Breakdown Voltage Prediction for Sphere and Semispheroid Geometries With Gaussian Process Regression-Based Model Under the Application of Lightning Impulses of Both Polarities Prévision de la tension de claquage pour les géométries sphérique et semi-sphéroïde avec un modèle basé sur la régression du processus Gaussien sous l'application d'impulsions d'éclair des deux polarités

Vidya M. S., *Graduate Student Member, IEEE*, Sunitha K., *Member, IEEE*, Deepa S. Kumar, *Senior Member, IEEE*, Deepak Mishra, *Senior Member, IEEE*, and Ashok S., *Senior Member, IEEE*

***Abstract*— The design of high-voltage (HV) systems is principally dependent on the discharge voltage of their insulation. Sphere geometry and semispheroid geometry are extremely important in HV systems, such as ground rods and gas-insulated substations (GISs). Hence, in this work, a machine learning algorithm is proposed to develop a model to predict the discharge characteristics of air for sphere and semispheroid geometries. Finite element method (FEM) simulations have been performed to extract different electric fields and energy features of air gaps in the range of 5–40 mm under lightning impulses of both polarities. While developing the model, these features along with gap lengths are considered. The features have been used for training a machine learning algorithm based on the Gaussian process regression (GPR) to develop the model. The outcomes received from the model are ratified with measured experimental data. A good comparison between the two establishes the fidelity of the novel model. The proposed methodology is also compared with the other state-of-the-art techniques and found good. Remarkable performance has been acquired for other gap geometries as well.**

***Résumé*— La conception des systèmes haute tension (HT) dépend principalement de la tension de décharge de leur isolation. La géométrie des sphères et des demi-sphéroïdes est extrêmement importante dans les systèmes HT, tels que les tiges de terre et les sous-stations à isolation gazeuse (GIS). Par conséquent, dans ce travail, un algorithme d'apprentissage automatique est proposé pour développer un modèle permettant de prédire les caractéristiques de décharge de l'air pour les géométries sphérique et semi-sphéroïde. Des simulations par la méthode des éléments finis (FEM) ont été réalisées pour extraire les différents champs électriques et les caractéristiques d'énergie des fentes d'air dans la gamme de 5 à 40 mm sous des impulsions d'éclair des deux polarités. Lors de l'élaboration du modèle, ces caractéristiques ainsi que les longueurs d'entrefer sont prises en compte. Les caractéristiques ont été utilisées pour entraîner un algorithme d'apprentissage automatique basé sur la régression du processus Gaussien (GPR) pour développer le modèle. Les résultats obtenus à partir du modèle sont ratifiés avec les données expérimentales mesurées. Une bonne comparaison entre les deux établit la fidélité du nouveau modèle. La méthodologie proposée est également comparée à d'autres techniques de pointe et s'avère bonne. Des performances remarquables ont été obtenues pour d'autres géométries d'interstices.**

***Index Terms*— Gaussian processes (GPs), high-voltage (HV) techniques, insulation, lightning.**

I. Introduction

THE role of dielectric insulation is important in modern power systems. It is reported that most of the failures occurring in the power system are due to insulation failures [1]. The insulation system is always subjected to different kinds of stresses; the most important among them is due to overvoltages. Various kinds of overvoltages strike the insulation of power apparatus during their operation, the significant among them being lightning and switching overvoltages. Employment of air as insulation is dominant in many areas of the power system. When air insulation is subjected to different kinds of overvoltages, discharge occurs, and this may give rise to

Manuscript received June 10, 2020; revised October 23, 2020, June 6, 2021, and October 10, 2021; accepted January 3, 2022. *(Corresponding author: Vidya M. S.)*

Vidya M. S. is with the Department of Electrical Engineering, College of Engineering Trivandrum, Thiruvananthapuram 695016, India (e-mail: mvidya07@gmail.com).

Sunitha K and Ashok S are with the Department of Electrical Engineering, National Institute of Technology, Calicut, Kerala 673601, India (e-mail: karakkadsunitha@gmail.com; ashoks@nitc.ac.in).

Deepa S. Kumar is with the Power Networks Demonstration Center, University of Strathclyde, Glasgow, U.K. (e-mail: deepa.kumar@ieee.org)

Deepak Mishra is with the Department of Avionics, Indian Institute of Space Science and Technology, Valiamala, Kerala 695547, India (e-mail: deepak.mishra@iist.ac.in).

Associate Editor managing this article's review: Daniela Constantinescu.

Digital Object Identifier 10.1109/ICJECE.2022.3142891

This is a peer reviewed, accepted author manuscript of the following research article: S., V. M., K., S., Kumar, D. S., Mishra, D., & S., A. (2022). Breakdown voltage prediction for sphere and semispheroid geometries with gaussian process regression-based model under the application of lightning impulses of both polarities. IEEE Canadian Journal of Electrical and Computer Engineering, 45(2), 132-140. https://doi.org/10.1109/ICJECE.2022.3142891

effects such as corona and short circuits. The phenomena being random, it is difficult to predict the voltage at which it happens, principally for nonuniform gaps. Usually, a power system has an extensive range of air gaps from short to long, which may be uniform or nonuniform. Hence, for the efficient design of insulation, it is indispensable to predict the discharge characteristics precisely.

Prebreakdown phenomena in short rod–plane gaps under lightning voltages of positive and negative polarities have been explained in [2]. The negative discharge characteristics of air under lightning impulses are well explained in [3] by considering the discharge as an RLC network. Modeling and computation of discharge parameters of discharges propagating in the air have been done in [4]. The streamer development under impulse voltages for nonuniform air gaps is studied, and its classification has been done in [5] and [6]. One of the early models developed in the field of air breakdown, namely, the disruptive effect model, which is popularly known as the generalized integration method, predicts time to break down and not breakdown voltage (BDV). This method for modeling and prediction of impulse volt–time characteristics has been discussed by Darveniza and Vlastos [7] and Ancajima *et al.* [8]. A feedforward network-based model had been developed for the partial discharges in solid insulating materials in [9]. These models are based on real physical phenomena, and more simplified models have been developed based on machine learning methods. For such models, data collected experimentally from some known gap lengths are employed to predict the flashover voltage for the unknown gap lengths. The prediction of transformer oil BDV has been done in [10] using the artificial neural network (ANN). A support vector machine (SVM)-based model for power frequency BDV has been developed in [11]. Considering the energy-storage features, the switching impulse breakdown characteristics have been obtained for long air gaps by Qiu *et al.* [12]. In [13], a support vector regression (SVR)-based model has been developed for power frequency BDV for rod–plane air gaps.

The above literature review reveals that many of the researchers have developed predictive models of the breakdown of short and long air gaps under power frequency and switching voltages. However, it is important to mention here that a model for the prediction of discharge characteristics of 5–40-mm air gap under lightning impulses for sphere and semispheroid geometry has not been developed so far. Sphere–sphere geometries are found in protective devices, and semispheroid geometries are found in ground rods of power systems. In addition to that, conducting particles, which take the shape of spheroid geometry, are important in gas-insulated substations (GISs) as they may cause partial discharge breakdown in the gaseous dielectrics [14]. These electrodes are often subjected to overvoltages of different types. Because of these facts, an attempt has been made in our work to predict the discharge characteristics of air under lightning impulses of positive polarity and negative polarity for sphere–sphere geometry, and the work has been extended to semispheroid geometry. We have used a semiempirical approach mentioned in [12]. Different methods of prediction are based on fitting a regression model to experimental data and finding a relation between output and input variables. This method introduces much error as fitting a predefined relationship may not be accurate, particularly when the phenomena are mostly stochastic. Hence, a probabilistic prediction method based on a predictive distribution has gained much attention in many engineering applications. The fundamental principles of air discharge are well explained by Townsend's theory and the streamer theory [15]. Accordingly, the basic air discharge is explained by the presence of primary electrons, secondary electrons, ionization, excitation, electron attachment, and several other electron processes. The accumulative effect of all these processes along with other atmospheric conditions makes air discharge under different geometries and gap lengths possess a stochastic character, and it would be suitable to predict the discharge characteristics based on predictive distribution. The most recent development in predictive distribution is a Gaussian process (GP) distribution. Hence, in this work, a machine learning algorithm based on GP regression (GPR) is made use of. GPR-based models work as the Bayesian estimation, where a prior distribution is assumed, and observed data are relocated based on Baye's rule [16].

The proposed model is robust and pliable based on the following grounds.

1) The number of assumptions made about the shape of the estimator functions is less. Hence, the relationship between the known variables and unknown variables can be easily visualized.
2) As GPR is based on the dependencies between the features, it can be appropriate to datasets with a small number of features.
3) GPR is inherently probabilistic and is useful for predictions in phenomena such as discharge characteristics of air.

The main contributions of our work are the following: development of an effective model for predicting the breakdown between electrodes of the sphere and semispheroid geometry for medium-length air gaps under lightning impulses of positive polarity. The model possesses high accuracy with the reasonable number of input features: extending the work to negative polarity impulses and experimental validation of the model.

II. Prediction Methodology

The prediction of air flashover with the help of machine learning algorithms can be found in [11]–[13]. The method utilized in these papers is primarily based on feature extraction using the finite element method (FEM), and the same is employed to develop a novel model for the required geometric configurations of medium gap lengths in our work. However, it is to be noted that the number of input features required has been found only six to develop an efficient model. The effective employment of this method necessitates four steps. Initially, the computation of field and energy features throughout the length of the air gap needs to be carried out. Second, the values of the BDV of these gap lengths are collected experimentally. A consolidated dataset that comprises all the features and BDV magnitudes of gap distances is to be prepared next. The last step involves the development of the model using the prepared dataset.

A. Application of Finite Element Method to the Feature Extraction of Air Gaps

The performance of any machine learning algorithm relies on good features. Hence, feature extraction of the air gap of the required geometry is the first and an important step in the prediction of BDV. The distribution of different features across the air gap changes with the geometry, length of the gap, amplitude, and shape of the input voltage wave. Accuracy in the extracted parameters is mostly influenced by these factors, and hence, precision modeling of the electrode geometry is inevitable to get the desired results. Accordingly, modeling of the geometry is done by using standard dimensions of the electrodes available. For sphere geometry, spheres of the standard dimension of 100-mm diameter, and for semispheroid geometry, Verband Deutscher Electrotechniker (VDE) electrodes have been used, and the dimensions are as per IEC-60156 standard.

Computations of the field and energy features have been done by FEM using COMSOL Multiphysics 5.2 software. The geometry of the required electrode configuration is first created. Triangular meshing is done, and the size of the triangular elements is made extremely fine throughout the geometry to get accurate results. After grounding the bottom electrode, the other electrode is supplied with lightning impulses. Lightning impulse waveform applied is of standard 1.2/50 μs, as shown in the following equation (IS-20171):

$$u(t) = A(e^{-\alpha t} - e^{-\beta t}) \tag{1}$$

where $u(t)$ represents the instantaneous value of the wave, A represents the magnitude, and α and β represent constants. In this work, positive polarity and negative polarity lightning impulses have been used for both experiments and simulation. It is worthwhile to mention that about 90% of the lightning strikes in power systems are of negative polarity. The simulations are carried out using an electrostatics module with a time-dependent study. Five features (electric field and energy) have been extracted for each gap length. The features extracted are given as follows.

1) The electric field strength (E) characterizes the intensity of the field along the discharge path. The discharge path chosen is the shortest path between the two electrodes [11]. The variation of E over the discharge path between the two electrodes has been considered while developing the model.
2) The electric energy density (Ewe), represents the capacitive energy stored in the gap.
3) The electric potential (V) characterizes the potential distribution between the electrodes. When the lightning impulse voltage is applied between the electrodes, the potential distribution in the gap varies and is different for different gap lengths and geometries.
4) The current density (J) constitutes the electric current per cross-sectional area at a given point in space. After creating the model, lightning impulses have been given as the input, and a time-dependent study has been chosen in the simulation since lightning impulse varies with time. In the solution matrix, the current density also shows a small value and, hence, has been considered as a parameter in the training set.
5) The total electric energy (E_{twe}) is the total energy due to electric and magnetic fields.

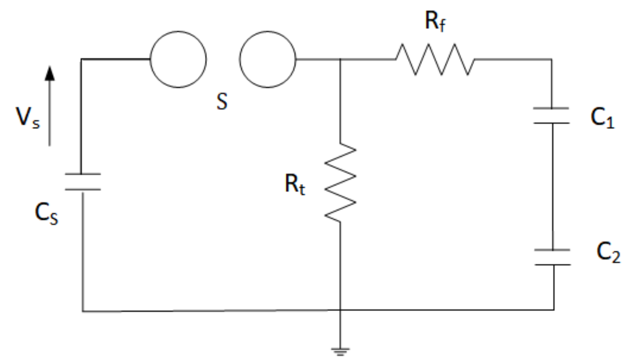


Fig. 1. Impulse generator circuit configuration.

The analysis is based on solving a set of equations in between the electrodes subjected to the following boundary conditions. i.e., the potential distribution is obtained by solving the Laplace equation at each node, i.e., $\nabla^2 V = 0$. Under the zero charge boundary condition, the equation $\nabla . V = 0$ is solved. Initial values are assumed as $V = 0$, where V is the applied potential. The ground boundary condition applied refers to $V = 0$. In the FEM, the potential distribution is obtained by considering the minimum energy criteria between different elements. For this, the entire solution space is divided into small elements, and (2) is solved as follows

$$W = \frac{1}{2}\epsilon E^2. \tag{2}$$

Equation (2) is written in Cartesian coordinates as follows:

$$W = \frac{1}{2}\epsilon \int\int\int \left[\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 + \left(\frac{\partial V}{\partial z}\right)^2 \right]. \tag{3}$$

Since the model is created in 2-D, the change in potential along the z-direction is zero, and hence, the equation becomes

$$W = \frac{1}{2}\epsilon \int\int\int \left[\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 \right]. \tag{4}$$

B. Experimental Setup

The training dataset required to train the machine learning model is obtained by conducting experiments on the predefined configurations of electrode geometries. Marx's single-stage impulse generator of rating 140 kV has been used to generate the standard lightning impulse waveform for conducting the experiments.

The circuit for the generation of impulses is shown in Fig. 1, and Fig. 2 shows the experimental setup. C_1 represents previously charged capacitor, and S represents the sphere gap. When the desired voltage is applied, the sphere gap breaks down, and the required voltage is applied across the test cell. C_2 is the load capacitor, and R_1 and R_2 determine the wave shape characteristics of the impulse waveform generated.

Fig. 2. Test setup.

1) Experimental Procedure: Experiments have been conducted as per IEC standards. Great care has been taken to ensure that the surfaces of the electrodes are clean, free from dust or deposited moisture. Also, the gap between the electrodes is kept free from floating dust particles, fibers, and so on. The up and down method has been used to find out the BDV. During this procedure, an approximate initial voltage (V) is selected. (ΔV) constitutes equally spaced voltage levels above and below the starting voltage. After applying the first shot at V, if a breakdown occurred, $V - \Delta V$ is chosen for the next shot. The value of voltage is increased to $V + \Delta V$ otherwise. An identical procedure is repeated for negative polarity impulses. As per standards, V_{50} voltage is established after a minimum of 20 applications of voltages for self-restoring insulation [15], [17]. Atmospheric correction factors are applied accordingly to the experimental data [18]. A consolidated set consisting of all the five features extracted by simulation and the corresponding gap lengths constitute the six input parameters of the training data. The BDVs obtained experimentally for each gap length are chosen as the output of the training data given to the model. The dataset is normalized and shuffled before being input into the model.

C. Gaussian Process Regression Model

The regression analysis has been proven as an efficient tool for addressing many engineering problems. Most of the physical phenomena occurring in nature can be described by a model, in which the dependent variables are related to independent variables by some relation that can be represented by a mathematical equation. In our study, the extracted features of the geometry considered as independent variables include gap length (G), E, E_{we}, V, J, and E_{twe}. The target variable considered as a dependent variable is the BDV. With all variables being real-valued, a regression model based on predictive distribution can be used. It is to be noted here that discharge of air is a probabilistic phenomenon that follows a Gaussian distribution. Based on the literature survey, the GPR model is observed to solve complicated nonlinear problems even for a small sample space. Furthermore, it has a strong theoretical basis in statistical learning. Hence the model is selected for prediction in this work.

GPs have been considered as an infinite extension of a multivariate normal distribution. The correlation between the input variable and the output variable can be written as follows:

$$y_i = f(x_i) + \varepsilon \tag{5}$$

where $f(x_i)$ is the function representing the independent variable for the ith observation and ε is the additive noise. For a zero mean value

$$\varepsilon \sim \mathcal{N}\big(0, \sigma_n^2\big) \tag{6}$$

where σ_n^2 represents the variance of noise and n is the number of observations. The prior distribution of the training sample is represented by the following equation:

$$y = \mathcal{N}\big(0, K + \sigma_n^2 I\big) \tag{7}$$

where I is the nth-order unit matrix.

The best estimate of the dependent variable f_* of a new test dataset is found out for a training dataset $D_s = \{X, y\}$. A Gaussian function is completely described by m_x and $k(x, x')$, where m_x is the mean and $k(x, x')$ is the covariance function or kernel.

For any x, x'

$$m_x = E(f(x)) \tag{8}$$

$$k(x, x\prime) = \text{Cov}(f(x), f(x\prime)). \tag{9}$$

In our work, a zero mean is assumed, and a squared exponential kernel function given by (10) is used. The choice of this squared exponential kernel is very useful for smooth functions

$$k\big(x, x'\big) = \sigma_f^2 \exp\bigg(-\frac{1}{2}\big(x - x'\big)^T M\big(x - x'\big)\bigg) \tag{10}$$

where σ_f^2 represents the signal variance, $M = \text{diag}(l)^{-2}$, where $l = \{l_k | k = 1, 2, \ldots, d\}$ represents characteristic length scale for each input dimension, and M forms a $d \times d$ matrix with its diagonal consisting of $(1/l_k^2)$ and zero elsewhere. l, σ_f, σ_n are called hyperparameters and are found out by the Markov chain Monte Carlo method [16]. For the training dataset D_s, the set of test input vector X, and a new set of inputs X_*, the joint distribution of y and f_* is formed as a matrix as

$$\begin{pmatrix} y \\ f_* \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} K(X, X) & K(X_*, X) \\ K(X, X_*) & K(X_*, X_*) \end{bmatrix} \tag{11}$$

where $K(X, X_*)$ describes the matrix of co variances computed at all pairs of training and test samples. Similarly, the other $K(,)$ expressions denote the matrix of covariances. Now, for X and y, the outputs f_* can be estimated from new sets of inputs X_* by modeling the function y as a GP. i.e., if the observed data are y and the unobserved data are f_*, coming from a GP, concatenating y and f_* results in a multivariate normal distribution with the mean and covariance structure

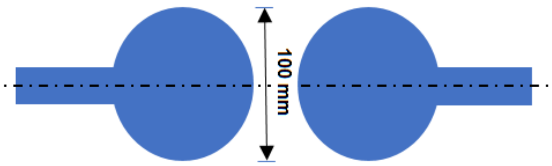


Fig. 3. Sphere–sphere geometry.

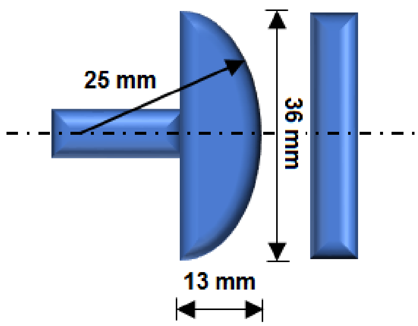


Fig. 4. Semispheroid–plane geometry.

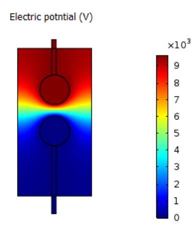


Fig. 5. Potential distribution (volts) of the sphere–sphere electrode configuration. The potential is decreasing toward the bottom electrode.

given by (8) and (9) [16], [19]. Now, because y is observed, f_* can be modeled as the conditional distribution of a multivariate normal using (12)

$$p(f_*|X_*, X, y) \sim \mathcal{N}(K(X_*, X)K(X, X)^{-1}y \tag{12}$$

$$K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)). \tag{13}$$

In this study, f_* represents the discharge voltage of the unknown gap distance.

III. Results and Discussion

To study the effect of various gap lengths on BDVs, the geometry of the two specific electrode configurations had been created as 2-D models in COMSOL software. Fig. 3 represents a sphere–sphere electrode configuration designed with the standard diameter of 100 mm, and Fig. 4 represents a semispheroid shape electrode designed with dimensions as per standards mentioned in Section II-A. Lightning impulses of unit magnitude have been applied to the top electrode, and with the other electrode grounded, the field and energy features along the length between the electrodes are extracted. Fig. 5 shows a sample plot of potential distribution for 40-mm gap sphere–sphere configuration. The potential is higher toward the two end electrodes. The sample plot of E for sphere–sphere geometry is shown in Fig. 6. From the figure, it is implied that the distribution of features varies widely in a nonlinear manner. The gap between sphere–sphere geometries is considered uniform when the distance between the spheres

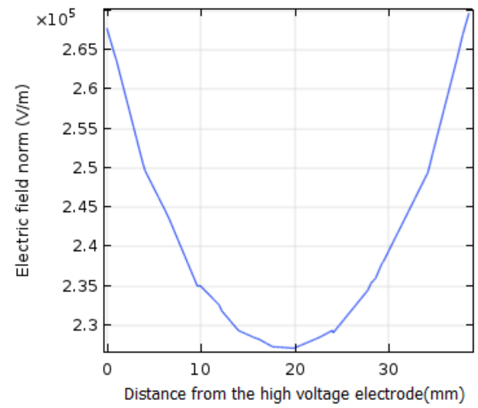


Fig. 6. Field distribution sample plot for the sphere–sphere geometry. The gap distance is 40 mm.

is sufficiently larger than the sphere diameter. In this work, as the selected gap lengths are less than sphere diameter, the field distribution obtained is mostly nonuniform for all the gap lengths. Similarly, the other features are also extracted by simulation. The proposed GPR model effectively takes care of the nonlinear variation in the features. Table I shows the gap distances and BDVs chosen for extracting the features for the preparation of the training dataset. The actual dataset includes hundreds of samples of the extracted features taken along the discharge path for each of the gap distances. The generation of the training dataset has been done by considering a total of 191 samples (total number of extracted original data for the sphere–sphere geometry) and 677 samples (total number of extracted original data for semispheroid–plane geometry) under impulses of either polarity. Known discharge voltages that have been generated for these gap distances experimentally are used as the target variable. A consolidated training dataset with gap distance as the sixth parameter is then prepared. For the sphere–sphere configuration, four different gap lengths (5, 15, 30, and 40 mm) and, for the semispheroid–plane configuration, three different gap lengths (5, 15, and 25 mm), which is shown in Table I, have been chosen for generating the training data. From the training data collected from experiments, it is observed that the discharge voltage of gaps for both the geometries is more for negative polarity impulses. The experimental results reveal that the discharge voltage of air is greatly influenced by the changes in the geometry and polarity of the waveform applied. The GPR model is trained with the prepared dataset for creating the model. For the prediction of the BDVs, features of the unknown gap geometries are extracted, a consolidated dataset is prepared and is given as the input to the model, and the results are obtained. Computations have been done using a regression learner in the MATLAB version R2018a.

A. Assessment of the Model

Cross-validation of a machine learning model has been adopted to estimate the effectiveness of the model in making predictions with the new dataset while developing the model. This is done by partitioning the dataset into training and test sets. k-fold cross-validation is done by dividing the dataset into

TABLE I

GAP LENGTH AND BDV VALUES USED FOR GENERATING THE TRAINING DATASET

Gap Geometry	Positive polarity		Negative Polarity	
	Gap Length (mm)	BDV (kV)	Gap Length (mm)	BDV (kV)
sphere-sphere	5	15.1	5	15.9
	15	46.2	15	47.6
	30	82.4	30	84.5
	40	102.4	40	105.8
semi-spheroidplane	5	12.1	5	13.1
	15	37.6	15	43.2
	25	55.7	25	64.6

TABLE II

GAP LENGTH VALUES USED FOR GENERATING THE VALIDATION TEST DATASET

Gap Geometry	Positive polarity			Negative Polarity		
	Gap Length (mm)	Exp	Predicted	Gap Length (mm)	Exp	Predicted
sphere-sphere	10	31.2	31.5	10	32.2	33
	20	56.8	59.1	20	58.2	60.6
	25	72.3	71.2	25	75.6	72.9
	35	94.1	92.8	35	97.8	95.5
semi-spheroidplane	10	27.6	27.2	10	30.8	30.9
	20	47.2	46.9	20	54.2	54.3

TABLE III

PERFORMANCE METRIC OF THE TRAINING DATA

Gap Geometry	Positive polarity			Negative Polarity		
	RMSE	MSE	MAE	RMSE	MSE	MAE
semi-spheroid plane	0.0079	0.000062	0.0021	0.01	0.0001	0.0078
Sphere-Sphere	0.024	0.0006	0.015	0.028	0.0008	0.0166

k partitions. In the first run, the first partition is taken as test data, and $k-1$ partitioned sets are used to train. In the next run, the second partition is taken as the test data, and with the other partitioned sets, training is done. This procedure is repeated for all the k partitions. The performance indices in terms of errors of the developed model are obtained by taking the average of cross-validation errors computed in all the iterations. In our work, the performance of the model has been assessed by fivefold cross-validation between the samples. i.e., the dataset is arbitrarily partitioned into five sets, and the training is done by using four out of five sets (80%) keeping one (20%) of the five partitioned sets as the test data. The model is obtained by storing all the results in a 4×1 cell array. The cross-validation errors computed are mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). The errors of the prediction model are described by (13)–(15)

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y})^2 \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y})^2} \tag{15}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} \mid y_i - \widehat{y} \mid \tag{16}$$

where y_i represents the actual output and $\widehat{y}_i$ represents the predicted value. A detailed description is given in [20]. Table II shows the gap length values used for generating the validation test dataset. The comparison of the performance indices of the developed model for the two geometries is shown in Table III for both polarities. Data given in the table indicate the comparison of the model errors with training samples. The closeness of fit is evident from the table, which shows very less values of MSE, MAE, and RMSE. It may be noted that the absolute values are less than 1 in all the cases for training data.

TABLE IV

PERFORMANCE METRIC OF THE VALIDATION TEST DATA

Gap Geometry	Positive polarity			Negative Polarity		
	RMSE	MSE	MAE	RMSE	MSE	MAE
semi-spheroid plane	1.98	3.93	1.65	0.01	0.1	0.1
Sphere-Sphere	1.43	2.07	1.25	2.17	4.74	2.05

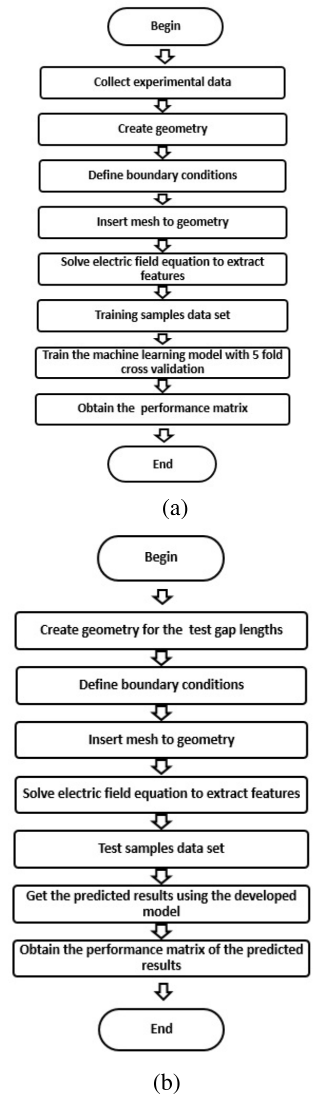
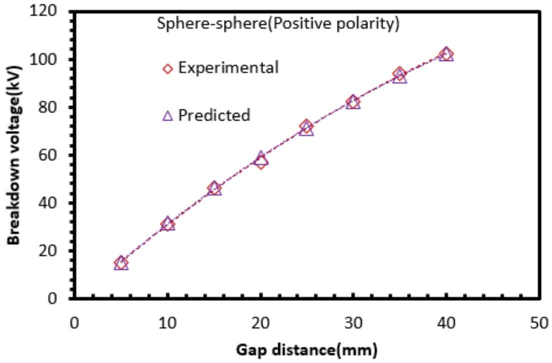


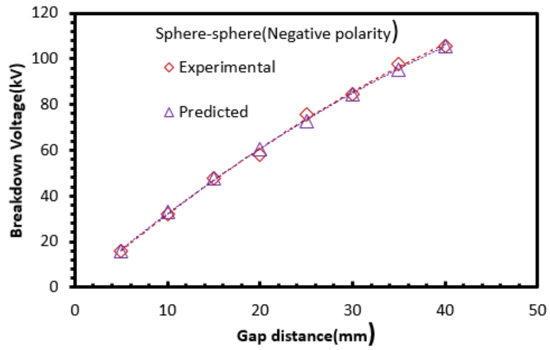
Fig. 7. Flowchart. (a) Training phase. (b) Validation phase.

B. Validation of Results

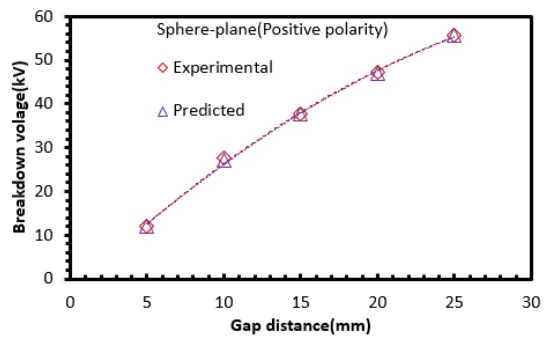
The consistency of the model is established through experimental validation. Simulations have been done with the geometries separately for the required gap distances, and new validation test datasets have been prepared. The datasets include the five electric field and energy features given in Section II-A. The input to the model includes these datasets along with the gap length. However, the dataset does not



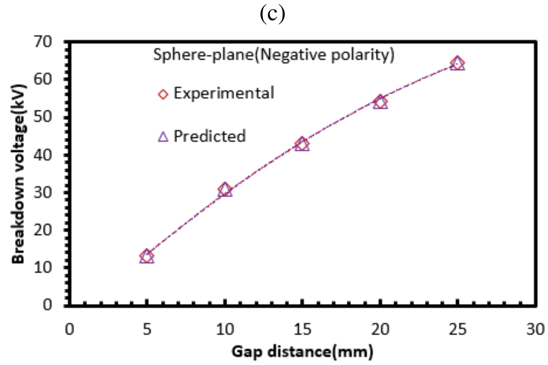
(a)



(b)



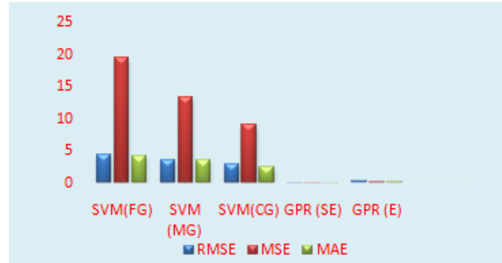
(c)



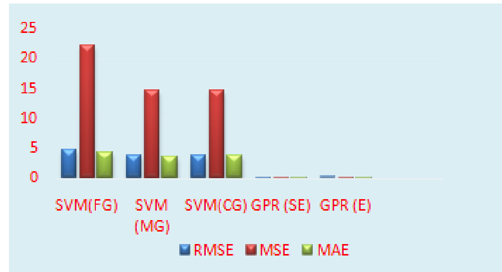
(d)

Fig. 8. Comparison between experimental and predicted BDVs. (a) For sphere–sphere positive polarity. (b) For sphere–sphere negative polarity. (c) For semispheroid–plane positive polarity. (d) For semispheroid–plane negative polarity.

include the target variable. BDV has been predicted for sphere–sphere configuration (10, 20, 25, and 35 mm) and semispheroid–plane configuration (10 and 20 mm) using the developed model. These gap lengths had been used to generate the validation test dataset. Table II shows the gap lengths

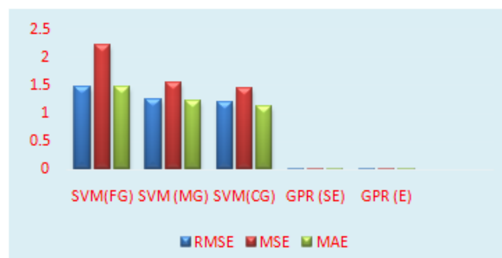


(a)



(b)

(c)



(d)

Fig. 9. Comparison of error indices. The GPR models show the least error. (a) Sphere–sphere positive. (b) Sphere–sphere negative. (c) Semispheroid–plane positive. (d) Semispheroid–plane negative.

chosen for generating the validation test dataset, and the experimental and predicted values of BDVs. It may be noted that the test data mentioned in this section are different from the test dataset considered while developing the model. For the validation of the results, gap lengths that have not been considered for training are taken. That is, it is the unknown gap lengths for the model. The values of the predicted BDVs have been compared with the data obtained from actual experiments. Table IV indicates the performance metric computed from the predicted results of the validation test dataset shown in Table II. From the table, it can be observed that the deviation in the predicted value and the actual data is very small. The predictions are in concurrence with the experimental data.

The flowcharts representing the training phase and validation phase are shown in Fig. 7. Necessary data for generating

TABLE V
COMPARISON OF PERFORMANCE INDICES OF GPR MODEL WITH OTHER MODELS

Parameters	Models				
sphere-sphere positive					
	SVMR(FG)	SVMR (MG)	SVMR(CG)	GPR(SE)	GPR(E)
RMSE	4.41	3.64	3.03	0.024	0.289
MSE	19.44	13.27	9.18	0.0006	0.083
MAE	4.13	3.56	2.51	0.015	0.071
sphere-sphere negative					
	SVMR(FG)	SVMR (MG)	SVMR(CG)	GPR(SE)	GPR(E)
RMSE	4.72	3.83	3.84	0.028	0.246
MSE	22.3	14.72	14.81	0.0008	0.06
MAE	4.41	3.75	3.8	0.0166	0.057
semi-spheroid-plane positive					
	SVMR(FG)	SVMR (MG)	SVMR(CG)	GPR(SE)	GPR(E)
RMSE	1.26	1.07	1.02	0.0079	0.011
MSE	1.59	1.15	1.04	0.000062	0.00013
MAE	1.24	1.05	0.966	0.0021	0.0059
semi-spheroid-plane negative					
	SVMR(FG)	SVMR (MG)	SVMR(CG)	GPR(SE)	GPR(E)
RMSE	1.48	1.24	1.2	0.01	0.015
MSE	2.21	1.54	1.45	0.0001	0.0002
MAE	1.46	1.22	1.13	0.0078	0.0089

TABLE VI
PERFORMANCE INDICES OF THE MODEL
(ROD–ROD AND ROD–PLANE)

Error	Rod-Plane	Rod-Rod
RMSE	1.11	3.3
MSE	1.23	10.9
MAE	0.142	0.43

the training dataset have been obtained by conducting in-house experiments for some gap distances. The feature extraction has been carried out by creating the required geometries in COMSOL Multiphysics. This involves applying boundary conditions that are specified in Section II. Extra fine meshing has been applied to the geometry to divide the solution space into discrete domains accurately. By solving the field equations, required features have been extracted from the geometry. From the features extracted, the training dataset has been prepared to train the model. Error indices are obtained to evaluate the performance of the model. Now, for the validation phase, gap lengths have been changed to the required value in the geometry created. After applying boundary conditions to the created geometric model, meshing is done. Features are extracted by solving the required electric field equations, and a consolidated validation test dataset has been prepared for each of the test gap lengths selected for validation. The prepared dataset does not include the target variable, which is the value of the BDV to be predicted. After getting the predicted results, the error indices of the validation test data have been obtained. Fig. 8(a)–(d) shows the comparison plots of the predicted results with experimental data. The closeness of fit is evident from the figures.

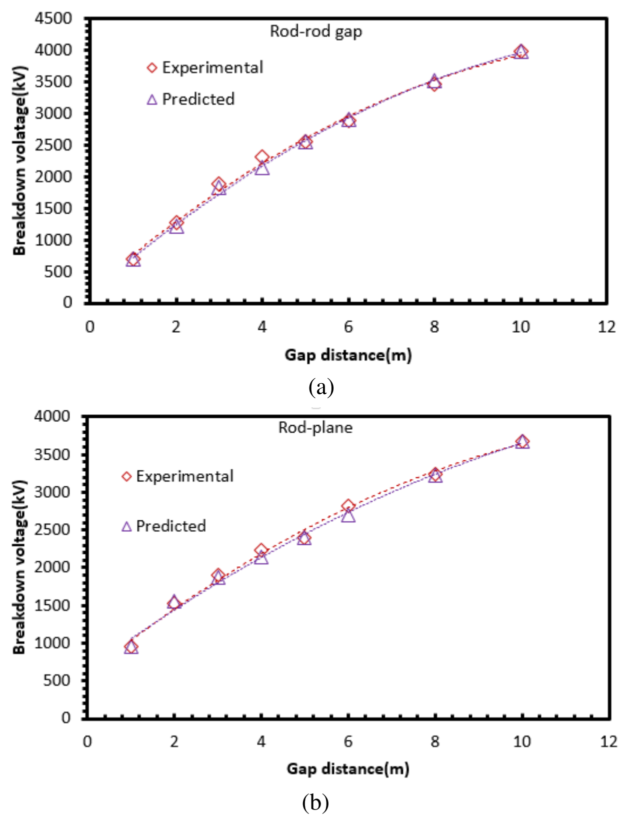


Fig. 10. Comparison of predicted and experimental results. (a) For long air gaps with the GPR model (rod–rod). (b) For long air gaps with the GPR model (rod–plane).

IV. COMPARISON WITH STATE-OF-THE-ART MODELS

The model is compared with the GPR model of the exponential Kernel function and also with other models, and verifies the efficacy. Other models chosen for comparison are SVM regression–fine Gaussian (FG), medium Gaussian (MG), and coarse Gaussian (CG). Table V shows the performance parameters of the GPR model with other models. It indicates the efficacy of GPR in terms of very small error values in prediction. The graphical comparison of results is shown in Fig. 9(a)–(d). It can be observed that the GPR models show the least error compared to other models for the chosen geometric configurations. The GPR model with the squared exponential kernel is marginally more accurate than exponential kernel and, hence, proves the best fit.

The GPR model performance is evaluated for long air gaps [12] using rod–plane and rod–rod geometries. The rod–rod and rod–plane geometries are created in COMSOL as the first step. The training dataset is prepared from the extracted features of 1-, 5-, and 10-m gap lengths. The BDVs have been predicted with the developed GPR model as per the procedure discussed in Section II for gaps of lengths of 2, 4, 6, 8, and 10 m for the rod–rod and rod–plane configurations. For the rod–rod geometry, 1763 samples, and rod–plane geometry, 1545 samples had been used for training. From the various evaluation metric shown in Table VI, it is evident that the proposed model is equally pertinent to long air gaps also. The comparison of predicted results with experimental data is shown in Fig. 10.

V. Conclusion

In this work, a GPR-based machine learning model for the prediction of discharge voltage of air under lightning impulses of both polarities has been developed. Since a machine learning model demands a large number of inputs for accurate prediction, FEM-based feature extraction is adopted to effectively increase the input parameters of the model. The extracted features, namely, electric field strength (E), electric energy density (E_{we}), electric potential (V), current density (J), and total electric energy (E_{twe}) of the gaps and corresponding gap lengths, are given as the input parameters of the model. From the extracted features using the FEM method, it is perceptible that the discharge characteristics of different electrode geometries with applied voltage vary in a stochastic manner. It is found that the proposed model effectively predicts the BDV of air for different gap geometries and gap lengths. The model is validated using experimental tests. The predicted results using the model show good concurrence with test results. Accurate predictions have been achieved with a reasonable number of input parameters. The effectiveness of the model is established by comparison with GPR models of different kernel functions and with the other state-of-the-art models. The results show the efficacy of the model with very small values of error indices. The GPR model when applied to long air gaps (rod–plane and rod–rod) is found to give promising results.

References

[1] I. A. Metwally, "Technology progress in high-voltage gas-insulated substations," *IEEE Potentials*, vol. 29, no. 6, pp. 25–32, 2010.

[2] P. N. Mavroidis, P. N. Mikropoulos, and C. A. Stassinopoulos, "Discharge characteristics in short rod-plane gaps under lightning impulse voltages of both polarities," in *Proc. 42nd Int. Universities Power Eng. Conf.*, Sep. 2007, pp. 1070–1074.

[3] J. H. Rakotonandrasana, A. Beroual, and I. Fofana, "Modelling of the negative discharge in long air gaps under impulse voltages," *J. Phys. D, Appl. Phys.*, vol. 41, no. 10, May 2008, Art. no. 105210.

[4] M. A. Douar, A. Beroual, and X. Souche, "Creeping discharges features propagating in air at atmospheric pressure on various materials under positive lightning impulse voltage–Part 2: Modelling and computation of discharges' parameters," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 6, pp. 1429–1437, 2017.

[5] K. Hotta *et al.*, "Streamer development mechanism under non-uniform electric field in air," in *Proc. Int. Conf. High Voltage Eng. Appl.*, Sep. 2012, pp. 501–504.

[6] H. Kojima *et al.*, "Classification of impulse breakdown mechanisms under non-uniform electric field in air," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 1, pp. 194–201, Feb. 2016.

[7] M. Darveniza and A. E. Vlastos, "The generalized integration method for predicting impulse volt-time characteristics for non-standard wave shapes–A theoretical basis," *IEEE Trans. Elect. Insul.*, vol. 23, no. 3, pp. 373–381, Jun. 1988.

[8] A. Ancajima, A. Carrus, E. Cinieri, and C. Mazzetti, "Behavior of MV insulators under lightning-induced overvoltages: Experimental results and reproduction of volt–time characteristics by disruptive effect models," *IEEE Trans. Power Del.*, vol. 25, no. 1, pp. 221–230, Jan. 2010.

[9] S. Mohanty and S. Ghosh, "Artificial neural networks modelling of breakdown voltage of solid insulating materials in the presence of void," *IET Sci., Meas. Technol.*, vol. 4, no. 5, pp. 278–288, Sep. 2010.

[10] S. S. M. Ghoneim, S. S. Dessouky, A. A. Elfaraskoury, and A. B. A. Sharaf, "Prediction of insulating transformer oils breakdown voltage considering barrier effect based on artificial neural networks," *Electr. Eng.*, vol. 100, no. 4, pp. 2231–2242, Dec. 2018, doi: 10.1007/s00202-018-0697-5.

[11] Z. Qiu *et al.*, "Hybrid prediction of the power frequency breakdown voltage of short air gaps based on orthogonal design and support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 2, pp. 795–805, Apr. 2016.

[12] Z. Qiu, J. Ruan, W. Xu, and C. Huang, "Energy storage features and a predictive model for switching impulse flashover voltages of long air gaps," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 5, pp. 2703–2711, Oct. 2017.

[13] Z. Qiu, J. Ruan, W. Xu, X. Wang, and D. Huang, "Electrostatic field features on the shortest interelectrode path and a SVR model for breakdown voltage prediction of rod–plane air gaps," *IET Sci., Meas. Technol.*, vol. 12, no. 7, pp. 886–892, 2018.

[14] V. Q. Huynh, B. Techaumnat, and K. Hidaka, "Analysis on electrostatic behavior of a conducting prolate spheroid under an electric field," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 6, pp. 2230–2238, Dec. 2013.

[15] E. Kuffel, W S Zaengl, and J. Kuffel, *High Voltage Engineering Fundamentals*. Amsterdam, The Netherlands: Elsevier, 2000.

[16] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[17] M. Hikita, S. Ohtsuka, N. Yokoyama, S. Okabe, and S. Kaneko, "Effect of electrode surface roughness and dielectric coating on breakdown characteristics of high pressure CO_2 and N_2 in a quasi-uniform electric field," *IEEE Trans. Dielectr. Elect. Insul.*, vol. 15, no. 1, pp. 243–250, Feb. 2008.

[18] *High-Voltage Test Techniques Part 1: General Definitions and Test Requirements*, European Standard, Standard IEC 60060-1, 2010.

[19] E. Schulz, M. Speekenbrik, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.

[20] G. Drakos. (2018). *How to Select the Right Evaluation Metric for Machine Learning Models: Part 1 Regression Metrics*. Towards Data Science. Saatavissa. Accessed: Mar. 2019. [Online]. Available: https://towardsdatascience. com/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-1-regrression-metrics-3606e25beae0