Naive Bayesian classifier for hydrophobicity classification of overhead polymeric insulators using binary image features with ambient light compensation

Rajamohan Jayabal\textsuperscript{1}, Vijayarekha Karuppiyan\textsuperscript{1}, Rakesh Kumar Sidharthan\textsuperscript{1}

\textsuperscript{1}School of EEE, SASTRA Deemed to be University, Trumalaisamudram, Thanjavur, India

E-mail: rajamohan@eee.sastra.edu

Abstract: Dispersion nature of water droplets over the insulator surface is used for hydrophobicity classification. Stochastic nature of water dispersions makes naive Bayesian classifier a preferable choice, which has been investigated in this work. About 12 features describing the characteristics of water droplets are extracted from the binary image using binary large objects analysis. Ambient light intensity is a significant factor that affects the binary image quality. As these insulators are installed in the outside environment, variations in ambient light intensity are inevitable. An adaptive threshold technique is proposed to compensate for ambient light variations. Six classes of various ambient light intensities have been considered in this study, and the proposed adaptive threshold technique can produce quality binary image consistently. Features extracted from the binary image are ordered according to their principal components (PCs) using PC analysis. Improvement in classification accuracy with the accumulation of ordered features is analysed. Results illustrate the use of the first eight features provides a reliable classification accuracy of 97.6\% for test image samples. In comparison to the other existing classifiers, the proposed classifier illustrates optimal performance in terms of classification accuracy and computational time.

1 Introduction

Overhead insulators are installed in electrical transmission lines to provide isolation to poles from high voltage. They play a dominant role in determining the safety of electrical transmission and distribution systems. Porcelain, glass and polymeric insulators are the types of insulators used in electrical systems [1]. Among them, polymeric insulators exhibit lower failure rate as compared with non-polymeric insulators. Polymeric insulators are lightweight, less costly, less possibility of breakage, easy to mount and mould and exhibit low surface energy. They have better hydrophobic surface even in contaminated state and faster recovery of hydrophobicity. Polymeric insulators also provide better flashover characteristics than other types of insulators. It makes them widely used in electrical transmission lines [2].

As these insulators are installed in the outside environment, they are affected by various environmental factors such as ultraviolet radiation, moisture, acidic components in rain, thermal stress, corona discharges, dry band arcing and pollutant [3]. It degrades the hydrophobic property of the insulator over time. Loss of hydrophobicity in insulators leads to reduction of flashover voltage even under a constant degree of pollution [4, 5]. It may lead to flashover/short circuit and eventually the failure of the electrical systems [6]. Therefore, monitoring the hydrophobicity level of the polymeric insulator is essential and it can be used to predict the lifetime of the insulator, which can ensure reliable operation of the electrical system [7].

Contact angle, surface tension and spray method are the standard techniques (IEC 62073) used to identify hydrophobicity class (HC) of insulator [8]. Contact angle and surface tension methods require a detailed inspection and demands the insulators to be shifted to the laboratory. Therefore, they are considered as offline methods [9, 10] and may lead to misclassification due to subjective analysis. Spray method analyses the dispersion pattern of water droplets to classify hydrophobicity by comparing with standard patterns available in Swedish Transmission Research Institute (STRI) guide [11]. This method makes the insulator to be tested on the field itself. HC is identified by the inspector, which makes it a subjective analysis leading to incorrect classification.

Digital image processing technique has been a promising tool to analyse the water dispersion pattern and to provide reliable identification of insulator’s hydrophobicity [12]. Water is sprayed over the insulator surface, and images are acquired using the digital camera. Acquired images are pre-processed and features are extracted. The literature illustrates the use of colour parameters, statistical features and geometrical features for hydrophobicity classification [13, 14]. Distilled water is coloured in contrast with the insulator colour and sprayed to measure the colour parameter. Dispersion of coloured water is identified by using the colour-based filter. Colour parameters depend on the concentration of colour mixed in the water, ambient light and spray method, which makes them highly stochastic.

Statistical features are extracted from the grey-scaled image. Various statistical features including fractal dimension, standard deviation, entropy, maximum intensity, kurtosis, skewness, variance, homogeneity, contrast, correlation and energy [15] are widely used. Grey-scale intensity level is the vital parameter to determine the quality of these features, which depends on the ambient light. Since insulators are installed in the outside environment, the image acquired is subjected to ambient light variations and dust. It makes statistical features suitable only for laboratory test conditions [14]. The literature illustrates the use of pre-processing techniques such as histogram equalisation, white top hat filter, Sobel edge operator and digital filters, which are used to minimise illumination effect [16, 17].

Unlike grey-scale images, the binary image has two levels (black and white) of intensity. It facilitates extraction of features describing the count, size, shape and dispersion of white spaces (corresponds to water) from its black (corresponds to insulator surface) background. It makes the binary image features a more reliable choice for hydrophobicity classification [18]. Binary large object (BLOB) analysis is one of a preferred technique for binary image feature extraction, which has been least investigated for hydrophobicity classification. It isolates the group of connected pixels called BLOB objects. Then, the properties of each BLOB objects are analysed individually.

Even though binary image provides several advantages, selection of a threshold for binary image conversion remains a
Table 1 Criteria of evaluation HC

| HC description | Description |
|----------------|-------------|
| only discrete droplets are formed | $\theta = 80^\circ$ or larger for the majority of droplets (0% alcohol) |
| only discrete droplets are formed | $50^\circ < \theta < 80^\circ$ for the majority of droplets (10% and 20% alcohol) |
| only discrete droplets are formed | $20^\circ < \theta < 50^\circ$ for the majority of droplets. Usually, they are no longer circular (30 and 40% alcohol) |
| both discrete droplets and wetted traces from the water runnels are observed | (i.e. $\theta = 0^\circ$). Completely wetted areas $< 2 \text{ cm}^2$. Together they cover $< 90\%$ of the tested area (50 and 60% alcohol) |
| some completely wetted areas $> 2 \text{ cm}^2$, which cover $< 90\%$ of the tested area (70 and 80% alcohol) |
| wetted areas cover $> 90\%$, i.e. small un-wetted areas (spots/traces) are still observed (90% alcohol) |
| continuous water film over the whole tested area | (100% alcohol) |

2 Experimental procedure

The proposed work employs spray method as described in the STRI guide to determine hydrophobicity. The STRI guide provides a standard procedure and apparatus to be used for hydrophobicity classification, which makes the HC depend only on the nature of the insulator. In this method, water is sprayed on the insulator surface, and the image of water dispersion pattern is acquired. An operator classifies the HC of the insulator from the acquired image by comparing with the standard reference image provided in the STRI guide. Classification of hydrophobicity by the operator is carried out by visual comparison, which may lead to misclassification of HC due to subjective analysis.

Automatic inspection of insulators is required for monitoring it periodically, which is done by image processing and learning techniques, and lead to accurate classification of HC. Table 1 shows the criteria for evaluation of HC according to STRI guide.

Acquiring a large number of insulated samples with various hydrophobicity classes is a challenging task. A solution of isopropyl alcohol and distilled water with different concentrations is prepared. It is sprayed on the surface of a new polymeric insulator to generate various classes of hydrophobicity. Alcohol tends to lower the surface tension of water and lower its density. This makes the solution to spread uniformly over the surface emulating the same effect as if the surface is less hydrophobic. The literature [22] illustrates the percentage of alcohol by volume in the spraying solution and their corresponding hydrophobic class as in Table 1. The concentration of isopropyl alcohol is inversely proportional to the hydrophobicity due to the degradation of water surface tension. For instance, a 0% alcohol concentration produces a high class of hydrophobicity (HC-1) and 100% concentration of alcohol emulates a hydrophilic surface with a very low grade of hydrophobicity (HC-7) as illustrated in Fig. 1.

A specimen is cut from the fresh polymer insulator and placed on flat surface. In case of field specimen, contamination is inevitable and will lead to a false classification of hydrophobicity [23]. Hence, the polymer insulators are cleaned to remove the contaminants from the surface before proceeding with the test. Solutions of different alcohol concentrations are sprayed on the polymer surface. A camera is mounted vertically 25 cm above the insulator surface, and images are captured for various hydrophobic classes. In the proposed work, a total of 414 images are captured with at least 30 images per HC. About 330 images are considered for training the Bayesian classifier and the remaining 84 images are used to generate testing data for validation.

The proposed work uses binary image of insulator to evaluate the hydrophobicity. Extraction of binary image uses threshold to determine the white and black pixels. Hence, an accurate threshold is required to provide a binary image of high quality. As the threshold depends on the ambient light intensity, an adaptive threshold technique is adopted. It involves two stages, namely stage-1 to evaluate the ambient light intensity and stage-2 to provide threshold based on the evaluated ambient light intensity.

At stage-1, an image of the insulator before spraying of water is acquired. The acquired colour image is converted to grey scale ($I_g$) and its mean intensity ($I_{\text{avg}}$) is evaluated as in (1), which quantitatively provides the ambient light intensity. This mean intensity ($I_{\text{avg}}$) is used to calculate the binary threshold level ($T_b$) using the pre-defined calibrated curve, which is discussed in Section 3.2. Stage-2 involves the acquisition of insulator image after spraying water. Acquired image is converted to binary form using the calculated threshold (from stage-1) as in (2), which can provide ambient light compensation and improve the image quality.

$$I_{\text{avg}} = \frac{1}{M \times N} \sum_{n=1}^{N} \sum_{m=1}^{M} I_g(m, n)$$

(1)

$$I_b(m, n) = \begin{cases} 1, & I_g(m, n) \geq T_b \\ 0, & I_g(m, n) < T_b \end{cases}$$

(2)

Binary features are extracted using BLOB analysis from the binary image ($I_b$) [24]. Total of 12 features describing the distribution of...
3 Methodology

In the proposed work, the methodology involves pre-processing of the acquired image, adaptive threshold selection, extraction of binary features and design of Bayesian classifier, which has been explained as follows.

3.1 Pre-processing

Images of the insulator surface with water droplets are acquired from the digital camera. The images are converted to grey scale to make it colour independent. Variations in intensity distribution due to shadows and ambient lights are compensated by using histogram equalisation. Histogram equalisation performed on the grey-scale image tends to equalise the intensity of the image and eventually enhances the features as illustrated in Fig. 3. Then, the images are cropped with the pre-defined region and size making it uniform for further processing.

3.2 Adaptive binary threshold using ambient light intensity

In the proposed work, the grey-scale image \(I_g\) of the insulator is converted to binary image \(I_b\) which requires a threshold \(T_b\) as in (1). Value of threshold plays a vital role in determining the image quality and majorly depends on the ambient light illumination, in which the image is acquired. As these polymer insulators are installed in transmission lines, which are predominantly in the outside environment, the acquired images are subjected to variations in ambient light intensity. It demands intensity-based binary threshold that can compensate for intensity variations, which has been addressed in this work.

To find a correlation between the ambient light intensity and binary threshold, six classes of various ambient light intensities prevail in the outside environment are considered in this paper (see Figs. 4a, 5a and 6a for classes 1, 3 and 6, respectively). About 20 images for each class are acquired before and after water spraying. Images acquired before water spraying (stage-1, see Fig. 2) are used to calculate the mean ambient light intensity \(I_{al}\) level. The binary threshold \(T_b\) producing optimal binary image after water spraying are also determined. The correlation between the intensity level \(I_{al}\) and the binary threshold \(T_b\) are determined and it is found to be a linear relation. Fig. 7 illustrates intensity level and their corresponding thresholds of the images obtained under various ambient lighting conditions (20 images \(\times\) 6 lighting conditions = 120 image samples). Mean intensity levels of these images across the ambient light classes are obtained. A linear relation is observed between the intensity level \(I_{al}\) and the binary threshold \(T_b\). This leads to fit a linear curve using least-square technique with the reliable level of root-mean-square error 0.02155 as in (3). Using this curve, the binary threshold for the given ambient light intensity conditions can be determined, which can provide a binary image of reliable quality

\[
T_b = 0.004291 I_{al} - 0.01041
\]  

(3)

To ascertain the efficiency of the proposed ambient light compensation technique, the effect of binary threshold on binary features is evaluated. Number of water droplets \(N_{wd}\) is one of the important binary feature, which is considered in this paper. The acquired image of insulator after water spraying (see Fig. 8a) is converted to grey-scale image and histogram equalisation is applied as described in Fig. 8b. Histogram equalisation enhances the image and improves the sharpness of the edges leading to distinguish water droplets from its background effectively. A pre-
The defined region is cropped from the image as shown in Fig. 8c. The choice of region is capable of avoiding any insulator edge and observing the water dispersion pattern. To demonstrate the impact of threshold, the cropped images are converted to binary image with both fixed threshold (see Fig. 9) and with the proposed adaptive threshold (see Fig. 10).

BLOB analysis is performed to isolate the group of white space regions, which corresponds to the water droplets. Area-based filtering technique is used to denoise the segmented binary image. Binary objects with the lesser area often correspond to noise, which is removed by a pre-defined area constraint. The number of water droplets is identified as illustrated in Figs. 9d and 10d. It is observed that the adaptive threshold is capable of accurately identifying the water droplets, whereas the fixed threshold tends to group multiple water droplets along with its background.

Furthermore, the robustness of the proposed adaptive threshold technique is evaluated across all the six classes of ambient light intensities ($C_a$) considered in this paper. These classes of intensities indicate discrete samples of increasing ambient lights that occur in an outdoor environment at a forenoon session of day time. Ambient light intensity class ($C_a = 1$) corresponds to low light intensity that occurs at dawn/dusk and $C_a = 6$ indicates a high light intensity at noon. Since the same class of lighting conditions will occur in decreasing order on the afternoon session, it becomes redundant and hence they are omitted. The number of water droplets in the same insulator is determined across various light intensities with area constraint. It is observed that the number of water droplets identified remains consistent and close to the actual number of water droplets ($N_{wd} = 68$) as in Fig. 11.

3.3 Feature extraction using BLOB analysis

BLOB analysis primarily groups the cluster of white pixels (indicating water droplets) and determines their properties. This makes BLOB a preferred tool for hydrophobic classification as it predominantly depends on the dispersion pattern of water droplets [25]. It uses a template of connected pixels may be either four connected or eight connected pixels to detect the region. Once the region is segmented, the properties of these regions are determined and can be used as features for hydrophobicity classification. 12 features describing the dispersion pattern of water droplets are extracted from the BLOB analysis as explained below.

3.3.1 Number of water drops: Number of water droplets in the given area is a key marker to indicate the dispersion of water over the hydrophobic surface. As the water is sprayed uniformly, they tend to cluster locally to provide a minimum contact area with the insulator. This leads to an increase in number of water droplets (with smaller area) for higher HC as illustrated in Fig. 12a. In a low hydrophobic insulator, water tends to spread all over the surface leading to reduced number of water droplets (with larger area) as in Fig. 12d.
In BLOB analysis, the number of water droplets \( N_{\text{wd}} \) is calculated by the number of binary objects \( N_{\text{bl}} \) detected. However, BLOB has a potential to identify even a smaller cluster of white pixels, which may correspond to a high-frequency noise. A threshold for the area \( A_{\text{bl}} \) is set to filter out the noise and the binary objects having area higher than the threshold are considered as water droplets as in (4). It is possible to define a minimum pixel that corresponds to water droplet as the camera is mounted in accordance to the standards at a fixed distance of 25 cm from the insulator. As there is no relative movement/translation between the camera and insulator, the minimum pixel value corresponding to the water droplet is constant and set as area threshold.

The segmented image with identified water droplets for various HC classes of hydrophobicities is illustrated in Fig. 12. It is clearly evident that the number of water droplets significantly reduces with an increase in its area. This makes the number of water droplets a clear marker of hydrophobicity

\[
N_{\text{wd}} = \left| \{ A_{\text{bl}} \mid A_{\text{bl}} \geq A_{\text{bl}}^{\text{thre}}, i \in N_{\text{bl}} \} \right|
\]

### 3.3.2 Circular factor \((C_f)\):

Circular factor is one of the crucial shape features to classify hydrophobicity. It describes the closeness of the water droplet to a perfect circle. It is directly related to the surface tension existing between water and the insulator surface. If the water droplet appears to be circular, then the insulator exhibits a higher hydrophobicity. Water droplets may be traced out (non-circular) in a lower hydrophobic surface

\[
\text{Circularity} = \frac{\text{perimeter of BLOB}}{2\sqrt{\pi} \times \text{area of BLOB}} \quad (5)
\]

### 3.3.3 Coverage rate of water:

In the given binary image, the white pixels correspond to the presence of water and a black pixel represents the background surface of the insulator. Coverage rate of water \((C_i)\) is calculated as a ratio of number of white pixels \(N_{\text{wpix}}\) to the total number of pixels \(N_{\text{pix}}\) in the given image. In a highly hydrophobic surface, the water tends to present in the minimal area due to its higher surface tension. This makes coverage rate of water to be lesser for a good hydrophobic insulator. On the other hand, water tends to spread over the lesser hydrophobic surface making a higher coverage rate

\[
C_i = \frac{N_{\text{wpix}}}{N_{\text{pix}}} \quad (6)
\]

### 3.3.4 Coverage rate of maximum water droplet:

The water droplet with a maximum area can provide some insight into the distribution of water over the insulator surface. The area of this maximum water droplet is calculated as its coverage rate and used for hydrophobicity classification

\[
C_{\text{max}} = \frac{N_{\text{wpix}}}{N_{\text{pix}}} = \max(A_{\text{bl}}) \quad (7)
\]

### 3.3.5 Solidity:

Solidity is the measure of droplet shape. It is an area section of the object related to its convex hull. A more circular water drop produces the solidity value closer to one and indicates a higher hydrophobic surface. If it is stretched, then the value is less than one indicating a lesser hydrophobic insulator.

### 3.3.6 Maximum perimeter:

Perimeter of the binary objects \(P_{\text{max}}\) can provide information about the shape of water dispersed over the insulator surface. On a high hydrophobic surface, the water droplet experiences a circular shape with the minimal perimeter. Similarly, the water tends to disperse on a lower hydrophobic surface having a higher perimeter.

### 3.3.7 Shape factor:

Shape factor is a standard metric to determine the shape of the binary objects. In the proposed work, the shape factor of the maximum water droplet is considered as a feature. It is proportional to the ratio of diameter \(D_{\text{max}}\) to the perimeter of the maximum binary objects in the given image as in the equation below:

\[
S = \frac{4 \pi D_{\text{max}}}{P_{\text{max}}} \quad (8)
\]

### 3.3.8 Euler number:

Euler number is the difference between the number of objects and the number of holes in that image. Negative sign shows that the sum of holes is higher than the sum of objects. It is one of a significant feature that describes the reflectivity of the water droplet. On a higher hydrophobic surface, the water droplet tends to form a convex shape and reflects the light to create a single brighter spot. The reflection of the insulator surface in the water droplet creates a darker spot just near the brighter spot as shown in Fig. 13. It creates a hole in the binary image of the water droplet, which can be measured using Euler number. The water tends to disperse on a lesser hydrophobic surface. It will create a uniform reflection of light and also no reflection of the insulator surface is observed. The binary image is also observed to be with no or a minimal number of holes as shown in Fig. 13.

### 3.3.9 Eccentricity mean:

Unlike the conventional technique, BLOB analysis provides the elliptical fit of the identified binary objects. It provides a major \(r_m\) and minor \(r_m\) radii that cover the possible area of the ith binary objects. Thus, the eccentricity mean for the given image is calculated as the ratio of minor to the major radius, which is close to unity for a circular object. It provides the overall inferences about the shape of the water droplets (Fig. 14)

\[
E_m = \frac{1}{N_{\text{wd}}} \sum_{i=1}^{N_{\text{wd}}} \frac{r_m}{r_M} \quad (9)
\]
using the given features. This Bayesian classifier is designed. Once the NBC is trained, its classification accuracy is evaluated using the predicted HC of the insulator.

As the insulator surface loses hydrophobicity, the number of water droplets falls and the isolation of water tends to decrease as the insulator loses its hydrophobicity as observed in Fig. 15.

4 Bayesian classifier

Naive Bayesian classification (NBC) is a statistical supervised classification method [26]. It is one of the widely used machine learning technique for both binary classification and multi-class classification problems. NBC is based on Bayes theorem to calculate the conditional probability \( P(H_c|f_j) \) for the given image features \( (f_j) \) belong to a particular class of hydrophobicity \( (H_c) \) as in (10). It is simple to be constructed and suitable for high dimensionality systems.

\[
P(H_c|f_j) = \frac{P(f_j|H_c)P(H_c)}{P(f_j)} \tag{10}
\]

NBC finds its application in challenging image classification problem as reported in the literature [27]. NBC is preferred for image classification because of its higher classification accuracy and training speed [28]. It provides the likelihood of class for the extracted features, which enables variable thresholds and makes reliable for real-time applications [27, 29]. NBC outperforms with higher accuracy than other conventional techniques [30] in texture classification, which is closely related to the proposed work. It makes NBC a preferred choice for analysing water dispersion patterns in the proposed work.

In the proposed work, binary features are extracted from the acquired insulator image and its corresponding hydrophobic classes are determined by its alcohol concentration. The extracted features and their corresponding HCs are used as training data to design multiple Gaussian models, which represent the a priori knowledge \( P(f_j|H_c) \). Once the NBC is trained, its classification accuracy is evaluated using the features available in the training data. For testing, new images for various classes of hydrophobicities are acquired. Binary features are extracted after pre-processing with an adaptive threshold. These features are used to determine the HC of the insulator.

5 Optimal feature selection

Among the 12 features extracted from BLOB analysis, optimal features required for classification are selected as follows. Initially, PCA is used to order the feature set \( (f_{blr}) \) in accordance with the magnitude of eigenvectors \( (E_g) \). Eigenvectors provide information about the interdependencies among features. The independent feature will have a higher magnitude of eigenvector co-efficient and are considered as preferred features. Eigenvectors \( (E_g) \) are sorted in the increasing order of its magnitude and the corresponding feature indices \( (s_i) \) are identified. Next, the features are reordered according to the identified indices and a new set of features \( (f_{bs}) \) is created.

This reordered feature is accumulated one by one and a Bayesian classifier \( (B_c) \) is designed. This Bayesian classifier is used to predict the HC \( (H_{cp}) \) using the given features \( (f_{bs}) \). Classification accuracy is evaluated using the predicted \( (H_{cp}) \) and

3.3.11 Histogram-based major area: Histogram is used to classify the water droplets by their area. The binary objects identified from the BLOB analysis are classified into 20 area bins. The number of water droplets in each of the area bins is found, and a histogram is plotted as in Fig. 15. The area bin having the highest number of water droplets is considered as major area and used as a feature to classify hydrophobicity. As the insulator surface loses hydrophobicity, the number of water droplets falls and the dispersion/area coverage by water increases as illustrated in Fig. 15.

3.3.12 Histogram-based major water droplets: The number of water droplets having the majority area is also considered to be a feature. A good class of insulator produces a higher number of water droplets belonging to a smaller area bin, and the isolation of water tends to decrease as the insulator loses its hydrophobicity as observed in Fig. 15.

**Fig. 14 Eccentricity**

Water droplets are fitted into an elliptical shape. Major and minor axes of the ellipse are calculated. Eccentricity is defined as the ratio of these two axes. Water droplets having more circular nature produces the eccentricity value closer to one. In this work, both mean and maximum values of eccentricity are considered as a separate feature.

3.3.10 Eccentricity maximum: The maximum value of the eccentricity is also considered as one of the features for hydrophobic classification. For a perfect circle, the major and minor axes are identical leading the eccentricity equals to one. Eccentricity maximum describes the availability of any water droplet closer to a circular shape.

**Fig. 15 Area histogram**

(a) Low hydrophobicity, (b) High hydrophobicity
Input: $f_{BL}, H_{ins}$ // BLOB feature set and Actual Hydrophobicity class  
Output: $C_{m}$ // Confusion matrix  

Do:  
1. Compute: $E_g$ ← PCA($f_{BL}$) // performs PCA to determine Eigenvectors  
2. Compute: $E_g$ ← sort($E_g$, 'Ascending') // Arrange the values according to the magnitude  
3. Set: $i_f$ ← FindIndex($E_{gs}$) // Find all the features index  
4. Set: $f_{bl}$ ← $f_{BL}(i_f)$ // Reorder the feature set using the new index  
5. Initialize: $i = 1$  
6. While: $i \leq N_f$ // Do for each feature  
   Do:  
   1. Set: $f_{bs}$ ← $f_{bl}(i) [i \leq i]$ // Create a new feature set by accumulation  
   2. Compute: $B_c$ ← DesignBayes($f_{bs}$) // Design Bayesian classifier  
   3. Compute: $H_{cp}$ ← predict($B_c, f_{bs}$) // Predict using Bayesian classifier  
   4. Compute: $C_m(i)$ ← confusionmatrix($H_{ca}, H_{cp}$) // Determine confusion matrix  
   5. Increment: $i ← i + 1$;  
End  

Fig. 16 Algorithm 1: optimal features selection  

| Table 2 | Eigenanalysis of the correlation matrix |  |
|--------|----------------------------------------|---|
| PC     | Eigenvalue | Proportion | Cumulative |
| PC1    | 7.6616     | 0.6385     | 0.6385    |
| PC2    | 1.6843     | 0.1374     | 0.7758    |
| PC3    | 0.9018     | 0.0751     | 0.8510    |
| PC4    | 0.7452     | 0.0621     | 0.9131    |
| PC5    | 0.3782     | 0.0315     | 0.9446    |
| PC6    | 0.2943     | 0.0245     | 0.9691    |
| PC7    | 0.1295     | 0.0108     | 0.9799    |
| PC8    | 0.1111     | 0.0093     | 0.9892    |
| PC9    | 0.0754     | 0.0063     | 0.9955    |
| PC10   | 0.0465     | 0.0039     | 0.9993    |
| PC11   | 0.0080     | 6.6856×10^{-04} | 1.0000   |
| PC12   | 7.2582×10^{-09} | 6.0485×10^{-10} | 1.0000   |

actual class of hydrophobicity ($H_{ca}$) of the insulator. The improvement is classification accuracy with the introduction of additional feature is analysed by using the confusion matrix ($C_m$). This procedure is repeated for all the identified 12 features as illustrated in Algorithm 1 (see Fig. 16).  

5.1 Feature ordering using PCA  

Eigenanalysis of the correlation matrix (formulated using the feature sets) is carried out using PCA [31]. The variations in eigenvalues across all the PCs of the features are identified. These values are used to identify the predominant components to evaluate the eigenvectors. From Table 2, it is observed that around 85% of the features lies in the first three PCs and remaining PCs are eliminated due to its lesser significance.  

Eigenvectors for these first three PCs are evaluated as in Table 3. The coefficients of these eigenvectors having a significantly higher magnitude are identified. The features are reordered in accordance with the strength of its coefficients as illustrated in Table 3.  

5.2 Feature evaluation using accumulation effect  

The improvements in classification accuracy with the accumulation of features in the order prescribed by PCA are evaluated. Confusion matrix has been used as a performance index to evaluate the classification accuracy for training and testing image sets. Fig. 17 shows the improvement in accuracy with a sequential inclusion of one feature at a time. It is observed that the inclusion of feature beyond eight has no significant improvement in classification accuracy. Hence, the first eight feature ordered by PCA is considered to be optimal and the remaining four features are made obsolete.  

![Fig. 17 Effect of feature accumulation on classification accuracy of NBC](http://creativecommons.org/licenses/by/3.0/)  

Table 3 Eigenvectors of PC1–PC3  

| Features              | Feature order | PC1   | PC2   | PC3   |
|-----------------------|---------------|-------|-------|-------|
| number of water drops  | 10            | -0.3202 | 0.1021 | 0.1226 |
| coverage rate of water | 3             | 0.2615 | 0.0520 | -0.4524 |
| water drop            | 11            | 0.3188 | 0.2418 | -0.0036 |
| circular factor       | 6             | -0.3144 | 0.3417 | 0.0324 |
| solidity              | 4             | -0.2978 | -0.0282 | -0.4177 |
| maximum perimeter     | 8             | 0.3340 | 0.0654 | 0.2198 |
| shape factor          | 7             | -0.3415 | 0.0945 | -0.0119 |
| Euler number          | 5             | -0.2617 | -0.3641 | -0.3250 |
| eccentricity mean     | 9             | 0.3269 | -0.2755 | -0.0288 |
| eccentricity maximum  | 2             | 0.1284 | -0.6286 | -0.1455 |
| histogram-based major | 1             | -0.1355 | -0.3681 | 0.6548 |
| area                  | 12            | 0.3188 | 0.2418 | -0.0037 |

Identified feature used for training and testing of classifier.  

6 Results and discussion  

A total of 330 training images and 84 test images of the insulator are considered in this paper. These images are compensated for ambient light variations and converted to binary images. Identified eight BLOB features (see Table 2) have been extracted for these images and the training feature set is used to design NBC. The performance of NBC is evaluated using the testing data set with a confusion matrix as in Table 4. It is observed that the proposed NBC can classify HC accurately at higher classes (HC-1–3) and produce minor classification error at a lower class of...
hydrophobicity. It is mainly due to the minimal shape variations of water droplets in lower HCs.

Furthermore, the influence of the training sample size is analysed as follows. The training samples \((N_{tr})\) are chosen as per the ratio \((R)\) of testing samples \((N_{te})\) as in (11). The classification accuracy for the proposed NBC trained using the samples selected based on the ratios is illustrated in Fig. 18. The ratio less than unity corresponds to lesser training samples than testing samples \((N_{tr} < N_{te})\) and vice versa. At unity ratio, the number of training data is equal to the testing data \((N_{tr} = N_{te})\). At lower ratios, it is observed that the proposed classifier tends to overfit, which is indicated by a large deviation between the classification accuracies for training and testing data. With a higher number of training data, the NBC is capable of learning the likelihood of features amidst outliers. This improves the classification accuracies at higher ratios

\[ N_{te} = R \times N_{te} \quad (11) \]

Thus, the proposed NBC trained with all the training samples \((R = 4)\) produces a classification accuracies of 99.1 and 97.6% for training and testing samples, respectively. Multi-fold cross-validation \((N = 10 \text{ folds})\) is also performed to assess classification accuracy. These folds provide an ordered selection of training data and testing data from the same set of samples. This makes each data to be used for both training and testing (not simultaneously) at least once. Classification accuracy of 97.82% is observed, which indicates the reliability of the proposed NBC.

Furthermore, the performance of the proposed Bayesian classifier is compared with other closely related classical techniques. This includes decision tree, linear discriminate analysis (LDA), K-nearest neighbourhood (KNN) and SVM. These techniques are also found to be employed for hydrophobicity classification with different features [18, 21, 32]. Similar to the feature selection procedure used in the proposed NBC, features ordered by PCA are sequentially accumulated for other classifiers. Their classification accuracies for testing data are evaluated as illustrated in Fig. 19. It is observed that the decision tree and SVM are aligned with the proposed NBC in providing optimal classification accuracy for eight features. LDA and KNN produce higher classification accuracy for 9 and 12 features, respectively.

To evaluate the performance of the classifiers, confusion matrix and computational time requirement are considered as performance metrics. Confusion matrices for both training and testing data are evaluated. The computational time required for the design of classifier and evaluation of HC for a given feature set is used for performance analysis. Mean and variance are determined for 100 samples to capture the stochastic nature of computational time variations. Comparative analyses of these classifiers are described in Table 5. SVM, being a quadratic programming-based algorithm, evolve to be complex with the number of training data. This makes SVM to have a higher computational time for training and linear prediction also makes larger testing time as compared with other classifiers. Although decision tree and KNN can produce 100% classification accuracy for training data, failure of these classifiers

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
AC/PC & HC-1 & HC-2 & HC-3 & HC-4 & HC-5 & HC-6 & HC-7 \\
\hline
HC-1 & 30 (12) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) \\
HC-2 & 0 (0) & 60 (11) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) \\
HC-3 & 0 (0) & 0 (0) & 60 (18) & 0 (0) & 0 (0) & 0 (0) & 0 (0) \\
HC-4 & 0 (0) & 0 (0) & 0 (0) & 59 (11) & 0 (0) & 0 (0) & 0 (0) \\
HC-5 & 0 (0) & 0 (0) & 0 (0) & 1 (0) & 59 (10) & 0 (0) & 1 (1) \\
HC-6 & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 29 (11) \\
HC-7 & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 0 (0) & 29 (11) \\
\hline
\end{tabular}
\caption{Confusion matrix for all training and testing data – after reduction of feature by PCA}
\end{table}

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Classifier & Classification, % & Computational time, ms & & \\
 & Training & Testing & Training & Testing \\
\hline
decision tree & 100 & 91.7 & 46.6 ± 0.01 & 2.0 ± 0.007 \\
LDA & 88.2 & 88.1 & 46.5 ± 3.0 & 6.7 ± 0.005 \\
KNN & 100 & 72.6 & 18.9 ± 0.50 & 7.2 ± 0.01 \\
SVM & 94.8 & 94.0 & 6334.2 ± 2350.0 & 35.4 ± 0.02 \\
NBC (proposed) & 99.1 & 96.4 & 30.8 ± 0.10 & 13.4 ± 0.03 \\
\hline
\end{tabular}
\caption{Comparative analysis of various classifiers}
\end{table}
to provide reliable accuracy for testing data makes them overfitted. Parametric nature of LDA makes it produce least classification accuracy for the proposed classification problem. Among the classifiers considered in this paper, stochastic NBC is capable of providing reliable classification accuracy with optimal computational time.

7 Conclusion

In this paper, the HC of polymer insulator is identified by the spray method. Dispersion of water droplets over the insulator surface is a direct marker of HC. Features are extracted from the digital image of the insulator to infer the water dispersion pattern. Ambient light variations have been measured, and the threshold for binary images is determined adaptively. Six classes of ambient light variations are used to validate the adaptive threshold technique. Results illustrate a robust extraction of feature (number of water droplets) amidst variations in ambient light intensity.

NBC is used to build a priori model of features for the given HC and used to classify the insulator's HC. About 12 features are extracted from the binary image by BLOB analysis. Features are ordered by PCA and improvements in classification accuracy with the feature population of features are analysed for possible feature reduction. Experimental results demonstrate that the use of first eight features can produce a significant classification accuracy of 97.6% for testing data. The proposed classifier is equipped with ambient light compensation, which makes it suitable for field testing of insulators. Automated acquisition of insulator images and online classification using an unmanned aerial vehicle is the future scope of the proposed work.

Furthermore, binary image-based features used in this proposed work are best suited for evaluation of pattern, shape, size and location of the foreground objects. Hence, the proposed procedure can be applied for image-based object classification. Identification of various elements in transmission lines, evaluation of flashover pattern in contaminated insulators and assessment of transmission line layout using aerial images are some of the potential applications that can use the proposed methodology.

8 References

[1] Papailiopoulos, K.O., Schmuck, F.: ‘Silicone composite insulators: materials, design, applications’ (Springer Verlag, Berlin, Germany, 2013)
[2] CIGRE TB 442: ‘Evaluation of dynamic hydrophobicity properties of polymeric materials for non-ceramic outdoor insulation, retention and transfer of hydrophobicity’, CIGRE WG D1.14, 2010
[3] Yoshimura, N., Kamagai, S., Nishimura, S.: ‘Electrical and environmental aging of silicone rubber used in outdoor insulation’, IEEE Trans. Dielectr. Electr. Insul., 1999, 6, (5), pp. 632–650
[4] Xidong, L., Shaowu, W., Lengceng, H., et al.: ‘Artificial pollution test and pollution performance of composite insulators’, Int. Symp. High Voltage Eng., UK, August 1999, pp. 337–340
[5] Xidong, L., Shaowu, W., Ju, F., et al.: ‘Development of composite insulators in China’, IEEE Trans. Dielectr. Electr. Insul., 1999, 6, (5), pp. 586–594
[6] Liu, Y., Du, B.X.: ‘Recurrent plot analysis of leakage current in dynamic drop test for hydrophobicity evaluation of silicone rubber insulators’, IEEE Trans. Power Del., 2013, 28, (4), pp. 1996–2003
[7] Gubanski, S.M., Demfalk, A., Anderson, J., et al.: ‘Diagnostic methods for outdoor polymeric insulators’, IEEE Trans. Dielectr. Electr. Insul., 2007, 14, (5), pp. 1065–1080
[8] ICTS 62073: ‘Guidance on the measurement of wettability of insulator surfaces’, 2003
[9] Xu, Z., Liu, F.: ‘A static contact angle algorithm and its application to hydrophobicity measurement in silicone rubber corona using test’, IEEE Trans. Dielectr. Electr. Insul., 2013, 20, (5), pp. 1820–1831
[10] Du, B.X., Han, T., Cheng, X.X., et al.: ‘Characterization of surface discharge as indicator for hydrophobicity evaluation of silicone rubber insulators’, IEEE Trans. Dielectr. Electr. Insul., 2012, 19, (3), pp. 1708–1714
[11] ‘Hydrophobicity classification guide, STRI guide 92-1’, Swed. Transm. Res. Inst., 1992
[12] Thomazini, D., Gelfuso, M.V., Altfrun, R.A.C.: ‘Classification of polymers insulators hydrophobicity based on digital image processing’, Mater. Res., 2012, 15, (3), pp. 365–371
[13] Du, B.X., Ma, Z.L., Cheng, X.X., et al.: ‘Hydrophobicity evaluation of silicone rubber insulators using LBG induced electromagnetic wave’, IEEE Trans. Dielectr. Electr. Insul., 2012, 19, (3), pp. 1060–1067
[14] Li, C., Huang, X., Zhao, L.: ‘Image analysis on the surface hydrophobicity of polluted silicone rubber insulators’. IEEE Int. Conf. Condition Monitoring and Diagnosis, Beijing, China, April 2008, pp. 389–391
[15] Jayabal, R., Vijayaraghavan, K., Rakesh Kumar, S.: ‘Design of ANFIS for hydrophobicity classification of polymeric insulators with two-stage feature reduction technique and its field deployment’, Energies, 2018, 11, (12), p. 3391
[16] Wei, Z., Lidong, H., Jun, W., et al.: ‘Entropy maximisation histogram specification method for image enhancement’, IET Image Proc., 2014, 9, (3), pp. 226–235
[17] Lidong, H., Wei, Z., Jun, W., et al.: ‘Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement’, IET Image Proc., 2015, 9, (10), pp. 908–915
[18] Dong, Z., Fang, Y., Wang, X., et al.: ‘Hydrophobicity classification of polymeric insulators based on embedded methods’, Mater. Res., 2015, 18, (1), pp. 127–137
[19] Bedruz, R.A., Sybingco, E., Bandala, A., et al.: ‘Real-time vehicle detection and tracking using a mean-shift based blob analysis and tracking approach’. Int. Conf. Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, Manila, Philippines, December 2017
[20] Pylarinus, D., Lazaro, S., Marmidis, G.: ‘Classification of surface condition of polymeric line insulators using wavelet transform and neural networks’. Int. Conf. Wavelet Analysis and Pattern Recognition, Beijing, China, November 2007, pp. 658–663
[21] Jarrar, I., Assaleh, F., Al-Hadidi, M.H.: ‘Using a pattern recognition-based test to assess the hydrophobic class of silicone rubber materials’, IEEE Trans. Dielectr. Electr. Insul., 2014, 21, (6), pp. 2611–2618
[22] Thomazini, D., Gelfuso, M.V., Altfrun, R.A.C.: ‘Hydrophobicity classification of polymeric materials based on fractal dimension’, Mater. Res., 2008, 11, (4), pp. 415–419
[23] Zhao, L., Li, C., Xiong, J., et al.: ‘Online hydrophobicity measurement for silicone rubber insulators on transmission lines’, IEEE Trans. Power Del., 2009, 24, (2), pp. 806–813
[24] Yoon, Y., Ban, K.D., Yoon, H., et al.: ‘Blob extraction based character segmentation method for automatic license plate recognition system’. Int. Conf. Systems Man and Cybernetics, Anchorage, AK, USA, November 2011, pp. 2192–2196
[25] Wang, Q., Huang, Y., Mo, X., et al.: ‘The hydrophobic detection of transformer composite insulator bushing based on digital image processing technique’. IEEE Int. Conf. High Voltage Engineering and Application (ICHVE), Chengdu, China, September 2016, pp. 1–4
[26] Duyur, B., Valverde, R.J., Moura, M.F., et al.: ‘A survey of the applications of Bayesian networks in agriculture’, Eng. Appl. Artif. Intell., 2017, 65, pp. 29–42
[27] Han, T., Cheng, X., Huang, C.L.: ‘Image classification using Naïve Bayes classifier with pairwise local observations’, J. Inf. Sci. Eng., 2017, 33, (5), pp. 1–15
[28] Park, D.C.: ‘Image classification using Naïve Bayes classifier’, Int. J. Comput. Sci. Electron. Eng., 2016, 4, (3), pp. 135–139
[29] Kasat, N.R., Thepade, S.D.: ‘Novel content based image classification method using LBG vector quantization method with Bayes and lazy family data mining classifiers’. Procedia Computer Science, Int. Conf. Communication, Computing and Virtualization, 2016, vol. 79, pp. 483–489
[30] Mansour, A.M.: ‘Texture classification using Naïve Bayes classifier’, Int. J. Comput. Sci. Netw. Secur., 2018, 18, (1), pp. 112–120
[31] Kumar, D., Singh, R., Kumar, A., et al.: ‘An adaptive method of PCA for minimization of classification error using Naïve Bayes classifier’. Procedia Computer Science, Int. Conf. Eco-friendly Computing Communication Systems, 2015, vol. 70, pp. 9–15
[32] Wang, Q.D., Zhong, Z.F., Wang, X.P.: ‘Design and implementation of insulators material hydrophobicity measure system by support vector machine decision tree learning’. IEEE Int. Conf. Machine Learning Cybernetics, Guangzhou, China, August 2005, vol. 7, pp. 4328–4334