



A NOVEL QUALITY PREDICTION METHOD BASED ON FEATURE SELECTION CONSIDERING HIGH DIMENSIONAL PRODUCT QUALITY DATA

JUNYING HU^{1,3}, XIAOFEI QIAN^{1,2,3,5,*}, JUN PEI^{1,2,3}
CHANGCHUN TAN⁴, PANOS M. PARDALOS², XINBAO LIU^{1,3,*}

¹School of Management, Hefei University of Technology, Hefei 230009, China

²Center for Applied Optimization, Department of Industrial and Systems
Engineering University of Florida, Gainesville, FL 32611, USA

³Key Laboratory of Process Optimization and Intelligent Decision-making
of the Ministry of Education, Hefei 230009, China

⁴School of Economics, Hefei University of Technology, Hefei 230009, China

⁵Ministry of Education Engineering Research Center for Intelligent Decision-Making
& Information System Technologies, Hefei 230009, China

(Communicated by Biao Luo)

ABSTRACT. Product quality is the lifeline of enterprise survival and development. With the rapid development of information technology, the semiconductor manufacturing process produces multitude of quality features. Due to the increasing quality features, the requirement on the training time and classification accuracy of quality prediction methods becomes increasingly higher. Aiming at realizing the quality prediction for semiconductor manufacturing process, this paper proposes a modified support vector machine (SVM) model based on feature selection, considering the high dimensional and nonlinear characteristics of data. The model first improves the Radial Basis Function (RBF) in SVM, and then combines the Duelist algorithm (DA) and variable neighborhood search algorithm (VNS) for feature selection and parameters optimization. Compared with some other SVM models that are based on DA, genetic algorithm (GA), and Information Gain algorithm (IG), the experiment results show that our DA-VNS-SVM can obtain higher classification accuracy rate with a smaller feature subset. In addition, we compare the DA-VNS-SVM with some common machine learning algorithms such as logistic regression, naive Bayes, decision tree, random forest, and artificial neural network. The results indicate that our model outperform these machine learning algorithms for the quality prediction of semiconductor.

1. Introduction. A semiconductor is a highly miniaturized integrated circuit consisting of thousands of components. Semiconductor manufacturing is considered to be one of the most complicated high-tech manufacturing processes. Semiconductor manufacturing process starts with raw wafers, thin discs made of silicon or gallium arsenide [45]. In general, a semiconductor production line consists of several hundred equipments. Considering the scale of integration, the type of package, and

2020 *Mathematics Subject Classification.* Primary: 62C99, 62P30; Secondary: 68T20.

Key words and phrases. Support vector machine, feature selection, kernel function, hybrid intelligent algorithm, quality predictions.

* Corresponding author: Xiaofei Qian, Xinbao Liu.

customer specifications, there are more than 600 single process steps in a whole manufacturing process [51]. The four main stages of semiconductor manufacturing are shown in Fig.1. Hundreds of processing steps, months of processing time, and re-entrant process flows on different equipments all together have posed great difficulties in predicting the quality of semiconductors.

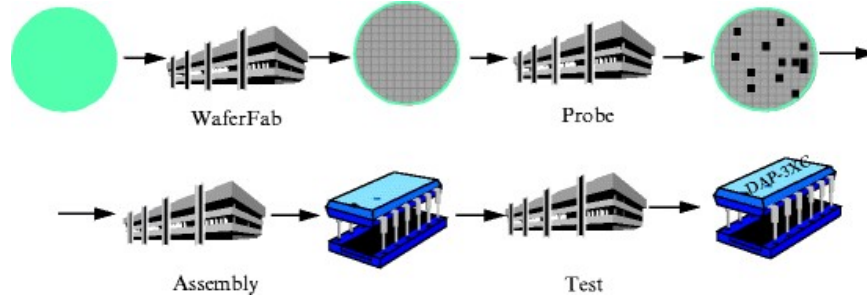


FIGURE 1. Stages of semiconductor manufacturing

The semiconductor is one of the essential parts for an electronic product. With the increasing popularity and rapid development of electronic products, semiconductor quality attracts a lot of attention in semiconductor manufacturing industry [42]. The fast technological update, sophisticated manufacturing process, short product life cycle, and variable market demand bring several difficulties to the quality control of semiconductors. In addition, the semiconductor manufacturing process is complicated and constantly changing. Hence, it is difficult to ensure the quality prediction accuracy in reality. Because of the complexity of semiconductors, it is difficult to describe a semiconductor's internal mathematical model accurately, making some system-model-based methods ineffective. On one hand, system modeling requires setting a large number of assumptions, and then testing whether the assumptions are true through data collection and analysis. Model analysis is meaningless when too many assumptions are made to suit the model. On the other hand, the development of information technology enlarges the scale of data. Most system models can not describe these various, huge, and uncertain data based on simple relationships between variables [35].

On the contrary, new generation of sensor technology, mass data storage and transmission technology, and information analysis and processing technology provide the possibility for the real-time perception of product quality data, and thus the quality inspectors can obtain a large amount of monitoring data to supervise and maintain the regular operation of semiconductor manufacturing process. Furthermore, some hidden knowledge behind the massive monitoring data can be used to describe the operational status of semiconductors. Therefore, the data-driven modeling method which can make up for the shortcomings of model-based technology to a certain extent is a natural choice for realizing the monitoring and maintenance of semiconductor quality.

Semiconductor manufacturing process produces a large number of quality features. The quality data produced in such process contain high dimensionality, high feature correlation, non-stationarity, nonlinear and large amount of noise [69]. With the development of computer and database technology, all quality data can

be retained, however, in reality the data set often contains redundant information. Therefore, it is necessary to remove the redundant information before detecting product quality. There are two commonly used feature reduction techniques. The first is to generate a new feature set with low dimensional and unrelated features from the original feature set. This is called feature synthesis, and each new feature is a function of all the original features. It has high computational efficiency. However, the disadvantage of feature synthesis is that the numerical coefficients of new features can not explain the importance of individual features in the original set. Feature selection, the other reduction technique, is the process of selecting features that are relevant in explaining the patterns present in the data [26]. Feature selection refers to selecting a subset of features with optimal output performance from a set of high-dimensional features according to specific evaluation criteria. It reduces the dimension of feature space by deciding which features to keep and which ones to disregard [46]. In this paper, feature selection is used to reduce attributes of semiconductor manufacturing process. The purpose is to find out the factors that affect the quality of semiconductor manufacturing process and the order of their importance, so as to guide the semiconductor manufacturing better.

As a part of classification rule, feature selection is subsequently followed by a standard classifier that is applied to the selected subset of features [31]. With the increase of the number of features, the search space of the feature subset grows exponentially. Most traditional feature selection algorithms are low-efficiency, so many scholars turn to use intelligent algorithms with stronger search ability. Zhang et al. [70] used genetic algorithm (GA) for feature selection of mitochondrial toxicity. Li et al. [44] proposed an effective feature selection method for hyperspectral image classification based on GA. Huang and Dun [32] combined the discrete particle swarm optimization (PSO) with the continuous-valued PSO to select the input feature subset. Allias et al. [7] studied the performance of PSO for feature selection on different classifiers and different population sizes. Miguel et al. [28] developed a hybrid metaheuristic based on variable neighborhood search (VNS) and tabu search (TS) for feature selection in classification. Garcia et al. [29] compared the performance of the metaheuristic strategies (best first, GA, scatter search and VNS) for feature selection in proteomic mass spectrometry data.

Semiconductor quality prediction is a kind of classification problem. The commonly used classification algorithms mainly include logistic regression (LR) [34], naive Bayes (NB) [59], decision tree (DT) [57], artificial neural network (ANN) [2], SVM and so on. These algorithms do not require the participation of professionals, but only extract the model from the relevant data. Logistic regression is a method of classifying by establishing a regression model on the boundary line based on data. Naive Bayes is a classification method based on Bayes' theorem and assuming that features are independent of each other. Decision tree is a non-parametric classifier, which does not need any prior assumptions about the data. Artificial neural network is biologically inspired method designed to simulate the way the human brain processes information. SVM, which is developed on the basis of statistical learning theory, is a powerful machine learning method in dealing with classification problem [17]. Many successful applications in nonlinear classification and function estimation show that SVMs can handle high dimensional data better even with relatively fewer training samples and exhibit a very good generalization ability for complex models [67]. SVM is capable of handling nonlinear and high-dimensional data and has high generalization ability. It gets the structural description of data

distribution by maximizing the interval, which reduces the requirement of data size and distribution. Considering the nonlinear and high-dimensional characteristics of semiconductor data, SVM is selected to solve the quality prediction problem of semiconductor. The kernel function and penalty parameters have remarkable impacts on the classification performance of SVM [55]. Choosing an appropriate kernel function and some suitable parameters can significantly improve the classification accuracy of SVM. A distance function is used by SVM to determine the proximity between each sample data. A common kind of distance is the Euclidean distance, which is usually applied when the training data set is assumed to be evenly distributed. However, for semiconductor, the quality data produced in semiconductor manufacturing process present the characteristics of non-linear and non-uniform distribution. Therefore, the Euclidean distance can not reflect the actual distance between two points in the semiconductor sample data. To address this issue, we instead use the geodesic distance proposed by Tenenbaum [61], which can estimate the actual distance better in a manifold structure.

It is proverbial that the prediction accuracy of the SVM model is not only related to the input characteristics of SVM, but also closely related to the selection of SVM model parameters. Some commonly used SVM parameter selection methods include the empirical method and the grid search combined with cross-validation method. The former method is too subjective and the latter one is time-consuming. In general, the above parameter selection methods can only modify the parameters individually, and can not achieve the collaborative optimization among the parameters. Recently, more and more researchers have applied intelligent optimization algorithms to parameter selection of the SVM model.

In this paper, a feature-selection-based method is proposed for the semiconductor quality prediction problem with high-dimensional quality features. The highlights of this paper can be summarized as follows: 1) Two improved kernel functions of the SVM is proposed to enhance the ability of handling nonlinear data. 2) A hybrid intelligent algorithm is designed to optimize SVM parameters and select features simultaneously. 3) The above method is applied to solve the quality prediction problem for semiconductor and achieves good results.

The rest of the paper is organized as follows. In Section 2, a sketchy review of previous theoretical achievements related to our research is presented. In Section 3, the modified SVM model is set up. In Section 4, a DA-VNS algorithm is designed for feature selection and optimization parameters. In Section 5, a real example about quality prediction of semiconductor products is given to illustrate the proposed model. Finally, we summarize this paper and put forward future research directions in Section 6.

2. Literature review. The relevant research on the quality prediction method are as follows. Xin et al. [68] presented a system based on the application of sequential analysis for the fault prediction of the main diesel engine. Fujimaki et al. [27] proposed a novel anomaly detection system for spacecrafts based on data mining techniques. Toscano [39] studied the potential of various artificial intelligence techniques to predict and analyze the damage for a ball bearing. Long et al. [47] proposed a hybrid evolutionary algorithm featuring a competitive swarm optimizer combined with a local search for intelligent diagnosis of machinery faults. Long et al. [48] applied a sparse echo autoencoder network to the fault diagnosis of delta 3-D printers using attitude data.

The quality prediction for semiconductors has also been extensively studied. Fridgeirsdottir et al. [73] used the statistical method of data mining to judge the faults of the inspection points in semiconductor production lines, which helped improve the qualified rate. Kim et al. [38] applied a polynomial neural network (PNN) to construct a predictive model of plasma etch processes. Bae et al. [9] applied dynamic polynomial neural network (DPNN) and decision tree for data modeling and rule extraction from the ingot fabrication data. Su et al. [60] proposed a new quality prognostics scheme (QPS) for plasma sputtering in thin-film transistor liquid crystal display (TFT-LCD) manufacturing process. Chou et al. [16] constructed a hybrid integrating support vector machine and a genetic algorithm model to implement dynamic wafer quality prediction system. Purwins et al. [56] predicted the Plasma Enhanced Chemical Vapor Deposition (PECVD) Silicon Nitride layer thickness based on metrology and production equipment Fault Detection and Classification (FDC) data. Melhem et al. [50] studied the application of product quality prediction method based on alarm data in the semiconductor manufacturing process. Alagic et al. [5] proposed an approach combining image processing and statistical modeling to quantify and predict the damage intensity in SAM images. Kharaz et al. [4] applied Artificial Neural Network (ANN) to reveal the relationship between semiconductor products end quality state and processes alarm events. Kim et al. [37] used an ensemble of ordinary least squares (OLS) regression and ridge regression to consider the wafer-level and field-level overlay error signatures. The aforementioned references on quality prediction in semiconductor manufacturing process are summarized in Table 1.

TABLE 1. Quality prediction problems in semiconductor manufacturing processes in recent years

Publications	Problems	Methods	Data Driven
Fridgeirsdottir [73]	Fault diagnosis	Data mining	✓
Kim [37]	Prediction of plasma etch processes	PNN	✓
Bae [9]	Modeling and rule extraction of the ingot fabrication	DPNN	✓
Su [59]	Quality prognostics for plasma sputtering	NN	✓
Chou [16]	Prediction of dynamic wafer quality	SVM	✓
Purwins [55]	Prediction of Silicon Nitride layer thickness	Collinearity regression	✓
Melhem [49]	Prediction of batch scrap	Regularized regression	✓
Alagic [5]	Prediction of the damage intensity	Image processing and statistical modeling	✓
Al-Kharaz [4]	Prediction of quality state	ANN	✓
Kim [36]	Prediction of wafers errors	Ordinary least squares regression and ridge regression	✓

Feature selection, a prerequisite for establishing classifiers, has been applied in many fields, such as image recognition, text mining, fault diagnosis, and so on. According to the theoretical principle, feature selection methods can be based on statistics, rough set, manifold, and information theory. Li et al. [43] designed a statistics-based wrapper for feature selection and implemented on financial distress identification with SVM. Derrac et al. [21] proposed a feature selection method based on rough set. Bonev [13] proposed a feature selection method for omnidirectional image classification based on information theory. With the increase of the amount of data, the research on applying intelligent algorithms to feature selection has gradually increased. As an intelligent algorithm with strong local search ability, VNS is favored by many scholars. Mucherino and Liberti [52] studied a VNS-based

heuristic for feature selection in data mining. Garcia et al. [30] designed a VNS strategy for tackling the feature selection problem on high-dimensional datasets. Costa et al. [18] proposed a feature selection method based on VNS for the hierarchical classification context. Chen et al. [14] identified the optimal feature subset by adopting a hybrid VNS with the elitist population strategy.

SVM is a learning method based on hyperplane classification, which is proved to be feasible and applicable. Since proposed in 1995, SVM has activated the attention of a large number of scholars. There are many references referring to the research on the improvement of SVM. Keelthi and Gilbert [36] improved the SMO algorithm and solved the Vo-SVM classifier problem. Cherkassky and Ma [15] investigated a practical selection of hyper-parameters for support vector machines regression (that is, insensitive zone and regularization parameter). Janik and Lobos [33] proposed a new method for power quality classification using SVM neural network based on Radial Basis Function (RBF). Adankon and Cheriet [1] presented an approach for model selection of LS-SVM considering empirical error criteria. Because of its high performance and strong generalization ability, SVM has been widely used in many fields. Lai et al. [41] applied hybrid Support Vector Machine/Gaussian Mixture Model (SVM/GMM) classification model to design a multi-home appliance identification system. Wu et al. [66] established a risk assessment model for textile clothing safety by using support vector machine and verified the effectiveness of the model. Some other references concerning the application of SVM are listed in [54, 22, 11, 72, 73, 25, 40, 63, 10].

With a limited number of samples, designing the SVM classifier with a large number of features is computationally expensive and poorly classified. As a dimension reduction method, feature selection can significantly improve the performance of SVM. Julia et al. [53] presented four unique continuous feature selection approaches directly minimizing the classifier performance to accomplish the feature selection and classification simultaneously. Weber and Basak [64] proposed a support vector machine method with kernel penalty (KP-SVM) to optimize the shape of an anisotropic RBF kernel. Maulik et al. [49] used a novel approach to combine feature (gene) selection and transductive support vector machine (TSVM). Rahman et al. [58] established a new computational model based on feature selection, random forest, and SVM.

The selection of parameters and kernel functions has a significant impact on the performance of SVM. Using intelligent algorithms to select the parameters and kernel functions of SVM is a common and effective method. Zheng and Jiao [71] proposed an effective strategy for automatic selection of support vector machine parameters by using a Genetic algorithm. Unler et al. [62] presented a hybrid filter-wrapper feature subset selection algorithm based on particle swarm optimization (PSO) for support vector machine classification. Ahmadi and Bahadori [3] used a genetic algorithm (GA) to select and optimize the hyperparameters embedded in the LSSVM model. Dong et al. [23] calculated the weights of some factors affecting the return on investment (ROI) based on LSSVM-PSO.

The main work of this paper concerns the performance of a modified SVM coupling with an intelligent algorithm for parameter optimization and feature selection. First, we modified the kernel function of the SVM to improve the ability of handling nonlinear data. Second, a hybrid intelligent algorithm is designed which can optimize SVM parameters and select features simultaneously to get the schemes through combinatorial optimization. Third, we apply the above method to a real

case study to predict the quality of semiconductor products with high-dimensional quality features.

3. Examples. The basic idea of SVM is to nonlinearly map the training data into a higher dimensional feature space via a mapping function, and then find the separating hyperplane which can best divide the samples by solving in the feature space [17]. For linearly separable data, SVM obtains the hyperplane which maximizes the margin (distance) between the training samples and the class boundary. For nonlinearly separable cases, samples are mapped to a high dimensional space where such a hyperplane can be found [6]. The mapping is carried out by a mechanism called the kernel function. This section will introduce the necessary background knowledge and the modified SVM model.

3.1. SVM model. Suppose $(x_1, y_1), \dots, (x_n, y_n)$ are the sample data, where $x_i \in R^m$, $y_i \in -1, +1$, $i = 1, 2, \dots, n$. SVM aims to find a hyperplane that maximizes the interval between the positive data and the negative data in the sample set. The hyperplane in the n -dimensional space is determined by the following equation:

$$\omega^T x + b = 0, \quad (1)$$

where ω and x are both n -dimensional column vectors. ω is the normal vector that determines the direction of the hyperplane. b is the intercept term.

The problem of finding the optimal hyperplane can be considered as the following optimization problem

$$\begin{aligned} \min_{\alpha_i} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \kappa(x_i, y_i) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \\ & \sum_i \alpha_i y_i = 0, \end{aligned} \quad (2)$$

where α_i are the Lagrange multipliers. During the entire construction process of SVM, we only need to know the inner product in the new feature space, rather than the explicit form of nonlinear mapping $\phi(x)$. $\kappa(x, y) = \langle \phi(x), \phi(y) \rangle$ denotes the kernel function that can map points from low-dimensional space to high-dimensional space. According to Cristianini and Shawe-Taylor [19], the kernel function must satisfy the Mercer's theorem (The applicable kernels are limited to symmetric semi-positive definite ones).

Kernel selection is one of the key technologies to improve the ability of SVM classifier. Common SVM kernel functions are shown in Table 2. Among them, RBF is the most commonly used kernel function due to its ability to map sample data from the input space to high-dimensional feature space effectively and efficiently. In this paper, we improve the RBF kernel function in order to get a classifier with better performance.

TABLE 2. Common Kernel Function

Kernel function name	Kernel function representation
Radial basis function	$\kappa(x_i, x_j) = \exp(-\gamma x_i - x_j)$
Linear kernel function	$\kappa(x_i, x_j) = x_i \cdot x_j$
Polynomial kernel function	$\kappa(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Sigmoid kernel function	$\kappa(x_i, x_j) = \tanh[n < x_i \cdot x_j > + \theta]$

3.2. Geodesic distance for kernel functions. Geodesic distance is an important concept in mathematical morphology. For neighbor samples, the geodesic distance approximately reflects the distance between the samples along the data distribution surface using the Euclidean distance. For samples that are far apart, the geodesic distance is the length of the shortest spatial curve connecting the two sample points on the sample distribution surface.

If x_j is one of the k -nearest neighbors of x_i , then the geodesic distance $d_G(x_i, x_j)$ can be calculated by

$$d_G(x_i, x_j) = d_e(x_i, x_j) \quad (3)$$

where $d_e(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j .

If x_j is not one of the k -nearest neighbors of x_i , and $\{x_{i,1}, x_{i,2}, \dots, x_{i,k}\}$ are the k -nearest neighbors of x_i in $\{x_1, x_2, \dots, x_n\}$, then

$$d_G(x_i, x_j) = \min\{d_e(x_i, x_{i,1}) + d_G(x_{i,1}, x_j), \dots, d_e(x_i, x_{i,k}) + d_G(x_{i,k}, x_j)\} \quad (4)$$

Compared with the Euclidean distance, the geodesic distance can better reflect the shape information of the sample distribution and express the actual distance between the samples more accurately. Therefore, the geodesic distance is more suitable for non-linear sampling point distribution surface.

Kernel functions are divided into local kernel functions and global kernel functions. The global kernel function has weak interpolation ability and a strong generalization ability. Compared to global kernel function, the local kernel function only affects the samples in a small range near the test points. RBF kernel is a typical local kernel function, which has the characteristics of weak generalization ability and strong learning ability. For the above properties, we improve the RBF kernel function $\kappa(x_i, x_j) = \exp(-\gamma||x_i - x_j||)$ by using geodesic distance.

Using the geodesic distance to measure the similarity between samples, the RBF kernel function can be improved in the following two ways:

$$\kappa_1(x_i, x_j) = \exp(-\gamma \cdot d_G(x_i - x_j)) \quad (5)$$

and

$$\kappa_2(x_i, x_j) = \exp(-\gamma \cdot (d_G(x_i - x_j) + t_1) + t_2) \quad (6)$$

$\kappa_2(x_i, x_j)$ adds two regularization parameters on the basis of $\kappa_1(x_i, x_j)$ to increase the amplitude and displacement changes of the kernel function, so as to further enhance its generalization ability. To show that κ_1 and κ_2 satisfy the Mercer's theorem which is necessary for classification, we prove that they are all positive semi-definite functions.

Theorem 3.1. $\kappa_1(x_i, x_j) = \exp(-\gamma \cdot d_G(x_i - x_j))$ is a positive semi-definite function.

Proof. According to the definition of geodesic distance, the geodesic distance is the sum of a limited number of Euclidean distances. First, we prove the theorem when the limited number is 2. Without loss of generality, suppose that $\gamma = \frac{1}{2}$. Then we have

$$\begin{aligned}\kappa_1(x_i, x_j) &= \exp(-\gamma \cdot d_G(x_i - x_j)) \\ &= \exp(-\gamma \cdot [(x_i - x_p)^2 + (x_p - x_j)^2]) \\ &= \exp(-\frac{(x_i - x_p)^2}{2}) \cdot \exp(-\frac{(x_p - x_j)^2}{2})\end{aligned}$$

Note that

$$\begin{aligned}h(t_s) &= \exp(-\frac{(x_i - x_p)^2}{2}) = \exp(-\frac{t_s^2}{2}) = E[e^{it_s Z}] \\ h(t_k) &= \exp(-\frac{(x_p - x_j)^2}{2}) = \exp(-\frac{t_k^2}{2}) = E[e^{it_k Z}]\end{aligned}$$

where $h(t)$ is the characteristic function of a random variable Z with $N(0, 1)$ distribution, and i denotes the imaginary unit. For real numbers t_1, t_2, \dots, t_n and a_1, a_2, \dots, a_n , we have

$$\begin{aligned}\sum_{i,j=1}^n a_i a_j \kappa_1(x_i, x_j) &= \sum_{s,k=1}^n a_s a_k h(t_s) h(t_k) \\ &= \sum_{s,k=1}^n a_s h(t_s) \cdot a_k h(t_k) \\ &= \sum_{s=1}^n a_s h(t_s) \cdot \sum_{k=1}^n a_k h(t_k) \\ &= E[\sum_{s=1}^n a_s e^{it_s Z}] \cdot E[\sum_{k=1}^n a_k e^{it_k Z}]\end{aligned}$$

$$\begin{aligned}
&= E\left[\sum_{s=1}^n a_s e^{it_s Z}\right] \cdot E\left[\sum_{k=1}^n a_k e^{it_k Z}\right] \\
&= \left(E\left[\sum_{s=1}^n a_s e^{it_s Z}\right]\right)^2 \geq 0
\end{aligned}$$

which entails that $\kappa_1(x_i, x_j)$ is a positive semi-definite function, that is, the kernel.

In a similar way, we can prove that $\kappa_1(x_i, x_j)$ is a positive semi-definite function when the limited number is larger than 2. \square

Theorem 3.2. $\kappa_2(x_i, x_j) = \exp(-\gamma \cdot (d_G(x_i - x_j) + t_1) + t_2)$ is a positive semi-definite function.

Proof.

$$\begin{aligned}
\kappa_2(x_i, x_j) &= \exp(-\gamma \cdot (d_G(x_i - x_j) + t_1) + t_2) \\
&= \exp(-\gamma \cdot (d_G(x_i - x_j) + t_1 - t_2/\gamma))
\end{aligned}$$

Note that

$$t_1 - t_2/\gamma = D_t, -\infty < D_t < +\infty$$

That is, $d_G(x_i - x_j) + D_t$ is obtained by adding or subtracting a distance from $d_G(x_i - x_j)$. According to the proof process of Theorem 1, we know that $\kappa_2(x_i, x_j) = \exp(-\gamma \cdot (d_G(x_i - x_j) + t_1) + t_2)$ is a positive semi-definite function. \square

4. DA-VNS algorithm. Inspired by human combat and learning ability, the Du-elist algorithm (DA) is first proposed by Biyanto et al. [12] based on Genetic algorithm. The local search capability of DA is not as satisfactory as its global search capability. Variable neighborhood search (VNS) is a metaheuristic algorithm with strong local search capability for solving combinatorial problems. In this paper, VNS is integrated into DA to improve its local search ability.

When using SVM as the classifier, it is important to choose appropriate parameters for SVM since different values of parameters will affect the selection of a subset of sample features, and vice versa. Hence, to obtain a classifier with higher precision, it is necessary to select the feature subset and parameters simultaneously. The selected parameters through combinatorial optimization can make SVM have higher training accuracy, whereas unreasonable choices of parameters will reduce the learning ability and promotion ability of SVM. In our model, there are several specific parameters that will affect the performance of SVM. The regularization parameter C is used to balance the model complexity and the kernel function parameter γ defines the kernel width of the RBF kernel function. t_1 and t_2 are two regulating parameters to increase the amplitude and displacement changes of the kernel function, respectively. In this paper, a DA-VNS algorithm is designed for feature selection and optimizing parameters. The flowchart of the proposed method is designed in Fig. 2. And Fig.3 is the process of the DA-VNS algorithm.

4.1. Encoding. Encoding is an important step in DA-VNS algorithm, which converts the parameters (t_1, t_2, C, γ) and feature subsets into binary encoding. The encoding of DA-VNS algorithm is composed of parameters and feature subsets, as shown in Fig. 4.

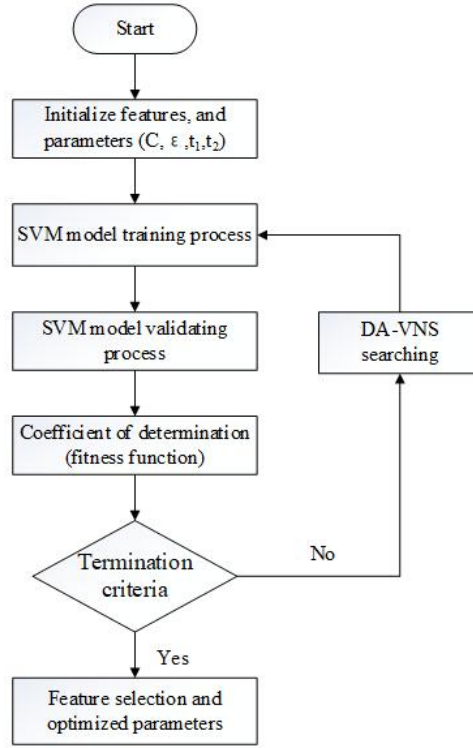


FIGURE 2. Flowchart of proposed method

$b_{t_1}^1 \sim b_{t_1}^{n_{t_1}}$, $b_{t_2}^1 \sim b_{t_2}^{n_{t_2}}$, $b_C^1 \sim b_C^{n_C}$, $b_\gamma^1 \sim b_\gamma^{n_\gamma}$, and $b_f^1 \sim b_f^{n_f}$ are the binary strings of parameters t_1, t_2, C, γ , and feature subsets, respectively. $n_{t_1}, n_{t_2}, n_C, n_\gamma$, and n_f are the numbers of binary digits of t_1, t_2, C, γ , and feature subsets, respectively. In this paper, the selection of parameters and the selection of features are carried out simultaneously. That is, we find the optimum parameters corresponding to the feature subset while selecting features.

4.2. Fighting capability function. The fighting capability function should consider the SVM classification accuracy rate, the selected feature number, and the feature cost. A high fighting capability depends on high classification accuracy, small feature number, and low feature cost. Let θ be a given model. $R(\theta)$ is the classification accuracy rate of model θ , and w_c is the weight of it, usually between 0.75 and 1. $C_f(\theta)$ is the complexity of model θ , and w_f is the weight of it. Then, the fighting capability function $q(\theta)$ is

$$q(\theta) = w_c \cdot R(\theta) + w_f \cdot 1/C_f(\theta), w_c \in [0.75, 1] \quad (7)$$

$$C_f(\theta) = c_{f1} \log_2(n_{SV}) + c_{f2} \log_2\left(\sum \beta_i k_i\right), \beta_i = 0, 1, \quad (8)$$

where β_i is a boolean variable representing that whether the i -th feature is selected. $\beta_i = 1$ represents that the i -th feature is selected. Constants c_{f1} and c_{f2} fix the trade-off between classification rate improvement and complexity reduction. n_{SV}

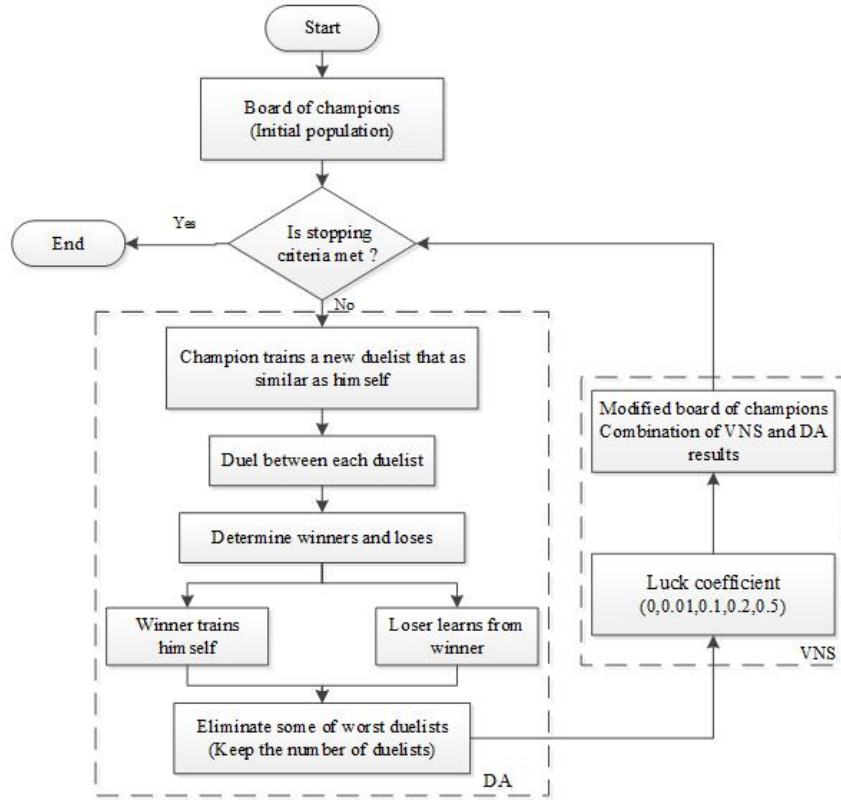


FIGURE 3. Flowchart of DA-VNS algorithm

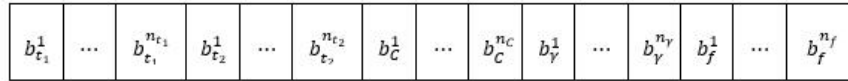


FIGURE 4. Encoding of DA-VNS algorithm

represents the number of the support vectors. k_i denotes the cost for the extraction of the i -th feature. When those costs are unknown, $k_i = 1$ is used for all features.

4.3. Duel scheduling between duelists. Board of champions is determined to keep the best duelist in the game. Each champion develops a new duelist, and the duelists will have a one-on-one duel. The duel schedule is set at random. Duelists will use their fighting capabilities and luck to decide winners and losers. If the sum of duelist A's fighting capability and luck is higher than B, then A is the winner and B is the loser. The pseudocode to determine the winner and the loser is shown in Algorithm 1.

4.4. Duelist's improvement. In this step, each loser and winner has an opportunity to improve their fighting capabilities. The loser learns from the winner to improve fighting capability (by replacing part of its binary array with the winner's

Algorithm 1. Determination of the winner and the loser

```

1.Duelist A and B, Luck_coeff
2.FC=Fighting capability; LC=Luck_coeff
3.A(Luck)=A(FC)*(LC+(random(0,1)*LC));
4.B(Luck)=B(FC)*(LC+(random(0,1)*LC));
5.if ((A(FC)+A(Luck))>= (B(FC)+B(Luck)))
6. A(Winner) = 1;
7. B(Winner) = 0;
8.else
9. A(Winner) = 0;
10. B(Winner) = 1;
11.end if

```

binary array value). At the same time, the winner will try to innovate to improve fighting capability (by changing the value of its binary array to a new value). To maintain the number of duelists in the game, some of the worst combatants will be eliminated. The pseudocode of duelist's improvement is displayed in Algorithm 2.

Algorithm 2. Duelist's Improvement

```

1.Duelist A and B, Duelist_length, Prob_innovate; Prob_learn
2.if A(Winner) = 1
3. for i=1:(Duelist_length)
4. r=random(0,1)
5. if r<Prob_innovate
6. if A[i]=1
7. A[i]=0
8. else A[i]=1
9. end if
10. end if
11. end for
12.else
13. for i=1:(Duelist_length)
14. r=random(0,1)
15. if r<Prob_learn
16. B[i]=A[i]
17. end if
18. end for
19.end if

```

4.5. Framework of DA-VNS. DA algorithm has strong global search ability. However, with the operation of the algorithm, DA is easy to rely on solution space, resulting in slow solution evolution. In addition, DA has poor local search ability, making the entire search process time consuming, thereby reducing the efficiency of the algorithm. In contrast, VNS has a higher local search ability, which can make up for the above defects. Combining DA and VNS, the proposed algorithm can effectively avoid the local optima and the blindness search of DA algorithm. An important step in DA algorithm is the one-to-one duel between the duelists. As mentioned above, luck is an influential factor in determining winners and losers

in a duel. Different luck coefficients may lead to different performance of DA in different experiments. It is difficult to find a fixed luck coefficient suitable for all cases. Thus, we define DA with a particular luck coefficient as a neighborhood of VNS. Five neighborhoods are set up for DA-VNS algorithm in this paper, that is, DA (Luckcoeff = 0), DA (Luckcoeff = 0.01), DA (Luckcoeff = 0.1), DA (Luckcoeff = 0.2), and DA (Luckcoeff = 0.5).

Algorithm 3. DA-VNS

```

1.Initial solution  $s$ , the neighborhoods  $N_k, k = 1, 2, 3, 4, 5$ 
2.Iteration
3.While stopping rule is not satisfied do
4.  $k = 1$ 
5. while  $k \leq 5$  do
6. if  $f(s') > f(s)$  then
7.  $s' = s, k = 1$ 
8. else
9.  $k = k + 1$ 
10. end if
11. end while
12.end while

```

5. Experiments and results. In this section, we conduct some experiments using the SECOM data set in the UCI (University of California, Irvine) database [8] to validate the performance of the proposed method for quality prediction. The data set is from a semiconductor manufacturing process, and a semiconductor's quality is divided into two categories: qualified and unqualified. The data set contains 1567 samples, of which the number of qualified samples is 1463 and the number of unqualified samples is 104. In addition, each sample has 590 quality features.

A difficulty in product quality classification and recognition lies in the fact that the variation in quality features of products in the same category may be very large, while the variation between products in different categories may be very small. The quality features of products in the same category may vary greatly due to the deployment of personnel or equipment in the production process. Products in different categories may have few differences in quality features due to equipment errors or resource matching deviations in a certain production process.

There is a manifold structure when quality feature data in same category of semiconductor products are arranged together according to their similarities [65]. Therefore, we can consider that the quality data of real semiconductor products are located in a low-dimensional manifold embedded in a high-dimensional space. It can be concluded that the data set of semiconductor quality is a non-linear manifold formed by coupling of some specific intrinsic variables. For effective semiconductor quality prediction, the improved SVM based on geodesic distance is a suitable model.

5.1. Data preparation. Since there are many missing values in the data set, we need to perform data cleaning operations on the original data before using the data set for subsequent mining. The data cleaning process in this experiment follows two principles: 1) the data set retains as many attributes as possible, and 2) the positive and negative samples in the data set can not be too unbalanced. Finally,

we obtain a sample data set with size and dimension of 240×225 . It means that the data set ultimately retains 240 samples, containing 140 positive samples and 100 negative samples, each with 225 quality features. This gives us a set of study data to be used in our experiments.

To eliminate the influence of different dimensions on the numerical values, further normalization of data is required. The normalization formula is as follows.

$$a'_{ij} = \frac{a_{ij} - a_{imin}}{a_{imax} - a_{imin}}, \quad (9)$$

where a_{ij} is the initial sample data to be normalized, a_{imin} and a_{imax} are the minimum and maximum values in the column sample values.

5.2. Experimental study. All experiments are run on an Intel Core i7-7700 8 GB, the Microsoft Windows 10 operating system and the development environment of Python 3.6.6, PyCharm 2018.3. The parameter settings are shown in Table 3.

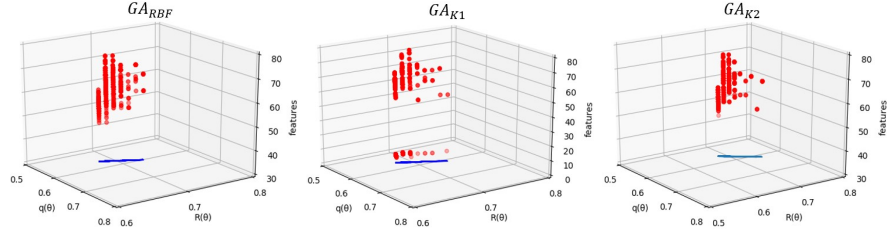
TABLE 3. List of preset parameters in DA-VNS

Parameters	$DA - VNS_{RBF}$	$DA - VNS_{\kappa_1}$	$DA - VNS_{\kappa_2}$
Population size	100	100	100
Iteration times	500	500	500
Nearest neighbor number	/	5	5
Learning probability	0.8	0.8	0.8
Innovation probability	0.1	0.1	0.1
Mutation probability	0.1	0.1	0.1
Search range of penalty parameter C	$[10^{-3}, 10^3]$	$[10^{-3}, 10^3]$	$[10^{-3}, 10^3]$
Search range of kernel width γ	$[2^{-6}, 2^6]$	$[2^{-6}, 2^6]$	$[2^{-6}, 2^6]$
Search range of amplitude regulating parameter t_1	/	/	$[-10, 10]$
Search range of displacement regulating parameter t_2	/	/	$[-10, 10]$
Luck coefficient	$\{0, 0.01, 0.1, 0.2, 0.5\}$	$\{0, 0.01, 0.1, 0.2, 0.5\}$	$\{0, 0.01, 0.1, 0.2, 0.5\}$
w_c, w_f, c_{f1}, c_{f2}	0.8, 0.2, 0.8, 0.2	0.8, 0.2, 0.8, 0.2	0.8, 0.2, 0.8, 0.2

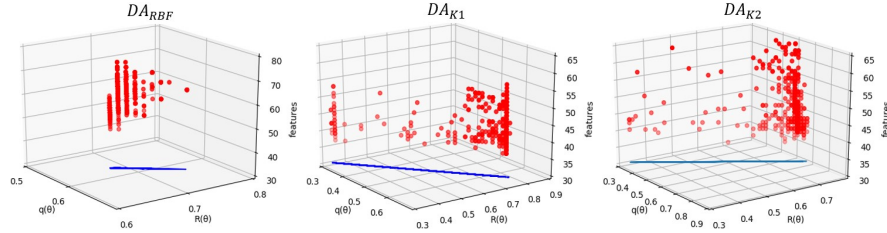
We design a set of comparative experiments to evaluate the performance of the proposed algorithm in solving SVM classification problems. Intelligent algorithms, i.e., GA and DA are used to compare with the proposed algorithm. According to the above specified parameters, 500 iterations have been performed based on the training sets and testing sets, and the running results of GA, DA and DA-VNS algorithm with different kernel functions are reported in Figs. 5 and 6.

As shown in Fig. 5, in general, the solutions obtained by DA-VNS algorithm is superior to DA, and the solutions obtained by DA algorithm is superior to GA. By comparing Fig. 5(c) with (a) and (b), the solutions obtained by DA-VNS algorithm are mostly clustered around the optimal solution. In addition, the solutions obtained by GA algorithm are more concentrated than those obtained by DA and DA-VNS algorithm, regardless of the kernel function. In other words, the search scope of the GA algorithm is smaller than the other two algorithms. From Fig. 5(a), (b) and (c), it is obvious to find that κ_2 kernel function is superior to κ_1 kernel function and κ_1 kernel function is superior to RBF kernel function for each algorithm.

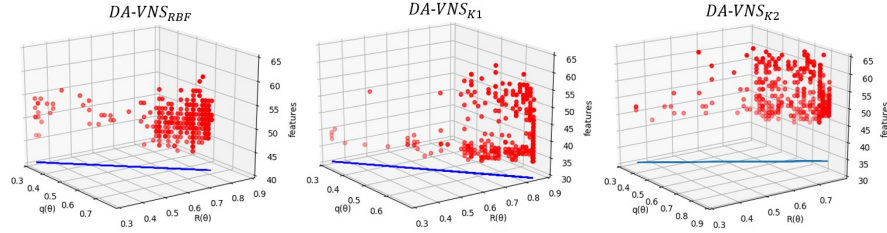
Fig. 5(a) shows that the distribution shapes of the solutions obtained by the three kinds of kernels are similar, indicating that for the GA algorithm, the influence of



(a) Running results of GA with different kernel functions



(b) Running results of DA with different kernel functions



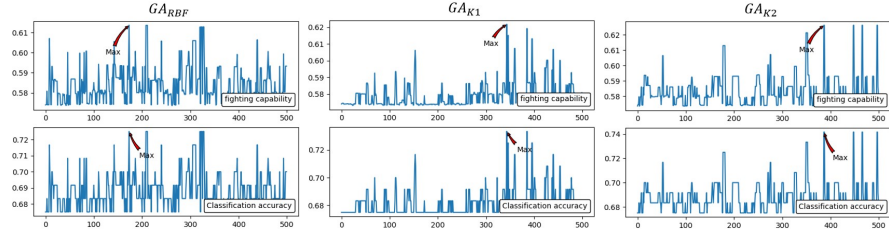
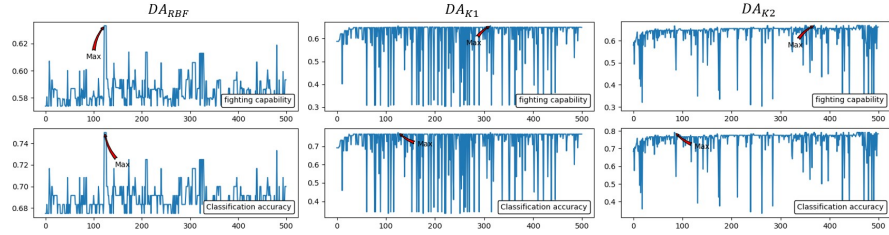
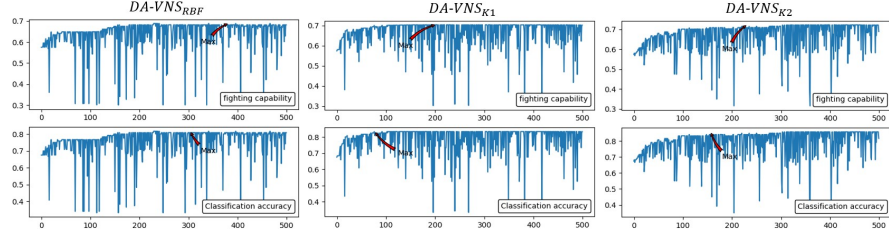
(c) Running results of DA-VNS with different kernel functions

FIGURE 5. Running results after 500 iterations of GA, DA and DA-VNS algorithms respectively (The blue points indicate the projection of the solutions on $(q(\theta), R(\theta))$.)

the kernel functions lies in the fact that a better solution can be found at a certain point. From Fig. 4(b), κ_1 and κ_2 kernel functions have changed the distribution shapes of the solutions, and κ_2 kernel function increases the search range of solutions compared with κ_1 . As shown in Fig. 4(c), on one hand, the solution obtained by $DA - VNS_{\kappa_2}$ is better than that obtained by $DA - VNS_{\kappa_1}$ and $DA - VNS_{RBF}$. On the other hand, the number of the feather set selected by $DA - VNS_{RBF}$ are centralized between 45 and 55 while the fluctuation ranges of the number of feature sets obtained by the other two algorithms are 22 and 25, respectively. It means that the modified kernel functions expand the search scope of solution and the kernel function κ_2 is slightly better than κ_1 .

To sum up, the modified kernel function can expand the search range of the solutions. DA-VNS algorithm not only broadens the search scope of solutions, but also always makes the solutions at a high level.

Fig. 6 shows the evolution of the best $q(\theta)$ and $R(\theta)$ calculated by GA, DA and DA-VNS over 500 iterations. From Fig. 6(a), although the GA algorithm with RBF kernel function finds the optimal solution earlier than GA algorithm with κ_1 and κ_2 ,

(a) The evolution of the best $q(\theta)$ and $R(\theta)$ calculated by GA with different kernel functions(b) The evolution of the best $q(\theta)$ and $R(\theta)$ calculated by DA with different kernel functions(c) The evolution of the best $q(\theta)$ and $R(\theta)$ calculated by DA-VNS with different kernel functionsFIGURE 6. The evolution of the best $q(\theta)$ and $R(\theta)$ for GA, DA and DA-VNS respectively over 500 iterations

the optimal solutions obtained by the GA algorithm with κ_1 and κ_2 is better than that with RBF kernel function. In Fig. 6(b), we can see that the modified kernel functions can make the solutions converge to the optimal solution continuously. Fig. 6(c) shows that DA-VNSRBF can successfully find the global optimum of input feature set and hyperparameters within 380 iterations. The values of best fighting capability gradually improve from 1st to 160th iteration and start to gradually converge to an optimum after around 160 generations. There are no significant improvements after the 200th iteration. In addition, the $DA-VNS_{RBF}$ algorithm uses 300 iterations to find the optimal classification accuracy. It is obvious to find that $DA-VNS_{\kappa_1}$ and $DA-VNS_{\kappa_2}$ converge to the best fighting capability with 200 and 230 iterations, and converge to the optimal classification accuracy with 80 and 150 iterations, respectively. That is to say, the modified SVM kernel functions can definitely improve the speed of convergence to the optimal solution.

Table 3 summarizes the results of the comparative experiments. In addition to GA and DA, we also designed the Information Gain (IG) (which is a commonly used algorithm in feature selection) to compare with the proposed algorithm. The optimal parameters selection for each algorithm are also shown in Table 4.

TABLE 4. Comparison of performance between DA-VNS and other algorithms

Algorithm	Optimal parameters				$q(\theta)$	$R(\theta)$	Selected features
	C	γ	t_1	t_2			
GA_{RBF}	78.54	3.36	/	/	0.6136	0.725	60
GA_{κ_1}	73.11	0.80	/	/	0.6217	0.7333	47
GA_{κ_2}	11.21	0.94	1.53	4.41	0.6262	0.7416	56
GA_{RBF}	96.10	0.52	/	/	0.6329	0.75	64
DA_{κ_1}	62.34	0.49	/	/	0.6559	0.775	43
DA_{κ_1}	42.04	0.49	9.81	9.96	0.6695	0.7917	41
IG	/	/	/	/	0.6957	0.8105	24
$GA - VNS_{RBF}$	21.32	1.31	/	/	0.6882	0.8033	48
$DA - VNS_{\kappa_1}$	60.82	1.60	/	/	0.7038	0.8333	36
$DA - VNS_{\kappa_2}$	96	0.91	-2.69	8.77	0.7221	0.8583	49

To show the effect of the DA-VNS algorithm and the modified kernel functions, we calculate the improvements on $q(\theta)$ and $R(\theta)$ for a clearer explanation according to Ellefsen et al. [24]. The improvements on $q(\theta)$ and $R(\theta)$ are noticeable when using the $DA - VNS_{\kappa_1}$ and the $DA - VNS_{\kappa_2}$ algorithm, as shown in Tables 5 and 6.

TABLE 5. Performance improvement of $q(\theta)$ between DA-VNS and other algorithms. $Improvement = \frac{q(\theta) - q'(\theta)}{q(\theta)} * 100\%$.

Improvement	GA_{RBF}	GA_{κ_1}	GA_{κ_2}	DA_{RBF}	DA_{κ_1}	DA_{κ_2}	IG	$DA - VNS_{RBF}$	$DA - VNS_{\kappa_1}$
$DA - VNS_{\kappa_1}$	12.82	11.66	11.02	10.07	6.81	4.87	1.15	2.22	/
$DA - VNS_{\kappa_2}$	15.03	13.90	13.28	12.35	9.17	7.28	3.66	4.69	2.53

TABLE 6. Performance improvement of $R(\theta)$ between DA-VNS and other algorithms. $Improvement = \frac{R(\theta) - R'(\theta)}{R(\theta)} * 100\%$.

Improvement	GA_{RBF}	GA_{κ_1}	GA_{κ_2}	DA_{RBF}	DA_{κ_1}	DA_{κ_2}	IG	$DA - VNS_{RBF}$	$DA - VNS_{\kappa_1}$
$DA - VNS_{\kappa_1}$	13.00	12.00	11.00	10.00	7.00	4.99	2.74	3.60	/
$DA - VNS_{\kappa_2}$	15.53	14.56	13.60	12.62	9.70	7.76	5.57	6.41	2.91

On an average of 500 iterations, $DA - VNS_{\kappa_1}$ outperforms eight optimization algorithms by 12.82%(GA_{RBF}), 11.66%(GA_{κ_1}), 11.02%(GA_{κ_2}), 10.07%(DA_{RBF}), 6.81%(DA_{κ_1}), 4.87%(DA_{κ_2}), 1.15% (IG), 2.22%($DA - VNS_{RBF}$) in the matter of $q(\theta)$. $DA - VNS_{\kappa_2}$ outperforms nine optimization algorithms by 115.03%(GA_{RBF}), 13.09%(GA_{κ_1}), 13.28%(GA_{κ_2}), 12.35%(DA_{RBF}), 9.17%(DA_{κ_1}), 7.28%(DA_{κ_2}), 3.66% (IG), 4.69%($DA - VNS_{RBF}$), 2.53%($DA - VNS_{\kappa_1}$) in the matter of $q(\theta)$. Similar results are shown in Table 6 about the improvements in the matter of $R(\theta)$, which will not be covered here.

Furthermore, we compare the performance between DA-VNS-SVM and some common machine learning algorithms such as logistic regression, naive Bayes, decision tree, random forest, and artificial neural network. For DA-VNS-SVM, we still choose the parameter settings in Table 3. The constructed ANN model includes 5 hidden layers with ReLU activation function. In the decision tree algorithm, the maximum depth of the decision tree is set to 5. The number of trees in the random forest is 30, and Gini index criterion function is selected to measure the quality of a split in the random forest algorithm. According to the above settings, the comparative experiments have been performed on SECOM data set and the results are reported in Table 7.

TABLE 7. Comparison of performance between DA-VNS-SVM and common machine learning algorithms

Algorithm	Accuracy
Logistic Regression (LR)	0.4917
Naive Bayes (NB)	0.6167
Artificial Neural Network (ANN)	0.6417
Decision Tree (DT)	0.658
Random Forest (RF)	0.667
DA-VNS $_{\kappa_1}$ -SVM	0.7038
DA-VNS $_{\kappa_2}$ -SVM	0.7221

Table 7 displays the results the classification accuracy of seven algorithms. Logistic regression has the worst classification accuracy among these algorithms. The possible reason is that it is a linear classifier, which is not suitable for nonlinear data. Random forest uses multiple decision trees to train and predict samples, so its classification accuracy is higher than that of decision trees. The DA-VNS-SVM has better classification accuracy compared with the other five machine learning algorithms. Furthermore, we outlines the performance improvement between DA-VNS-SVM and other algorithms in Figure 7. Specifically, DA-VNS $_{\kappa_1}$ -SVM outperforms the other algorithms by 30.14% (Logistic Regression), 12.38% (Naive Bayes), 8.82% (Artificial Neural Network), 6.51% (Decision Tree), and 5.23% (Random Forest), respectively. DA-VNS $_{\kappa_2}$ -SVM outperforms the other algorithms by 31.91% (Logistic Regression), 14.60% (Naive Bayes), 11.13% (Artificial neural network), 8.88% (Decision Tree), and 7.63% (Random Forest), respectively. Therefore, the proposed DA-VNS-SVM has achieved a good performance in solving the quality prediction problem for semiconductor.

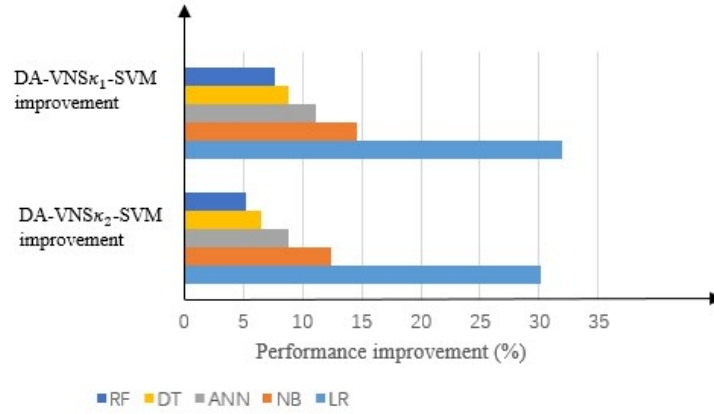


FIGURE 7. Performance improvement between DA-VNS-SVM and other algorithms

6. Conclusion. In this paper, a DA-VNS-SVM method is proposed for the semiconductor quality prediction with high-dimensional quality features. In the proposed method, the kernel function of SVM is improved. Meanwhile, a hybrid intelligent algorithm is designed to select features and optimize the parameters of SVM. The main contributions of this paper can be summarized as follows: 1) Considering the characteristics of semiconductor quality data, we use the geodesic distance instead of the Euclidean distance to modify the SVM kernel function. In addition, two regulating parameters are added to the modified kernel function to enhance its generalization ability. 2) A DA-VNS algorithm is designed to optimize the parameters and select features for SVM models. The combination of DA and VNS improves the local search capability and the speed of convergence to the optimal solution. 3) We compare our DA-VNS-SVM with other classification algorithms such as logistic regression, naive Bayes, decision tree, random forest, and artificial neural network. The experiment result shows that the DA-VNS-based SVM outperforms the other classification algorithms in terms of the quality prediction accuracy.

The method proposed in this paper aims at solving quality prediction problem for semiconductor. However, it also can be applied to other problems in other industries. Because the proposed DA-VNS-SVM method is essentially a classification algorithm for the datasets which contain high dimensional and nonlinear characteristics. It is applicable to any industry that fits both of these points, such as aircraft engine fault diagnosis, food safety inspection, and disaster weather prediction, etc.

Our study can be extended in the following future research directions. First, runtime may be further reduced by exploring more computational-efficient SVM algorithms, faster parameter tuning mechanism, and adaptive cross-validation. Second, parallelization techniques and methods are worth exploring and utilizing to improve learning performance and reduce the computational cost in the feature selection model, allowing the feasible application to larger data and complex data. Third, by committing both runtime reduction and algorithm stability, future research may aim at building a practical feature selection model based on heuristic techniques for solving the large-scale problems. Finally, with the development of data acquisition and processing technology, a large number of uncertain data such

as relational data, semi-structured data and streaming data are generated. How to preprocess and model uncertain data is also a future research direction for us.

Acknowledgments. This work is supported by The National Key Research and Development Program of China (2019YFB1705300), the Fundamental Research Funds for the Central Universities (Nos. JZ2020HGTB0035), the National Natural Science Foundation of China (Nos. 72071056, 71922009, 71801071, 71871080, 71601065, 71690235, 71501058, 71601060), and Innovative Research Groups of the National Natural Science Foundation of China (71521001), Anhui Province Natural Science Foundation (No. 1908085MG223), the Fundamental Research Funds for the Central Universities (Nos. JZ2019HGTA0051, JZ2019HGBZ0131), Base of Introducing Talents of Discipline to Universities for Optimization and Decision-making in the Manufacturing Process of Complex Product (111 project), the Project of Key Research Institute of Humanities and Social Science in University of Anhui Province, Open Research Fund Program of Key Laboratory of Process Optimization and Intelligent Decision-making (Hefei University of Technology), Ministry of Education.

REFERENCES

- [1] M. M. Adankon and M. Cheriet, [Model selection for the LS-SVM. Application to handwriting recognition](#), *Pattern Recognition*, **42** (2009), 3264–3270.
- [2] S. Agatonovic-Kustrin and R. Beresford, [Basic concepts of artificial neural network \(ANN\) modeling and its application in pharmaceutical research](#), *Journal of Pharmaceutical & Biomedical Analysis*, **22** (2000), 717–727.
- [3] M. A. Ahmadi and A. Bahadori, [A LSSVM approach for determining well placement and conning phenomena in horizontal wells](#), *Fuel*, **153** (2015), 276–283.
- [4] M. Al-Kharaz, B. Ananou, M. Ouladsine, M. Combal and J. Pinaton, [Quality prediction in semiconductor manufacturing processes using multilayer perceptron feedforward artificial neural network](#), *2019 8th International Conference on Systems and Control (ICSC)*, (2019), 423–428.
- [5] D. Alagić, O. Bluder and J. Pilz, [Quantification and prediction of damage in SAM images of semiconductor devices](#), *International Conference Image Analysis and Recognition*, Lecture Notes in Computer Science, 10882, Springer, Cham, 2018, 490–496.
- [6] E. Alba, J. Garcia-Nieto, L. Jourdan and E. Talbi, [Gene selection in cancer classification using GPSO/SVM and GA/SVM hybrid algorithms](#), *IEEE Congress on Evolutionary Computation, IEEE*, (2008), 284–290.
- [7] N. Alias, M. N. M. M. Noor, M. N. Ismail, K. de Silva, [A hybrid Gini PSO-SVM feature selection: An empirical study of population sizes on different classifier](#), *International Conference on Artificial Intelligence*, (2013), 107–110.
- [8] Asuncion A, Newman DJ. *UCI Machine Learning Repository*, 2007.
- [9] H. Bae, S. S. Kim, K. B. Woo, G. S. May and D. K. Lee, Fault detection, diagnosis, and optimization of wafer manufacturing processes utilizing knowledge creation, *International Journal of Control, Automation, and Systems*, **4** (2006), 372–381.
- [10] M. G. Bari, X. Ma and J. Zhang, [PeakLink: A new peptide peak linking method in LC-MS/MS using wavelet and SVM](#), *Bioinformatics*, **30** (2014), 2464–2470.
- [11] A. Bazzani, A. Bevilacqua, D. Bollini, et al., [An SVM classifier to separate false signals from microcalcifications in digital mammograms](#), *Physics in Medicine and Biology*, **46** (2001), 1651–1651.
- [12] T. R. Biyanto, et al., [Duelist algorithm: An algorithm inspired by how duelist improve their capabilities in a duel](#), *The Seventh International Conference on Swarm Intelligence*, 2016, 39–47.
- [13] B. Bonev, Feature selection based on information theory, *Universidad de Alicante*, 2010. Available from: <http://hdl.handle.net/10045/18362>.
- [14] W. Chen, Z. Li and J. Guo, [A VNS-EDA algorithm-based feature selection for credit risk classification](#), *Mathematical Problems in Engineering*, **2020** (2020), 14 pp.

- [15] V. Cherkassky and Y. Ma, [Practical selection of SVM parameters and noise estimation for SVM regression](#), *Neural Networks*, **17** (2004), 113–126.
- [16] P.-H. Chou, M.-J. Wu and K.-K. Chen, [Integrating support vector machine and genetic algorithm to implement dynamic wafer quality prediction system](#), *Expert Systems with Applications*, **37** (2010), 4413–4424.
- [17] C. Cortes and V. Vapnik, Support-Vector Networks, *Machine Learning*, **20** (1995), 273–297.
- [18] H. Costa, L. R. Galvao, L. H. C. Merschmann and M. J. F. Souza, A VNS algorithm for feature selection in hierarchical classification context, *Electronic Notes in Discrete Mathematics*, **66** (2018), 79–86.
- [19] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
- [20] K. Fridgeirsdottir, R. Akella, L.-M. Al, [Statistical methodology for yield enhancement via baseline reduction](#), *Advanced Semiconductor Manufacturing Conference & Workshop*, (1998), 77–81.
- [21] J. Derrac, C. Cornelis, S. García and F. Herrera, Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection, *Information Sciences*, **186** (2012), 73–92.
- [22] S. Dong, Y. Zhang, Z. He, et al., [Investigation of support vector machine and back propagation artificial neural network for performance prediction of the organic Rankine cycle system](#), *Energy*, **144** (2018), 851–864.
- [23] R. Dong, J. Xu and B. Lin, [ROI-based study on impact factors of distributed PV projects by LSSVM-PSO](#), *Energy*, **124** (2017), 336–349.
- [24] A. L. Ellefsen, E. Bjorlykhaug, V. Esøy, et al., [Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture](#), *Reliability Engineering & System Safety*, **183** (2019), 240–251.
- [25] A. O. Filho, A. C. Silva, A. C. de Paiva, et al., [Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM](#), *Journal of Digital Imaging*, **30** (2017), 812–822.
- [26] B. Frénay, G. Doquire and M. Verleysen, [Is mutual information adequate for feature selection in regression](#), *Neural Networks*, **48** (2013), 1–7.
- [27] R. Fujimaki, T. Yairi and K. Machida, [An anomaly detection method for spacecraft using relevance vector learning](#), *Pacific-Asia Conference on Knowledge Discovery & Data Mining*, Lecture Notes in Computer Science, 3518, Springer, Berlin, Heidelberg, 2005, 785–790.
- [28] M. Garcia-Torres, F. C. Garcia-López, B. Melián-Batista, J. A. Moreno-Pérez and J. M. Moreno-Vega, Solving feature subset selection problem by a hybrid, *Hybrid Metaheuristics*, (2004), 59–68.
- [29] M. Garcia-Torres, R. Armananzas, C. Bielza, et al., [Comparison of metaheuristic strategies for peakbin selection in proteomic mass spectrometry data](#), *Information Sciences*, **222** (2013), 229–246.
- [30] M. García-Torres, F. Gómez-Vela, B. Melián-Batista and M. Moreno-Vega, [High-dimensional feature selection via feature grouping: A variable neighborhood search approach](#), *Information Sciences*, **326** (2016), 102–118.
- [31] J. Hua, W. D. Tembe and E. R. Dougherty, [Performance of feature-selection methods in the classification of high-dimension data](#), *Pattern Recognition*, **42** (2009), 409–424.
- [32] C.-L. Huang and J.-F. Dun, [A distributed PSO-SVM hybrid system with feature selection and parameter optimization](#), *Applied Soft Computing*, **8** (2008), 1381–1391.
- [33] P. Janik and T. Lobos, [Automated classification of power-quality disturbances using SVM and RBF networks](#), *IEEE Transactions on Power Delivery*, **21** (2006), 1663–1669.
- [34] A. C. Janssens, Y. Deng, G. J. Borsboom, et al., [A new logistic regression approach for the evaluation of diagnostic test results](#), *Medical Decision Making*, **25** (2005), 168–177.
- [35] Y.-S. Jeong, B. Kim and Y.-D. Ko, [Exponentially weighted moving average-based procedure with adaptive thresholding for monitoring nonlinear profiles: Monitoring of plasma etch process in semiconductor manufacturing](#), *Expert Systems with Applications*, **40** (2013), 5688–5693.
- [36] S. S. Keerthi and E. G. Gilbert, [Convergence of a generalized SMO algorithm for SVM classifier design](#), *Machine Learning*, **46** (2002), 351–360.
- [37] S. J. Kim, H. G. Yoon, K. B. Lee, et al., [Hybrid overlay modeling for field-by-field error correction in the photolithography process](#), *IEEE Transactions on Semiconductor Manufacturing*, **33** (2020), 53–61.

- [38] B. Kim, D. W. Kim and G. T. Park, [Prediction of plasma etching using a polynomial neural network](#), *IEEE Transactions on Plasma Science*, **31** (2003), 1330–1336.
- [39] T. Kolodziejczyk, R. Toscano, S. Fouvry and G. Morales-Espejel, [Artificial intelligence as efficient technique for ball bearing fretting wear damage prediction](#), *Wear*, **268** (2010), 309–315.
- [40] M. Kumar, M. Bhasin, N. K. Natt, et al. BhairPred: prediction of β -hairpins in a protein from multiple alignment information using ANN and SVM techniques, *Nucleic Acids Research*, **33** (2015), 154–159.
- [41] Y.-X. Lai, C.-F. Lai, Y.-M. Huang and H.-C. Chao, [Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home](#), *Information Sciences*, **230** (2013), 39–55.
- [42] M. Last and A. Kandel, *Data Mining for Process and Quality Control in the Semiconductor Industry*, 3, Data Mining for Design and Manufacturing, Springer, Boston, MA, 2001.
- [43] H. Li, C.-J. Li, X.-J. Wu and J. Sun, [Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine](#), *Applied Soft Computing*, **19** (2014), 57–67.
- [44] S. Li, H. Wu, D. Wan and J. Zhu, [An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine](#), *Knowledge-Based Systems*, **24** (2011), 40–48.
- [45] Y. Li and X. Zhang, [Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection](#), *Chemometrics and Intelligent Laboratory Systems*, **136** (2014), 47–57.
- [46] H. Liu and H. Motoda, Feature extraction construction and selection: A data mining perspective, *Journal of the American Statistical Association*, **94** (1999), 014004.
- [47] J. Long, S. Zhang and C. Li, [Evolving deep echo state networks for intelligent fault diagnosis](#), *IEEE Transactions on Industrial Informatics*, **16** (2020), 4928–4937.
- [48] J. Long, Z. Sun, C. Li, Y. Hong, Y. Bai and S. Zhang, [A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D printers](#), *IEEE Transactions on Instrumentation and Measurement*, **69** (2020), 683–692.
- [49] U. Maulik, A. Mukhopadhyay and D. Chakraborty, [Gene-expression-based cancer subtypes prediction through feature selection and transductive SVM](#), *IEEE Transactions on Biomedical Engineering*, **60** (2013), 1111–1117.
- [50] M. Melhem, B. Ananou, M. Ouladsine, M. Combal and J. Pinaton, [Product quality prediction using alarm data : Application to the semiconductor manufacturing process](#), *2017 25th Mediterranean Conference on Control and Automation (MED)*, (2017), 1332–1338.
- [51] L. Mönch, J. W. Fowler and S. J. Mason, *Production Planning and Control for Semiconductor Wafer Fabrication Facilities: Modeling, Analysis, and Systems*, 52, Springer Science & Business Media, Springer, New York, 2012.
- [52] A. Mucherino and L. Liberti, [A VNS-based heuristic for feature selection in data mining](#), *Hybrid Metaheuristics*, Studies in Computational Intelligence, 434, Springer, Berlin, Heidelberg, 2013, 353–368.
- [53] J. Neumann, C. Schnorr and G. Steidl, [Combined SVM-based feature selection and classification](#), *Machine Learning*, **61** (2005), 129–150.
- [54] Y. Oh, K. Ransikarbum, M. Busogi, et al., [Adaptive SVM-based real-time quality assessment for primer-sealer dispensing process of sunroof assembly line](#), *Reliability Engineering & System Safety*, **184** (2019), 202–212.
- [55] A. V. Phan, M. L. Nguyen and L. T. Bui, [Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems](#), *Appl. Intell.*, **46** (2017), 455–469.
- [56] H. Purwins, et al., Regression methods for prediction of PECVD silicon nitride layer thickness, *2011 IEEE International Conference on Automation Science and Engineering*, (2011), 387–392.
- [57] J. R. Quinlan, [Induction on decision tree](#), *Machine Learning*, **1** (1986), 81–106.
- [58] M. S. Rahman, M. K. Rahman, M. Kaykobad and M. S. Rahman, [isGPT: An optimized model to identify sub-Golgi protein types using SVM and random forest based feature selection](#), *Artificial Intelligence in Medicine*, **84** (2017), 90–100.
- [59] I. Rish, An empirical study of the naive Bayes classifier, *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, **3** (2001), 41–46.

- [60] Y.-C. Su, M.-H. Hung, F.-T. Cheng and Y.-T. Chen, [A processing quality prognostics scheme for plasma sputtering in TFT-LCD manufacturing](#), *IEEE Transactions on Semiconductor Manufacturing*, **19** (2006), 183–194.
- [61] J. B. Tenenbaum, V. D. Silva and J. C. Langford, [A global geometric framework for nonlinear dimensionality reduction](#), *Science*, **290** (2000), 2319–2323.
- [62] A. Unler, A. Murat and R. B. Chinnam, [mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification](#), *Information Sciences*, **181** (2011), 4625–4641.
- [63] Y. Wang, Z. Xue, G. Shen, et al., [PRINTR: Prediction of RNA binding sites in proteins using SVM and profiles](#), *Amino Acids*, **35** (2008), 295–302.
- [64] R. Weber and J. Basak, [Simultaneous feature selection and classification using kernel-penalized support vector machines](#), *Information Sciences*, **181** (2011), 115–128.
- [65] C. Wei, J. Chen, Z. Song and C. Chen, [Soft sensors of nonlinear industrial processes based on self-learning kernel regression model](#), *Asian Control Conference, IEEE*, (2018), 1783–1788.
- [66] X. Wu, L. Chen, S. Pang and X. Ding, [A paratactic subjective-objective weighting methods and SVM risk assessment model applied in textile and apparel safety](#), *International Journal of Quality & Reliability Management*, **32** (2015), 472–485.
- [67] Y. Xiang and L. Jiang, [Water quality prediction using LS-SVM and particle swarm optimization](#), *2009 Second International Workshop on Knowledge Discovery and Data Mining*, (2009), 900–904.
- [68] S. D. Xin and C. B. Zhong, [Some practical application of sequential analysis to the fault prediction system of a main diesel engine](#), *Conference of the IEEE Industrial Electronics Society, IEEE*, **3** (2002), 2151–2156.
- [69] H. Zhang, Q. Li, Z. Sun and Y. Liu, [Combining data-driven and model-driven methods for robust facial landmark detection](#), *IEEE Transactions on Information Forensics and Security*, (2016), 2409–2422.
- [70] H. Zhang, Q.-Y. Chen, M.-L. Xiang, et al., [In silico prediction of mitochondrial toxicity by using GA-CG-SVM approach](#), *Toxicology in Vitro*, **23** (2009), 134–140.
- [71] C. Zheng and L. Jiao, [Automatic parameters selection for SVM based on GA](#), *Intelligent Control & Automation*, **2** (2004), 1869–1872.
- [72] X. Zhao, E. K. Wong, Y. Wang, et al., [A support vector machine \(SVM\) for predicting preferred treatment position in radiotherapy of patients with breast cancer](#), *Medical Physics*, **37** (2010), 5341–5350.
- [73] Y. Zhu, Y. Tan, Y. Hua, et al., [Feature selection and performance evaluation of support vector machine \(SVM\)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography](#), *Journal of Digital Imaging*, **23** (2010), 51–65.

Received December 2019; 1st revision September 2020; 2nd revision January 2021.

E-mail address: yhhjy001@126.com

E-mail address: qianxiaofei888@126.com

E-mail address: peijun@hfut.edu.cn

E-mail address: cctan@ustc.edu.cn

E-mail address: pardalos@ufl.edu

E-mail address: lxb@hfut.edu.cn