

A Random Forest Model Based Pollution Severity Classification Scheme of High Voltage Transmission Line Insulators

K.Kannan[†], R.Shivakumar* and S.Chandrasekar*

Abstract – Tower insulators in electric power transmission network play a crucial role in preserving the reliability of the system. Electrical utilities frequently face the problem of flashover of insulators due to pollution deposition on their surface. Several research works based on leakage current (LC) measurement has been already carried out in developing diagnostic techniques for these insulators. Since the LC signal is highly intermittent in nature, estimation of pollution severity based on LC signal measurement over a short period of time will not produce accurate results. Reports on the measurement and analysis of LC signals over a long period of time is scanty. This paper attempts to use Random Forest (RF) classifier, which produces accurate results on large data bases, to analyze the pollution severity of high voltage tower insulators. Leakage current characteristics over a long period of time were measured in the laboratory on porcelain insulator. Pollution experiments were conducted at 11 kV AC voltage. Time domain analysis and wavelet transform technique were used to extract both basic features and histogram features of the LC signal. RF model was trained and tested with a variety of LC signals measured over a lengthy period of time and it is noticed that the proposed RF model based pollution severity classifier is efficient and will be helpful to electrical utilities for real time implementation.

Keywords: Insulator, Flashover, Pollution, Discrete wavelet transform, Random forest model

1. Introduction

In today's world of ever increasing power demand from public and private sectors alike, it is necessary to preserve the reliability of electric power transmission system by carrying out proper preventive maintenance works. Tower insulators in electric power transmission network play a crucial role in preserving the reliability of the system. Flashover of polluted insulators in a transmission line is a major trouble faced by electrical utilities, which results in major economical losses. In general, these insulators are normally used for outdoor applications and hence deposition of pollution on the insulator surface is unavoidable. Under wet surface conditions, due to fog or drizzle, pollution layer becomes conducting and causes the flow of leakage current (LC) which finally leads to flashover. Hence many research works are in progress towards understanding the severity of pollution in transmission tower insulators both in the laboratory level and in the on-field level [1-3].

Various approaches were used to diagnose the pollution severity of transmission tower insulators [4-16]. Partial discharge measurements have been carried out in many research papers, but it is difficult to separate the partial discharges occurring on the surface of the insulators from

the corona discharges in the transmission line and external interferences which occurs in the same frequency band of original partial discharge signal [4-7]. Several works discuss the investigation of LC signals of insulators towards pollution severity studies [9-16]. T. Suda [13], classified the LC signals of cap and pin insulator at different pollution conditions into six stages and predicted the occurrence of flashover. M.Ugur et al., [14] classified the surface degradation characteristics of polymeric insulation structure during tracking resistance test in the laboratory. Features of LC signals were extracted using FFT approach and artificial neural network was used for classification. S. Chandrasekar et al., [16] investigated the surface pollution condition of silicone rubber insulator using multi resolution signal decomposition analysis of discrete wavelet transform (DWT) technique.

Since the LC signal is vastly irregular in nature and surface wetness depends on many environmental factors, it is very difficult to decide the surface pollution severity based on the measurement and analysis of individual LC signals over a short period of time. In order to perform better estimate of the pollution severity, it is necessary to measure LC signals over a long period of time at different pollution conditions in the laboratory. In this connection, few reports are only available in the literature regarding the collection and analysis of LC signals over a long period of time [1, 16]. Also, in a real time system, collection of LC signals over a lengthy period will result in huge amount of data, which makes inconvenience in the design of better

[†] Corresponding Author: Electrical Engineering, Anna University, Chennai, India. (kannankst@gmail.com)

* Dept. of Electrical and Electronics Engg., Sona College of Technology, Salem, Tamilnadu, India. (shivakumar@sonatech.ac.in, chandrukvt@gmail.com.)

Received: October 15, 2015; Accepted: January 7, 2016

flashover monitoring system.

Considering the importance of developing a better diagnostic tool to exactly predict the pollution severity of outdoor insulators, in the present work, an advanced classification technique, Random Forest (RF) method has been adopted. RF method has the capability to manage large data bases without any deletion of even a single input variable. It can be trained very fast and it has the capacity to handle large number of input variables [17-21]. DWT overcomes the drawbacks of FFT and it is well suited for time-frequency analysis of LC signals. Even under noisy environment, DWT technique gives accurate results of transient analysis of signals [10, 12]. Hence in this work, DWT technique has been used to extract salient features of the LC signals at different pollution levels. Basic features and histogram features are extracted from the time-frequency domain analysis and given as an input to the RF classifier.

2. Laboratory Experimental Setup

Porcelain insulators of 11 kV rating with shed diameter 254 mm and leakage distance 330 mm were used for pollution experiments. Laboratory experiments were conducted as per IEC 60507 test procedures [22]. Fig. 1 shows the picture of cap and pin porcelain insulator and laboratory setup used for pollution studies.

The single disc porcelain insulator was suspended vertically inside the fog chamber as shown in Fig. 2.

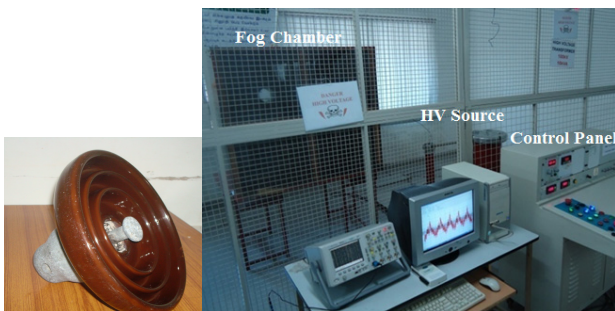


Fig. 1. Picture of porcelain insulator and laboratory experimental setup

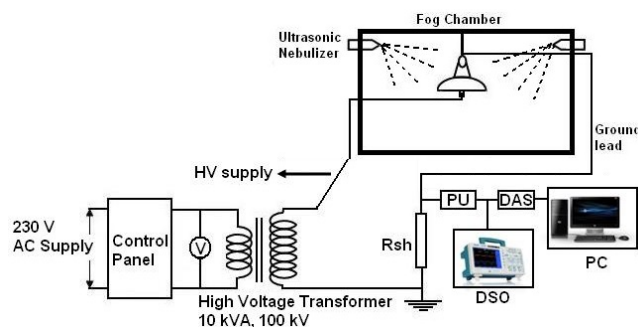


Fig. 2. Laboratory experimental set up

Initially, the insulator was cleaned with isopropyl alcohol and rinsed with distilled water to remove any trace of dirt and grease [16]. Insulator specimen was applied with 11 kVrms using a transformer of 100 kV, 10 kVA rating which is connected with a control panel to set the required voltage. To reproduce the saline pollution typical of coastal areas, pollution slurry was prepared using NaCl of required amount added with 40g of kaolin in 1 litre of distilled water. The amount of NaCl added in the pollution slurry was varied from 10 g to 100 g to reproduce Equivalent Salt Deposit Density (ESDD) in mg/cm^2 to 0.06 (lightly polluted-LP), 0.08 (moderately polluted-MP), 0.12 (heavily polluted-HP) and 0.25 (very heavily polluted-VHP). In order to maintain the wet condition of insulator, relative humidity inside the fog chamber was controlled using four ultrasonic nebulizers and a humidifier. A wall-mount hygrometer (HTC-1) instrument was used to measure the relative humidity [5]. A glass window was also provided on the fog chamber to monitor the discharges occurring on the insulator surface.

LC measuring system consists of a ground lead shunt resistance along with a protection unit connected with a data acquisition system as shown in Fig. 2. National Instruments USB 6251 data acquisition system was used in this work which has the capacity to capture LC signals at 1.25 MSa/sec. Sampling frequency of 5 kHz was maintained for the entire work. A LabVIEW software system was developed in the PC to store the entire LC signals and therefore it will be useful for further feature extraction analysis [1]. Agilent DSO of sampling rate 1 GSa/sec was also used to notice the LC signals.

3. Discrete Wavelet Transform

Time-frequency analysis of LC signals will provide valuable information to develop an effective condition monitoring system for transmission tower insulators. DWT is an efficient technique to extract salient time-frequency features from the LC data [10, 12]. Consider a discrete signal $c_0[n]$ captured from the experimental studies. Using multi resolution signal decomposition approach of DWT, $c_0[n]$ is decomposed into two different signals namely approximate signal $c_1[n]$ and detailed signal $d_1[n]$ in the first level of decomposition. Decomposition is carried out using quadrature mirror filters such as h and g as given by equation 1 and 2. In the next level, $c_1[n]$ is again decomposed into $c_2[n]$ and $d_2[n]$. This process continues until the entire frequency band of the original signal is covered. In the present work, since the sampling frequency is 5 kHz, decomposition is carried out upto 7 levels as shown in Table 1 [16]. The DWT multi resolution signal decomposition is given by,

$$c_m[n] = \sum_k h[k-2n]c_{m-1}[k] \quad (1)$$

Table 1. Frequency range of decomposed signal

Detailed components of DWT	Frequency bandwidth (Hz)
D1	1250-2500
D2	625-1250
D3	312.5-625
D4	156.25-312.5
D5	78.125-156.25
D6	39.0625-78.125
D7	19.53125-39.0625

$$d_m[n] = \sum_k g[k-2n]c_{m-1}[k] \quad (2)$$

where m is the scale of decomposition, n is the sampling points and k is translation coefficient. Daubechies 4 mother wavelet is used in this analysis [10, 12]. Energy content of the signal at different decomposition level is evaluated using standard deviation value [16]. Distortion ratio (DR), which is the ratio of standard deviation values of (D3+D4+D5) to fundamental component D6 [16] is also calculated and considered as an important feature in this work.

4. Concept of Random Forest Algorithm

Random forest (RF) is an efficient ensemble learning approach with exceptional performance in data exploration, predictive modeling and classification tasks [17-21]. RF is one of the most precise classifying algorithm available for large data bases. RF makes use of a forest of classification trees built on the bootstrap sample of the data. RF utilizes both bagging and random variable selection for tree building. Another interesting feature of RF method is the possibility to estimate important statistics from the left out samples such as out of bag estimate which is a measure of generalization error of the forest. RFs can be run either in unsupervised mode or in supervised mode. RF runs as a classifier in the supervised mode and it measures the importance and ability of all the variables for proper classification of given data. In RF approach, each tree is grown on different random subsamples and splitter party is determined at random.

RF algorithm is implemented as shown in Fig. 3. The training set for each tree is selected by bootstrap aggregating method, also called bagging. The training

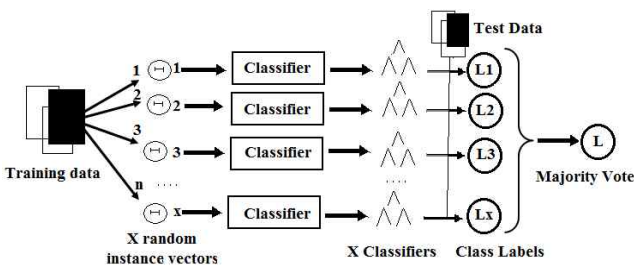


Fig. 3. Schematic representation of RF classifier

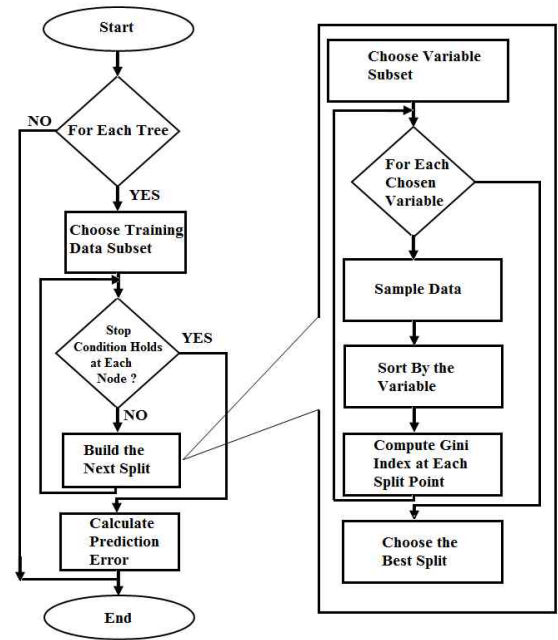


Fig. 4. Flow chart of construction process of trees random forest classifier

subset (X random instance vectors) is sampled from whole training dataset uniformly and with replacement [19]. Typically, it accounts for 63.2% of the whole dataset. Each tree is grown to as large size as possible without pruning.

Fig. 4 shows the flow chart of construction process of each tree in RF algorithm. Consider ‘ N ’ be the number of training and ‘ M ’ be the number of variables in the classifier. If m of input variables to be used to determine the decision at a node of the tree; m should be much less than M . Choose a training set for this tree by choosing n times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes. For each node of the tree, randomly choose m variables on which to base the decision at that node. Based on these m variables in the training set, best split is calculated [20].

In the case of classification process, a new sample is sent down the tree and vote of each tree is calculated. This process continues over all trees in the ensemble and based on the majority of vote of all trees RF algorithm gives the output result. Splits are chosen according to Gini Impurity Measure. The optimal feature chosen to split a node is the one that can generate a child node with lowest Gini impurity, $I(n)$.

$$I(n) = 1 - \sum_{j=0}^n p(\frac{j}{t}) \quad (3)$$

where $p(j/t)$ is the probability of a specific class within the child node. Gini impurity should be zero if the node is all one class. In normal decision tree, the optimal feature to split a node is selected from all features. However, for

random forests, the optimal feature is selected from a random subset of features. Typically, for a dataset with p features, \sqrt{p} features are used in each split. Based on the majority vote of each decision tree, output of RF classifier is determined.

5. Results and Discussion

5.1 Time domain analysis of LC signals

First of all, it is necessary to understand the variations in shape and magnitude of the LC signal at different pollution levels. Hence, tests were conducted with increasing pollution level at 11 kVrms, 95% relative humidity and the corresponding LC signals were captured.

Fig. 5 (a,b,c,d) shows the typical LC signal obtained with increase in pollution level from 0.06 ESDD (LP) to 0.25 ESDD (VHP). From Fig. 5a, it can be easily understood that during the initial stages of low pollution, the LC follows close to sinusoidal path. Fig. 5b and 5c shows presence of short time discharges with rise in pollution, which causes LC waveform distortions. When the pollution level reaches 0.12 ESDD (HP), multiple discharges in the LC waveform are clearly visible. At 0.25 ESDD, sudden increase in LC magnitude with a near sinusoidal pattern is noticed. These results confirms that increase in pollution level above 0.08 ESDD at 95 % relative humidity conditions, results in the flow of more leakage current with higher magnitude [16]. Fig. 6 (a, b) shows the photographs of insulator surface with arcing at different pollution levels.

It is noticed that LC flow is highly intermittent in nature and hence it is difficult to accurately classify the pollution level of insulator by performing time domain analysis of the trend followed by the LC peak or LC mean value. At highly polluted conditions, presence of large number of peaks and valleys are noticed, which makes it difficult to exactly predict the pollution level of insulator at a

particular time period [16]. It is necessary to perform both time and frequency domain analysis of LC signals in order to estimate the pollution severity of the insulators. Therefore, in the following sections, time-frequency analysis of LC signals using DWT technique is discussed.

5.2 Time-frequency domain analysis of LC signals

In order to understand the trend followed by DWT detailed coefficients, 500 LC signals were captured continuously at each pollution level and used for DWT analysis. Fig. 7 (a, b, c, d) shows the variations of DWT detailed coefficients at different pollution levels.

During lightly polluted conditions, D5 component is somewhat less than other components such as D1, D2 and D3 (Fig.7 a). During arcing formation rise in magnitude of D5 is observed, compared with D3, D2 and D1 (Fig. 7 b,c,d). Analysis of Fig. 7 also shows that whenever there is a formation of discharge due to rise in pollution on the wet surface, then corresponding rise in magnitude of D6 is observed (D6 corresponds to fundamental frequency of LC signal). This obviously shows that D5 and D6 has positive correlation. Therefore D5 and D6 components of DWT pattern are considered as an important features to identify pollution severity of insulators.

Similarly, DR values are calculated for each pollution

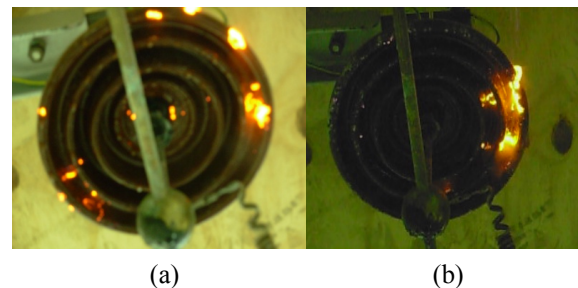


Fig. 6. Photographs of arcing on the surface of insulator with increase in pollution level: (a) MP; (b) HP

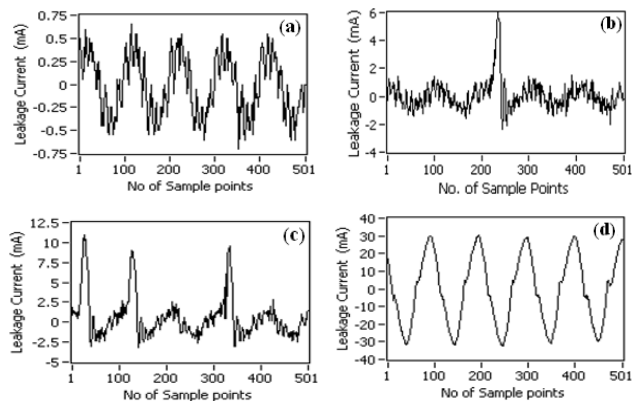


Fig. 5. Typical leakage current signals obtained during experimental study: (a) LP; (b) MP; (c) HP; (d) VHP

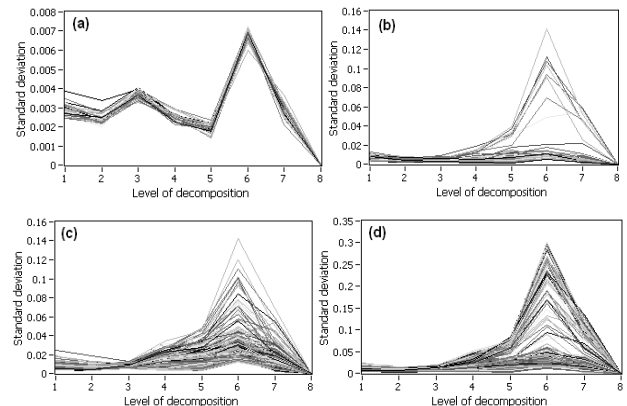


Fig. 7. Variations of DWT coefficients of LC signals captured over a long period of time at different pollution levels: (a) LP; (b) MP; (c) HP; (d) VHP

level and a typical DR plot is shown in Fig. 8 corresponding to medium and highly polluted cases. Whenever there is an arc formation due to increase in pollution, then DR value also increases above 60. However, when severe long arcs are formed due to high pollution, then the number of DR values below 40 range increases significantly [16]. From the Figs. 7 and 8, large variations in D5, D6 and DR value is noticed and hence it is understood that if only a few samples of LC signals over a short period of time are measured, then it is very difficult to predict accurate results.

5.3 Overall schematic of RF based classifier

Fig. 9 shows the overall schematic of the proposed RF based pollution severity classification system. In this approach, LC signal is continuously monitored for 5 hours

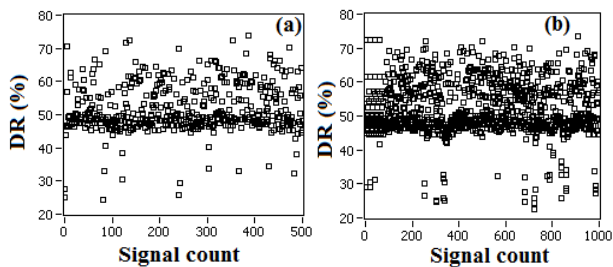


Fig. 8. Typical variations of DR values of LC signals captured over a long period of time at different pollution levels: (a) MP; (b) HP

at each pollution level and stored in PC. As discussed earlier, time domain analysis of LC signals is carried out to evaluate the peak value and mean value of each LC waveform. Similarly DWT analysis is carried out to evaluate the D5, D6 and DR values of each LC waveform. Then basic features such as minimum, maximum, mean and percent values above mean are extracted from the large data base of LCpeak, LCmean, D5, D6 and DR values captured for 5 hours. Similarly histogram features are extracted from the data base. Extracted features are given as an input to the RF classifier algorithm, which further classifies the pollution level of insulator as light, medium, heavy and very heavy pollution.

5.4 Extraction of basic features at various pollution levels

Two kinds of features are extracted from the time domain and DWT analysis, including basic features and histogram features. Basic features, such as minimum value, maximum value, mean value and percentage of higher margin values above mean value are extracted from the 5 hours continuous measurement data of LCpeak, LCmean, D5, D6 and DR parameters at each pollution level. In this study, 20 basic features (Four values such as min, max, mean, % above mean values for all 5 parameters) are extracted at each measurement.

Fig. 10 shows the variations of minimum, maximum and mean values of each parameter such as D5, D6, DR, LCpeak and LCmean. In general, it is noticed that there is a

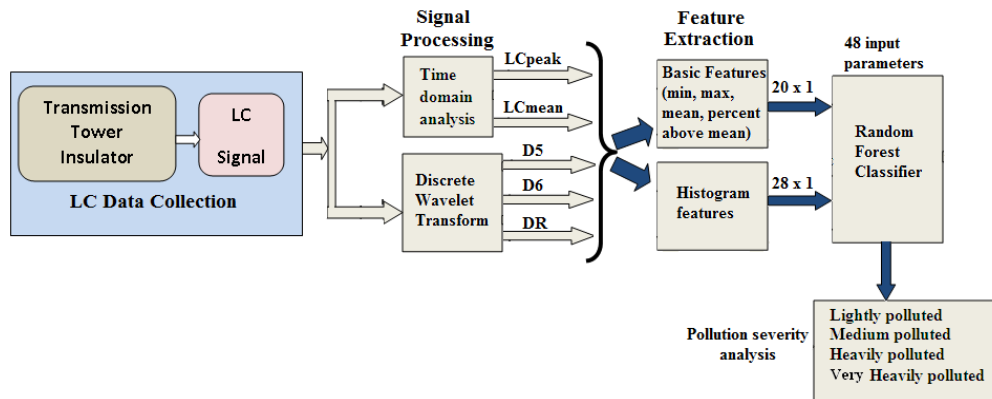


Fig. 9. Overall schematic diagram of the proposed RF based pollution severity classifier

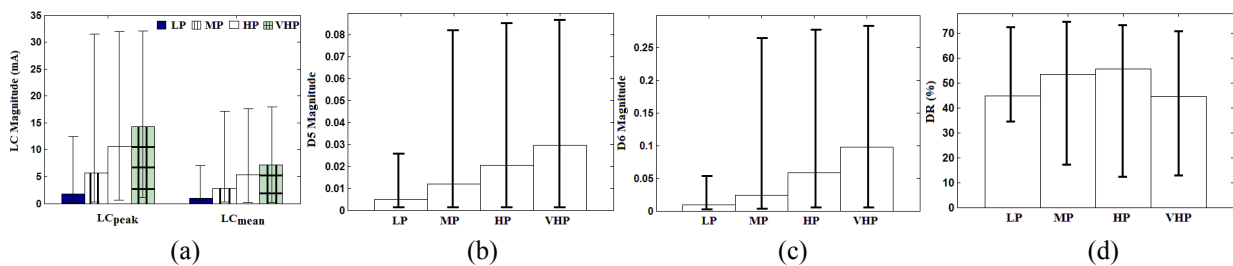


Fig. 10. Variations of minimum, maximum and mean values of (a) LCpeak and LCmean; (b) D5; (c) D6; (d) DR at different pollution conditions.

huge margin between the maximum values and mean values of each parameter at each pollution condition, which shows that the parameters are highly scattered. Fig. 11 shows the percentage of values above mean value of each parameter at different pollution conditions. When compared with lightly polluted conditions, percentage of values above mean value increases at high pollution levels. This strongly shows the presence of large number of discharges during highly polluted conditions.

5.5 Extraction of histogram features at various pollution levels

In order to develop a better diagnostic system, it is necessary to understand the distribution of features of LC data over a period of time. Hence histogram features are extracted from the 5 hours continuous measurement data of LCpeak, LCmean, D5, D6 and DR parameters at each pollution level. Width of histogram bins are selected according to the sensitivity of each parameter at both lower and higher amplitude areas. LCpeak is divided into 5 different bins, LC mean and D6 are divided into 6 different bins, D5 is divided into 4 different bins and DR is divided

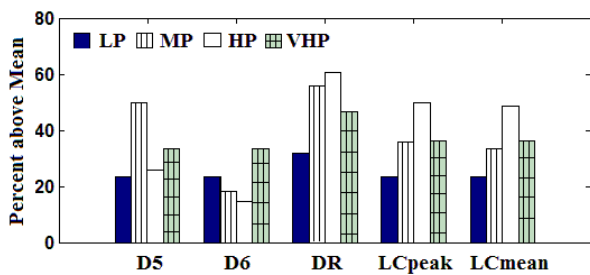


Fig. 11. Percentage of higher margin values above mean value of LCpeak, LCmean, D5, D6 and DR at different pollution conditions.

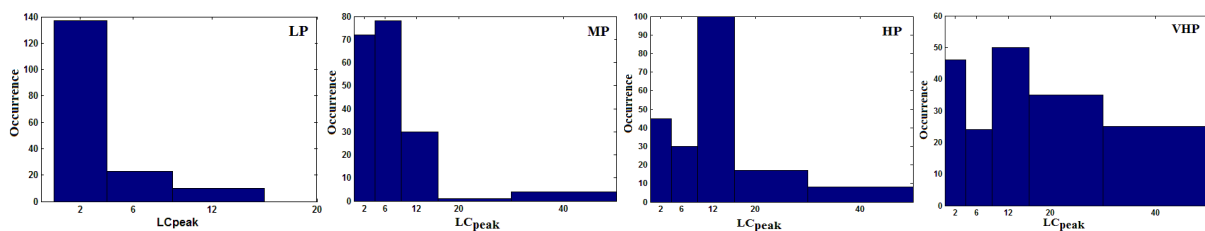


Fig. 12. Amplitude histograms of LCpeak at different pollution conditions.

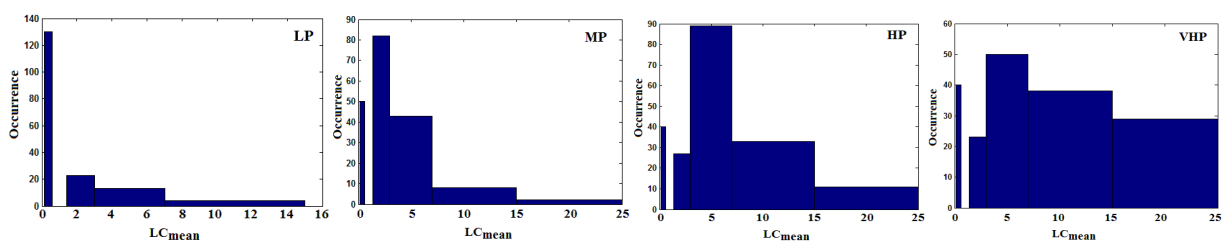


Fig. 13. Amplitude histograms of LCmean at different pollution conditions.

into 7 different bins.

Fig. 12 to 16 shows the histograms of LCpeak, LCmean, D5, D6 and DR values at different pollution conditions respectively. Large variations in histogram bin values at each pollution condition is noticed. Total 28 bin values are evaluated from these five parameters and they are considered as histogram features of LC data.

5.6 Pollution severity classification results of RF classifier

After collecting all the basic and histogram features, they were integrated into one combined feature set consisting of 48 features and given as an input to the RF classifier. MATLAB software was used to perform RF classifier analysis. Table 2 shows the various parameters used in the implementation of RF classifier algorithm. There are two major parameters of RF classifier, one is 'n_{tree}' representing the number of trees and another one is 'm_{try}' representing the number of variables used in determining the best split at each node. Leo Breiman [17] explained that the optimal value of m_{try} usually lies between log₂(n) and \sqrt{n} where n is the number of features. Therefore m_{try} is tested in the range of 5 to 7 and the optimal value of 6 is selected which gives less out of bag (OOB) error rate. Minimum size of terminal nodes is selected as 3 so that it causes smaller trees to be grown and

Table 2. Training parameters of RF classifier

No. of feature inputs	48
No. of outputs	4
No. of trees n _{tree}	500
No. of variables for best split at each node (m _{try})	6
Minimum node size	3
No. of training set	350
No. of test set	150

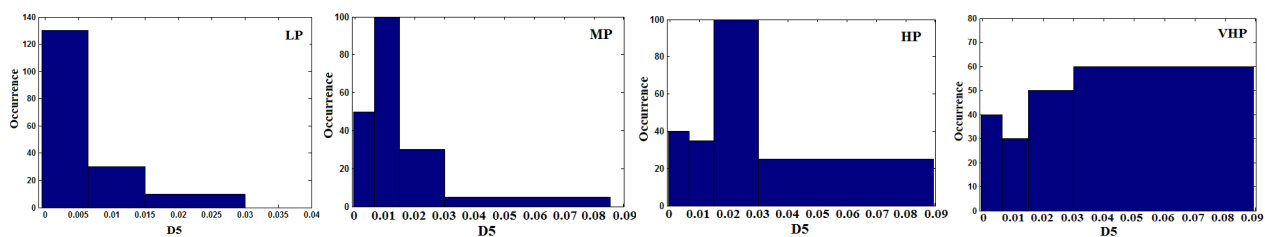


Fig. 14. Amplitude histograms of D5 value of DWT at different pollution conditions.

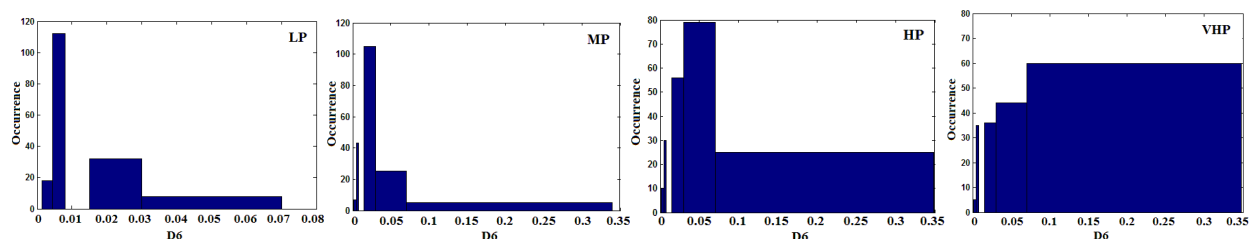


Fig. 15. Amplitude histograms of D6 value of DWT at different pollution conditions.

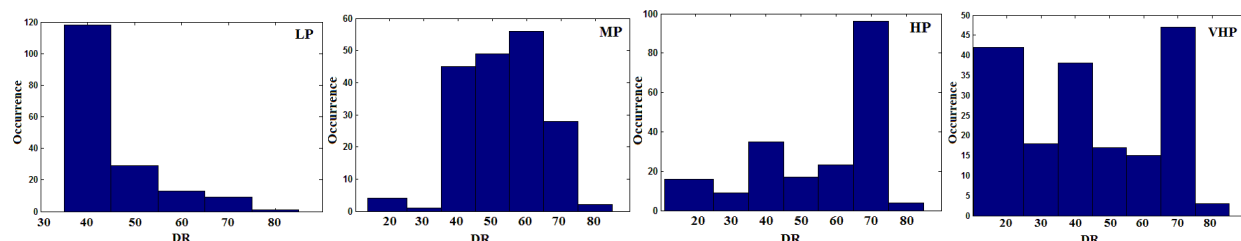


Fig. 16. Amplitude histograms of DR value of DWT at different pollution conditions.

Table 3. Trained output pattern

Type of output	Output pattern
LP	1
MP	2
HP	3
VHP	4

thus takes less time. Number of trees (n_{tree}) is varied upto 500. Table 3 shows the trained output pattern for each pollution condition.

In order to validate the trained RF classifier, instead of using a separate test set data, error can be estimated internally in RF [18]. Different bootstrap sample is used in the construction of each tree and close to $1/3^{rd}$ of the total samples are left out of the bootstrap sampling and they are not used in the construction of the tree, so the left out samples can be put into the tree as test samples. Total 500 data set is chosen and out of that 350 data set is used for training and remaining 150 used for testing purpose. By calculating the proportion of misclassified samples over all cases, it is possible to get the OOB error estimation, which has been proved to be an unbiased error estimation method for RF. Fig. 17 shows the OOB error rate of the RF method for different forest sizes. It is observed that the number of trees in the forest should be more than 70 in order to get a stable zero OOB rate.

While evaluating the OOB error, RF method also finds a

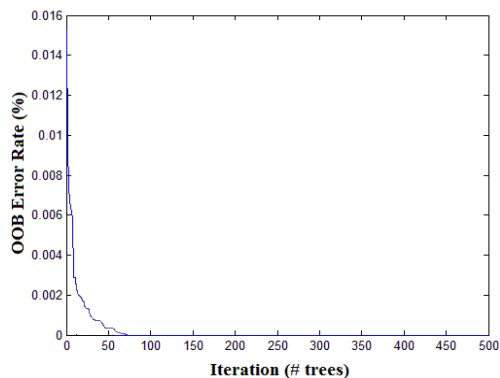


Fig. 17. Out of bag error rate of RF method for different forest sizes.

parameter namely ‘mean decrease in accuracy’. Fig. 18 shows the variations in mean decrease in accuracy values of entire feature set. This parameter is a measure of importance of each variable in the feature set, i.e. if the accuracy of RF decreases highly due to the elimination of a particular variable in the feature set, then that particular variable is very important for classification. Hence, features with higher mean decrease in accuracy are very important for accurate classification of given data.

In the RF algorithm, whenever a node is split using a particular feature variable, then the corresponding Gini index for each child nodes are evaluated, which varies from

0 to 1. If the value of Gini index is '0' it indicates homogeneous case and '1' indicates heterogeneous case. Gini index value is again compared with the original node. Then the contribution of each feature variable for the homogeneity of the RF nodes is evaluated using the parameter namely 'mean decrease in Gini index'. Fig. 19 shows the mean decrease in Gini index of entire set of feature data. The changes obtained in Gini index are summed for each feature and normalized at the end of the calculation. Large decrease in Gini index of a feature is a measure of higher purity of the nodes.

Table 4 shows the identification results of RF classifier model. In order to understand the performance of RF classifier, identification results are checked with different number of trees and also with different feature set. For example, number of trees are varied from 50 to 500.

Table 4. Identification results of RF classifier

No. of Trees	Feature set	Identification rate (%)			
		LP	MP	HP	VHP
50	Basic (20)	84	82	81	82
100	Basic (20)	84	82	82	82
200	Basic (20)	85	83	82	83
500	Basic (20)	85	84	82	83
50	Histogram (28)	87	86	86	85
100	Histogram (28)	88	86	87	86
200	Histogram (28)	90	88	89	87
500	Histogram (28)	90	88	89	87
50	Basic + Histogram (48)	98	96	98	97
100	Basic + Histogram (48)	99	97	98	97
200	Basic + Histogram (48)	99	99	99	98
500	Basic + Histogram (48)	100	100	99	98

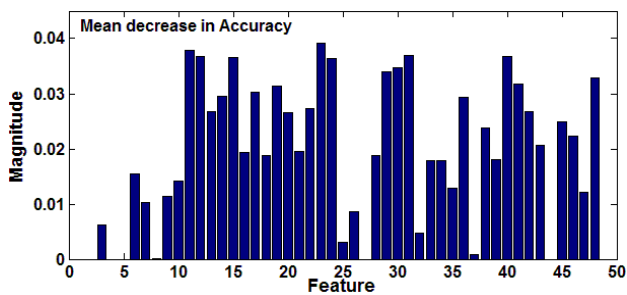


Fig. 18. Mean decrease in accuracy value of RF method for entire feature set.

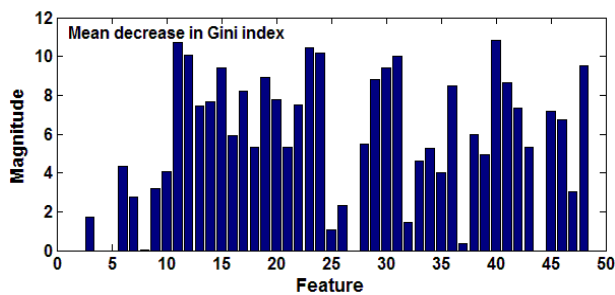


Fig. 19. Mean decrease in Gini index of RF method for entire feature set.

Similarly RF performance is evaluated with basic features, histogram features and combined basic and histogram features. When the number of trees increases above 50, then considerable increase in accuracy of identification results is noticed. Similarly when combined basic and histogram features are given as an input to the RF classifier, then significant increase in accuracy of identification rate is observed. When the RF classifier is trained with more than 200 trees with combined feature set, identification rate of the RF classifier lies above 98 % at all pollution conditions.

The majority votes secured by RF classifier at each pollution condition denotes the accuracy of classification. Fig. 20 shows the percentage of majority votes secured by RF classifier at different pollution conditions. Percentage values are calculated from the votes obtained at 500 trees. Fig. 20(i) shows the votes obtained at lightly polluted data

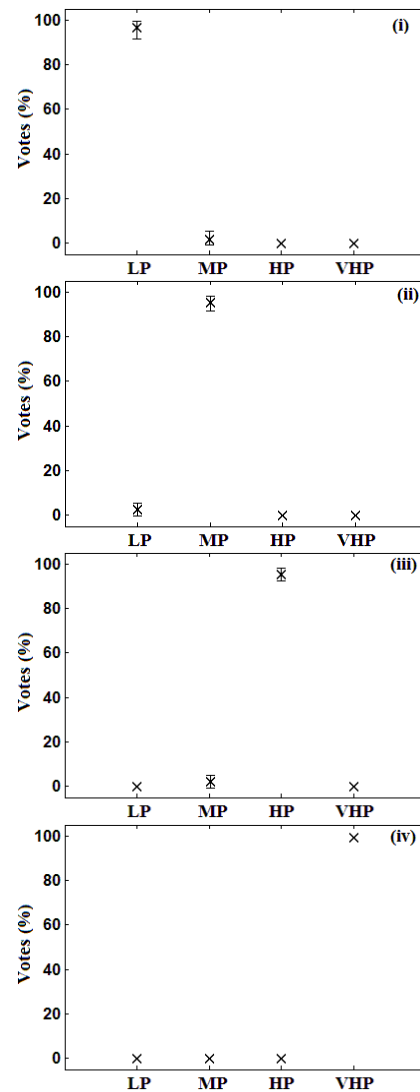


Fig. 20. Mean percentage of majority votes secured by RF classifier at each polluted case with respective minimum and maximum values (i) LP (ii) MP (iii) HP (iv) VHP

where percentage votes secured by HP and VHP is zero, whereas percentage of votes of MP lies below 5. It clearly shows more than 95 % of votes are secured by LP. Fig. 20(iv) shows the results of very high polluted case, in which 100 % of votes goes to VHP, whereas all other cases are 0%. It clearly shows the accuracy of RF classifier and there is no ambiguity in classification.

In many earlier works [10, 12, 14], mainly artificial neural networks are used for classification purpose and the number of inputs is limited to 5 to 10. It can not handle large data bases. When the number of inputs and number of training data set increases, then artificial neural networks takes long time for training and the number of iterations increases. In addition, accuracy of the results also reduces with large data bases. In earlier pollution severity diagnostic systems, mainly parameters of a single leakage current signal is considered at each time and it is given as an input to neural networks. Since LC is vastly irregular in nature, it is necessary to analyze variations of LC parameters over a lengthy time period. In the proposed RF based classifier model, it can handle large data bases effectively, it takes very less time for training and number of iterations are also very less, without compromising the accuracy of identification results. In the present work, large data base of leakage current signals collected over a period of time are given as an input to RF algorithm for training purpose, hence the accuracy of the proposed pollution severity diagnostic system is better compared with earlier works [10, 12, 14] and it can be used for real time applications.

6. Conclusion

Pollution severity analysis of power transmission line porcelain insulators from the LC measurement and its classification using RF model has been presented in this paper. Random forest model is able to clearly identify the pollution severity of outdoor insulators at different pollution conditions. Time domain analysis and discrete wavelet transform technique are used to extract both basic features and histogram features of the LC signal. Total 48 features are extracted from the LC data measured over a long period of time and given as an input to RF classifier. These results clearly show the ability of the RF classifier to handle large data base. The RF classifier based model is efficient and will be certainly useful for electrical utilities for real time applications.

References

[1] V. Jayaprakash Narayanan, M. Sivakumar, K. Karpagavani and S. Chandrasekar, "Prediction of Flashover and Pollution Severity of High Voltage Transmission Line Insulators Using Wavelet Trans-

- form and Fuzzy C-Means Approach", *Journal of Electrical Engineering & Technology*, vol. 9, no. 5, pp. 1677-1685, 2014.
- [2] L. H. Meyer, S. H. Jayaram and E. A. Cherney, "Correlation of damage, dry band arcing energy, and temperature in inclined plane testing of silicone rubber for outdoor insulation", *IEEE Trans. Dielectrics and Electr. Insul.*, vol. 11, no. 3, pp. 424-432, 2004.
- [3] S. H. Kim, E. A. Cherney and R. Hackam, "Hydrophobic behaviour of Insulators Coated with RTV Silicone Rubber", *IEEE Trans. on Electr. Insul.*, vol. 27, no. 3, pp. 610-622, 1992.
- [4] A. Cavallini, S. Chandrasekar and G. C. Montanari, "Inferring Ceramic Insulator Pollution by an innovative Approach Resorting to PD Detection", *IEEE Trans. Dielectrics and Electr. Insul.*, vol. 14, no. 1, pp. 23-29, 2007.
- [5] V. Jayaprakash Narayanan, B. Karthik and S. Chandrasekar, "Flashover Prediction of Polymeric Insulators Using PD Signal Time-Frequency Analysis and BPA Neural Network Technique", *Journal of Electrical Engineering & Technology*, vol. 9, no. 4, pp. 1375-1384, 2014.
- [6] S. Chandrasekar, C. Kalaivanan, Gian Carlo Montanari and Andrea Cavallini, "Partial Discharge Detection as a Tool to Infer Pollution Severity of Polymeric Insulators", *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 181-188, Feb. 2010.
- [7] K. Mekala, S. Chandrasekar and R. Samson Ravindran, "Investigations of Accelerated Aged Polymeric Insulators Using Partial Discharge Signal Measurement and Analysis", *Journal of Electrical Engineering & Technology*, vol. 10, no. 1, pp. 299-307, 2015.
- [8] R. S. Gorur and H. M. Schneider, "Surface resistance measurements on non-ceramic insulators", *IEEE Trans. Power Delivery*, vol. 16, pp. 801-805, 2001.
- [9] G. Montoya, I. Ramirez, J. I. Montoya, "Correlation among ESDD, NSDD and leakage current in distribution insulators", *IEE Proc. of Generation, Transmission and Distribution*, vol. 151, no. 3, pp. 334-340, 2004.
- [10] R. Sarathi, S. Chandrasekar and N. Yoshimura, "Investigations into the Surface Condition of the Silicone Rubber Insulation Material using Multiresolution Signal Decomposition", *IEEE Trans. Power Delivery*, vol. 21, pp. 243-252, 2006.
- [11] Shihua Zhao, Xingliang Jiang, Zhijing Zhang, Jianlin Hu, And Lichun Shu, "Flashover voltage prediction of composite insulators based on the characteristics of leakage current", *IEEE Trans. Power Delivery*, vol. 28, no. 3, pp. 1699-1708, 2013.
- [12] R.Sarathi and S.Chandrasekar, "Diagnostic study of the surface condition of the insulation structure using wavelet transform and neural networks", *Electric Power Systems Research, Elsevier*, vol. 68, pp. 137-147, 2004.

- [13] T.Suda, "Frequency characteristics of leakage current waveforms of an artificially polluted suspension insulator", *IEEE Trans. Dielectrics and Electr. Insul.*, vol. 8, no. 4, pp. 705-709, Aug 2001.
- [14] M.Ugur, D.W.Auckland, B.R.Varlow, and Z.Emin, "Neural Networks to Analyze Surface Tracking on Solid Insulators", *IEEE Trans. Dielectrics and Electr. Insul.*, vol. 4, no. 6, pp. 763-766, Dec 1997.
- [15] Suwarno, "Leakage Current Waveforms of Outdoor Polymeric Insulators and Possibility of Application for Diagnostics of Insulator Conditions", *Journal of Electrical Engineering & Technology*, vol. 1, no. 1, pp. 114-119, 2006.
- [16] S.Chandrasekar, C.Kalaivanan, Andrea Cavallini and Gian Carlo Montanari, "Investigations on Leakage Current and Phase Angle Characteristics of Porcelain and Polymeric Insulator under Contaminated Conditions", *IEEE Trans. Dielectr. Electr. Insul.*, vol. 16, no. 2, pp. 574-583, Apr. 2009.
- [17] Leo Brieman and E.Schapire, "Random Forests", *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [18] Weiting Chen, Yu Wang, Guitao Cao, Guoqiang Chen and Qiufang Gu, "A random forest model based classification scheme for neonatal amplitude-integrated ECG", *Biomedical Engineering Online*, vol. 13, no. 2, pp. 1-13, 2014
- [19] Long Zhenze, "Air-Gap Partial Discharge Development Stage Recognition for Power Transformer Insulation Monitoring Considering Effect of Cavity Size", *Theses and Dissertations*. Paper 502, UWM Digital Commons, 2014
- [20] Musa Al-Hawamdah, "Random Forest", *slideshare*, 2012
- [21] Jie Mei, Dawei He, Ronald Harley, Thomas Habetler and Guannan Qu, "A Random Forest Method for Real-Time Price Forecasting in New York Electricity Market", *Proceedings of IEEE PES General Meeting Conference*, pp.1-5, July 2014
- [22] IEC 60507, *Artificial Pollution tests on high voltage insulators to be used on AC systems*, 1991.



R. Shivakumar He received B.E. degree in Electrical and Electronics Engineering from the Shanmugha college of engineering, Tanjore, Bharadhidasan University, India, 1997. His M.E. degree is in Power System engineering from Annamalai University, Chidambaram, India, 1999. He received

Ph.D in the area of power systems from Anna University, Chennai in 2012. He has published 25 papers in national and international conferences. He has published 12 papers in reputed international journals. Under his supervision, currently 11 research scholars are working. His areas of interest are power system stability and control, Bio inspired optimization algorithms and renewable energy conversion system. He is member of ISTE life Member and IEEE Member.



S. Chandrasekar He received the B.E. degree in Electrical and Electronics Engineering from Thiagarajar college of Engineering, Madurai in 1996 and M.E degree in Power Systems from Coimbatore Institute of Technology, Coimbatore in India in 2001 and the Ph.D degree from Indian Institute of

Technology Madras, India in 2005. He was a postdoctoral research fellow at the University of Bologna, Italy from 2005 to 2006. Currently, he is working as a Dean (R&D) and Professor at Sona College of Technology and he is the Head of SonaPERT (Sona Electric Power Engineering Research and Testing Centre). His area of research interest is the high voltage insulation system, electric power engineering, signal processing and soft computing techniques. He is the author/coauthor of more than 125 research papers.



K. Kannan He received B.E. degree in Electrical and Electronics Engineering and M.E. degree in Power System engineering, both from the Sona College of Technology, Salem, Anna University Chennai, India, in 2003 and 2006 respectively and currently pursuing Ph.D. degree in the faculty of

electrical engineering from Anna University, Chennai, India. His research interests include High voltage systems, Electric power components, Power Quality, Fuzzy logic and AI techniques.