This article is a preprint. Please cite the published version: **https://doi.org/10.1016/j.epsr.2023.109340**

A Machine Learning Approach for Condition Monitoring of High Voltage Insulators in Polluted Environments

Héctor de Santos [1,2*], Miguel Á. Sanz-Bobi [2]

[1] Verescence, La Granja Insulators, San Ildefonso, Spain
[2] ICAI School of Engineering, Comillas Pontifical University, Madrid, Spain
[*]hector.desantos@verescence.com

Key-words: *Condition monitoring, imbalanced data, insulators, machine learning, pollution flashover*.

Abstract: This paper proposes a new approach for insulator condition monitoring based on the combination of the random under sampling technique with an adaptative boosting algorithm (RUSBoost) and aiming to estimate key condition indicators from the meteorological and environmental data. The research was conducted at a 245 kV test station located in a severely polluted area in France, where one glass insulator string and two mirroring strings, but composed by full and half silicone-coated (bottom surface only) glass insulators, were monitored in real operational conditions during two consecutive years. The definition of the condition indicators was carried out through the characterization of the leakage current obtained in laboratory tests, subjecting the glass insulator string to different artificial pollution levels until flashover. Afterwards, the performance of the new proposed RUSBoost approach was evaluated and compared with AdaBoost, Bagging, Random Subspace Ensemble with k-nearest neighbors (KNN) and support vector machines (SVM) algorithms. The results show the effectiveness of RUSBoost in addressing the estimation of the highly imbalanced insulator condition indicators and its advantage over other methods by achieving a macro-averaged F-score of 0.757 for the non-coated string and a F-score of 0.768 for the half-coated string and 0.792 for the full coated string.

1. Introduction

High-voltage glass insulators are an essential component of transmission and distribution overhead lines and their pollution performance is crucial in maintaining the reliability of the power supply. The combination of severe pollution with harsh environmental and weather conditions may cause insulator flashovers and undesirable line outages constituting a great threat to deliver reliable service for many electric utilities worldwide. The right selection and dimensioning of the outdoor insulators are important in limiting the number of pollution-induced flashovers, but even so, in certain areas having high natural and/or industrial pollution and low rainfall, preventive maintenance may be still required to keep the insulators within a desirable operational level [1]. The maintenance strategy depends on several aspects such as the site pollution severity, weather, live working or out-of-service conditions, strategic value, accessibility and logistics as well as economical requirements, which must be carefully considered to achieve the best cost-effective solution.

Among the available maintenance procedures, periodic washing is the most common method to remove the pollutants from the surface of the glass insulators [2]. The washing can be carried out in a number of ways [3], but at present, live or hot-line washing of insulators with high-pressure water systems mounted in trucks or helicopters, is largely preferred by electric utilities for rapid operations over long distances without service interruptions. In those cases where regular washing is not possible, or economically viable, the installation of pre-coated glass insulators with room temperature vulcanized (RTV) silicone rubber has been also successfully implemented to upgrade their pollution performance in several countries such as Spain, Italy, Saudi Arabia, the United States, or China [4,5]. Thanks to the hydrophobicity transfer capabilities of the silicone rubber coating, the formation of continuous and conductive water films into the polluted surface is inhibited to a large extent, thereby reducing the risk of having a flashover in service [6]. In this context, one of the main challenges to address when it comes to insulator maintenance is the scheduling of washing. Considering the large variability of the environmental pollution and weather conditions the outdoor insulators are exposed to, regular washing at fixed time intervals is not an efficient strategy. An interval length that proves to be appropriate at one period of time may not be suitable for another period with different conditions, and washing too frequently or too infrequently could result in excessive costs or in unacceptable consequences to the reliability of the power system. Therefore, maintenance efficiency can be sharply improved when it is based on the current pollution condition of the outdoor insulation and performed in a dynamic and predictive manner: only when required and initiated just-in-time [7].

While there are several methods to monitor insulator pollution [8], the most widely used are the equivalent salt deposit density (ESDD), the non-soluble deposit density (NSDD) and the leakage current. The first ones measure the surface contamination adhered to the insulator, but have important limitations as they do not reflect the wetting effect on pollution and they cannot provide real-time data. On the contrary, the leakage current can be continuously monitored by means of wireless sensors, providing a comprehensive reflection of the pollution accumulated and wetting events in real-time [9,10]. Leakage current has proven to be a meaningful pollution performance indicator as it gives a measure of how close the insulators string is to flashover, allowing for the early detection of developing pollution issues and delivering an effective trigger for maintenance before critical conditions arise [11–13].

The strong correlation observed between the insulator pollution and the meteorological variables such as relative humidity, ambient temperature, wind speed and direction, solar radiation and rain [14–16] has motivated research focused on regression techniques for the prediction of the ESDD, NSDD or leakage current from the weather and environmental conditions. In particular, some linear and non-linear regression methods were proposed to estimate the leakage current [17,18], but showed a limited applicability to very short periods of time and a lack of flexibility since the regression coefficients changed frequently. By contrast, more promising results were obtained by means of machine learning based regression techniques. These algorithms capture better the very complex functional relationships between the inputs and output, especially when large amounts of data are involved, making them more suitable to address the problem properly. In connection with this, previous studies indicated the effectiveness of artificial neural networks (ANN) [19–22], support vector machines (SVM) [23] and random forests (RF) [24] algorithms in the prediction of the ESDD in insulators. The application of machine learning algorithms in the estimation of leakage current is even better suited for the task because it is monitored in real-time and involves much larger datasets than in the case of ESDD. In this regard, recent research based on RF [25] demonstrated good suitability in handling the leakage current regression problem in an effective way.

Despite the remarkable progress achieved through these techniques, some enhancements are still necessary in different aspects. Firstly, since the primary motivation behind insulator condition monitoring is essentially practical, the leakage current data must be processed in such a way that it is easily comprehensible and rapidly translated into specific maintenance actions. The classification of the leakage current into a certain number of ranges or condition indicators, based on the risk of having a flashover, would improve its accuracy and functionality. And secondly, the frequency of occurrence of relevant leakage current peaks connected to pollution and wetting events is rare. This results in having highly imbalanced datasets and the need for monitoring long periods of time, typically covering more than one year, to collect these key data. Class imbalance is a major challenge in machine learning because these algorithms learn from the data and, in many cases, standard methods may have difficulties handling them adequately, resulting in minority classes seldom predicted or even overlooked as potential outliers [26]. The proper estimation of the minority instances becomes especially important in the task and, consequently, the machine learning techniques must be designed taking into account these particular circumstances.

This paper proposes a new machine learning classification approach for insulator condition monitoring based on the combination of the random under sampling technique with an adaptative boosting algorithm (RUSBoost) and aiming to estimate key condition indicators from the meteorological and environmental data. The experimental research was conducted at a 245 kV test station facility located in a severely polluted seashore area in the south of France, where one glass insulator string and two mirroring strings, but composed by full silicone-coated and half silicone-coated (bottom surface only) glass insulators, were monitored in real operational conditions during two consecutive years. The definition of the condition indicators was carried out through the characterization and analyses of the leakage current obtained in laboratory tests, subjecting the glass insulator string to different artificial pollution levels until flashover. Afterwards, the performance of the new proposed RUSBoost approach was evaluated and compared with AdaBoost, Bagging, Random Subspace Ensemble with k-nearest neighbors (KNN) and support vector machines (SVM) algorithms showing the effectiveness of RUSBoost in addressing the estimation of the highly imbalanced insulator condition indicators and its advantage over other methods.

The paper is organized as follows: Section 2 describes the experimental procedures. Section 3 presents the proposed approach, the results obtained, as well as the metrics used to evaluate the performance. Finally, Section 4 summarizes the main conclusions achieved.

2. Experimental procedures

2.1. Insulator samples

For this research, three identical strings composed of ten U160BSP insulator units were selected as specimens: the first string with non-coated glass insulators, the second one with silicone half-coated insulators, where the RTV silicone was applied to the bottom part, and the last one with silicone full-coated insulators. The anti-fog insulator type U160BSP is standardized as per IEC 60305 [27] and its design, with large under-ribs, provides longer creepage distance per unit and is suitable for polluted and/or coastal areas. The main dimensional features of the U160BS insulator type are presented in Table 1 and the drawing is shown in Fig. 1.

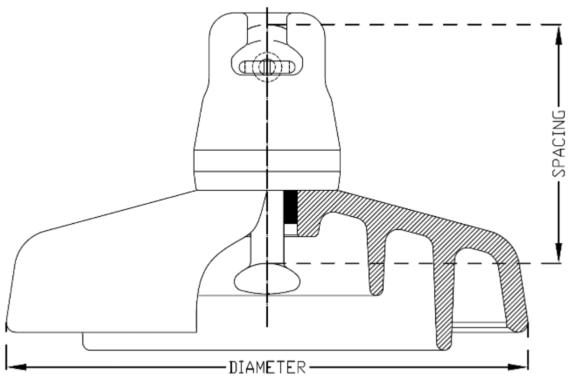

Fig. 1. Insulator type U160BSP anti-fog profile.

Table 1
Dimensional features

U160BSP anti-fog profile insulator	
Spacing	146 mm
Diameter	320 mm
Creepage distance	545 mm
▪ Top	30 %
▪ Bottom	70 %
Surface	3.428 cm^2
▪ Top	33 %
▪ Bottom	67 %
Protected creepage distance	355 mm
▪ Top	7 %
▪ Bottom	93 %

Fig. 2. Glass insulator string in test position.

Fig. 3. Insulator strings installed in the outdoor test station.

2.2. Laboratory tests

The artificial pollution tests were performed in the salt fog chamber of the FGH High Voltage laboratory in Mannheim, Germany. The dimensions of the chamber are 12x12x14 m (LxWxH) and the salt fog was prepared from salt (NaCl) mixed with de-ionized water made by an in-house reverse-osmosis system. The power frequency test transformer of 350 kV / 700 kVA met all the requirements of IEC 60507 [28] related to the minimum short-circuit current (Isc), resistance to reactance ratio (R/X) and capacitive current to short-circuit ratio (Ic/Isc) to ensure that the voltage drop during the tests was small and had no influence on the results.

The glass insulator string was mounted in vertical position, as shown in Fig. 2, and centred between the salt fog nozzles which were regulated individually by separate flow meters. The lower side of the insulator string was connected to the test voltage supply and the top was grounder over a measurement resistor for the leakage current detection. The leakage current was monitored during the whole test period and the flashovers were recorded by a digital oscilloscope. The objectives of these laboratory tests were to determine the maximum withstand salinity of the glass insulator string at the voltage level of 245/ √3 kV and to record the highest leakage current value in each test. Full and half silicone-coated strings were not tested since the test procedure is not applicable for polymeric surfaces and because the geometry, design and the string configuration were exactly the same. Therefore, the results obtained from the glass insulator string can be extrapolated as the worst in-service case condition corresponding to the complete loss of hydrophobicity for the silicone-coated strings. The salt fog method described in IEC 60507 was followed. The insulators were first thoroughly cleaned with water about 50 ºC mixed with detergent to remove any trace of dirt or grease the insulators may have. The preconditioning process consisted in subjecting the glass insulator string to the test voltage at the reference salinity of 80 kg/m^3 for 20 minutes and, then, raising the voltage in steps of 10 % of the test voltage every 5 minutes until flashover. After that, the string was re-energized and the voltage raised quickly to the 90% of the previously obtained flashover value and thereafter increased in steps of 5 % of the initial flashover voltage every 5 minutes until flashover.

This process was repeated until having eight flashovers. Then, the fog was cleared and the string washed down with tap water. The withstand tests were started immediately after the preconditioning process. A series of three 1-hour tests were performed at 245/ √3 kV per salinity level. The criteria were the following: if no flashover occurs at the end of all three withstand tests, it is considered as passed, and then the test is repeated at a higher salinity level. If only one flashover occurs, a fourth additional test is performed, and the test is passed if no flashover occurs. When the total number of flashovers reaches two, the test is not withstood and no further tests need to be carried out. At the end of each test the glass insulator string was completely cleaned using tap water. A total of eight different salinity levels were tested: 7, 20, 40, 80, 112, 136, 160 and 224 kg/m^3 and the leakage current was monitored and analyzed for each individual test.

2.3. Field monitoring program

The field monitoring program was carried out during two consecutive years at the Martigues test station established by Électricité de France (EDF). This facility is located in the south of France, facing the Mediterranean Sea, next to a thermal power plant, and in the vicinity of many petrochemical and heavy industries. The continuous exposure of the site to salt spray from the sea and industrial pollution led to very challenging outdoor conditions that were suitable for studying the pollution performance of the insulators. The Mediterranean climate of Martigues area corresponded to Csa subtype according to the Köppen–Geiger scale, with warm dry summers and mild and humid winters. As shown in Fig. 3, the glass insulator string and the two mirroring half-coated and full-coated insulator strings were installed in parallel and energized at 245/ √3 kV with the same conductor and resulting in a Unified Specific Creepage Distance (USCD) of 38.5 mm/kV. A fourth string with an experimental surface treatment was also installed, but it was not included in this work. The leakage current was monitored with individual sensors installed at the ground side of the string which guides the current through the measuring instrument associated to a data acquisition system [29]. The maximum current peaks exceeding 10 mA over a five-minute interval were recorded for each insulator string during the whole monitored period.

Table 2
Condition indicators, associated maintenance activities and detail of the highly imbalanced class distribution of the datasets

Condition indicator (CI)	Salinity range (kg/m^3)	Equivalent leakage current (mA)	Maintenance activity	Distribution of the field-monitored data		
				Non-coated string (%)	Half-coated string (%)	Full-coated string (%)
Excellent	0 – 10	0 – 41	No actions	96.52	99.55	99.82
Good	10 – 40	41 – 180	Surveillance	3.14	0.45	0.18
Medium	40 – 80	180 – 378	Preventive washing	0.32	-	-
Bad	80 – 160 (MWS)	378 – 794	Urgent washing	0.03	-	-

Two additional glass and full-coated insulator strings were installed at the back, in a non-energized area, to carry out complementary ESDD and NSDD pollution measurements. The station was equipped with one directional dust deposit gauge (DDDG) device in accordance with IEC 60815 [30] to gather the windborne dust from the north, south, east and west directions at monthly intervals. These measurements were intended to quantify the conductive pollutants in the air. Finally, a dedicated weather station was installed for monitoring the relevant meteorological parameters such as temperature, relative humidity, dew point, wind speed, wind direction and solar radiation every five-minutes.

3. Methodology

3.1. Condition Indicators set-up

Leakage current activity across an insulator string increases as the insulators become more polluted thus indicating a higher risk of having a flashover. The characterization of the leakage current, and the related condition indicators (CI), was carried out at the laboratory subjecting the glass insulator string to series of three 1-hour tests at different pollution levels in terms of salinity. The relationship between the highest leakage current peak (I_h) and the artificial salinity (S) level is shown in Fig. 4.

The salt fog tests were performed with increasing salinities to determine the maximum withstand salinity (MWS) and the corresponding maximum withstand leakage current. This information was used for adjusting the condition indicators properly as the results may differ for other insulator types, string arrangements or voltage levels. In the present investigation, the MWS was determined at 160 kg/m^3 for the non-coated glass insulator string energized at 245/ $\sqrt{3}$ kV. The results obtained from the glass insulator string were extrapolated as the worst in-service case condition corresponding to the complete loss of hydrophobicity for the silicone-coated strings because the geometry, design and configuration were the same [31].

The condition indicators were established, accordingly, considering the MWS and the related leakage current as the maximum permissible in service. Then, four classes were set-up linked to the most widely used maintenance activities for insulators as presented in Table 2. It is important to observe that the configuration of these condition indicators, i.e., the number of classes or their width in terms of salinity or current, can be tailor-made for each overhead power line based on the available maintenance procedures as well as the service reliability and design requirements imposed by the utility.

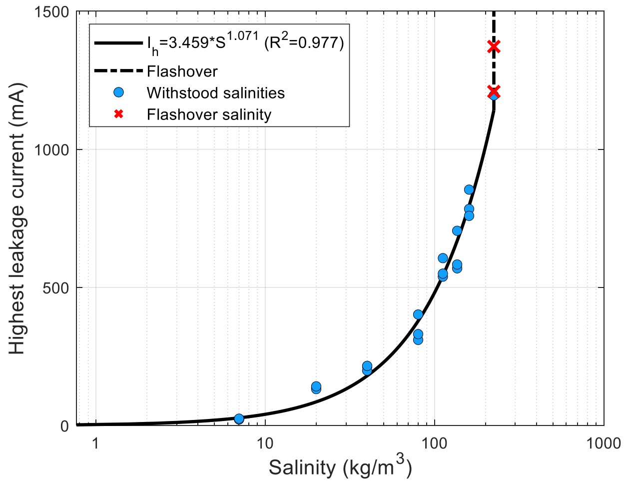

Fig. 4. Relationship between the maximum leakage current and salinity for the non-coated glass insulator string tested in laboratory.

3.2. Field-monitored data and inputs variables

Field-monitored data of the three energized insulator strings, weather and environmental conditions were collected during two consecutive years. The condition indicators, based on leakage current ranges, presented a highly imbalanced distribution among classes as shown in Table 2. This is directly connected to the behaviour of the leakage current in polluted conditions which is in the form of discrete rare events. The occurrence, duration and current level achieved during such events rely on the wetting effects over the accumulated pollution on the insulator string. This situation favours the prevalence of the CI Excellent as the majority class since most of the time the strings were in dry conditions. The minority classes, i.e., CI Good, CI Medium and CI Bad, represented a very tiny percentage of the data because of the uneven combination of pollution and wetting leading to relevant leakage current activity. Half and full silicone-coated insulator strings did not reach CI Medium and CI Bad levels on account of the hydrophobicity transfer properties of the silicone which inhibited the formation of continuous water paths and the related leakage currents.

Part of the weather and environmental data needed to be processed to be used as inputs for the model. In this regard, the Cumulative Pollution Index (CPI), which is the real-time estimation of the soluble pollution deposited on the insulator strings, was obtained from wind, directional dust from the DDDG and rain data as described in [25]. The CPI is based on the physics of pollution deposition, with the cube of the wind speed, and the natural self-cleaning of the insulators, as an exponential function of rain intensity and duration. The CPI can be linearly converted to ESDD data to make its interpretation easier. In this respect, it shall be noted that silicone-coated insulators tend to collect more pollution than the non-coated ones because of the different surface roughness. However, thanks to the hydrophobicity transfer properties of the silicone, this does not result in worse pollution performance. Relative humidity, temperature and solar radiation were other key weather inputs variables monitored. The summary statistics for the inputs variables is shown in Table 3

Data visualization revealed key relationships among the condition indicators and the inputs. For instance, as shown in Fig. 5 the most critical conditions for the non-coated insulator string, in terms of the CI Medium and Bad, were reached at the higher ESDD levels, around 0.1 mg/cm2, and with the relative humidity above 80%. By contrast, the full and half silicone-coated strings, despite collecting more pollution because of the roughness of the silicone, only reached the CI Good under similar wetting conditions evidencing the effectiveness of the coating. Other interesting relationships were observed among the condition indicators and the difference between the ambient temperature and the dew point, i.e., dew point depression. The smaller this difference, the more moisture content there is in the air. As shown in Fig. 6, when it is analyzed in combination with the solar radiation and the time of day, it provides some useful insights into the performance of the strings under moisture condensation. Solar radiation thermally heated the insulators, particularly during the peak sun-hours, and it had a direct effect on the wetting of the pollution deposits and the resulting condition indicators.

Table 3 Summary statistics for inputs variables

Variable	Mean ± SD	Min	Max	Percentiles				
				5th	25th	50th	75th	95th
Relative humidity (%)	74.01 ± 14.02	18.28	98.00	47.00	65.21	76.21	85.00	92.62
Temperature (ºC)	14.63 ± 5.86	-3.90	32.86	4.89	10.70	14.45	18.86	24.20
Solar radiation (W/m^2)	161.09 ± 245.20	0.00	1200.50	0.00	0.00	2.35	272.12	732.29
ESDD monitored data linearly converted from the CPI (mg/mm^2):								
▪ Non-coated string	0.05 ± 0.03	0.00	0.12	0.01	0.03	0.05	0.07	0.10
▪ Half-coated string	0.09 ± 0.05	0.00	0.24	0.02	0.06	0.09	0.14	0.18
▪ Full-coated string	0.10 ± 0.06	0.00	0.25	0.02	0.06	0.09	0.15	0.19

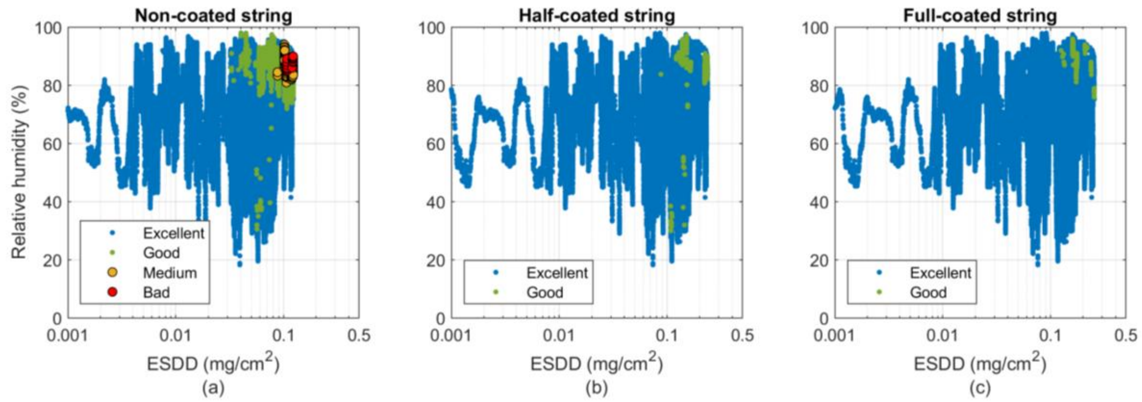

Fig. 5. Condition indicators as a function of relative humidity and ESDD converted from the CPI.

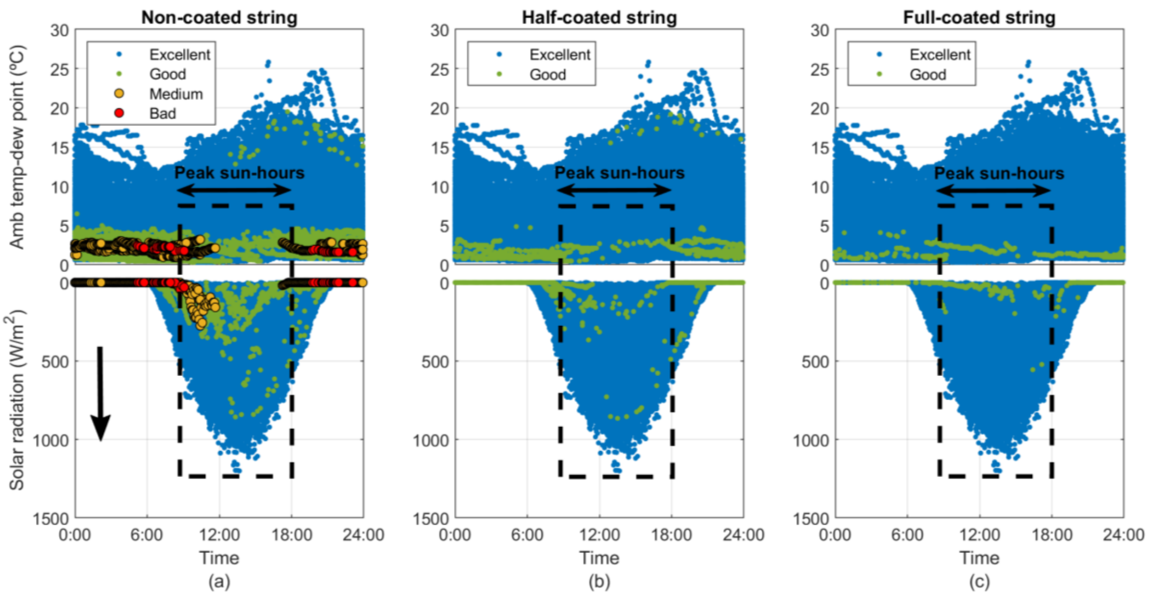

Fig. 6. Condition indicators as a function of the difference between temperature and dew point, solar radiation and time of day.

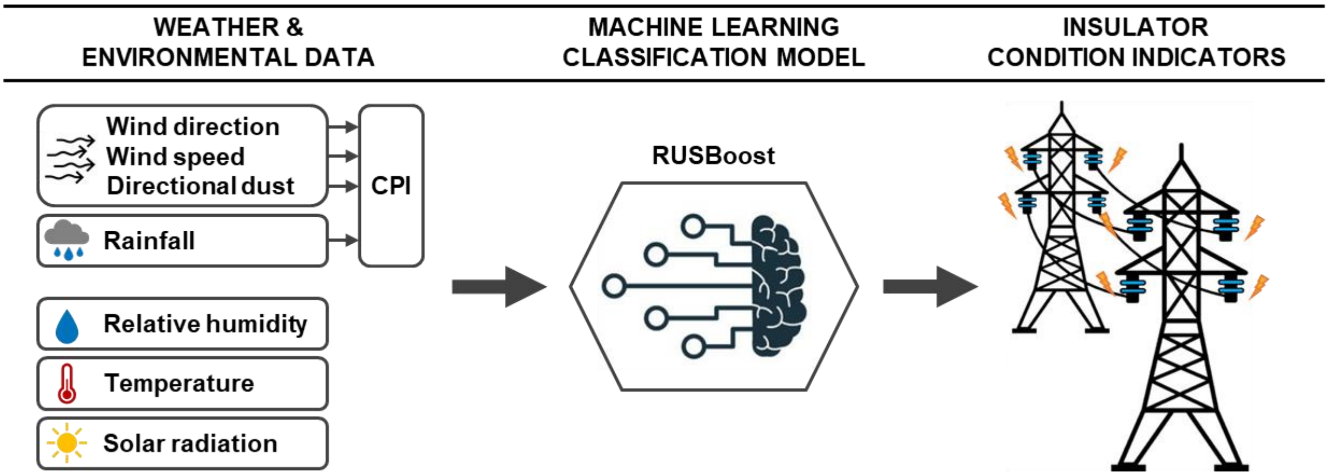

Fig. 7. Proposed machine learning approach for condition monitoring of high-voltage insulators in polluted environments.

Algorithm 1 RUSBoost adapted from [32]

Given: Set S of examples $(x_1, y_1), \ldots, (x_m, y_m)$ with minority class $y^r \in Y$, $|Y| = 4$

Weak Learner using decision trees, *WeakLearn*

Number of iterations, T

Desired percentage of total instances to be represented by the minority class, N

1. Initialize $D_1(i) = \frac{1}{m}$ for all i.

2. for $t = 1, 2, \ldots, T$ do
   a. Create temporary training dataset S'_t with distribution D'_t using random undersampling
   b. Call *WeakLearn*, providing it with examples S'_t and their weights D'_t.
   c. Get back a hypothesis $h_t : X \times Y \rightarrow [0,1]$.
   d. Calculate the pseudo-loss (for S and D_t):
   $$\varepsilon_t = \sum_{(i,y):y_i \neq y} D_t(i)(1 - h_t(x_i, y_i) + h_t(x_i, y)).$$
   e. Calculate the weight update parameter:
   $$\alpha_t = \frac{\varepsilon_t}{1 - \varepsilon_t}.$$
   f. Update D_t:
   $$D_{t+1}(i) = D_t(i)\alpha_t^{\frac{1}{2}(1 + h_t(x_i, y_i) - h_t(x_i, y: y \neq y_i))}$$
   g. Normalize D_{t+1}: Let $Z_t = \sum_i D_{t+1}(i)$.
   $$D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_t}.$$

   end for

3. Output the final classifier:
$$H(x) = \arg\max_{y \in Y} \sum_{t=1}^{T} h_t(x, y) \log \frac{1}{\alpha_t}.$$

3.3. Machine learning classification approach

The purpose of the machine learning (ML) classification approach is the prediction of the condition indicators for the insulator strings from the weather and environmental data as summarized in Fig. 7. The complex relationships between the condition indicators and the weather conditions cannot be depicted with analytical expressions and ML opens up new possibilities in the matter. These algorithms learn and improve their performance from data, however, when dealing with imbalanced datasets, most of the existing ML classification algorithms introduce a bias in favour of the majority class and overlook the minority class instances as potential outliers. The data collected in the field presented a highly imbalanced distribution and the ML classifier must be designed taking into account these particular circumstances in order to accurately classify the data samples of the minority classes which are, precisely, the classes of interest for this application. Furthermore, this classification problem is characterized by low variance and high bias and a relative reduced dataset for learning and, in this case, ML algorithms of the boosting family are recommended to use.

To handle adequately this problem the RUSBoost hybrid resampling/boosting algorithm was used [32]. It combines the random under sampling (RUS) technique with an extension of the adaptative boosting algorithm for multiple classes (AdaBoost.M2) [33]. In this method, the RUS technique is employed as a preprocessing stage to randomly remove majority class instances to balance the classes before applying the learning algorithm. This is appropriate because, despite removing data, the loss of information is very limited and without any significant impact due to the great similarity among the data collected for the majority class. Potential alternatives to RUS based on oversampling can include more sophisticated methods such as the synthetic minority oversampling technique (SMOTE) [34] which generates new minority instances by interpolating between several minority class instances that lie relatively close to each other. However, the very few instances of the minority classes, linked to the high degree of data imbalance, may lead to overfitting, poor generalization and a lack of effectiveness which make them less suitable for this particular application.

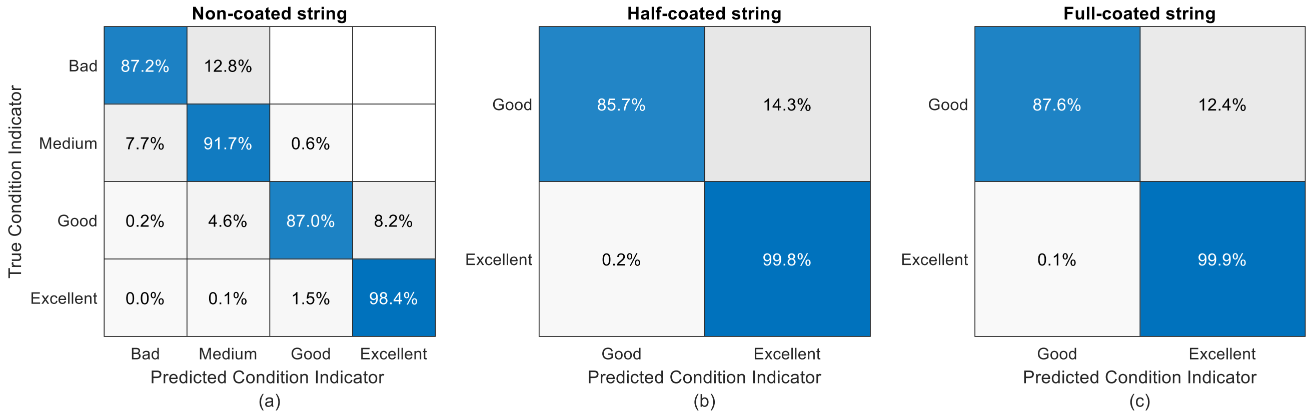

Fig. 8. Confusion matrices of RUSBoost for the three classifiers, i.e., (a) non-coated, (b) half-coated and (c) full-coated insulator strings.

As presented in Algorithm 1, the RUSBoost method trains an ensemble of decision trees, which are weak classifiers, using a random under-sampled subset of the data. During the training, the weight of every sample is adjusted iteratively so that the weights of the misclassified instances are increased while the correctly classified instances are decreased. Given the higher weights to the misclassified instances in the boosting method, they are more prone to be correctly classified in the subsequent iterations. The final classifier output is obtained by an ensemble of majority learners. The accuracy of the model can be further enhanced tuning the model hyperparameters, i.e., the number of ensemble learning cycles, learning rate for shrinkage and the maximal number of decision splits (branch nodes). Instead of the manual tuning of such hyperparameters which is a time-consuming task, the Bayesian optimization was employed to obtain the optimal values. The proposed ML classification approach was trained and validated using the ten-fold cross validation procedure on the three datasets obtained from the field monitoring program. The benefit of k-fold cross validation lies in the fact that the entire dataset is used for training and testing simultaneously and each observation is used for testing once [35], making efficient use of the data.

4. Results and discussion

4.1. Performance metrics

The normalized classification confusion matrices were computed for the three classifications models and the results are presented in Fig. 8. This visual technique summarizes the performance of the three classification models where the diagonal elements of the matrix represent the correctly classified elements. It can be observed that the multi-class classifier built for the non-coated glass insulator string presented quite good results and the misclassifications were mostly for neighboring condition indicators. Note that it was one of the main targeted factors when designing the ML classification model. Half and full silicone-coated strings operated the whole monitored period under two condition indicators only and, therefore, their respective classifiers were binary. In both cases the results were very similar presenting a high ratio of true positive values.

Four metrics were used for the evaluation performance of the three classification models: specificity, recall or sensitivity, precision and F-score, which can be also adapted to accommodate multi-class problems [36]. For a given class C_i the metrics are built based on the type of classification, i.e., true positives (TP_i), true negatives (TN_i), false positives (FP_i) and false negatives (FN_i):

$$Specificity_i = \frac{TN_i}{TN_i + FP_i} \tag{1}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{2}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

The F-score is calculated from the recall and precision and it is widely used for the evaluation of imbalanced datasets.

$$F\text{-}score_i = \frac{2 \cdot recall_i \cdot precision_i}{recall_i + precision_i} \tag{4}$$

In the multi-class case, the overall classification was assessed by means of macro-averaging which is based on computing the metric independently for each class and taking the average to treat all classes equally. The performance of the proposed RUSBoost classifiers were compared with support vector machines (SVM) and with three traditional ensemble methods: standard AdaBoost, Bagging and Random Subspace Ensemble with KNN algorithm for classification [37]. In all cases, similar ten-fold validation scheme was used as well as the Bayesian optimization to obtain the hyperparameters. The metrics were calculated and the results are presented in Tables 4, 5 and 6 for the three classifiers, i.e., non-coated, half-coated and full coated insulator strings. These results show the advantage of RUSBoost over the other four methods by achieving a macro-averaged F-score of 0.757 for the non-coated string and a F-score of 0.768 for the half-coated string and 0.792 for the full coated string. Furthermore, when it comes to dealing with the high criticality classes of the multi-class model for the non-coated string, namely CI Medium and CI Bad, RUSBoost notably outperformed the other methods, which had serious limitations in handling those classes as a consequence of the highly imbalanced class distribution of the datasets.

Table 4
Performance comparison of RUSBoost with other machine learning classifiers developed for the non-coated insulator string

Condition indicator (CI)	Performance metrics	Non-coated insulator string				
		RUSBoost	AdaBoost	Bagging	Subspace KNN	SVM
Excellent	Specificity	0.926	0.879	0.882	0.608	0.870
	Recall	0.984	0.993	0.996	0.999	0.994
	Precision	0.997	0.996	0.996	0.986	0.995
	F-score	0.991	0.994	0.996	0.992	0.995
Good	Specificity	0.985	0.993	0.996	0.999	0.994
	Recall	0.870	0.863	0.854	0.601	0.851
	Precision	0.650	0.807	0.863	0.937	0.821
	F-score	0.744	0.834	0.859	0.732	0.836
Medium	Specificity	0.998	0.999	0.999	0.999	0.999
	Recall	0.917	0.665	0.648	0.575	0.690
	Precision	0.577	0.630	0.639	0.873	0.779
	F-score	0.708	0.647	0.643	0.693	0.701
Bad	Specificity	0.999	0.999	0.999	0.999	0.999
	Recall	0.872	0.359	0.308	0.308	0.231
	Precision	0.370	0.264	0.250	0.364	0.209
	F-score	0.519	0.304	0.276	0.333	0.220
Macro-average	Recall	0.911	0.720	0.701	0.621	0.691
	Precision	0.648	0.674	0.687	0.790	0.701
	F-score	**0.757**	0.696	0.694	0.695	0.696

Table 5
Performance comparison of RUSBoost with other machine learning classifiers developed for the half-coated insulator string

Condition indicator (CI)		Performance metrics	Half-coated insulator string				
			RUSBoost	AdaBoost	Bagging	Subspace KNN	SVM
Excellent	Good	Specificity	0.998	0.999	0.999	0.999	0.999
		Recall	0857	0.681	0.657	0.468	0.538
		Precision	0.696	0.734	0.742	0.813	0.729
		F-score	**0.768**	0.707	0.697	0.594	0.619

Table 6
Performance comparison of RUSBoost with other machine learning classifiers developed for the full-coated insulator string

Condition indicator (CI)		Performance metrics	Full-coated insulator string				
			RUSBoost	AdaBoost	Bagging	Subspace KNN	SVM
Excellent	Good	Specificity	0.999	0.999	0.999	0.999	0.999
		Recall	0.876	0.661	0.664	0.420	0.445
		Precision	0.723	0.784	0.752	0.846	0.649
		F-score	**0.792**	0.717	0.705	0.561	0.528

4.2. Predictor importance

Finally, Figure 9 shows the relative importance of each input of the model, i.e.: the relative humidity, the CPI (which can be linearly converted into ESDD) the solar radiation and the temperature. It was found that the relative humidity and the pollution, in terms of CPI, are the most relevant ones. Solar radiation and temperature, related to wetting phenomena by condensation, had less significance importance for the prediction of the Condition Indicators.

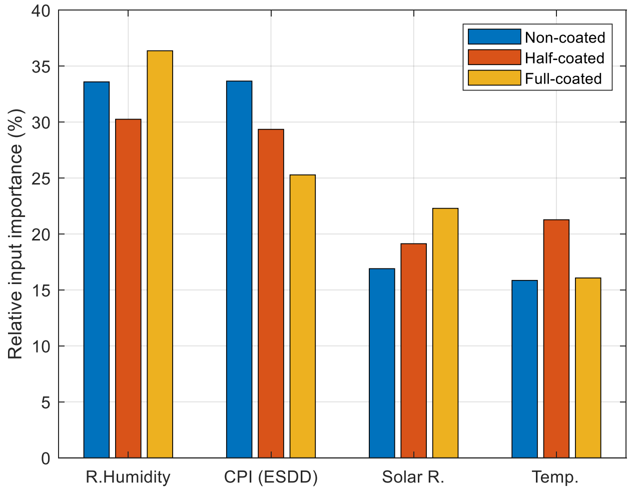

Fig. 9. Relative importance of the inputs.

5. Conclusions

This paper has proposed a novel machine learning classification approach for insulator condition monitoring based on the combination of the random under sampling technique with an adaptative boosting algorithm and aiming to estimate key condition indicators from the meteorological and environmental data. The investigation is supported by laboratory tests and data collected during two consecutive years in an outdoor test station facility located in France. The main findings are summarized below:

I. The definition of key condition indicators linked to maintenance activities, such as washing, was carried out through the characterization and analyses of the leakage current obtained in laboratory tests, where the glass insulator string was subjected to different artificial pollution levels until flashover.

II. Field-monitored data collected during two years revealed that the condition indicators presented a very highly imbalanced distribution among classes. The prediction of such condition indicators from the weather and environmental data was addressed through the RUSBoost algorithm. The results show the effectiveness of the new proposed RUSBoost approach in addressing the estimation of the highly imbalanced insulator condition indicators and its advantage over other methods by achieving a macro-averaged F-score of 0.757 for the non-coated string and a F-score of 0.768 for the half-coated string and 0.792 for the full coated string.

III. The proposed ML classification approach proved to be valid for insulators made from different materials such as glass and RTV silicone-coated insulators with hydrophobicity transfer properties. Full and half silicone-coated strings had similar performance and did not reach condition indicators linked to washing activities in contrast to the glass insulator string.

IV. Insulator maintenance can be sharply improved by means of the proposed approach to carry out the maintenance in a dynamic and predictive manner instead of the regular washing at fixed time intervals.

References

[1] CIGRE Working Group C4.303. Outdoor insulation in polluted conditions: Guidelines for selection and dimensioning. Part 1: General principles and the AC case. Tech Broch 361 2008.
[2] CIGRE Task Force 33.04.01. Polluted insulators: A review of current knowledge. Tech Broch 158 2000.
[3] IEEE Std 957. IEEE Guide for Cleaning Insulators 2005. https://doi.org/10.1109/IEEESTD.2005.339594.
[4] Cherney EA, Gorur R, Marzinotto M, El-Hag A, Meyer L, George JM, et al. RTV Silicone rubber pre-coated ceramic insulators for transmission lines. IEEE Trans Dielectr Electr Insul 2013;20:237–44. https://doi.org/10.1109/TDEI.2013.6451363.
[5] Liang X, Li S, Gao Y, Su Z, Zhou J. Improving the outdoor insulation performance of Chinese EHV and UHV AC and DC overhead transmission lines. IEEE Electr Insul Mag 2020;36:7–25. https://doi.org/10.1109/MEI.2020.9111096.
[6] de Santos H, Á. Sanz-Bobi M. Research on the pollution performance and degradation of superhydrophobic nano-coatings for toughened glass insulators. Electr Power Syst Res 2021;191:106863. https://doi.org/10.1016/j.epsr.2020.106863.
[7] Lannes W, Schneider H. Pollution severity performance chart; key to just-in-time insulator maintenance. IEEE Trans Power Deliv 1997;12:1493–500. https://doi.org/10.1109/61.634166.
[8] CIGRE Task Force 33.04.03. Insulator pollution monitoring. Electra 1994:79–90.
[9] Phillips AJ, Bologna FF, Major JM, Engelbrecht CS. Development and demonstration of low cost robust leakage current sensors for evaluating contaminated insulators. Proc. 16th Int. Symp. High Volt. Eng., 2009, p. 1–6.
[10] Harid N, Bogias AC, Griffiths H, Robson S, Haddad A. A wireless system for monitoring leakage current in electrical substation equipment. IEEE Access 2016;4:2965–75. https://doi.org/10.1109/ACCESS.2016.2577553.
[11] Ramirez I, Hernández R, Montoya G. Measurement of leakage current for monitoring the performance of outdoor insulators in polluted environments. IEEE Electr Insul Mag 2012;28:29–34. https://doi.org/10.1109/MEI.2012.6232007.
[12] Gao S, Zhou Z, Jiang H, Wang L. Research and application on power grid contamination flashover pre-warning technology. 2014 China Int. Conf. Electr. Distrib., 2014, p. 1654–9. https://doi.org/10.1109/CICED.2014.6991987.
[13] Salem AA, Abd-Rahman R, Al-Gailani SA, Salam Z, Kamarudin MS, Zainuddin H, et al. Risk assessment of polluted glass insulator using leakage current index under different operating conditions. IEEE Access 2020;8:175827–39. https://doi.org/10.1109/ACCESS.2020.3026136.
[14] Castillo Sierra R, Oviedo-Trespalacios O, Candelo JE, Soto JD. The influence of atmospheric conditions on the leakage current of ceramic insulators on the Colombian Caribbean coast. Environ Sci Pollut Res 2015;22:2526–36. https://doi.org/10.1007/s11356-014-3729-3.
[15] Wang J, Xi Y, Fang C, Cai L, Wang J, Fan Y. Leakage current response mechanism of insulator string with ambient humidity on days without rain. IEEE Access 2019;7:55229–36. https://doi.org/10.1109/ACCESS.2019.2910660.
[16] Waluyo, Fauziah D, Khaidir IM. The evaluation of daily comparative leakage currents on porcelain and silicone rubber insulators under natural environmental conditions. IEEE Access 2021;9:27451–66. https://doi.org/10.1109/ACCESS.2021.3057626.
[17] Vosloo WL, Holtzhausen JP. The prediction of insulator leakage currents from environmental data. IEEE Africon Conf. Africa, 2002. https://doi.org/10.1109/afrcon.2002.1159978.
[18] Zhao L, Jiang J, Duan S, Fang C, Wang J, Wang K, et al. The prediction of post insulators leakage current from environmental data. 2011 Int Conf Electr Control Eng 2011:5103–6. https://doi.org/10.1109/ICECENG.2011.6057235.
[19] Ahmad AS, Ghosh PS, Shahnawaz Ahmed S, Aljunid SAK. Assessment of ESDD on high-voltage insulators using artificial neural network. Electr Power Syst Res 2004;72:131–6.

https://doi.org/10.1016/j.epsr.2004.03.009.
[20] Karamousantas DC, Chatzarakis GE, Oikonomou DS, Ekonomou L, Karampelas P. Effective insulator maintenance scheduling using artificial neural networks. IET Gener Transm Distrib 2010;4:479–84. https://doi.org/10.1049/iet-gtd.2008.0657.
[21] Thanh-Phuong Nguyen, Chao-Tsung Yeh, Ming-Yuan Cho, Chun-Lung Chang M-JC. Convolutional neural network bidirectional long short-term memory to online classify the distribution insulator leakage currents. Electr Power Syst Res 2022;208:107923.
[22] Phuong Nguyen Thanh M-YC. Online leakage current classification using convolutional neural network long short-term memory for high voltage insulators on web-based service. Electr Power Syst Res 2023;216:109065.
[23] Shangbin Jiao DL, Guo Xie YD. Assessment of contamination condition of insulator based on PSO-SVM. 2009 4th IEEE Conf Ind Electron Appl 2009:358–62. https://doi.org/10.1109/ICIEA.2009.5138228.
[24] Ren A, Li Q, Xiao H. Influence analysis and prediction of ESDD and NSDD based on random forests. Energies 2017;10:878. https://doi.org/10.3390/en10070878.
[25] de Santos H, Sanz-Bobi MÁ. A Cumulative Pollution Index for the estimation of the leakage current on insulator strings. IEEE Trans Power Deliv 2020;35:2438–46. https://doi.org/10.1109/TPWRD.2020.2968556.
[26] Fernández A, García S, Galar M, Prati RC. Learning from Imbalanced Data Sets. 10.1007/978-3-319-98074-4; 2018. https://doi.org/10.1007/978-3-319-98074-4.
[27] Insulators for overhead lines with a nominal voltage above 1000 V - Ceramic or glass insulator units for a.c. systems - Characteristics of insulator units of the cap and pin type. IEC Stand 60305 1995.
[28] Artificial pollution tests on high-voltage ceramic and glass insulators to be used on a.c. systems. IEC Stand 60507 2013.
[29] GIGRE Working Group B2.03. Guide for the establishment of naturally polluted insulator testing stations. Tech Broch 333 2007.
[30] Selection and dimensioning of high-voltage insulators intended for use in polluted conditions - Part 1: Definitions, information and general principles. IEC Tech Specif 60815-1 2008.
[31] CIGRE Working Group B2.69. Coating for improvement of electrical performance of outdoor insulators under pollution conditions. Tech Broch 837 2021.
[32] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Trans Syst Man, Cybern - Part A Syst Humans 2010;40:185–97. https://doi.org/10.1109/TSMCA.2009.2029559.
[33] Schapire RE, Freund Y. Boosting: Foundations and Algorithms. MIT Press; 2012.
[34] Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique Nitesh. J Artif Intell Res 2002;16:321–57. https://doi.org/10.1613/jair.953.
[35] Vakharia V, Gujar R. Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques. Constr Build Mater 2019;225:292–301. https://doi.org/10.1016/j.conbuildmat.2019.07.224.
[36] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45:427–37. https://doi.org/10.1016/j.ipm.2009.03.002.
[37] Rokach L. Ensemble Learning: Pattern Classification Using Ensemble Methods. Second Edi. World Scientific Publishing Company Incorporated; 2019. https://doi.org//10.1142/11325.