Applying the support vector machine with optimal parameter design into an automatic inspection system for classifying micro-defects on surfaces of light-emitting diode chips

Chung-Feng Jeffrey Kuo[1] · **Chun-Ping Tung[2]** · **Wei-Han Weng[1]**

Abstract This study discusses the optimal design of an automatic inspection system for processing light-emitting diode (LED) chips. Based on support vector machine (SVM) with optimal theory, the classifications of micro-defects in light area and electrode area on the chip surface, and develop a robust classification module will be analyzed. In order to design the SVM-based defect classification system effectively, the multiple quality characteristics parameter design. The Taguchi method is used to improve the classifier design, and meanwhile, PCA is used for analysis of multiple quality characteristics on influence of characteristics on multi-class intelligent classifier, to regularly select effective features, and reduce classification data. Aim to reduce the classification data and dimensions, and with features containing higher score of principal component as decision tree support vector machine classification module training basis, the optimal multi-class support vector machine model was established for subdivision of micro-defects of electrode area and light area. The comparison of traditional binary structure support vector machine and neural network classifier was conducted. The overall recognition rate of the inspection system herein was more than 96%, and the classification speed for 500 micro-defects was only 3 s. It is clear that we have effectively established an inspection process, which is highly effective even under disturbance. The process can realize the subdivision of micro-defects, and with quick classification, high accuracy, and high stability. It is applicable to precise LED detection and can be used for accurate inspection of LED of mass production effectively to replace visual inspection, economizing on labor cost.

Keywords Taguchi methods · Principal component analysis · Decision tree support vector machine · Multi-class support vector machine model

Introduction

The use of LED technology has been rapidly increasing to meet high-efficiency lighting and energy saving requirements. Micro-defects on the surface of LED chips directly affect the luminous efficiency. To improve the process yield, increase the quality, and reduce the manufacturing cost, micro-defects need to be well-characterized and identified for modifications.

In the current industrial process to inspect micro-defects on surfaces of LED chips, image processing with intelligent classifier is often adopted to determine whether the chip is defective.

Chang et al. (2008) used learning vector quantization neural network for LED wafer defect inspection, the dies could be classified as either acceptable or defective. Lin (2009) employed the wavelet-based neural network (WNN) and wavelet-based multivariate statistical approaches are proposed individually to detect the surface blemishes. However, the aforesaid references only identify if there is defect or not, they cannot solve the problem of multiple defects in the actual process of LED.

✉ Chung-Feng Jeffrey Kuo
jeffreykuo@mail.ntust.edu.tw

[1] Department of Materials Science and Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan, ROC

[2] Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 10607, Taiwan, ROC

Pan and Chen (2012) presented the features of defects and the generalized regression neural networks for the defect inspection of LED chip. Kuo et al. (2014) used two-step back-propagation neural network for recognizing the appearance and internal structure defects of LED chip. Karimi and Asemani (2014) investigated different surface defect detection methods based on statistical pattern recognition, feature vector extraction and image classification. Zhong et al. (2015) presented that the image processing method can locate the LED chips and exclude the abnormal chips, but the analysis of improving the defective process of LED is insufficient. How to achieve the optimal design of an automatic inspection system for industrial implementation still needs to be considered.

According to literatures, no general algorithm has yet been proposed accounting for all different defect types at the same time. Some methods have fast speed but low accuracy while other methods are associated with high accuracy but restricted by complex computations to a lower speed. The intelligent classifier was mostly used to discuss and develop the application of LED defect detection, the classification was accurate to some extent, but the required training time was relatively long. When it is used industrial application, facing the continuously updated process and chip architecture, the neural network as classifier shall be improved.

In addition, the defective sample in the industrial inspection of LED is the minority in the massive database, resulting in data unbalance, which is a big challenge to the application of machine learning. Occasionally, only few records that indicate defective units are available and they are classified as a minority group in a large database. This situation causes an imbalanced data set problem and lead to a problem when applying machine-learning techniques for obtaining effective solution. Tan et al. (2015) proposed evolutionary artificial neural networks to classify an imbalanced data set. Although the combination of two intelligent algorithms can solve the data unbalance, the complicated computations needed more time to calculate. It is inapplicable to industrially mass real-time inspection and classification.

Following the requirement of real-time processing and existence of different patterns in LED chips industry, high speed and high accuracy are essential challenges to be afforded at the same time.

According to the aforesaid literatures, the appropriate classifier must be selected according to the kind, number and distribution of the applied feature values.

Support vector machine (Vapnik 1998) has a rapid calculation speed and small errors in independent test sets according to decision rules from limited training samples. Azadeh and Nasser (2011) indicated that the standard SVM algorithm is originally designed for two-class classification. He proposed an algorithm based on decision tree (DT) for multi-class SVM. The tree-structure constructs according to distances

among center point classes and range of spread data in those classes, but it might cause to allocate more weights to those features with better distinctions among other classes. Song et al. (2014) proposed SVM classifiers utilizing binary DT to solve multiclass problems. The decision tree is constructed according to the degree of difficulty to classify classes which is measured by Euclidean distance and standard deviation. Jair et al. (2015) presented a data selection method based on decision tree for SVM classification on large data sets. A data filter based on a DT that scans the entire data and obtains a small subset of data points. Chang et al. (2010) uses a decision tree to decompose a given data space and train SVMs on the decomposed regions. The tree decomposition method can derive a generalization error bound for the classifier. Mulay et al. (2010) proposed the hierarchical and tree structured multiclass SVM algorithms for implementation of intrusion detection system. The results for the proposed system are not available but it seems that multi-class pattern recognition problems can be solved using the tree-structured binary SVMs and the resulting intrusion detection system could be faster than other methods. Modi et al. (2011) presented a framework to integrate various algorithms to infer the human state using various cues. It is concluded that using several algorithms to infer the human state was significantly better than using one algorithm only owing to the diversity in the technique of inference. Especially, combination of high ranked experts yields a better result.

According to the above mentioned literatures, several algorithms, including one-against-one (OAO), one-against-all (OAO), and directed acyclic graph support vector machines (DAGSVM) were developed for solving a multi-class problem by SVM. And the combination of DT and SVM is a feasible method to solve the multi-class classification, especially, DTSVM and tree structured SVM will become favorable when the number of classes in recognition increases.

In the multiclass SVM, the weights of the classifications need to be considered, in order to obtain higher accuracy, easier separable class should be separated in the upper node. Due to the major drawback of SVM occurs in its training phase, which is computationally expensive and highly dependent on the size of input data set (Jair et al. 2015), the excessive input data will be avoided to influence the classification result error or increase the computation complexity. Therefore, in this study, DTSVM is developed for the purpose of classification, and compared with NN in effect of classification. In addition, if the classification data can be analyzed and evaluated in advance, and the configuration of the characteristics such as the weights of features of defects in classification training can be solved. The reliable analysis data will be used to build the SVM multi-class classification architecture without huge but not very independent input data, so as to increase the calculation rate and robustness.

To reduce the classification input data and establish the optimal multi-class classification module with only high weight of features of defect, generally, the experimental analysis is designed with try and error method and empirical approach, but no benefit is available.

Xu et al. (2015) using the focused on the optimization of perpendicular magnetizing parameters, using the Taguchi method. The average magnetic flux density was increased. However, Taguchi method only can be used in single quality characteristic problem, and it cannot be used in multiple-quality-characteristic optimal problem, and it needs to combine with multi-criteria optimization method. Fang et al. (2014) proposed a method to obtain the required optical characteristics at maximal performance for LED. After integrated optical design, the Taguchi method in combination with principal component analysis (PCA) led to an improvement of the systematic performance in terms of brightness, the thickness of the backlight module and uniformity.

Yingjie et al. (2015) proposed a robust state-based structured SVM tracking algorithm combined with incremental PCA. The incremental PCA was used to update the virtual feature vector corresponding to the virtual state and the principal subspace of the object's feature vectors. Nandi et al. (2009) aimed to search an optimal process environment, capable of producing desired bead geometry parameters of the weldment. Taguchi's robust optimization technique has been applied to determine the optimal setting, which can maximize the composite principal component. PCA has been adapted to covert multiple objectives of the optimization problem into a single objective function. A confirmatory test showed satisfactory result of PCA-based hybrid Taguchi method for parametric optimization.

Therefore, ranking the classifier effectiveness as quality characteristic and using process optimization theory for optimization are a feasible method.

According to the literatures, through the combination of Taguchi method and PCA, the experimental quality purpose can be effectively improved and it also can be used with SVM. In industry, to achieving high speed detection, most of company use try and errors to build the database for standard of determination, and avoid using intelligent algorithm. However, try and errors is time and human costing, and it's hard to create a standard operation procedure, and not favorable for different application of product. Therefore, it's not a good method in industry.

This study develops automatic optical inspection for LED chip process. Responding to practical application, it must be rapid, precise and robust, so as to replace manpower effectively. The key point of this study is to combine Taguchi with principal component to optimize the multi-class intelligent classifier, constructing a standard detection process is favorable for the development of different process applications.

Research method

Based on the LED chip surface micro-defects inspection process development, the application of intelligent algorithm in industrial automation LED defect inspection and subdivision will be discussed. The process flow is as shown in Fig. 1 After image acquisition, chip locating, and characteristic acquisition of light area and electrode area micro-defects, the planning on image characteristic value and classifier design planning is analyzed with Taguchi method combined with PCA. As for chip light area, two-dimensional linear SVM is used to distinguish surface color aberration and breakdown, and as for chip electrode area, DTSVM is used to distinguish normal, contamination, scrape and non-probe, and finally, through comparison with traditional DT, NN, LIB-SVM (Chang and Lin 2012), it is verified that the classifier established herein has a better speed, accuracy and reliability, and it is applicable to industrial automation inspection.

Figure 2 shows the normal LED chip structure, including the overall shape of chip, backlight image and front light image of light area, and electrode area. The front light detects electrode area micro-defects, highlighting electrode surface micro-defects directly. As the electrode is non-light-tight material, the image feature information of micro-defects cannot be obtained effectively by using backlight. The types of micro-defects on the electrode area and light area are shown in Fig. 3.

A. Electrode area
   (1) Contamination: The smudge area is larger than 1/10 of electrode area.
   (2) Scrape: The accumulative area of probe scrape is larger than 1/5 of electrode area. The metal electrode scrape is not allowed to expose to the substrate, and the scrape length does not exceed the diameter length of electrode).
   (3) Non-probe: P/N-electrode surface is smooth, and there is no traces of probe measurement. The chip of the electrical property and optical activity have not been detected.

B. Light area
   (1) Breakdown (including protective layer is spalled off): The back aluminum plating peels off. The epitaxy is defective, resulting in non-uniform light area thickness and color difference in light area.
   (2) Color aberration: The inconsistent light area color does not over 1/10 of area).

Taguchi method

This study adopts Taguchi method for the experimental planning of feature's PCA, to make the designed classifier quality

Fig. 1 Defect inspection flow
chart
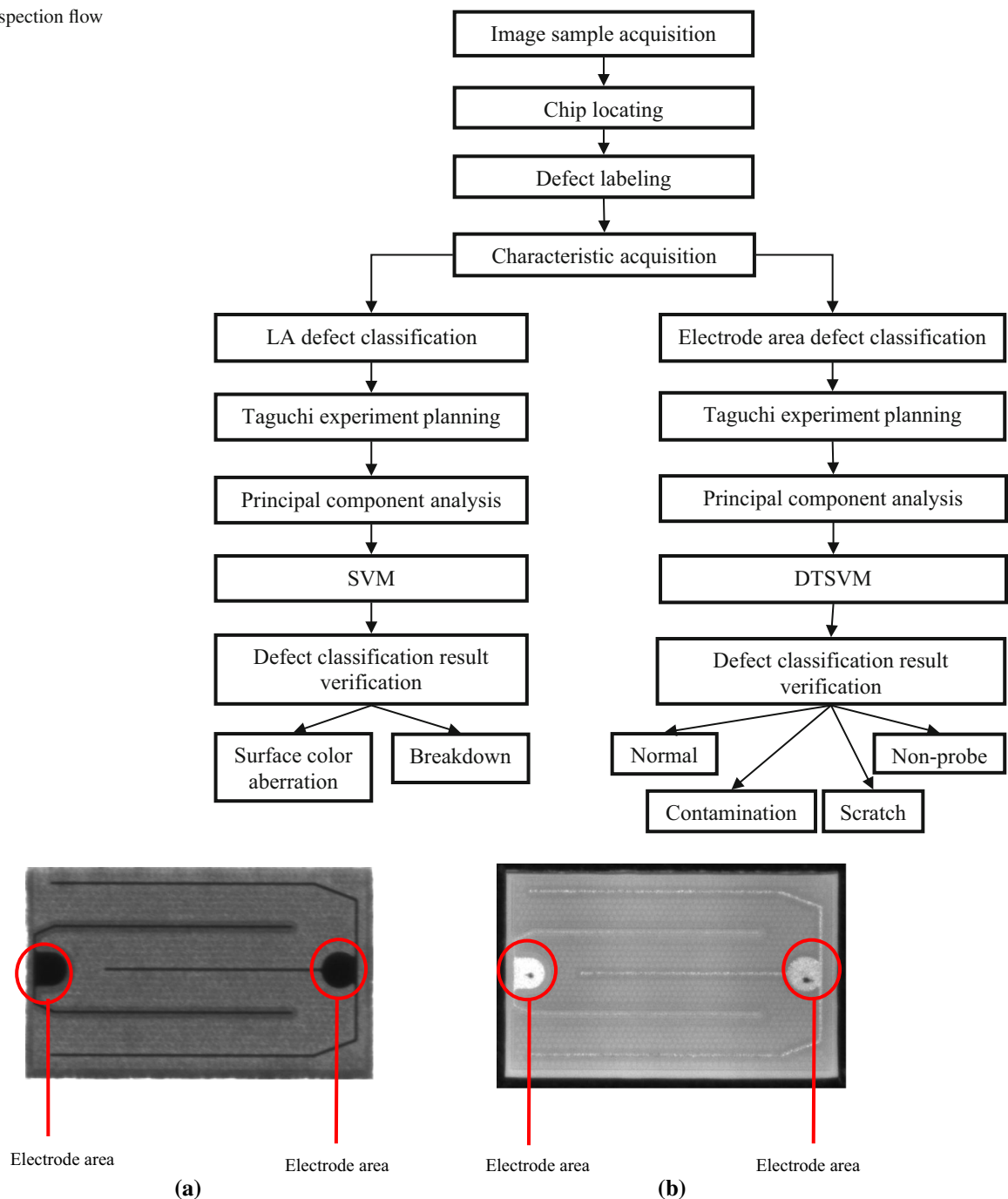
Fig. 2 Normal LED structure. **a** From backlight, **b** from front light

reach stable and robustness, to avoid the use of trial and error method, therefore, saving the analysis time for develop a large inspection system.

The implementation steps of Taguchi method are as shown in Fig. 4 (Ross 1996):

Lacking multiple quality measurement performance indicators, Taguchi method is unable to select the optimal parameter combination. In this study, PCA is added, to guarantee the reliability of Taguchi method multiple quality characteristic analysis results, and experiment is repeatedly conducted for confirmation and optimal parameter evaluation, so as to verify the reproducibility and stability of classifier.

Fig. 3 Micro-defects in LED chips. **a** Electrode area, **b** light area

Electrode area	Image of defect
Contamination	
Scrape	
Non-probe	

(a)

Light area	Image of defect
Breakdown	
Color Aberration	

(b)

## Principal component analysis (PCA)

PCA is used for characteristics preprocessing of micro-defects, to improve the accuracy, effectiveness and measurability of subsequent classification. As SVM training samples and training support vector number as well as data dimension affect the classification calculation time, this study finds out the most typical characteristics with PCA method. Through PCA method, the square error of data variable is minimized, so that after dimension reduction, the most typical characteristics can be held.

Step 1:

List the classifier indicators, good and poor, obtained in each experiment. The data represented by the form are not measured only through one-time measurement, and they are the measured values estimated of different characteristics in the experiment for times.

Step 2:

Make the original data be subject to normalization and all the data be ranged from 0 to 1, with the maximum value 1 and the minimum value 0, aiming to eliminate the analysis problem due to difference in unit. In this study, it is better to have a higher accuracy of classifier, and the amplified nor-
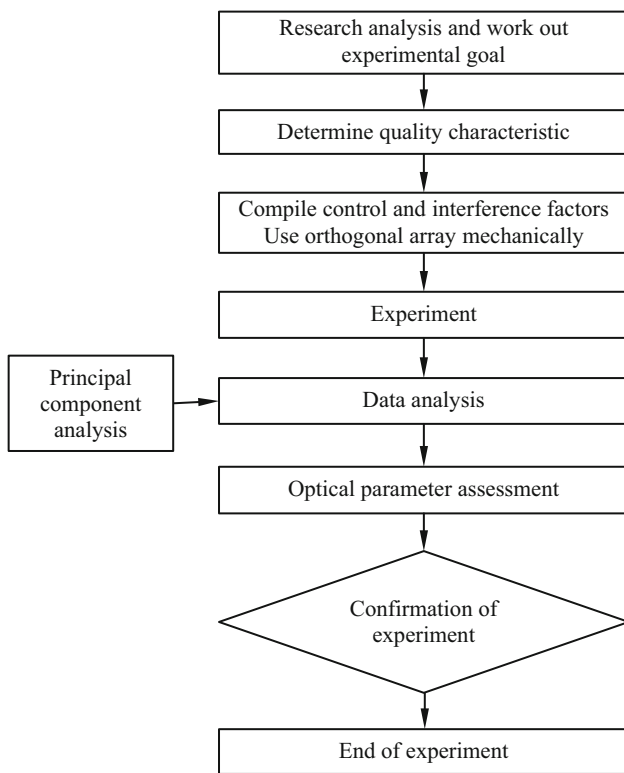
Fig. 4 Taguchi planning experiment process

malization is as shown in the following formula: $\max_{\forall k} x_i(k)$ and $\min_{\forall k} x_i(k)$ represent the maximum and minimum values of one list.

$$x_i^*(k) = \frac{x_i(k) - \min_{\forall k} x_i(k)}{\max_{\forall k} x_i(k) - \min_{\forall k} x_i(k)} \tag{1}$$

Step 3:

The correlation coefficient matrix obtained with the data after normalization R is shown as below:

$$\rho_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \tag{2}$$

ρ_{xy} = the correlation coefficient of x to y, $\bar{x}$ is the average value of x.

As $\rho_{xx} = 1$ and $\rho_{xy} = \rho_{yx}$, a square matrix will be formed, in which, the oblique line from upper left to lower right is 1.

Step 4:

The eigenvector β of eigenvalue λ and place rate C is obtained according to the correlation matrix:

$$\left| R - \lambda_k I_{p \times p} \right| = 0 \tag{3}$$

$$\sum_{k=1}^{P} \lambda_k = P \tag{4}$$

$$C_k = \frac{\lambda_k}{P} \tag{5}$$

where, R is correlation matrix, I is unit vector equal to R in dimension, and P is quality characteristic quantity.

$$(R - \lambda_k I_{p \times p})\beta_j = 0 \tag{6}$$

Step 5:

The principal component point is obtained, for eigenvector is a $P \times P$ (quality characteristic quantity) matrix, which is equal to the column number of normalization data (quality characteristic quantity), and the following formula shall be adopted:

$$Y_k = x_i^* \times \beta_j \tag{7}$$

Y_k is principal component score, x_i^* is data after normalization, β_j is eigenvector corresponding to x. According to the above equations, the principal component score table same as the original normalization dimension can be obtained, and then through the principal component score value, the component influencing degree can be understood.

Support vector machine (SVM)

In this study, by use of SVM, chip image characteristic is mapped to feature space, so as to find out the optimal hyperplane to divide the different sets, therefore, achieving chip micro-defects classification. SVM can solve the identification problem of small-sample, non-linear and high dimensional models.

Decision function:

$$D(x) = w^t x + b \tag{8}$$

w is the vector of dimension m; b is offset, to make the decision function realize translation, so as to find out the corresponding w and b, and the data are classified into two parts, to get the so-called predict model. The division hyperplane shall satisfy:

$$y_i(w^t x_i + b) \geq 1 \quad for \quad i = 1, \ldots, M. \tag{9}$$

As shown in Fig. 5, square dot and round dot indicate two kinds of samples, the line in the middle is classification line, and the two lines nearby are respectively the straight lines passing the samples nearest to the classification line parallel to classification line, and the distance between them is classification margin. The optical classification line has training
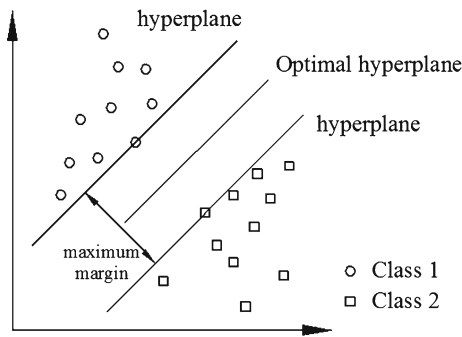
Fig. 5 the optimal hyperplane of SVM

error rate 0 and the largest classification interval. Through normalization treatment of Eq. 8, the linear separable sample set S meets Eq. 9. At this time, classification has the maximum interval $2/||w||$, and after $2/||w||$ is minimized, the optimal classification face is obtained, and the training sample points on the lines at two sides are support vectors.

Optimal hyperplane is established according to linear separable condition, and the quadratic programming optimal problem is shown as below:
Minimize:

$$\frac{1}{2}\|w\|^2 \tag{10}$$

Subject to:

$$y_i\left(w^t x_i + b\right) \geq 1 \tag{11}$$

Through Lagrange multiplier method (Vapnyarskii 2001), Eqs. (10), (11) are converted to (12), to minimize L to obtain the w, b, α_i (α_i is Lagrange multiplier).

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left[y_i\left(w^T x_i - b\right) - 1\right] \tag{12}$$

The obtained $L(w, b, \alpha)$ is used for partial differential of w and b. Assign w and b in L as minimum value $w = w*$ and $b = b*$, as well as α as maximum value $\alpha = \alpha*$, to get the following:

$$\frac{\partial L\left(w*, \alpha*\right)}{\partial w} = w* - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$w* = \sum_{i=1}^{n} \alpha_i^* y_i x_i \tag{13}$$

$$\frac{\partial L\left(w*, \alpha*\right)}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\sum_{i=1}^{n} \alpha_i^* y_i = 0 \tag{14}$$

Equations (13), (14) are substituted into (12) to obtain the minimum value, which can solve the quadratic programming problem: if the maximum value is obtained, it can be converted to the SVM function of one equal dual form:
Maximize:

$$Q(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i,j=0}^{M} \alpha_i \alpha_j y_i y_j x_i^t x_j \tag{15}$$

Subject to:

$$\sum_{i=1}^{M} y_i \alpha_i = 0, \quad \alpha_i \geq 0 \ \ for \ \ i = 1, \ldots, M. \tag{16}$$

As for the linear separable data, maximum α_i can be obtained from dual form, and then according to Eq. 14, w can be obtained from Eq. 13. According to Kuhn–Tucker theory (Kuhn and Tucker 1951), when training data are input, some points will meet the Kuhn–Tucker conditions, and through the optimal theory, the following is obtained:

$$y_i\left(w^t x_i - b\right) - 1 \geq 0, \quad \forall i \tag{17}$$

$$\alpha_i \geq 0, \quad \forall i \tag{18}$$

$$\alpha_i \left[y_i\left(w^t x_i - b\right) - 1\right] = 0, \quad \forall i, \alpha_i \geq 0 \tag{19}$$

According to Kuhn–Tucker conditions, it is known that, when $\alpha_i > 0$, the corresponding data is defined as support vector, and such data will have a decisive effect on the values of w and b. After support vector is obtained, it can be used to judge which set the newly added point belongs to.

$$f(x) = w^{*T} x - b^* = \sum_{I} \alpha_I y_i x_i^T x - b^* \tag{20}$$

$$b^* = \left(\sum_{j} \alpha_j y_j x_j^T x_i - y_i\right) \tag{21}$$

where i is support vector.

SVM has a better classification ability when solving the identification problem with few samples, and it is solved after transformation to quadratic programming problem, so it is the globally optimal solution.

Nonlinear kernel function is used for cut-out of the optimal space after the multidimensional characteristic data is converted to hyperplane. Although it can realize multiclassification, it has a slow calculation speed, and moreover, different parameters affect the classification results. Therefore, this study takes linear SVM as classifier and combines with DT, to reach the multi-classification function. To avoid parameter adjustment inconformity, Taguchi method is used to plan linear classifier design experiment, and PCA is used to

reduce the characteristic data, and the experiment is repeatedly confirmed, so as to improve the linear SVM accuracy, reliability and speed, and moreover, there is a higher reproducibility.
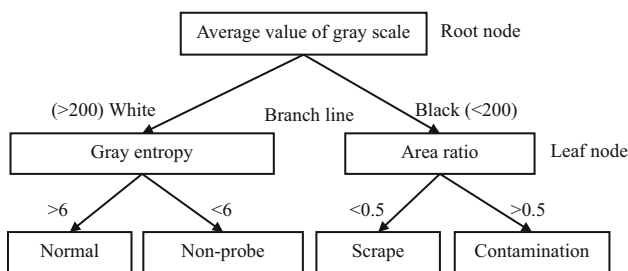
Decision tree (DT)

As it is rapid to learn DT induction method and classification, this study combines DT and SVM, aiming to design various micro-defects identification architectures, to replace the nonlinear kernel function's SVM which is weak point in multi-classification calculation time. Figure 6 shows the DT process architecture.

Each internal node indicates one test attribute, each branch indicates one possible test output result, and each leaf node indicates the class label. DT algorithm is defined as below:

1. Data setting: The original data are classified into two groups, namely training data and testing data.
2. DT generation: Training data are used to establish DT, and in each internal node, indicators are selected based on different attributes, from the existing attributes, the one with the best classification ability is picked out to be taken as the internal nodes of tree (root node and intermediate node), and the values of internal noses generate corresponding branches, and they are called as splitting nodes. As for each new branch, the training data shall be rearranged, and the following internal node generation will be conducted, which shall be repeated, until the termination conditions are satisfied.
3. Pruning: DT pruning shall be conducted by use of testing data.

   Repeat the steps 1 to 3, until all newly generated nodes become leaf nodes; and moreover, the data in such group are all classified into the corresponding class, and there are no data not yet processed, and in the data of such group, no new attributes can be found out for node splitting.

   In this study, C4.5 algorithm is taken as DT algorithm (Quinlan 1993).

Fig. 6 Decision tree process architecture

$$GainRatio\,(A) = \frac{Gain\,(A)}{SplitInfo\,(A)} \qquad (22)$$

$$SplitInfo(A) = -\sum_{j=1}^{} \frac{|S_j|}{|S|} \times \log_2\left(\frac{|S_j|}{|S|}\right) \qquad (23)$$

$Gain\,(A)$: Information gain value of A attribute for splitting of S object set.
$SplitInfo(A)$: A attribute information value.
$|S_j|$: Total number of objects belonging to attribute values in S set.
$|S|$: Total number of S.

The gain of one node under different split methods is defined as: the average value of sum obtained according to the formula: entropy of such node before separation deducts entropy of each sub-node after separation. It is expected that, it is better that the total average value of sub-node entropy after separation is smaller, and therefore, a bigger gain is better. The gain formula is shown as below:

$$Gain = Entropy_{Before} - Entropy_{After} \qquad (24)$$

As for excessive training problem in the example of DT learning, the model processing is not the general characteristic of training data, and instead, it is the local characteristic of training data, so it is inaccurate in classification of new samples. Therefore, this study conducts data preprocessing with PCA and uses effective characteristics to train, to effectively reduce the excessive fitness problem.

Results and discussions

Based on the micro-defects characteristic data, this paper firstly uses Taguchi method to conduct experimental planning of feature and classifier design, to improve the time consumption under trial and error method, and increase the reliability of classifier. Then, lists the quality characteristic indicators of eigenvalue in classifier, and according to the quality characteristics obtained in experiment. After that, calculates the principal component score of defect feature with PCA, and selects the feature with a higher principal component point as classifier input, to realize the data reduction and effectively save the classifier calculation, as well as improve the classifier accuracy and reliability. Finally, defect characteristic analysis result is used to design classifier including OAOSVM and DTSVM, and after comparison of the designed two structures of classifiers, a more stable algorithm will be adopted, and then through comparison with neural network, DT and multi-dimensional SVM. The advantages and disadvantages of different classifiers are discussed, to verify the accuracy and calculation time of DTSVM defect classification and

prove that the micro-defects inspection system herein is more robust.

Micro-defects features analysis

Taguchi methods and principal component analysis for light area defects

By use of orthogonal arrays of Taguchi method, the highest benefit is obtained with the fewest experiments. According to light area inspection process, based on back light image characteristic and micro-defects characteristic, the obtained classification characteristics include area ratio, entropy, mean and deviation, and such experiment is aimed to reduce the two characteristics for two-dimensional SVM design, therefore, L_4 orthogonal arrays are used for experimental planning. Table 1 shows the classification quality characteristic indicators obtained after Taguchi experiment, where Recall represents classification accuracy of total defects, P_B represents classification accuracy of light area breakdown, P_C represents classification accuracy of light area surface color aberration.

Light area micro-defects classification aims at two classes, chip surface color aberration and chip breakdown. Through Taguchi orthogonal arrays experiment, it can be clearly discovered that the characteristics enabling classification to have the best effect are mean and area ratio. Through the calculation of principal component score, the degree of influence of micro-defects on SVM, as shown in Table 2. Through repeated confirmation, it is better to have mean and area ratio as the characteristics of SVM.

Taguchi methods and principal component analysis of electrode area defects

Electrode Area classification has 4 types, including electrode normal, scrape, contamination, and non-probe. Image

inspection process based on front light source image feature and micro-defect characteristics, respectively. There are 12 micro-defect characteristics: electrode gray scale mean (F_G_M) and standard deviation (F_G_D), defect gray scale mean (D_G_M) and standard deviation (D_G_D), electrode defect area ratio (A_R), defect compactness (COMP), defect length-width ratio (R_R), defect gray scale entropy (D_Entropy), defect contrast (D_Con), defect correlation (D_Cor), defect homogeneity (D_Homo), and defect energy (D_Energy). According to DTSVM used herein, the micro-defects, which are greatly different, are firstly separated with DT, and then subdivided with two-dimensional linear SVM, as shown in Fig. 6. DT is upper layer, so Taguchi and principal component experiment planning analysis is firstly conducted to DT, and Table 3 shows the Taguchi experiment planning.

Larger-the-better method is adopted for normalization of **Taguchi method** experiment results. Therefore, a better effect will be achieved when there is better micro-defects classification. Then, through calculation according to Eq. (2), correlation matrix can be obtained, and subject to such matrix, its eigenvector is solved. Finally, through calculation of principal component score according to Eq. (7), the degree of influence of characteristics on DT can be obtained through principal component score, as shown in Table 4.

It is shown from the table that the optimal combination of classification characteristic used by DT is F_G_M, D_G_D, A_R, D_Cor, D_Homo and D_Energy. Therefore, it's taken as the classification characteristic of DT.

In the classification at the second layer, two-dimensional linear SVM is adopted, and classification is conducted respectively to contamination, normal, scrape, and non-probe, as shown in Fig. 6. Therefore, two SVMs are respectively subject to Taguchi experiment planning and PCA. In the classification part of contamination and normal, Table 5 shows the Taguchi experiment planning, the reason that orthogonal arrays of L_9 is SVM is subjected to the input of the two with the largest influence, and the obtained feature basis of orthogonal arrays is the optimal group Exp. No. 1 in DT with Taguchi experiment results, and then the characteristic of Exp. No. 1 is substituted into orthogonal arrays of L_9, to be subjected to further multi-quality PCA, as shown in Table 5.

After normalization of the result obtained in Taguchi experiment, according to Eqs. (2) and (7), principal component score is calculated, to obtain the degree of influence of the feature of corresponding factor level of each principal component score given in Table 6 on normal and contamination classification. The results indicate that, F_G_D and A_R have the largest influence on normal and contamination classification.

The PCA analysis steps of electrode scrape and non-probe characteristics are same those of electrode normal and con-

Table 1 SVM L_4 Taguchi orthogonal arrays

Exp. No.	1	2	Recall	P_B	P_C
1	Deviation	Entropy	0.845	0.809	0.917
2	Deviation	Area ratio	0.901	0.851	0.964
3	Mean	Entropy	0.859	0.813	0.957
4	Mean	Area ratio	0.93	0.907	0.964

Table 2 Influencing degree of color aberration and breakdown characteristics

Deviation	1.65172
Mean	3.036791
Entropy	1.458237
Area ratio	3.230274

Table 3 Decision tree L$_8$ Taguchi orthogonal arrays

Exp. No.	1	2	3	4	5	6
1	F_G_D	D_G_D	A_R	D_Cor	D_Con	D_Energy
2	F_G_D	D_G_D	A_R	R_R	D_Homo	D_Entropy
3	F_G_D	D_G_M	COMP	D_Cor	D_Con	D_Entropy
4	F_G_D	D_G_M	COMP	R_R	D_Homo	D_Energy
5	F_G_M	D_G_D	COMP	D_Cor	D_Homo	D_Energy
6	F_G_M	D_G_D	COMP	R_R	D_Con	D_Entropy
7	F_G_M	D_G_M	A_R	D_Cor	D_Homo	D_Entropy
8	F_G_M	D_G_M	A_R	R_R	D_Con	D_Energy

Table 4 Influencing degree of characteristics in DT

F_G_D	3.447845	D_Cor	4.16999363
F_G_M	4.043113	R_R	3.32096426
D_G_D	4.186103	D_Con	3.33381992
D_G_M	3.304855	D_Homo	4.15713796
A_R	4.215712	D_Energy	4.31313749
COMP	3.275246	D_Entropy	3.1778204

Table 5 L$_9$ Taguchi orthogonal arrays of SVM normal and contamination

Exp. No.	1	2
1	F_G_D	A_R
2	F_G_D	D_Cor
3	F_G_D	D_Con
4	D_G_D	A_R
5	D_G_D	D_Cor
6	D_G_D	D_Con
7	D_Energy	A_R
8	D_Energy	D_Cor
9	D_Energy	D_Con

Table 6 Degree of influence of SVM characteristic on normal and contamination

1		2	
F_G_D	5.0854	A_R	4.56403
D_G_D	1.6213572	D_Cor	1.6747082
D_Energy	1.2679661	D_Con	1.7359837

Table 7 Degree of influence of SVM characteristic on scrape and non-probe

F_G_D	−1.86803	A_R	2.019466
D_G_D	1.41459	D_Cor	2.30687
D_Energy	0.215066	D_Con	−1.835978

gray scale mean, and gray scale deviation have the largest influence.

Micro-defects classification

This study adopts the combination of two dimensional linear SVM and DT, to reach the multi-classifier function, aiming to replace SVM non-linear multi-classifier and neural network, and through Taguchi method and PCA, reliable multi-class classifier architecture design can be provided, to realize high speed calculation, high accuracy, high stability, and high reproducibility.

Result of micro-defects classification in light area

According to the inspection process planned herein, in the light area defect inspection, classification is required for chip surface color aberration and breakdown, so only single class of SVM needs classification, and Mean and Area Ratio obtained according to Taguchi experiment on principal component are taken as classification characteristics, and compared with multi-dimensional DT (MDDT) and multi-dimensional SVM (MDSVM).

Figure 7 shows light area micro-defects (a) training and (b) prediction manifestations, detail of this training and prediction data is shown in Table 8. After inspection process planned, the SVM has more regular distribution in two dimensional space. The recognition rate of light area breakdown is 94.12%, the recognition rate of chip surface color aberration is 92.1%, and the total recognition rate is 93.26%.
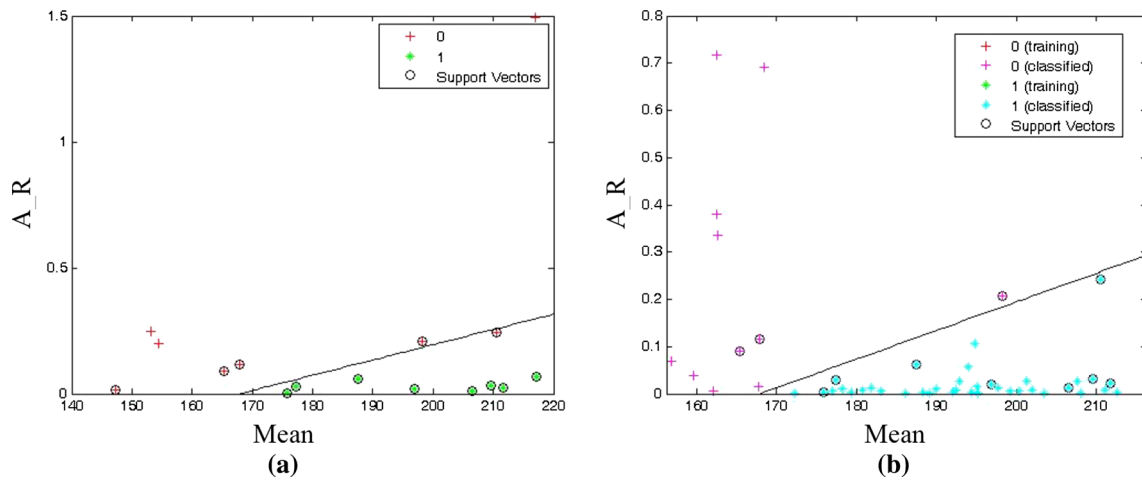
tamination. According to Table 7, A_R and D_COR have the largest influence on scrape and non-probe classification.

According to Taguchi experiment analysis of principal component characteristics, the effective characteristics include defect area ratio, defect correlation, gray scale deviation, gray scale mean, gray scale energy, and gray scale homogeneity, in which, defect area ratio, defect correlation,

Fig. 7 Two dimension linear SVM of light area. **a** Training, **b** prediction results

Table 8 Micro-defects classification in light area

Error type	Items			
	Training data	Prediction data	Fall	Accuracy (%)
Breakdown	8	51	3	94.12
Surface color aberration	8	38	3	92.1

Table 9 Comparisons of micro-defects classification in light area

	2D-SVM	DT	MSVM
Recall rate	0.9326	0.873	0.915
Breakdown precision rate	0.9412	0.919	0.925
Color aberration precision rate	0.921	0.824	0.903
Time cost (s)	1	2	435

According to Table 9, through Taguchi method and PCA, two features of defects are selected for classification, and there is an improvement in accuracy and each type of classification, which is better than DT, and the features after data preprocessing have a relatively regular distribution, and there is no need to use nonlinear kernel function for SVM calculation, to reduce classification time, which is faster than multi-dimensional characteristic space SVM (MSVM), therefore, reaching the effect to improve classification.

Result of micro-defects classification in electrode area

The defect classification of electrode area is normal, contamination, scrape, and non-probe. In this study, the obtained feature through PCA is taken as the training characteristic
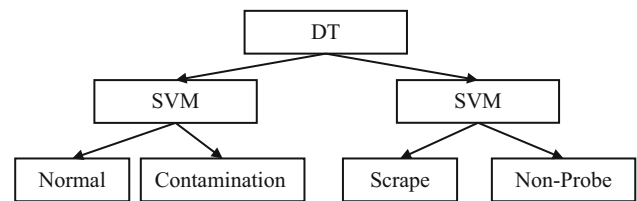

**Fig. 8** Electrode area DTSVM structure

of SVM, and then DT is combined with SVM for two-dimensional hierarchical classification, and it is compared with OAOSVM architecture, to discuss the classification efficiency of different structures.

Figure 8 shows the classification of tree-like architecture through combination of DT and SVM. The first layer is DT, and the second layer is two SVMs. By use of the characteristic that DT has high speed computing power to many characteristics, the obtained features by PCA include gray scale mean, gray scale deviation, defect area ratio, defect correlation, defect homogeneity, defect energy, which are input to DT, and by use of the characteristic that DT has a very quick instruction cycle to many characteristic points, rigorous defect characteristic information is used to distinguish the defects that are uneasy to be distinguished based on contamination and scrape, normal and non-probe, and then the two characteristics with higher point values are taken out through PCA, including defect area ratio (A_R), defect correlation (D_Cor), gray scale deviation (F_G_D) are taken as linear two-dimensional SVM for classification of normal and contamination, scrape and non-probe, aiming to obtain high-efficiency instruction cycle and precision.

In the SVM classification of normal and contamination, the used features are A_R and F_G_D, and the training and prediction results are as shown in Fig. 9; in the SVM classification of scrape and non-probe, the used features are D_Cor
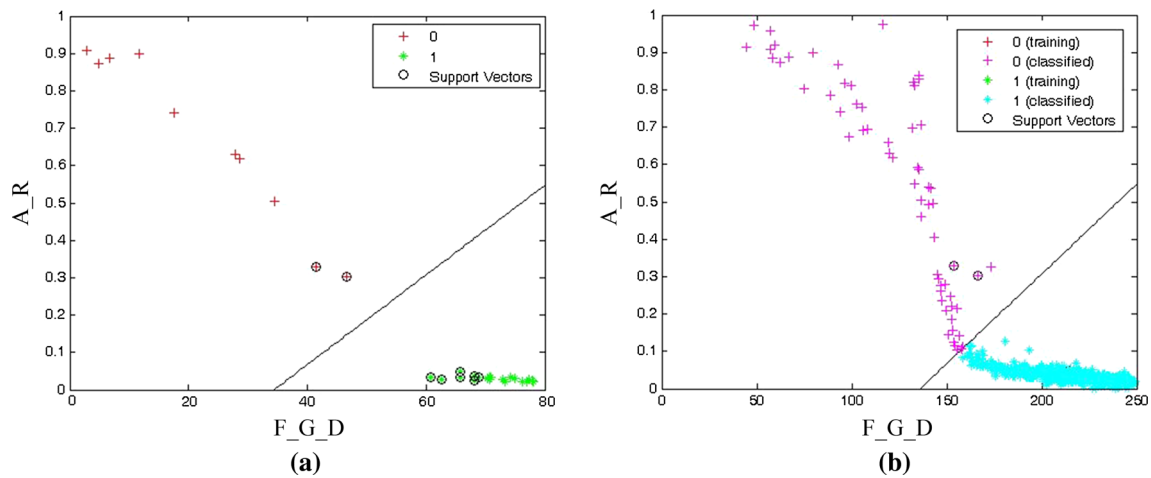
Fig. 9 DTSVM—normal and contamination. **a** Training, **b** prediction results
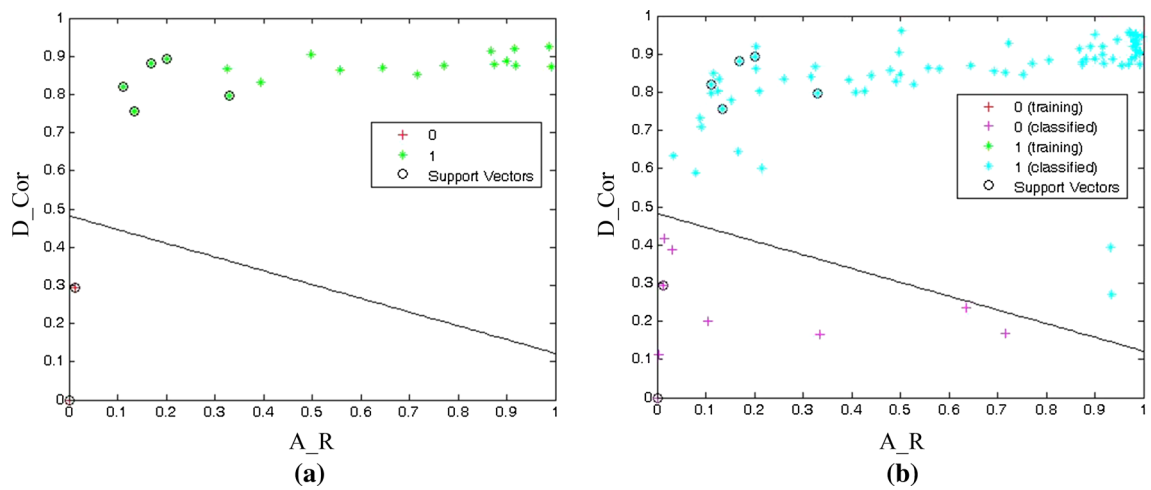
Fig. 10 DTSVM—scrape and non-probe. **a** Training, **b** prediction results

and A_R, and the training and prediction results are as shown in Fig. 10. It can be known that through the classification based on the features obtained from PCA, different types can be effectively separated, and there is a more regular distribution, and there is no need to use nonlinear kernel function as SVM, therefore, linear SVM classification can be used to reduce the classification calculation time, and the total calculation time is about 3 s, and meanwhile, this can avoid multilayered structure design and parameter setting inconformity.
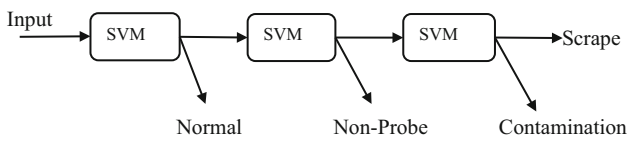
According to the classification results of DTSVM **in electrode area,** the recognition rate of electrode normal is 97.3%, contamination 96%, scrape 95.7%, and non-probe 95.1%, and the total recognition rate of such classifier is 96.83%. There are 1262 samples in total, including 53 misjudgments, the detail information is shown in Table 10.

Figure 11 shows one-against-one cascade structure SVM (OAOSVM), which can separate out the types with great dif-

Table 10 Classification performance of DTSVM in electrode area

Class	Items		
	Samples	Success	Accuracy (%)
Normal	934	909	97.3
Contamination	50	48	96
Scratch	93	89	95.7
Non-probe	185	176	95.1

ference layer by layer. In the first layer, A_R and F_G_D are used as characteristics for training of electrode area normal identification, and the training and prediction results are as shown in Fig. 12; in the second layer, A_R and D_Cor are used as classification characteristics, for non-probe identification of abnormal classification, and the training and prediction results are as shown in Fig. 13; and in the third layer, F_G_D and A_R are used as classification charac-

Fig. 11 OAOSVM structure

teristics, for scrape and contamination identification for the classes beyond normal and non-probe, and the training and prediction results are as shown in Fig. 14. Each layer is linear two-dimensional SVM, to avoid parameter design difference.

According to OAOSVM classification results, the recognition rate of electrode normal is 96.0%, non-probe electrode 97.3%, electrode contamination 98.8%, and electrode scrape 89.24%. There are 1262 samples in total, including 53 misjudgments, with the total recognition rate 95.8%, detail information is shown in Table 11.

Table 12 shows electrode area classifier efficiency comparison and it is proved that the DTSVM classifier developed herein features quick speed and high accuracy, and the difference between DTSVM and OAOSVA lies in that DTSVM has a relatively uniform accuracy in each class, which can reach more than 95.8%, more stable than the result of OAOSVM classification. For this reason, we selected DTSVM to our LED micro-defect detect system in electrode area. Through Taguchi method planning classifier design, after efficient repetition of experiment and experimental confirmation, the reproducibility of classifier can be improved. Through PCA, the influence of each feature on classification is evaluated, and the unnecessary characteristic data are reduced, to effectively improve the system accuracy in both DTSVM and OAOSVM. Compared with multidimensional DT, DTSVM classifier has the same recognition rate, but the accuracy is each class is relatively mean, that is DTSVM has more stable classification results, while compared with multidimensional
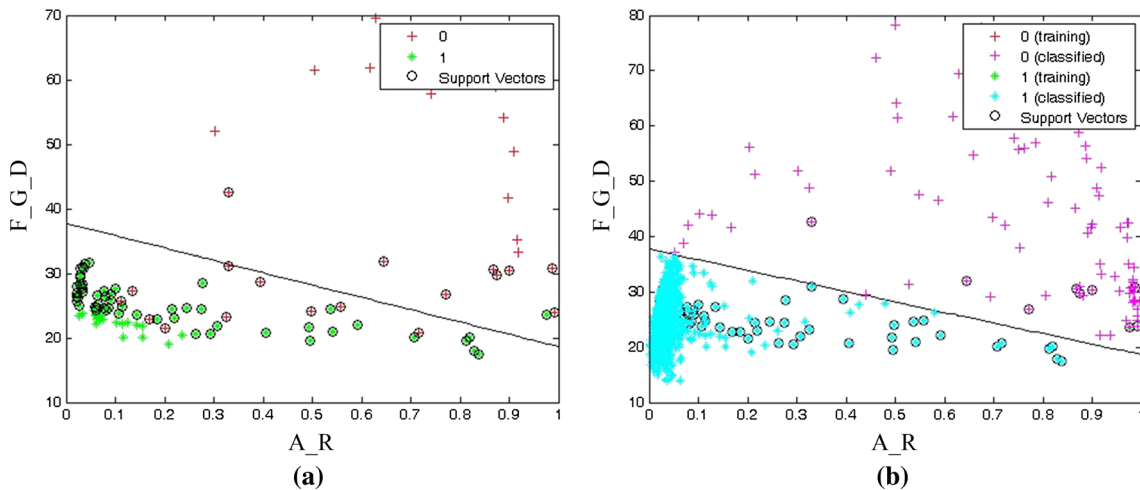
Fig. 12 OAOSVM—normal in first layer. **a** Training, **b** prediction results
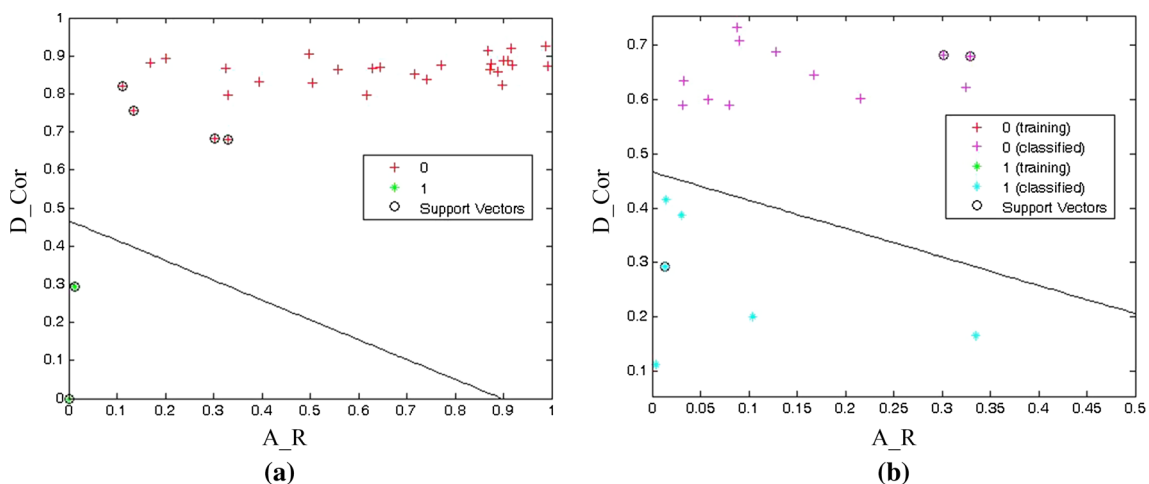
Fig. 13 OAOSVM—non-probe in second layer. **a** Training, **b** prediction results
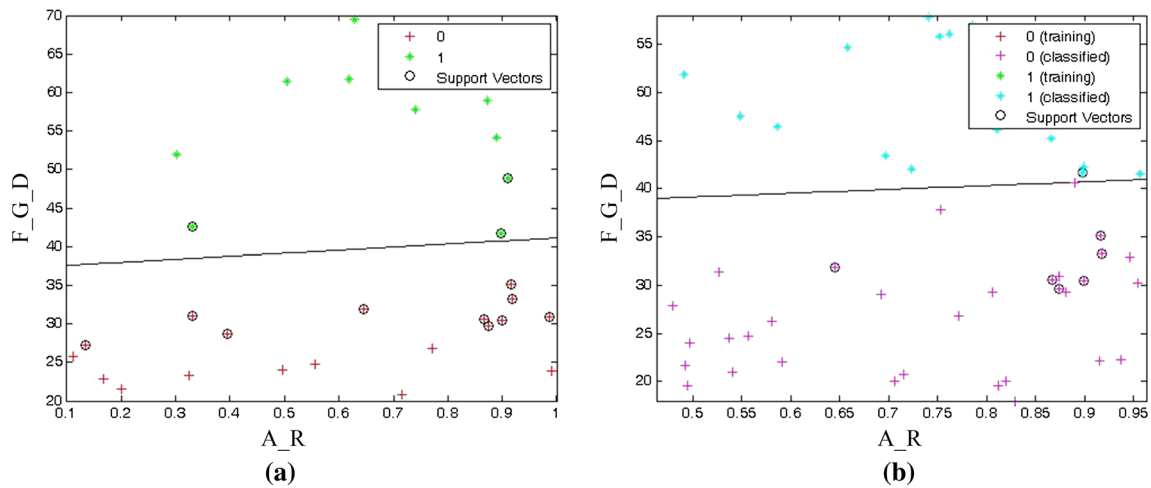
Fig. 14 OAOSVM scrape and contamination in third layer. **a** Training, **b** prediction results

Table 11 Classification performance of OAOSVM in electrode area

Class	Items		
	Samples	Error	Accuracy (%)
Normal	934	37	96
Non-probe	185	5	97.3
Contamination	50	1	98.8
Scratch	93	10	89.24

Table 12 comparison of electrode area classifier efficiency

	DTSVM	OAOSVM	DT	MSVM	NN
Recall	0.968	0.958	0.968	0.951	0.967
Normal	0.973	0.96	0.975	0.962	0.978
Non-probe	0.96	0.973	1	0.961	1
Contamination	0.957	0.988	0.833	0.838	0.809
Scratch	0.951	0.892	0.892	0.737	0.829
Time cost (s)	1	1	2	1087	956

SVM (MSVM) and neural network, DTSVM on the basis of combination of two-dimensional linear SVM and DT has an excellent calculation speed, and it is more suitable for LED defect inspection of lots of chip samples.

Conclusions

This system develops LED surface micro-defects identification inspection system. It is a multiple quality characteristics programming for industrial defect detection, and the standard of LED micro-defect classification system optimization is provided. The overall recognition rate of the inspection system based on SVM was more than 96%, and the classifica-

tion speed for 500 micro-defects was only 3 s. This automatic inspection can replace human's detection with quick classification, high accuracy, and high stability, it is applicable in LED manufacturing to reduce the cost of human.

With incisive analysis data and in the accuracy rate up to standard, it can rapidly identify defect and realize micro-defects classification Based on image inspection process, in which, chip can be located without the disturbance influence of image light source, to effectively classify defect images and completely obtain the characteristic information of micro-defects. A program for multiple quality characteristics parameter design is proposed by combining PCA with Taguchi Method, to design the SVM-based defect classification system effectively. The Taguchi method is used to improve the classifier design with trial-and-error method, and meanwhile, PCA is used for analysis of multiple quality characteristics on influence of characteristics on classifier, to regularly select effective features, and reduce classification data volume. In addition, in terms of the SVM, the selection of feature value has significant effect on the classification result. There are 12 image defect features extracted in this study. These good features selection result in better hyperplane margin from SVM training, and better classification result is obtained. The experimental results show:

(1) Under light area micro-defects classification, the recognition rate is 93.28%, and when data volume is reduced by use of Taguchi method and PCA, the accuracy and precision actually can be improved, and there has good performance in classification reproducibility.

(2) Under electrode area micro-defects classification, the recognition rate of the proposed DTSVM is 96.8%, and compared with OAOSVM, multidimensional DT, multidimensional SVM and NN using selected features, this study adopts quality analysis theory in effective char-

acteristic selection in classifier optimal design, and the classification precision at each class can reach more than 95%. Except for high precision, high stability and high reproducibility, a high instruction cycle is available, to reach the optimal classifier, so it is more applicable to industrial automation defect inspection system.

References

Azadeh, O., & Nasser, S. (2011). An efficient hybrid algorithm for multi-class support vector machines. In: *International conference on computer science and information technology* June 10–12 (pp. 24–28). Chengdu, China.

Chang, C. C., & Lin, C. J. (2012). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1–27:27. (2011).

Chang, C. Y., Chang, C. H., Li, C. H., & Jeng, M. D. (2008). Learning vector quantization neural networks for LED wafer defect inspection. *International Journal of Innovative Computing, Information and Control*, *4*(10), 2565–2579.

Chang, F., Guo, C. Y., Lin, X. R., & Lu, C. J. (2010). Tree decomposition for large-scale SVM problems. *Journal of Machine Learning Research*, *11*, 2855–2892.

Fang, Y. C., Tzeng, Y. F., & Wu, K. Y. (2014). A study of integrated optical design and optimization for LED backlight module with prism patterns. *Journal of Display Technology*, *10*(10), 812–818.

Jair, C., Farid, G. L., Asdrúbal, L. C., Lisbeth, R. M., & Sergio, J. R. (2015). Data selection based on decision tree for SVM classification on large data sets. *Applied Soft Computing*, *37*, 787–798.

Karimi, M. H., & Asemani, D. (2014). Surface defect detection in tiling industries using digital image processing methods: Analysis and evaluation. *ISA Transactions*, *53*, 834–844.

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley symposium* (pp. 481–492). Berkeley: University of California Press.

Kuo, C. F., Hsu, C. T., Liu, Z. X., & Wu, H. C. (2014). Automatic inspection system of LED chip using two-stages back-propagation neural network. *Journal of Intelligent Manufacturing*, *25*, 1235–1243.

Lin, H. D. (2009). Automated defect inspection of light-emitting diode chips using neural network and statistical approaches. *Expert Systems With Applications*, *36*(1), 219–226.

Modi, S., Lin, Y., Cheng, L., Yang, G., Liu, L., & Zhang, W. (2011). A socially inspired framework for human state inference using expert opinion integration. *IEEE ASME Transactions on Mechatronics*, *16*(5), 874–878.

Mulay, S. A., Devale, P. R., & Garje, G. V. (2010). Decision tree based support vector machine for intrusion detection. In *IEEE international conference on networking and information technology* (pp. 59–63).

Nandi, G., Datta, S., Bandyopadhyay, A., & Pal, P. K. (2009). Application of PCA-based hybrid Taguchi method for correlated multicriteria optimization of submerged arc weld: A case study. *International Journal of Advanced Manufacturing Technology*, *45*(3–4), 276–286.

Pan, Z. L., & Chen, L. (2012). Defect inspection of LED chips using generalized regression neural network. *Solid State Phenomena*, *181–182*, 212–215.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Ross, P. J. (1996). *Taguchi techniques for quality engineering*. New York: McGraw-Hill.

Song, X., Xiaojun, J., Songlin, S., & Hai, H. (2014). Binary decision tree based multiclass support vector machines. In *International symposium on communications and information technologies*. September 24–26 (pp. 85–89). Incheon, Songdo.

Tan, S. C., Junzo, W., Zuwairie, I., & Marzuki, K. (2015). Evolutionary fuzzy ARTMAP neural networks for classification of semiconductor defects. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(5), 933–950.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.

Vapnyarskii, I. B. (2001). *Lagrange multipliers. Encyclopedia of mathematics*. Hazewinkel: Springer.

Xu, Z. H., Wang, S. C., Zhang, Z. W., Chin, T. S., & Sung, C. K. (2015). Optimization of magnetizing parameters for multipole magnetic scales using the Taguchi method. *IEEE Transactions on Magnetics*, *51*(11), 1–4.

Yingjie, Y., De, X., Xingang, W., & Mingran, B. (2015). Online state-based structured SVM combined with incremental PCA for robust visual tracking. *IEEE Transactions on Cybernetics*, *45*(9), 1988–2000.

Zhong, F., He, S., Li, B., & Chen, M. S. (2015). Blob analyzation-based template matching algorithm for LED. *Internal Journal of Advanced Technology*,. doi:10.1007/s00170-015-7638-5.