# A Data Mining Approach for Developing Quality Prediction Model in Multi-Stage Manufacturing

**Article** *in* International Journal of Computer Applications · May 2013
DOI: 10.5120/12106-8375



3 authors:


Fahmi Arif
Institut Teknologi Nasional



N. Suryana
Technical University of Malaysia Malacca



Burairah Hussin
Univesiti Teknikal Malaysia, Melaka

International Journal of Computer Applications (0975 – 8887)
Volume 69– No.22, May 2013

# A Data Mining Approach for Developing Quality Prediction Model in Multi-Stage Manufacturing

Fahmi Arif, Nanna Suryana, Burairah Hussin

Faculty of Information and Communication Technology, Univerisiti Teknikal Malaysia Melaka
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

## ABSTRACT
Quality prediction model has been developed in various industries to realize the faultless manufacturing. However, most of quality prediction model is developed in single-stage manufacturing. Previous studies show that single-stage quality system cannot solve quality problem in multi-stage manufacturing effectively. This study is intended to propose combination of multiple PCA+ID3 algorithm to develop quality prediction model in MMS. This technique is applied to a semiconductor manufacturing dataset using the cascade prediction approach. The result shows that the combination of multiple PCA+ID3 is manage to produce the more accurate prediction model in term of classifying both positive and negative classes.

## General Terms
Data Mining, Prediction Model.

## Keywords
Principal Component Analysis, ID3, Quality Prediction, Data Mining, Multi-stage Manufacturing.

## 1. INTRODUCTION
In order to realize the on-line quality monitoring activity, the ability to predict the finished product quality from manufacturing operation condition is required. This ability can be enabled by providing a formulation or mathematical model which can relate the manufacturing operation condition to the product quality [1]. This model is called quality prediction model. Using quality prediction model, process engineers are able to monitor product quality level by evaluating the manufacturing operation.

Recently, various data mining techniques have been employed to develop quality prediction model from manufacturing historical dataset. For example, clustering [2], [3], classification [4–16], association rules [17], [18], and regression have been applied in various industries. These techniques were implemented in injection molding industry, semiconductor manufacturing, slider manufacturing, machining process, hard disk manufacturing, loudspeaker manufacturing, and food processing industry. Most of the prediction models were developed in Single-stage Manufacturing System (SMS).

Recently, multi-stage manufacturing system (MMS) becomes more common in real-world industrial setting [19]. MMS refers to the manufacturing system which involves more than one workstation to produce a complex product [20], [21]. Since customer's taste has become more sophisticated, the complexity in product structure was growing, hence MMS becomes more popular. Various products such as printed circuit board (PCB), semiconductor, automotive products and aerospace device, also need several stages to be produced due to their complex structures [22]. In trying to achieve the faultless manufacturing, quality prediction models are also developed in MMS.

Most of quality prediction model in MMS is developed using SMS approach. In MMS, final product is produced through a series of manufacturing operation performed in several workstations. Therefore, the use of SMS approach to measure quality in MMS can be misleading and ineffective due to the cumulative effect in a workstation as the result of the existence of preceding manufacturing operation in previous workstation [23].

Reference [24] proposed a framework of Cascade Quality Prediction Method (CQPM) for developing quality prediction. However, the accuracy of the prediction model that has been developed using CQPM has not been proved and the techniques that can be employed by this model have not been investigated as well. This study aims to propose a data mining technique developing quality prediction model for MMS based on CQPM.

## 2. RELATED WORKS
In developing quality prediction model in MMS, there are two alternative approaches. First alternative is developing one prediction model for the whole manufacturing line. This approach, called single-point approach, treats manufacturing operation that is performed in every workstation as happened in a workstation. Various data mining technique such as classification [12], [14], clustering [3], and association rules [17], [18] have been employed to develop quality prediction model using this approach.

Another approach is developing one prediction model for every workstation. This approach is called multi-point approach. Using this approach, there will be several prediction models for the whole manufacturing line. Clustering [2], Principal Component Analysis (PCA) [25], and Partial Least Square (PLS) [26], [27] are some of techniques employed to develop prediction model using this approach.

Single-point approach that has been applied using multivariate statistics or data mining techniques assumed that each manufacturing workstation has an independent effect to the product quality level. Moreover, this model has difficulty to reveal the correlation between manufacturing operations from workstation to workstations [2], [25], [26]. From the point of view of partial and total quality as explained by [15], this approach can only explains the partial quality at the last workstation

Multi-point approach is able to model the behavior of a particular workstation. In other word, this approach produced the model that is able to explain the relationship among manufacturing operation variables in a workstation. However, this approach can be misleading and ineffective considering that the measurement of a workstation is probably confounded by the cumulative effect from the previous workstation [23].

Moreover, the individual model for each stage is only able to explain the partial quality for specific workstation.

Beside single-point and multi-point, different approach is proposed by [24]. They proposed a cascade approach to develop quality prediction model for MMS. This method named Cascade Quality Prediction Method (CQPM). At the end, the quality of the final product is represented by the product characteristics. This method is design based on the variable relationships in MMS. Reference [24] explained the condition of variable relationships in MMS as follows:

- Manufacturing operation variables in a workstation are related to each other and influencing the quality of the output from that particular workstation.
- Quality of the output from a workstation are influenced not only by the manufacturing in that particular workstation but also by the output from the previous workstation
- Quality of the final product is influenced by the entire manufacturing operation variables.

Based on that condition, the complex variable relationships in MMS can be explained as follows:

- Relationship among manufacturing operation in a workstation ($R_1$)
- Relationship among workstations ($R_2$)
- Relationship between manufacturing operation variables and final product quality ($R_3$)

CQPM introduces the utilization of latent variables, named product characteristics to model the relationship in every workstation hence the complexity in variable relationship can be reduced. By employing the latent variables named product characteristic ($c_{i,k}$), $R_1$ and $R_2$ are represented by the $c_{i,k}$, $R_3$ is represented by the relationship between $c_{i,k}$ and the product quality level (q). These relationships can be expressed as follows:

$$c_{i,k} = f\left(c_{(i-1),k}, x_{i,j}\right) \quad (1)$$

and

$$q = f\left(c_{n,k}\right) \quad (2)$$

$c_{i,k}$ $k^{th}$ product characteristic of the output from $i^{th}$ workstation
$i = 1, 2, 3, \ldots, n$
$k = 1, 2, 3, \ldots.$

$x_{i,j}$ $j^{th}$ manufacturing operation variable in $i^{th}$ workstation
$j = 1, 2, 3, \ldots$

In order to reveal the Equation (1), without any underlying knowledge of the relationships among $x_{i,j}$, the process of finding the relationships of inter-correlated variables is the same with extracting those variables into some sets of new dimensions. This idea is exactly the same idea with the PCA technique. PCA has been widely used as a method to extracting relevant information from a confusing datasets [28]. Hence it is useful to find latent pattern in high dimensional data [29].

In quality prediction, PCA is used to define the new set of variables by transforming several correlated manufacturing operation variables. PCA is used to develop a prediction model from a historical dataset when product quality data are not available [30]. The product quality is monitored based on the transformed manufacturing operation variables. For example, suppose there are two inter-correlated variables ($x_1$ and $x_2$). Using PCA a set of inter-correlated variables can be transformed into a new set of uncorrelated variables, usually named as principal component (PC), as follows:

$$PC_j = \sum_{i=1}^{n} a_{j,i} x_i \quad (3)$$

where:

$a_{j,i}$ = $j^{th}$ weight of $x_i$ to $PC_j$
$x_i$ = $i^{th}$ inter-correlated variable
$PC_j$ = $j^{th}$ Principal Component

In PCA, $a_{j,i}$ is the constant to be determined. The value of $a_{j,i}$ indicates the amount of contribution of $x_i$ to the $PC_j$. In order to determine $a_{j,i}$, covariance matrix of the involved variables should be calculated first. Then, the eigenvectors of the covariance matrix should be calculated. The eigenvector of the covariance matrix is used as the weight or constant ($a_{j,i}$). Hence, the principal component ($PC_j$) is the linear combination of the original variables ($x_i$) with its eigenvectors [28], [31], [32].

In revealing Equation (2), classification techniques can be used because product quality level is often expressed in category either accepted or rejected while manufacturing operation data might be expressed in numerical or nominal variables. One of those classification techniques is decision tree. Using decision tree, relationship between manufacturing operation variables and product quality level can be expressed in several if-then rules.

According to [33], recently, there are various decision tree algorithms such as Iterative Dichotomiser 3 (ID3), C4.5, Decision Stump (DS), Chi-squared Automatic Interaction Decision (CHAID), Random Tree (RT), and Random Forest (RF). These algorithms are categorized as supervised learning classification algorithm that can be used to build decision tree from the dataset. From a given set of pre-classified cases, these algorithms build diagram to map the attribute values to classes [34]. Those six decision tree algorithms work in similar way, but use different split criteria and different splitting methods.

Amongst many decision tree algorithms, ID3 is the most commonly used [35]. Summarizing from several works [35–39] the advantages and disadvantages of ID3 algorithm can be described as follows:

- Advantages: simple, able to handle large quantities of objects, high speed classification (computing time increases only linearly with the difficulty of the problem), and produces the easy-to-understand classification rules.
- Disadvantages: can only deal with the categorical variables, sensitive to the noise, and miss-classification often happen in handling attribute with too many values.

Based on CQPM, this study proposes a combination of multiple PCA and ID3 algorithm to develop a quality prediction model in MMS. The process applying multiple PCA and ID3 is explained in the following section.

## 3. METHODOLOGY

The process of developing quality prediction model basically is the process of mathematical modeling or rule extraction from manufacturing historical dataset. This process can be illustrated as shown in Figure 1.

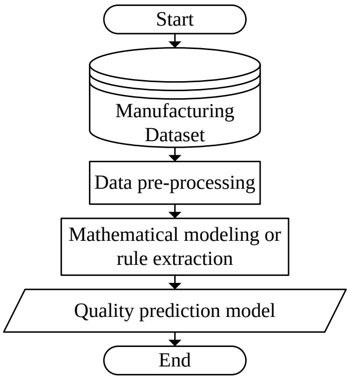


**Fig 1: The process of developing quality prediction model**

Manufacturing dataset from an MMS usually consists of manufacturing operation data and the quality level of some products. Manufacturing operation data usually captured from some monitoring devices or sensors whereas product quality level usually gathered from quality inspection activity. In other word, the manufacturing dataset show the value of manufacturing operation given to a product and the quality level of that product as shown in Figure 2.

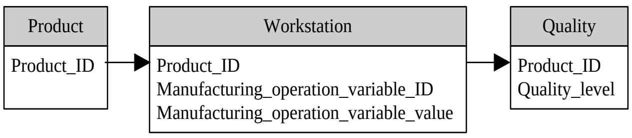


**Fig 2: Data relationship diagram for MMS dataset**

CQPM works to model the quality prediction from the dataset as shown in Figure 2. Using CQPM, basically the process of modeling is the process of revealing Equation (1) for every workstation and revealing Equation (2) at the end. In this study, Equation (1) is modeled using PCA and Equation (2) is modeled using ID3 algorithm. Therefore the process of modeling can be illustrated as shown in Figure 3.

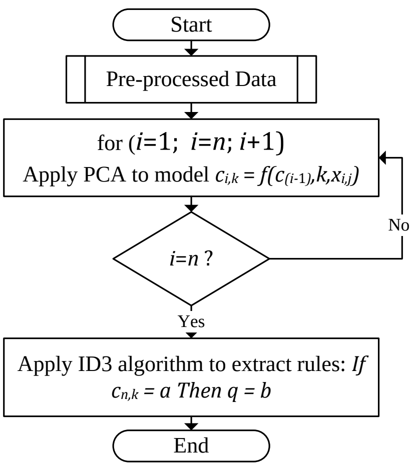


**Fig 3: The process of modeling quality prediction using CQPM**

Using the process as shown in Figure 3, if there is an MMS with two workstations, the process can be illustrated as follows:

- Workstation 1: Manufacturing operation variables in this workstation are interacted each other to change the product characteristics. Using PCA, the relationship between product characteristics and manufacturing operation variables as shown in Equation (1) can be expressed as follows:

$$c_{1,k} = \sum_{j=1}^{m_1} a_{k,j} x_{1,j}$$

$c_{1,k}$ = $k^{th}$ characteristic of the output from $1^{st}$ workstation, $k = 1,2,3,\ldots,m_1$
$a_{k,j}$ = $k^{th}$ weight of $j^{th}$ operation variables
$x_{1,j}$ = $j^{th}$ operation variable in $1^{st}$ workstation, $j = 1,2,3,\ldots,m_1$

- Workstation 2: In workstation 2, product characteristics from the first workstation together with the manufacturing operation variables in this workstation are interacted each other. The product characteristics from this workstation can be expressed as follows:

$$c_{2,k} = \sum_{j=1}^{m_1} a_{k,j} c_{1,k} + \sum_{j=m_1+1}^{m_2} a_{k,j} x_{2,j}$$

$c_{2,k}$ = $k^{th}$ characteristic of the output from $2^{th}$ workstation, $k_2 = 1,2,3,\ldots,m_2$
$c_{1,k}$ = $k^{th}$ characteristic of the output from $1^{st}$ workstation, $k_1 = 1,2,3,\ldots,m_1$
$x_{2,j}$ = $j^{th}$ operation variable in $2^{nd}$ workstation, $j = 1,2,3,\ldots,m_2$
$a_{k,j}$ = $k^{th}$ weight of $j^{th}$ operation variables

This workstation is the last workstation, therefore the product characteristics can be used to explain the final product quality as shown in Equation (2), hence using ID3 algorithm the relationships between product characteristics and product quality level can be expressed as follows:

IF $c_{2,k} = a$ THEN $q = b$

$c_{i,k}$ = $k^{th}$ characteristics of the output from $i^{th}$ workstation, $k = 1,2,3,\ldots$
$q$ = final product quality result
$a$ = class for $c_{n,k}$
$b$ = class for $q$

## 4. CASE STUDY: SEMICONDUCTOR MANUFACTURING

In order to evaluate the performance of multiple PCA+ID3, as well as the CQPM, in developing quality prediction model, a historical dataset from semiconductor manufacturing is applied in a case study. This dataset, namely SECOM dataset [40], consists of manufacturing operation data and the semiconductor quality data. This dataset consists of 590 manufacturing operation variables and 1 quality variable for 1115 instances.

As a real life dataset, SECOM contain of some irrelevant variables and missing value data. A data cleansing procedure discards 452 instances with null and missing values. Regarding the irrelevant variables, since not all 590 sensors were used to gather quality-related data, [40] suggested the simple feature selection technique to select 40 variables that is highly related to the quality variables. These 40 variables are divided into five workstations based on the typical semiconductor manufacturing monitoring process as explained by [41].

As the dataset has been pre-processed, the process as shown in Figure 3 is applied to the dataset. This process is applied using MATLAB and RapidMiner 5 on 2.20 GHz computer with 2.0 GB memory. In the first workstation, four manufacturing operation variables ($x_{1,1}$, $x_{1,2}$, $x_{1,3}$, $x_{1,4}$) are transformed into three product characteristics variables ($c_{1,1}$, $c_{1,2}$, $c_{1,3}$). The number of usable product characteristic variables are selected based on the threshold of "minimum cumulative variance = 0.9".

In the second workstation, manufacturing operation variables, together with the product characteristics of the output from the first workstation, are transformed into product characteristic variables. This process is repeated in every workstation hence the product characteristics for the output of every workstation are produced. The number of involved variables in every workstation is shown in Table 1.

**Table 1. Involved variables in every workstation**

| Work-station | Input Variable | | Output Variable | |
|---|---|---|---|---|
| | No. | Name | No. | Name |
| 1 | 4 | $x_{1,1}$ until $x_{1,4}$ | 3 | $c_{1,1}$ until $c_{1,3}$ |
| 2 | 12 | $c_{1,1}$ until $c_{1,3}$ and $x_{2,1}$ until $x_{2,9}$ | 9 | $c_{2,1}$ until $c_{2,9}$ |
| 3 | 19 | $c_{2,1}$ until $c_{2,9}$ and $x_{3,1}$ until $x_{3,10}$ | 13 | $c_{3,1}$ until $c_{1,13}$ |
| 4 | 21 | $c_{3,1}$ until $c_{1,13}$ and $x_{4,1}$ until $x_{4,8}$ | 17 | $c_{4,1}$ until $c_{4,17}$ |
| 5 | 26 | $c_{4,1}$ until $c_{4,17}$ and $x_{5,1}$ until $x_{5,9}$ | 22 | $c_{5,1}$ until $c_{5,22}$ |

In the fifth workstation, 22 product characteristic variables are produced. ID3 algorithm is applied to reveal the relationships between these variables and the product quality level. Since ID3 algorithm can only deal with nominal variable, the product characteristic variables are categorized into six class from very low to very high using the division of control chart area as explained by [42]. As the result, the relationship between product characteristics and product quality level is represented by 219 rules. The result statistics is shown in Table 2.

**Table 2. Summary of ID3 result**

| | |
|---|---|
| TP | 942 |
| FP | 53 |
| FN | 53 |
| TN | 14 |
| Accuracy | 0.9002 |
| Error Rate | 0.0998 |
| No. of Leaves | 219<br>- 164 positive class (accepted product)<br>- 55 negative class (rejected product) |
| Depth of tree | 6 level |
| No. of Correct Sample | 1115 |
| Computation Time | 0.2945 s |

The result as shown in Table 2 is the results of 10-fold cross validation which usually come as a confusion matrix consist of the number of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) classes. In this case, positive class is representing the accepted product and negative class is representing the rejected product.

Table 2 shows that the accuracy of the model is 90.02%. However, this number cannot explain the actual performance of the model since the dataset is imbalanced. In case of manufacturing quality, imbalanced distribution is the nature of its data, where the proportion of bad product is usually very small [14]. In facing imbalanced dataset, [43] suggest the geometric mean ($G_{mean}$) measurement to explain the performance of decision tree algorithm in classifying both positive and negative classes. For more detail explanation about $G_{mean}$ measurement, ones can refers to [43].

In order to evaluate the performance of multiple PCA+ID3, this combination of technique is compared to other techniques. Multiple PCA+ID3 is compared to C4.5, DS, CHAID, RT, and RF. The comparison result is shown in Table 3.

**Table 3. Comparison of performance of decision tree algorithms**

| | $TP_{rate}$ | $TN_{rate}$ | $FP_{rate}$ | $FN_{rate}$ | $G_{mean}$ |
|---|---|---|---|---|---|
| ID3 | 0.94 | 0.20 | 0.79 | 0.05 | 0.4448 |
| C4.5 | 0.98 | 0.08 | 0.91 | 0.01 | 0.2948 |
| CHAID | 0.98 | 0.05 | 0.94 | 0.01 | 0.2390 |
| DS | 1.00 | 0.00 | 1.00 | 0.00 | 0.0000 |
| RT | 0.99 | 0.00 | 1.00 | 0.00 | 0.0000 |
| RF | 1.00 | 0.00 | 1.00 | 0.00 | 0.0000 |

AS shown in Table 3 The highest $TP_{rate}$ (1.0) is obtained by DS and RF algorithms. RT, CHAID, C4.5 are following with 0.999, 0.9853 and 0.9852. ID3 obtains the lowest (0.9467). Based on the performance in classifying the positive class, the order from the highest to the lowest is: DS and RF, TR, CHAID, C4.5 and ID3. On the other hand, ID3 algorithm achieves the highest $TN_{rate}$ (0.2090), followed by C4.5 (0.0882) and CHAID (0.0580). The lowest (0.0) is obtained by the DS, RT, and RF algorithms. Based on the measurement of $G_{mean}$ as shown in, ID3 algorithm, with $G_{mean}$ = 0.4448, performs better than others. C4.5 and CHAID are following with 0.294 and 0.2390. The others (DS, RT and RF) get 0 for $G_{mean}$. Therefore, based on the performance in classifying the negative class, ID3 works better than others.

Beside the comparison of the techniques used, this study also compares the method to develop quality prediction. CQPM is compared to single-point prediction method. The first comparison model is developed using single-point approach with ID3 algorithm (SP-ID3) whereas the other is developed using single-point approach with PCA+ID3 (SP-PCA+ID3). The comparison result is shown in Table 4.

**Table 4. Comparison of performance of different prediction method**

| Measurement | Prediction Method | | |
|---|---|---|---|
| | CQPM | SP-ID3 | SP-PCA+ID3 |
| $TP$ | 942 | 943 | 900 |
| $FP$ | 53 | 59 | 51 |
| $FN$ | 53 | 69 | 37 |
| $TN$ | 14 | 3 | 4 |
| $G_{mean}$ | 0.9002 | 0.8808 | 0.9113 |

Table 3 shows that SP-ID3, SP-PCA+ID3, and CQPM obtains the different $G_{mean}$. SP-ID3 treats all manufacturing operation variable as having equal contribution to the final product quality level. It assumes that every manufacturing operation variable has individual effect to the final product quality. As the result, final product quality can be directly estimated by evaluating the value of manufacturing operation variable using the extracted rules. Different with those techniques, SP-PCA+ID3 and CQPM are considering the interaction effect of the manufacturing operation variables to the final product quality. Using this method, manufacturing operation variables which are probably considered as not having individual effect to the final product quality by the SP-ID3 method are still taken into account.

SP-PCA+ID3 and CQPM treat manufacturing operation variables differently. SP-PCA+ID3 assume that all manufacturing operation variables are interacted each other in the same time as in single manufacturing system. On the other hand, CQPM employ multiple PCA from workstation to workstation so that the cumulative effect of manufacturing operation variables to the final product quality can be captured. Since CQPM achieve the highest $G_{mean}$, it can be concluded that the model that has been developed using CQPM is performed better in classifying both accepted and rejected class compared to SP-ID3 and SP-PCA+ID3.

## 5. CONCLUSION

The comparison of CQPM and single-point prediction method shows that CQPM is managed to develop the quality prediction model for semiconductor manufacturing. The comparison result shows, CQPM achieve the highest $G_{mean}$. It indicates that CQPM is able to produce a better prediction model compare to single-point method in terms of classifying both majority and minority classes.

In this study, decision tree algorithms are used to reveal the relationship between product characteristics variables and final product quality level. All product characteristic variables as the attribute for the decision tree algorithm have the uniform number of values, which are *very low*, *low*, *lower medium*, *upper medium*, *high*, and *very high*. It can be summarized that in the situation of imbalanced dataset with uniform number of attribute values, ID3 performs better than C4.5 and CHAID, while DS, RT and RF are totally failed in classifying the minority class.

This study shows that combination of multiple PCA+ID can produce better quality prediction model compare to others. However, considering the relatively high accuracy (0.9002) and the relatively low *Gmean*, (0.4448), it can be concluded that the probability of miss-classification in negative class is high. Further improvement in technical level to increase the performance of this method is still possible. Additional technique might be combined to improve prediction model performance for imbalanced dataset.

## 6. REFERENCES

[1] Kano, M., and Nakagawa, Y. 2008. Data-Based Process Monitoring, Process Control, and Quality Improvement: Recent Developments and Applications in Steel Industry. Computers & Chemical Engineering 32 (1–2), 12–24.

[2] Guo, X., Wang, F., and Jia, M. 2005. A Stage-Based Quality Prediction and Control Method for Batch Processes. In Proceeding of the International Conference on Machine Learning and Cybernetics.

[3] Kim, D., Kang, P., Cho, S., Lee, H., and Doh, S., 2012. Machine Learning-Based Novelty Detection for Faulty Wafer Detection in Semiconductor Manufacturing, Expert Systems with Applications. 39 (4), 4075–4083.

[4] Chen, P, and Luo, J. 2008 Data Detection and Pattern Recognition on FMS Control Charts. In Proceedings of the International Conference on Industrial Technology.

[5] Chen, W., Tai, P., Wang, M., Deng, W., and Chen, C. 2008. A Neural Network-Based Approach for Dynamic Quality Prediction in A Plastic Injection Molding Process. Expert Systems with Applications. 35 (3), 843–849.

[6] Ho, W., Tsai, J. , Lin, B., and Chou, J. 2009. Adaptive Network-Based Fuzzy Inference System for Prediction of Surface Roughness in End Milling Process Using Hybrid Taguchi-Genetic Learning Algorithm. Expert Systems with Applications. 36 (2), 3216–3222.

[7] Ho, G. T. S. , Lau, H. C. W. , Lee, C. K. M. , Ip, A. W. H., and Pun, K. F. 2005. An Intelligent Production Workflow Mining System for Continual Quality Enhancement. The International Journal of Advanced Manufacturing Technology. 28(7–8), 792–809.

[8] Ho, G. T. S. , Lau, H. C. W. , Kwok, S. K. , Lee, C. K. M., and Ho, W. 2009. Development of a Co-Operative Distributed Process Mining System for Quality Assurance. International Journal of Production Research. 47(4), 883–918.

[9] Ho, G. T. S. , Lau, H. C. W. , Chung, S. H. , Fung, R. Y. K., Chan, T. M. , and Lee, C. K. M. 2008. Fuzzy Rule Sets for Enhancing Performance in A Supply Chain Network. Industrial Management & Data Systems. 108 (7), 947–972.

[10] Hua, D. , Shi-yuan, Y., and De-hui, W. 2007. Intelligent Prediction Method for Small-Batch Producing Quality Based on Fuzzy Least Square SVM. Systems Engineering-Theory & Practice. 27(3).

[11] Huang, Y., Cheng, F., and Hung, M. 2009. Developing A Product Quality Fault Detection Scheme. In Proceedings of International Conference on Robotics and Automation.

[12] Johnston, A.B., Maguire, L.P., and McGinnity, T. M. 2009. Downstream Performance Prediction for A Manufacturing System Using Neural Networks and Six-Sigma Improvement Techniques. Robotics and Computer-Integrated Manufacturing. 25(3), 513–521.

[13] Melin, P., and Castillo, O. 2007. An Intelligent Hybrid Approach for Industrial Quality Control Combining Neural Networks, Fuzzy Logic and Fractal Theory. Information Sciences. 177 (7), 1543–1557.

[14] Rokach, L., and Maimon, O. 2006. Data Mining for Improving the Quality of Manufacturing: A Feature Set Decomposition Approach. Journal of Intelligent Manufacturing. 17 (3) 285–299.

[15] Xu, X. 2004. Fuzzy Control for Manufacturing Quality Based on Variable Precision Rough Set. In Proceeding of the Fifth World Congress on Intelligent Control and Automation.

[16] Yu, J., and Xi, L. 2009. A Neural Network Ensemble-Based Model for On-Line Monitoring and Diagnosis of Out-of-Control Signals in Multivariate Manufacturing

Processes. Expert Systems with Applications. 36 (1), 909–921.

[17] Chen, W., Tseng, S., and Wang, C. 2005. A Novel Manufacturing Defect Detection Method Using Association Rule Mining Techniques. Expert Systems with Applications. 29 (4), 40–52..

[18] Lau, H., Ho, G., Chu, K., Ho, W., and Lee, C. 2009. Development of An Intelligent Quality Management System Using Fuzzy Association Rules. Expert Systems with Applications. 36 (2), 1801–1815.

[19] Kaya, I., and Engin, O. 2007. A New Approach to Define Sample Size at Attributes Control Chart in Multistage Processes: An Application in Engine Piston Manufacturing Process. Journal of Materials Processing Technology. 183 (1), 38–48.

[20] Li, J. T., Freiheit, J., Hu, S. J., and Koren Y.. 2007. A Quality Prediction Framework for Multistage Machining Processes Driven by an Engineering Model and Variation Propagation Model. Journal of Manufacturing Science and Engineering. (129. (6), 1088–1100, 2007.

[21] Huang, Q., and Shi, J. 2004. Stream of Variation Modeling and Analysis of Serial-Parallel Multistage Manufacturing Systems. Journal of Manufacturing Science and Engineering, 126 (3), 611.

[22] Tsung, F., Li, Y., and Jin, M. 2008. Statistical Process Control for Multistage Manufacturing and Service Operations: A Review and Some Extensions. International Journal of Services Operations and Informatics. 3 (2), 449–452.

[23] Zou, C., and Tsung, F. 2008. Directional MEWMA schemes for multistage process monitoring and diagnosis. Journal of Quality Technology. 40 (4), 407–427.

[24] Arif, F., Suryana, N., and Hussin, B. 2013. Framework of Cascade Quality Prediction Method Using Latent Variables for Multi-Stage Manufacturing. International Journal of Management Theroy and Application. 1 (1).

[25] Zhao, C., Wang, F., Lu, N., and Jia M. 2007. Stage-Based Soft-Transition Multiple PCA Modeling and On-Line Monitoring Strategy for Batch Processes. Journal of Process Control. 17 (9), 728–741

[26] Qi, Y., Wang, P., and Gao, X. 2011. Enhanced Batch Process Monitoring and Quality Prediction Using Multi-Phase Dynamic PLS, Control Conference (CCC).

[27] Ge, Z., Zhao, L., Yao, Y., Song, Z., and Gao, F., Utilizing Transition Information in Online Quality Prediction of Multiphase Batch Processes, Journal of Process Control.

[28] Shlens, J., 2003, A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition, 2003. [Online]. Available: http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf. [Accessed: 01-Dec-2011].

[29] Salahshoor, K., Alaei, H. K., and Alaei, H. K., 2010. A New On-line Predictive Monitoring Using Integrated Approach Adaptive Filter and PCA. In Proceeding of International Workshop on Soft Computing Applications.

[30] Bersimis, S., Psarakis, S., and Panaretos, J. 2007. Multivariate Statistical Process Control Charts: An Overview, Quality and Reliability Engineering International. 23 (5), 517–543.

[31] Abdi, H., and Williams, L.J. 2010. Principal Component Analysis, Wiley Interdiciplinay Reviews: Computational Statistics. New York: John Wiley & Sons.

[32] Smith, L.I.. 2002. A Tutorial on Principal Components Analysis, [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

[33] Akthar, F., and Hahne, C., 2012. RapidMiner 5: Operator Reference. Rapid-I GmbH..

[34] Ture, M., Tokatli, F., and Kurt, I. 2009. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. Expert Systems with Applications. 36 (2), 2017–2026.

[35] Shrivastava, S. K., and Tantuway, M. 2011. A Decision Tree Algorithm Based on Rough Set Theory after Dimensionality Reduction. International Journal of Computer Applications. 17(7), 29–34.

[36] Anyanwu, M.N., and Shiva, S.G. 2009. Comparative Analysis of Serial Decision Tree Classification Algorithms. Journal of Computer Science and Security, 3, 230–240.

[37] Du, J., Cui, H., H. Li, H., and Zhao, Z. 2011. Fault Diagnosis of Vacuum Circuit Breakers Based on ID3 Method. In Proceeding of the 1st International Conference on Electric Power Equipment.

[38] Jin, C., De-lin, L., and Fen-xiang, M. 2009. An Improved ID3 Decision Tree Algorithm. In Proceeding of the 4th International Conference on Computer Science & Education.

[39] Jingjun, F., and Shuting, Y. 2009. Alumina Production Operations Management Information System Based on Data Mining Technology. In Proceeding of the International Forum on Computer Science-Technology and Applications.

[40] McCann, M. and Johnston, A., 2008. Secom Dataset, UCI Machine Learning Repository, [Online]. Available: http://archive.ics.uci.edu/ml/datasets.html.

[41] May, G.S., and Spanos, C.J. 2006. Fundamentals of Semiconductor Manufacturing and Process Control. New Jersey: John Wiley & Sons

[42] DeVor, R.E., Chang, T., and Sutherland, J.W. 2007. Statistical Quality Design and Control: Contemporary Concepts and Methods, 2nd ed. Prentice-Hall.

[43] García, S., Fernández, A. and Herrera, F., 2009) Enhancing the Effectiveness and Interpretability of Decision Tree and Rule Induction Classifiers with Evolutionary Training Set Selection Over Imbalanced Problems. Applied Soft Computing. 9 (4), 1304–1314..