Neural Computation, Article No. 1403/109, 2000.