


Article

Forecasting Flashover Parameters of Polymeric Insulators under Contaminated Conditions Using the Machine Learning Technique

Arshad ^{1,*}, Jawad Ahmad ², Ahsen Tahir ³, Brian G. Stewart ¹ and Azam Nekahi ⁴

¹ Institute for Energy and Environment, University of Strathclyde, Glasgow G1 1XQ, UK; brian.stewart.100@strath.ac.uk

² School of Computing, Edinburgh Napier University, Edinburgh EH11 4DY, UK; Jawadkhattak@ieeee.org

³ Department of Electrical Engineering, University of Engineering and Technology, Lahore 54890, Pakistan; ahsenkn001@gmail.com

⁴ School of Computing Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, UK.; azam.nekahi@gcu.ac.uk

* Correspondence: arshad.100@strath.ac.uk

Received: 28 May 2020; Accepted: 27 July 2020; Published: 30 July 2020



Abstract: There is a vital need to understand the flashover process of polymeric insulators for safe and reliable power system operation. This paper provides a rigorous investigation of forecasting the flashover parameters of High Temperature Vulcanized (HTV) silicone rubber based on environmental and polluted conditions using machine learning. The modified solid layer method based on the IEC 60507 standard was utilised to prepare samples in the laboratory. The effect of various factors including Equivalent Salt Deposit Density (ESDD), Non-soluble Salt Deposit Density (NSDD), relative humidity and ambient temperature, were investigated on arc inception voltage, flashover voltage and surface resistance. The experimental results were utilised to engineer a machine learning based intelligent system for predicting the aforementioned flashover parameters. A number of machine learning algorithms such as Artificial Neural Network (ANN), Polynomial Support Vector Machine (PSVM), Gaussian SVM (GSVM), Decision Tree (DT) and Least-Squares Boosting Ensemble (LSBE) were explored in forecasting of the flashover parameters. The prediction accuracy of the model was validated with a number of error cost functions, such as Root Mean Squared Error (RMSE), Normalized RMSE (NRMSE), Mean Absolute Percentage Error (MAPE) and R. For improved prediction accuracy, bootstrapping was used to increase the sample space. The proposed PSVM technique demonstrated the best performance accuracy compared to other machine learning models. The presented machine learning model provides promising results and demonstrates highly accurate prediction of the arc inception voltage, flashover voltage and surface resistance of silicone rubber insulators in various contaminated and humid conditions.

Keywords: silicone rubber; NSDD; ESDD; surface resistance; flashover; machine learning; bootstrapping

1. Introduction

Outdoor insulators are some of the most important high voltage and medium voltage components, playing a key role in the reliability of power transmission and distribution. Power system failures are mostly due to the contamination of outdoor insulators [1,2]. Due to the increasing demand for energy in the last few decades, the demand for reliable insulator designs and diagnostic tools is increasing. Outdoor insulators are subjected to various environmental conditions during service [3,4]. The pollution constituents arise from industrial, sea and dust deposits on the insulator surface as

a dry pollution layer. Although dry pollution does not have a significant effect on the insulation strength of the materials, during fog, rain, or moisture, the pollution constituents dissolve in water, leading to a thin conductive layer on the insulator surface [5]. The presence of a conductive layer encourages the flow of current along the insulator surface due to an applied electric field stress [6]. The superior performance of polymeric materials under contamination is due to their hydrophobicity. Due to the hydrophobic nature of silicone rubber, it resists the formation of a uniform water layer, leading to reduced leakage current and high surface resistance. Silicone rubber offers many advantages, e.g., low weight, easy handling, etc., as compared to porcelain, glass and other polymeric materials; however, the ageing of silicone rubber materials is unavoidable due to their organic nature [7–9]. The flashover process of polymeric insulators is also different than porcelain and glass insulators due to their hydrophobic nature [10–12]. Previous studies have shown that the electric field stress at the junction of the water droplet, air and insulator is very high, leading to corona and partial arcs [13,14]. Elongation of the water droplet occurs in the direction of the applied electric field [15]. These partial arcs and discharges at the edges of water droplets decrease the surface hydrophobicity, leading to the formation of a uniform water layer. Once a uniform water film forms on the insulator surface, leakage current flows through the conductive layer, resulting in Joule heating and dry band formation [16–18].

When the voltage drop across the dry band exceeds the withstanding capability of the dry band, partial arcs initiate and propagate, which may lead to flashover under certain conditions. The performance of outdoor insulators under contamination has been investigated by many researchers [19–22]. Mathematical models have been developed to predict the critical flashover voltage, critical leakage current and pollution resistance [23–28]. However, these models cannot be directly applied to silicone rubber and other polymeric insulators. The Equivalent Salt Deposit Density (ESDD) of an insulator is considered a reliable method for measuring the pollution severity [29,30]. However, silicone rubber material interacts with the pollution constituents, and measuring only ESDD may not give accurate results. It was proposed by an IEEE task force on insulator contamination that surface resistance can be used to predict the ageing and contamination severity of polymeric insulators [31].

The influence of pollution severity and fog rate on surface resistance was reported in [32,33]. Laboratory experiments of outdoor insulators are time consuming and need sophisticated hardware and software. Therefore, most of the previous work has utilized mathematical modelling to predict the flashover parameters of outdoor insulators. However, the existing mathematical models are used for ceramic and glass insulators and cannot be directly applied to polymeric insulators.

Intelligent systems based on neural network, fuzzy logic and Support Vector Machine (SVM) have been used to predict the critical flashover parameters of outdoor insulators [34–36]. The critical flashover voltage for outdoor insulators was predicted using Artificial Neural Networks (ANN) in [37,38]. Similar models have been developed to predict the critical flashover voltage and leakage current of outdoor insulators [39–44]. However, the training dataset used in these studies consisted of the insulator diameter, height, ESDD, creepage distance and form factor. Furthermore, the data were derived from the combination of laboratory experiments and using existing mathematical models of critical flashover voltage. There appears to be no experimental results of flashover parameters based on environmental conditions (humidity and temperature) and inert pollution for training the machine learning algorithms. The major limitation of previous mathematical and regression models is that they cannot be applied directly to polymeric insulators due to their unique property of hydrophobicity. Furthermore, the existing intelligent system applications in this area do not consider arc inception voltage and surface resistance. Measuring and predicting inception voltage and surface resistance are both critically important parameters to understand the process leading to flashover.

Previous statistical models predicting the surface resistance of polymeric insulators [30,45,46] are based on linear regression techniques and consider pollution severity and pollution type as input parameters. Currently, there appears to be no literature available on the use of experimental data and bootstrapping methods to increase the machine learning samples for insulator applications, thus providing the motivation for the work presented in this paper. Section 2 of this paper discusses

the problem formulation followed by the experimental setup and test method in Sections 3 and 4, respectively. Section 5 includes the details of the bootstrapping method and machine learning algorithms utilised in this work. Section 6 presents the results and discussion followed by the conclusions in Section 7.

2. Problem Description

Efforts have been undertaken in the past to predict the flashover voltage and ESDD of outdoor insulators using mathematical models and other statistical approaches. In this paper, an intelligent model is proposed to predict arc the inception voltage, flashover voltage and surface resistance of silicone rubber insulators based on environmental (humidity, temperature) and contaminated (ESDD, Non-Soluble Salt Deposit Density (NSDD)) conditions. The data were obtained from extensive experimentation in the laboratory under controlled environmental conditions. A total of four parameters and their influence on surface resistance, arc inception and flashover voltage were investigated.

To this end, the main contributions of this work are:

- An understanding of the effect of pollution severity, NSDD, humidity and temperature on arc inception voltage, flashover voltage and surface resistance through laboratory experiments.
- The implementation of machine learning algorithms such as ANN, Polynomial SVM (PSVM), Gaussian SVM (GSVM), Decision Tree (DT) and Least-Squares Boosting Ensemble (LSBE) to predict arc inception voltage, flashover voltage and surface resistance based on the experimental data.
- Accuracy evaluation of the proposed model with the actual measured dataset and also through bootstrapping methods using criteria such as Root Mean Squared Error (RMSE), Normalised RMSE (NRMSE), Mean Absolute Percentage Error (MAPE) and the Regression value (R).

3. Experimental Setup

The schematic diagram of the experimental setup is shown in Figure 1. High Temperature Vulcanized (HTV) silicone rubber samples having dimensions of 10 cm × 4 cm × 0.6 cm were polluted using the modified solid layer method based on IEC 60507. The test sample configuration is shown in Figure 2. The polluted samples were exposed to high voltage stress using a 0–100 kV, 100 kVA power frequency test transformer. All tests were performed in a climate chamber under controlled environmental conditions. The samples were placed on a horizontal table made of acrylic glass, and the high voltage was provided using a finely polished hemisphere electrode made of copper, 1.2 cm in diameter. The ground electrode was a copper plate of 1 cm in width. The voltage and current waveforms were measured using a high voltage capacitive divider and current transformer, respectively. The data were acquired using an NI data acquisition card and LabVIEW software. As the silicone rubber loses its hydrophobicity after exposure to high voltage stress, multiple samples were prepared with various ESDD and NSDD levels. Further details regarding the experimental setup and sample configuration can be found in [47,48]. Additionally, experimental data is provided in Appendix A.

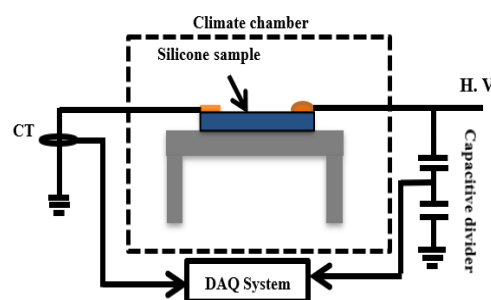


Figure 1. Test setup.

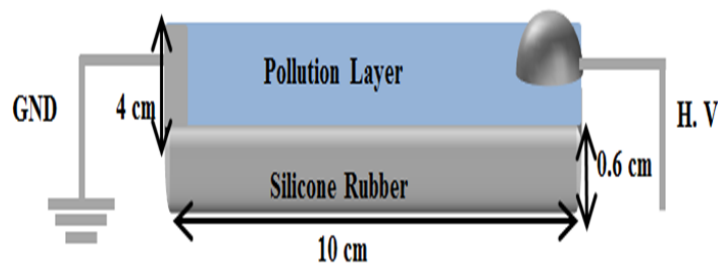


Figure 2. Test sample configuration.

4. Test Method

Humidity, temperature, fog rate and rain intensity were set and controlled within the climate chamber. All tests were performed under clean fog (solid layer method) conditions, and the fog rate was kept constant at 4 L/h. Before energizing, samples were placed in the chamber for 20 min to facilitate uniform wetting of the surface and avoid unintentional dry band formation. A visual inspection was performed before energizing to make sure the samples were uniformly wetted. Initially, tests were performed on a uniformly polluted sample to determine the typical flashover voltage. Once the predicted flashover voltage was determined, the applied voltage was increased in steps of 5% of the predicted flashover voltage starting from zero and until flashover occurred. Each step change in voltage was maintained for 2 min, and if no arcs appeared on the insulator surface, the voltage was increased further. When the first arc appeared on the insulator surface, a note of voltage magnitude was made, and this was defined as the arc inception voltage. At the arc inception stage, the voltage was again kept constant for 2 min until either the arc disappeared or led to flashover. Each test was repeated multiple times to generate sufficient data for machine learning algorithms to be effectively trained.

Surface resistance was measured according to the guidelines presented in the IEEE task force on insulator contamination [31]. One of the major requirements in surface resistance measurement is that the applied voltage should be low enough that it does not lead to partial arcs, but high enough to facilitate the flow of leakage current. Therefore, the applied voltage was kept constant at 15 V/mm, and the insulators were stressed for 1 h under clean fog conditions. Pollution was reapplied after each test so that a uniform pollution layer could be maintained during the test. Multiple tests were performed on each sample under different ESDD and NSDD conditions and under different humidity and temperature environments.

Once enough experimental data were generated, a machine learning model was developed based on four input parameters (ESDD, NSDD, humidity and temperature) and three output parameters (inception voltage, surface resistance and flashover). The performance of each algorithm was evaluated on four performance metrics, namely RMSE, NRMSE, MAPE and R. The details of the proposed machine learning algorithms are presented in the following section.

5. Machine Learning Algorithms

A number of machine learning algorithms such as ANN, PSVM, GSVM, DT and LSBE were applied to predict the flashover parameters of the silicone rubber insulation. To improve the accuracy of the proposed algorithms, bootstrapping was used to increase the number of data samples. Details about the bootstrapping method and machine learning algorithms are given in the following section.

5.1. Bootstrap Method

The bootstrap method is used to provide a better approximation of the actual distribution of the parameters of estimation. It expands upon the single realization of the actual distribution to generate the bootstrap sample, which provides improved understanding of the actual distribution or process. Sampling with a replacement procedure is used to generate bootstrap samples. This involves randomly

selecting one sample from the dataset containing N samples and placing it in the bootstrap dataset. Another observation is then randomly sampled and is placed in the bootstrap dataset. This process continues until a new bootstrap dataset is formed with N samples. It was reported in [49] that bootstrapping is highly effective when applied to a small sample size ($N < 30$). In the case of a larger sample size, bootstrapping offers the benefits of improved accuracy in statistical estimation and increased percentage of estimation effectiveness and is independent of the sample distribution. Due to random selection and replacement, the bootstrap dataset may include multiple copies of the same observation and no copies of other observations. Therefore, it is important to select the appropriate number of bootstrap samples for increased confidence in the obtained results. In this study, the random selection of bootstrap samples was performed using a two-dimensional chaos map also known as TD-ERCS (Tangent Delay Ellipse Reflecting Cavity Map System). TD-ERCS is used for random number generation and permutation [50]. Such a chaos system provides an equiprobability distribution and hence can be effectively used in bootstrapping methods. Through such a map, one can select random samples from a provided dataset. The TD-ERCS method is expressed through [50]:

$$\begin{cases} x_n = -\frac{2k_{n-1}y_{n-1} + x_{n-1}(\mu^2 - k_{n-1}^2)}{\mu^2 + k_{n-1}^2} \\ y_n = k_{n-1}(x_n - x_{n-1}) + y_{n-1}, \quad n = 1, 2, 3, \dots \end{cases} \quad (1)$$

where:

$$k_n = \frac{2k'_{n-m} - k_{n-1} + k_{n-1}(k'_{n-m})^2}{1 + 2k_{n-1}k'_{n-m} - k(k'_{n-m})^2} \quad (2)$$

$$k'_{n-m} = \begin{cases} -\frac{x_{n-1}}{y_{n-1}}\mu^2 & n < m \\ -\frac{x_{n-m}}{y_{n-m}}\mu^2 & n \geq m \end{cases} \quad (3)$$

$$y_0 = \mu\sqrt{1 - x_0^2} \quad (4)$$

$$k'_0 = -\frac{x_0}{y_0}\mu^2 \quad (5)$$

$$k_0 = -\frac{\tan\alpha + k'_0}{1 - k'_0\tan\alpha} \quad (6)$$

$$\begin{cases} \mu \in (0, 1) \\ x_0 \in [-1, 1] \\ \alpha \in (0, \pi) \\ m = 2, 3, 4, 5, \dots \end{cases} \quad (7)$$

Here, μ , x_0 , α and m are known as seed parameters of the TD-ERCS map and x_n and y_n are random sequences, respectively. These seed parameters are utilised in random number generation. The number of bootstrap samples used to train the machine learning algorithms was 100. Each bootstrap sample consisted of 44 observations, which were used for training, while the observations not selected during bootstrapping from the original dataset were used for testing. The performance metrics were calculated for each bootstrap sample, and the average values along with the Standard Deviation (STD) are reported in this paper. The mean value and 95% confidence interval are also calculated and presented in this paper.

$$\bar{y} = \frac{1}{B} \sum_{b=1}^B y \quad (8)$$

$$STD = \sqrt{\frac{\sum_{b=1}^B (y - \bar{y})^2}{B - 1}} \quad (9)$$

where \bar{y} is the mean value and B is the number of bootstrap samples.

5.2. Artificial Neural Network

In machine learning algorithms, one of the main tools is ANN, which mimics biological networks. Unlike programmed task rules, ANN learns tasks from a set of examples. Generally, the output data are manually labelled. For example, in image classification algorithms, the output is labelled as “cat” and “dog” with example images provided for training the ANN algorithm. After training a network, data are provided to the ANN, and the output “cat” and “dog” is predicted. A simple neural network can be represented as shown in Figure 3. In ANN, a neural network consists of the input, hidden and output layers, respectively.

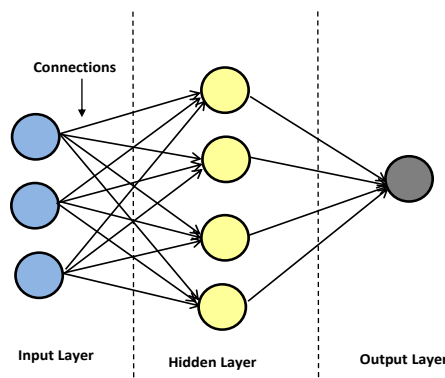


Figure 3. ANN network.

The available dataset is mainly divided into training data and testing data. In the training phase, ANN is trained with a learning algorithm such as gradient descent, the Newton method, the conjugate gradient, the quasi-Newton method or the Levenberg–Marquardt (LM) algorithm. In this work, we used gradient descent due to the faster convergence and less computational time. Mathematically, gradient descent minimises a cost function, and the derivative of the error function is updated based on the gradient. The cost function in this case is the RMSE. The updated weights are calculated as:

$$\alpha(x + 1) = \alpha(x) - \nabla \alpha(x), \quad (10)$$

where $\alpha(x)$ is the weight at iteration x and $\nabla \alpha(x)$ is the gradient of $\alpha(x)$, which can be written as:

$$\nabla \alpha(x) = \mu \frac{\partial f}{\partial \alpha} \quad (11)$$

where μ is the learning rate and $\frac{\partial f}{\partial \alpha}$ is the loss function. According to the chain rule, $\frac{\partial f_m}{\partial \alpha_m}$ at iteration m is written as:

$$\frac{\partial f}{\partial \alpha} = \frac{\partial f_m}{\partial y_m} \frac{\partial y_m}{\partial w_m}, \quad (12)$$

where y_m is the output. In this research, the sigmoid activation function is used. Mathematically, the sigmoid function (β) can be written as:

$$\beta(y) = \frac{1}{1 + \exp^{-y}}, \quad (13)$$

where exp is the exponential function. In this paper, input neurons to the ANN model consist of pollution severity, NSDD, relative humidity and ambient temperature, while the output neurons are surface resistance, inception voltage and flashover voltage. The proposed ANN model can be seen in Figure 4. The ANN training and testing steps are shown in Figure 5.

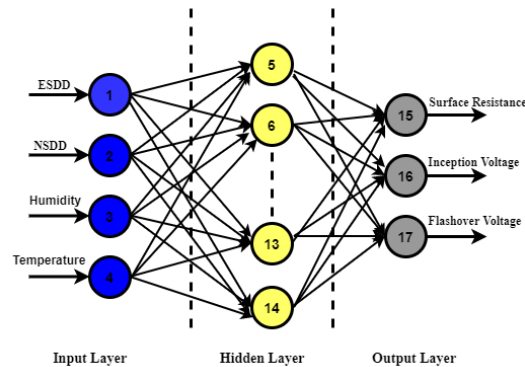


Figure 4. The proposed ANN model for flashover parameters.

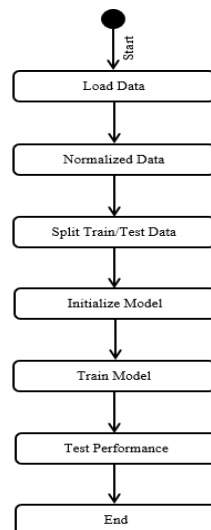


Figure 5. Steps involved in the proposed model.

5.3. Support Vector Machine

SVM is a machine learning algorithm used for both classification and regression problems. In SVM, there is a starting point containing all the dataset records. Then, the SVM algorithm finds the optimal hyper-plane, on the basis of the highest value of dissimilarity possible, which linearly separates the data points. Unlike the classical neural network approach, the SVM formulation of the learning problem leads to Quadratic Programming (QP) with a linear constraint. However, the size of the matrix involved in the QP problem is directly proportional to the number of training points. Consider a given training set x_i, y_i , where $i = 1, 2, \dots, N$ with input data $x_i \in R$ and output data $y_i \in R$. The following regression model can be constructed by using a non-linear mapping function $\Phi(x)$:

$$y = w^T \Phi(x) + b \quad (14)$$

where w is the weight vector and b is the bias term. As in ANN, it is necessary to minimize a cost function C containing a penalized regression error, as follows:

$$\min C(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=0}^N e_i^2 \quad (15)$$

subject to equality constraints:

$$y = w^T \Phi(x) + b + e_i. \quad (16)$$

The first part of this cost function is a weight decay, which is used to regularize weight sizes and penalize large weights. Due to this regularization, the weights converge to a similar value. In this work, two different SVM kernels were used, namely polynomial and Gaussian, for predicting the flashover parameters.

5.4. Decision Tree

DT is a supervised learning method used for classification and regression using a model of decisions or tree-like graph. DT is basically a contrasting machine learning technique as compared to ANN or SVM. The other approaches deal with the data numerically, while the theory of DT does not depend on the assumption of numerical data. In DT, there are nodes, branches (links) and leaves; each node represents an attribute (feature); branches denote a decision (rule); and a leaf signifies an outcome. Any interested reader can find more details in [51]. DT mainly work on an “if, then, else” programmable structure. The inputs are compared against constant values, and the tree is split based on the less than, equal to or greater than value from the constant. The Classification and Regression Trees (CART) algorithm is utilised in this work to select the best split input at each node. Gini’s diversity Index (GI) is used as the split criterion, with $N - 1$ maximum number of spits, where N is the input set size for input $x_i \in R^d$ and output data $y_i \in R$. The GI value is calculated from Equation (17) as:

$$GI = 1 - \sum_{i=1}^N p^2(i) \quad (17)$$

In (17), $p(i)$ represents the probability of the i th input. The tree is grown by estimating an optimal pruned sequence of sub-trees, and the leaves, which originate from the same parent node, are combined together. During testing, the test inputs are routed down the tree according to their values and are compared at each node against constants. The final value is obtained once the leaf node is reached.

5.5. Least-Squares Boosting Ensemble

Instead of utilizing a single tree, multiple regression trees can be combined for higher accuracy. Such techniques are known as regression tree ensembles. In the ensemble technique, weak learners are combined to get a stronger learner. Mainly two types of decision tree ensemble methods are used: (i) bagging and (ii) boosting. In both cases, the aim is to decrease variances by combing several learners. In bagging, the responses on N learners are averaged, while in boosting, the results are calculated based on the weighted average. In the boosting regression tree ensemble, higher weights are allocated to good classification/regression. In both cases, the aim is to decrease variances by combing several learners. In our work, the best results were obtained with the regression tree ensemble generated with the LSBE algorithm. LSBoost fits tree learners to regression ensembles. The ensemble generates an aggregate prediction $g(x_i)$ of the previously grown regression tree learners. For every new learner, LSBoost fits the learner to the difference between the obtained output and the aggregated prediction $g(x_i)$ of previously grown learners. The LSBE algorithm fits tree learners to minimize the mean squared error, as given in Equation (18):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \eta g(x_i))^2 \quad (18)$$

The LSBE algorithm was tested with various parameters for leaf size, number of regression tree learners and the learning rate η . The best results were obtained with the parameter values illustrated in Table 1.

Table 1. Least-Squares Boosting Ensemble (LSBE) algorithm parameters.

LSBE Parameters	Values
Minimum Leaf Size	8
Tree Learners	30
Learning rate, η	0.1

6. Results and Discussions

The above machine learning methods were applied to the flashover parameter datasets. The learning rate for ANN was 0.005, and the learning algorithm was gradient decent with 10 hidden neurons and 1000 epochs. The predicted results were close to the actual results. However, a numeric value will be more accurate for the comparison of different machine learning algorithms. The efficiency and strength of each algorithm was evaluated on four metrics, namely RMSE, NRMSE, MAPE and R . The accuracy of each algorithm was tested using the aforementioned metrics. RMSE was used to calculate the average magnitude error, while NRMSE was used to calculate the normalised square root of the average of squared differences between the actual and predicted flashover parameters. MAPE is the percentage representation of the mean of the sum of absolute differences between the actual and predicted parameters. These parameters can be represented mathematically using the following expressions.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (FP_{Ai} - FP_{Pi})^2} \quad (19)$$

$$NRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{FP_{Ai} - FP_{Pi}}{FP_{Ai}} \right)^2} \quad (20)$$

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{FP_{Ai} - FP_{Pi}}{FP_{Ai}} \right| \times 100\% \quad (21)$$

$$R = 1 - \left(\frac{\sum_{i=1}^N (FP_{Ai} - FP_{Pi})}{\sum_{i=1}^N FP_{Ai}} \right)^2 \quad (22)$$

where N is the number of samples and FP_A and FP_P are the actual and predicted flashover parameters i.e., surface resistance, inception voltage and flashover voltage at time i , respectively. For machine learning algorithms to be efficient and reliable, the values of RMSE, NRMSE and MAPE should be as close to zero as possible, while the value of R should be as close to one as possible.

The 95% confidence interval, which was centred on the bootstrap estimator and the width of which was proportional to the standard deviation, was calculated for all performance metrics mentioned above. However, the obtained frequency distribution was very similar in all cases and can be better represented with STD. Therefore, only confidence interval histograms for surface resistance are reported, which were obtained with DT algorithms, in Figure 6. Figure 6 shows the distribution of RMSE, NRMSE, MAPE and the R value. The vertical lines on the histograms represent the 95% confidence level on the obtained results.

Figures 7–9 show the performance metrics for the different machine learning algorithms. It can be observed from Figure 7 that the lowest RMSE, NRMSE and MAPE values were obtained in the case of PSVM for surface resistance. For the prediction of surface resistance, the performance of ANN was low when compared to the other machine learning algorithms. Moreover, the R value was also very close to the ideal value of one for PSVM. A similar study was performed to predict the inception

voltage. The performance metrics obtained are shown in Figure 8. The RMSE, NRMSE and MAPE values obtained in the case of PSVM were lower when compared to the other algorithms. However, the results obtained by applying LSBE were close to PSVM in this case. Again, the performance of ANN was low in comparison to other machine learning algorithms. The value of R was comparably similar in the case of LSBE and GSVM.

The different performance metrics obtained for flashover voltage are presented in Figure 9. In this case, again, the PSVM outperformed the other machine learning algorithms. However, for flashover prediction, GSVM performed better than LSBE. Comparing the obtained results, it was observed that PSVM performed better in all cases. Additionally, the performance of each algorithm was different for each output parameter. The best prediction accuracy was obtained for surface resistance as compared to inception and flashover voltage.

The vertical lines on the graphs show the variation in the mean values when bootstrapping was applied. It can be observed from Figure 7 that apart from the mean values of RMSE, NRMSE, MAPE and R , the STD was lower in the case of PSVM, DT, LSBE and GSVM as compared to ANN. The lowest variation was noted in the case of PSVM followed by GSVM and LSBE in the case of surface resistance. A similar trend in performance metrics variation was observed in Figures 8 and 9 for inception voltage and flashover voltage. However, the proposed machine learning algorithms performed better for surface resistance as compared to inception and flashover voltage. Comparing the results presented in Figures 7–9, it can be observed that ANN performed relatively better in the case of inception voltage as compared to surface resistance and flashover voltage.

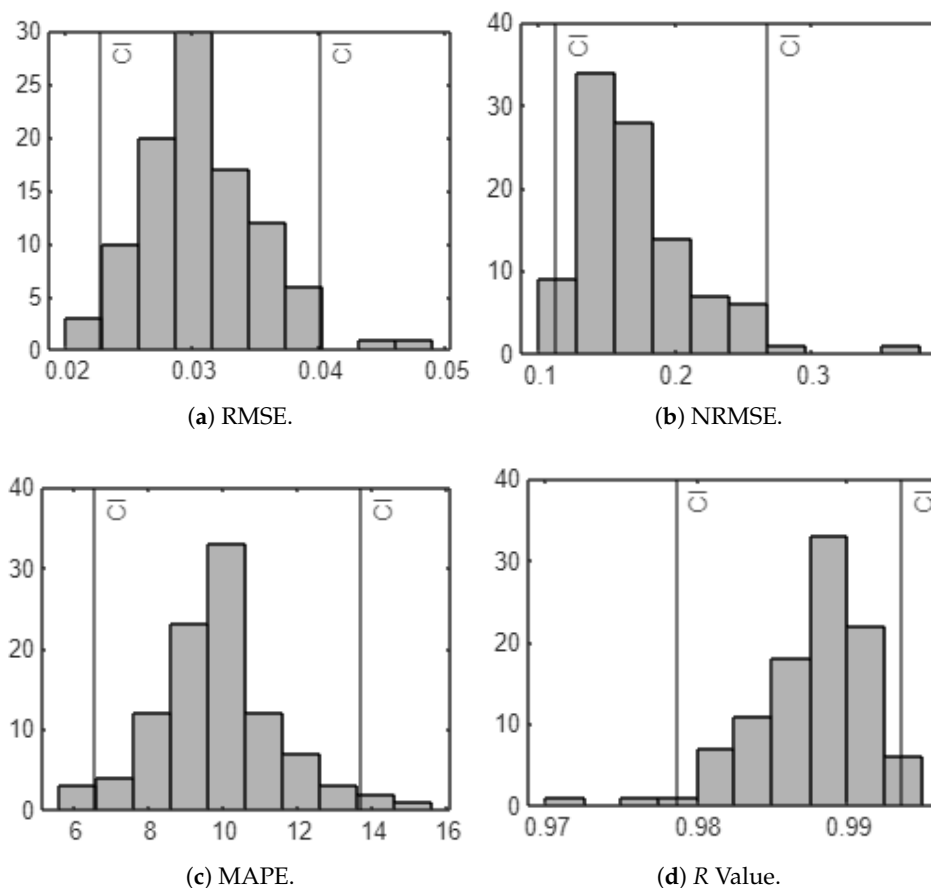


Figure 6. Histograms of the 95% confidence interval of surface resistance using the DT algorithm: (a) RMSE, (b) NRMSE, (c) MAPE, (d) R Value.

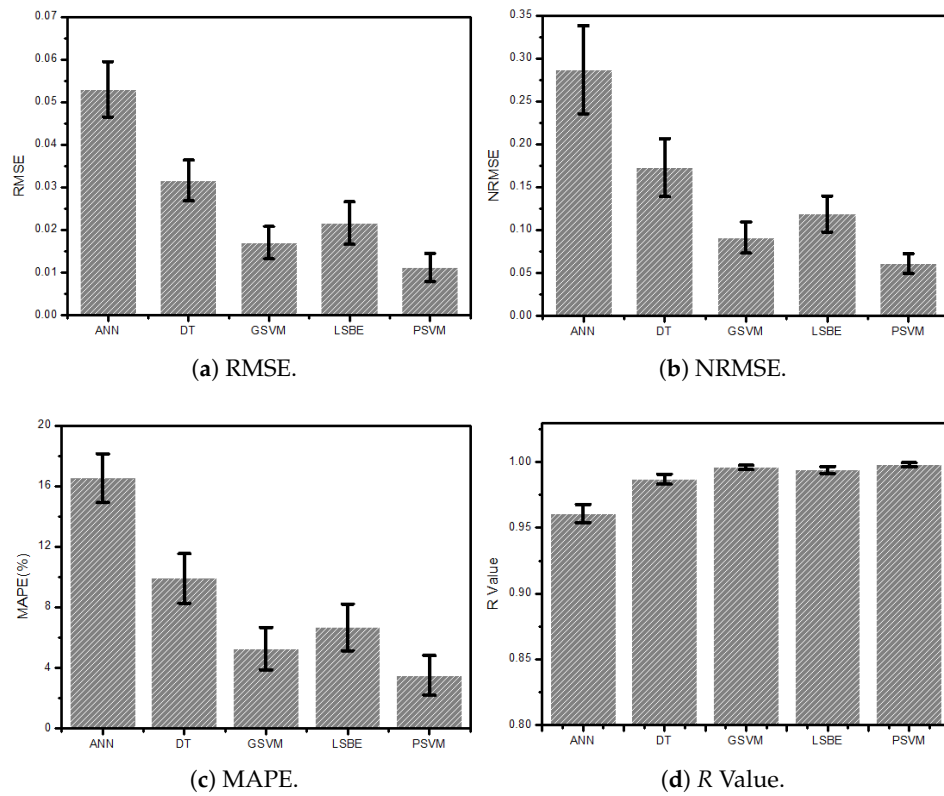


Figure 7. Comparison of different machine learning algorithms for surface resistance: (a) RMSE, (b) NRMSE, (c) MAPE, (d) R Value.

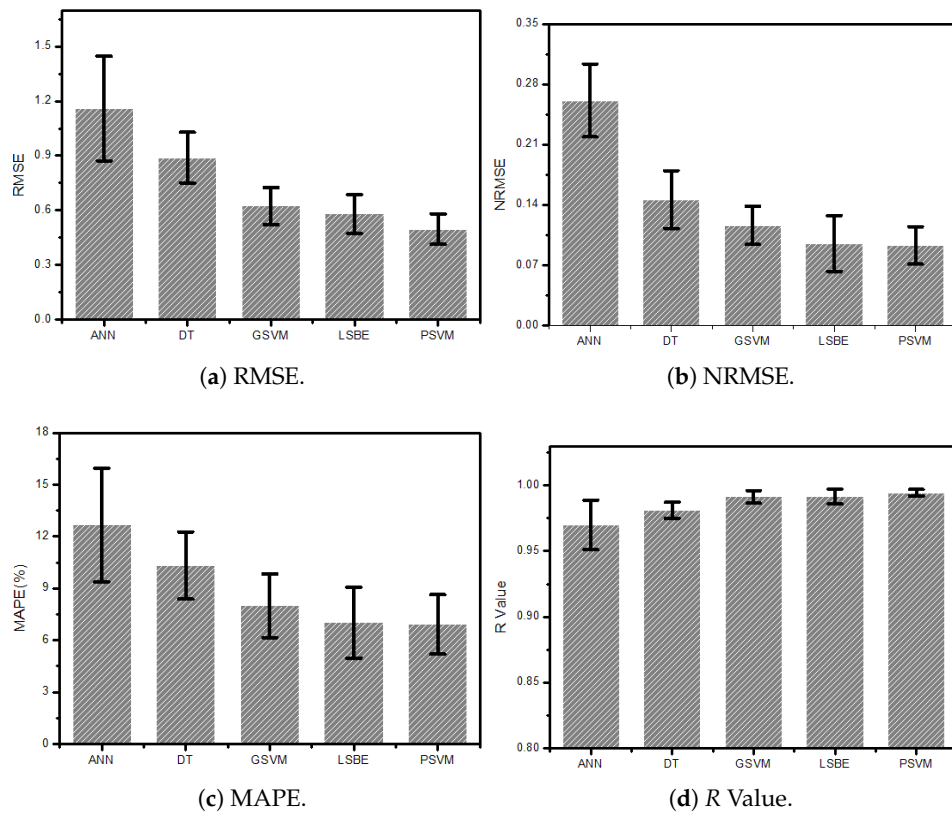


Figure 8. Comparison of different machine learning algorithms for inception voltage: (a) RMSE, (b) NRMSE, (c) MAPE, (d) R Value.

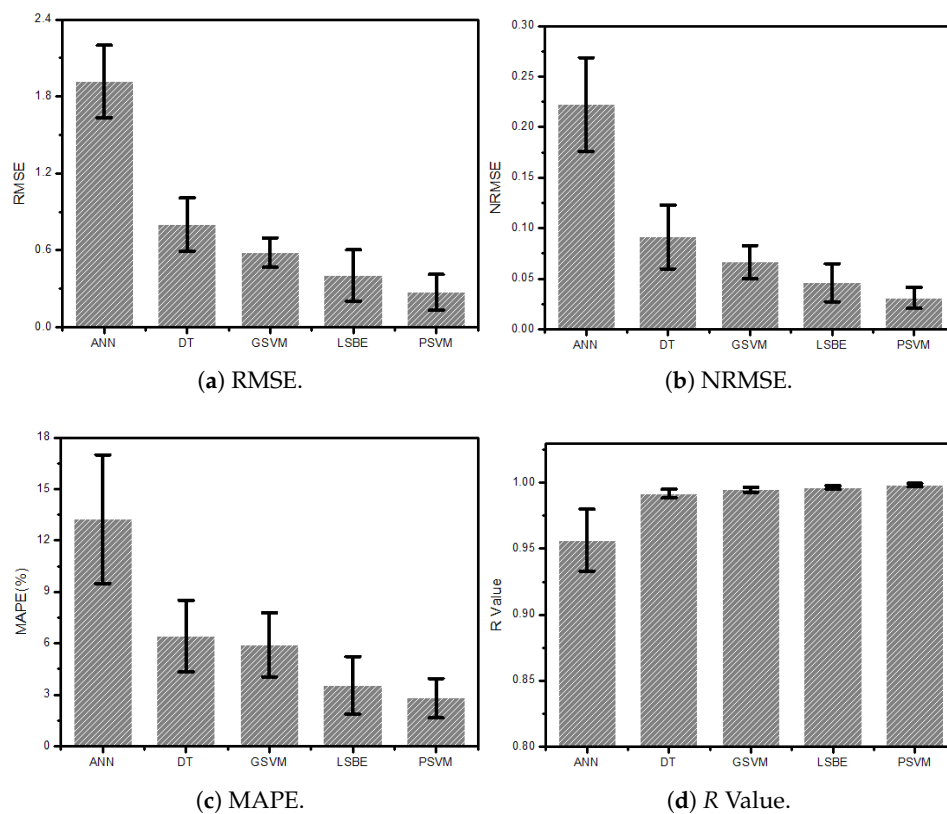


Figure 9. Comparison of different machine learning algorithms for flashover voltage: (a) RMSE, (b) NRMSE, (c) MAPE, (d) R Value.

Comparing the mean and STD of the RMSE from Figure 7 to Figure 9, it can be observed that the RMSE was lowest for PSVM in all three cases. However, comparing on the basis of output parameter, it was noted that the surface resistance RMSE was lowest followed by flashover voltage and inception voltage in all cases, except ANN. The NRMSE comparison shows that PSVM showed better performance for all machine learning algorithms and output parameters. However, in contrast to RMSE, NRMSE was noted to be low for flashover voltage as compared to inception voltage and surface resistance. Similar to RMSE and NRMSE, MAPE was lower in the case of PSVM when compared to other machine learning algorithms. The results obtained for R value showed that ANN had the lowest R values for all flashover parameters as compared to other techniques. However, the overall results showed that the R values obtained for flashover voltage were higher than inception voltage and surface resistance in all cases, except ANN. Furthermore, the computational complexity of the machine learning algorithms was also evaluated using MATLAB R2019b on a PC with a 2.90 GHz processor and 8.0 GB memory. The computational times for ANN, PSVM, GSVM, DT and LSBE were 1.47, 0.06, 0.03, 0.05 and 0.58 s, respectively. Based on the computational time, it was noted that GSVM outperformed all other machine learning algorithms.

These results highlight that the prediction performance of PSVM and LSBE was better when compared to the other machine learning algorithms based on the accuracy metrics. However, in terms of computational complexity, GSVM performed better followed by PSVM and DT, while ANN had the highest computational complexity. All test parameters indicated that the machine learning schemes provided results that were close to actual values, and hence, a machine learning algorithm such as PSVM may be employed for predicting flashover parameters of high voltage insulators. It was also observed from the obtained results that selecting the best machine learning algorithm was mainly dependent on the type of performance evaluation metric and the flashover parameter. For example, LSBE showed better performance than GSVM for inception voltage and flashover voltage, but GSVM

outperformed LSBE in the case of surface resistance. It is therefore recommended to use multiple performance metrics while choosing the best machine learning algorithm for a specific application. The proposed machine learning algorithms were effectively applied to lab data utilised in this paper. It is anticipated that the methodology should be effective for field insulator data, but this requires validation on captured field datasets. Obtaining a large workable dataset from a field environment is always a challenge; therefore, bootstrapping may also be used to effectively increase the sample space for these datasets.

7. Conclusions

Flashover is one of the major causes of insulation failure in outdoor environments. This paper presents machine learning algorithms for predicting flashover parameters of HTV silicone rubber insulators under varying environmental and polluted conditions. Five different types of machine learning algorithms were used, and the performance of each algorithm was tested on a set of criteria. It was observed from the obtained results that machine learning can be effectively utilised to predict the flashover performance of outdoor high voltage insulation under humid and contaminated conditions. Additionally, it was concluded that the performance of machine learning algorithms depends on the size of the training set as well as the flashover parameter chosen for prediction. The prediction performance of all machine learning algorithms was better for surface resistance as compared to inception and flashover voltage. It was also concluded that PSVM outperformed the other four algorithms in all cases. This model can be effectively used to predict the flashover parameters of outdoor insulators under varying environmental and polluted conditions. This will also help asset managers and utility engineers in the selection of outdoor insulators for contaminated conditions, as well as in the ability to predict when flashover performance may start to be an issue.

Author Contributions: Conceptualization, A. and A.N.; formal analysis, A. and B.G.S.; investigation, A. and A.N.; methodology, A.; project administration, A.N.; resources, A.N.; software, J.A. and A.T.; supervision, B.G.S.; validation, J.A.; visualization, A.T.; writing, original draft, A. and J.A.; writing, review and editing, A.T. and B.G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported at the University of Strathclyde through EPSRC Grant EP-T001445-1.

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

Appendix A

Table A1 shows the experimental data used to train the machine learning algorithms.

Table A1. Experimental data.

Temperature (°C)	Humidity (%)	NSDD (mg/cm ²)	ESDD (mg/cm ²)	SR(Mohm/cm)	IV(kV)	FV(kV)
5	70	0.75	0.05	0.33	8.35	11.9
5	70	0.75	0.1	0.285	6.3	10.7
5	70	0.75	0.2	0.247	5.16	8.1
5	70	0.75	0.4	0.2	4.2	5.6
10	70	0.75	0.05	0.323	8.1	11.45
10	70	0.75	0.1	0.28	6.2	10.3
10	70	0.75	0.2	0.24	5	7.56
10	70	0.75	0.4	0.19	4	5.4
15	70	0.75	0.05	0.317	7.9	10.8
15	70	0.75	0.1	0.273	6.02	9.83
15	70	0.75	0.2	0.231	4.85	7.31
15	70	0.75	0.4	0.179	3.84	5.12
20	70	0.75	0.05	0.31	7.67	10.6
20	70	0.75	0.1	0.268	5.85	9.39
20	70	0.75	0.2	0.225	4.71	6.8
20	70	0.75	0.4	0.172	3.68	4.75
10	70	0.1	0.05	0.392	11.2	14.3
10	70	0.1	0.1	0.356	9.3	13.5

Table A1. Cont.

Temperature (°C)	Humidity (%)	NSDD (mg/cm ²)	ESDD (mg/cm ²)	SR (Mohm/cm)	IV (kV)	FV (kV)
10	70	0.1	0.2	0.32	7.8	12.17
10	70	0.1	0.4	0.27	6.65	10.1
10	70	0.3	0.05	0.37	10.2	13.38
10	70	0.3	0.1	0.33	8.12	12.64
10	70	0.3	0.2	0.29	6.9	11.3
10	70	0.3	0.4	0.25	5.8	8.86
10	70	0.5	0.05	0.351	9.15	12.52
10	70	0.5	0.1	0.31	7.35	11.38
10	70	0.5	0.2	0.268	6.1	10.2
10	70	0.5	0.4	0.22	5.05	7.8
10	70	0.75	0.05	0.323	8.1	11.45
10	70	0.75	0.1	0.28	6.2	10.3
10	70	0.75	0.2	0.24	5	7.5
10	70	0.75	0.4	0.19	4	5.4
10	50	0.75	0.05	0.35	8.5	11.45
10	50	0.75	0.1	0.31	6.45	10.3
10	50	0.75	0.2	0.261	5.2	7.56
10	50	0.75	0.4	0.215	4.2	5.4
10	70	0.75	0.05	0.323	7.9	10.82
10	70	0.75	0.1	0.28	5.8	9.66
10	70	0.75	0.2	0.24	4.75	7.45
10	70	0.75	0.4	0.19	3.87	4.86
10	90	0.75	0.05	0.29	7.1	10.43
10	90	0.75	0.1	0.25	5.3	9.12
10	90	0.75	0.2	0.206	4.2	6.38
10	90	0.75	0.4	0.167	3.4	4.7

References

- Li, Q.F.; Fan, Z.; Wu, Q.; Gao, J.; Su, Z.Y.; Zhou, W.J. Investigation of ice-covered transmission lines and analysis on transmission line failures caused by ice-coating in China. *Power Syst. Technol.* **2008**, *32*, 33–36.
- Kim, C.J.; Shin, J.H.; Yoo, M.H.; Lee, G.W. A study on the characterization of the incipient failure behavior of insulators in power distribution line. *IEEE Trans. Power Deliv.* **1999**, *14*, 519–524. [[CrossRef](#)]
- Yoshimura, N.; Kumagai, S.; Nishimura, S. Electrical and environmental aging of silicone rubber used in outdoor insulation. *IEEE Trans. Dielectr. Electr. Insul.* **1999**, *6*, 632–650. [[CrossRef](#)]
- Hussain, M.M.; Farokhi, S.; McMeekin, S.G.; Farzaneh, M. Risk assessment of failure of outdoor high voltage polluted insulators under combined stresses near shoreline. *Energies* **2017**, *10*, 1661. [[CrossRef](#)]
- Hackam, R. Outdoor HV composite polymeric insulators. *IEEE Trans. Dielectr. Electr. Insul.* **1999**, *6*, 557–585. [[CrossRef](#)]
- Hampton, B. Flashover mechanism of polluted insulation. *IET Digit. Libr.* **1964**, *5*, 985–990.
- Fernando, M.; Gubanski, S. Ageing of silicone rubber insulators in coastal and inland tropical environment. *IEEE Trans. Dielectr. Electr. Insul.* **2010**, *17*, 326–333. [[CrossRef](#)]
- Amin, M.; Amin, S. Aging research on SIR and TPE insulators (an overview). *Rev. Adv. Mater. Sci.* **2014**, *36*, 29–39.
- Gorur, R.S.; Karady, G.G.; Jagota, A.; Shah, M.; Yates, A.M. Aging in silicone rubber used for outdoor insulation. *IEEE Trans. Power Deliv.* **1992**, *7*, 525–538. [[CrossRef](#)]
- Karady, G.G.; Shah, M.; Brown, R. Flashover mechanism of silicone rubber insulators used for outdoor insulation-I. *IEEE Trans. Power Deliv.* **1995**, *10*, 1965–1971. [[CrossRef](#)]
- Shah, M.; Karady, G.G.; Brown, R. Flashover mechanism of silicone rubber insulators used for outdoor insulation-II. *IEEE Trans. Power Deliv.* **1995**, *10*, 1972–1978. [[CrossRef](#)]
- Mughal, M.A.; Nekahi, A.; Khan, M.; Umer, F. Influence of Single and Multiple Dry Bands on Critical Flashover Voltage of Silicone Rubber Outdoor Insulators: Simulation and Experimental Study. *Energies* **2018**, *11*, 1335.
- Que, W.; Sebo, S.A. Electric field and potential distributions along non-ceramic insulators with water droplets. In Proceedings of the Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Conference (Cat. No. 01CH37264), Cincinnati, OH, USA, 18 October 2001; pp. 441–444.
- Guan, Z.; Wang, L.; Yang, B.; Liang, X.; Li, Z. Electric field analysis of water drop corona. *IEEE Trans. Power Deliv.* **2005**, *20*, 964–969. [[CrossRef](#)]

15. Imano, A.M.; Beroual, A. Deformation of water droplets on solid surface in electric field. *J. Colloid Interface Sci.* **2006**, *298*, 869–879. [[CrossRef](#)] [[PubMed](#)]
16. Waters, R. Formation and characterization of dry bands in clean fog on polluted insulators. *IEEE Trans. Dielectr. Electr. Insul.* **1999**, *6*, 714–731.
17. Arshad; Nekahi, A.; McMeekin, S.; Farzaneh, M. Influence of dry band width and location on flashover characteristics of silicone rubber insulators. In Proceedings of the 2016 IEEE Electrical Insulation Conference (EIC), Montreal, QC, Canada, 19–22 June 2016; pp. 73–76.
18. Zhou, J.B.; Gao, B.; Zhang, Q.G. Dry band formation and its influence on electric field distribution along polluted insulator. In Proceedings of the 2010 Asia-Pacific Power and Energy Engineering Conference, Chengdu, China, 28–31 March 2010; pp. 1–5.
19. Farzaneh, M.; Kienicki, J.; Martin, R. A laboratory investigation of the flashover performance of outdoor insulators covered with ice. In Proceedings of the 4th International Conference on Properties and Applications of Dielectric Materials (ICPADM), Brisbane, Queensland, Australia, 3–8 July 1994; Volume 2, pp. 483–486.
20. Nekahi, A.; McMeekin, S.; Farzaneh, M. Investigating flashover behaviour of silicone rubber insulators under contaminated conditions. In Proceedings of the 2015 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), Ann Arbor, MI, USA, 18–21 October 2015; pp. 856–859.
21. Nekahi, A.; McMeekin, S.; Farzaneh, M. Numerical computation of electric field and potential along silicone rubber insulators under contaminated and dry band conditions. *3D Res.* **2016**, *7*, 25.
22. Liu, Y.; Xia, B.; Du, B.; Farzaneh, M. Influence of Fine Metal Particles on Surface Discharge Characteristics of Outdoor Insulators. *Energies* **2016**, *9*, 87. [[CrossRef](#)]
23. Wilkins, R. Flashover voltage of high-voltage insulators with uniform surface-pollution films. *IET Digit. Libr.* **1969**, *3*, 457–465. [[CrossRef](#)]
24. Mekhaldi, A.; Namane, D.; Bouazabia, S.; Beroual, A. Flashover of discontinuous pollution layer on HV insulators. *IEEE Trans. Dielectr. Electr. Insul.* **1999**, *6*, 900–906. [[CrossRef](#)]
25. Li, Y.; Yang, H.; Zhang, Q.; Yang, X.; Yu, X.; Zhou, J. Pollution flashover calculation model based on characteristics of AC partial arc on top and bottom wet-polluted dielectric surfaces. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 1735–1746. [[CrossRef](#)]
26. Hadjrioua, F.; Mahi, D.; Slama, M.E.A. Application of the analytical arc parameters on the dynamic modelling of HVDC flashover of polluted insulators. In Proceedings of the 2014 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM), Tunis, Tunisia, 3–6 November 2014; pp. 1–5.
27. Dhahbi-Megrache, N.; Beroual, A. Predictive dynamic model of the leakage current and flashover voltage of discontinuously polluted insulators under ac voltage: experimental validation. *J. Phys. Appl. Phys.* **2007**, *40*, 7782. [[CrossRef](#)]
28. Hussain, M.M.; Chaudhary, M.A.; Razaq, A. Mechanism of Saline Deposition and Surface Flashover on High-Voltage Insulators near Shoreline: Mathematical Models and Experimental Validations. *Energies* **2019**, *12*, 3685. [[CrossRef](#)]
29. Lambeth, P.; Schneider, H.; Beausejour, Y.; Dumora, D.; Kawamura, T.; Marrone, G.; Moran, J.; Naito, K.; Nigbor, R.; Sakich, J.; et al. Final report on the clean fog test for HVAC insulators. *IEEE Trans. Power Deliv.* **1987**, *2*, 1317–1326. [[CrossRef](#)]
30. Venkataraman, S.; Gorur, R. Prediction of flashover voltage of non-ceramic insulators under contaminated conditions. *IEEE Trans. Dielectr. Electr. Insul.* **2006**, *13*, 862–869. [[CrossRef](#)]
31. Gorur, R.; Schneider, H.; Cartwright, J.; Beausajour, Y.; Kondo, K.; Gubanski, S.; Hartings, R.; Shah, M.; McBride, J.; De Tourreil, C.; et al. Surface resistance measurements on nonceramic insulators. *IEEE Trans. Power Deliv.* **2001**, *16*, 801–805. [[CrossRef](#)]
32. Gorur, R.; Subramanian, K. Use of surface resistance for assessing vulnerability of HV outdoor insulators to contamination flashover. In Proceedings of the 2003 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Albuquerque, NM, USA, 19–22 October 2003; pp. 406–409.
33. Gorur, R.S.; Olsen, R.; Crane, J.; Adams, T.; Journey, J. Prediction of flashover voltage of insulators using low voltage surface resistance measurement. In *Power System Engineering Research Center; Final Project Report*; Arizona State University: Tempe, AZ, USA, 2006; pp. 154–163.
34. Salam, M.; Al-Alawi, S.; Maqrashi, A. Prediction of equivalent salt deposit density of contaminated glass plates using artificial neural networks. *J. Electrostat.* **2008**, *66*, 526–530. [[CrossRef](#)]

35. Shuai, H.; Gong, Q. Insulator ESDD forecasting under complex climate conditions on the basis of LSSVM. In Proceedings of the 2009 IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende, Italy, 21–23 September 2009; pp. 313–317.
36. Almad, A.; Ahmad, H.; Salam, M.; Ahmad, S. Regression technique for prediction of salt contamination severity on high voltage insulators. In Proceedings of the 2000 Annual Report Conference on Electrical Insulation and Dielectric Phenomena (Cat. No. 00CH37132), Victoria, BC, Canada, 15–18 October 2000; Volume 1, pp. 218–221.
37. Kontargyri, V.; Gialketsi, A.; Tsekouras, G.; Gonos, I.; Stathopoulos, I. Design of an artificial neural network for the estimation of the flashover voltage on insulators. *Electr. Power Syst. Res.* **2007**, *77*, 1532–1540. [[CrossRef](#)]
38. Proietti, A.; Leccese, F.; Caciotta, M.; Morresi, F.; Santamaria, U.; Malomo, C. A new dusts sensor for cultural heritage applications based on image processing. *Sensors* **2014**, *14*, 9813–9832. [[CrossRef](#)]
39. Gençoğlu, M.; Cebeci, M. Investigation of pollution flashover on high voltage insulators using artificial neural network. *Expert Syst. Appl.* **2009**, *36*, 7338–7345. [[CrossRef](#)]
40. Asimakopoulou, G.; Kontargyri, V.; Tsekouras, G.; Asimakopoulou, F.; Gonos, I.; Stathopoulos, I. Artificial neural network optimisation methodology for the estimation of the critical flashover voltage on insulators. *IET Sci. Meas. Technol.* **2009**, *3*, 90–104. [[CrossRef](#)]
41. Li, J.; Sun, C.; Sima, W.; Yang, Q.; Hu, J. Contamination level prediction of insulators based on the characteristics of leakage current. *IEEE Trans. Power Deliv.* **2009**, *25*, 417–424.
42. Zegnini, B.; Belkheiri, M.; Mahi, D. Modeling flashover voltage (FOV) of polluted HV insulators using artificial neural networks (ANNs). In Proceedings of the 2009 International Conference on Electrical and Electronics Engineering-ELECO 2009, Bursa, Turkey, 5–8 November 2009; p. I-336.
43. Bashir, N.; Ahmad, H. A neural network based method for the diagnosis of ageing insulators. In Proceedings of the 2009 IEEE Symposium on Industrial Electronics & Applications, Kuala Lumpur, Malaysia, 4–6 October 2009; Volume 1, pp. 41–46.
44. El-Hag, A.H.; Jahromi, A.N.; Sanaye-Pasand, M. Prediction of leakage current of non-ceramic insulators in early aging period. *Electr. Power Syst. Res.* **2008**, *78*, 1686–1692. [[CrossRef](#)]
45. Venkataraman, S.; Gorur, R. Non linear regression model to predict flashover of nonceramic insulators. In Proceedings of the 2006 38th North American Power Symposium, Carbondale, IL, USA, 17–19 September 2006; pp. 663–666.
46. Venkataraman, S.; Gorur, R.; Mishra, A. Impact of weathering on flashover performance of nonceramic insulators. *IEEE Trans. Dielectr. Electr. Insul.* **2008**, *15*, 1073–1080. [[CrossRef](#)]
47. Arshad; Nekahi, A.; McMeekin, S.G.; Farzaneh, M. Flashover characteristics of silicone rubber sheets under various environmental conditions. *Energies* **2016**, *9*, 683.
48. Arshad; Nekahi, A.; McMeekin, S.; Farzaneh, M. Measurement of surface resistance of silicone rubber sheets under polluted and dry band conditions. *Electr. Eng.* **2018**, *100*, 1729–1738.
49. Lola, M.S.; Zainuddin, N.H.; Ramlee, M.N.A.; Sofyan, H. Double bootstrap control chart for monitoring sukuk volatility at Bursa Malaysia. *J. Teknol.* **2017**, *79*. [[CrossRef](#)]
50. Ahmad, J.; Hwang, S.O. A secure image encryption scheme based on chaotic maps and affine transformation. *Multimed. Tools Appl.* **2016**, *75*, 13951–13976. [[CrossRef](#)]
51. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]

