FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

Evaluation of Random Forest Algorithm Performance in Predicting the Flashover Voltage of Polluted Insulators

Rama Alkhtiar[1*], Professor Jamal Alnasseir[2], and Professor George Isber[3]

[1]PhD Candidate, Department of Electrical Engineering, University of Latakia, Syria; Email: rama.alkhtiar@latakia-univ.edu.sy
[2]Professor, Department of Electrical Power, Faculty of Mechanical and Electrical Engineering, Damascus University, Damascus, Syria; Email: jamal.nassier@damascusuniversity.edu.sy
[3]Professor, Department of Electrical Power, Faculty of Mechanical and Electrical Engineering, Latakia University, Latakia, Syria; Email: George.Isber@latakia-univ.edu.sy

***Correspondence:** *Rama Alkhtiar, rama.alkhtiar@latakia-univ.edu.sy*

ABSTRACT- This study aims to evaluate the capability of the Random Forest model to predict the flashover voltage of polluted insulators, with particular emphasis on the effect of hyperparameter tuning strategies on model accuracy and stability. A two-stage methodology was adopted. In the first stage, Grid Search and Particle Swarm Optimization were compared for tuning the model hyperparameters using a published dataset of cap-and-pin insulators. The results showed close agreement between the two methods in terms of the mean root mean square error, with a slight accuracy advantage for Particle Swarm Optimization, whereas Grid Search provided higher stability and greater computational simplicity. Accordingly, the Grid Search–tuned Random Forest model was adopted in the second stage, where its performance was evaluated after merging the published data with local laboratory measurements obtained at the High Voltage Laboratory of Damascus University, Syria. The model demonstrated high predictive performance under 70/30 and 80/20 training/testing splits. In addition, ten-fold cross-validation confirmed the stability of the model performance across different data partitions. The feature-importance analysis revealed that surface conductivity was the most influential factor affecting flashover voltage, followed by the geometrical characteristics of the insulator. The results confirm that the Grid Search–tuned Random Forest model provides an effective and initially generalizable tool for predicting the flashover voltage of polluted insulators. However, further expansion of the laboratory database is recommended to improve the reliability of practical applications.

Keywords: Polluted Insulators, Flashover Voltage, Random Forest, Grid Search, Particle Swarm Optimization, Surface Conductivity.

ARTICLE INFORMATION

Author(s): Rama Alkhtiar, Jamal Alnasseir, and George Isber;
Received: 10/03/26; **Accepted**: 01/06/26; **Published**: 30/06/26;
E- ISSN: 2347-470X;
Paper Id: IJEER 1003B08;
Citation: 10.37391/ijeer.140226
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-14/ijeer-140226.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Electrical insulators are essential components in power transmission and distribution systems, as they contribute to ensuring electrical insulation and the operational stability of power networks. However, the performance of these insulators is strongly affected by environmental factors, particularly the accumulation of pollutants and moisture. Wet pollution layers increase surface conductivity and leakage currents, thereby creating favorable conditions for the initiation and development of surface discharges, ultimately leading to electrical flashover. Accurate prediction of flashover voltage under polluted conditions is therefore an important issue for improving insulator design and enhancing the reliability of power systems operating in harsh environments [1–2].

Previous studies have confirmed that pollutant accumulation produces significant leakage currents on insulator surfaces, resulting in severe electrical discharges, i.e., flashover, which threaten the reliability of power systems [3]. Therefore, accurate prediction of flashover voltage represents a major challenge for improving insulator design and ensuring network stability under severe operating conditions [1, 3].

Considerable research efforts have been devoted to understanding the behavior of polluted insulators using different approaches. From a mathematical perspective, several models based on the Obenaus model have been developed to estimate the critical flashover voltage as a function of the insulator geometrical characteristics (L, D_m), pollution severity C, and arc constants (A, n) [4]. However, the predictions of these models may vary due to differences in the constants associated with each insulator type [5]. Experimental studies have shown a close relationship between flashover voltage and different humidity and temperature levels, where increased humidity leads to a reduction in the flashover voltage of polluted insulators, while elevated temperature may reduce the dielectric strength of insulators and increase the risk of electrical flashover [6]. In parallel, finite element method (FEM) simulations have revealed that the presence of a pollution layer causes sharp variations in the electric field distribution, which contributes to validating

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

experimental findings [7].

Despite these efforts, experimental investigations aimed at determining the critical flashover voltage are time-consuming and costly [5], whereas simulations always require experimental validation to ensure accuracy [7].

To overcome the limitations of traditional models in capturing the complex nonlinear relationships between insulator geometrical characteristics and the environmental factors affecting flashover voltage, artificial intelligence techniques have emerged as effective tools with strong modeling and prediction capabilities [1–3].

Artificial neural networks (ANNs) represented the starting point in this field, as they were successfully applied to estimate flashover voltage using the geometrical characteristics of insulators as input variables [8]. These applications were later extended to the simultaneous prediction of both flashover voltage and equivalent salt deposit density (ESDD) [9]. Adaptive algorithms implemented using advanced programming languages were also developed to improve ANN performance [10], and multilayer neural network architectures were adopted to model complex relationships with higher accuracy [11, 12].

In a related context, the adaptive neuro-fuzzy inference system (ANFIS), which combines the advantages of neural networks and fuzzy logic, was applied and demonstrated advanced predictive capability using different membership functions [13]. A subsequent study proved the superiority of ANFIS over both conventional neural networks and fuzzy logic in estimating flashover voltage [14]. Fuzzy logic (FL) was also used to estimate flashover voltage with confidence intervals, showing effectiveness in handling uncertainties associated with this complex physical phenomenon [15, 16].

With the increasing use of machine learning techniques, support vector machines (SVMs) emerged as an effective tool for predicting flashover voltage, demonstrating their ability to provide better predictive results compared with some other machine learning tools [17]. The least-squares support vector machine (LS-SVM) formulation was developed as a simplified version, characterized by requiring only the solution of a set of linear equations, which makes it computationally more efficient for modeling critical flashover voltage [17, 18]. LS-SVM showed high predictive efficiency, outperforming conventional neural networks in several applications [18], and the effect of different kernel functions on its performance was also investigated [19]. In a later development, an improved LS-SVM methodology based on a fixed set of support vectors selected according to the quadratic Renyi criterion was proposed, contributing to enhanced model generalization and improved prediction accuracy of flashover voltage under polluted conditions [20]. LS-SVM was also applied under various environmental conditions, such as dry and rainy conditions, clearly demonstrating the effect of humidity in reducing flashover voltage values [21].

These basic models paved the way for the development of more advanced hybrid models that combine artificial intelligence techniques with intelligent optimization algorithms. Previous studies have demonstrated the success of such hybridization strategies, such as combining neural networks with genetic algorithms [22] or integrating whale optimization and particle swarm optimization algorithms [5], which contributes to improving prediction accuracy and accelerating model convergence compared with conventional models [5, 16, 22].

Hybrid models combining artificial intelligence techniques with Particle Swarm Optimization (PSO) have also been developed to improve the prediction accuracy of flashover voltage. For instance, the LS-SVM-PSO model was applied to optimize hyperparameters and achieved higher prediction accuracy compared with non-optimized models [23]. Improved versions focusing on adaptive parameter selection were subsequently proposed [24], while PSO was also used to directly determine the arc constants in the mathematical model [25]. In more recent studies, the limitations of conventional artificial neural networks, including their tendency to become trapped in local minima, motivated the development of ANN-PSO models. One such model outperformed LS-SVM-PSO and LS-SVM-GWO [26], achieving an R^2 value of 0.997 [27]. Another recent study reported very high accuracy, with R^2 = 0.999, for different insulator profiles under dry and rainy conditions [1]. Despite the clear progress in applying artificial intelligence techniques to flashover voltage prediction, recent studies indicate the importance of developing hybrid models or ensemble learning models capable of improving prediction accuracy and generalization ability in polluted-insulator problems [27]. Nevertheless, most previous studies have focused on neural networks, support vector machines, and their optimized variants, with a common reliance on a single dataset split for performance evaluation [23, 27]. This approach may lead to an optimistic estimate of the model's generalization ability, particularly when dealing with limited-size datasets. Moreover, ensemble learning models, especially Random Forest, have not received sufficient attention in modeling the flashover voltage of polluted insulators, despite their recent use in predicting flashover characteristics of insulators under lightning impulse conditions [28, 29].

In a recent conference contribution by the present authors, the XGBoost algorithm was applied to the same problem without hyperparameter optimization using PSO, achieving R^2 = 0.9945 [31]. Building on that preliminary work, the present study investigates the Random Forest algorithm with a more rigorous evaluation framework and an expanded dataset that includes local laboratory measurements.

Accordingly, this study proposes a systematic framework for evaluating the Random Forest model in predicting the flashover voltage of polluted insulators. This is achieved by comparing Grid Search and Particle Swarm Optimization strategies for hyperparameter tuning, analyzing performance stability across multiple random data splits, and then testing the model's generalization capability after integrating published data with local laboratory measurements.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

2. MATERIALS AND METHODS

2.1. Proposed Methodology

This study adopted a two-stage experimental–computational methodology. The first stage was devoted to selecting the most appropriate hyperparameter tuning strategy for the Random Forest model using a published dataset of polluted insulators. In the second stage, the selected model was evaluated after expanding the dataset by integrating the published data with local laboratory measurements. This stage aimed to assess the model's prediction accuracy, performance stability, and generalization capability.

2.2. Published Dataset

In the first stage, a published dataset of cap-and-pin polluted insulators was used. This dataset has previously been employed in studies related to the modeling of flashover voltage in polluted insulators. The input variables included the geometrical characteristics of the insulator, namely diameter *D*, height *H*, leakage distance *L*, and form factor *F*, in addition to a variable representing the severity of surface pollution. The critical flashover voltage (FOV) was considered as the output variable. This dataset was used to evaluate the sensitivity of the Random Forest model to variations in training/testing data splits and to compare two hyperparameter tuning strategies, namely Grid Search and Particle Swarm Optimization (PSO).

The dataset was constructed using the same sources and methodologies adopted in previous studies [14, 27]. It included both experimental measurements and data generated by applying the mathematical model. Experimentally, the tests were conducted according to international standards. Artificial pollution was applied using a mixture of silica powder, kaolin clay, and sodium chloride suspended in isopropyl alcohol for 30 minutes, after which the insulators were left to dry for one hour. The equivalent salt deposit density (ESDD) on the insulator surface was then determined [27]. The calculated data were generated based on the equivalent circuit mathematical model used to evaluate the critical flashover voltage [14]. This dataset was used only in the first stage for the systematic comparison of hyperparameter tuning strategies, and not for deriving the final model based on the integrated dataset.

2.2.1. Mathematical Model of the Polluted Insulator

The Obenaus model is a theoretical framework developed to examine flashover phenomena on polluted insulators. It conceptualizes the process by considering a partial arc that bridges a dry zone on the insulator surface, in combination with the resistive properties of the pollution layer. These two elements are commonly modeled as a series circuit [14]. The model is particularly significant in defining the critical flashover voltage c, which represents the threshold voltage at which the partial arc develops into a complete flashover. This critical voltage is determined using a formula that combines the physical and electrical characteristics of the insulator and the pollution layer. Therefore, the Obenaus model provides a structured basis for understanding and predicting the flashover performance of polluted insulators [14, 27]. At the critical flashover condition, the critical voltage Uc (V) is expressed as [27].

$$Uc = \frac{A}{n+1}(L + \pi . D_m . F . K . n)(\pi . D_{m.}\sigma_s . A)^{-n/(n+1)} \quad (1)$$

This equation involves several key parameters: D_m is the maximum diameter of the insulator disc (cm), *F* is the form factor representing the geometric profile of the insulator, *K* is the pollution layer resistance coefficient, and σ_s is the surface conductivity of the pollution layer (Ω^{-1}) [14, 27].

Studies have shown that the values of the arc constants *A* and *n* vary significantly across previous research [14], posing a major challenge in modeling the phenomenon. In this study, the values calculated using genetic algorithms, which yielded results consistent with experiments, were adopted *A*=124.8 and *n*=0.409 [27].

The form factor *F* is defined by the integral along the creepage path [14, 27]:

$$F = \int_0^l \frac{dl}{\pi D(l)} \quad (2)$$

where $D(l)$ is the insulator diameter varying along the creepage path. The surface conductivity σ_s is related to the equivalent salt deposit density *C* (mg/cm²) by the empirical relationship [27]:

$$\sigma_s = (369.05\ C + 0.42) \times 10^{-6} \quad (3)$$

For cap-and-pin insulators, the pollution layer resistance coefficient *K* is given by [14, 27]:

$$K = 1 + \frac{n+1}{2.\pi.F.n} ln\left(\frac{L}{2.\pi.R.F}\right) \quad (4)$$

where R is the radius of the arc foot (cm), calculated from [27]:

$$R = 0.469(\pi A\ D_m \sigma_s)^{1/(2(n+1))} \quad (5)$$

Experimental evidence has shown that the flashover voltage of polluted insulators is not constant even under identical conditions, due to the stochastic nature of arc phenomena, such as rib bridging or arc deviation from the insulator surface [14]. Therefore, a comprehensive training dataset containing a sufficient number of representative data points was prepared. The input variables included leakage distance *L*, height *H*, maximum diameter D_m, Form Factor F, and surface conductivity σ_s , while the output variable was the flashover voltage FOV [27]. The flashover voltage was calculated using *equation (1)* based on the following values of C: 0.02, 0.03, 0.04, 0.05, 0.06, 0.13, 0.16, 0.23, 0.28, 0.34, 0.37, 0.49, 0.52, and 0.55 mg/cm² [14, 27]. The C values were converted into σ_s using *equation (3)*. *Tables A1* and *A2* summarize the geometrical characteristics of the insulators used in the mathematical model and the experimental data.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

2.3. Integrated Dataset and Local Laboratory Measurements

In the second stage, the most stable optimization strategy identified in the first scenario was used to evaluate the model performance after integrating data from previous studies [14, 27] with local laboratory measurements conducted at Damascus University. This integration aimed to assess the generalization capability of the model by increasing the diversity of data sources. The Random Forest Regressor was adopted to predict the flashover voltage (FOV) using the integrated dataset.

2.3.1. Laboratory Description and Experimental Setup

Laboratory tests were carried out at the High Voltage Laboratory of Damascus University using a test transformer with a maximum voltage of 100 kV. The experiments were performed by exposing the insulator surface to saline solutions with different conductivities, followed by the application of an alternating voltage that was gradually increased until flashover occurred. Temperature and relative humidity were recorded during the tests to characterize the accompanying environmental conditions. This experimental setup was designed to represent the behavior of insulators under different wet pollution conditions and to provide local data that could be integrated with the published dataset to improve the diversity of the training database.

The experimental setup used in the laboratory tests is shown in *figure 1*.

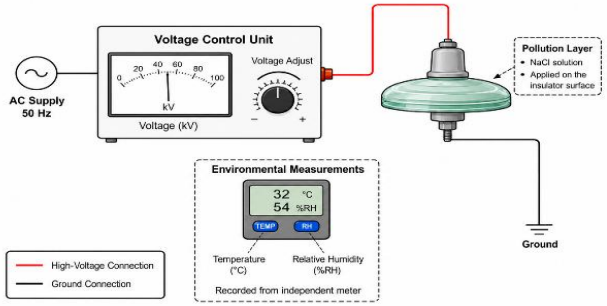

Figure 1. Experimental setup in the High Voltage Laboratory of Damascus University

2.3.2. Tested Insulators and Measurement Variables

The experiments were conducted on single insulator discs in order to represent the basic unit of insulator strings. This approach allowed the physical behavior of the insulator to be studied more accurately, without the additional complexities associated with voltage distribution along complete insulator strings. The tested specimens included a cap-and-pin glass insulator. The tested glass insulator is shown in *figure 2*.

Figure 2. Tested glass insulator

Table 1 presents the main geometrical characteristics of the tested insulator, including leakage distance, insulator diameter, height, and form factor.

Table 1. Geometrical characteristics of the locally tested insulator

Insulator type	Leakage distance[cm]	Diameter [cm]	Height[cm]	Form factor
cap-and-pin	23	20	8	0.36

2.3.3. Construction of the Local Dataset

Pollution tests were conducted using a sodium chloride (NaCl) solution. Before each test, the insulator was carefully cleaned to ensure consistent measurement conditions. Pollution was then applied non-uniformly to the insulator surface in order to simulate the realistic distribution of contaminants on insulators under operating environments. Measurements were performed over a wide range of surface conductivity values, extending from 5.5 to 1300 µS, allowing different degrees of pollution severity to be represented.

These tests produced 23 polluted samples for the glass insulator. All samples were organized into a local dataset that included, for each sample, the geometrical characteristics of the insulator, temperature, relative humidity, and surface conductivity, while the breakdown voltage was considered the target variable to be predicted.

Table 2 shows the distribution of pollution tests according to the insulator type and the environmental condition represented by temperature and relative humidity, along with the number of tests performed in each condition.

Table 2. Pollution test conditions for the tested insulator

Insulator type	Temperature (°C)	Relative humidity (%)	Number of pollution tests	Conductivity range (µS)
Glass insulator	32	54	23	5.5–1300

The detailed values of the local laboratory dataset are provided in *table A3*.

2.4. Random Forest Algorithm and Hyperparameter Tuning

Random Forest belongs to the family of ensemble learning models and was first introduced by Breiman in 2001. The algorithm is based on constructing multiple decision trees using subsets of the training data and then aggregating their outputs to obtain the final prediction [30]. In regression problems, the output is typically computed as the average of the predictions produced by the individual trees, which helps reduce variance and improve the model's generalization capability. Random Forest models have recently been used in applications related to predicting flashover characteristics of electrical insulators, supporting their suitability for this type of nonlinear regression problem [28, 29].

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

To improve the model performance, the hyperparameters of the RF algorithm were tuned in the first stage using two methods: Grid Search and Particle Swarm Optimization (PSO). Grid Search is based on testing all predefined combinations within the search space and selecting the combination that achieves the lowest error according to the adopted performance criterion. In contrast, PSO is a metaheuristic optimization algorithm in which each potential solution is represented by a particle moving within the search space. The particle positions are updated based on the best individual solution and the best global solution obtained by the swarm.

PSO has been used in previous studies to tune the parameters of artificial intelligence models and improve the prediction accuracy of flashover voltage in polluted insulators [23, 26, 27]. In this study, the particles represented different combinations of RF hyperparameters, and RMSE was used as the objective function to be minimized in order to select the optimal combination. After comparing the two methods in terms of accuracy and stability, the more balanced strategy was adopted for evaluating the model on the integrated dataset in the second stage.

2.5. Evaluation Strategies

Model performance was evaluated using the coefficient of determination (R^2), which measures the proportion of variance explained by the model, and the root mean square error (RMSE), which quantifies the average prediction error. These performance metrics were adopted following the methodology reported in [14]. This study adopted a two-stage evaluation framework. In the first stage, RF-Grid Search and RF-PSO were compared across ten independent random splits with a 70%/30% training/testing ratio. This procedure aimed to evaluate the stability of each tuning strategy with respect to variations in the distribution of training and testing data.

In the second stage, the selected RF-Grid Search model was evaluated on the integrated dataset using three complementary procedures. The first and second procedures involved fixed training/testing splits of 70%/30% and 80%/20%, respectively, where the larger portion of the data was used for training and the remaining portion was reserved for independent testing. The third procedure consisted of applying 10-fold cross-validation to the entire integrated dataset, which consisted of 191 samples, using the same optimal hyperparameters. In this procedure, each fold was used once for testing, while the remaining nine folds were used for training. The mean and standard deviation of R^2 and RMSE across the ten folds were then calculated to quantitatively express the stability of the model.

Hyperparameter tuning was performed independently in each stage. In the first stage, the search space reported in Table 3 was used to compare RF-Grid Search and RF-PSO on the published dataset. In the second stage, after integrating the published and local laboratory datasets, a new GridSearchCV procedure was applied using an initial 70% training subset with five-fold internal cross-validation. The search space included n_estimators = {100, 200, 300}, max_depth = {None, 5, 10, 20}, and min_samples_split = {2, 5, 10}. The optimal hyperparameters obtained in this stage were then fixed and used in all subsequent evaluation experiments.

2.6. Computational Environment and Implementation Details

The computational implementation was carried out in Google Colaboratory using Python 3.12.13. The main libraries used were NumPy 2.0.2, Pandas 2.2.2, scikit-learn 1.6.1, Matplotlib 3.10.0, and SciPy 1.16.3. Random Forest modeling and GridSearchCV-based hyperparameter tuning were implemented using scikit-learn. All computations were performed on a CPU-based Google Colab environment with an Intel Xeon processor and 12.7 GB RAM.

3. RESULTS AND DISCUSSION

3.1. First-Stage Results: Comparison of Hyperparameter Tuning Strategies

Table 3 presents the search space used for tuning the hyperparameters of the Random Forest model, together with the optimal values obtained using Grid Search and Particle Swarm Optimization.

Table 3. First-stage search space and optimal hyperparameters for RF-Grid Search and RF-PSO

Method	Optimized RF parameter	Search range	Best value
RF + Grid Search	n_estimators	50, 100, 200	200
	max_depth	5, 10, 15	10
	min_samples_split	2, 5	2
	min_samples_leaf	1, 3	1
RF + PSO	n_estimators	50–200	194
	max_depth	3–15	11
	min_samples_split	2–10	2
	min_samples_leaf	1–5	1

As shown in *table 3*, both tuning strategies selected a relatively large number of trees and a moderate tree depth, while the minimum values of min_samples_split and min_samples_leaf were preferred in both cases. This indicates a clear similarity in the optimal model structure obtained by the two approaches.

Figure 3 compares the RMSE values of the RF-Grid Search and RF-PSO models across ten random 70/30 training/testing splits.

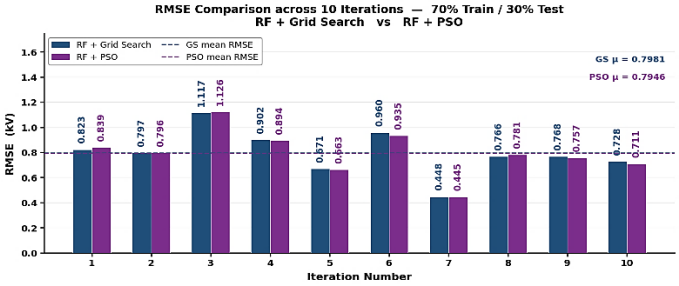

Figure 3. Comparison of RMSE values between RF-Grid Search and RF-PSO across ten random data splits

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

The results shown in *figure 3* indicate a close agreement between the two optimization strategies. The mean RMSE was approximately 0.7981 kV for RF-Grid Search and 0.7946 kV for RF-PSO. Although PSO achieved a slightly lower mean RMSE, the improvement was very limited. In contrast, RF-Grid Search showed slightly higher stability, with a standard deviation of 0.1688 kV, compared with 0.1702 kV for RF-PSO. Considering the marginal difference in prediction accuracy, together with the computational simplicity and reproducibility of Grid Search, RF-Grid Search was adopted for the second stage of the study.

3.2. Second-Stage Results: Evaluation on the Integrated Dataset

The Grid Search procedure applied to the integrated dataset yielded the following optimal hyperparameters: n_estimators = 300, max_depth = 10, and min_samples_split = 2.

3.2.1. Model Performance under Fixed Training/Testing Splits

3.2.1.1 Training/testing split of 70/30

Figure 4 shows the predicted and actual flashover voltage values obtained using the 70/30 training/testing split.

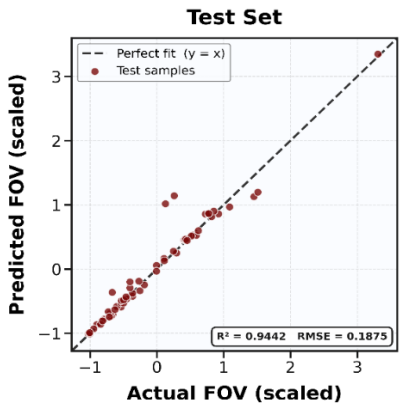

Figure 4. Actual and predicted flashover voltage values for the 70/30 split

3.2.1.2. Training/testing split of 80/20

Figure 5 shows the predicted and actual flashover voltage values obtained using the 80/20 training/testing split.

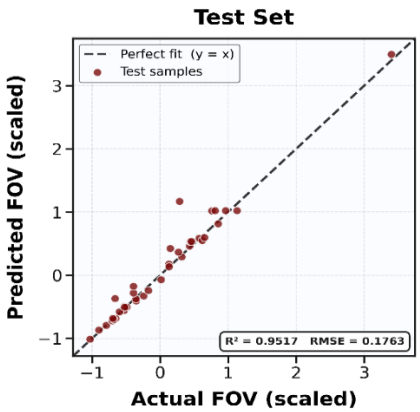

Figure 5. Actual and predicted flashover voltage values for the 80/20 split

The RF-Grid Search model demonstrated high predictive performance under both fixed data-splitting scenarios. For the 70/30 split, the test coefficient of determination was R^2 = 0.9442, while the RMSE was 0.1875. When the 80/20 split was used, the test performance improved, with R^2 = 0.9517 and RMSE = 0.1763. This improvement indicates that increasing the training data size enhanced the model's ability to capture the nonlinear relationship between the geometrical characteristics, surface conductivity, and flashover voltage. Moreover, the difference between training and testing performance remained within an acceptable range, suggesting that the model did not suffer from severe overfitting.

3.2.2. Residual Error Analysis

Residual error analysis was performed for the RF-Grid Search model under the 80/20 split, as this split achieved the best testing performance among the fixed data-splitting scenarios. The residual errors versus the predicted flashover voltage values are shown in *figure 6*.

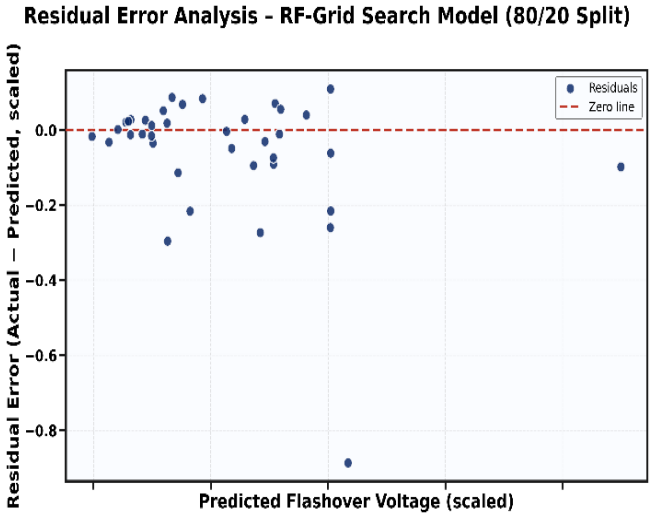

Figure 6. Residual errors versus predicted FOV under the 80/20 split

Figure 6 shows the residual error analysis of the RF-Grid Search model under the 80/20 split, where the residuals were calculated as the difference between the actual and predicted flashover voltage values. Most residuals are concentrated close to the zero-error line, indicating the absence of a clear systematic bias for most test samples. Some larger negative residuals, including one evident outlier, indicate overprediction for a limited number of samples. This behavior may be attributed to the limited size and heterogeneous nature of the integrated dataset. Overall, the residuals do not show a clear systematic pattern across the predicted range, supporting the reliability of the model within the limits of the available data.

3.2.3. Ten-Fold Cross-Validation Results

Figure 7 presents the ten-fold cross-validation results obtained using the complete integrated dataset.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

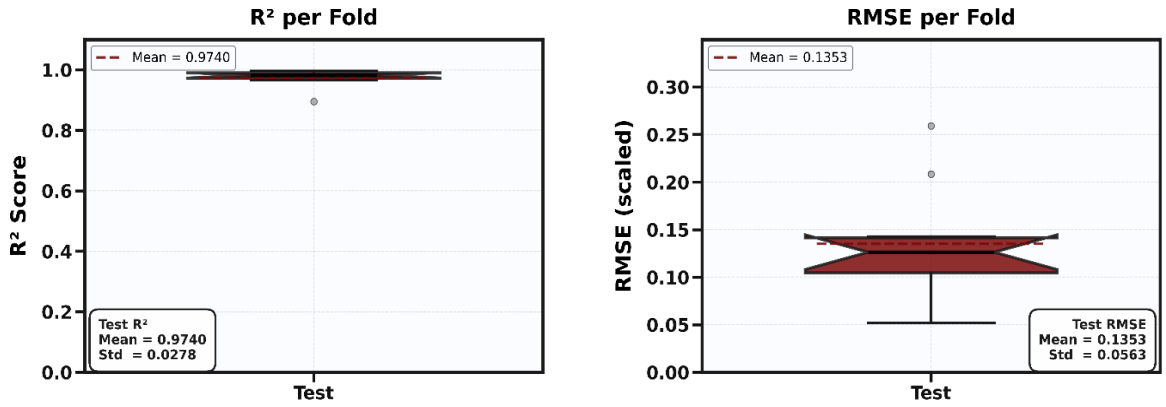

Figure 7. Ten-fold cross-validation results

The ten-fold cross-validation results showed an average coefficient of determination of $R^2 = 0.9740 \pm 0.0278$ and an average RMSE = 0.1353 ± 0.0563 across the test folds. These values confirm the stability of the model performance across different data partitions and support the generalization capability of the RF-Grid Search model within the limits of the available dataset. The lower RMSE observed in cross-validation, compared with some fixed splits, can be attributed to the fact that the model benefits from a larger portion of the dataset for training in each fold.

3.2.4. Feature-Importance Analysis

To interpret the behavior of the RF-Grid Search model and identify the most influential input variables in predicting flashover voltage, feature-importance analysis was performed using the model trained under the 80/20 split, which achieved the best testing performance among the two fixed splits. The relative importance of the input variables obtained from this analysis is shown in *figure 8*.

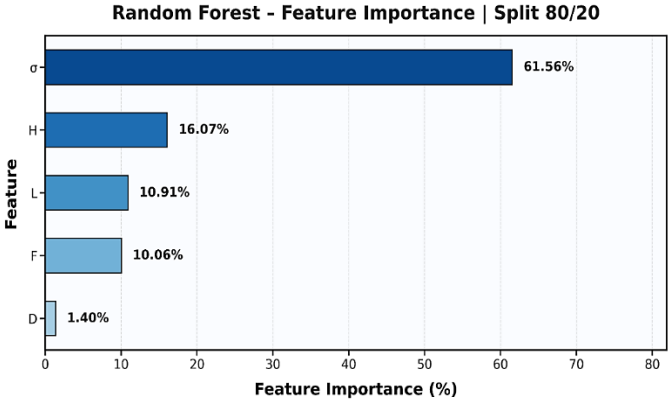

Figure 8. Relative importance of input variables in the RF-Grid Search model

Figure 8 shows that surface conductivity was the most influential variable in predicting flashover voltage, with a relative importance of 61.56%. This result is consistent with the physical interpretation of flashover in polluted insulators, since increasing surface conductivity reduces the resistance of the pollution layer and increases leakage current, thereby promoting the development of surface discharge.

The geometrical variables had lower but still meaningful contributions. The height of the insulator contributed 16.07%, the leakage distance 10.91%, and the form factor 10.06%, whereas the diameter had a relatively limited effect of 1.40%. These results indicate that surface pollution severity is the dominant factor in the studied dataset, while the geometrical characteristics contribute to improving the model's representation of the nonlinear relationship between insulator structure and flashover voltage.

3.2.5. Comparison with Previous Studies

A direct quantitative comparison between the performance indicators obtained in this study and those reported in previous works should be interpreted with caution because of differences in evaluation protocols. Most previous studies relied on a single fixed data split, which may lead to an optimistic estimation of model generalization, especially when limited datasets are used. In contrast, the present study adopted a more rigorous evaluation framework based on multiple random splits and ten-fold cross-validation.

Table 4 provides a reference comparison between the proposed Random Forest model and representative models reported in the literature.

Table 4. Reference comparison of model performance with previous studies

Model	R^2	Evaluation protocol	Reference
ANN-PSO	0.997	fixed split	[27]
ANFIS	0.989	Single fixed split	[14]
XGBoost (authors)	0.9945	fixed split	[31]
RF-Grid Search, present study	0.974 ± 0.028	10-fold cross-validation	—

The proposed RF-Grid Search model achieved a promising predictive performance, with an average coefficient of determination of 0.974 across the ten test folds.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

The accompanying standard deviation of 0.028 indicates stable performance across different data partitions. Therefore, the obtained results should not be interpreted as inferior to those of previous models, but rather as more conservative and statistically reliable due to the stricter evaluation procedure adopted in this study.

3.2.6. Study Limitations and Future Scope

Although the proposed RF-Grid Search model demonstrated stable and promising predictive performance, several limitations should be acknowledged. First, the integrated dataset is still relatively limited in size, particularly with respect to the local laboratory measurements. Second, the local experimental campaign was limited to one cap-and-pin glass insulator type due to the limited availability of different insulator geometries. Third, the local measurements were conducted under one temperature and relative humidity condition, which limits the ability to fully assess the influence of broader environmental variations. Therefore, future work should expand the experimental database to include different insulator materials, geometrical profiles, pollution severities, humidity levels, and temperature conditions. Further comparisons with other ensemble learning algorithms are also recommended to improve the robustness and practical applicability of flashover voltage prediction models.

4. CONCLUSIONS

This study concluded that the Grid Search–tuned Random Forest model represents an effective tool for predicting the flashover voltage of polluted insulators. The initial comparison between Grid Search and Particle Swarm Optimization showed a high degree of similarity in prediction accuracy, with a slight advantage for PSO in terms of mean RMSE, whereas Grid Search provided higher stability and greater computational simplicity. Accordingly, the RF-Grid Search model was adopted to evaluate performance on an integrated dataset combining published data with local laboratory measurements.

The model achieved high performance under fixed data-splitting scenarios. The 80/20 split produced the best testing performance, with a coefficient of determination of 0.9517 and an RMSE of 0.1763. In addition, ten-fold cross-validation confirmed the stability of the model across different data partitions, yielding an average coefficient of determination of 0.9740 ± 0.0278 and an average RMSE of 0.1353 ± 0.0563. Feature-importance analysis showed that surface conductivity was the most influential factor affecting flashover voltage, followed by the geometrical characteristics to varying degrees.

The main contribution of this study lies in combining the evaluation of hyperparameter-tuning stability with testing the model on an integrated dataset that includes local laboratory measurements. These findings support the potential of the RF-Grid Search model as a reliable data-driven approach for flashover voltage prediction in polluted insulators.

Appendix A

Table A1. Geometrical characteristics of the insulators used in the mathematical model.

D_m(cm)	*H (cm)*	*L (cm)*	*F*
26.8	15.9	33.0	0.79
26.8	15.9	40.6	0.86
25.4	16.5	43.2	0.90
25.4	14.6	31.8	0.72
29.2	15.9	47.0	0.92
27.9	15.6	36.8	0.76
32.1	17.8	54.6	0.96
28.0	17.0	37.0	0.80
25.4	14.5	30.5	0.74
20.0	16.5	40.0	1.29

Table A2. Geometrical characteristics of the insulators used in the experiment data.

Insulator Type	*L (cm)*	D_m *(cm)*	*H (cm)*	*F*	*C (mg/cm²)*	*Uc (kV)*
Type 1	27.9	25.4	14.6	0.68	0.13	12.0
Type 1	27.9	25.4	14.6	0.68	0.16	11.1
Type 1	27.9	25.4	14.6	0.68	0.23	8.10
Type 1	27.9	25.4	14.6	0.68	0.28	9.10
Type 1	27.9	25.4	14.6	0.68	0.34	7.50
Type 1	27.9	25.4	14.6	0.68	0.37	7.80
Type 1	27.9	25.4	14.6	0.68	0.49	6.80
Type 1	27.9	25.4	14.6	0.68	0.52	6.20
Type 1	27.9	25.4	14.6	0.68	0.55	6.10
Type 2	30.5	25.4	14.6	0.70	0.02	22.0
Type 2	30.5	25.4	14.6	0.70	0.05	16.0
Type 2	30.5	25.4	14.6	0.70	0.10	13.0
Type 2	30.5	25.4	14.6	0.70	0.16	11.0
Type 2	30.5	25.4	14.6	0.70	0.22	10.0
Type 2	30.5	25.4	14.6	0.70	0.30	8.50
Type 3	43.2	25.4	14.6	0.92	0.02	26.0
Type 3	43.2	25.4	14.6	0.92	0.05	19.0
Type 3	43.2	25.4	14.6	0.92	0.10	15.0
Type 3	43.2	25.4	14.6	0.92	0.16	13.0
Type 3	43.2	25.4	14.6	0.92	0.22	12.0
Type 3	43.2	25.4	14.6	0.92	0.30	10.5
Type 4	43.2	22.9	16.6	1.38	0.02	23.5
Type 4	43.2	22.9	16.6	1.38	0.03	20.9
Type 4	43.2	22.9	16.6	1.38	0.04	19.4
Type 4	43.2	22.9	16.6	1.38	0.05	18.3
Type 4	43.2	22.9	16.6	1.38	0.06	16.9
Type 4	43.2	22.9	16.6	1.38	0.10	15.8
Type 4	43.2	22.9	16.6	1.38	0.20	13.6

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Review Article | Volume 14, Issue 2 | Pages 507-516 | e-ISSN: 2347-470X

Table A3. Detailed Values of the Local Laboratory Dataset

Temp. (^{0}C)	Relative Humidity %	D_m (cm)	H (cm)	L (cm)	F	Conductivity (µs)	FOV (kV)
32	54	20	8	23	0.36	5.5	45
32	54	20	8	23	0.36	13.5	40
32	54	20	8	23	0.36	29.1	38
32	54	20	8	23	0.36	62.5	36
32	54	20	8	23	0.36	70	35
32	54	20	8	23	0.36	83	34
32	54	20	8	23	0.36	83.7	33
32	54	20	8	23	0.36	110	32
32	54	20	8	23	0.36	144.5	30
32	54	20	8	23	0.36	180	29
32	54	20	8	23	0.36	213	28
32	54	20	8	23	0.36	280	24
32	54	20	8	23	0.36	284	23
32	54	20	8	23	0.36	320	17
32	54	20	8	23	0.36	406	16
32	54	20	8	23	0.36	500	15
32	54	20	8	23	0.36	593	14
32	54	20	8	23	0.36	600	13
32	54	20	8	23	0.36	625	12
32	54	20	8	23	0.36	720	11
32	54	20	8	23	0.36	1013	9
32	54	20	8	23	0.36	1200	7
32	54	20	8	23	0.36	1300	5

Author Contributions: *Rama Alkhtiar:* Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. *Jamal Alnasseir:* Conceptualization, Resources, Investigation, Validation, Supervision. *George Isber:* Conceptualization, Resources, Investigation, Validation, Supervision.

Acknowledgments: The authors gratefully acknowledge the High Voltage Laboratory of Damascus University, Syria, for providing the experimental facilities and technical support that enabled the local laboratory measurements conducted in this study.

Data Availability Statement: All experimental data supporting the findings of this study are included in the manuscript (Tables A1–A3 in the Appendix). The Python code used for model training, hyperparameter tuning, and evaluation is available from the corresponding author upon reasonable request.

Conflicts of Interest: "The authors declare no conflict of interest."

REFERENCES

[1] L. Taibaoui, A. Mahdjoubi, and B. Zegnini, "Optimizing artificial neural networks with particle swarm optimization for accurate prediction of insulator flashover voltage under dry and rainy conditions," ITEGAM-JETIA, vol. 11, no. 51, pp. 213-219, 2025. doi: 10.5935/jetia.v11i51.1467.

[2] D. Doufene, S. Benharat, S. Bouazabia, and S. A. Bessedik, "Hybrid Grey Wolf and Finite Element Method (GWO-FEM) Algorithm for Enhancing High Voltage Insulator String Performance in Wet Pollution Conditions," Engineering, Technology & Applied Science Research, vol. 12, no. 3, pp. 8765-8771, Jun. 2022. doi: 10.48084/etasr.4978.

[3] L. Taibaoui, A. Mahdjoubi, and B. Zegnini, "Optimizing the prediction of lightning impulse withstand voltage for glass insulators using ANFIS enhanced with particle swarm optimization (PSO)," TEM Journal, vol. 14, no. 2, pp. 1725-1732, 2025. doi: 10.18421/TEM142-70.

[4] F. V. Topalis, I. F. Gonos, and I. A. Stathopulos, "Dielectric behaviour of polluted porcelain insulators," IEE Proceedings - Generation, Transmission and Distribution, vol. 148, no. 4, pp. 269-274, Jul. 2001. doi: 10.1049/ip-gtd:20010258.

[5] S. Kherfane, R. L. Kherfane, M. A. Moussa, and B. Toual, "Determining critical flashover voltage for various contaminated insulators using a hybrid approach of whale optimization and particle swarm optimization," Gongcheng Kexue Yu Jishu/Advanced Engineering Science, vol. 55, no. 2, pp. 260-274, Sep. 2023. doi: 10.5281/zenodo.4651203.

[6] R. Zahedi Khatir, M. Mirzaie, and H. Mahdavi, "Analysis of Flashover Voltage of Porcelain and Glass Insulators under Different Temperatures with Various Levels of Pollution and Humidity," Journal of Operation and Automation in Power Engineering, vol. 14, no. 4, pp. 257–266, Dec. 2026. doi: 10.22098/JOAPE.2025.15614.2201. [Online]. Available: https://joape.uma.ac.ir/article_4248.html

[7] A. Ali, A. R. Bhatti, A. Rasool, F. U. Rehman, M. A. Khan, A. Ali, and A. Sherefa, "Performance analysis of high voltage disc insulators with different profiles in clean and polluted environments using flashover, withstand voltage tests and finite element analysis," Scientific Reports, vol. 14, no. 1, p. 20299, 2024. doi: 10.1038/s41598-024-71392-5.

[8] A. A. Gialketsi, V. T. Kontargyri, I. F. Gonos, and I. A. Stathopulos, "Estimation of the flashover voltage on insulators using artificial neural networks," WSEAS Transactions on Circuits and Systems, vol. 4, no. 5, pp. 373-378, May 2005.

[9] S. Al Alawi, M. A. Salam, A. A. Maqrashi, and H. Ahmad, "Prediction of flashover voltage of contaminated insulator using artificial neural networks," Electric Power Components and Systems, vol. 34, no. 8, pp. 831-840, Aug. 2006. doi: 10.1080/15325000600561563.

[10] V. T. Kontargyri, G. J. Tsekouras, A. A. Gialketsi, and P. A. Kontaxis, "Comparison between artificial neural networks algorithms for the estimation of the flashover voltage on insulators," in Proc. 9th WSEAS Int. Conf. Neural Networks (NN'08), Sofia, Bulgaria, May 2008, pp. 225-230.

[11] B. Zegnini, M. Belkheiri, and D. Mahi, "Modeling flashover voltage (FOV) of polluted HV insulators using artificial neural networks (ANNs)," in Proc. Int. Conf. Electrical and Electronics Engineering (ELECO), Bursa, Turkey, Dec. 2009, pp. I-336-I-340. doi: 10.1109/ELECO.2009.5355301.

[12] U. Sajjad, A. Arshad, J. Ahmad, and S. Shoaib, "Application of artificial neural network in predicting flashover behaviour of outdoor insulators under polluted conditions," in Proc. IEEE Conf. Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Russia, Jan. 2021, pp. 2868-2873. doi: 10.1109/ElConRus51938.2021.9396388.

[13] K. Erenturk, "Adaptive-network-based fuzzy inference system application to estimate the flashover voltage on insulator," Instrumentation Science & Technology, vol. 37, no. 4, pp. 446-461, Jul. 2009. doi: 10.1080/10739140903087873.

[14] S. A. Bessedik and H. Hadi, "Prediction of flashover voltage of insulators using adaptive neuro-fuzzy inference system," Journal of Electrical Engineering, vol. 13, no. 1, pp. 1-8, Jan. 2013.

[15] G. E. Asimakopoulou, V. T. Kontargyri, G. J. Tsekouras, C. N. Elias, F. E. Asimakopoulou, and I. A. Stathopulos, "A fuzzy logic optimization methodology for the estimation of the critical flashover voltage on insulators," Electric Power Systems Research, vol. 81, no. 2, pp. 580-588, Feb. 2011. doi: 10.1016/j.epsr.2010.10.024.

[16] Y. Bourek, N. M'Ziou, and H. Benguesmia, "Prediction of flashover voltage of high-voltage polluted insulator using artificial intelligence," Transactions on Electrical and Electronic Materials, vol. 19, no. 1, pp. 1-10, Jan. 2018. doi: 10.1007/s42341-018-0010-3.

[17] B. Zegnini, A. H. Mahdjoubi, and M. Belkheiri, "A least squares support vector machines (LS-SVM) approach for predicting critical flashover voltage of polluted insulators," in Proc. IEEE Conf. Electrical Insulation and Dielectric Phenomena (CEIDP), Cancun, Mexico, Oct. 2011, pp. 403-406. doi: 10.1109/CEIDP.2011.6232680.

[18] A. Mahdjoubi, B. Zegnini, and M. Belkheiri, "A LS-SVM (least squares support vector machines) approach for predicting critical flashover voltage of polluted insulators," Journal of Energy and Power Engineering, vol. 7, no. 2, pp. 355-360, Feb. 2013.

[19] A. Mahdjoubi, B. Zegnini, and M. Belkheiri, "Kernels functions for squares support vector machines (LS-SVM) to diagnose HV polluted insulator," in Proc. 9ème Conf. Nationale sur la Haute Tension (CNHT'2013), Laghouat, Algeria, Apr. 2013, pp. 269-274.

[20] A. Mahdjoubi, B. Zegnini, M. Belkheiri, and T. Seghier, "Fixed least squares support vector machines for flashover modelling of outdoor insulators," Electric Power Systems Research, vol. 173, pp. 29-37, Aug. 2019. doi: 10.1016/j.epsr.2019.03.010.

[21] A. Mahdjoubi, B. Zegnini, and M. Belkheiri, "Prediction of critical flashover voltage of polluted insulators under sec and rain conditions using least squares support vector machines (LS-SVM)," Diagnostyka, vol. 20, no. 1, pp. 49-54, Nov. 2019. doi: 10.29354/diag/99854.

[22] H. Zhao, "Prediction of pollution flashover voltage of insulators based on genetic algorithm," in Proc. 2020 Int. Symp. Advances in Informatics, Electronics and Education (ISAIEE 2020), Dec. 2020, pp. 46-53. doi: 10.25236/isaiee.2020.010.

[23] S. A. Bessedik and H. Hadi, "Prediction of flashover voltage of insulators using least squares support vector machine with particle swarm optimisation," Electric Power Systems Research, vol. 104, pp. 87-92, Nov. 2013. doi: 10.1016/j.epsr.2013.06.013.

[24] S. A. Bessedik, H. Hadi, and R. Djekidel, "Improved least squares support vector machines to estimate flashover voltage of insulators," in Proc. 4th Int. Conf. Electrical Engineering (ICEE), Apr. 2016.

[25] S. Kherfane, R. L. Kherfane, A. Amari, N. Kherfane, F. Khoudja, B. Toual, and M. A. Moussa, "Estimation of the critical flashover voltage for different polluted insulators by particle swarm optimization," International Journal of Advanced Studies in Computer Science & Engineering, vol. 11, no. 7, pp. 1-8, 2022.

[26] L. Taibaoui, A. Mahdjoubi, and B. Zegnini, "LS-SVM improvement using PSO and GWO to determine flashover voltage of polluted insulators," in 1st International Conference on Innovative Academic Studies (ICIAS), Konya, Turkey, Sep. 2022.

[27] L. Taibaoui, A. Mahdjoubi, and B. Zegnini, "Enhanced prediction of insulator flashover voltage using artificial neural networks optimized with particle swarm optimization," Engineering, Technology & Applied Science Research, vol. 15, no. 4, pp. 25710-25718, Aug. 2025. doi: 10.48084/etasr.10330.

[28] S. He, Y. Han, Z. Zhao, W. Huang, et al., "Experimental database and prediction model of insulators flashover under lightning impulse," IEEE Transactions on Instrumentation and Measurement, vol. PP, no. 99, pp. 1–1, Jan. 2025, doi: 10.1109/TIM.2025.3568940.

[29] S. He, Y. Han, Z. Zhao, G. Liu, L. Qu, Z. Huang, Y. Zhang, B. Liu, Z. Wu, and L. Li, "Intelligent prediction of 110kV insulator lightning flashover criteria based on random forest," Electric Power Systems Research, vol. 232, p. 110423, Jul. 2024, doi: 10.1016/j.epsr.2024.110423.

[30] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[31] R. M. Alkhtiar, J. Alnassier, G. Isber, and I. Alwazah, "Flashover Voltage Prediction of Polluted Insulators Using Extreme Gradient Boosting (XGBoost)," in Proc. 2026 8th Int. Youth Conf. Radio Electronics, Electrical and Power Engineering (REEPE), 2026, doi: 10.1109/REEPE69046.2026.11481319.