Electric Power Components and Systems, 42(2):180–189, 2014

ISSN: 1532-5008 print / 1532-5016 online
DOI: 10.1080/15325008.2013.853217

Taylor & Francis
Taylor & Francis Group

A Hybrid Support Vector Fuzzy Inference System for the Classification of Leakage Current Waveforms Portraying Discharges

Konstantinos Theofilatos,[1] Dionisios Pylarinos,[2] Spiros Likothanassis,[1] Damianos Melidis,[1] Kiriakos Siderakis,[3] Emmanuel Thalassinakis,[4] and Seferina Mavroudi[5]

[1]Pattern Recognition Lab, Department of Computer Engineering and Informatics, University of Patras, Greece
[2]Researcher/Consultant for P.P.C./HEDNO, Terma Kastorias Str, Katsambas, Iraklion, Greece
[3]Electrical Engineering Department, School of Applied Technology, Technological Educational Institute of Crete, Iraklion, Greece
[4]Public Power Corporation/HEDNO, Terma Kastorias Str, Katsambas, Iraklion, Greece
[5]Department of Social Work, School of Sciences of Health and Care, Technological Educational Institute of Patras, Greece

CONTENTS

Keywords: leakage current, insulators, field, discharges, feature selection, classification, fuzzy logic, genetic algorithms, support vector machines
Received 14 March 2013; accepted 5 October 2013

Address correspondence to Dr. Dionisios Pylarinos, Vartholomio, Ilia, 27050, Greece. E-mail: dpylarinos@yahoo.com

Abstract—Several techniques have been applied on leakage current waveforms in order to extract information regarding electrical activity on high-voltage insulators. However, a fully representative value is yet to be defined. In this article, a hybrid support vector fuzzy inference system is introduced as a classification tool. The system incorporates fuzzy logic, genetic algorithms, and support vector machines. Apart from the classification accuracy achieved, the system also produces a set of fuzzy rules under which the classification is made, allowing a further insight of the process. A comparison is made to other classification tools previously applied on the same data set.

INTRODUCTION

The performance of insulators is a matter of great concern for system operation. A single insulator failure, especially when the insulator is located in a high-voltage (HV) station or substation, can result to an excessive outage of the power system. Several factors connected with local operation conditions affect the insulators' performance, with pollution being probably the most significant one [1–3]. Several standardized tests are employed in order to investigate insulators' performance in the lab, *e.g.*, [4–6]. However, since insulators' performance is strongly correlated to environmental conditions, field testing is also employed with a guide for the establishment of HV insulator test stations having recently been published [7].

Leakage current measurements are commonly employed to monitor and investigate the performance of insulators in both lab and field [8]. The basic stages of activity have been well correlated with certain waveform shapes during lab tests [9–12]. An investigation of field waveforms recently showed that similar shapes are recorded and should be expected in the field, along with waveforms rarely recorded in the lab [13–16]. Several techniques have been applied on leakage current waveforms in order to extract and record information connected to surface activity. In the field, the most commonly extracted values are the peak value, the charge, and the number of pulses exceeding pre-defined thresholds, whereas the harmonic content is commonly investigated in lab measurements [8]. However, it is commonly accepted that it is the shape of the leakage current waveform that corresponds to the experienced electrical activity, and a fully representative value of the waveforms' shape is yet to be defined. Several signal analysis and classification techniques have been applied on LC measurements with different values considered as inputs and aiming to different goals and a thorough review can be found in [8].

Recently, a new approach has been proposed for the classification of leakage current waveforms [13, 17, 18]. According to this approach, 20 different features are extracted from the leakage current waveform in order to be used for the classification. The features selected are commonly used in the literature [8] and equally represent the time and the frequency domain (ten features from each domain). Classification techniques are used to classify each waveform in two different classes depending on the duration of discharges [13, 17, 18]. At first, a linear classification was attempted employing a Euclidian classifier and a simple genetic algorithms (GAs) approach, and results were not that encouraging [17]; this was attributed to the non-linearity of the problem and the absence of an effective feature selection scheme. Then, non-linear classification techniques were employed, including three different classification algorithms (k-nearest neighbors [$k$nn], Naïve Bayes, support vector machines [SVMs]) and two feature extraction techniques (student's t-test and minimum redundancy maximum relevance [mRMR]) [13]. Results showed the superior performance of SVMs and of the feature set provided by the mRMR algorithm. Then, a new GAs approach was applied [18] with GAs used for both feature selection and classification and the accuracy percentage achieved was significantly higher compared to the previous GA approach [17] and slightly inferior to the SVM-mRMR approach [13].

Although the mRMR-SVM classification scheme offered the best results, there were still some drawbacks. Specifically, mRMR is a multivariate filtering feature selection technique which could not incorporate the technical characteristics of the classifier in its feature selection mechanism. Moreover, SVM classifiers are highly non-linear classifiers and the extracted models can not be interpreted. For these reasons, in the present article an evolutionary hybrid methodology is proposed, which deploys a GA to optimize the following: the feature subset, which should be used as an input to the classifier, the parameters of the SVM classifier, and the parameters of a methodology which extracts interpretable fuzzy classification rules directly from the extracted SVM model. The aim of the study remains the classification of field waveforms portraying discharges in two different classes depending on the duration of discharges. A comparison with results from previous implementations is shown and discussed. Besides the superior classification achieved, the system outputs a set of fuzzy classification rules that offer, for the first time, an insight of the classification process.

2. EXPERIMENTAL SETUP

The LC waveforms investigated in this article have been recorded in two 150 kV substations of the transmission system of Crete, in Greece during a period exceeding six years. The Cretan Transmission Network is exposed to intense marine pollution and several techniques have been employed by the Greek Public Power Corporation (PPC) to cope with the problem [19–22], including the construction of a HV Test Station in Iraklion, Crete [23–25].

The waveforms investigated in this article have been recorded on 18 different 150-kV post insulators (porcelain, RTV SIR coated, and composite) that were part of the grid [13, 14, 17, 18]. A collection ring was installed at the bottom side of each monitored insulator and the current was driven through a Hall current sensor. The acquired data was then transmitted to a commercially available data acquisition system (DAQ). Sampling was performed continuously and simultaneously for all monitored insulators, at a rate of 2 kHz and resolution of 12 bit. Each waveform recorded has a duration of 480 ms. The monitoring system incorporated the time-window technique [15, 16] to record waveforms. The waveform portraying the highest peak value in the considered time window is recorded. A schematic representation of the measuring system is shown in Figure 1. Detailed specifications of the DAQ, pictures, and more information can be found in [13–18].

3. PROBLEM DESCRIPTION

The correlation of surface activity with the shape of leakage current waveforms has been well established, especially in case of lab tests [1–3, 8–12]. The basic discrete stages consists of: sinusoid waveforms due to the presence of conductive film on the insulator surface, distorted sinusoid waveforms as an intermediate stage, and dry band discharges that causes a time lag of current onset. Recent research showed that the same basic waveforms' shapes should also be expected in the field [13]; however, the reality of field conditions results to an

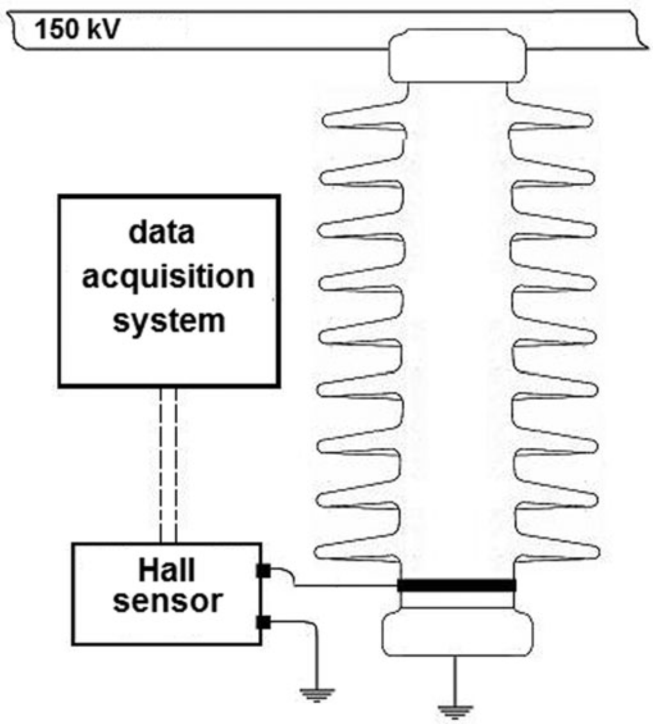

FIGURE 1. Schematic representation of the measuring system.

increased complexity of field recordings [13–16]. Some of the facts that should be considered is the presence of noise [15, 16], spikes [13], sinusoids of amplitude similar or greater than that of dischargers [13, 14], and also the complexity of discharge portraying waveforms [13, 14].

The data set considered in this article consists of 387 discharge portraying waveforms. The noise reduction/removal techniques described in [15, 16] have been applied in order to remove noise related waveforms, the S_R ratio [13, 16] has been used in order to remove isolated spikes and the D3/D5 ratio derived from wavelet analysis has been used to remove sinusoids waveforms [13], in order to isolate only discharge portraying waveforms. The waveforms have been classified in two different classes, depending on the duration of discharges. Class C1 includes waveforms that portray discharges that last four half-cycles or less, whereas class C2 includes waveforms that portray discharges that last five or more half-cycles. Some examples are shown in Figure 2. The waveforms in Figures 2(a) and 2(b) illustrate clearly the grouping criterion: If a waveform portrays discharges that last four or less consecutive half-cycles is identified as class C1 (Figure 2(a)), if the waveform portrays a discharge that lasts five or more consecutive half-cycles, then it is identified as class C2 (Figure 2(b)). However, such simple shapes are rather the exception and not the rule. Waveforms recorded in the field, frequently portray a complex shape, with two examples shown in Figures 2(c) and 2(d), whereas a greater selection of waveforms can be seen in [8, 14, 25]. It should be noted that all waveforms shown in Figure 2 have been selected so as to be similar in shape and peak value and yet belong to different classes, in order to underline the need of advanced techniques for the classification.

Recorded data are converted to mat files, using custom made software [26], and are then processed off line with the use of MATLAB, a software used in various applications in insulators' research [27–30]. A set of 20 features are extracted and used for the classification. The features can be seen in

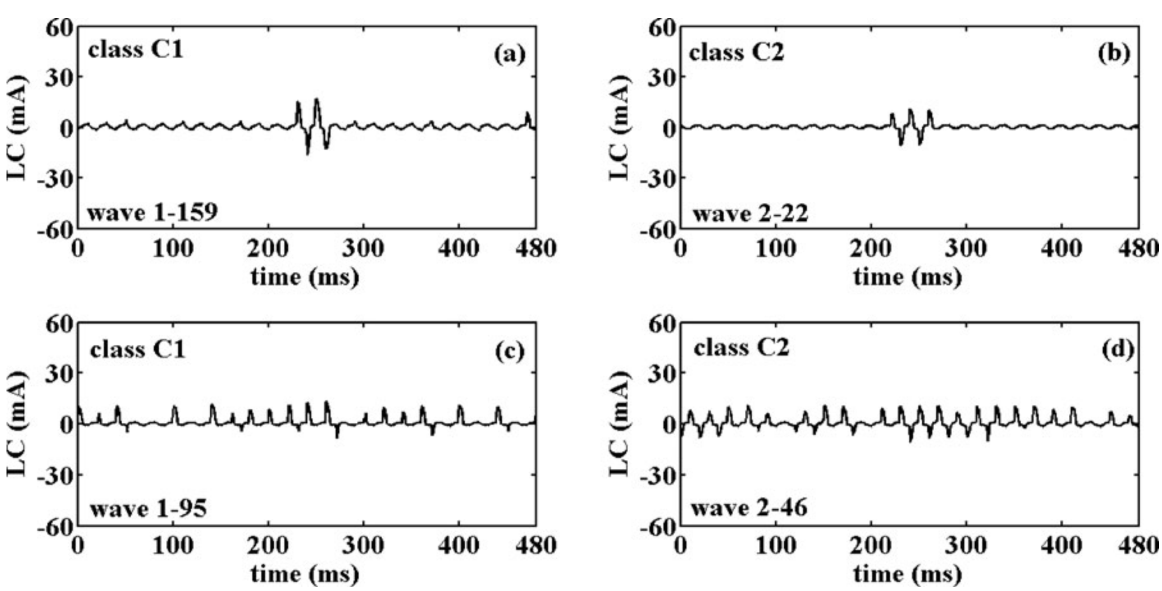

FIGURE 2. Waveforms from the investigated dataset: (a)–(c) class C1 and (b)–(d) class C2.

No.	Feature (time domain)	No.	Feature (frequency domain)
1	Amplitude	11	Third to first harmonic ratio: K3/K1
2	Mean	12	Fifth to first harmonic ratio: K5/K1
3	Median	13	Fifth to third harmonic ratio: K5/K3
4	Variance	14	Total harmonic distortion ratio: THD
5	Standard deviation (STD)	15	Harmonic distortion ratio: HD
6	Median absolute deviation (MAD)	16	STD_MRA VECTOR ratio: D1/D5
7	Skewness	17	STD_MRA VECTOR ratio: D2/D5
8	Kurtosis	18	STD_MRA VECTOR ratio: D3/D5
9	Interquartile range (IQR)	19	STD_MRA VECTOR ratio:D4/D5
10	Charge	20	Distortion ratio: D_R

TABLE 1. Employed features

Table 1. Features 1–10 derived from the time domain, and features 11–20 from the frequency domain, and have been selected in order to evenly represent both domains and also considering the literature [8].

Regarding the time domain features, frequently used values such as the amplitude and charge, along with commonly used [8] statistical values are employed. Regarding the feature domain features, it was considered that the content of odd harmonics is commonly correlated to the occurrence of discharges and the distortion of the waveforms' shape and therefore several commonly used ratios of odd harmonics [8, 12, 30] are employed. It should be noted that the fundamental frequency is 50 Hz and that the harmonic distortion (HD) ratio is similar to the total harmonic distortion (THD) ratio, with the numerator being the sum of the odd harmonics' content. Further, wavelet analysis and especially multi resolution analysis (MRA) is employed in order to acquire the standard deviation (STD) _MRA VECTOR [13, 16, 31, 32]. The STD_MRA VECTOR contains the STD of the details of each level of the wavelet MRA of the original waveform, with D1 referring to the first decomposition level, D2 to the second level etc [13, 16]. The distortion ratio [8] given by: $D_R = (D_1 + D_2 + D_3 + D_4)\big/D_5$, is also considered. The frequency bands of the STD_MRA VECTOR's components are shown in Table 2.

Decomposition level	(A) Approximation (Hz)	(D) Details (Hz)
1	0–500	500–1000
2	0–250	250–500
3	0–125	125–250
4	0–62.5	62.5–125
5	0–31.25	31.25–62.5
6	0–15.625	15.625–31.25

TABLE 2. Frequency bands of MRA

The same data set and features that have been used in previous implementations [13, 17, 18] are considered, so that comparisons can be made.

4. HYBRID SUPPORT VECTOR FUZZY INFERENCE SYSTEM

SVMs are considered as one of the most accurate machine learning classifiers [33], with a variety of applications including insulation evaluation (*e.g.*, [34, 35]). The SVM algorithm is a supervised learning method that addresses the problem of linear and non-linear classification by finding the maximum margin hyperplane that best separates the classes. Nonlinear SVMs map the training samples from the input space into a higher-dimensional feature space with the use of some mapping function, also known as the kernel function. Several kernel functions can be used and the radial base function has been employed in this article as being the most commonly used kernel function in non-linear classification problems. The mapping procedure resembles the hidden neuron layer of neural networks. However, SVMs do not suffer from local minima or overfitting, as neural networks do. They have the advantage of automatically selecting their model size and provide superior generalization ability by maximizing the margin of separation.

The main disadvantage of SVM classifiers is their black box nature, which does not allow user to extract interpretable inferences from the final classification models. Moreover, SVMs' performance deteriorates when non informative features are used as inputs raising the problems' dimensionality. Furthermore, the SVMs' parameters should be tuned effectively. Grid search and other heuristic approaches [36, 37] have been developed to solve this problem, however they are inefficient in terms of computational cost and they do not search in parallel for the optimal feature subset.

Fuzzy rules' language is considered to be among the closest computer languages to the natural human one. Thus, fuzzy rules can easily be interpreted by domain experts to extract useful conclusions. Furthermore, fuzzy systems present the ability to hide imprecise knowledge through fuzziness. This property allows for the extraction of novel knowledge from the initial row data. Furthermore, fuzzy systems can model non-linear functions. Their drawbacks include low classification performance, overfitting, and thus absence of generalization properties.

Extracting fuzzy rules from trained SVM classification models was a great challenge for the scientific community as this could be a step toward interpreting the high classification performance of SVMs and extracting useful domain information from them. In the last decade many approaches have been developed to accomplish this goal [38]. In the present article, the methodology which was proposed in [39], and has been successfully applied for other classification tasks [40, 41], is used. This methodology uses the technique proposed in [42] to describe the SVM classification model as a set of SVFI rules which are proved to be equivalent to the initial classification model.

The SVFI rules are in this form:

$$\text{Rule } k\text{: if } P_1^k \text{ and } P_2^k \text{ and } P_N^k \text{ then } C_k,$$

where $P_i^k, i = 1, \ldots, N$ are fuzzy clauses, having the form where x_i is *CloseToSV*(k,i). These fuzzy clauses examine the membership of the ith input value in the ith fuzzy set of the kth support vector. The support vectors are the training samples that are selected by the SVM algorithm to define the final classification hyperplane. The sets $CloseToSV(k, i)$ are the fuzzified numerical distance of x_i in the x_i^k component of the k-th support vector. A Gaussian function of the form $\mu_i^k(x_i) = \exp(-\frac{1}{2}(\frac{x_i^k - x_i}{\sigma_k})^2)$ estimates the membership function by quantifying the distance of the inputs component x_i from the value x_i^k of the ith component of SV_k. The parameters σ_K are real constant numbers ($\sigma_k \in R$).

The SVFI rules are large in number and hard to interpret. Thus, in [39] a methodology is proposed to derive a simpler fuzzy system that approximates the accurate set of rules keeping only the more important aspects of the data. This methodology not only reduces the extracted fuzzy rules but also replace *CloseToSV*(k, i) with linguistic clauses (low, medium, high). The fuzzy sets low, medium, and high which are used to linguistically represent the values of specific features have Gaussian participation functions with centers which should be either user-defined or algorithmically optimized. The performance of this method is highly depending of the optimal selection of its parameters β, δ and of the appropriate setting of the linguistic sets. β is a threshold used for discarding a fuzzy clause due to its membership value and δ is a threshold for discarding a fuzzy clause due to its significance (the higher its langrage multiplier the higher its significance). To the best knowledge of the authors, no effective analytical method exists thus far to locate optimal values for these parameters.

GAs are general optimization meta-heuristic algorithms based on the initial creation of a population of candidate solutions, called chromosomes, and their iterative differentiation using the operators of evaluation, selection, crossover, and mutation until some termination criteria are reached [43]. GAs have been proved useful and efficient in optimization problems where the search space is big and complicated or there is not any available mathematical analysis of the problem.

In the present article, GAs were used in the optimization of a variety of variables. These variables include feature variables to define if a feature should be used as input, the parameters C and gamma of the RBF-SVM classifier, the parameters β, δ and the centers of the linguistic clauses of the fuzzy rule extraction methodology [39]. Specifically, the chromosome of the proposed GA consists of 20 binary genes to determine which features should be used as inputs for the classifier and seven real-valued genes to optimize C, gamma, β, δ parameters and the centers of the three fuzzy sets low, medium, and high (Table 3). The feature selection genes take values 0 or 1 and force the classifier to use a specific feature as input if the feature value for this input is 1.

The crossover operator which was used was the one-point crossover with a crossover probability of 90%. As for the mutation operator (mutation probability: 10%), the binary mutation operator was applied for the feature genes and the Gaussian mutation operator is applied for the other genes because they are real valued genes. The binary mutation randomly alters a gene value from 0 to 1 and opposite. The Gaussian mutation operator adds a random number in a randomly selected gene. This random number is taken from the Gaussian distribution using as center the zero value and as width the interval of allowed values for this gene divided by 10.

Gene	Position in chromosome	Allowed values
Feature genes	1–20	0 or 1
Regularization parameter C	21	[0–1024]
RBF parameter gamma	22	[0–1024]
Threshold β	23	[0–1]
Threshold δ	24	[0–1]
Center of fuzzy set low	25	[0–1]
Center of fuzzy set medium	26	[0–1]
Center of fuzzy set high	27	[0–1]

TABLE 3. Chromosomes representation

Classification technique	Euclidian [16]	GA classification [16]	GA feature selection and classification [17]	SVM classification and mRMR feature selection [12]	Hybrid support vector fuzzy inference system
Best accuracy	77.20%	56.12%	88.48%	90.21%	94.36%
Number of features	20	20	11	10	9

TABLE 4. Best results for different classification techniques

The fitness function which was used to measure the performance of each individual is shown in Eq. (1):

$$\begin{aligned} Fitness = {} & Accuracy\_SVM \\ & + 0.5 \bullet Accuracy\_Interpretable\_Rules \\ & + 0.1 \bullet \left(1 - \frac{\#Interpretable\_Rules}{\#Initial\_SV\_Rules}\right) \\ & - 0.01 \bullet (\#Selected\_Features). \end{aligned} \quad (1)$$

The multipliers on the terms of the fitness function are selected to state the importance that is given in each different goal. Thus, the most important goal is the accuracy of the SVM classifier, with the accuracy of the interpretable rules, the complexity of the fuzzy rules, and the number of selected features being the other goals from the most important to the least significant one.

The population of the proposed evolutionary algorithm was set to 100 after thorough experimentation using the training set. The termination criteria of the algorithm were a combination of the maximum number of generations (1000) to be reached and a convergence criterion. The convergence criterion is satisfied when the fitness of the best solution found so far is less than 5% away from the mean fitness of the population in a specific iteration of the algorithm.

5. RESULTS AND DISCUSSION

5.1. Overall Results and Comparison

Ten runs were conducted. In each run, 40% of the data was used as the training set, 10% as the evaluation set (selecting optimal values for C and gamma parameters using grid search) and 50% as the test set. The mean identification success rate (percentage) for the 10 runs is 91.19%, the mean geometric mean is 90.90%, and the average number of features used is 9.2. The best accuracy achieved in a single run was 94.36% with a geometric mean of 94.33% and nine features used. A comparative table of the best classification accuracy achieved and the number of features used with previously applied techniques for the same data set is shown in Table 4. It is shown that the hybrid SVFI system achieves the best accuracy percentage and also that it uses the less features compared to the other techniques.

5.2. Overall Feature Selection

The stochastic nature of the proposed methodology provided a variety of final solutions which use different feature subsets as inputs. This fact was expected as many of the examined features are highly dependent and share mutual information. The percentages of selection for every feature are shown in Table 5. Despite the stochastic nature of the proposed methodology, some of the features are selected in most executions while others are rarely selected, which indicates the robustness of the proposed methodology and the importance of some specific features in the classification model. The frequent use of the odd harmonics ratios is in agreement with the commonly accepted relationship of their content with surface activity [8, 10, 12, 44–52]. The third to first harmonic ratio is the most frequently selected feature which should be expected since it has been well correlated with the presence of discharges [8–12, 44–52]. It should be noted that the odd harmonic ratios derived from Fourier analysis are more frequently considered compared to the wavelet STD_MRA VECTOR components, which provide a ratio of frequency band contents, and that the THD, HD, and D_R have a relatively low selection percentage. This was to be expected since such ratios give a "wider" view of the picture and although they may be preferred if a single indication is required, they were bound to be left out when fuzzy sets are employed in favor of more precise features as the harmonic ratios.

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Selection percentage for every feature (%)	60	30	30	50	50	50	30	40	50	50	90	70	80	40	10	20	10	50	60	50

TABLE 5. Percentages of selection for every feature

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Result (class)	Aggregated strength
Rule 0						L		L	L	L	L	L	L								C1	1.017
Rule 1						L	H		L	L	H		L								C1	0.308
Rule 2						L		L	L	L		L	L			L					C2	0.567
Rule 3						L			L	L	L	L	L								C1	0.321
Rule 4						L	H		L	L	L	L	L								C1	0.522
Rule 5						L			L	L		L	L								C1	0.864
Rule 6						L		L	L	L		L	L								C2	0.873
Rule 7								L				L	L			L					C2	0.429
Rule 8						L		L	L	L			L								C2	0.799
Rule 9						L		L	L	L	H		L								C2	0.589
Rule 10						L		L	L	L	H	L	L								C2	0.305

TABLE 6. Rules for best run (accuracy 94.36%)

5.3. Best Run

The fuzzy rules for the best run are shown in Table 6. The features used hint some interesting results. First of all, the amplitude is not considered in any rule which is an added indication of the peak value being misleading in regard to the waveform shape [13]. Instead, more robust features resilient to data set outliers such as the interquartile range (IQR), the median absolute deviation (MAD), and the charge are considered in almost every rule. The fifth to third harmonic ratio is considered in every rule. This may hint to the importance of this ratio, which has also been correlated to ageing [52], but the fact that it always has the same value shows that it probably plays a minor role in this classification. A closer look to Rules 0 and 10 hints that the third to first ratio has a more decisive impact, as a change in its value results to a change in the classifier's output, with all other features remaining the same. Further, the frequent use of the harmonic content ratios instead of the STD_MRA VECTOR ratios and the THD, HD, and D_R ratios underlines the above said about such features in fuzzy classification.

6. CONCLUSION

Leakage current monitoring is a commonly employed tool for the investigation of insulators' performance. Several values may be used as an indication of electrical activity, but it is actually the shape of leakage current waveforms that is correlated to the experienced electrical phenomena. However, automating the classification of waveforms' shapes can be a rather complex task, especially in the case of field waveforms. In this article, a hybrid SVFI system is employed for the classification of leakage current waveforms portraying discharges. A number of 387 waveforms recorded on live HV post insulators installed in 150 kV substations is used as a data set. Twenty different features are extracted from each waveform, ten from the time and ten from the frequency domain. The waveforms are classified in two classes based on the duration of discharges. The hybrid system employed uses GAs and SVMs and provides a set of fuzzy logic rules for the classification, offering an insight to the process. Results for overall classification and the run achieving the highest accuracy percentage are shown. Comparisons are made with other classification schemes previously applied on the same data and feature set. Overall feature selection and the feature set and fuzzy rules providing the best accuracy are further investigated. Results show that the considered hybrid system offers the best classification (reaches 94.36%) accuracy compared to previous classification schemes, while using less features, and that it is also able to offer an insight to the classification process.

REFERENCES

[1] CIGRE, "The measurement of site pollution severity and its application to insulator dimensioning for AC systems, CIGRE WG 33-04," *Electra*, Vol. 64, pp. 101–116, 1979.

[2] CIGRE, "A review of current knowledge: Polluted insulators," CIGRE WG 33-04, TF 01, 1998.

[3] IEC, "Selection and dimensioning of high-voltage insulators intended for use in polluted conditions," IEC/TS 60815, 2008.

[4] IEC, "Artificial pollution tests on high-voltage insulators to be used on AC systems," IEC 60507, 1991.

[5] IEC, "Electrical insulating materials used under severe ambient conditions—test methods for evaluating resistance to tracking and erosion," IEC 60587, 2007.

[6] IEC, "Polymeric insulators for indoor and outdoor use with a nominal voltage > 1000 V—general definitions, test methods and acceptance criteria," IEC 62217, 2005.

[7] CIGRE, "Guide for the establishment of naturally polluted insulator testing stations," CIGRE WG B2.03, 2007.

[8] Pylarinos, D., Siderakis, K., and Pyrgioti, E., "Measuring and analyzing leakage current for outdoor insulators and

specimens," *Rev. Advanc. Mater. Sci.*, Vol. 29, No. 1, pp. 31–53, 2011.

[9] Fernando, M. A. R. M., and Gubanski, S. M., "Leakage current patterns on contaminated polymeric surfaces," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 6, No. 5, pp. 688–694, 1999.

[10] Suda, T., "Frequency characteristics of leakage current waveforms of an artificially polluted suspension insulator," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 8, pp. 705–709, 2001.

[11] Li, J., Sima, W., Sun, C., amd Sebo, S. A., "Use of leakage currents of insulators to determine the stage characteristics of the flashover process and contamination level prediction," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 17, No. 2, pp. 490–501, 2010.

[12] Samimi, M. H., Mostajabi, A. H., Ahmadi-Joneidi, I., and Shayegani-Akmala, A. A., "Performance evaluation of insulators using flashover voltage and leakage current," *Elect. Power Compon. Syst.*, Vol. 41, No. 2, pp. 221–233, 2013.

[13] Pylarinos, D., Theofilatos, K., Siderakis, K., Thalassinakis, E., Vitellas, I., Alexandridis, A. T., and Pyrgioti, E., "Investigation and classification of field leakage current waveforms," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 19, No. 6, pp. 2111–2118, 2012.

[14] Pylarinos, D., Siderakis, K., Thalassinakis, E., Pyrgioti, E., and Vitellas, I., "Investigation of leakage current waveforms recorded in a coastal high voltage substation," *Eng. Technol. Appl. Sci. Res.*, Vol. 1, No. 3, pp. 63–69, 2011.

[15] Pylarinos, D., Siderakis, K., Thalassinakis, E., Pyrgioti, E., Vitellas, I., and David, S. L., "Online applicable techniques to evaluate field leakage current waveforms," *Elect. Power Syst. Res.*, Vol. 84, No. 1, pp. 65–71, 2012.

[16] Pylarinos, D., Siderakis, K., Pyrgioti, E., Thalassinakis, E., and Vitellas, I., "Impact of noise related waveforms on long term field leakage current measurements," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 18, No. 1, pp. 122–129, 2011.

[17] Pylarinos, D., Theofilatos, K., Siderakis, K., Pyrgioti, E., Papazoglou, T., Vitellas, I., and Thalassinakis, E. "Classification of field leakage current waveforms using genetic algorithms and an euclidian classifier," *DEMSEE 7th International Workshop on Deregulated Electricity Market Issues in South-Eastern Europe*, Bucharest, Romania, 20–21 September 2012.

[18] Pylarinos, D., Theofilatos, K., Siderakis, K., Pyrgioti, E., Papazoglou, T., Vitellas, I., and Thalassinakis, E., "Feature selection and classification of field leakage current waveforms using genetic algorithms," *CIGRE Lisbon Symposium*, Portugal, 22–24 April 2013.

[19] Thalassinakis, E., Siderakis, K., and Agoris, D. "Experience with new solutions to combat marine pollution in the power system of the Greek islands," *World Conference and Exhibition on Insulators, Arresters and Bushings (INMR 2003)*, Malaga, Spain, 16–19 November 2003.

[20] Siderakis, K., Pylarinos, D., Thalassinakis, E., and Vitellas, I., "High voltage substation pollution maintenance: The use of RTV silicone rubber coatings," *J. Elect. Eng.*, Vol. 11, No. 2, Article 11.2.22, pp. 1–6, 2011.

[21] INMR, "Greek utility battles pollution affecting island transmission system," *Insulator News Market Rep. Issue 78*, Vol. 15, No. 4, p. 24, 2009.

[22] Siderakis, K., Pylarinos, D., Thalassinakis, E., Pyrgioti, E., and Vitellas, I., "Pollution maintenance techniques in coastal high voltage installations," *Eng. Technol. Appl. Sci. Res.*, Vol. 1, No. 1, pp. 1–7, 2011.

[23] TALOS high voltage test station, available at: www.talos-ts.com

[24] INMR, "Greek utility readies to energize new insulator test station," *Insulator News Market Rep. Issue 82*, Vol. 16, No. 4, p. 32, 2008.

[25] Pylarinos, D., Siderakis, K., Thalassinakis, E., Vitellas, I., and Pyrgioti, E. "Recording and managing field leakage current waveforms in Crete. Installation, measurement, software development and signal processing," *ISAP 16th International Conference on Intelligent System Applications to Power Systems*, Hersonissos, Crete, Greece, 25–28 September, 2011.

[26] Pylarinos, D., "A custom-made MATLAB based software to manage leakage current waveforms," *Eng. Technol. Appl. Sci. Res.*, Vol. 1, No. 2, pp. 36–42, 2011.

[27] Izgi, E., Inan, A., and Ay, S., "The analysis and simulation of voltage distribution over string insulators using MATLAB/simulink," *Elec. Power Compon. Syst.*, Vol. 36, No. 2, pp. 109–123, 2008.

[28] Ashouri, M., Mirzaie, M., and Gholami, A., "Calculation of voltage distribution along porcelain suspension insulators based on finite element method," *Elect. Power Compon. Syst.*, Vol. 38, No. 7, pp. 820–831, 2008.

[29] Chen, W., Wang, W., Xia, Q., Luo, B., and Li, L., "Insulator contamination forecasting based on fractal analysis of leakage current," *Energies*, Vol. 5, No. 7, pp. 2594–2607, 2012.

[30] Pratomosiwi, F., and Suwarno, "Performance improvement of the ceramic outdoor insulators located at highly polluted environment using room temperature vulcanized silicone rubber coating," *Int. J. Elect. Eng. Informatics*, Vol. 2, No. 1, pp. 15–28, 2010.

[31] Mallat, S. G., *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[32] Mallat, S. G., "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 11, pp. 674–693, 1989.

[33] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Angus Ng, B., Liu, P. S., Yu, Z. H., Zhou, M., Steinbach, D. J., Hand, and Steinberg, D., "Top 10 algorithms in data mining," *Knowl. Info. Sys.*, Vol. 14, No. 1, pp. 1–37, 2008.

[34] Umamaheswari, R., and Sarathi, R., "Identification of partial discharges in gas-insulated switchgear by ultra-high-frequency technique and classification by adopting multi-class support vector machines," *Elect. Power Compon. Syst.*, Vol. 39, No. 14, pp. 1577–1595, 2011.

[35] Ashkezari, A. D., Ma, H., Saha, T. K., and Ekanayake, C., "Application of fuzzy support vector machine for determining the health index of the insulation system of in-service power transformers," *IEEE Trans. Dielect. Elect. Insulat.*, Vol. 20, No. 3, pp. 965–973, 2013.

[36] Huang, C., and Wang, C., "A GA-based feature selection parameter optimization for support vector machines," *Expert Sys. Appl.*, Vol. 31, No. 2, pp. 231–240, 2006.

[37] Keerthi, S., and Lin, C., "Asymptotic behavior of support vector machines with Gaussian kernel," *Neural Computat.*, Vol. 15, pp. 1667–1689, 2003.

[38] Moewes, C., and Kruse, R., "On the usefulness of fuzzy SVMs and the extraction of fuzzy rules from SVMs," *EUSFLAT Adv. Intell. Syst. Res.*, Vol. 1, No. 1, pp. 943–948, 2011.

[39] Papadimitriou, S., and Terzidis, K., "Efficient and interpretable fuzzy classifiers from data with support vector learning," *Intell. Data Anal.*, Vol. 9, No. 6, pp 527–550, 2005.

[40] Mavroudi, S., Katsanos, P., Papadimitriou, S., and Likothanassis, S., "Transparent classification process of bioinformatics data with an approximated support vector fuzzy inference system," *International Special Topic Conference on Information Technology in Biomedicine (ITAB)*, Ioannina–Epirus, Greece, 26–28 October, 2006.

[41] Papadimitriou, S., Terzidis, K., Mavroudi, S., Lambros, S., and Likothanassis, S., "Efficient and interpretable fuzzy classifiers from data with support vector learning" *Proceedings of 9th World Scientific and Engineering Society (WSEAS) International Circuits, Systems, Communications and Computers (CSCC)*, Vouliagmeni, Athens, Greece, 11–16 July 2005.

[42] Chen, Y., and Wang, J., "Support vector learning for fuzzy rule-based classification," *IEEE Trans. Fuzzy Syst.*, Vol. 11, No. 6, pp. 716–728, 2003.

[43] Holland, J. H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Cambridge: MA: MIT Press, 1995.

[44] Siderakis, K., and Agoris, D., "Performance RTV silicone rubber coatings installed in coastal systems," *Elect. Power Syst. Res.*, Vol. 78, No. 2, pp. 248–254, 2008.

[45] El-Hag, A. H., Jayaram, S. H., and Cherney, E. A., "Fundamental and low frequency harmonic components of leakage current as a diagnostic tool to study aging of RTV and HTV silicone rubber in salt-fog," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 10, pp. 128–136, 2003.

[46] Kordkheili, H. H., Abravesh, H., Tabasi, N., Dakhem, M., and Abravesh, M. M., "Determining the probability of flashover occurrence in composite insulators by using leakage current harmonic components," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 17, No. 2, pp. 502–512, 2010.

[47] Garniwa, I., Sudiarto, B., and Ansorulah, R. S., "Effect of pollutant type and concentration on harmonic characteristic of leakage current on resin epoxy insulator," *2nd Indonesia Japan Joint Scientific Symposium*, Indonesia, 6–8 September 2006.

[48] Du, B. X., and Liu, Y., "Frequency distribution of leakage current on silicone rubber insulators in salt-fog environments," *IEEE Trans. Power. Deliv.*, Vol. 24, No. 3, pp. 1458–1464, 2009.

[49] Jiang, X., Shi, Y., Sun, C., and Zhang, Z., "Evaluating the safety condition of porcelain insulators by the time and frequency characteristics of LC based on artificial pollution tests," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 17, No. 2, pp. 481–489, 2010.

[50] Meghnefi, F., Volat, C., and Farzaneh, M., "Temporal and frequency analysis of the leakage current of a station post insulator during ice accretion," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 14, No. 6, pp. 1381–1389, 2007.

[51] Douar, M. A., Mekhaldi, A., and Bouzidi, M. C., "flashover process and frequency analysis of the leakage current on insulator model under non-uniform pollution conditions," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 17, No. 4, pp. 1284–1297, 2010.

[52] Bashir, N., and Ahmad, H., "Odd harmonics and third to fifth harmonic ratios of leakage currents as diagnostic tools to study the ageing of glass insulators," *IEEE Trans. Dielect. Elect. Insul.*, Vol. 17, No. 3, pp. 819–832, 2010.

BIOGRAPHIES

Konstantinos Theofilatos was born in Patras in 1983. He graduated in 2006 from the Department of Computer Engineering and Informatics of the University of Patras, Greece. In 2009, he received a Master's degree from the same department. Since 2009, he has been a Ph.D. degree candidate in the same department. He is a member of the Pattern Recognition Laboratory (prlab.ceid.upatras.gr) since 2006. His research interests include computational intelligence, machine learning, data mining, bioinformatics, web technologies, time series forecasting and signal processing.

Dionisios Pylarinos was born in Athens, Greece in 1981. He received a Diploma degree in Electrical and Computer Engineering in 2007 and the Ph.D. degree in the same field in 2012 from the University of Patras, Greece. He is a researcher/consultant for Public Power Corporation Greece since 2008 in the field of HV insulator testing and monitoring and he is a part of the team behind Talos High Voltage Test Station. He is a member of the Technical Chamber of Greece. His research interests include outdoor insulation, electrical discharges, leakage current, signal processing and pattern recognition.

Spiros Likothanassis is currently Professor and Director of the Pattern Recognition Laboratory, Department of Computer Engineering and Informatics, University of Patras. His research interests include Intelligent Signal Processing and Adaptive Control, Neural Networks, Genetic Algorithms and applications, Intelligent Agents and applications, Bioinformatics, Web based Applications, Virtual e-Learning Environments, Artificial Intelligence/Expert Systems, Data and Knowledge Mining and Intelligent Tutoring Systems.

Damianos Melidis is a Computer Engineer. He received his diploma for the Department of Computer Engineering and Informatics (University of Patras). His diploma thesis was about "Development of a system of efficient and interpretable classification using Support Vector Machines (SVMs) and genetic algorithms". His general research interests include Algorithms, Data Mining, Biological Data Analysis and Structural Bioinformatics.

Kiriakos Siderakis was born in Iraklion in 1976. He received a Diploma degree in Electrical and Computer Engineering in 2000 and the Ph.D. degree in 2006 from the University of Patras. Presently, he is a Lecturer at the Department of Electrical Engineering, at the Technological Educational Institute of Crete. He is a member of the Greek CIGRE and of the Technical Chamber of Greece. His research interests include outdoor insulation, electrical discharges, high voltage measurements and high voltage equipment diagnostics and reliability.

Emmanuel Thalassinakis received the Diploma in Electrical and Mechanical Engineering and also the Ph.D. degree from the National Technical University of Athens. After working for the Ministry of the Environment, in 1991 he joined the Public Power Corporation (PPC) where he is now Assistant Director of the Islands Network Operations Department.

Seferina Mavroudi is a lecturer in the Department of Social Work, School of Sciences of Health and Care, Technological Educational Institute of Patras, Greece. She graduated in 1998 from the Department of Electrical and Computer Engineering, School of Engineering of the Aristotles University of Thessaloniki. In 2000 she received a Master's degree from the European Postgraduate Program on Biomedical Engineering, organized by the Faculty of Medicine of the University of Patras, the Faculty of Mechanical Engineering and the Faculty of Electrical and Computer Engineering of the National Technical University of Athens. In the same program, in February of the year 2003 she completed her Ph.D. Her research interests include computational intelligence, bioinformatics and scientific computing.