

Review of data mining applications for quality assessment in manufacturing industry: support vector machines

Hamidey Rostami^{1,2,*}, Jean-Yves Dantan¹, and Lazhar Homri¹

¹ LCFC, Arts et Métiers Paris Tech, HESAM, ENIM Metz, France

² Université de Lorraine, Ile du Saulcy, Metz, 57045, France

Received: 26 May 2015 / Accepted: 8 September 2015

Abstract. In many modern manufacturing industries, data that characterize the manufacturing process are electronically collected and stored in databases. Due to advances in data collection systems and analysis tools, data mining (DM) has widely been applied for quality assessment (QA) in manufacturing industries. In DM, the choice of technique to be used in analyzing a dataset and assessing the quality depend on the understanding of the analyst. On the other hand, with the advent of improved and efficient prediction techniques, there is a need for an analyst to know which tool performs better for a particular type of dataset. Although a few review papers have recently been published to discuss DM applications in manufacturing for QA, this paper provides an extensive review to investigate the application of a special DM technique, namely support vector machine (SVM) to deal with QA problems. This review provides a comprehensive analysis of the literature from various points of view as DM concepts, data preprocessing, DM applications for each quality task, SVM preliminaries, and application results. Summary tables and figures are also provided besides to the analyses. Finally, conclusions and future research directions are provided.

Keywords: Data mining, quality assessment, manufacturing industry, support vector machine

1 Introduction

In order to meet the unique needs, improve the customer satisfaction, enhance investor confidence, ensure effective management and efficient administration, quality assessment (QA) needs to be considered and penetrate in all sectors of each manufacturing organization. Due to the increasing expectations of customers, quality must be continually improved and new processes of conformity assessment should be adopted in manufacturing organizations regarding to the products and services they have to offer. It is noteworthy that high quality never happens by chance; it evolves over some time as a result of experience. A vital question for industry to be addressed is how a manufacturer can improve its product quality and yield. This question arises repeatedly during new product development, in which a faster and more systematic way is needed [1]. Although using precise first-principle models is the most reliable approach to improve the quality, not only there is no developed model for most newly developed processes, but also modeling of a complex industrial process is very difficult and timely inefficient. Particularly, it is difficult for manufacturers to provide a precise first-principle model that can optimally explain why defects appear in products.

An effective way to solve this difficult problem is using of operation data. Recently, data-based approaches have

been widely accepted for QA in various industries. Many organizations have built integrated data-bases to store many operation data from production sectors and product features data from inspection. This storage involves several input and output variables with a huge volume of data that are not possible for accurately being modeled and/or optimized. Accordingly, data mining (DM) and knowledge discovery from databases (KDD) have been successfully applied for solving QA and control problems [2].

In order to select a proper technique and method to mine quality data and extract valuable knowledge in assessing production systems, manufacturers need to first know about the quality tasks in manufacturing. This knowledge provides how the input quality parameters are related to the final outputs and how to predict the output quality from the input quality parameters; how to classify the quality of the products to find most important affecting parameters; and how to optimize the input to obtain optimal output. The knowledge is gained by reviewing the QA tasks in manufacturing. Accordingly, four criteria: *quality of product and process*, *quality prediction*, *quality classification*, and *parameter optimization*, are described in this paper. The KDD and DM techniques which help manufacturer to extract the knowledge have been reviewed. Since several DM techniques have been developed in the literature and finding the most proper one is a challenging work, this paper suggests an efficient DM technique, namely support vector machine (SVM) to the

* Correspondence: hamideh.rostami@ensam.eu

manufacturer handle QA issues. Researchers report that SVM show superior predictive power to other data mining methods and techniques used in the domains [3]. Its importance lies in the need to develop predictive models from a reduced, minimal number of input variables which best summarize the overall input data resulting in maximal predictive power. Dimensionality reduction is the process undertaken in order to reduce the number of independent variables utilized within a data mining exercise. The variables within the dataset are examined to see if and how they relate to, and influence other variables. A SVM-based technique has been selected due to high classification rates and to their high generalization abilities.

According to the above-enumerated issues, the rest of the paper is organized as follows. Section 2 describes QA in manufacturing including the role of QA in manufacturing and different quality related criteria that should be described through the production system such as *quality of product and process*, *predicting quality*, *classification of quality*, *parameter optimization* and *key parameter extraction*. For better understanding of how KDD and DM techniques are utilized to handle elaborated QA tasks in Section 2, Section 3 explains the application of KDD and DM techniques in solving QA issues. For suggesting a proper DM technique to manufacturer to handle QA issue, Section 4 discusses the SVM technique and its features including soft and hard margins. Section 5 address a comprehensive review of different papers applying SVM in their work. Finally, the conclusion and further research directions are provided in Section 6.

2 QA in manufacturing

This section provides an overview of QA tasks in manufacturing in order to better discover the production system and how QA positively contributes to the quality of the final products. First, a general introduction of QA application in manufacturing is provided. Next, as mentioned in Section 1, four criteria, namely *quality of product and process*, *quality prediction*, *quality classification* and *quality optimization* are described in this section.

Nowadays in many modern manufacturing organizations, data are electronically collected and stored to characterize the manufacturing process. In these organizations, data mining techniques can be utilized to discover interesting and useful rules in the manufacturing processes. These rules can be subsequently extracted to enhance the whole manufacturing process in areas such as defect prevention and detection, reducing flow-time, increasing safety, etc. [4]. There are several studies in the literature that examine the implementation of data mining tools in manufacturing, while this paper focuses on the application of data mining techniques in QA through manufacturing processes.

In QA of manufacturing systems, the most important task is to collect and analyze the data to solve quality related problems. Quality assessment programs such as design for six sigma (DFSS), six sigma (6σ) and kaizen

are encouraging quality managers to collect data to attack quality problems, while advances in automation and computer systems, data from manufacturing processes is becoming more and more available [5]. Although traditionally several analysis tools have been used in QA, but better techniques have been developed and now are used to discover knowledge from massive data sets.

Nowadays quality problems involve several input and output variables that are not possible for accurately being modeled and/or optimized. Accordingly, DM and KDD have been successfully applied for solving QA and control problems [2]. Although classification of DM techniques for a particular application area is challenging, DM techniques have been typically classified as data summarization, classification, prediction, clustering and so on. However, these classes highly depend on the type of knowledge that can be discovered in databases [6,7]. In this report, DM techniques are classified based on [5].

In the following, a classification of QA tasks within the scope of this paper is given first. Though many QA tasks take place throughout the life cycle of a product, Phadke [8] classified the product life cycle process into three phases: *product and manufacturing process design*, *manufacturing of the product*, and *usage of the product by the customer*. In this report, the main focus is on certain QA tasks occurring during the first two phases.

Through product and manufacturing process design phase, three stages are typically involved as: (1) conceptual design of the products/processes; (2) setting values of their design parameters; and (3) setting tolerances for the design parameters [2,5,8].

Regarding to QA activities during the product manufacturing phase, following functions can be introduced as: (1) determining factors that significantly impact quality; (2) modelling relationships between input and output characteristics of quality; and (3) predicting quality levels for a given set of input parameters [5]. Such quality analyses not only provide invaluable feedback to product/process design phase, but also can be used as corrective actions for QA. Accordingly, due to ease of collecting relevant data, most of DM applications are expected to occur in the phase of product manufacturing [5].

The final phase of product development process considers the usage of the product by the customers. In this situation, quality perception can be improved through repair or replacement of the product under warranty, and other services beyond warranty [5]. We have excluded DM applications involving QA at after-sales services.

Broadly speaking, the QA and control activities that we have considered in this paper include some tasks (see Fig. 1) in manufacturing phase of the product life cycle as follows [5]:

- *Describing the quality of product and process*: quality of products or processes can be defined or characterized by performing the following tasks:
 - identifying factors and variables that significantly affect quality;
 - ranking the factors and variables based on their significance;

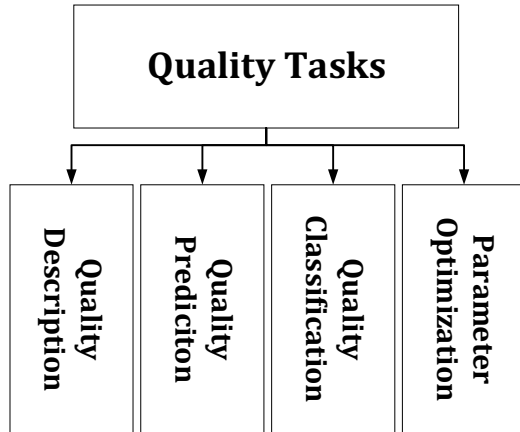


Fig. 1. The position of our work in quality tasks.

- identifying how low, medium and high yielding products can be typically grouped in data, and finding the most effective factor(s) that distinguishes low and high yielding products.
- *Predicting quality*: in production systems that quality of output products can be stated as real numbers, quality prediction involves developing models that relate input characteristics of quality to the output, and using such models to predict what the resulting quality characteristic value will be for a given set of input parameter values.
- *Classification of quality*: this activity includes classifying the interested quality characteristic such as nominal, binary or ordinal outputs (such as defects) into a set of given classes of the quality using a set of input parameters.
- *Parameter optimization*: after finding the most important factors that significantly affect the cases yielding high quality, parameter optimization can be introduced as finding optimal levels of these process/product-related factors.

After describing the QA tasks in manufacturing, now is the time to know about different methods and techniques that can be utilized to extract knowledge for the above-mentioned criteria. These methods allow studying the field of deployment of SVM. For this aim, the following Section 3 explains these methods.

3 KDD in QA

This section provides a general introduction on KDD and DM techniques to handle enumerated QA tasks in Section 2. First, KDD and DM are described; next, different criteria of KDD and DM techniques are elaborated.

Nowadays, the amount of data stored in databases is increasing at a tremendous speed. Since data has become one of the most valuable assets of a business, organizations attempt to collect all kinds of information about the business process such as financial, payroll, and customer data, while in some specific communities, a single experiment can lead to terabytes of data. However, even storing,

let alone analyzing, such huge amounts of data presents many new obstacles and challenges for us when we are drowning in data, and yet starving for knowledge.

This growing need leads to birth of a new research field called Knowledge Discovery from Data (KDD) or Data Mining (DM), which has attracted attention from researchers in many different fields including database design, statistics, pattern recognition, machine learning, and data visualization. Among different research fields, industrial engineers whose duties are to optimize processes in order to add more values to the system, prevent loss, and make future uncertainties ineffective, data-mining plays as a key and powerful tool for evaluating and making the best decisions based on stored data. However, the potential of data-mining for industrial managers has yet to be fully undiscovered.

Organizations always encounter needs of planning for the future and estimating the future values of business variables has become a necessity for their survival. These organizations require prediction or forecasting of supply, sales, and demand for production planning, marketing and financial decisions [9]. In production and manufacturing systems, we face with variety of issues such as supply-chain management, human factors, cost estimations, demand forecasting, material requirements and enterprise resource planning, process optimization, scheduling, sequencing, cell organization, and quality control, but the knowledge of data mining tools that could reduce the common nightmares in these areas is not widely available. In the following, detail explanations are provided regarding data mining definitions and its tools.

3.1 General data mining process

Data Mining is the process of analyzing data from different perspectives and summarizing the results as useful information. Frawley et al. [10], Fayyad et al. [11], Kantardzic [12], Gorunescu [13], Romero and Ventura [14], Larose [15] and McCue [16] have defined DM as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data”. The DM’s process utilizes machine learning, statistics, and visualization techniques to discover and present comprehensible knowledge. Moreover, the word “Knowledge” in KDD means the investigating and figuring out patterns which are extracted from the processed data, in which a pattern is an expression describing facts in a subset of the data. Thus, the difference between KDD and data mining is that “KDD refers to the overall process of discovering knowledge from data while data mining refers to application of algorithms for extracting patterns from data without the additional steps of the KDD process” (see [11, 14–16]). These two different terms are used interchangeably by most researchers, although DM is a crucial and important part of KDD process. Accordingly, Figure 2 shows the overall steps of the KDD.

According to Figure 1, the process of KDD can be divided into three main steps as Table 1, the first step is to pre-process data that includes several activities such as

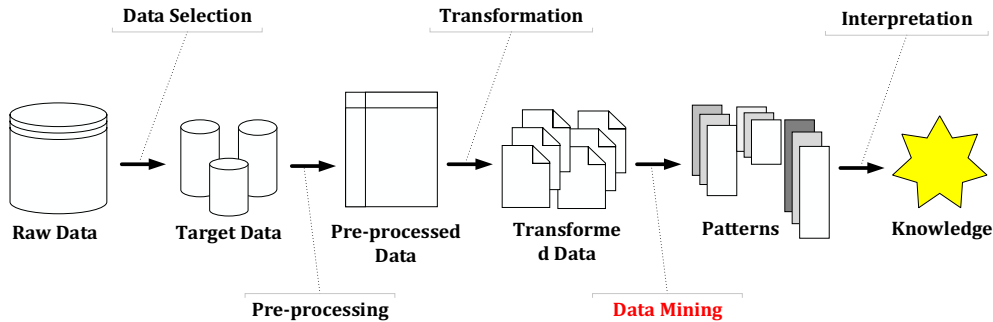


Fig. 2. KDD Process' steps.

Table 1. KDD's main steps.

KDD's Main Step	Details
I. Data Preprocessing	<ul style="list-style-type: none"> ✓ Data preparation ✓ Data reduction ✓ Noise elimination
II. Data Mining	<ul style="list-style-type: none"> ✓ Data-Mining techniques
III. Data Post-processing	<ul style="list-style-type: none"> ✓ Result interpretation ✓ Knowledge extraction

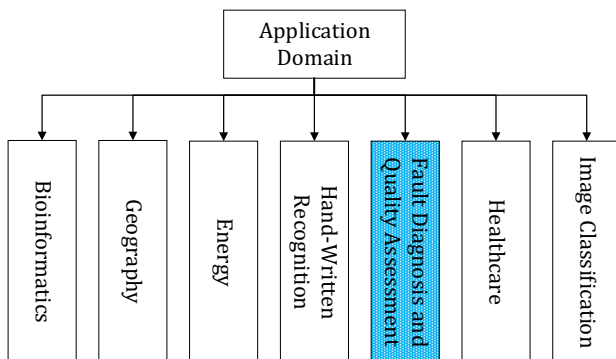


Fig. 3. KDD application in quality domain.

data collection, data smoothing, data cleaning, data transformation and data reduction. The second step, normally called data mining (DM), involves data modeling and prediction. Some researchers have also considered data classification and prediction as activities of DM, in which classification methods encompass several sub-activities such as database segmentation, clustering, deviation detection and etc., while prediction methods are consisting of (1) mathematical modeling approaches such as linear and nonlinear scoring (neural nets), and advanced statistical methods including the multiple adaptive regression by splines; (2) distance methods involving the nearest-neighbor approach; (3) logic-based methods like decision trees. The third step is data post-processing that is interpreting the information resulted from DM tools, then concluding or transforming this information into valuable knowledge. These steps and their sub-activities have been schematically shown in Figure 3.

The main goals of data mining are detection, interpretation, and prediction of qualitative and/or quantitative patterns in data. To achieve these goals, data mining so-

lutions employ a wide variety of methods to both predict and describe interpretable and valuable information. Since the importance of description and prediction can vary between different applications, they can be fulfilled by any of data mining tasks including: classification, clustering, regression, summarization, deviation detection, and dependency modeling [13, 17, 18]. A brief explanation for different tasks of descriptive data mining can be provided:

- *Classification* – Classification is a data mining (machine learning) method used to predict group membership for data instances ([12–16, 19, 20]).
- *Regression* – Considering either a linear or nonlinear model of dependency, regression is to predict a value of a given continuously valued variable based on the values of other variables [12–16, 19].
- *Deviation Detection* – Through deviation detection, the most significant changes in data from previously measured or normative values are discovered [11, 14–16].
- *Clustering* – Clustering is a division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar to one another and dissimilar to objects of other groups [13, 15, 16, 21].
- *Summarization* – Summarization is to find a concise description for a subset of data [11, 13, 16].
- *Dependency modeling* – In some datasets, there are significant dependencies between variables that dependency modeling task is to find a model that describes significant dependencies between these variables [13, 15, 16, 21].

3.2 DM application in QA

Since KDD has been applied in various domains as Figure 4, this paper mainly focus on the domain of quality data. In this section, we discuss different papers from various points of view as elaborated in Section 3. Data pre-processing before performing the DM techniques in the quality tasks are discussed first. Then, DM in each of the quality tasks (i.e., see Sect. 2) in manufacturing systems are reviewed. It should be noted that the third step (post-processing) of KDD will not be reviewed in this section.

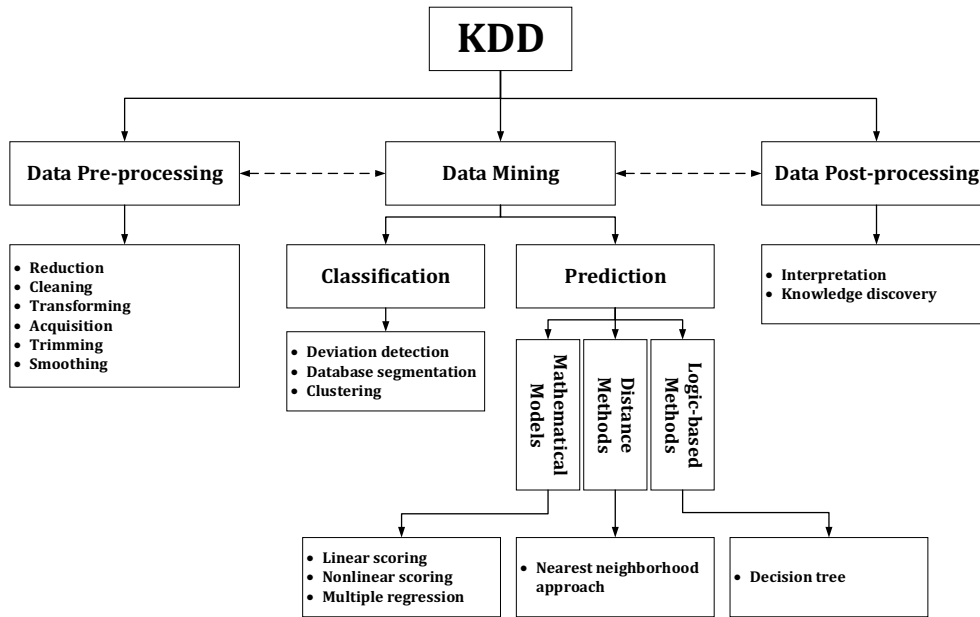


Fig. 4. KDD main process and sub activities.

3.2.1 Pre-processing of quality data

Two steps that are highly necessary before performing DM techniques are data preparation and data preprocessing. Although there is less attention regarding these steps in KDD process in the literature. Below, some explanations are provided regarding data preparation and preprocessing steps [22].

Quality data preparation

Through reviewing different QA-related papers in the literature, it can be understood that data have been collected by statistically designed experiments (DOE) such as 2- and 3-level full factorials, fractional factorials, Taguchi orthogonal designs, and so on. Another way to prepare data is done through observation. This method is also common in the literature and these data are obtained during online sensor measurements or from historical logs. Alternatively, process simulations and simulated data sets also existed and these were used to test the performance of the algorithms used [5, 22, 23].

Quality data pre-processing

Data pre-processing techniques have been performed in order to purify datasets from missing values, outliers, incomplete and inconsistent data. Most of the purifying techniques are applied to the historical-observational type data sets. On the other hand, data transformation could be a part of data pre-processing step, in which normalization of the data has been the most common transformation technique rather than other techniques such as smoothing, transforming categorical data to numerical, and adjusting abnormal data, etc. another technique of data pre-processing is data reduction. In data reduction step, dimension reduction is the most form used through variety

of techniques like: Subjective Evaluation (SE), Analysis of Variance (ANOVA) and Artificial Neural Network (ANN). Subjective Evaluation SE is based on the subjectivity (understandability) of the user who evaluates the patterns on the basis of novelty, actionability. . . ANOVA is a statistical method used to test differences between two or more means. ANOVA is a particular form of statistical hypothesis heavily used in the analysis of experimental data. A statistical hypothesis test is performed for making decisions using data. A test result (calculated from the null hypothesis and the sample) is called statistically significant if it is deemed unlikely to have occurred by chance, assuming the truth of the null hypothesis. A statistically significant result, when a probability is less than a threshold (significance level), justifies the rejection of the null hypothesis. ANN are computational models that consist of a number of simple processing units that communicate by sending signals to one another over a large number of weighted connections.

In addition, data compression techniques such as Principal Component Analysis (PCA) to obtain manageable sizes of data have been also commonly utilized for data reduction [5, 22, 23]. PCA, used for dimension reduction, is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

3.2.2 DM in quality data

The number of DM applications in quality data, has been increasing for the last 10 years. This section provides a brief description of some studies regarding the application of DM in the quality tasks mentioned in the beginning of

this section. Interested readers are referred to [5, 22, 23] for comprehensive review from 2007 to date.

Describing the quality of product and process

Gardner and Bieker [24] studied the application of data mining in semiconductor wafer manufacturing problems through combination of self-organizing neural networks and rule induction to identify the critical poor yield factors from normally collected wafer manufacturing data. Karim et al. [25] modified the original Growing self-organizing map (GSOM) for manufacturing yield improvement through introduction of a clustering quality measure to evaluate the performance of the program in separating good and faulty products and a filtering index to reduce noise from the dataset. Results show that the proposed method is able to effectively differentiate good and faulty products. Huang et al. [26] explained that finding the sources of electromagnetic noise is a time-consuming process and they developed an attribute selection and fault diagnosis based on the advantage of rough set theory (RST) which is a novel data mining approach for dealing with vagueness and uncertainty.

Liu and Chien [27] developed a knowledge-based intelligent system for WBM defect diagnosis for yield enhancement in wafer fabrication. Their proposed system consists of three parts as: graphical user interface, the wafer bin maps (WBM) clustering solution, and the knowledge database. In particular, their developed WBM clustering approach integrates spatial statistics test, cellular neural network (CNN), adaptive resonance theory (ART) neural network, and moment invariant (MI) to cluster different patterns effectively. In a similar work, Chien et al. [28] developed a manufacturing intelligence solution that integrates spatial statistics and neural networks for the detection and classification of WBM patterns to construct a system for online monitoring and visualization of WBM failure percentages and corresponding spatial patterns with an extended statistical process control chart.

Data mining techniques have been also applied in metal product manufacturing industry in order to describe product and process quality, where product/process attributes or variables affecting quality are determined, ranked based on significance, or grouped based on similarities. In packaging manufacturing, De Abajo et al. [29] developed a tinplate quality diagnostic model utilizing SE, genetic algorithm (GA), SOM, and decision tree (DT). In addition, in a cutting process, Chang and Jiang [30] applied PCA for dimension reduction, and for a hot rolling process.

Predicting quality

Several papers have been published with the goal of quality prediction in computer and electronic product manufacturing industry. Chen et al. [31] employed ANN in a chemical vapor deposition process to establish a predictor of product quality. Yang et al. [32] used ANN to propose a quality control system for the solder stencil printing process. Similarly, Shi et al. [33] used also ANN for

modelling nonlinear cause-effect relationships in printed circuit board (PCB) manufacturing, whereas general regression neural networks (GRNNs) were used in [34] for predictions in a spot welding process.

In metal product manufacturing industries, some of the applications of DM techniques are as follows. Sung et al. [35] used well-known fuzzy regression (FR) to predict the bead geometry in robotic welding process that deals with a high degree of fuzziness and also to build the relationship between four process variables and the four quality characteristics, respectively. DM techniques have been also applied to predict quality in other industries such as plastics manufacturing industries [36, 37], chemical process industries [38], and healthcare [39].

Classification of quality

DM has been extensively applied in different fields of manufacturing in order to classify the quality characteristics or predicting the class of quality. Several papers have been published with the goal of quality classification in computer and electronic product manufacturing industry. As an example, Chen et al. [31] established a quality predictor for analyzing the relationship between manufacturing process parameter setting and final product quality in the plasma-enhanced chemical vapor deposition (PECVD) of semiconductor manufacturing by applying the back-propagation neural network (BPNN) algorithm and Taguchi method.

In metal manufacturing, quality classification has not been applied as much as quality prediction. For instance, Wang [40] discussed the nature and implications of DM techniques in manufacturing and their implementations on product design and manufacturing. In addition, Hou et al. [41] employed RST together with fuzzy set theory (FST) and ANN in conveyor belts manufacturing for rule induction. Regarding to other application of DM techniques in quality classification, DT was used in ultra-precision manufacturing [42]. An integration of SVM and DT has been considered for continuous stirred tank reactor [43]. Sarimveis et al. [44] considered ANN and FST for classifying the product quality in paper manufacturing.

Parameter optimization

In computer and electronic product manufacturing industry, Hung [45] combined the Taguchi method with GA and ANN to use them as the optimal tool in wire bond designing parameters for an ultra-thin chip scale package (CSP), and then construct a set of the optimal parameter analysis flow and steps.

In metal manufacturing, Hamed et al. [46] considered optimization of three important process parameters in the spot welding of the body components, namely welding current, welding time, and gun force. They first experimentally investigated the effects of these parameters on deformation of the sub-assemblies. Next, neural networks and multi-objective genetic algorithms were utilized to select the optimum values of welding parameters

that yield the least values of dimensional deviations in the sub-assemblies.

In other application, Nedic et al. [47] presented a novel approach for modelling traffic noise that is based on optimization. They also applied four optimization techniques in simulation of their work including: genetic algorithms, Hooke and Jeeves algorithm, simulated annealing and particle swarm optimization. They also compared their work with classical statistical model, and superior capabilities of proposed model were demonstrated. In addition, Shen et al. [37] determined the optimal values of process parameters in plastic injection molding. Chiang et al. [48] used ANN for optimizing multiple quality characteristics of a polymerization process. Ho et al. [49] considered ANN together with fuzzy rule sets (FRS) and expert systems (ESs) in a slider manufacturing process. Manimala et al. [50] combined Wavelet Packet Transform (WPT) and Support Vector Machines (SVM) to classify the power quality events. In their model, the features of the disturbance signals are extracted using WPT and then given to the SVM for effective classification.

Based on the literature review of KDD and DM application in quality data, most of researchers have demonstrated the efficiency and applicability of SVM in handling quality tasks. Therefore in the following, this paper provides a comprehensive review of SVM's features and its application in QA issues.

4 Support vector machine – SVM

The Support Vector Machine (SVM) is an advanced classification method introduced by Boser et al. [51]. The SVM classifier has been widely employed in different disciplines specifically bioinformatics due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data [52].

Although the SVM has been developed for solving linearly separable problems, it is possible to generalize it in order to classify non-linear problems by employing the kernel method [53]. The kernel method is an algorithm that maps the original data points into a higher dimensional feature space in which they can be separated by a linear classifier. The projection of a linear classifier on the feature space is a nonlinear one in the original (input) space [54]. According to [55], using kernel method has two advantages including: (1) the ability to generate non-linear decision boundaries using methods designed for linear classifiers; and (2) the use of kernel functions allows the user to apply a classifier to data that have no obvious fixed-dimensional vector space representation.

In order to use SVM effectively, it is required to understand how it functions. Through training an SVM, the user needs to make a number of decisions consisting of: how to preprocess the data, what kernel to use, and setting the parameters of the SVM and the kernel. In the following, the preliminaries of SVM linear classifier and kernel method are briefly explained.

4.1 Preliminaries: linear vs. nonlinear classifiers

This section explains the preliminaries of SVM and why SVMs are in the category of linear two-class classifiers. In the data for a two class learning problem, objects are labeled with two labels corresponding to the two classes; for convenience we assume the label +1 for positive examples and the label -1 for negative examples. The boldface X denotes a vector with components x_i . Let X_i denote the i th vector in a dataset $\{(X_i, y_i)\}_{i=1}^n$, where y_i is the label associated with X_i . Accordingly, the objects X_i are called patterns or examples [56]. Besides, the set of all examples is shown as P . It is noteworthy that assuming the examples as vectors will be relaxed once the kernel methods are employed, in which the examples could be any kind of continuous/discrete objects.

Through defining a linear classifier, a key concept is the dot product between two vectors (i.e., inner product or scalar product) which is defined as $W^T X = \sum_i w_i x_i$. Accordingly, a linear classifier can be proposed as a linear discriminant function (1):

$$f(x) = W^T X + b, \quad (1)$$

where W and b are the weight vector and the bias, respectively. In a special case when $b = 0$ since we have $W^T X = 0$, the all points X are perpendicular to W and go through the origin. That is, a line in two dimensions, a plane in three dimensions, and more generally, a hyperplane. The bias b translates the hyperplane deviation from the origin [56]. The following hyperplane (2) divides the space into two different regions and one region for one class as shown in Figure 5:

$$\{X : f(x) = W^T X + b = 0\}. \quad (2)$$

In addition, the sign of the discriminant function $f(x)$ for a specific point X_i denotes the side of the hyperplane that X_i belongs to (see also Fig. 5). The boundary between regions classified as positive and negative is called the decision boundary of the classifier. A decision boundary is called linear when it is defined base on a linear function of the input examples as equation (1). Accordingly, a classifier with a linear decision boundary is called a linear classifier [56]. Conversely, the decision boundary of a classifier defined based on non-linear function of input examples is called non-linear classifier as illustrated in Figure 6.

4.2 Large margin classification

In some cases, data are linearly separable and there exists a linear decision boundary that separates positive from negative examples (Fig. 7). In this section, we first consider that data are linearly separable, and later indicate how to handle data that is not linearly separable.

In Figure 7, the circled data points are the support vectors that are closest examples to the decision boundary. These examples determine the margin with which the two classes are separated.

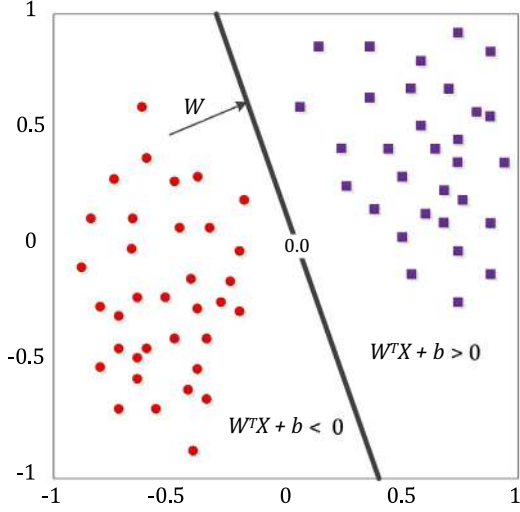


Fig. 5. Classifier with linear decision boundary.

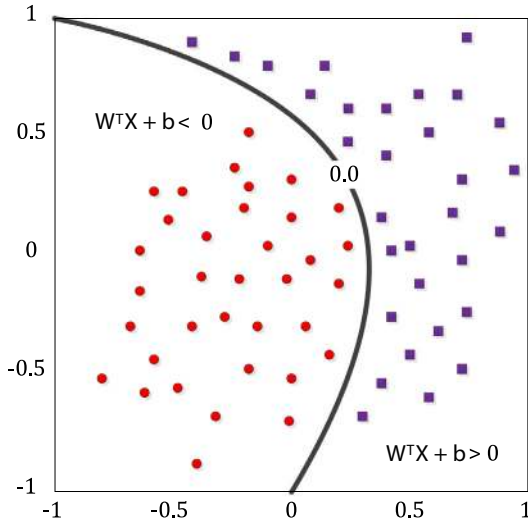


Fig. 6. Classifier with non-linear decision boundary.

In order to find the margin in SVM, consider a given hyperplane and let X^+ and X^- are the closest points to the hyperplane among the positive and negative examples, respectively. In addition, let consider the norm of a vector W denoted by $\|W\|$ is its length, and can be stated as $\sqrt{W^T W}$. Besides, let \hat{W} is a unit vector in the direction of W and is calculated as $W/\|W\|$, in which $\|\hat{W}\| = 1$. It is obvious that the margin of the hyperplane for a dataset \mathbb{P} , can be easily obtained as equation (3):

$$m_{\mathbb{P}}(f) = \frac{1}{2} \hat{W}^T (X^+ - X^-), \quad (3)$$

where X^+ and X^- are equidistant from the decision boundary that, for a constant d , leads to equations (4) and (5) for X^+ and X^- , respectively:

$$f(X^+) = W^T X^+ + b = d, \quad (4)$$

$$f(X^-) = W^T X^- + b = -d. \quad (5)$$

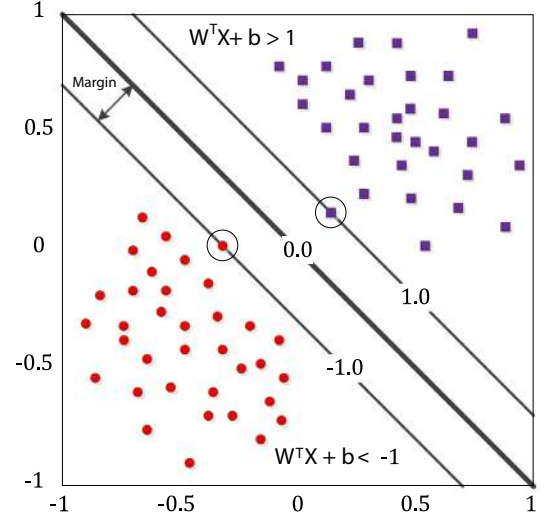


Fig. 7. Classifier with linear decision boundary and large margin.

Making the geometric margin meaningful, the value of the $f(X)$ at the points closest to the hyperplane are set equal to 1 (i.e., $d = 1$) in equations (4) and (5). By adding the two equations (4) and (5) and dividing them by $\|W\|$, the following equation (6) is obtained:

$$m_{\mathbb{P}}(f) = \frac{1}{2} \hat{W}^T (X^+ - X^-) = \frac{1}{\|W\|}. \quad (6)$$

4.3 SVM: the maximum margin classifier

Having the concept of a margin, now we can formulate the SVM as a maximum margin classifier. Two different margin SVM can be defined as hard margin SVM and soft margin SVM. The hard margin SVM is applicable for a linearly separable dataset, while for non-separable data, the soft margin SVM is developed [57]. We first explain the hard margin SVM, and then propose the soft version. The maximum margin classifier is the discriminant function that maximizes the geometric margin $1/\|W\|$ which is equivalent to minimizing $\|W\|^2$. This issue leads to the following optimization model:

$$\min \frac{1}{2} \|W\|^2 \quad (7)$$

s.t.

$$y_i (W^T X_i + b) \geq 1 \quad \forall i = 1, \dots, n \quad (8)$$

$$Wb \in R. \quad (9)$$

Constraint (8) guarantees that each example is classified correctly. It is noteworthy that constraint (8) is applicable when the data is linearly separable. In practice, data is often not linearly separable; and even if it is, a greater margin can be achieved by allowing the classifier to misclassify some points. For this aim, the new constraint (10) is proposed that leads to a soft margin SVM.

$$y_i (W^T X_i + b) \geq 1 - e_i \quad \forall i = 1, \dots, n, \quad (10)$$

where $e_i \geq 0$ is slack variable that allows an example to be in the margin ($0 \leq e_i \leq 1$) or to be misclassified ($e_i \geq 1$). In addition, by limiting the value of $\sum e_i$, we can bound the number of misclassified examples. Accordingly, the objective function (7) is modified as new objective function (11). Finally, the optimization model is proposed as follow:

$$\min \frac{1}{2} \|W\|^2 + \lambda \sum_{i=1}^n e_i \quad (11)$$

s.t.

$$y_i (W^T X_i + b) \geq 1 - e_i \quad \forall i = 1, \dots, n \quad (12)$$

$$Wb \in R, e_i \geq 0 \quad \forall i = 1, \dots, n, \quad (13)$$

where the constant λ is a penalty to avoid misclassification and sets the relative importance of maximizing the margin and minimizing the amount of slack.

5 SVM in QA and data classification

The SVM is a popular supervised learning algorithm that has been employed in many real-world problems such as fault diagnosis [58], image classification [59], bioinformatics [60], geographical analysis [61] and hand-written character recognition [62]. Originally, the SVM is designed to solve binary classification problems, but multi-class extensions are also available [63]. Hereafter, the most related studies in SVM techniques and uncertainty of input data are reviewed. The section deals with a review of SVM applications according to the classification mentioned in Section 2.

5.1 Quality description

Taboada et al. [64] employed SVM to create a quality map of a slate deposit. After defining the quality of the slate and the location of the sampling points, they used different kinds of support vector machines (SVMs) including: SVM classification, ordinal SVM and SVM regression to draw up the quality map. They also compared the results with those for kriging (i.e., Gaussian process regression) and concluded that the SVM regression and ordinal SVM outperform kriging and have higher ability in terms of control of outliers. Lo [65] proposed an auto mechanism to classify customer messages based on the techniques of text mining and SVM. Their proposed mechanism can filter the messages into the complaints automatically and appropriately to enhance service department productivity and customer satisfaction. They also used the p -control chart to control the complaining rate under the expected service quality level for the website execution. In the experimental data, Lo [65] builds a 224-keyword database, and the correctness of SVM classification exceeds 80%, compared to 76% for other Chinese classification. Generally, text mining reduces the need for human effort in message recognition and accelerates message handling by

customer service departments. This mechanism can promote the level of service in website management.

In SVM training, since the kernel parameters and feature selection have very important roles for higher accuracy, Moravej et al. [66] used two stage of feature selection including mutual information feature selection (MIFS) and correlation feature selection (CFS) techniques, respectively, in order to get optimal features for the classifier, in which MIFS can reduce the dimensionality of inputs, speed up the training of the network and get better performance and with CFS can get optimal features. Their results show that the proposed classifier has an excellent performance on training speed, reliability and accuracy. They considered several typical power quality disturbances and the comparison between their proposed method and those of in the literature shows that the proposed method, effectively classify different kind of power quality disturbances correctly with the accuracy of 98.40%. They claimed that the accuracy of the proposed method depends on the selected features used for classifier.

Most similar to our work, Chou et al. [67] proposed a virtual metrology (VM) system to increase the processing quality by employing online quality measurements during the manufacturing process. Their approach outperforms classical offline methods such as statistical process control and predictive maintenance, in which when an abnormal condition occurs, it causes a large number of scrapped substrates and the costs will be seriously raised. The proposed VM system not only fulfills real-time quality measurement of each wafer, but also detects the performance degradation of the corresponding machines from the information of manufacturing processes. The VM system benefits from hybridization of different methods such as: (1) principal component analysis (PCA) to obtain higher accuracy by the kernel function approach; (2) support vector data description (SVDD) to obtain higher accuracy we obtain in novelty detection module; and finally (3) genetic algorithm (GA) and incremental learning methods to increase the training/learning of support vector machine (SVM) model. They finally concluded that the SVM approach obtains better prediction accuracy than the radial basis function neural network (RBFN) and back-propagation neural network (BPNN) approaches.

5.2 Quality classification

Singh et al. [68] constructed and applied support vector classification (SVC) and regression (SVR) models to the surface water quality data to optimize the monitoring program. The data set comprised of 1500 water samples representing 10 different sites monitored for 15 years. The objectives of their study were to classify the sampling sites (spatial) and months (temporal) to group the similar ones in terms of water quality with a view to reduce their number; and to develop a suitable SVR model for predicting the biochemical oxygen demand (BOD) of water using a set of variables. Consequently, the SVC model achieved a data reduction of 92.5% for redesigning the future monitoring program and the SVR model provided a tool for the

prediction of the water BOD using set of a few measurable variables.

It is well-known that, machine's vibration is the best indicator of its overall mechanical condition, and an earliest indicator of arising defects. Accordingly, Baccarini et al. [69] used machine's vibrations alongside with SVM for a new intelligent mechanical fault diagnostic method. The proposed method becomes more attractive for online monitoring without maintenance specialist intervention. Widodo and Yang [70] developed an intelligent machine prognostics system using survival analysis (SA) and support vector machine (SVM). First, SA utilizes censored and uncensored data collected from condition monitoring (CM) routine and then estimates the survival probability of failure time of machine components. Next, SVM is trained by data input from CM histories data that corresponds to target vectors of estimated survival probability. After validation process, SVM is employed to predict failure time of individual unit of machine component.

Cui and Wang [71] investigated the performance of the support vector machines classifier in the application of analog fault diagnosis. In order to improve the conventional "one against rest" SVC, they employed the label analysis mechanism that generates the refusal area, which was then further resolved by the introduction of high-dimensional space discriminant analysis. They proved the proposed method is successful in diagnosing two actual analog circuits. Also, a comparison with other classifiers, including some usually used SVCs and ANN, were carried out. Saimurugan et al. [72] proposed a two-level diagnostic system including DT and SVM. They used c-SVC and nu-SVC models of SVM with four kernel functions for classification of faults using statistical features extracted from vibration signals under good and faulty conditions of rotational mechanical system. Initially, DT algorithm was used to select the prominent features. Next, these features were given as inputs for training and testing the c-SVC and nu-SVC model of SVM and their fault classification accuracies were compared.

Wu [73] presented a new version of fuzzy wavelet support vector classifier machine to diagnosing the nonlinear fuzzy fault system with multi-dimensional input variables in car assembly line diagnosis. They handled fuzzy uncertainty in input and output variables by integrating the fuzzy theory, wavelet analysis theory, Gaussian loss function and v -support vector classifier machine. In order to optimize the unknown parameters of the proposed fuzzy Gaussian wavelet v -support vector classifier machine (TFGW v -SVCM), genetic algorithm (GA) was presented. The results confirmed the feasibility and the validity of the proposed diagnosing method. In a similar work, Wu and Ni [74] presented a new version of fuzzy support vector classifier machine by integrating the fuzzy theory and v -support vector classifier machine (TF v -SVCM), in which input and output variables are defined as triangular fuzzy numbers. To seek the optimal parameters of TF v -SVCM, particle swarm optimization (PSO) was utilized to optimize parameters of TF v -SVCM. In addition, Wu et al. [75] combined the fuzzy theory, wavelet anal-

ysis theory and v -support vector classifier machine (FW v -SVCM). GA was used to optimize the optimal parameters of FW v -SVCM. In a more simple work, Wu [76] combined the fuzzy theory with v -support vector machine and applied GA to determine optimal parameters.

Gryllias and Antoniadis [77], proposed a hybrid two stage one-against-all SVM approach for the automated diagnosis of defective rolling element bearings. In their model, the SVM is first trained using simulation data. Then, vibration measurements, resulting from the machine under condition monitoring, are imported and processed directly by the already trained SVM. This method eliminates the need of training the SVM with experimental data of the specific defective bearing. They also employed the data preprocessing approach in order to overcome problems related to sudden changes of the shaft rotating speed. In their two-stage model, at the first stage, a SVM classifier separates the normal condition signals from the faulty signals. Next, at the second stage, a SVM classifier recognizes and categorizes the type of the fault.

Ekici [78] presented a new two-stage approach by combining Wavelet Transform (WT) and SVM to classify fault types and predict the fault location in the high-voltage power transmission lines. WT is applied in the first stage to reduce the size of feature vector before classification and prediction by SVM in the second stages. Salem et al. [79] suggested an original fault signature based on an improved combination of Hilbert and Park transforms. They created two fault signatures including: Hilbert modulus current space vector (HMCSV) and Hilbert phase current space vector (HPCSV). These two fault signatures were subsequently analyzed using the classical fast Fourier transform (FFT).

Zhu and Hu [80] stated that as a black box, the support vector machine is difficult for users' understanding and explanation. To improve the comprehensibility of SVMs, they proposed a rule extraction technique from support vector machines via analyzing the distribution of samples. They also defined the consistent region of samples in terms of classification boundary, and form a consistent region covering of the sample space. Then a covering reduction algorithm was developed for extracting compact representation of classes.

Shokri et al. [81] used support vector regression (SVR) for online estimation of product quality in refining processes. As choosing optimal hyper-parameter values for the SVR is a hard optimization problem, they developed some Hybrid Meta-Heuristic (HMH) algorithms based on GA and PSO to determine the parameters as fast and accurate as possible. In the literature, in order to seek the feature set used to construct a classifier, most strategies have proposed the elimination of features independently of classifier construction by exploiting statistical properties of each of the variables, or via greedy search. In a different work, Maldonado et al. [82] proposed two different Mixed Integer Linear Programming (MILP) formulations based on extensions of Support Vector Machines. The proposed approaches perform variable selection simultaneously with classifier construction using optimization models.

Saidi et al. [83] presented a novel pattern classification approach for bearings diagnostics, which combines the higher order spectra analysis (HOSA) features and SVM classifier. The use of non-linear features motivated by HOSA is a promising approach to analyze the non-linear and non-Gaussian characteristics of the mechanical vibration signals. The output of HOSA (i.e., extracted bispectrum features) is subject to PCA for dimensionality reduction. Next, the output of PCA was fed to SVM to distinguish four kinds of bearing faults. Yin et al. [84] proposed a new online fault diagnosis system that is capable of updating the class numbers in fault diagnosis in a dynamic state. Their system is based on Incremental Support Vector Data Description (ISVDD) and Extreme Learning Machine with incremental output structure (IOELM). ISVDD is used to find a new failure mode quickly in the continuous condition monitoring of the equipment.

Using the information contained in the vibration signals, Fernández-Francos et al. [85] presented an automatic method for bearing fault detection and diagnosis. Initially, a one-class ν -SVM is used to discriminate between normal and faulty conditions. Next, in order to build a model of normal operation regime, only data extracted under normal conditions is used. Besides, Band-pass filters and Hilbert Transform were then used sequentially to obtain the envelope spectrum of the original raw signal that will finally be used to identify the location of the problem. Nasiri et al. [63] formulated a least squares version of twin K -class support vector classification (Twin- K SVC) called as LST- K SVC that leads to extremely simple and fast algorithm. They examined the proposed LST- K SVC on face datasets indicate that the proposed method has comparable accuracy in classification to that of Twin- K SVC but with remarkably less computational time.

Pooyan et al. [58] developed Multi Label SVM approach (MLSVM) for Process Fault Diagnosis (PFD) in a Dew Point process that overcomes the difficulties of the Mono Label Artificial Neural Network (MLANN) approach including the needs for a large number of data points with difficult data gathering procedure and time consuming computation. In this paper, simulation results showed that the classification performance of the proposed MLSVM approach with multiple regulation parameters is similar to MLANN approach and better than MLSVM with single regulation parameter. However, MLSVM with multiple regulation parameters has major advantages in comparison with the MLANN approach including requirement of smaller number of data, easy data gathering and lower computational time.

Similar to Yin et al. [86], Marnerides et al. [87] presented a fundamentally new approach for classifying known digital subscriber line (DSL) level anomalies by exploiting the properties of novelty detection via the employment of one-class support vector machines (SVMs). Tomar and Agarwal [88] extended the formulation of binary least squares twin support vector machine (LSTSVM) classifier to multi-class by using the concepts such as “One-versus-All”, “One-versus-One”, “All-versus-One” and directed acyclic graph (DAG). They also con-

ducted a comparative analysis of these multi-classifiers in terms of their advantages, disadvantages and computational complexity. Finally, they concluded that all the proposed multi-classifiers have shown better performance as compared to the typical multi-classifiers based on SVM and “Twin Support Vector Machine”.

Jegadeeshwaran and Sugumaran [89] proposed an on-line condition monitoring by using SVM for the hydraulic brakes in automobiles. Since there are several numbers of features, they employed decision tree for classification of the given problem as well as SVM. The selected features were classified using the C-SVM and Nu-SVM with different kernel functions. For more accurate fault diagnosis, they considered seven good statistical features including minimum, standard error, sample variance, kurtosis, skewness, standard deviation, and mean. By taking the top five features into account, the RBF kernel of the C-SVM classification model reported better results, compared to other kernel functions. They also concluded that in comparison of the SVMs based on the kernel functions, radial basis function (RBF) has more classification accuracy in both types of SVM, and can be suggested for practical applications. In order to investigate the classification performance of SVMs for multiple classes, Fu et al. [90] combined auxiliary algorithms with SVMs including: Recursive Feature Elimination (RFE) algorithm, parameters optimizing methods and Two-Step Classification strategy. They concluded that classification ability and operation efficiency of SVMs are both improved when dimension is reduced and two-step classification SVM (TSC-SVM) works well under circumstances that samples overlap with each other. Tian et al. [91] studied a fault diagnosis method based on modified SVM, in which the dimension of samples is effectively reduced by recursive feature elimination (RFE) algorithm, and computing time is saved at the same time. Couellan et al. [92] proposed a collaborative variant of SVM namely self-adaptive SVM that can handle binary classification, hyper parameters selection, multiclass learning as well as unsupervised learning. The proposed self-adaptive SVM provides simple formulations of complex learning tasks that are sometimes very difficult to solve with classical optimization strategies.

In the one-against-all SVM for multi-label classification, the main problem is that a data sample may be classified into a multi-label class that is not defined or it may not be classified into any class. Accordingly, Abe [93] proposed fuzzy SVMs (FSVMs) for multi-label classification, in which for each multi-label class, a region with the associated membership function was defined and a data point was classified into a multi-label class whose membership function was the largest.

5.3 Quality prediction

Liu et al. [94] extended SVM to the doubly regularized SVM (DrSVM) based on the elastic net penalty. They successfully applied DrSVM in handling variables election while retaining (or discarding) correlated variables. Since

solving this model is challenging, they develop an iterative l_2 -SVM approach to implement DrSVM over high-dimensional datasets. Fei et al. [95] proposed a wavelet relevance vector machine in order to predict product quality and optimize production process. However, there are complex nonlinear relationship among the product quality parameters and the production process variables. The results demonstrated that the proposed model has a higher prediction precision than the common methods such as partial least squares (PLS), orthogonal signal correction-partial least squares (OSC-PLS), Quadratic-PLS, kernel partial least squares (KPLS), orthogonal signal correction-kernel partial least squares (OSC-PLS), least squares-support vector machines (LS-SVM) and ordinary relevance vector machines (RVM).

Chen et al. [96] proposed a novel intelligent diagnosis model based on wavelet support vector machine (WSVM) and immune genetic algorithm (IGA) for gearbox fault diagnosis, in which WSVM is used to solve the diagnosis problem with small sampling, nonlinearity and high dimension and IGA is to determine the optimal parameters for WSVM.

Bansal et al. [97] studied an application of the SVM in multiclass gear-fault diagnosis when the gear vibration data in frequency domain averaged over a large number of samples is used. Since the training data may not always be available at all angular speeds of the gear, hence, they proposed two novel techniques, namely the interpolation and the extrapolation methods. These techniques that help the SVM classifier perform multiclass gear fault diagnosis with noticeable accuracy, even in the absence of the training data at the testing angular speed. Similarly, Zhang and Zhou [98] presented a novel procedure based on ensemble empirical mode decomposition (EEMD) and optimized SVM for multi-fault diagnosis of rolling element bearings.

In quality control domain, manual inspection and evaluation of quality control data has become a tedious task that requires the undistracted attention of specialized personnel. On the other hand, automated monitoring of a production process is necessary, not only for real time quality assessment, but also for potential machinery malfunction diagnosis. For this reason, control chart pattern recognition (CCPR) methods have received a lot of attention over the last two decades [99]. Consequently, Xanthopoulos and Razzaghi [99] proposed the use of weighted support vector machines (WSVM) for automated process monitoring and early fault diagnosis, while original SVM demonstrate poor performance when applied directly to these problems.

5.4 Parameter optimization

Chen et al. [100] proposed a parallel time variant particle swarm optimization (TVPSO) algorithm to simultaneously perform the parameter optimization and feature selection for SVM, termed PTVPSO-SVM. Their algorithm was implemented in a parallel environment using Parallel Virtual Machine (PVM). In the proposed method,

a weighted function was adopted to design the objective function of PSO, which took into account the average classification accuracy rates (ACC) of SVM, the number of support vectors (SVs) and the selected features simultaneously. Yin et al. [86] constructed a new fault detection scheme based on the robust one class support vector machine (1-class SVM). 1-class SVM is a special variant of the general SVM and since only the normal data is required for training, 1-class SVM is widely used in anomaly detection. Since the 1-class SVM is sensitive to the outliers included in the training dataset, they proposed a robust 1-class SVM through designing penalty factors.

Since the Least squares SVM (LS-SVMs) are sensitive to outliers or noise in the training dataset, several researchers have developed the weighted least squares SVM (WLS-SVMs) that can partly overcome this shortcoming by assigning different weights to different training samples. However, it is a difficult task for WLS-SVMs to set the weights of the training samples, which greatly influences the robustness of WLS-SVMs. Accordingly, Yang et al. [101] presented a novel robust LS-SVM (RLS-SVM) based on the truncated least squares loss function for regression and classification with noise. They also proposed an iterative algorithm based on the concave-convex procedure (CCCP) and the Newton algorithm in order to solve the proposed RLS-SVM. Finally, they concluded the robustness of RLS-SVM is higher than that of LS-SVMs and WLS-SVMs.

Zhang and Zhang [102] employed the social emotional optimization algorithm (SEOA) for machine training and parameter settings for SVM. They modeled machine training for SVM as a multi-parameter optimization problem which is solved by SEOA. They also utilized SEOA not only for setting the SVM's parameter, but also for optimizing the kernel function parameter and error penalty parameter of SVM.

As a result of this section, a gap analysis is done for the literature of SVMs in QA problems. Accordingly, Table 2 have listed the detail of the more related previous works. In order to better analyzing the gaps in the literature, Figure 8 illustrates the percentage of studies in different details. In addition, to better illustrate the interest in using SVM technique in QA problems, Figure 9 shows the number of papers applied SVM in QA domain during the last years.

6 Conclusions and future research directions

In many modern manufacturing industries, data that characterize the manufacturing process are electronically collected and stored in the databases. Therefore, data mining tools can be employed for automatically discovering interesting and valuable knowledge and patterns in the manufacturing processes. These knowledge and patterns can be subsequently exploited to boost the whole manufacturing process in such areas as defect prevention and detection, quality improvement, reducing flow-time, increasing safety, etc.

Table 2. Details of related works.

Author	Year	Quality goal		Quality task				System		SVM		Parameter settings		Uncertainty		
		Predictive	Descriptive	Quality	prediction	classification	optimization	stage One	processing With Prr-	Linear	Non-linear	heuristics Meta-	MILP	Fuzzy	Robust	Mixed
Taboada et al.	2007	✓	✓	✓	✓	✓	✓	✓	✓	✓						
Lo	2008	✓	✓					✓	✓							
Moravej et al.	2009	✓	✓					✓								
Chou et al.	2010	✓	✓					✓								
Singh et al.	2011	✓	✓					✓								
Baccarini et al.	2011	✓	✓					✓								
Widodo and Yang	2011	✓	✓					✓								
Cui and Wang	2011	✓	✓					✓								
Saimurugan et al.	2011	✓	✓					✓								
Wu (a)	2011	✓	✓					✓						✓		
Wu and Ni	2011	✓	✓					✓						✓		
Wu et al.	2011	✓	✓					✓						✓		
Wu (b)	2011	✓	✓					✓						✓		
Gryllias and Antoniadis	2012	✓	✓					✓						✓		
Ekici	2012	✓	✓					✓						✓		
Salem et al.	2012	✓	✓					✓						✓		
Liu et al.	2013	✓	✓					✓						✓		
Fei et al.	2013	✓	✓					✓						✓		
Zhu and Hu	2013	✓	✓					✓						✓		
Fernández-Francos et al.	2013	✓	✓					✓						✓		
Chen et al.	2013	✓	✓					✓						✓		
Bansal et al.	2013	✓	✓					✓						✓		
Zhang and Zhou	2013	✓	✓					✓						✓		
Shokri et al.	2014	✓	✓					✓						✓		
Maldonado et al.	2014	✓	✓					✓						✓		
Saidi et al.	2014	✓	✓					✓						✓		
Yin et al.	2014	✓	✓					✓						✓		
Xanthopoulos and Razzaghi	2014	✓	✓					✓						✓		
Chen et al.	2014	✓	✓					✓						✓		
Yin et al.	2014	✓	✓					✓						✓		
Yang et al.	2014	✓	✓					✓						✓		
Marnierdes et al.	2015	✓	✓					✓						✓		
Tomar and Agarwal	2015	✓	✓					✓						✓		
Jegadeeshwaran and Sugumaran	2015	✓	✓					✓						✓		
Nasiri et al.	2015	✓	✓					✓						✓		
Pooyan et al.	2015	✓	✓					✓						✓		
Zhang and Zhang	2015	✓	✓					✓						✓		
Fu et al.	2015	✓	✓					✓						✓		
Tian et al.	2015	✓	✓					✓						✓		
Couellan et al.	2015	✓	✓					✓						✓		
Abe	2015	✓	✓					✓						✓		

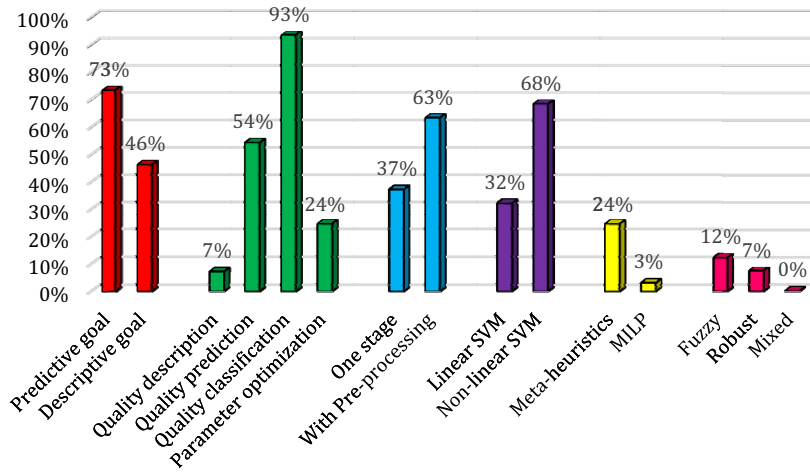


Fig. 8. Percentage of different studies with different detail.

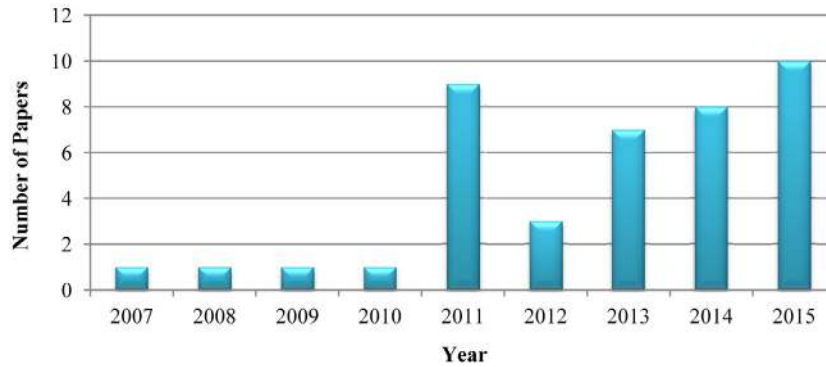


Fig. 9. Number of studied papers about SVM technique application in QA.

According to Section 4, it can be stated that the predictive and classifying aspects of data mining alongside with parameter optimization are probably the most developed parts. In data mining, it is obvious that the choice of technique to use in analyzing a dataset and investigating different classes of quality depend on the understanding of the analyst. On the other hand, a lot of time is wasted in testing every single prediction technique to find the best solution that fits the analyst’s needs. Hence, with the advent of improved and efficient prediction techniques, there is a need for an analyst to know which tool performs best for a particular type of data set.

Recently, as one of the strongest prediction and classifier tools, SVM is widely used on diverse industrial data sets to predict and classify the quality of the products. This paper presented the results of a comprehensive review of SVM application in manufacturing industries to the QA problems. Based on the results of Table 2 and Figure 8, the following concluding remarks can be stated:

- In many of the studies reviewed, data preprocessing methods have been used, where the success of a DM study depends heavily on the success in preprocessing of the data. Using different methods on data handling practices can be stated as an important future direction.
- According to comprehensive review in the application of DM methods and specially SVM in QA problems, it has been observed that there is an increasing trend in the use of SVM method for QA. This can be attributed to the availability of massive data sets in some manufacturing domains and the need to improve processes with the intense competition in the industries. In addition, among different DM techniques, it has been shown that SVM is an easy-to-use method that is able to handle large dimensional dataset with different features.
- In most of the studies reviewed, non-linear SVM and the use of kernel functions are not reported at all, even though not only most of real datasets are not linearly separable, but also the success of a SVM method depends heavily on the success in utilizing proper kernel function. Using different well-known kernel functions and developing pre-computed kernel functions on data classifying practices can be stated as important future directions.
- Regarding to performance comparisons, a few of papers have compared their results with those of well-known DM methods in the literature as well as with traditional statistical methods, while traditional statistical methods may still provide valuable information.

Comparison between SVM and other methods could be important for further researches.

- In many of the studies reviewed, parameter settings have been neglected, while the success of a SVM method depends deeply on the success in setting optimal values for SVM parameters. Some papers have applied meta-heuristics to set the parameters. Accordingly, using efficient meta-heuristic algorithms on parameter settings practices can be stated as an important future direction.
- SVM methods have been developed either as stand-alone or more interestingly as hybrid. According to the results, quality classification is the most frequently used SVM function observed, alone or as a prior/integral part of optimization. In most of the applications SVMs were used for quality prediction purposes as well as parameter optimization in a few numbers of the others.
- Since in many industries, input data are affected by uncertain environment and managers have no control on determining exact value for them, data uncertainty is an inherent property in various applications due to reasons such as outdated sources or imprecise measurement. When data mining techniques are applied to these data, their uncertainty has to be considered to obtain high quality results. In order to cope with uncertain input, the classical SVM can be modified to obtain a robust classifier that is insensitive to and able to handle data uncertainty.

References

1. M. Kano, Y. Nakagawa, Data-based process monitoring, process control and quality improvement: Recent developments and applications in steel industry, *Comput. Chem. Eng.* **32**, 12–24 (2008)
2. K. Yang, J. Trewn, *Multivariate Statistical Methods in Quality Management* (McGraw-Hill, New York, 2004)
3. A. Nachev, B. Stoyanov, Product quality analysis using Support Vector Machines. *Information Models and Analyses* **1**, 179–192 (2012)
4. L. Rokach, O. Maimon, Data mining for improving the quality of manufacturing: a feature set decomposition approach, *J. Intell. Manufact.* **17**, 285–299 (2006)
5. G. Köksal, I. Batmaz, M.C. Testik, A review of data mining applications for quality improvement in manufacturing industry, *Expert Syst. Appl.* **38**, 13448–13467 (2011)
6. C.P. Wei, S. Piramuthu, M.J. Shaw, Knowledge discovery and data mining. In *Handbook on Knowledge Management*, edited by C.W. Holsapple (Springer-Verlag, Berlin, 2003), Vol. 2, pp. 157–189
7. A.K. Choudhary, J.A. Harding, M.K. Tiwari, Data mining in manufacturing: A review based on the kind of knowledge, *J. Intell. Manufact.* (2008)
8. M.S. Phadke, *Quality Engineering using Robust Design* (Prentice-Hall, Englewood Cliffs, NJ, 1989)
9. A. Berson, K. Thearling, J. Stephen, *Building Data Mining Applications for CRM* (USA, McGraw-Hill, 1999)
10. W. Frawley, G. Piatetsky-Shapiro, C. Matheus, Knowledge Discovery in Databases: An Overview, *AI Magazine* (1992), pp. 213–228
11. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurasamy, *Advances in Knowledge Discovery and Data Mining* (AAAI/MIT Press, 1996)
12. M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms* (John Wiley & Sons, 2011)
13. F. Gorunescu, *Data mining* (Springer-Verlag Berlin An, 2013)
14. C. Romero, S. Ventura, Data Mining in Education, *Wiley Interdiscipl. Rev.: Data Mining Knowl. Discovery* **3**, 12–27 (2013)
15. D.T. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining* (John Wiley & Sons, 2014)
16. C. McCue, *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis* (Butterworth-Heinemann, 2015)
17. D. Montgomery, C. Peck, A. Elizabeth, V.G. Geoffrey, *Introduction to Linear Regression Analysis* (John Wiley & Sons, Inc., New York, 2001)
18. Harrell, Jr., E. Frank, *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis* (Springer-Verlag, New York, 2001)
19. N. Mastrogiannis, B. Boutsinas, I. Giannikos, A method for improving the accuracy of data mining classification algorithms, *Comput. Oper. Res.* **36**, 2829–2839 (2009)
20. A. Chemchem, H. Drias, From data mining to knowledge mining: Application to intelligent agents, *Expert Syst. Appl.* **42**, 1436–1445 (2015)
21. P. Berkhin, A Survey of Clustering Data Mining Techniques. *Grouping Multidimensional Data* (2006) 25–71
22. N. Padhy, D. Mishra, R. Panigrahi, The survey of data mining applications and feature scope, *arXiv:1211.5723* (2012)
23. S.L. Ting, Y.K. Tse, G.T.S. Ho, S.H. Chung, G. Pang, Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. *Int. J. Prod. Econ.* **152**, 200–209 (2014)
24. M. Gardner, J. Bieker, Data mining solves tough semiconductor manufacturing problems. In *S.J. Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining*, KDD 2000 (ACM, Boston, New York, 2000), pp. 376–383
25. M.A. Karim, S. Halgamuge, A.J.R. Smith, A.L. Hsu, Manufacturing yield improvement by clustering. In *Neural information processing: 13th international conference proceedings Part III*, edited by I. King et al., 3–6 October 2006, Hong Kong, China. In LNCS (Springer, Berlin, 2006), Vol. 4234, pp. 526–534
26. C. Huang, T. Li, T. Peng, Attribute selection based on rough set theory for electromagnetic interference (EMI) fault diagnosis, *Quality Eng.* **18**, 161–171 (2006)
27. C.W. Liu, C.F. Chien, An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing, *Eng. Appl. Artif. Intell.* **26**, 1479–1486 (2013)
28. C.F. Chien, S.C. Hsub, Y.J. Chen, A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence, *Int. J. Prod. Res.* **51**, 2324–2338 (2013)
29. De N. Abajo, A.B. Diez, V. Lobato, S.R. Cuesta, ANN quality diagnostic models for packaging manufacturing: an industrial data mining case study. In *KDD-2004: Proc. of the tenth ACM SIGKDD int. Conf. on Knowledge*

- Discovery and Data Mining*, 22–25 August 2004, Seattle Washington, edited by R. Kohavi et al. (ACM Press, New York, 2004), pp. 799–804
30. D.S. Chang, S.T. Jiang, Assessing quality performance based on the online sensor measurements using neural networks, *Comput. Ind. Eng.* **42**, 417–424 (2002)
 31. W.C. Chen, A.H.I. Lee, W.J. Deng, K.Y. Liu, The implementation of neural network for semiconductor PECVD process, *Expert Syst. Appl.* **32**, 1148–1153 (2007)
 32. T. Yang, T. Tsai, J. Yeh, A neural network-based prediction model for fine pitch stencil printing quality in surface mount assembly, *Eng. Appl. Artif. Intell.* **18**, 335–341 (2005)
 33. X. Shi, P. Schillings, D. Boyd, Applying artificial neural networks and virtual experimental design to quality improvement of two industrial processes, *Int. J. Prod. Res.* **42**, 101–118 (2004)
 34. H.Y. Tseng, Welding parameters optimization for economics design using neural approximation and genetic algorithm, *Int. J. Adv. Manufact. Technol.* **27**, 897–901 (2006)
 35. B.S. Sung, I.S. Kim, Y. Xue, H.H. Kim, Y.H. Cha, Fuzzy regression model to predict the bead geometry in the robotic welding process, *ACTA Metall. Sinica* **20**, 391–397 (2007)
 36. B. Ozcelik, T. Erzurumlu, Comparison of the warpage optimization in the plastic injection molding using ANOVA, neural network model and genetic algorithm, *J. Mater. Process. Technol.* **171**, 437–445 (2006)
 37. C. Shen, L. Wang, Q. Li, Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method, *J. Mater. Process. Technol.* **183**, 412–418 (2007)
 38. S. Guessasma, Z. Salhi, G. Montavon, P. Gougeon, C. Coddet, Artificial intelligence implementation in the APS process diagnostic, *Mater. Sci. Eng. B* **110**, 285–295 (2004)
 39. P.C. Austina, J.V. Tua, J.E. Hoe, D. Leveye, D.S. Lee, 2013. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, *J. Clin. Epidemiol.* **66**, 398–407 (2013)
 40. K. Wang, S. Tong, B. Eynard, L. Roucoules, N. Matta, Review on application of data mining in product design and manufacturing. In *Proc. of the fourth Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, 24–27 August, Haikou Hainan, China, edited by J. Lei, J. Yu, S. Zhou (IEEE, 2007), pp. 613–618
 41. T.H. Hou, W.L. Liu, L. Lin, Intelligent remote monitoring and diagnosis of manufacturing processes using an integrated approach of neural networks and rough sets. *J. Intell. Manufact.* **14**, 235–239 (2003)
 42. H. Huang, D. Wu, 2005. Product quality improvement analysis using data mining: A case study in ultra-precision manufacturing industry. In *Proc. of the Second Int. Conference on Fuzzy Systems and Knowledge Discovery*, 22–29 August 2005, Changsha, China, edited by L. Wang, Y. Jin, FSKD 2005. In *Lecture notes in computer science* (Springer, Berlin, 2005), Vol. 3614, pp. 577–580
 43. G.T. Jemwa, C. Aldrich, Improving process operations using support vector machines and decision trees, *AICHE J.* **51**, 526–543 (2005)
 44. H. Sarimveis, P. Doganis, A. Alexandridis, A classification technique based on radial basis function neural networks, *Adv. Eng. Software* **37**, 218–221 (2006)
 45. Y.H. Hung, Optimal process parameters design for a wire bonding of ultrathin CSP package based on hybrid methods of artificial intelligence, *Microelectron. Int.* **24**, 3–10 (2007)
 46. M. Hamedi, M. Shariatpanahi, A. Mansourzadeh, Optimizing spot welding parameters in a sheet metal assembly by neural networks and genetic algorithm, *Proc. Instit. Mech. Eng. Part B J. Eng. Manufact.* **221**, 1175–1184 (2007)
 47. V. Nedic, S. Cvetanovic, D. Despotovic, M. Despotovic, S. Babic, Data mining with various optimization methods, *Expert Syst. Appl.* **41**, 3993–3999 (2014)
 48. T.L. Chiang, C.T. Su, T.S. Li, R.C.C. Huang, Improvement of process capability through neural networks and robust design: A case study, *Quality Eng.* **14**, 313–318 (2002)
 49. G.T.S. Ho, H.C.W. Lau, C.K.M. Lee, A.W.H. Ip, K.F. Pun, 2006. An intelligent production workflow mining system for continual quality enhancement, *Intell. J. Adv. Manufact. Technol.* **28**, 792–809 (2006)
 50. K. Manimala, K. Selvi, R. Ahila, Optimization techniques for improving power quality data mining using wavelet packet based support vector machine, *Neurocomputing* **77**, 36–47 (2012)
 51. B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers. In *5th Annual ACM Workshop on COLT*, edited by D. Haussler (Pittsburgh, PA. ACM Press, 1992), pp. 144–152
 52. B. Scholkopf, K. Tsuda, J.P. Vert, *Kernel Methods in MIT Press series on Computational Molecular Biology* (MIT Press, 2004)
 53. N. Cristianini, J. Shawe-Taylor, *An introduction to support vector machines: And other kernel-based learning methods* (Cambridge University Press, 2000)
 54. J. Shawe-Taylor, N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge UP, Cambridge, UK, 2004)
 55. B. Scholkopf, A. Smola, *Learning with Kernels* (MIT Press, Cambridge, MA, 2002)
 56. A. Ben-Hur, J. Weston, A User's Guide to Support Vector Machines, Technical report, <http://pym1.sourceforge.net/doc/howto.pdf> (2012)
 57. C. Cortes, V. Vapnik, Support vector networks, *Machine Learning* **20**, 273–297 (1995)
 58. N. Pooyan, M. Shahbazian, K. salahshoor, M. Hadian, Simultaneous fault diagnosis using multi class support vector machine in a Dew point process, *J. Natural Gas Sci. Eng.* **23**, 373–379 (2015)
 59. S. Li, H. Wu, D. Wan, J. Zhu, An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine, *Knowledge-Based Systems* **24**, 40–48 (2011)
 60. G. Orrù, W. Pettersson-Yeo, A.F. Marquand, G. Sartori, A. Mechelli, Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review, *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012)
 61. B. Pradhan, A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Comput. Geosci.* **51**, 350–365 (2013)

62. J. John, K.V. Pramod, K. Balakrishnan, Unconstrained handwritten Malayalam character recognition using wavelet transform and support vector machine classifier, *Procedia Eng.* **30**, 598–605 (2012).
63. J.A. Nasiri, N.M. Charkari, S. Jalili, Least squares twin multi-class classification support vector machine, *Pattern Recognition* **48**, 984–992 (2015)
64. J. Taboada, J.M. Matías, C. Ordóñez, P.J. García, Creating a quality map of a slate deposit using support vector machines, *J. Comput. Appl. Math.* **204**, 84–94 (2007)
65. S. Lo, Web service quality control based on text mining using support vector machine, *Expert Syst. Appl.* **4**, 603–610 (2008)
66. Z. Moravej, S.A. Banihashemi, M.H. Velayati, Power quality events classification and recognition using a novel support vector algorithm, *Energy Convers. Manag.* **50**, 3071–3077 (2009)
67. P.H. Chou, M.J. Wua, K.K. Chen, Integrating support vector machine and genetic algorithm to implement dynamic wafer quality prediction system, *Expert Syst. Appl.* **37**, 4413–4424 (2010)
68. K.P. Singh, N. Basant, S. Gupta, Support vector machines in water quality management, *Anal. Chim. Acta* **703**, 152–162 (2011)
69. L.M.R. Baccarini, V.V.R. Silva, B.R. Menezes, W.M. Caminhas, SVM practical industrial application for mechanical faults diagnostic, *Expert Syst. Appl.* **38**, 6980–6984 (2011)
70. A. Widodo, B.S. Yang, Machine health prognostics using survival probability and support vector machine, *Expert Syst. Appl.* **38**, 8430–8437 (2011)
71. J. Cui, Y. Wang, A novel approach of analog circuit fault diagnosis using support vector machines classifier, *Measurement* **44**, 281–289 (2011)
72. M. Saimurugan, K.I. Ramachandran, V. Sugumaran, N.R. Sakthivel, Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine, *Expert Syst. Appl.* **38**, 3819–3826 (2011)
73. Q. Wu, Car assembly line fault diagnosis model based on triangular fuzzy Gaussian wavelet kernel support vector classifier machine and genetic algorithm, *Expert Syst. Appl.* **38**, 14812–14818 (2011)
74. Q. Wu, Z. Ni, Car assembly line fault diagnosis based on triangular fuzzy support vector classifier machine and particle swarm optimization, *Expert Syst. Appl.* **38**, 4727–4733 (2011)
75. Q. Wu, R. Law, S. Wu, Fault diagnosis of car assembly line based on fuzzy wavelet kernel support vector classifier machine and modified genetic algorithm, *Expert Syst. Appl.* **38**, 9096–9104 (2011)
76. Q. Wu, Hybrid fuzzy support vector classifier machine and modified genetic algorithm for automatic car assembly fault diagnosis, *Expert Syst. Appl.* **38**, 1457–1463 (2011)
77. K.C. Gryllias, I.A. Antoniadis, A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments, *Eng. Appl. Artif. Intell.* **25**, 326–344 (2012)
78. S. Ekici, Support Vector Machines for classification and locating faults on transmission lines, *Appl. Soft Comput.* **12**, 1650–1658 (2012)
79. S.B. Salem, K. Bacha, A. Chaari, Support vector machine based decision for mechanical fault condition monitoring in induction motor using an advanced Hilbert-Park transform, *ISA Trans.* **51**, 566–572 (2012)
80. P. Zhu, Q. Hu, Rule extraction from support vector machines based on consistent region covering reduction, *Knowledge-Based Syst.* **42**, 1–8 (2013)
81. S. Shokri, M.T. Sadeghi, M.A. Marvast, High reliability estimation of product quality using support vector regression and hybrid meta-heuristic algorithms, *J. Taiwan Instit. Chem. Eng.* **45**, 2225–2232 (2014)
82. S. Maldonado, J. Pérez, R. Weber, M. Labbé, Feature selection for support vector machines via mixed integer linear programming, *Information Sci.* **279**, 163–175 (2014)
83. L. Saidi, Ben J. Ali, F. Fnaiech, Application of higher order spectral features and support vector machines for bearing faults classification, *ISA Transactions*. Article in press. DOI: 10.1016/j.isatra.2014.08.007 (2014)
84. G. Yin, Y.T. Zhang, Z.N. Li, G.Q. Ren, H.B. Fan, Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure, *Neurocomputing* **128** 224–231 (2014a)
85. D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class *v*-SVM, *Comput. Ind. Eng.* **64**, 357–365 (2013)
86. S. Yin, X. Zhu, C. Jing, Fault detection based on a robust one class support vector machine, *Neurocomputing* **145**, 263–268 (2014b)
87. A.K. Marnerides, S. Malinowski, R. Morla, H.S. Kim, Fault diagnosis in DSL networks using support vector machines, *Comput. Commun.* DOI: 10.1016/j.comcom.2015.01.006 (2015)
88. D. Tomar, S. Agarwal, A comparison on multi-class classification methods based on least squares twin support vector machine, *Knowledge-Based Systems*. DOI: 10.1016/j.knosys.2015.02.009 (2015)
89. R. Jegadeeshwaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, *Mech. Syst. Signal Process.* **52–53**, 436–446 (2015)
90. M. Fu, Y. Tian, F. Wu, Step-wise support vector machines for classification of overlapping samples, *Neurocomputing* **155**, 159–166 (2015)
91. Y. Tian, M. Fu, F. Wu, Steel plates fault diagnosis on the basis of support vector machines, *Neurocomputing* **151**, 296–303 (2015)
92. N. Couellan, S. Jan, T. Jorquera, J.P. Georgé, Self-adaptive Support Vector Machine: A multi-agent optimization perspective, *Expert Syst. Appl.* **42**, 4284–4298 (2015)
93. S. Abe, Fuzzy support vector machines for multi-label classification, *Pattern Recognition* DOI: 10.1016/j.patcog.2015.01.009 (2015)
94. D. Liu, H. Qian, G. Dai, Z. Zhang, An iterative SVM approach to feature selection and classification in high-dimensional datasets, *Pattern Recogn.* **46**, 2531–2537 (2013)
95. H. Fei, X. Jinwu, L. Min, Y. Jianhong, Product quality modelling and prediction based on wavelet relevance vector machines, *Chemometrics and Intelligent Laboratory Systems* **121**, 33–41 (2013)

96. F. Chen, B. Tang, R. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, *Measurement* **46**, 220–232 (2013)
97. S. Bansal, S. Sahoo, R. Tiwari, D.J. Bordoloi, Multiclass fault diagnosis in gears using support vector machine algorithms based on frequency domain data, *Measurement* **46**, 3469–3481 (2013)
98. X. Zhang, J. Zhou, Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines, *Mech. Syst. Signal Process.* **41**, 127–140 (2013)
99. P. Xanthopoulos, T. Razzaghi, A weighted support vector machine method for control chart pattern recognition, *Comput. Ind. Eng.* **70**, 134–149 (2014)
100. H.L. Chen, B. Yang, S.J. Wang, G. Wang, D.Y. Liu, H.Z. Li, W.B. Liu, Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy, *Appl. Math. Comput.* **239**, 180–197 (2014)
101. X. Yang, L. Tan, L. He, A robust least squares support vector machine for regression and classification with noise, *Neurocomputing* **140**, 41–52 (2014)
102. Y. Zhang, P. Zhang, Machine training and parameter settings with social emotional optimization algorithm for support vector machine, *Pattern Recognition Lett.* **54**, 36–42 (2015)
103. P. Cheeseman, On Finding the Most Probable Model. In *Computational Models of Scientific Discovery and Theory Formation*, edited by J. Shrager, P. Langley (San Francisco, Calif.: Morgan Kaufmann, 1990), pp. 73–95