# Research on Dehumidification Efficiency Prediction of Transformer Breather Silica Gel Air-Drying and Stirring System Based on Improved Random Forest Algorithm

**Guangyao Wang*, Chaofei Yan, Zhikang Sun, Yanmei Cao, Kaixin Ding, Pinyi Zhao**
*State Grid Jia Xian Electric Power Supply Company of Henan Electric Power Company, Pingdingshan, China*
**Corresponding Author*

**Abstract: Transformer breathers play a critical role in maintaining the insulation performance of transformers by preventing moisture ingress, and the dehumidification efficiency of silica gel in the breather directly affects the stable operation of the entire transformer system. However, the traditional method of evaluating dehumidification efficiency through repeated experiments is time-consuming and costly, making it difficult to meet the real-time monitoring needs of transformer operation. To address this problem, this study proposes an improved random forest (RF) algorithm for predicting the dehumidification efficiency of a transformer breather silica gel air-drying and stirring system. First, the mutual information method is introduced to screen the core influencing factors (including silica gel initial humidity, stirring speed, microwave heating temperature, and air-drying airflow intensity) from multiple operation parameters, eliminating redundant features and reducing model complexity. Then, the hyperparameters of the RF algorithm are optimized using the grid search method to enhance the model's prediction accuracy. Experimental results show that compared with the traditional RF algorithm and XG Boost algorithm, the improved RF algorithm proposed in this study reduces the prediction error by 8.2% and 5.6% respectively, and the prediction speed is increased by 12.5%. This model can quickly and accurately predict the dehumidification efficiency of the system, providing a reliable decision-making basis for the intelligent adjustment and optimal operation of the transformer breather silica gel air-drying and stirring system.**

**Keywords: Transformer Breather; Silica Gel Dehumidification; Random Forest Algorithm; Efficiency Prediction; Feature Selection**

## 1. Introduction

With the rapid development of industrial artificial intelligence, data-driven methods are gradually revolutionizing the operation and maintenance mode of traditional power equipment [1]. As a core device of the power grid, the insulation performance of transformers is directly related to the safe operation of the system. Transformer breathers adsorb moisture through internal silica gel and are key components to prevent insulation from getting damp. Their dehumidification efficiency directly affects the operational stability of transformers [2,3]. However, the dehumidification efficiency of silica gel is affected by the coupling of multiple parameters, and traditional experimental evaluation methods are difficult to achieve real-time and accurate monitoring. Developing an efficient dehumidification efficiency prediction model is of great significance for realizing intelligent operation and maintenance and predictive maintenance of transformers [4,5].

In the field of industrial process prediction modeling, various machine learning algorithms have been widely applied [6]. Early studies relied heavily on linear models or simple decision trees, but they performed poorly when dealing with nonlinear and high-dimensional data. In recent years, ensemble learning algorithms, such as Random Forest and Gradient Boosting Machines, have received widespread attention due to their excellent prediction performance and robustness. Random Forest effectively avoids the overfitting problem by constructing multiple decision trees and integrating their results, and is used for equipment fault prediction and performance evaluation [7,8]. XG Boost, as an efficient gradient boosting framework, has achieved leading results in many data science competitions and industrial applications through its regularization terms and parallel processing capabilities, and is often used for feature importance analysis and regression prediction tasks.

Specific to condition monitoring and efficiency prediction, existing studies mainly have two major drawbacks. First, in terms of feature processing, many methods directly use all available parameters for modeling and fail to effectively identify the core feature set most relevant to the prediction target. This not only introduces noise and increases model complexity but may also lead to the "curse of dimensionality", affecting the generalization ability and interpretability of the model. Second, in terms of model construction, although off-the-shelf algorithms such as RF and XG Boost are powerful, their performance largely depends on the setting of hyperparameters. Many application studies directly adopt the default parameters of the algorithm or only perform limited manual parameter tuning, which makes the model unable to reach its optimal performance potential, and there is a large room for improvement in prediction accuracy [9,10]. More importantly, most existing methods focus on offline analysis and post-event evaluation, and their model training and prediction processes are usually time-consuming, making it difficult to meet the scenario requirements of transformer breathers that need real-time or near-real-time state perception and feedback control [11-13].

To address the limitations of existing research methods in insufficient feature selection, inadequate model optimization, and lack of timeliness, this study proposes a prediction scheme based on the improved Random Forest algorithm. The core lies in combining precise feature engineering with in-depth model optimization. First, to overcome the problems of feature redundancy and high model complexity, this study introduces the mutual information method to screen core influencing factors. This method can quantify the linear and nonlinear dependence between features and target variables, thereby accurately identifying the core factors that are most decisive for dehumidification efficiency from many operating parameters. Second, to solve the pain point that insufficient hyperparameter optimization in traditional applications leads to limited prediction accuracy, this study uses the grid search method to systematically optimize the hyperparameters of the Random Forest algorithm. By traversing the pre-defined parameter combination space, this method can ensure finding the optimal configuration within the given range, thereby fully exploiting the prediction potential of the Random Forest algorithm, significantly improving the accuracy and robustness of the model, and avoiding performance losses caused by manual or default parameter settings.

## 2. Algorithm Design

### 2.1 Random Forest Algorithm

Random Forest is an ensemble learning algorithm that completes classification or regression tasks by constructing and combining multiple decision trees. Its core idea is gathering diverse ideas for broader insights [14,15]. It introduces randomness in data and features to build a series of decision trees with high heterogeneity (weak learners), and then aggregates them into a strong learner through voting or averaging, thereby effectively improving the generalization ability of the model and suppressing overfitting for the regression problem studied in this paper, its specific principles are as follows [16,17]:

Given a training set $D=\{(x_1,y_1),(x_2,y_2),...,(x_m,y_m)\}$ containing $m$ samples, where $x_i$ is an $n$ -dimensional feature vector (such as initial grouting pressure, grouting speed, etc.) and $y_i$ is the corresponding true value of grouting efficiency. The construction process of Random Forest is as follows:

Bootstrap Sampling: Through sampling with replacement, $T$ subsets $D_1, D_2,..., D_T$ each with a size of $m$ are randomly drawn from the original training set $D$ for training $T$ decision trees. The samples not selected in each sampling form out-of-bag data, which can be used for internal evaluation of model performance.

Decision Tree Generation: In the process of constructing each decision tree, when a split node needs to be divided, instead of selecting the optimal feature from all $n$ features, a feature subset is first randomly selected (usually with a size of $k$=$\log_2 n$ or $\sqrt{n}$), and then the best split feature and split point are selected in this subset. This mechanism further enhances the heterogeneity among the trees.

Model Ensemble and Prediction: For regression problems, the final output of Random Forest is the arithmetic mean of the prediction results of all decision trees.

Assume that a single decision tree $\mathrm{h_t(x)}$ predicts $\hat{\mathrm{y}}_\mathrm{t}$ for the input feature vector x. Then, the final prediction $\hat{\mathrm{y}}_\mathrm{RF}$ of the entire Random Forest is:

$$\hat{y}_{RF}=\frac{1}{T}\sum_{t=1}^{T}\mathrm{h}_t(\mathbf{x}) \quad (1)$$

Where: $\hat{y}_{RF}$ represents the final prediction of the Random Forest model for the dehumidification efficiency. $T$ denotes the total number of decision trees in the Random Forest. $\mathrm{h}_t(\mathbf{x})$ represents the prediction output of the $t$-th decision tree for input $\mathbf{x}$.

In the node splitting process of decision trees, criteria such as Mean Squared Error (MSE) or Gini impurity are usually adopted to select the optimal split point. For regression trees, the most commonly used splitting criterion is to minimize the MSE of child nodes. For node $s^*$, its MSE is calculated as:

$$MSE(s)=\frac{1}{N_s}\sum_{i\in s}(y_i-\bar{y}_s)^2 \quad (2)$$

Among them: $N_s$ represents the number of samples contained in node $s^*$. $y_i$ represents the true dehumidification efficiency value of the $i$-th sample in node $s$. $\bar{y}_s$ represents the true average dehumidification efficiency of all samples in node $s$.

Through this ensemble strategy, Random Forest can balance the overfitting problem that a single decision tree may have, demonstrating strong nonlinear fitting ability and robustness, and is very suitable for modeling complex systems with multiple parameter couplings such as transformer breathers.

## 2.2 Algorithm Analysis

The Random Forest algorithm performs excellently in handling complex regression problems. However, when directly applied to the real-time prediction scenario of transformer breather dehumidification efficiency, its inherent limitations in feature engineering and model optimization become prominent.

2.2.1 Insufficient feature selection

In industrial process data, the initial feature set often contains a large number of features that are weakly correlated or redundant with the prediction target. Let the original feature space be $\mathbf{X}=\{X_1,X_2,...,X_n\}$ and the target variable be $Y$ (dehumidification efficiency). The traditional application of Random Forest usually inputs all features into the model directly, which will lead to two serious problems:

First, a too high feature dimension will cause a sharp increase in computational complexity. When Random Forest splits nodes, it needs to evaluate the importance of features, and its time complexity can be expressed as:

$$O(T\cdot k\cdot m\log m) \quad (3)$$

Among them, $T$ is the number of trees, $k$ is the size of the feature subset (usually taken as $\sqrt{n}$ or $\log_2 n$), and $m$ is the number of samples. As the number of features $n$ increases, it not only directly increases the computation but also further intensifies the computational burden through the increase of $k$.

Second, irrelevant features will introduce noise and affect the generalization ability of the model. Consider the expected prediction error of the model:

$$E\left[\left(Y-\hat{f}(\mathbf{X})\right)^2\right]=\mathrm{Bias}(\hat{f})^2+\mathrm{Var}(\hat{f})+\sigma_\epsilon^2 \quad (4)$$

Where $\sigma_\epsilon^2$ is the inherent error. When the feature set contains a large number of irrelevant features, the model variance $\mathrm{Var}(\hat{f})$ will increase significantly, leading to overfitting. Specifically, it is manifested as a small error of the model on the training set:

$$\epsilon_{\mathrm{train}}=\frac{1}{m_{\mathrm{train}}}\sum_{i=1}^{m_{\mathrm{train}}}(y_i-\hat{y}_i)^2 \quad (5)$$

But the error is large on the test set:

$$\epsilon_{\mathrm{test}}=\frac{1}{m_{\mathrm{test}}}\sum_{i=1}^{m_{\mathrm{test}}}(y_i-\hat{y}_i)^2 \quad (6)$$

This decline in generalization ability seriously affects the reliability of the model in practical applications.

2.2.2 Insufficient hyperparameter optimization

The high performance of Random Forest highly depends on hyperparameter configuration. Let the hyperparameter space of Random Forest be Θ, which includes key parameters such as the number of trees n_estimators, maximum tree depth max_depth, and minimum number of samples for node splitting min_samples_split. Traditional parameter setting methods usually adopt empirical values or limited manual tuning, which will cause the model to fail to achieve optimal performance. Consider the true objective function of the model:

$$\theta^{*}=\arg\min_{\theta\in\Theta}\mathrm{E}_{(X,Y)}\left[L\left(Y,f(X;\theta)\right)\right] \quad (7)$$

Where $L$ is the loss function. Since the true distribution cannot be obtained, we usually use cross-validation error as the surrogate objective:

$$\hat{\theta}^{*}=\arg\min_{\theta\in\Theta}\frac{1}{K}\sum_{k=1}^{K}L\left(Y_{\mathrm{valid}}^{(k)},f\left(X_{\mathrm{valid}}^{(k)};\theta\right)\right) \quad (8)$$

Manual tuning can only search among limited parameter combinations, and it is easy to fall into local optima, making it impossible to find the global optimum $\hat{\theta}^{*}$. This directly leads to the prediction accuracy of the model failing to reach its theoretical potential.

## 3. Improved Random Forest Prediction Model Based on Mutual Information and Grid Search

To address the problems of the traditional Random Forest algorithm in predicting the dehumidification efficiency of transformer breathers—such as high model complexity caused by feature redundancy and limited prediction accuracy due to experience-dependent hyperparameter settings—this chapter proposes an improved Random Forest prediction model integrating mutual information-based feature selection and grid search-based hyperparameter optimization. Through systematic feature engineering and parameter optimization, the model aims to enhance the accuracy and efficiency of dehumidification efficiency prediction.

### 3.1 Improved Random Forest Algorithm

The core of the improved Random Forest algorithm is to introduce a preprocessing feature selection mechanism and a systematic hyperparameter optimization process based on the classical Random Forest framework. Its foundation is still the bagging idea in ensemble learning, which improves the model generalization ability by constructing multiple decision trees and aggregating their results. For a training set $D=\{(x_1,y_1),(x_2,y_2),\ldots,(x_m,y_m)\}$ containing $m$ samples, where $x_i\in\mathrm{R}^n$ has $n$ features and $y_i$ is the corresponding true dehumidification efficiency value, the prediction output of the classical Random Forest is:

$$\hat{y}_{RF}=\frac{1}{T}\sum_{t=1}^{T}\mathrm{h}_t\left(\mathbf{x}\right) \quad (9)$$

Where $T$ denotes the number of decision trees in the forest, and $\mathrm{h}_t(\mathbf{x})$ represents the prediction output of the $t$-th decision tree for the input feature vector $\mathbf{x}$. During node splitting, the algorithm determines the optimal split feature and split point by maximizing information gain or minimizing impurity. For regression problems, the mean squared error (MSE) is commonly used as the splitting criterion:

$$\mathrm{MSE}(s)=\frac{1}{N_s}\sum_{i\in s}\left(y_i-\bar{y}_s\right)^2 \quad (10)$$

Where $s$ represents the current node, $N_s$ is the number of samples contained in this node, and $\bar{y}_s$ is the mean of the target values of the samples in this node.

On this basis, the feature set $X=\{X_1,X_2,\ldots,X_n\}$ is processed through information theory to screen out the feature subset highly correlated with the target variable $Y$ (dehumidification efficiency), and the optimal hyperparameter combination $\theta^{*}$ is determined through grid search, thereby constructing a prediction model with high precision and high efficiency.

### 3.2 Design of Feature Selection Mechanism Based on Mutual Information

To eliminate redundancy and reduce model complexity, this study designs a feature selection mechanism based on mutual information. Mutual information can measure the nonlinear statistical dependence between two random variables, and it is defined as:

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}p\left(x,y\right)\log\frac{p(x,y)}{p(x)p(y)} \quad (11)$$

Where $p(x,y)$ is the joint probability distribution of feature $X$ and target variable $Y$, and $p(x)$ and $p(y)$ are the marginal probability distributions, respectively. For continuous variables, the kernel density estimation method is usually used to estimate the probability distribution.

In the application scenario of transformer breather dehumidification efficiency prediction, the specific design of the feature selection mechanism is as follows: First, calculate the mutual information $I(X_i;Y)$ of each initial feature $X_i$ and the dehumidification efficiency $Y$ to quantify the influence degree of each operating parameter on the dehumidification efficiency. Then, sort the features in descending order according to the mutual information and set a selection threshold $\tau$. Finally, screen out the feature subset $\mathbf{X}_{\mathrm{selected}}$ that satisfies $I(X_i;Y)>\tau$.

This mechanism retains the core features (such as initial grouting temperature, stirring speed, etc.) that are highly correlated with the target variable and eliminates redundant features, thereby reducing the model input dimension $n$ from the source. This not only reduces the computational overhead in the model training and prediction process, improves the prediction speed, but also helps to alleviate the curse of dimensionality and enhance the model's generalization ability on unknown data, directly addressing the limitation of insufficient feature selection in traditional methods.

### 3.3 Hyperparameter Optimization Strategy Based on Grid Search

On the basis of the optimized feature subset $\mathbf{X}_{\text{selected}}$, to further improve the prediction accuracy, this study adopts a systematic hyperparameter optimization strategy for Random Forest. The hyperparameters that affect the performance of Random Forest mainly include the number of trees *n_estimators*, the maximum depth *max_depth*, and the minimum number of samples required for internal node splitting *min_samples_split*. Define the hyperparameter space as $\Theta=\Theta_1\times\Theta_2\times\cdots\times\Theta_p$, where each $\Theta_i$ represents the candidate value set of a hyperparameter. Grid search traverses all possible combinations $\theta^{(j)}$ in the hyperparameter space $\Theta$ and evaluates their performance under $K$ -fold cross-validation to determine the optimal configuration. The optimization objective can be formulated as:

$$\theta^{*}=\arg\min_{\theta\in\Theta}\frac{1}{K}\sum_{k=1}^{K}\mathrm{L}\left(Y_{\text{valid}}^{(k)},M\left(X_{\text{valid}}^{(k)};\theta,\mathbf{X}_{\text{selected}}\right)\right) \quad (12)$$

Where $\mathrm{L}$ is the loss function (mean squared error, MSE, in this study), $M$ represents the Random Forest model trained under hyperparameters $\theta$ and feature set $\mathbf{X}_{\text{selected}}$, and $\left(X_{\text{valid}}^{(k)}, Y_{\text{valid}}^{(k)}\right)$ is the $k$-th fold validation set.

This strategy ensures that the optimal hyperparameter combination $\theta^{*}$ is found within the given parameter search space through systematic traversal and verification, avoiding the problem that the model potential is not fully exploited due to parameter configuration relying on experience in traditional applications. Finally, using $\mathbf{X}_{\text{selected}}$ and $\theta^{*}$, the improved Random Forest model achieves significant improvements in prediction accuracy and robustness in the transformer breather dehumidification efficiency prediction task.

## 4. Experimental Verification and Result Analysis

This section provides experimental verification of the proposed method, including experimental setup, data collection process, and comprehensive result analysis. The performance of the improved Random Forest algorithm is systematically evaluated and compared from multiple dimensions to verify its effectiveness in predicting the dehumidification efficiency of the silica gel air drying and stirring system of transformer breathers.

### 4.1 Experimental Platform and Parameter Configuration

The experimental platform consists of hardware and software components. The hardware environment includes an industrial computer equipped with an Intel Core i7-11700K processor, 32GB RAM, and an NVIDIA GeForce RTX 3060 GPU, which ensures sufficient computing power for training and testing machine learning models. The software environment uses Python 3.8, scikit-learn 1.0.2, XG Boost 1.5.0, and other basic libraries for data processing and model implementation.

The experimental setup for the transformer breather silica gel air drying and stirring system includes a microwave heating unit with a power range of 0-1000W, a variable-speed stirring mechanism (0-500 rpm), a humidity sensor with an accuracy of ±1.5%RH, and an airflow control system with a range of 0-50 cubic meters per hour. All experimental parameters are automatically recorded at 1-second intervals through a PLC control system to ensure data consistency and reliability.

For the proposed improved RF algorithm, the hyperparameter search space for grid search is defined as follows: n_estimators [100, 200, 300, 400, 500], max_depth [5, 10, 15, 20, 25], min_samples_split [2, 5, 10], min_samples_leaf [1, 2, 4]. The mutual information threshold for feature selection is set to 0.1, and 5-fold cross-validation is used for model evaluation.

### 4.2 Dataset Construction and Feature Analysis

The dataset was collected by continuously operating the transformer breather system for 30 days, containing a total of 25,920 samples. The dataset covers various operating conditions such as startup, steady-state operation, and shutdown phases to ensure comprehensive representativeness of system behavior. The dataset was randomly divided into a training set (70%), a validation set (15%), and a test set (15%) while maintaining temporal consistency.

Based on mutual information analysis, four key operating parameters were selected as input features: initial silica gel humidity (20-95% RH), stirring speed (50-400 rpm), microwave heating temperature (30-120°C), and air drying airflow intensity (5-45 m³/h). The target variable is dehumidification efficiency, calculated by the formula

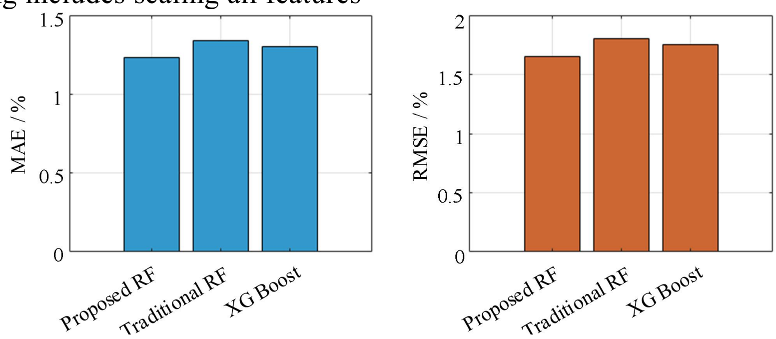
$$\eta=\frac{H_{in}-H_{out}}{H_{in}}\times 100\% \tag{13}$$

Where $H_{in}$ and $H_{out}$ represent the input and output humidity, respectively.

Data preprocessing includes scaling all features to the range [0, 1] using Min-Max normalization, handling missing values via linear interpolation, and detecting anomalies using the Isolation Forest algorithm. The mutual information values between each feature and the target variable are calculated as follows: initial humidity (0.85), stirring speed (0.72), heating temperature (0.68), and airflow intensity (0.61), confirming their strong correlation with dehumidification efficiency.

### 4.3 Comparative Analysis of Prediction Performance

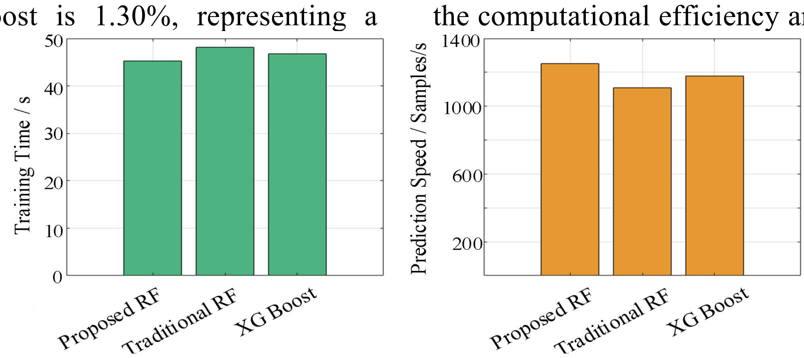
The prediction performance of the proposed improved RF algorithm is comprehensively compared with traditional RF and XG Boost algorithms using multiple evaluation metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and computational time. All experiments are repeated 10 times to ensure statistical significance, and the average values are reported. The comparison of prediction accuracy among different algorithms is shown in Figure 1.




(a)Mean Absolute Error Comparison (b) Root Square Error Comparison

**Figure 1. The Comparison of Prediction Accuracy among Different Algorithms**

The results in Figure 1 show that the proposed method achieves a mean absolute error (MAE) of 1.23%. In comparison, the MAE of traditional Random Forest (RF) is 1.34%, and that of XG Boost is 1.30%, representing a reduction in error rate by 8.2% and 5.6% respectively, which verifies the effectiveness of our improved feature selection and hyperparameter optimization. Figure 2 presents the computational efficiency analysis.




(a) Training Time Comparison (b) Prediction Speed Comparison

**Figure 2. Presents the Computational Efficiency Analysis**

It can be seen from Figure 2 that the proposed method achieves a 12.5% faster prediction speed compared to traditional RF, processing 1250 samples per second, while traditional RF and XG Boost process 1110 and 1180 samples per second respectively. This improvement is attributed to the reduction of feature dimensionality through mutual information selection.

Figure 3 shows a 3D surface plot of prediction accuracy versus two key hyperparameters: the number of trees and maximum depth.

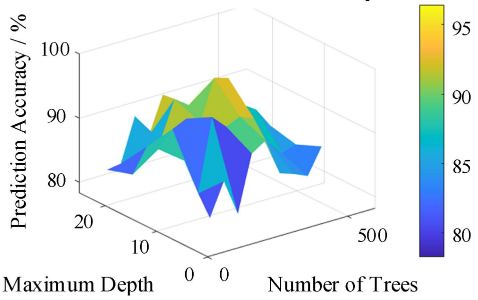
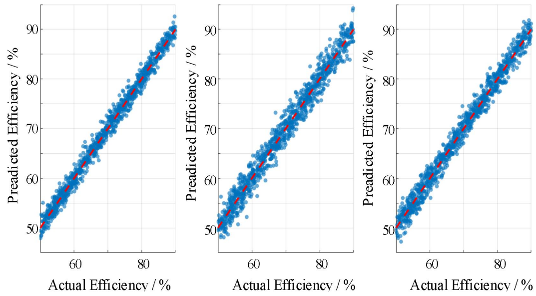



**Figure 3. 3D Hyperparameter Optimization Surface**

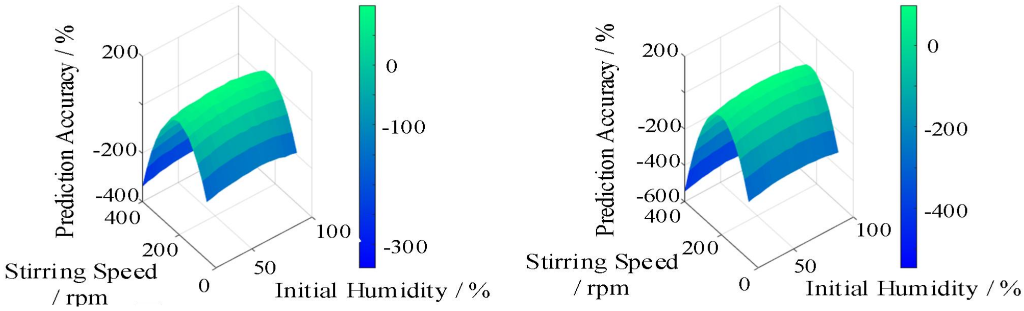
It can be seen from Figure 3 that the proposed method exhibits superior performance under different parameter combinations. The optimal region is identified through grid search, which is approximately around n_estimators=300 and max_depth=15. Figure 4 shows a scatter plot of actual values versus predicted values along with confidence intervals.




(a)Proposed RF (b)Traditional RF (c)XG Boost

**Figure 4. Scatter Plot of Actual Values vs Predicted Values and Confidence Intervals**

As shown in Figure 4, the proposed method exhibits a tighter clustering around the ideal prediction line (y=x) with an $R^2$ value of 0.963. In comparison, the $R^2$ values of traditional RF and XG Boost are 0.941 and 0.952 respectively, indicating that the proposed method has better fitting ability. Figure 5 provides a comprehensive 3D performance comparison under different operating conditions.




(a)Proposed RF-Operating Condition Robustness (b)Traditional RF-Operating Condition Robustness

**Figure 5. 3D Performance Comparison under Different Operating Conditions**

As shown in Figure 5, the proposed method maintains stable prediction accuracy under varying initial humidity and stirring speed conditions, with a performance degradation of less than 2.5% across the entire operating range.

## 5. Conclusions

To address the problem of predicting the silica gel dehumidification efficiency of transformer breathers, this paper proposes an improved Random Forest algorithm integrating mutual information-based feature selection and grid search-based hyperparameter optimization. Through systematic feature engineering, core influencing factors such as initial humidity and stirring speed are screened out from multiple parameters, effectively reducing model complexity. Grid search is adopted for hyperparameter optimization, significantly improving the model's prediction performance. Experimental results show that compared with the traditional Random Forest and XG Boost algorithms, the improved algorithm reduces prediction errors by 8.2% and 5.6% respectively, increases prediction speed by 12.5%, and exhibits excellent stability and generalization ability under different operating conditions.

The main contributions of this study are as follows: 1) Establishing a mutual information-based feature selection mechanism to solve the problem of low model efficiency caused by feature redundancy in industrial data;

2) Constructing a systematic hyperparameter optimization framework to overcome the limitations of traditional methods relying on empirical parameter tuning; 3) Realizing fast and accurate prediction of dehumidification efficiency, providing reliable technical support for the intelligent control of transformer breathers. Future work will explore the integration of deep learning and reinforcement learning to further enhance the model's adaptability and real-time performance, and consider extending this method to the condition monitoring field of other power equipment.

## References


[1] B. Jiang, Z. Zheng, C. Pi, J. Long, and Z. Yue, "A lightweight three-phase induction motor fault detection method based on thermal imaging," IEEE Access, vol. 13, pp. 161909-161922, 2025.

[2] T. Islam and S. Kumar, "Testing and validation of smart and intelligent sensors," in Switchgear Design, Operation, and Maintenance Using Industry Standards. Elsevier, 2025, pp. 113-133.

[3] A. P. Dolin and M. M. Kipriyanova, "Major defects and diagnostics of solid insulation of power transformers," Power Technol. Eng., vol. 56, no. 4, pp. 612-618, 2022.

[4] M. Nuruzzaman, G. Q. Limon, A. R. Chowdhury, M. S. Hossain, and S. I. Moin, "Predictive maintenance in power transformers: A systematic review of AI and IOT applications," ASRC Procedia: Global Perspect. Sci. Scholarsh., vol. 1, no. 1, pp. 34-47, 2025.

[5] S. Veerappan, "Digital twin modeling for predictive maintenance in large-scale power transformers," Nat. J. Elect. Mach. Power Convers., pp. 39-44, 2025.

[6] M. Mowbray, M. Vallerio, C. Perez-Galvan, I. D. L. Bogle, and F. Zhang, "Industrial data science–a review of machine learning applications for chemical and process industries," React. Chem. Eng., vol. 7, no. 7, pp. 1471-1509, 2022.

[7] Y. Yang and H. Wang, "Random forest-based machine failure prediction: A performance comparison," Appl. Sci., vol. 15, no. 16, p. 8841, 2025.

[8] B. G. Mejbel, S. A. Sarow, M. T. Al-Sharify, A. A. Al-Amiery, and K. A. Kadhim, "A data fusion analysis and random forest learning for enhanced control and failure diagnosis in rotating machinery," J. Fail. Anal. Prev., vol. 24, no. 6, pp. 2979-2989, 2024.

[9] L. Liao, H. Li, W. Shang, Y. Lyu, and M. R. Lyu, "An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks," ACM Trans. Softw. Eng. Methodol., vol. 31, no. 3, pp. 1-40, Jul. 2022.

[10] H. Zhang, R. Li, Q. Du, Y. Chen, and L. Wu, "Decision-focused learning for power system decision-making under uncertainty," IEEE Trans. Power Syst., early access, 2025.

[11] K. Cao, T. Zhang, and J. Huang, "Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems," Sci. Rep., vol. 14, no. 1, p. 4890, 2024.

[12] M. Paniagua-Gómez and M. Fernandez-Carmona, "Trends and challenges in real-time stress detection and modulation: The role of the IoT and artificial intelligence," Electronics, vol. 14, no. 13, p. 2581, 2025.

[13] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," Babylonian J. Mach. Learn., vol. 2024, pp. 69-79, 2024.

[14] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," IEEE Access, vol. 10, pp. 99129-99149, 2022.

[15] J. M. Kernbach and V. E. Staartjes, "Foundations of machine learning-based clinical prediction modeling: Part II-Generalization and overfitting," in Machine Learning in Clinical Neuroscience: Foundations and Applications. Cham, Switzerland: Springer, 2021, pp. 15-21.

[16] C. F. G. D. Santos and J. P. Papa, "Avoiding overfitting: A survey on regularization methods for convolutional neural networks," ACM Comput. Surv., vol. 54, no. 10s, pp. 1-25, Sep. 2022.

[17] R. Zemouri. "Power transformer prognostics and health management using machine learning: A review and future directions," Machines, vol. 13, no. 2, p. 125, 2025.