Research Article
Abnormal Control Chart Pattern Classification Optimisation Using Multi Layered Perceptron

Mitra Mahdiani, Hairulliza Mohd. Judi and Noraidah Sahari Ashaari
Department of Industrial Computing, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Abstract: In today's industry, control charts are widely used to monitor production process. The abnormal patterns of a quality control chart could reveal problems that occur in the process. In the recent years, as an alternative of the traditional process quality management methods, such as Shewhart Statistical Process Control (SPC), Artificial Neural Networks (ANN) have been widely used to recognize the abnormal pattern of control charts. Various types of patterns are observed in control charts. Identification of these Control Chart Patterns (CCPs) can provide clues to potential quality problems in the manufacturing process. Each type of control chart pattern has its own geometric shape and various related features can represent this shape. Feature-based approaches can facilitate efficient pattern recognition since extracted shape features represent the main characteristics of the patterns in a condensed form. The objective of this study was to evaluate the relative performance of a feature-based CCP recognizer compared with the raw data-based recognizer. The study focused on recognition of six commonly researched CCPs plotted on the Shewhart X-bar chart. The ANN-based CCP recognizer trained using the nine shape features resulted in significantly better performance and generalization compared with the raw data-based recognizer.

Keywords: Artificial neural networks, control chart patterns, shape features, statistical process control

INTRODUCTION

Control charts are the one of the simplest monitoring tools and by the implementations of this tool the information obtained about the process is either is in control or out of control. If the process is in control, the operator or user can precede the process under the same conditions. Conversely, the operator or user must identify the root causes of events happened in the process.

CCPs can exhibit six types of pattern: Normal (NR), Cyclic (CC), Upward Trend (UT), Downward Trend (DT), Upward Shift (US) and Downward Shift (DS) (Montgomery, 2008). Except for normal patterns, all other patterns indicate that the process being monitored is not functioning correctly and requires adjustment. Figure 1 shows six pattern types of control chart.

In recent years, several studies have been performed for recognition of the unnatural patterns. Some of the researchers used the expert systems (Swift and Mize, 1995; Evans and Lindsay, 1988). The advantage of an expert system or rule-based system is that it contains the information explicitly. If required, the rules can be modified and updated easily. However, the use of rules based on statistical properties has the difficulty that similar statistical properties may be derived for some patterns of different classes, which may create problems of incorrect recognition.

Also, Artificial Neural Networks (ANNs) have been widely applied for classifiers. ANNs can be simply categorized into two groups comprising supervised and unsupervised. Most researchers (Ebrahimzadeh and Ranaee, 2010; Ebrahimzadeh et al., 2012; Le et al., 2004; Pham and Oztemel, 1992; Cheng and Ma, 2008; Sagiroujlu et al., 2000; Pham and Oztemel, 1994) have used supervised ANNs, such as Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Learning Vector Quantization (LVQ), to classify different types of CCPs. Furthermore, unsupervised methods, e.g., Self-Organized Maps (SOM) and Adaptive Resonance Theory (ART) have been applied to fulfill the same objective in other studies (Wang et al., 2007). The advantage with neural network is that it is capable of handling noisy measurements requiring no assumption about the statistical distribution of the monitored data. It learns to recognize patterns directly through typical example patterns during a training phase.

Some of the researchers used the support vector machine to CCP recognition (Ranaee et al., 2010; Ranaee and Ebrahimzadeh, 2011). The accuracy of an SVM is dependent on the choice of kernel function and the parameters (e.g., cost parameter, slack variables,

Corresponding Author: Mitra Mahdiani, Department of Industrial Computing, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia, Tel.: +60176361445
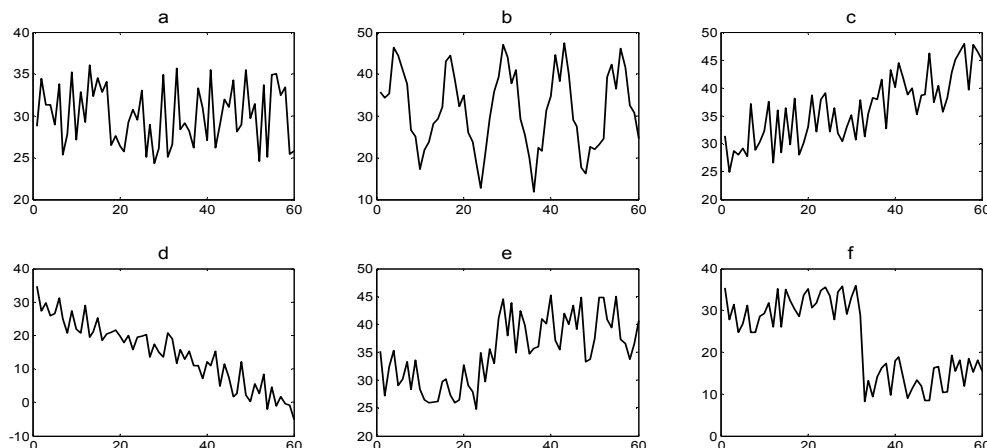
Fig. 1: Six various basic patterns of control charts, (a) normal pattern, (b) cyclic pattern, (c) upward trend, (d) downward trend, (e) upward shift, (f) downward shift

margin of the hyper plane, etc.). Failure to find the optimal parameters for an SVM model affects its prediction accuracy (Campbell and Cristianini, 1998).

Most the existing techniques used the unprocessed data as the inputs of CCPs recognition system. The use of unprocessed CCP data has further problems such as the amount of data to be processed is large. On the other hand, the approaches which use features are more flexible to deal with a complex process problem, especially when no prior information is available. If the features represent the characteristic of patterns explicitly and if their components are reproducible with the process conditions, the classifier Recognition Accuracy (RA) will increase (Pacella *et al.*, 2004). Features could be obtained in various forms, including shape features (Wani and Rashid, 2005; Gauri and Chakraborty, 2009; Pham and Wani, 1997), multi-resolution wavelet analysis (Ebrahimzadeh and Ranaee, 2010; Ranaee and Ebrahimzadeh, 2011) and statistical features (Hassan *et al.*, 2003).

Based on the published articles, there exist some important issues in the design of automatic CCPs recognition system which if suitably addressed, lead to the development of more efficient recognizers. One of these issues is the extraction of the features. In this study for obtaining the compact set of features which capture the prominent characteristics of the CCPs, a proper set of the shape features and are proposed.

Another issue is related to the choice of the classification approach to be adopted. The developed method uses a Multiplayer Perceptrons (MLP) as pattern recognizer. The MLP architecture has been successfully applied to solve some difficult and diverse problems in modeling, prediction and pattern classification (Haykin, 1999).

In recent years, control chart patterns have been widely used to solve the existing problems in the production process, so apart from the normal pattern, each pattern is indicative of a particular problem in the manufacturing process. In this study an automatic and accurate recognition system for control chart pattern

recognition based on Artificial Neural Network (ANN) has been proposed and investigated. In this research, ANN was used for intelligent classification.

METHODOLOGY

Data description: For this study, each pattern was taken as a time series of 60 data points. The following equations were used to create the data points for the various patterns (Ebrahimzadeh and Ranaee, 2010; Pham and Wani, 1997):

- Normal patterns:

$$p(t) = \eta + r(t)\sigma \tag{1}$$

- Cyclic patterns:

$$p(t) = \eta + r(t)\sigma + a\sin(2\pi t/T) \tag{2}$$

- Increasing trend patterns:

$$p(t) = \eta + r(t)\sigma + gt \tag{3}$$

- Decreasing trend patterns:

$$p(t) = \eta + r(t)\sigma - gt \tag{4}$$

- Upward shift patterns:

$$p(t) = \eta + r(t)\sigma + bs \tag{5}$$

- Downward shift patterns:

$$p(t) = \eta + r(t)\sigma - bs \tag{6}$$

where,
μ : The process variable's nominal mean value, which is the nominal mean under examination (set to 80)

σ : Process variable's under standard deviation (set to 5)

a : Cyclic variations' amplitude in cyclic pattern

g : Decreasing or increasing patterns' slope (set in the range between 0.2 and 0.5)

b : The shift position in an upward shift pattern and a downward shift pattern ($b = 0$ before the shift and $b = 1$ at the shift and thereafter)

s : The shift's magnitude (set between 7.5 and 20)

$r(t)$: A function that produces random numbers usually distributed between -3 and 3

t : The discrete time at which the observed process variable is sampled (set within the range of 0 to 59)

T : Cycle's period (set between 4 and 12 sampling intervals)

$P(t)$: Sampled data point's value at time t

Shape features: The shape features used by the CCP recognizer in this study are such that they facilitate recognition of CCPs quickly and accurately. The six types of CCP considered in this work have different forms, which can be characterized by a number of shape features. In Wani and Rashid (2005) and Pham and Wani (1997), the authors have introduced nine shape features for discrimination of the CCPs. These features are as follows.

S: Can be described as the slope of the least square line that represents the pattern and the magnitude of which is nearly zero for cyclic and normal patterns whereas for upward shift patterns its value is more than zero and it is less than zero for downward shift patterns together with decreasing trend. Therefore, it can be regarded as a suitable candidate for differentiating normal together with cyclic pattern from trend and shift patterns.

NC1: Can be described as the number of mean crossings, that is to say, pattern's crossing with the mean line. This feature can separate normal patterns from cyclic patterns together with cyclic and normal patterns from shift and trend patterns as its number of crossings for trend and shift patterns is small whereas this number for normal patterns is the highest and for cyclic patterns is intermediate, which means between those for shift or trend patterns and normal patterns.

NC2: Can be described as the number of least square line crossings. This feature can separate trend and natural patterns from other patterns as its value is highest for normal and trend patterns and the lowest for shift and cyclic patterns.

Cyclic membership (cmember): Is used to explain how closely a pattern is similar to a cyclic pattern. If complete cycles are not in existence or are unavailable, the slope of cycle's value may not be equal to zero but may be in the area of trend and shift patterns. The function of cyclic membership is defined for such a situation, which comes up with the extent of how closely a pattern is similar to a cycle. This feature separates cyclic patterns from other patterns as membership function, which will be described in detail later to create a positive value for cyclic patterns and a negative value for all other patterns.

AS: Can be described as the slope of the line segment in which each pattern will have two line segments which are appropriate to data, match the data and start from either end of the pattern in addition to the least square line that is close to a complete pattern. This feature separates the trend pattern from other patterns due to the fact that the average slope of line segments for a trend pattern will be higher than other patterns.

SD: Can be described as the slope difference between the least square line and the line segments depicting a pattern whose value is acquired through deducting the average slope of the two line segments from the least square line's slopes. The line segments and the least square line will not be the same for cyclic, normal and trend patterns. Therefore this feature separates a shift pattern from other patterns due to the fact that it will have a high value for a shift pattern and small value for other patterns.

APML: Can be described as the area between the pattern and the mean line which separate normal pattern from other patterns due to the fact that it is the lowest for a normal pattern.

APSL: Can be described as the area between the pattern and its least square line and can be utilized to separate cyclic together with shift patterns from normal and trend patterns due to the fact that cyclic and shift patterns have a greater APSL value than normal and trend patterns.

ASS: Can be described as the area between the line segment and the least square line, the value of which is nearly zero for a trend pattern and more than zero for an upward shift pattern. Therefore, this feature separates trend patterns from shift patterns.

Multi-Layer Perceptron (MLP) neural networks: An MLP neural network consists of an input layer (of source nodes), one or more hidden layers (of computation nodes) and an output layer. The recognition basically consists of two phases: training and testing. In the training stage, weights are calculated according to the chosen learning algorithm. The issue of training algorithm and its speed is very important for the MLP model. It is very difficult to know which training algorithm will be the fastest for a given problem. It depends on many factors, including the complexity of the problem, the number of data points in the training set, the number of weights and biases in the network, the error goal and whether the network is

being used for pattern recognition (discriminant analysis) or function approximation (regression). In this study the following training algorithms are considered.

Resilient back-Propagation (RPROP) algorithm: Derivatives' sign is considered as the sign for the direction of updating weight in RPROP (Riedmiller and Braun, 1993). In doing so, the weight step is not influenced by partial derivatives. The equation below puts the adjustment concerning update values of Δ_{ij} (weight change) for the RPROP algorithm on view. For initializing, all values are set to small positive ones:

$$\Delta_{ij}(t) = \begin{cases} \eta^+ \times \Delta_{ij}(t-1); & if \ \dfrac{\delta E}{\delta W_{ij}}(t-1)\dfrac{\delta E}{\delta W_{ij}}(t) > 0 \\ \eta^- \times \Delta_{ij}(t-1); & if \ \dfrac{\delta E}{\delta W_{ij}}(t-1)\dfrac{\delta E}{\delta W_{ij}}(t) < 0 \\ \eta^0 \times \Delta_{ij}(t-1); & otherwise \end{cases} \quad (7)$$

$\eta^0 = 0, 0 < \eta^- < 1 < \eta^+, \eta^{-,0,+}$ stand for update factor. Whenever the sign of the corresponding weight's derivative is changed, this indicates that the prior update value is too high and it has omitted a minimum. As a result, the value of update is decreased (η^-), as illustrated above. However, if the sign of derivative is not changed, the value of update is raised (η^+), which will help to speed up convergence in areas which are not deep. To avoid over-acceleration, for the purpose of avoiding over-acceleration, the value of new update in the epoch after applying (η^+) neither increases nor reduces (η^0) from the prior one. In every epoch, the values Δ_{ij} continue to be non-negative. The process of updating value adaptation is then followed by the actual process of updating weight which is determined through Eq. (8):

$$\Delta W_{ij}(t) = \begin{cases} -\Delta_{ij}; & if \ \dfrac{\delta E}{\delta W_{ij}}(t) > 0 \\ +\Delta_{ij}; & if \ \dfrac{\delta E}{\delta W_{ij}}(t) < 0 \\ 0 & ; \ otherwise \end{cases} \quad (8)$$

The values of the training parameters considered for the algorithms were set empirically.

Levenberg-Marquardt (LM) algorithm: In computing and mathematics, a numerical solution to the problem concerning minimizing a function, particularly appears in nonlinear programming and least square curve fitting, over a space of the function's parameter is provided by the Levenberg-Marquardt Algorithm (LMA) which is also recognized as Damped Least-Squares (DLS) method. Levenberg-Marquard algorithm is mainly applied in the problems of least curve fitting (Saravanan and Nagarajan, 2013).

In mathematics, the Hessian Matrix or Hessian can be defined as square matrix related to a function's second order partial derivative which explains the many variables local curvature of a function.

Hessian matrix's approximation is utilized by Hessian Matrix (Hagan and Menhaj, 1994) in the following update of Newton-like Eq. (9):

$$w_{ij}(t+1) = w_{ij}(t) - \left[J^T J + \mu I \right]^{-1} J^T e \quad (9)$$

where,
J = Jacobian matrix
e = A vector of network errors
μ = A constant

Scaled Conjugate Gradient algorithm (SCG): The Scaled Conjugate Gradient algorithm (SCG), developed by Moller (1990), was designed to avoid the time-consuming line search. This algorithm combines the model-trust region approach (used in the Levenberg-Marquardt algorithm, described in Levenberg-Marquardt), with the conjugate gradient approach. See (Moller, 1990) for a detailed explanation of the algorithm.

Conjugate Gradient Backpropagation with Fletcher-Reeves updates (CGBFR): More detail regarding the Conjugate gradient backpropagation with Fletcher-Reeves updates can be found in (Scales, 1985).

BFGS Quasi-newton Backpropagation (BFGSQB): Newton's method is an alternative to the conjugate gradient methods for fast optimization. In optimization, quasi-Newton methods (a special case of variable metric methods) are algorithms for finding local maxima and minima of functions. Quasi-Newton methods are based on Newton's method to find the stationary point of a function, where the gradient is 0. Newton's method assumes that the function can be locally approximated as a quadratic in the region around the optimum and uses the first and second derivatives to find the stationary point. In higher dimensions, Newton's method uses the gradient and the Hessian matrix of second derivatives of the function to be minimized.

In quasi-Newton methods the Hessian matrix does not need to be computed. The Hessian is updated by analyzing successive gradient vectors instead. Quasi-Newton methods are a generalization of the secant method to find the root of the first derivative for multidimensional problems. In multi-dimensions the secant equation is under-determined and quasi-Newton methods differ in how they constrain the solution, typically by adding a simple low-rank update to the current estimate of the Hessian. More detail regarding the BFGS quasi-Newton backpropagation can be found in The Numerical Algorithms Group (2012).

One-Step Secant backpropagation (OSS): The One Step Secant (OSS) method is an attempt to bridge the gap between the conjugate gradient algorithms and the quasi-Newton (secant) algorithms. This algorithm does not store the complete Hessian matrix; it assumes that at each iteration, the previous Hessian was the identity matrix. This has the additional advantage that the new search direction can be calculated without computing a matrix inverse.

More detail regarding the OSS can be found in (Battiti, 1992).

SIMULATION RESULTS

In this section we evaluate the performance of proposed recognizer. For this purpose we have used the generated patterns (see section of data description). This dataset contains 600 examples of control charts. For this study, we have used 40% of data for training the classifier and the rest for testing.

Performance comparison of different training algorithms with raw data: First we have evaluated the performance of the recognizer with raw data. The training parameters and the configuration of the MLP used in this study are shown in Table 1. The MLP classifiers were tested with various neurons for a single hidden layer and the best networks are selected.

Table 2 shows the Recognition Accuracy (RA) of different systems. In this table, NNHL means the number neurons in the hidden layers. The obtained results are the average of 50 independent runs. As it is depicted in Table 2, using various training algorithms and raw data, the highest accuracy is 98.55%, which is achieved by SCG training algorithms.

Performance comparison of different training algorithms with shape features: Table 3 shows the recognition accuracy of different systems. As it is depicted in Table 3, using various training algorithms and shape feature as input of MLP, the highest accuracy is 99.28%, which is achieved by SCG training algorithms.

Table 1: MLP architecture and training parameter

MLP architecture	Training parameter
The number of layers	2
Number of output neurons	6
Training algorithm	RPROP
	LM
	SCG
	CGBFR
	BFGSQB
	OSS
The initial weights and basis	Random
Activation function (hidden layer)	Tangent-sigmoid
Activation function (output layer)	Linear

Table 2: Recognition accuracy different systems using raw data

Training algorithm	RA (%)	NNHL	Run time (sec)	S.D.
RPROP	97.67	20	2	2.30
LM	96.43	24	18	9.50
SCG	98.55	20	2	1.60
CGBFR	95.64	18	3	12.80
BFGSQB	96.34	32	118	6.76
OSS	97.02	19	2	4.90

S.D.: Standard deviation

Table 3: Recognition accuracy different systems using shape feature

Training algorithm	RA (%)	NNHL	Run time (sec)	S.D.
RPROP	99.06	20	1	1.40
LM	98.52	16	6	2.70
SCG	99.28	20	1	0.90
CGBFR	98.84	18	2	7.40
BFGSQB	98.94	24	45	3.87
OSS	98.54	16	1	2.80

S.D.: Standard deviation

Table 4: A summary of different classification algorithms together with their reported results used measures of the accuracy

Ref. No.	Year	Classifier	RA (%)
Pham and Oztemel	1992	MLP	94.30
Sagiroujlu *et al.*	2000	MLP	93.73
Pham and Oztemel	1994	LVQ	97.70
Pham and Wani	1997	MLP	99.00
Hassan *et al.*	2003	MLP	97.18
Le *et al.*	2004	MLP (RSFM)	97.46
Cheng and Ma	2008	PNN	95.58
Gauri and Chakraborty	2009	MLP	97.22
Addeh *et al.*	2011a	RBF	99.26
Addeh *et al.*	2011b	MLP	99.21
This study	2013	MLP (SCG)	99.28

COMPARISON AND DISCUSSION

For comparison purposes, Table 4 gives the classification accuracies of our method and previous methods applied to the same database. As can be seen from the results, the proposed method obtains an excellent classification accuracy.

CONCLUSION

With the widespread usage of automatic data acquisition system for computer charting and analysis of manufacturing process data, there exists a need to automate the analysis of process data with little or no human intervention. This study presents methods for improving ANN performance in two aspects: feature extraction and ANN training algorithm. The highest level of accuracy obtained by MLP with SCG training algorithm using unprocessed data was 98.55%. The proposed method improves the accuracy up to 99.28% by using shape features as the classifier inputs.

REFERENCES

Addeh, J., A. Ebrahimzadeh and V. Ranaee, 2011a. Application of the PSO-RBFNN model for recognition of control chart patterns. Proceeding of the 2011 2nd International Conference IEEE Conference Control, Instrumentation and Automation (ICCIA), pp: 742-746.

Addeh, J., A. Ebrahimzadeh and V. Ranaee, 2011b. Control chart pattern recognition using adaptive back-propagation artificial neural networks and efficient features. Proceeding of the 2011 2nd IEEE International Conference on Control, Instrumentation and Automation (ICCIA), pp: 747-752.

Battiti, R., 1992. First and second order methods for learning: Between steepest descent and Newton's method. Neural Comput., 4(2): 141-166.

Campbell, C. and N. Cristianini, 1998. Simple learning algorithms for training support vector machines. Technical Report, University of Bristol.

Cheng, Z. and Y. Ma, 2008. A research about pattern recognition of control chart using probability neural network. Proceeding of the ISECS International Colloquium on Computing, Communication, Control and Management, 2: 140-145.

Ebrahimzadeh, A. and V. Ranaee, 2010. Control chart pattern recognition using an optimized neural network and efficient features. ISA T., 49(3): 387-393.

Ebrahimzadeh, A., J. Addeh and Z. Rahmani, 2012. Control chart pattern recognition using K-MICA clustering and neural networks. ISA T., 51(1): 111-119.

Evans, J.R. and W.M. Lindsay, 1988. A framework for expert system development in statistical quality control. Comput. Ind. Eng., 14(3): 335-343.

Gauri, S.K. and S. Chakraborty, 2009. Recognition of control chart patterns using improved selection of features. Comput. Ind. Eng., 56(4): 1577-1588.

Hagan, M.T. and M.B. Menhaj, 1994. Training feed-forward networks with the Marquardt algorithm. IEEE T. Neural Networ., 5(6): 989- 993.

Hassan, A., M.S.N. Baksh, A.M. Shaharoun and H. Jamaluddin, 2003. Improved SPC chart pattern recognition using statistical features. Int. J. Prod. Res., 41(7): 1587-1603.

Haykin, S.S., 1999. Neural Network: A Comprehensive Foundation. 2nd Edn., Prentice Hall, Upper Saddle River, NJ, pp: 842.

Le, Q., X. Goal, L. Teng and M. Zhu, 2004. A new ANN model and its application in pattern recognition of control charts. Proceeding of the 5th World Congress on Intelligent Control and Automation (WCICA, 2004), 2: 1807-1811.

Moller, M., 1990. A scaled conjugate gradient algorithm for fast supervised learning. Computer Science Department, Aarhus University, Vol. 339, DAIMI PB, ISSN: 0105-8517.

Montgomery, D.C., 2008. Introduction to Statistical Quality Control. 6th Edn., Technology and Engineering. John Wiley and Sons Inc., Hoboken, NJ., USA.

Pacella, M., Q. Semeraro and A. Anglani, 2004. Adaptive resonance theory-based neural algorithms for manufacturing process quality control. Int. J. Prod. Res., 42(21): 4581-4607.

Pham, D.T. and E. Oztemel, 1992. Control chart pattern recognition using neural networks. J. Syst. Eng., 2(4): 256-62.

Pham, D.T. and E. Oztemel, 1994. Control chart pattern recognition using linear vector quantization networks. Int. J. Prod. Res., 23: 256-262.

Pham, D.T. and M.A. Wani, 1997. Feature-based control chart pattern recognition. Int. J. Prod. Res., 35(7): 1875-1890.

Ranaee, V. and A. Ebrahimzadeh, 2011. Control chart pattern recognition using a novel hybrid intelligent method. Appl. Soft Comput., 11(2): 2676-2686.

Ranaee, V., A. Ebrahimzadeh and R. Ghaderi, 2010. Application of the PSO-SVM model for recognition of control chart patterns. ISA T., 49(4): 577-586.

Riedmiller, M. and H. Braun, 1993. A direct adaptive method for faster back-propagation learning: the rprop algorithm. Proceeding of the IEEE International Conference on Neural Networks (ICNN, 1993), 1: 586-591.

Sagiroujlu, S., E. Besdoc and M. Erler, 2000. Control chart pattern recognition using artificial neural networks. Turk. J. Electr. Eng. Co., 8(2): 137-146.

Saravanan, A. and P. Nagarajan, 2013. Performance of ANN in pattern recognition for process improvement using levenberg-marquardt and quasi-newton algorithms. IOSR J. Eng., 3(3): 08-13.

Scales, L.E., 1985. Introduction to Non-linear Optimization. Springer-Verlag, New York.

Swift, J.A. and J.H. Mize, 1995. Out-of-control pattern recognition and analysis for quality control charts using lisp-based systems. Comput. Ind. Eng., 28(1): 81-91.

The Numerical Algorithms Group, 2012. Keyword Index: Quasi-Newton. NAG Library Manual, Mark 23. The Numerical Algorithms Group Ltd., Oxford, UK.

Wang, C.H., W. Kuo and H. Qi, 2007. An integrated approach for process monitoring using wavelet analysis and competitive neural network. Int. J. Prod. Res., 45(1): 227-244.

Wani, M.A. and S. Rashid, 2005. Parallel algorithm for control chart pattern recognition. Proceedings of the 4th International Conference on Machine Learning and Applications (ICMLA '05). IEEE Computer Society, Washington, DC, USA, pp: 89-96.