$\mathcal{AP}$
ijpam.eu

MONITORING OF STRUCTURAL INTEGRITY USING UNSUPERVISED DATA CLUSTERING TECHNIQUES

Simone S.F. Souza[1] [§], Fernando P.A. Lima[2], Fábio R. Chavarette[3]

[1,2]Electrical Engineering Department
UNESP – Univ. Estadual Paulista
Brazil Avenue, 56, 15385-000, Ilha Solteira, SP, BRAZIL

[3]Department of Mathematics
UNESP – Univ. Estadual Paulista
Brazil Avenue, 56, 15385-000, Ilha Solteira, SP, BRAZIL

Abstract: This work presents a comparative study of three unsupervised data clustering techniques used to perform the monitoring of the structural integrity of an agricultural tractor. The techniques used in this study are: K-Means, Fuzzy C-Means and Kohonen artificial neural network. These techniques are intelligent learning tools, which provide a classification of the information based on the similarity clustering. The main application of these tools is to assist in structures inspection process in order to identify and characterize flaws as well as assist in making decisions, avoiding accidents. To evaluate these algorithms the modeling was performed and signs of simulation from a numerical model of an agricultural tractor. The results obtained by the methodologies presented a comparative study.

[§]Correspondence author

1. Introduction

Agriculture is one of the main economic activities of modern society and, as in all areas, computerization has been incorporated in the production process over the years. The incorporation of technology has enabled the expansion of the cultivated area and increased productivity.

Even greater growth of agricultural production depends, among other factors, the level of technology on tractors used in agriculture. Agricultural tractors must meet requirements of cost and productivity, so that they are able to operate at peak performance in the field tasks. However, in some of the route and speed conditions, the tractor can be unstable and provoke vibrations that are harmful to the integrity of the tractor structure or, in extreme cases, its overturning, causing an operational accident [6].

In this sense it is necessary to develop efficient ways to monitor the health of this type of structure, to detect flaws in initial states, intervene in its spread and consequently prevent wear or even your tipping. Diagnostic systems failures, or as more known, "Structural Health Monitoring (SHM)" carry out tasks such as data acquisition and processing, analysis and validation, detection, characterization and interpretation of adverse changes in a structure of to aid decision-making and identify structural failure [4].

The flaws in structures such as agricultural tractors can occur as a consequence of factors such as component wear, cracks, loosening of screw connections, or simply the combination of these. The fault cause a variation in the structure of spatial parameters, generating a reduced structural rigidity, mass reduction and also increase the damping, so that the dynamic behavior of the structure is changed, which may occur an accident [18].

To solve this problem, various solutions have been proposed, such as traditional SHMS based on ultrasonic inspection, radiography X, acoustic emission testing, among others. However, none of these techniques can meet the needs of a structure that is always in motion as agricultural tractors. Thus, a solution to develop modern and efficient SHMS is the use of intelligent techniques such as artificial neural networks and data clustering tools, which are unsupervised tools for automating the decision-making.

In literature can cite many studies that use intelligent techniques to solve this problem, among the most important we have approaches such as artificial neural networks [9], genetic algorithms [17], artificial immune systems [10], among others.

In this paper, we present three approaches to perform monitoring of the structural integrity of an agricultural tractor. In this sense, are used the K-

means algorithm [11], Fuzzy C-Means [3] and the Kohonen neural network [7]. These are unsupervised data clustering techniques that provide a classification based on the analysis of similarity between the patterns. To evaluate the proposed algorithm one simulated signals database was used numerically from a mathematical model of an agricultural tractor. The tractor was modeled and was simulated [12]. Through the results was carried out a comparative study of the proposed algorithms, to identify the best unsupervised data clustering tool to perform the monitoring of structural integrity. The results showed efficiency and accuracy, justifying the applicability of the proposed tools.

This paper is organized as in Section 2 describes the data clustering algorithms used in this study. Section 3 presents the mathematical model and simulations. The proposed methodology is presented in section 4. Results and applications are presented in Section 5. Finally, Section 6 presents the conclusion of this work.

2. Data Clustering

The data clustering tools have the objective provide an classification of a set of information using as basis of comparison operations of numeric values, so that it is possible to classify the data automatically without the need for supervision, which featuring a unsupervised learning paradigm.

This paper presents the K-Means, Fuzzy C-Means and Kohonen neural network. These algorithms are described in the following sections.

2.1. k-Means

The k-means is an unsupervised machine learning technique that was proposed by [11], and aims to create partitions of an n-dimensional population into k groups in a certain database.

The k-means algorithm uses an input parameter (k) determining the amount of clusters, and these clusters have (n) elements. After the execution, intend to obtain a high similarity of the elements of a group and lower similarity entrees clusters created by the algorithm.

The K-means algorithm can be summarized as follows [15, 16]:

Step 1. Choose random k objects to the database as the initial centers of the clusters;
Step 2. Assign each object to the cluster to which the object is more similar,

according to the average value of the objects in the cluster;

Step 3. Update the average of the clusters, namely, calculate the average for each cluster of objects;

Step 4. Tests the stopping criterion and exit or returns to item 2. The stopping criterion can be a defined number of iterations or threshold of convergence ($\varepsilon > 0$).

2.2. Fuzzy C-Means

The Fuzzy C-means data clustering algorithm (FCM) was proposed in [3] and extended by [1]. The FCM is an algorithm that divides a data set $X = \{x_1, x_2, x_3, ..., x_n\}$ in c groups. The result of the grouping is expressed in pertinence degrees in the matrix μ, where μ_{ij} is the degree of relevance of the object x_i to the *j-th* group.

The FCM algorithm tries to find a fuzzy partition that represents the structure of the data, minimizing the objective function defined as:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}^m d(x_i; c_j)^2 \tag{1}$$

Subject to restrictions:

$$\sum_{j=1}^{c} \mu_{ij} = 1, \forall i \in \{1, ..., n\}$$

$$0 < \sum_{i=1}^{n} \mu_{ij} < n, \forall j \in \{1, ..., c\} \tag{2}$$

where n is the number data; c the number of groups considered in the algorithm; $m > 1$ the weighting parameter that controls how is the fuzzy partition (fuzziness value). Usually, m is in the range $[1.25, 2]$; $x_i \in R^s (i = 1, ..., n)$ is the data vector, where each vector represents a position in the attribute data; $c_j \in R^s (j = 1, ..., c)$ is the centroid of the *j-th* group; $d(x_i; c_j)$ is the distance between x_i and c_j ;

The FCM algorithm is to update the pertinence degrees and values of the centroids, using (3) and (4).

In (3) the goal is to assign a degree of relevance to the data x_i relative to j group that is proportional to their degree of relevance to the other groups.

The input parameters to the algorithm are the n data, the number of groups is c and the value of the fuzzification parameter is m.

The main steps of the FCM are described below:

Step 1: Initialize the array of pertinence μ with continuous random numbers in the range $[0, 1]$;

Step 2: Calculate the centroid of the group j using (3);

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m x_i}{\sum_{i=1}^{n} \mu_{ij}^m} \tag{3}$$

Step 3: Calculate the initial value to J using (1);

Step 4: Calculate the array of pertinence fuzzy μ this way:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d(x_i; C_j)}{d(x_i; C_k)} \right)^{\frac{2}{m-1}}} \tag{4}$$

Step 5: If $d(J_U; J_A) \leq \varepsilon$ stop. Otherwise return to step 2.

The convergence criterion is the threshold $\varepsilon > 0$. Another possible criterion is stopped when a fixed number of iterations is executed.

The array of pertinence μ required by the algorithm can be substituted by the initial centroids cluster, random values within the data range. For this, the second step must be replaced by step 4 in each iteration.

2.3. Criteria for Validation of Clustering

In this section, we present two criteria used to validate the clusters. The purpose of these criteria is to check the quality of the partitions made by clustering algorithms. In this work, we used the simplified silhouette criteria (SSC) [5, 13] and the Fuzzy silhouette criterion (CSF) [2] to assess the results of the k-means algorithms and c-means fuzzy, respectively.

The SSC used in the K-means algorithm is given by:

$$CSS = \frac{1}{N} \sum_{q=1}^{n} s(q) \tag{5}$$

where $s(q)$ is the silhouette q-th pattern obtained by the following equation:

$$s(q) = \frac{b(q) - a(q)}{max\{a(q), b(q)\}} \tag{6}$$

$a(q)$ is the dissimilarity of the default q-th to the centroid of the group to which it is associated and $b(q)$ is the dissimilarity of the q-th standard the centroid of the nearest cluster.

The purpose of this silhouette is to identify a partition on which (6) is maximized. In [13] it is defined that $-1 \leq s(q) \leq 1$, then the smaller the value of $a(q)$ and the greater the value of $b(q)$ is closer to 1 is $s(q)$, so how much lower the value of $a(q)$ and the higher the value of $b(q)$, closest of 1 is $s(q)$, implying that the group of data determined by the algorithm is satisfactory.

In turn, the CSF is calculated using the equation (7):

$$CSF = \frac{1}{N} \frac{\sum_{q=1}^{n}(b\,(q) - a\,(q))^{\alpha}s(q)}{\sum_{q=1}^{n}(b\,(q) - a\,(q))^{\alpha}} \tag{7}$$

where $\alpha \geq 0$ is $a\,(q)$ is the degree of pertinence of q-th pattern of the centroid of the group to which it is associated and $b\,(q)$ is the degree of pertinence of standard q-th to the centroid of the most similar grouping. In fuzzy silhouette, $s(q)$ is calculated as (6).

2.4. Neural Network Kohonen

The Kohonen neural network was originally proposed in [7], and is composed of a two-dimensional grid of neurons, having an input set (x_j) associated with the weight vector (w_j). The weight vectors are initially set at random values in the range $[0, 1]$. Equations (8) and (9) show the definition of the set of input and weight vectors [7, 8].

$$w_j = [w_{1j}, w_{2j}, ..., w_{nj}] \tag{8}$$
$$x_j = [x_{1j}, x_{2j}, ..., x_{nj}] \tag{9}$$

Thus, the output y_j (j-th neuron activity) can be calculated by the following equation [7, 8]:

$$y_i = [w_{1j}x_{1j} + w_{2j}x_{2j} + ... + w_{nj}x_{nj}] \tag{10}$$

The training is a process through which the parameters of a neural network are adjusted by using a continuous stimuli. In the case of training for competition, as a stimulus to the network (input vector), the output neurons compete with each other to be activated. Thus, the neuron winner will update your weights in the training process. To update the value of the weights is used the following equation [7, 8]:

$$w_{h+1} = w_h + \alpha(x - w_h) \tag{11}$$

where α is the training rate ($0 \leq \alpha \leq 1$) and h is the updating index.

The winner neuron (NV) is chosen as being one that presents greater activity (y) on the Kohonen grid, as defined below [7, 8]:

$$NV = \max_j\{< w_j, x_j >\} \tag{12}$$

$$NV = \min_j\{\|w_j - x_j\|\} \tag{13}$$

where in $\|.\|$ is the operator of the norm Euclidean.

The winner neuron attributed to the index k. Then using the Kohonen rule proposed occurs the adjustment of the weights considering the neighborhood ($S1$, $S2$ and $S3$) around the winning neuron, as illustrated in Figure 1.
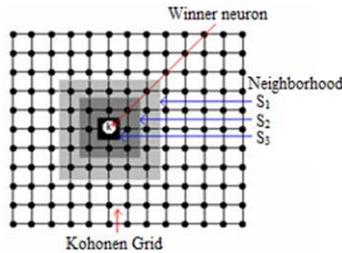

Figure 1: Vicinity of the winner neuron.

The steps used for training the Kohonen neural network are described below [7, 8]:

Step 1. Initialize the weight vector of the Kohonen neural network;
Step 2. Normalize the weight vectors;
Step 3. Normalize all input vectors;
Step 4. Show an input vector to be analyzed;
Step 5. Calculate the distance (inner product) for all neurons of Kohonen grid;
Step 6. Find the winning neuron using equations (12) or (13). Indicate the winner neuron through the index k;
Step 7. Update the weight vectors of the winner neuron and other neurons contained in the neighborhood chosen using equation (11) (see Figure 2);
Step 8. Re-normalize all vectors adapted weights at step 7. This procedure is necessary in order that the weight vector after adaptation, it has no unit length;

Step 9. Return to Step 4. This procedure should be repeated, considering a fixed number of iterations, or until changes (modules) of the weights is below a certain predetermined tolerance.

3. Mathematical Modeling and Simulations

3.1. Mathematical Model of Agricultural Tractor

To evaluate the proposed algorithms we realized simulations using a numerical model, built by ordinary differential equations and this model illustrated in Figure 2a [14].

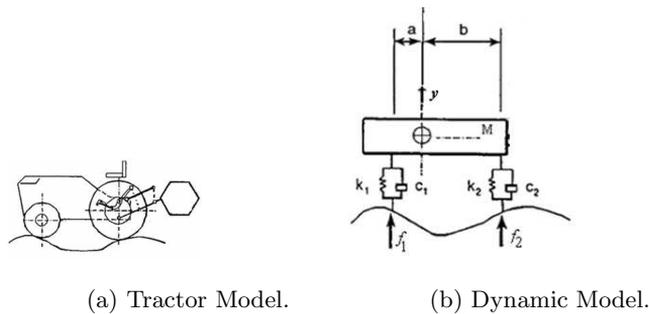(a) Tractor Model.   (b) Dynamic Model.

Figure 2: Model

In carrying out the work in the countryside the front and rear, wheels of a tractor may lose contact with the floor. This phenomenon can occur in three ways. Only the front wheels leave the floor (tractor rears up forward); only the rear wheels leave the floor (tractor rears back); the rear and front wheels leave the floor at the same time (the tractor takes a leap).

Given this, a simplified mathematical model of two dimensions (horizontal and vertical) which describes the dynamics of a tractor during the jump shown in Figure 2b.

The balance of power in the vertical direction and the jump time is equated by:

$$m\ddot{y} + F_1(y, \dot{y}, \theta, \ddot{\theta}) + F_2(y, \dot{y}, \theta, \ddot{\theta}) + mg = 0 \tag{14}$$

$$J\ddot{\theta} + aF_1(y, \dot{y}, \theta, \ddot{\theta}) - bF_2(y, \dot{y}, \theta, \ddot{\theta}) = 0 \tag{15}$$

where J is the angular momentum of the center of mass inertia; $F_1(\cdot)$ and $F_2(\cdot)$ are the dynamic loads on the front and rear tires of the tractor, given by:

$$F_1(y, \dot{y}, \theta, \dot{\theta}) = k_1(y + a\theta - h_1(t)) + c_1(\dot{y} + a\theta - \dot{h_1}(t)) \tag{16}$$

$$F_2(y, \dot{y}, \theta, \dot{\theta}) = k_2(y - b\theta - h_2(t)) + c_2(\dot{y} - b\theta - \dot{h_2}(t)) \tag{17}$$

$h_1(t)$ and $h_2(t)$ are periodic excitations that simulate the unevenness of the surface, and are defined by:

$$h_1(t) = h_0 sen(\omega t) \tag{18}$$

$$h_2(t) = h_0 sen(\omega t - \phi) \tag{19}$$

where $\omega = \frac{2\pi v}{L}$ and $\phi = \frac{2\pi W_B}{L}$.

It is observed that when the tractor:

- rears up forward ($F_1(\cdot) = 0$). Soon, $k_1 = c_1 = 0$;

- rears up back ($F_2(\cdot) = 0$). Soon, $k_2 = c_2 = 0$;

- jumps ($F_1(\cdot) = F_2(\cdot) = 0$). Soon, $k_1 = c_1 = k_2 = c_2 = 0$;

3.2. Simulations

To make the simulation of normal conditions (*base-line*) and abnormal (*structural failure*) on the tractor changes the distance between consecutive protrusions ($L = 0.6m, L = 0.4m, L = 0.3m, L = 0.2m$ and $L = 0.15m$) and speed-up range ($v = 1m/s$ to $10m/s$) for each L. The tractor operates in normal conditions with $L = 0.6m$ and variance $v = 1m/s$ to $3m/s$, and the others being considered abnormal, with $L = 0.15m$ and speed up $v = 6m/s$, that under certain conditions of the route and speed, the tractor may present instability in the vibrations damaging the control of the operator, or even cause an accident.

In this way, we simulate 500 signs in the modeled structure, signal 200 being in normal condition, i.e., without failure (*base-line*) and 300 signals structural failure. The signals were simulated considering the variation of the parameters (v and L). The tractor was excited by periodic functions is, as shown in [18] and [?].

After exciting the structure obtains the frequency response of the structure using fast Fourier transform (FFT). The frequency response is obtained from

the speed signals and the displacement of the structure to be excited. Each captured signal has 512 points. In Figure 3 illustrate two examples of signals collected in the simulations, Figure 3a shows a normal condition signal and 3b in a signal failure condition. Figure 3c illustrates a comparison of the frequency responses shown in 3a and 3b.
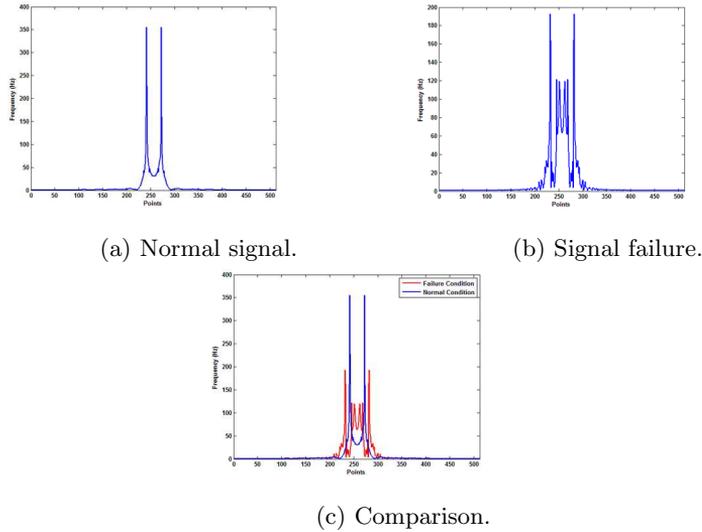

(a) Normal signal.   (b) Signal failure.

(c) Comparison.

Figure 3: Signal

4. Testing Methodology

All algorithms presented in this article were developed in MATLAB [12]. All tests were performed using a PC Intel core i7, 3.1 GHz, with 8 GB of main memory. The following is the database settings, the initialization parameters of the algorithms and the configuration of the tests.

4.1. Settings Test Base

The database for analysis and monitoring of the structural integrity of the agricultural tractor was simulated and obtained by the process described in section 3 of this article.

This data base consists of signals 500, 200 of the structure signals in normal condition (*base-line*) and 300 signs the structure in structural failure condition.

The signals represent the frequency response obtained from the FFT. All signals are real nature and has a length of 512 points.

To standardize the signals used in the tests used the z-score normalization according to the following equation:

$$v = \frac{v - u}{\sigma} \qquad (20)$$

where v the normalized value; v the value to be normalized; μ the mean population value and σ the population standard deviation.

4.2. Initialization Parameters

This section presents the initial parameters for algorithms k-means and FCM, and the Kohonen neural network.

For the k-means algorithm takes two parameters, as follows: k (number of clusters) and ε (convergence criteria). The values used for this problem are: $k = 2$, because we have two situations in the structure (normal and failure) and $\varepsilon = 0.01$.

For the FCM algorithm takes three parameters, as follows: c (number of clusters), m (fuzziness value) and ε (convergence criteria). The values used for this problem are: $c = 2$, $m = 1.25$ and $\varepsilon = 0.01$.

For Kohonen neural network requires four parameters for each stage of implementation (training and testing), as follows: the size of the grid (DG), the learning rate (α), the number of clusters (k) and the maximum number iterations (maxit). After conducting several tests, the architecture that showed the best result was configured using the $DG = 20x15$ parameters, $\alpha = 0.1$, $k = 2$ and $maxit = 10000$ for training phase. For the test phase the $GD = 20x15$ parameters were used, $\alpha = 0.1$, $k = 2$ and $maxit = 1000$.

4.3. Settings Test

To make a comparison among the three data clustering algorithms, several tests analyzing the computational time and the number of executed iterations have been performed.

In this regard a cross-reference test was performed where each algorithm was run 30 times. The test results obtained are presented in the following section.

5. Analysis and Results

This section presents the results obtained by applying the data clustering algorithms in the set of simulated data.

5.1. Results

Table 1 shows the number of grouped standards in each cluster of k-means algorithm, FMC and the network Kohonen.

Table 1: Number of grouped standards

Cluster	k-means		FCM		Kohonen	
	Standards	%	Standards	%	Standards	%
1	186	37.2	197	39.4	202	40.4
2	214	62.8	303	60.6	298	59.6

Patterns grouped by each algorithm the number of patterns classified correctly and incorrectly in each class are given in Table 2.

Table 2: Number of grouped standards

Cluster	Tests	k-means		FCM		Kohonen	
	Standards	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
1	200	187	13	194	6	199	1
2	300	285	15	295	5	296	4
Total	500	472	28	489	11	495	5
% accuracy and error		94.44	5.66	97.80	2.20	99.00	1.00

As shown in Table 2, the k-means algorithm, FCM and Kohonen properly grouped 94.44%, 97.80% and 99.00% of the analyzed signals.

The Table 3 shows the number of iterations performed and the computational algorithm each time it took to converge.

Table 3: Performance dos algorithms

	kmeans	FCM	Kohonem
Number of iterations	26	18	480
Computational time (s)	1.92	0.86	7.23

It is observed that on average the FCM algorithm with 18 iterations converge, the k-means algorithm with 26 iterations, and the Kohonen neural network with 480 iterations. The FCM algorithm has the lowest computational

time, then the k-means algorithm and finally the Kohonen neural network, with 0.86, 1.92 and 7.23 seconds respectively.

5.2. Discussion of Results

After performing all tests, it is observed that the Kohonen neural network has a better success rate compared to other algorithms. However, for computational efficiency FCM algorithm is the best, with the best performance.

Apart from this analysis is important to highlight the strengths and weaknesses of the implemented algorithms.

Strong points:

- The presented algorithms are easy to implement and provide efficient characterization of signals, and alternative tools for monitoring the structural integrity;

- The algorithms k-means and FCM have low computational time;

- The algorithms k-means and FCM have few parameters to be calibrated; Weak points:

- The Kohonen neural network has a high computational performance;

- The Kohonen neural network to be adjusted until an optimal architecture is found and the calibrated parameters;

6. Conclusion

In this article we were presented three approaches unsupervised data grouping to perform the monitoring of the structural integrity of an agricultural tractor. In this context, we tested the k-means algorithm and the FCM, and the Kohonen neural network. The three algorithms have a good success rate, with 94.44%, 97.80% and 99.00%, respectively.

The FCM algorithm and k-means have the best computational efficiency, however, the accuracy is low when compared to the Kohonen neural network. Already the Kohonen neural network has a high computational performance, plus it is justified because it is necessary to conduct a learning phase (training), which ensures a better accuracy in the diagnostic phase.

Finally, it is concluded that the presented algorithms are alternative tools to perform the monitoring of structural integrity, showing excellent performance.

Acknowledgment

References

[1] J. C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences*, **10**, No. 2 (1984), 191-203.

[2] R. J. G. B. Campello, E. R. Hruschka, A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, **157** (2006), 2858-2875.

[3] J. C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics*,**3**, No. 3, (1973), 32-57.

[4] S. R. Hall, The effective management and use of structural health data, *Proceedings of the International Workshop on Structural Health Monitoring* (1999), 265-275.

[5] E. R. Hruschka, L. N. de Castro, R. J. G. B. Campello, Evolutionary algorithms for clustering gene-expression data. *Proc. IEEE International Conference on Data Mining*, (2004), 403-406.

[6] J. K JR Khoury, G. P. Dias, R. R. Cordeiro, C. M. A. Souza, Modelagem da estabilidade de tratores agrcolas de pneus. *Pesquisa Agropecuria Brasileira*, **39**, No. 5 (2004), 459-468.

[7] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, 2nd Edition, Berlin, Germany, (1988).

[8] T. Kohonen, The Self-Organizing Map, *Proceedings of IEEE*, (1990), 1464-1480.

[9] F. P. A. Lima, F. R. Chavarette, S. S. F. Souza, M. L. M. Lopes, A. E. Turra, V. Lopes Jr., Monitoring and Fault Identification in Aeronautical Structures Using an ARTMAP-Fuzzy-Wavelet Artificial Neural Network,*Advanced Materials Research*, **1025-1026** (2014), 1107-1112.

[10] F. P. A. Lima, F. R. Chavarette, S. S. F. Souza, A. S. Souza, M. L. M. Lopes, Artificial Immune Systems Applied to the Analysis of Structural Integrity of a Building, *Applied Mechanics and Materials*, **472** (2014), 544-549.

[11] J. MacQueen, Some Methods for classification and analysis of multivariate observations. *Proceedings of the Berkley Symposium on Mathematical, Statistics and Probability*, (1967), 281-297.

[12] MATLAB 7.8 Version, Mathworks Company (2011).

[13] P. J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.*Proc. IEEE International Conference on DataMining*, (1987), 53-65.

[14] K. Sakai, *Nonlinear Dynamics and in Agricultural Systems*. Elsevier Science (2006).

[15] A. S. B. Samma, R. A. Salam, Adaptation of k-means algorithm for image segmentation. *Proceedings of World Academy of Science, Engineering and Technology*, (2009), 58-62.

[16] P. N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*. Pearson, 1st edition, (2006).

[17] F. L. Wang, T. H. T. Chan, D. P. Thambiratnam, A. C. C. Tan, Damage Diagnosis for Complex Steel Truss Bridges Using Multi-Layer Genetic Algorithm, *Journal of Civil structural Health Monitoring*, Springer-Verlag, (2013), 117-217.

[18] S. Zheng, X. Wang, L. Liu, Damage detection in composite materials based upon the computational mechanics and neural networks, *Proceedings of the European Workshop on Structural Health Monitoring*, Munich, (2004), 609-615.

134