

## Research Article

# An Ensemble Learning for Predicting Breakdown Field Strength of Polyimide Nanocomposite Films

Hai Guo,<sup>1,2</sup> Jinghua Yin,<sup>1</sup> Jingying Zhao,<sup>2,3</sup> Lei Yao,<sup>1,4</sup> Xu Xia,<sup>1</sup> and Hao Luo<sup>1</sup>

<sup>1</sup>School of Applied Science, Harbin University of Science and Technology, Harbin 150080, China

<sup>2</sup>College of Computer Science and Engineering, Dalian Nationalities University, 18 Liaohe West Road, Dalian Development Zone, Dalian 116600, China

<sup>3</sup>Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China

<sup>4</sup>Faculty of Engineering, Mudanjiang Normal College, Mudanjiang 157012, China

Correspondence should be addressed to Hai Guo; guohai@dlnu.edu.cn

Received 8 April 2015; Accepted 31 May 2015

Academic Editor: Mircea Chipara

Copyright © 2015 Hai Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Using the method of Stochastic Gradient Boosting, ten SMO-SVR are constructed into a strong prediction model (SGBS model) that is efficient in predicting the breakdown field strength. Adopting the method of in situ polymerization, thirty-two samples of nanocomposite films with different percentage compositions, components, and thicknesses are prepared. Then, the breakdown field strength is tested by using voltage test equipment. From the test results, the correlation coefficient (CC), the mean absolute error (MAE), the root mean squared error (RMSE), the relative absolute error (RAE), and the root relative squared error (RRSE) are 0.9664, 14.2598, 19.684, 22.26%, and 25.01% with SGBS model. The result indicates that the predicted values fit well with the measured ones. Comparisons between models such as linear regression, BP, GRNN, SVR, and SMO-SVR have also been made under the same conditions. They show that CC of the SGBS model is higher than those of other models. Nevertheless, the MAE, RMSE, RAE, and RRSE of the SGBS model are lower than those of other models. This demonstrates that the SGBS model is better than other models in predicting the breakdown field strength of polyimide nanocomposite films.

## 1. Introduction

As one of material products that have been developed for a long time, polyimide film (PI film) has been mainly applied in high and new technology industries such as aerospace, machinery, electrical and electronics engineering, optical communication, LCD, automobile, precision instrument, gas separation, and microelectronics [1]. With the development of nanotechnology, nanoparticles of different sizes, percentage compositions, and components have been mixed with PI by more and more researchers to produce high-quality polyimide nanocomposite films [2, 3].

Breakdown field strength is an important characteristic parameter to characterize polyimide nanocomposite films. It can be calculated by (breakdown voltage)/(film thickness). Many researchers have already made studies and analysis on the breakdown field strength of nanocomposite films [4–6].

There are many factors that can impact the breakdown field strength, including the type of nanoparticle, dielectric constant, electric conductivity, coefficient of thermal conductivity, composition, nanoparticles' size and specific area, and composite film thickness. In order to establish the knowledge base for the material property regarding breakdown field strength of polyimide nanocomposite films, large quantities of experiments have to be prepared and measure related characteristics. Nevertheless, it is also well known that getting properties data is very costly in terms of time and materials. It is for this reason that developing a fast and efficiency method to predict the breakdown field strength of polyimide nanocomposite films is very much in demand.

Intelligent computing and neuronal network have been widely applied in performance prediction, identification, and optimization of nanocomposite films. Yang et al. used a generalized regression neural network (GRNN) to predict

the friction coefficient of  $\text{Cr}_{1-x}\text{Al}_x\text{C}$  film [7]. Cho et al. optimized the characteristics of ITO/Al/ITO multilayer films by advantages of neural network and genetic algorithm [8]. Bahramian made use of an artificial neural network to predict the growth rate of  $\text{TiO}_2$  nanostructured film [9]. Ensemble learning has now become a new hotspot of intelligent computing [10–13]. By using ensemble, several weak classifiers can be constructed into a strong classifier. Some common methods of ensemble learning include boosting, bagging, and stacking. Boosting [14] keeps upgrading weights during data extraction and revises the weights of data set that have been classified wrong. In the end, several weak classifiers are obtained and can be constructed into a strong classifier. Bagging [15] is used the reiterative training to get several classifiers based on a training set. Stacking [16] contains two layers. In the first layer, different algorithms are employed to generate several weak classifiers. At the same time, a new data set with the same size of the original data set is also generated. Then, the new data set together with a new algorithm can be used to construct the classifier of the second layer.

The purpose of this paper is to develop a Stochastic Gradient Boosting + SMO-SVR model (SGBS model) to predict the breakdown field strength of polyimide nanocomposite films. In what follows, film preparation and prediction model are introduced first. The experimental details for sputtering systems and materials are described next. Then, the experimental results are described and the establishment of SGBS model with 10-fold cross validation results is carried out. Comparison experiments between linear regression, BP neural network (BP), general regression neural network (GRNN), SVR (support vector regression), and SMO-SVR models are conducted.

## 2. Nanocomposite Film Preparation and Prediction Model

*2.1. Preparation of Nanocomposite Films.* The method of in situ polymerization is used to prepare the polyimide matrix inorganic nanocomposite film. Experimental materials include 4,4'-diaminodiphenyl ether (ODA), pyromellitic dianhydride (PMDA), dimethylacetamide (DMAc),  $\text{Al}_2\text{O}_3$ , rutile  $\text{TiO}_2$ ,  $\text{BaTiO}_3$ ,  $\text{SiO}_2$ , and ethanol. The detailed purchase information of the above experimental materials is shown in Table 1. Firstly, put PMDA into the solution of ODA in DMAc to produce an amount of polyamide acid of certain viscosity. Secondly, add in different nanoparticles. Finally, let the mixture go through paving membrane heat treatment and imidization transform.

To verify our SGBS model, we have designed and prepared in this paper nanoparticle samples with different types, sizes, ratios, and thicknesses. The SEM image of surface appearance of pure PI is given in Figure 1(a). The surface of film is smooth and its tightness is well. Figure 1(b) shows the SEM image of surface appearance of PI/ $\text{BaTiO}_3$  composite film doped with content of 60 wt%. There are large amounts of  $\text{BaTiO}_3$  nanoparticles exposing on the surface of film with the size from 100 nm to 300 nm. Comparing with the other films, the surface appearance of film has been changed to be

TABLE 1: The detailed purchase information of experimental materials.

Material	Companies
4,4'-Diaminodiphenyl ether	Sinopharm Chemical Reagent Co., Ltd.
Pyromellitic dianhydride	Sinopharm Chemical Reagent Co., Ltd.
Dimethylacetamide	Tianjin Fuyu Fine Chemical Co., Ltd.
$\text{Al}_2\text{O}_3$	Beijing DK Nano Technology Co., Ltd.
Rutile $\text{TiO}_2$	Beijing DK Nano Technology Co., Ltd.
$\text{BaTiO}_3$	Beijing DK Nano Technology Co., Ltd.
$\text{SiO}_2$	Beijing DK Nano Technology Co., Ltd.
Ethanol	Tianjin Fuyu Fine Chemical Co., Ltd.

rough due to the dopant of  $\text{BaTiO}_3$  nanoparticles. Its density has also been decreased. Figure 1(c) shows the SEM image of surface appearance of PI/ $\text{TiO}_2$  composite film doped with content of 5 wt%.  $\text{TiO}_2$  can absorb ultraviolet ray to change the color of film into brown. The surface appearance of film is smooth compared with pure PI film. Figure 1(d) shows the SEM image of surface appearance of PI/ $\text{Al}_2\text{O}_3$  composite film doped with content of 20 wt%. There are some nanoparticles appearing in the surface of film, uniformly, where the diameter of nanoparticles is observed to be 70 nm and some particles are not cluster. In conclusion, the prepared hybrid PI particles are evenly distributed.

*2.2. Standard SMO-SVR Model.* SVM (support vector machine) [17] is one of the most typical machine learning methods in field of statistical learning theory. It has a very good learning ability in a small sampling space. Compared with BP, GRNN, and other neural networks, it has advantages such as structural risk minimization, global optimization solution, high dimensional space, and being linearly separable. It has been widely used for recognitions and regressions.

The core of support vector regression (SVR) is to find out a hyperplane, namely, to a minimized linear function with inequality constraints, giving a training sample

$$S = (x_1, y_1), \dots, (x_n, y_n), \quad (1)$$

where  $x$  means input and  $y$  means output. The training sample needs to be turned into a linearly separable problem when it is a linear inseparability. Generally, the sample space can be mapped from a low-dimensional space into a high dimensional space by using a kernel function. The nonlinear regression function (estimation) is

$$f(x) = \omega\varphi(x) + b, \quad (2)$$

where  $b$  is a threshold,  $\omega$  is a weight, and  $\varphi(x)$  is the nonlinear mapping function. The loss function is shown as follows:

$$c(x, y, f) = |y - f(x)|_\varepsilon. \quad (3)$$

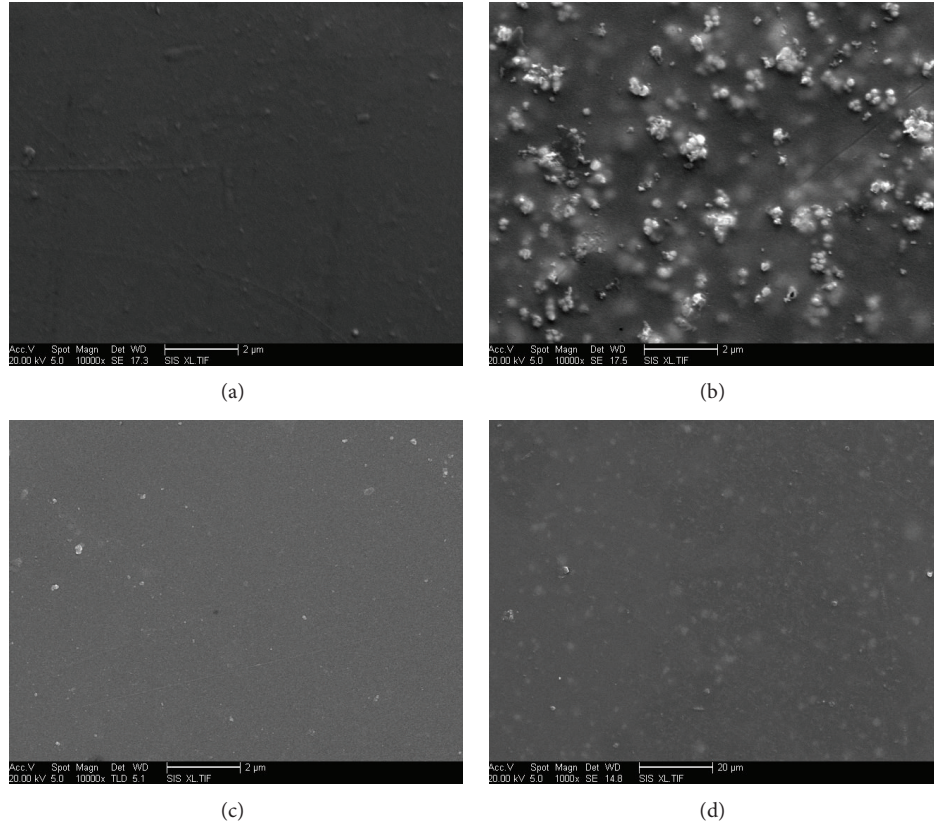


FIGURE 1: SEM image of the nanocomposite film. (a) The SEM image of pure PI film; (b) the SEM image of surface appearance of PI/BaTiO<sub>3</sub> composite film doped with content of 60 wt%; (c) the SEM image of surface appearance of PI/TiO<sub>2</sub> composite film doped with content of 5 wt%; (d) the SEM image of surface appearance of PI/Al<sub>2</sub>O<sub>3</sub> composite film doped with content of 20 wt%.

By importing a Lagrange multiplier, the minimization of the objective function can be expressed as

$$\begin{aligned}
 \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi'_i) \\
 \text{subject to} \quad & y_i - \omega \varphi(x_i) - b \leq \varepsilon + \xi_i \\
 & \omega \varphi(x_i) + b - y_i \leq \varepsilon + \xi'_i \\
 & \xi_i, \xi'_i \geq 0, \quad i = 1, 2, \dots, n.
 \end{aligned} \quad (4)$$

Transform (4) into a dual problem, and we get

$$\max \left[ -\frac{1}{2} \sum_{i,j}^n (a_i^* - a_i) (a_j^* - a_j) K(x_i, x_j) - \varepsilon \sum_{i=1}^n (a_i^* + a_i) + \sum_{i=1}^n y_i (a_i - a_i^*) \right], \quad (5)$$

where  $C$  is a penalty factor and  $a_i^*$  and  $a_i$  are Lagrangians; then the regression function is

$$f(x) = \sum_{i=1}^N (a_i - a_i^*) K(x_i, x_j) + b, \quad (6)$$

where  $N$  is the number of support vectors in (6) [18]. In this paper, a normalized polynomial kernel has been taken as the kernel function.

John Platt, from Microsoft Research, proposed the SMO (sequential minimal optimization) algorithm in order to shorten the training time of support vector regression in 1998 [19]. SMO can optimize the  $\alpha$  value of two samples at one time. Through the loop iteration with a given times  $M$ , the  $\alpha$  value of all samples can be optimized, as shown in Figure 2. For the data in Table 2, the method of 10-fold cross validation is used to make the model training on the standard SVR and SMO-SVR. The time of modeling SVR and SMO-SVR is 0.26 s and 0.03 s, respectively. SMO can promote the training efficiency of support vector regression while reducing the training time of the model.

*2.3. Promoting the Model by Stochastic Gradient Boosting.* Boosting [20], as one of the most important ensemble learning methods, is to obtain a predictive function by construction and ensemble of a series of predictive functions. The core idea of gradient boosting, proposed by Friedman [21], is to construct an ensemble learning machine by calculating a loss function and letting the function descend along its gradient. In other words, it is to calculate the loss function of previous model, so as to build a new model along the descending gradient direction of the loss function. Eventually, a regression

TABLE 2: Data of breakdown field strengths of nanocomposite films.

Samples	Type of nanoparticles	Dielectric constant	Input					Output	
			Thickness of films ( $\mu\text{m}$ )	Thermal conductivity ( $\text{W/cm}\cdot\text{k}$ )	Electrical resistivity ( $\Omega\cdot\text{m}$ )	Ratio of nanoparticles (wt%)	Size of nanoparticles (nm)	Specific area of nanoparticles ( $\text{m}^2/\text{g}$ )	Breakdown field strength of the hybrid PI ( $\text{kV/mm}$ )
1	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	10	100	12	157
2	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	15	100	12	137
3	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	20	100	12	125
4	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	25	100	12	119
5	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	30	100	12	105
6	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	50	100	12	72
7	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	60	100	12	54
8	BaTiO <sub>3</sub>	1400	25	0.50	$1 \times 10^5$	70	100	12	41
9	Rutile TiO <sub>2</sub>	100	25	0.40	$9 \times 10^7$	0	35	70	230
10	Rutile TiO <sub>2</sub>	100	25	0.63	$9 \times 10^7$	1	35	70	240
11	Rutile TiO <sub>2</sub>	100	25	0.63	$9 \times 10^7$	3	35	70	210
12	Rutile TiO <sub>2</sub>	100	25	0.63	$9 \times 10^7$	4	35	70	177
13	Rutile TiO <sub>2</sub>	100	25	0.63	$9 \times 10^7$	5	35	70	170
14	Rutile TiO <sub>2</sub>	100	25	0.63	$9 \times 10^7$	7	35	70	160
15	SiO <sub>2</sub>	1.56	25	160	$1 \times 10^{16}$	5	40	300	313
16	SiO <sub>2</sub>	1.56	25	160	$1 \times 10^{16}$	10	40	300	307
17	SiO <sub>2</sub>	1.56	25	160	$1 \times 10^{16}$	15	40	300	327
18	SiO <sub>2</sub>	1.56	25	160	$1 \times 10^{16}$	20	40	300	252
19	SiO <sub>2</sub>	1.56	25	160	$1 \times 10^{16}$	25	40	300	240
20	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	4	30	25	275
21	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	8	30	25	263
22	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	12	30	25	234
23	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	16	30	25	212
24	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	20	30	25	235
25	$\alpha\text{Al}_2\text{O}_3$	10	30	4.10	$1 \times 10^{14}$	24	30	25	203
26	Al <sub>2</sub> O <sub>3</sub>	8	30	29.31	$1 \times 10^{14}$	15	13	100	190
27	Al <sub>2</sub> O <sub>3</sub>	8	30	29.31	$1 \times 10^{14}$	20	13	100	189
28	Al <sub>2</sub> O <sub>3</sub>	8	30	29.31	$1 \times 10^{14}$	25	13	100	178
29	SiO <sub>2</sub>	1.56	30	160	$1 \times 10^{16}$	10	7	350	307
30	SiO <sub>2</sub>	1.56	30	160	$1 \times 10^{16}$	15	7	350	327
31	SiO <sub>2</sub>	1.56	30	160	$1 \times 10^{16}$	20	7	350	252
32	SiO <sub>2</sub>	1.56	30	160	$1 \times 10^{16}$	25	7	350	240

model can be generated through  $N$  times of iterations. In this paper, we use the Stochastic Gradient Boosting to turn ten SMO-SVR models into a strong predictive model (SGBS model). The algorithmic steps of the SGBS model are shown as follows:

- (1)  $\{X, Y\}$  is a data sample of breakdown field strengths of different nanocomposite thin films, in which  $X = (x_1, x_2, \dots, x_n)$  is input and  $Y = (y_1, y_2, \dots, y_n)$  is output. The loss function is  $L(y_i, F(x_i))$ ,  $i \in \{1, 2, \dots, n\}$ , and the prediction model of SMO-SVR is  $h(X)$ .

- (2) Initialize the model

$$F_0(x) = \arg \min_{\beta} \sum_{i=1}^N L(y_i, \beta). \quad (7)$$

- (3) Calculate the upper limit ( $M$ ) of iterations by cross validation.

- (4) Calculate the gradient direction

$$\bar{y}_i = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}, \quad i = 1, \dots, N. \quad (8)$$

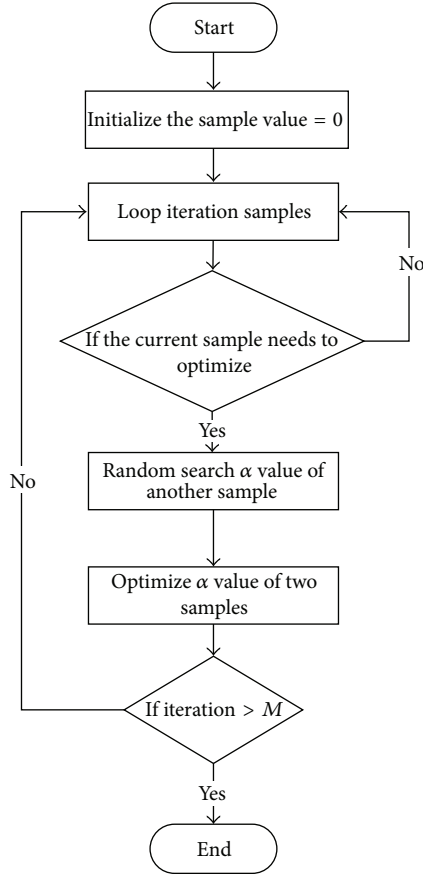


FIGURE 2: Flowchart of the SMO optimization algorithm.

- (5) Fit the SMO-SVR by using least squares to obtain the fitting model  $h(x_i, \alpha_m)$ .
- (6) Figure out new step length based on the loss function

$$\beta_m = \arg \min \sum_{i=1}^N L(y_i, F_{m-1}(x_i)) + \beta h(x_i, \alpha_m). \quad (9)$$

- (7) Generate a new model

$$F_m(x) = F_{m-1}(x) + \beta_m h(x, \alpha_m). \quad (10)$$

- (8) Stochastically extract  $f\%$  of the training sample to fit the SMO-SVR during every time of iteration; then we can get the SGBS model after  $M$  times of iterations:

$$F(x) = \sum_{m=1}^M \beta_m h(x, \alpha_m). \quad (11)$$

### 3. Experiment and Result Analysis

**3.1. Construction of the Experimental Sample.** Test the breakdown field strength by using a withstanding voltage tester (type: CS2674C). Voltage range is 0–50 kV; test error is  $\pm 5\%$ ; leakage current measurement range is 0.5–20 mA; measurement error is  $\pm 5\%$ . Take polymethylphenylsiloxane fluid with

high insulation strength (which can be up to 16 kV/mm) as the medium, with the rate of voltage rise of 500 V/s. Test the breakdown field strength of the nanocomposite films prepared in Section 2.1. Select 20 samples as a group for testing each hybrid PI, and take the mean value of the 10 middle breakdown field strengths as the breakdown field strength of the group. Then, calculate the standard deviation. After the experiment we get the data of breakdown field strengths as shown in Table 2. The input  $X$  includes type ( $X_1$ ), doping ratio ( $X_2$ ), electrical resistivity ( $X_3$ ), dielectric constant ( $X_4$ ), thermal conductivity ( $X_5$ ), size ( $X_6$ ), and specific area ( $X_7$ ) of nanoparticles and thickness of the film ( $X_8$ ). The output  $Y$  is the breakdown field strength of the hybrid PI.

**3.2. Evaluation Indicators of the Prediction Performance.** There are many indicators for evaluating the prediction performance, such as MAPE (mean absolute percentage error), MAE (mean absolute error), RMSE (root mean squared error), CC (correlation coefficient), RAE (relative absolute error), and RRSE (root relative squared error). In this paper, CC, MSE, RMSE, RSE, and RRSE are taken as evaluation indicators.

CC (correlation coefficient) represents the relevancy of the linear regression relationship between  $f_i$  (independent variable) and  $y_i$  (response variable) of the model. It is a coefficient of a simple linear correlation between  $y_i$  and its estimated values. CC is larger than zero and ranges from 0 to 1. A larger CC means a more relevant linear regression relationship.

MSE (mean absolute error) is to judge the different degree between predicted values and real ones. It is inversely proportional to prediction accuracy. A smaller MSE means a better effect the predictor can be with. It can be expressed as

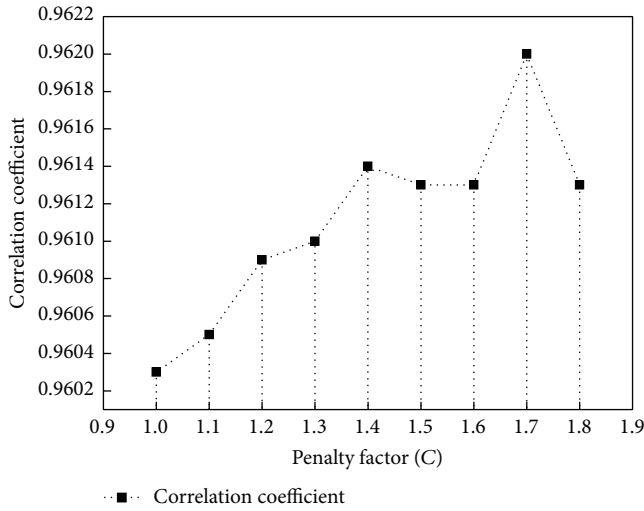
$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| = \frac{1}{n} \sum_{i=1}^n |e_i|, \quad (12)$$

where  $f_i$  represents predicted values and  $y_i$  represents real values.

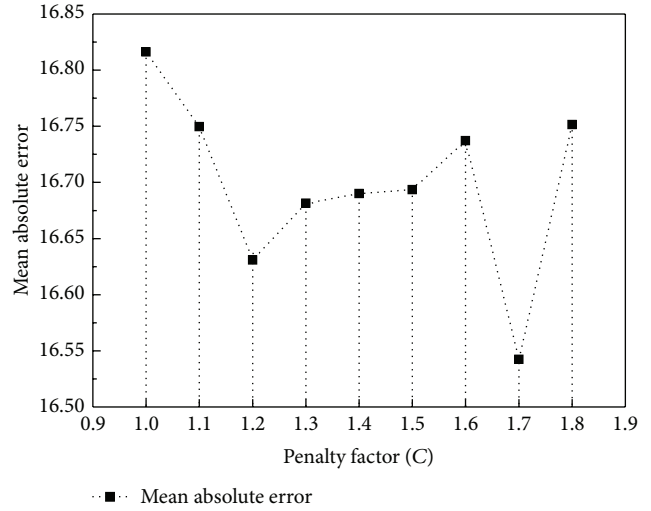
RMSE (root mean squared error) is the square root of the ratio of the quadratic sum of deviations between predicted values and real values to the times ( $n$ ) of predictions. It is sensitive to maximum or minimal errors of a group of predicted values and therefore can well reflect prediction accuracy. RMSE is inversely proportional to the prediction accuracy. The smaller the RMSE is, the more accurate the predictor can be. It can be expressed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (f_i - y_i)^2}{n}}. \quad (13)$$

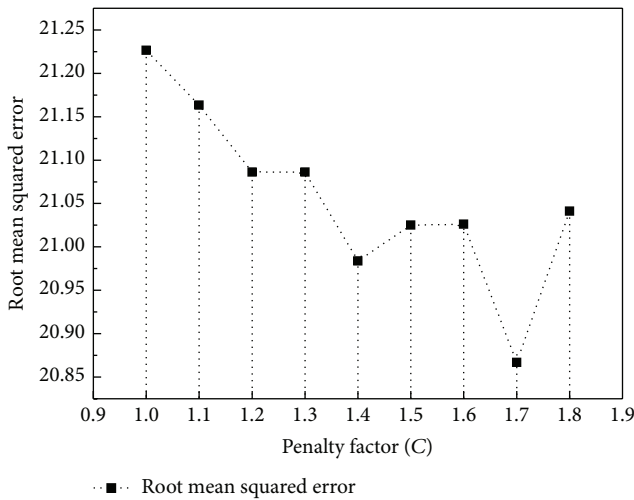
RSE (relative absolute error) is the different degree between the absolute deviation obtained from the prediction model and the absolute deviation obtained by directly speculating the training sample. It is inversely proportional to



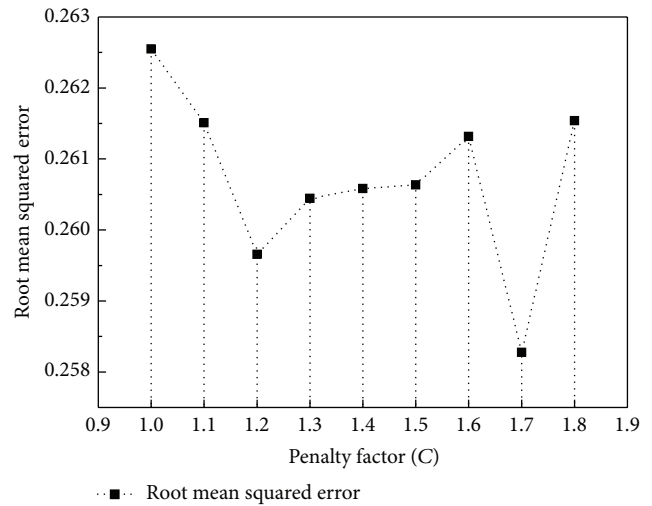
(a)



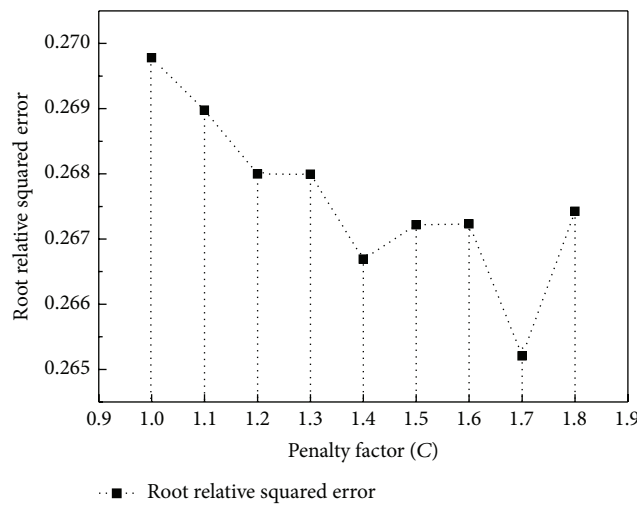
(b)



(c)

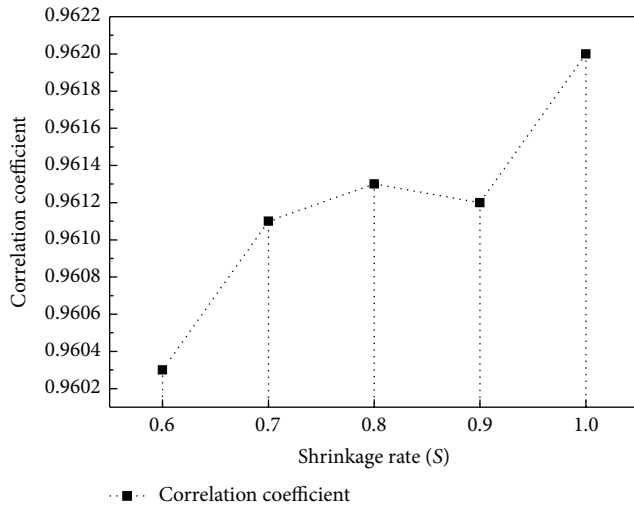


(d)

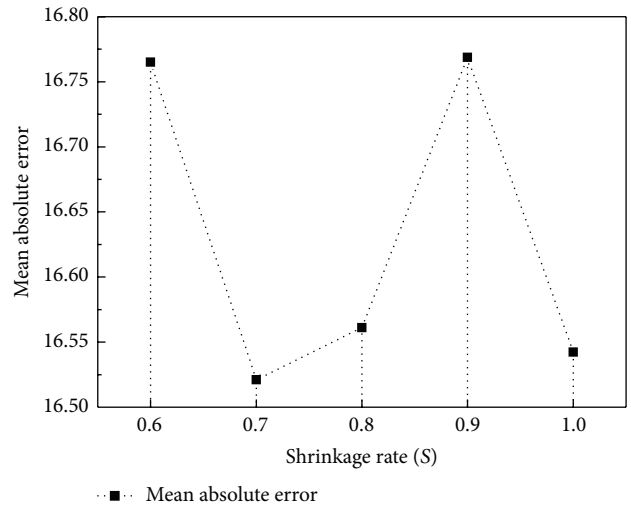


(e)

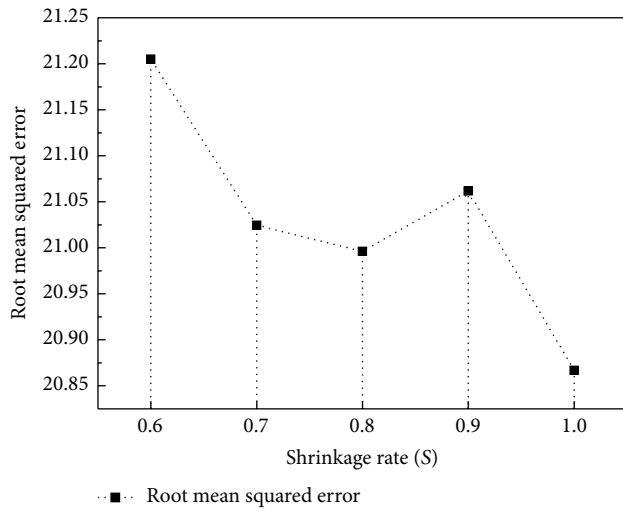
FIGURE 3: The impacts of variation of penalty factor ( $C$ ) on prediction performance. (a) The correlation coefficient with different penalty factors ( $C$ ); (b) the mean absolute error with different penalty factors ( $C$ ); (c) the mean absolute error with different penalty factors ( $C$ ); (d) the root mean squared error with different penalty factors ( $C$ ); (e) the root relative squared error with different penalty factor ( $C$ ).



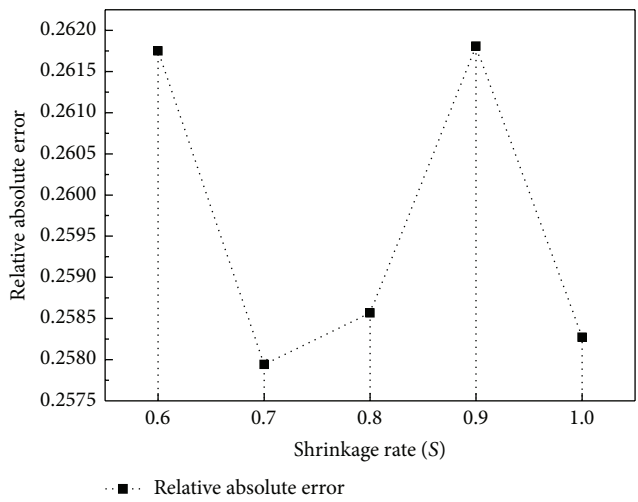
(a)



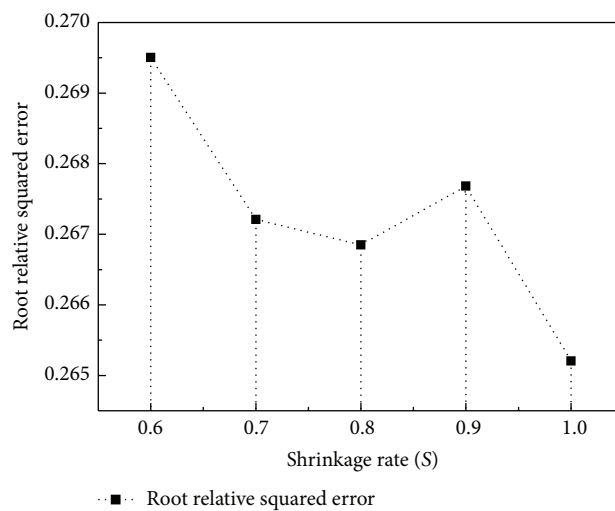
(b)



(c)



(d)



(e)

FIGURE 4: Impact of  $S$  on prediction performance. (a) Correlation coefficient relative to shrinkage rate ( $S$ ) of training samples; (b) mean absolute error relative to shrinkage rate ( $S$ ) of training samples; (c) root mean squared error relative to shrinkage rate ( $S$ ) of training samples; (d) relative absolute error relative to shrinkage rate ( $S$ ) of training samples; (e) root relative squared error relative to shrinkage rate ( $S$ ) of training samples.

TABLE 3: Comparison on prediction performances of three kernel functions.

Indicators	Normalized polynomial	Polynomial	RBF
Correlation coefficient	0.9603	0.9236	0.8901
Mean absolute error	16.8162	23.5927	29.6448
Root mean squared error	21.2266	29.2044	34.7791
Relative absolute error	26.255%	36.8351%	46.2842%
Root relative squared error	26.9779%	37.1173%	44.2024%

prediction accuracy. The smaller the RSE is, the higher the prediction accuracy can be:

$$\text{RSE} = \frac{\sum_{i=1}^n |f_i - y_i|}{\sum_{i=1}^n |\bar{f}_i - y_i|}. \quad (14)$$

RRSE (root relative squared error) can be calculated as follows:

$$\text{RRSE} = \frac{\sum_{i=1}^n |f_i - y_i|^2}{\sum_{i=1}^n |\bar{f}_i - y_i|^2}. \quad (15)$$

RRSE is also inversely proportional to prediction accuracy. The smaller the RRSE is, the higher the prediction accuracy can be [22].

**3.3. Experimental Results and Analysis.** To verify the SGBS prediction model, the type, dielectric constant, electrical resistivity, thermal conductivity, size and specific area of nanoparticles, and the thickness of films in Table 2 are taken as the input  $X$ , and the breakdown field strength of hybrid PI is taken as the output  $Y$ . In this paper, we use Macbook Pro (CPU: Intel I7-2640 M; memory: 16G) as the hardware for experiments and use Matlab 2012a to program prediction model.

*Experiment 1.* Use the method of 10-fold cross validation to fit the data in Table 2. Main parameters of the SGBS model include kernel function,  $C$  and  $S$ , of which  $C$  is the penalty factor of SMO-SVR and  $S$  is the extracting ratio when training the sample. First of all, normalize the sample data. Three kernel functions—normalized polynomial kernel, polynomial kernel, and RBF kernel—are employed to test the sample. The results are listed in Table 3. From the table we know that the normalized polynomial kernel has the highest correlation coefficient but lowest mean absolute error, root mean squared error, relative absolute error, and root relative squared error. Therefore, we choose it as the kernel function for the SGBS model.

For optimization function, the penalty factor of outliers is selected by experience as well as by experiments. Figure 3 shows the comparison on prediction performances of SGBS model when  $C$  varies from 1 to 1.8. Figure 3(a) demonstrates that the correlation coefficient reaches its peak values when  $C = 1.6, 1.7,$  and  $1.8,$  which are chosen as the value of  $C$ , for

correlation coefficient is proportional to prediction accuracy. By analyzing Figures 3(b) and 3(d) we know that the mean absolute error and relative absolute error reach their minimal values when  $C = 1.7,$  only larger than the values when  $C = 1.6$  and  $1.8.$  Therefore, they reach optimal performance when  $C = 1.7,$  for mean absolute error and relative absolute error are inversely proportional to prediction accuracy.

By analyzing Figures 3(c) and 3(d) we know that the smallest root mean squared error and root relative squared error appear at  $C = 1.6,$  followed by  $C = 1.7$  and  $C = 1.8.$  Based on the above analysis, this paper takes  $C = 1.7$  as the optimum value.

$S$  is the specify shrinkage rate in Stochastic Gradient Boosting, namely, the proportion of the stochastically extracted training sample. Figure 4 shows the impact of variation of  $S$  on prediction performance when  $C = 1.7.$  When  $S = 1,$  correlation coefficient reaches its peak value, and the prediction attains the best performance as shown in Figure 4(a). By analyzing Figures 4(b)–4(e) we know that when  $S = 1,$  mean absolute error, root mean squared error, relative absolute error, and root relative squared error reach their minimal values, signifying the best prediction performance.

According to the above analysis we know that when  $C = 1.7$  and  $S = 1$  are taken as the optimum values of the SGBS model, the prediction performance is the best. Figure 5 shows the degree of fitting between predicted values and real ones. Figure 6 shows the absolute error ratio of prediction. The tables and figures demonstrate that the error ratios of sample 8 (70 wt%, 100 nm, BaTiO<sub>3</sub>), sample 17 (15 wt%, 40 nm, SiO<sub>2</sub>), and sample 30 (15 wt%, 7 nm, SiO<sub>2</sub>) are all larger than 15%, indicating an ordinary fitting, while the error ratios of other samples are less than 15%, indicating a better fitting. For all the six multicomponent TiO<sub>2</sub> film samples with the thickness of 35 μm, the prediction errors are not larger than 15%. For the eight BaTiO<sub>3</sub> film samples, there is one sample of which the prediction errors are larger than 15%. For all the nine multicomponent Al<sub>2</sub>O<sub>3</sub> film samples, the prediction errors are not larger than 15%, while for the nine multicomponent SiO<sub>2</sub> film samples, there are also two samples of which the prediction errors are larger than 15 wt%. In this model, the prediction performance of the multicomponent Al<sub>2</sub>O<sub>3</sub> and rutile TiO<sub>2</sub> with different thicknesses is better than that of the BaTiO<sub>3</sub> and SiO<sub>2</sub> film samples. For the thirty-two nanocomposite films with different components, mixtures, and thicknesses, there are twenty-nine film samples of which the prediction errors are lower than 15%, proving that the model is of practical value in actual engineering works.

*Experiment 2.* In order to further verify the SGBS model, it needs to be compared to other models, namely, the linear regression, BP neural network, GRNN neural network, SVR (support vector regression), and SMO-SVR, under the same conditions. Comparison results of the prediction performance of these models are shown in Table 4. The correlation coefficient of SGBS model is 0.962, larger than that of the models of linear regression [23], BP neural network [24], GRNN neural network [7], SMO-SVR [25], and SVR [26], proving that the linear regression relationship of the SGBS model is better than that of the other five models. In the SGBS



TABLE 4: Comparisons of models' prediction performance.

Indicators	SGBS	Linear regression	BP	GRNN	SVR	SMO-SVR
CC	0.962	0.9438	0.960	0.957	0.459	0.917
RMSE	20.866	25.235	21.350	21.932	79.662	30.401
RRSE	26.520%	32.072%	27.135%	27.874%	101.246%	38.639%

TABLE 5: The comparison of the predicted and tested values of breakdown field strength.

Measured value [27] (kV/mm)	Predicted value (kV/mm)	MAE	RMSE	RAE (%)	RRSE (%)	Error ratio (%)
233	225.9667	7.0334	7.0334	24.5976	24.5976	3.0185

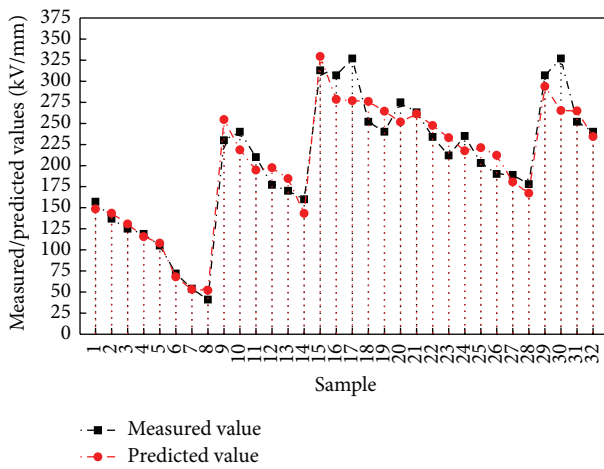


FIGURE 5: Comparison between real values and predicted values of the SGBS mode.

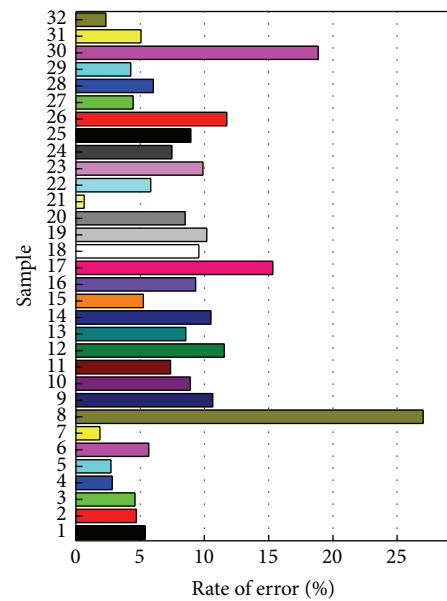


FIGURE 6: Analysis on sample errors.

model, the root mean squared error and root mean squared error are 20.8668 and 26.520%, respectively, lower than those of the other models, certifying its better prediction performance.

*Experiment 3.* In order to validate the generosity of the model, the actual measurement results conducted by Shi et al. in [27] are chosen to do the prediction. In this preference, the particle size is  $30\ \mu\text{m}$ , and the doping ratio is 2 wt% with PI/nano- $\text{Al}_2\text{O}_3$  composite films. The thickness of film is  $30\ \mu\text{m}$ . Its breakdown field strength is 233 kV/mm. Adopting the samples in Table 2 as the training set of SGBS model. For the parameters of the model,  $C = 1.7$ ,  $S = 1$ . Using the 2 wt%-PI/nano- $\text{Al}_2\text{O}_3$  composite films in [27] as testing sample, the results could be obtained as shown in Table 5. For the test result of breakdown field strength that is 233 kV/mm, its prediction value is 225.9667 kV/mm by SGBS model. The error ratio is 3.0185%. The MAE, RMSE, RAE, and RRSE are 7.0334, 7.0334, 24.5976, and 24.5976. The prediction and actual data are in good agreement. This method could predict the PI/nano- $\text{Al}_2\text{O}_3$  composite films effectively.

## 4. Conclusions

This paper presents an ensemble learning method for predicting breakdown field strength of polyimide nanocomposite films. By using the method of Stochastic Gradient Boosting, ten SMO-SVR prediction models are constructed into a strong prediction model (SGBS model) that is efficient in predicting the breakdown field strength. Through analyzing the experiment data we obtain following conclusions:

- (1) In prediction of thirty-two nanocomposite films of different components, particles, and thicknesses (25– $30\ \mu\text{m}$ ) by using the method of 10-fold cross validation, there are twenty-nine samples of which the prediction errors are lower than 15%, proving that the SGBS model is efficient in predicting the breakdown field strength of polyimide nanocomposite films.
- (2) Comparisons show that the SGBS model has a larger correlation coefficient than that of linear regression, BP, GRNN, SVR, and SMO-SVR models but smaller

root mean squared error and root relative squared error. Hence, prediction performance of the SGBS model is better than that of the other five models.

- (3) The SGBS model shows a better prediction on  $\text{Al}_2\text{O}_3$  and rutile  $\text{TiO}_2$  films than on  $\text{BaTiO}_3$  and  $\text{SiO}_2$  films.

Next, some other ensemble learning methods will be employed to predict the corona resistance, dielectric constant, dielectric constant, and thermal properties of polyimide nanocomposite films.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

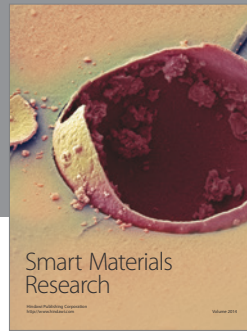
## Acknowledgments

This paper has obtained the support of the National Natural Science Foundation of China (51077028, 51307046, and 60803096) and Heilongjiang Natural Science Foundation of China (A201006), National Basic Research Program of China (2012CB723308), Fund of the State Ethnic Affairs Commission of China, and the Fundamental Research Funds for the Central Universities. Authors also gratefully acknowledge the helpful comments and suggestions of the reviewers who improved the presentation.

## References

- [1] W. J. Lee and Y. B. Kim, "Adhesion and interfacial characteristics of metal/PI composite film modified by  $\text{O}_2$  ion beam," *Thin Solid Films*, vol. 517, no. 3, pp. 1191–1194, 2008.
- [2] N. An, H. Liu, Y. Ding, M. Zhang, and Y. Tang, "Preparation and electroactive properties of a PVDF/nano- $\text{TiO}_2$  composite film," *Applied Surface Science*, vol. 257, no. 9, pp. 3831–3835, 2011.
- [3] K. Sivudu, Y. Mahajan, and S. Joshi, "Nano-enabled tribological thin film coatings: global patent scenario," *Recent Patents on Nanotechnology*, vol. 8, no. 2, pp. 97–116, 2014.
- [4] Y. Song, Y. Shen, H. Liu, Y. Lin, M. Li, and C.-W. Nan, "Improving the dielectric constants and breakdown strength of polymer composites: effects of the shape of the  $\text{BaTiO}_3$  nano-inclusions, surface modification and polymer matrix," *Journal of Materials Chemistry*, vol. 22, no. 32, pp. 16491–16498, 2012.
- [5] L. Jiang, Y. Zhao, and J. Zhai, "A lotus-leaf-like superhydrophobic surface: a porous microsphere/nanofiber composite film prepared by electrohydrodynamics," *Angewandte Chemie*, vol. 43, no. 33, pp. 4338–4341, 2004.
- [6] Z. Zhou, M. Mackey, K. Yin et al., "Fracture phenomena in micro- and nano-layered polycarbonate/poly(vinylidene fluoride-co-hexafluoropropylene) films under electric field for high energy density capacitors," *Journal of Applied Polymer Science*, vol. 131, no. 4, Article ID 39877, 2014.
- [7] Y.-S. Yang, J.-H. Chou, W. Huang, T.-C. Fu, and G.-W. Li, "An artificial neural network for predicting the friction coefficient of deposited  $\text{Cr}_{1-x}\text{Al}_x\text{C}$  films," *Applied Soft Computing Journal*, vol. 13, no. 1, pp. 109–115, 2013.
- [8] E. N. Cho, P. Moon, C. E. Kim, and I. Yun, "Modeling and optimization of ITO/Al/ITO multilayer films characteristics using neural network and genetic algorithm," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8885–8889, 2012.
- [9] A. Bahramian, "Study on growth rate of  $\text{TiO}_2$  nanostructured thin films: simulation by molecular dynamics approach and modeling by artificial neural network," *Surface and Interface Analysis*, vol. 45, no. 11, pp. 1727–1736, 2013.
- [10] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: the contribution of ensemble learning," *Decision Support Systems*, vol. 57, no. 1, pp. 77–93, 2014.
- [11] M. Claesen, F. De Smet, J. A. K. Suykens, and B. De Moor, "EnsembleSVM: a library for ensemble learning using support vector machines," *Journal of Machine Learning Research*, vol. 15, pp. 141–145, 2014.
- [12] N. Rooney, H. Wang, and P. S. Taylor, "An investigation into the application of ensemble learning for entailment classification," *Information Processing & Management*, vol. 50, no. 1, pp. 87–103, 2014.
- [13] K. P. Singh, S. Gupta, and D. Mohan, "Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches," *Journal of Hydrology*, vol. 511, pp. 254–266, 2014.
- [14] J. Xu, Q. Wu, J. Zhang, F. Shen, and Z. Tang, "Boosting separability in semisupervised learning for object classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 1197–1208, 2014.
- [15] F. Mordelet and J.-P. Vert, "A bagging SVM to learn from positive and unlabeled examples," *Pattern Recognition Letters*, vol. 37, no. 1, pp. 201–209, 2014.
- [16] T. P. Williams and J. Gong, "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers," *Automation in Construction*, vol. 43, pp. 23–29, 2014.
- [17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [18] M. Maalouf, N. Khoury, J. G. Laguros, and H. Kumin, "Support vector regression to predict the performance of stabilized aggregate bases subject to wet-dry cycles," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 36, no. 6, pp. 675–696, 2012.
- [19] C. J. C. Burge, "Sequential minimal optimization: a fast algorithm for training support vector machines," NeuroCOLT2 Technical Report Series NC2-TR-1998030, October 1998.
- [20] E. Bauer and R. Kohavi, "Empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Machine Learning*, vol. 36, no. 1, pp. 105–139, 1999.
- [21] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [22] T. E. Hoff, R. Perez, J. Kleissl, D. Renne, and J. Stein, "Reporting of irradiance modeling relative prediction errors," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 7, pp. 1514–1519, 2013.
- [23] P. Y. Lai and S. M. S. Lee, "Estimation of central shapes of error distributions in linear regression problems," *Annals of the Institute of Statistical Mathematics*, vol. 65, no. 1, pp. 105–124, 2013.
- [24] Y. Lou, W. Wu, and L. Li, "Inverse identification of the dynamic recrystallization parameters for  $\text{AZ}_{31}$  magnesium alloy using BP neural network," *Journal of Materials Engineering and Performance*, vol. 21, no. 7, pp. 1133–1140, 2012.

- [25] L. Dong, Z. Ren, and Q. Li, "Fault prediction for aircraft control surface damage based on SMO-SVR," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 38, no. 10, pp. 1300–1305, 2012 (Chinese).
- [26] H. Guo, J. Yin, J. Zhao, Z. Huang, and Y. Pan, "Prediction of fatigue life of packaging EMC material based on RBF-SVM," *International Journal of Materials & Product Technology*, vol. 49, no. 1, pp. 5–17, 2014.
- [27] H. Shi, L. Liu, L. Weng, J. Ding, and W. Cui, "The effects of coupling agents on the properties of polyimide/nano- $\text{Al}_2\text{O}_3$  composite films," in *Proceedings of the 6th International Forum on Strategic Technology (IFOST '11)*, vol. 1, pp. 33–36, Harbin, China, August 2011.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

