

Machine Learning Methods Applied on Long Term Data Analysis for Rain Detection in a Partial Discharge Sensor Network

Leandro H. S. Silva, Sergio C. Oliveira

Polytechnic School of Pernambuco
University of Pernambuco
Recife - PE, Brazil
{lhss, scampello}@ecomppoli.br

Eduardo Fontana

Department of Electronics and Systems
Federal University of Pernambuco
Recife-PE, Brazil
fontana@ufpe.br

Abstract — Partial discharges (PD) on the surface of high voltage insulators are directly related with the accumulation of pollution. A complete partial discharge sensor network was previously developed and has been in operation for approximately three years. This system records the PD activity, classifying it into four levels. As the PD activity is influenced by the weather conditions, each sensor system in the network also measures the one-hour average temperature and relative humidity. Also a fuzzy inference system was developed to extract the flashover occurrence risk level based on the partial discharge activity recorded. However, a strong rain event can wash insulators almost instantaneously, in turn decreasing the risk level. For a correct interpretation of the results it is important to properly analyze the weather data to detect the rain occurrence. This paper presents a comparison among three machine learning techniques for rain detection from humidity and temperature data, namely, Naïve Bayes Classifier, Support Vector Machines and Multilayer Perceptron Neural Network. These are trained on data gathered by meteorological stations located nearby the PD sensors and used in conjunction with the data obtained by the Sensor Network. Studies on the generalization training power and long term data analysis on sensor data are performed and presented.

Keywords- *Partial discharges; rain detection; pattern recognition; leakage current; insulators; sensor network.*

I. INTRODUCTION

High-voltage transmission lines are affected by many problems. One of them is the pollution accumulated on the surface of the insulators that support the conducting cables in the towers. When combined with high relative humidity the pollution layer becomes a conductive layer. A leakage current flows through this conductive layer causing irregular heating and then humidity evaporation, creating thin dry bands. The increase of electric charges in the boundaries of dry bands causes high electric fields that produce partial discharges (PD) across dry bands [1], [2]. The PD phenomenon can increase in rate and intensity until a complete discharge, known as flashover, bypasses all insulators, in turn causing an outage in power transmission [3].

One way to avoid flashover events is to remove the pollution layer deposited over the insulator string by periodic

washing. However, this is a high-cost operation and failures may occur during the procedure.

Aiming to assist the decision regarding the need for maintenance, a sensor network was previously developed to detect and classify partial discharges according to their frequency of occurrence and intensity [4]. This system comprises an optical sensor coupled to an optical fiber, which transmits the leakage current signal [5] to an electronic processing module, which has also a temperature and a humidity sensor [6]. The collected data are transmitted via satellite and stored in a database.

A fuzzy inference system has been developed in order to extract the flashover risk occurrence. The risk level is incremented and decremented according to the level of partial discharge activity considering its intrinsic relation with relative humidity [7]. The use of a fuzzy system has the advantage of being able to represent uncertainties of natural language.

However, on strong rain events the insulators are washed ceasing the risk of flashover. This almost instantaneous risk variation is not reflected on the fuzzy risk level. This work aims to develop a system capable of detecting the instantaneous cleaning of the insulator by strong rains, based on the available humidity and temperature data. Rain detection would make the fuzzy risk classification system more precise and turn the maintenance schedule more robust, reducing costs due to unnecessary washes.

Common electronic rain sensors are only capable of detecting rain in a small surface and are not capable of quantifying the event [8]. Electromechanical rain sensors are capable of easily detecting and quantifying rain. Nevertheless, when installed in outdoor environments this kind of sensor accumulates water, in turn attracting infestation by wasps or bees. The presence of these insects increases the risk for operators of the power transmission company and increases the failure rate of the rain sensor itself once the hives might block the mechanical parts of the sensor.

Temperature and humidity data gathered by the sensor network exhibits a daily regular pattern. This pattern is changed by rain events and a new rain pattern starts to occur. So, a pattern recognition system can be applied to detect the insulator washing by rain. A pattern recognition prototype system was developed based on the reliable data obtained from the Brazilian Institute of Meteorology (INMET)

database. This database has humidity and temperature information as well as the amount of rain precipitation per hour.

This paper compares three well know machine learning algorithms applied for the task of rain detection: Naive Bayes Classifier, Support Vector Machine (SVM) and Artificial Neural Network Multilayer Perceptron (ANN MLP) [1]. Based on previous results [1], the MLP was applied in a data set gathered by the partial discharge sensor network and visual inspections were carried out to ensure empirically the success of the proposed rain detection strategy.

II. SATELLITE SENSOR SYSTEM NETWORK

The sensor network is composed by six monitoring nodes and it has been in operation for three years in the Northeast region of Brazil. Each node is composed by an optical sensor, an electronic processing module and a satellite transmission modem [4], as illustrated in Fig. 1.

Each hour the sensor node transmits the partial discharges activities, average temperature and average humidity. The partial discharge activity is classified into four current ranges named N1 to N4, which are related to current pulses larger than 5, 10, 20 and 40 mA, respectively [4].

The information gathered by each sensor is organized into two 64-bit packets and transmitted via satellite each half hour. After reception the data are stored in a database. The access to this database is provided by a system called ADECI (a portuguese acronym for Electric Performance Evaluation on Insulator Strings). Only identified employees of CHESF (the generation and distribution company in the Northeast region of Brazil) can access the information.

III. DATA SETS AND RAIN PATTERN

The temperature and humidity have an almost regular daily behavior. During the day, the temperature is high and the humidity is low; at night the temperature falls down and the humidity goes up. During rain events this behavior is modified because the rain causes an immediate increase in humidity and decrease in temperature. This behavior can be seen in Fig. 2 – during rain events, that start to occur beyond the dashed line time point, the temperature falls down and the humidity goes up. This behavior is better observed in heavy rain events.

The INMET meteorological stations data is organized as daily 24 string data containing average temperature and humidity as well as the amount of rain precipitation in millimeters per hour. Linear interpolations were used to complete the series on every data missing less than 5 consecutive hours. When the time period of the missing data was larger than 5 hours, data for the full day were excluded from the database.

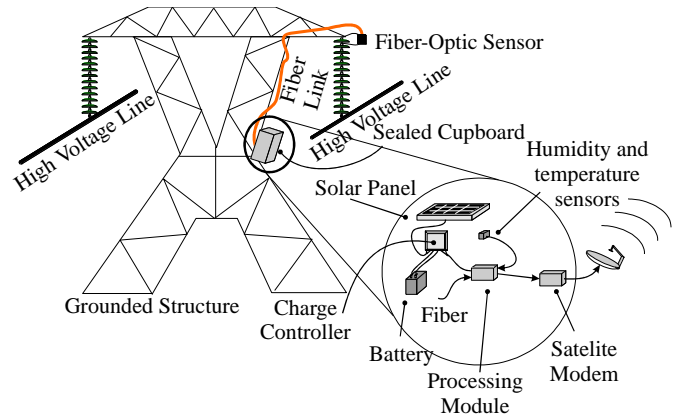


Figure 1. Sensor node for partial discharge monitoring.

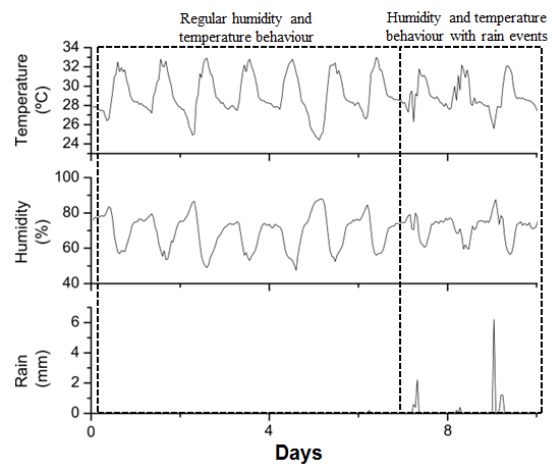


Figure 2. Plots of temperature and humidity patterns and the respective amount of rain.

The INMET database was used to train each detection rain model for further use on the ADECI database. Fig. 3 shows the sensor network topology and each node of the nearest INMET meteorological station. Although each sensor node has a near INMET station, the distance between them is about tens of kilometers and a rain event in the INMET station does not imply a similar occurrence in the nearest sensor location.

The data set was organized on day-long vectors as shown in Table I. Parameters T0 to T23 represent the temperatures uniformly distributed in 24 hours, as well as U0 to U23 represent the corresponding humidity values. If the day has a total rain precipitation larger than 1 mm, the day is classified as rainy. Otherwise it is classified as *no-rain*.

TABLE I. DATA SET ATTRIBUTES AND CLASS.

Attributes					Class	
T0	...	T23	U0	...	U23	[rain / no-rain]

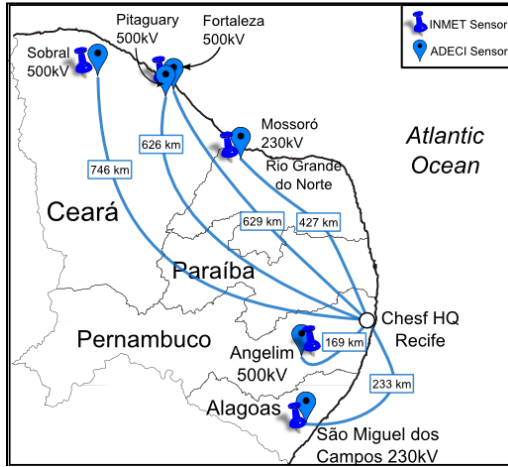


Figure 3. Sensor node and INMET station location.

IV. APPLIED TECHNIQUES

A. Naïve Bayes Classifier

A Naïve Bayes Classifier [9] is a supervised-learning statistical technique. A vector x represents m features (x_1, x_2, \dots, x_m) . In this work, each dimension of vector x comprehends an attribute of the database. The a posteriori probability of having rained in a specified day can be calculated using Bayes theorem as

$$P(\text{rain}|x) = \frac{P(\text{rain})P(x|\text{rain})}{P(x)}. \quad (1)$$

In (1), $P(x)$ is the probability of x occurring in the data set and $P(x|\text{rain})$ is the likelihood probability of x occurring in the *rain* class.

By using the naïve assumption, i.e., the attributes are conditionally independent, the likelihood probably of $P(x|\text{rain})$ is

$$P(x|\text{rain}) = \prod_{i=1}^m P(x_i|\text{rain}). \quad (2)$$

It means that under the naïve assumption, the conditional distribution over the *rain* class can be expressed as

$$P(\text{rain}|x) = \frac{1}{Z} P(\text{rain}) \prod_{i=1}^m P(x_i|\text{rain}), \quad (3)$$

where Z , the evidence, is a scaling factor dependent only on the features of the x vector.

All the Naïve Bayes Classifier parameters (the class prior and feature probability distributions) can be approximated with relative frequencies from the training set. In this work the continuous values associated with each class were considered to have a Gaussian distribution.

B. Multilayer Perceptron Neural Network

The ANN MLP [10] is an artificial neural network whose architecture is based on multiple layers of neurons: an input layer, one or more hidden layers and an output layer. The number of hidden layers can be changed depending on the application.

Each neuron can be seen as an element with inputs, weights, one activation function and the output signal. The output signal of each neuron is given by

$$y_j = f \left(\sum_{i=1}^n x_{ji} w_{ji} \right), \quad (4)$$

where y_j is the output signal of the j -th neuron, x_{ji} is the i -th entry of the j -th neuron, w_{ji} is the i -th weight of the j -th neuron and f is the activation function. In this work the sigmoid function was used as activation function [10]. The signal is propagated from the input layer to the output layer – where the classifier result is available.

The training of an MLP consists on adjusting the weights. The objective is to train the MLP network to achieve a balance between the ability to respond correctly to the input patterns used for training and the ability to provide good results for other similar inputs, i.e., train the network to be capable of performing generalization. For this task, the classic backpropagation algorithm was used to realize the training of the neural network [10].

C. Support Vector Machine

The SVM [11] is a statistically robust learning method in which the training process consists of finding an optimal hyperplane which maximizes the margin between two classes of data in the kernel induced feature space.

Given an input data of n samples x_i ($i = 1, \dots, n$) classified into two classes, each one of the classes associated with labels are $y_i = +1$ for the positive class (*rain*) and $y_i = -1$ for the negative class (*no-rain*), respectively. For linear data, it is possible to determine the hyperplane

$$f(x) = xw + b = 0, \quad (5)$$

where w is an M -dimensional vector and b is a scalar. This separating hyperplane should satisfy the constraints

$$\begin{aligned} x_i w + b &\geq 1, \text{ if } y_i = +1 \\ x_i w + b &\leq -1, \text{ if } y_i = -1 \end{aligned} \quad (6)$$

Furthermore, as the SVM searches for an optimal hyperplane, the margin width between the support vectors and the optimum hyperplane must be maximized, as shown in Fig. 4. The margin is calculated as

$$2 \cdot d = \frac{2}{\|w\|}, \quad (7)$$

so $\|w\|$ must be minimized.

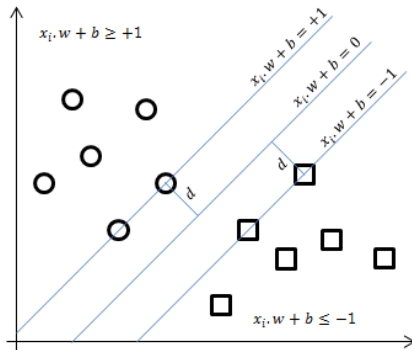


Figure 4. Support Vectors and separating hyperplane.

There is also the introduction of positive slack variables ξ_i , to measure the distance between the margin and the vectors x_i , which means that some mistakes can be tolerated. The optimal hyperplane separating the data can be obtained by solving the optimization problem

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \xi_i, \quad (8)$$

subject to

$$y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0 \quad (9)$$

The constraints aim to put the instances with positive label at one side of the margin of the hyperplane, and the ones with negative labels at the other side. Parameter C represents the cost, which is a positive constant specified by the user.

The optimization problem of the SVM is usually solved by introducing the lagrangian multipliers α_i , transforming the problem on the dual quadratic optimization.

The SVM method can also be used to classify nonlinear problems. By using a nonlinear mapping function, called Kernel function, the original data are mapped into a high-dimensional feature space, where the linear classification is possible. There are different Kernel functions used in SVMs, such as linear, polynomial, sigmoidal and Gaussian RBF. The selection of the better Kernel function is very important, since this function will define the feature space in which the training set examples will be classified [11].

V. METHODOLOGY

A. Experiments to setup parameters

At first, some experimental arrangements were made in order to evaluate the best setup parameter for the ANN MLP and for the SVM approaches.

For the ANN MLP, the number of hidden layers was limited in two. The tested topologies are shown in Table II. There are two MLP output neurons, one indicates the *rain* class and the other indicates the *no-rain* class. The validation set, necessary to avoid overfit was generated by selecting randomly 30% of the normalized complete data set.

TABLE II. EXPERIMENTAL ARRANGEMENT FOR MLP.

Neuron quantity	
First hidden layer	Second hidden layer
10	0
20	0
30	0
40	0
5	5
10	10
20	20
30	30

For the SVM, four kernel functions were tested: radial basis, linear, sigmoid and polynomial. For each kernel function the C parameter assumed, respectively, the values 1, 5, 10 and 30. The ϵ parameter was fixed at 0.001. And for the Naïve Bayes Classifier a Gaussian distribution function was assumed.

The test method for all experiments was the stratified cross-validation 5-fold. For the MLP the experiment was repeated twenty times. The Coruripe INMET database (near São Miguel dos Campos in the map of Fig. 3) was used to evaluate the best setup parameter for each technique.

The metrics used to compare the three techniques are the TP (True Positive) rate and the F-Measure. The F-Measure is an accuracy evaluation that considers the precision generating an overall score about the classifier. For this application, the TP of *no-rain* class is a very important measure, and this rate must be maximized. A false positive for the *rain* class will cause a decrease of the risk level of a flashover and the prediction system can miss the flashover event because of this false positive rain detection.

B. Experiments to evaluate the training applied to other data bases

With the best setup parameters, all three techniques were trained with the data from Coruripe INMET station and the trained models were applied in all others INMET stations.

The main objective was to evaluate if a training performed on one station could be applied to another one. The geographic limits of the training and the influence of the climate were also investigated.

C. Results on ADECI data

The trained models were applied on ADECI databases aiming to verify if the rain detection was satisfactorily. The analysis of these experiments could not be measured quantitatively because the ADECI data does not include the rain information. Instead careful visual inspections were made to identify the temperature and humidity behavior changes in order to qualitatively verify the results obtained. Those visual inspections will be better described in section VI-C.

VI. RESULTS

A. Evaluation of setup parameters

Table III presents the results for the Naïve Bayes Classifier. There are no parameters to adjust on this

classifier. The numbers in the table indicate that the Naïve Bayes Classifier achieves TP rates over 0.5 for both classes. However, the FP (false positive) rate of the *no-rain* class is still high for the application (the FP for the *no-rain* class is 0.227). The high result of FP *no-rain* is a bad issue as it can lead to unnecessary maintenance action for insulators wash.

Table IV presents the results for all ANN MLP topologies experimented.

TABLE III. EXPERIMENTAL ARRANGEMENT FOR NAÏVE BAYES CLASSIFIER.

TP rate <i>rain</i>	TP rate <i>no-rain</i>	F-Measure <i>rain</i> class
0.807	0.798	0.746

TABLE IV. RESULTS FOR ANN MLP.

Topology (as in Table II)	TP rate <i>rain</i> class	TP rate <i>no-rain</i> class	F-Measure <i>rain</i> class
10, 0	0.802 (0.047)	0.873 (0.020)	0.790 (0.016)
20, 0	0.793 (0.049)	0.877 (0.022)	0.788 (0.016)
30, 0	0.784 (0.049)	0.878 (0.021)	0.783 (0.016)
40, 0	0.784 (0.049)	0.878 (0.021)	0.783 (0.016)
5, 5	0.810 (0.050)	0.866 (0.025)	0.791 (0.016)
10, 10	0.810 (0.051)	0.869 (0.024)	0.792 (0.017)
20, 20	0.814 (0.049)	0.867 (0.022)	0.793 (0.016)
30,30	0.812 (0.051)	0.867 (0.022)	0.793 (0.018)

In order to choose the best topology for the ANN MLP, statistical tests were made. With the Shapiro Wilk test [12] all samples follow the normal distribution, and with the F test, all samples have the same variance. Complying with these assumptions, the T-Student test was applied to evaluate the best topology with statistical significance. The result of the T-Student test proves that there is no statistical difference between the topologies. So, the topology with fewer neurons in one layer was chosen. As shown in the highlighted cells in Table IV the results of the ANN MLP were better than those of the Naïve Bayes Classifier.

Table V presents the results for the SVM. In this table, only the best results for each kernel function are presented.

As the SVM classifier presents a unique solution, the set of parameters that resulted on the highest F-Measure was chosen (Radial Basis kernel function and *C* equals 10.0).

The results obtained with training and execution of the classifiers within the same database show that the rain pattern recognition is possible.

B. Training Generalization

A complete investigation about the best data training set is presented in Figs. 5 to 9. In these figures each INMET station was used as the data training set and the trained classifiers were tested on all stations, including that selected for training.

Fig. 5 exhibits the results obtained using Sobral INMET station as the training data set. All three classifiers were able to achieve good results of the True Positive *no-rain* patterns but the True Positive rates for the *rain* pattern were too low and unacceptable. It might be due to the few rain patterns on the INMET Sobral database. Instead, the small generalization power of Sobral station, all three classifiers present acceptable results when evaluated on the Sobral database itself.

TABLE V. RESULTS FOR SVM.

Kernel Function	C	TP rate <i>rain</i>	TP rate <i>no-rain</i>	F-Measure <i>rain</i>
Linear	1	0.758	0.896	0.781
	5	0.754	0.880	0.767
	10	0.754	0.880	0.767
Polynomial (3 degree)	1	0.256	0.973	0.393
	5	0.575	0.929	0.676
	10	0.643	0.916	0.717
Radial Basis	1	0.720	0.910	0.766
	5	0.749	0.889	0.777
	10	0.758	0.902	0.785
Sigmoidal	1	0.671	0.921	0.741
	5	0.744	0.905	0.778
	10	0.754	0.905	0.784

Fig. 6 exhibits the results obtained using Fortaleza INMET station as the training data set. Likewise as the training with the data of Sobral, this training with Fortaleza data exhibits good results for the *no-rain* class for all three classifiers on all locations, but Sobral. For the *rain* class, acceptable results were achieved on all databases and classifiers, except for Mossoró. The SVM and MLP classifiers produced better results for both *rain* and *no-rain* classes.

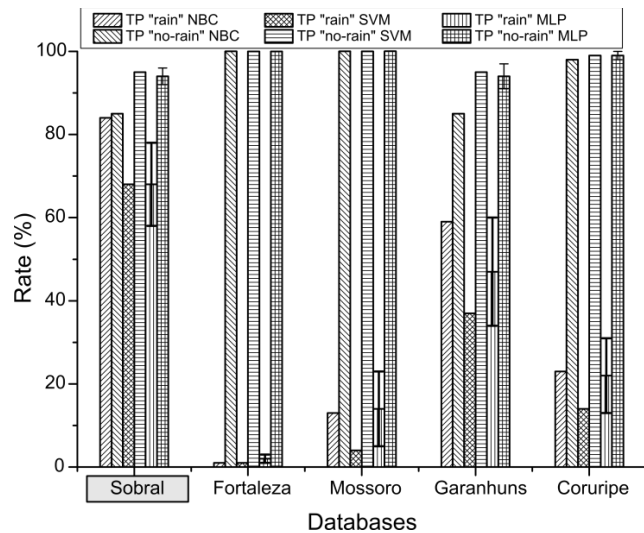


Figure 5. True positive rates for all nodes with Sobral data set training.

Using the Mossoró INMET station as the training database, only acceptable results for the *no-rain* class were obtained. Results for the Mossoró station itself, presented in Fig. 7, are worse than those of Fig. 6, obtained by training with the Fortaleza database.

For the Garanhuns and Coruripe INMET stations the best results obtained are with the respective databases. Figs. 8 and 9 clearly show that all three classifiers yielded excellent results for both the *rain* and *no-rain* classes when the trained classifiers were applied to the corresponding database.

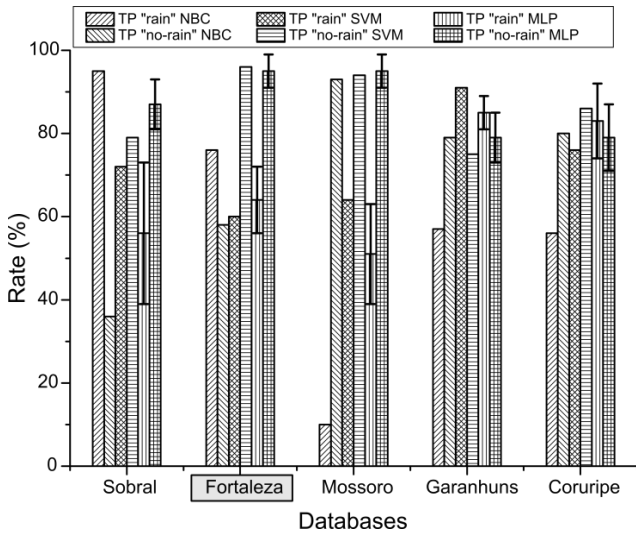


Figure 6. True positive rates for all nodes with Fortaleza data set training.

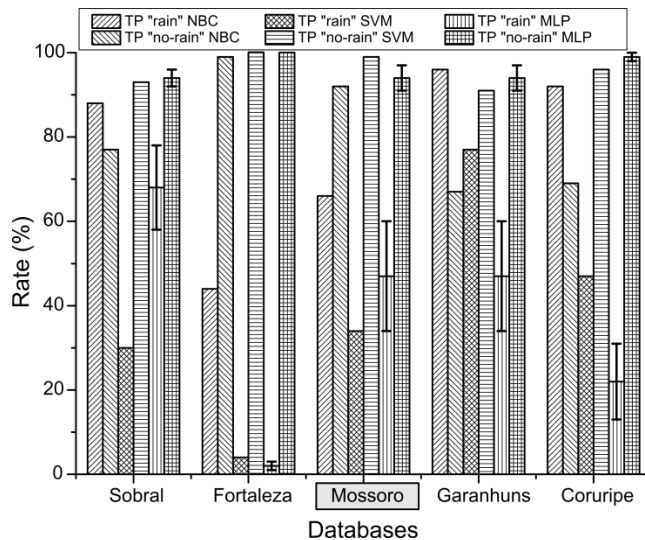


Figure 7. True positive rates for all nodes with Mossoró data set training.

Comparing all databases, the best option to predict the rain event in a given node of the sensor network is to use the database of its nearest INMET station.

Given that only the SVM and MLP classifiers presented good results, both of them could be chosen for the following analysis.

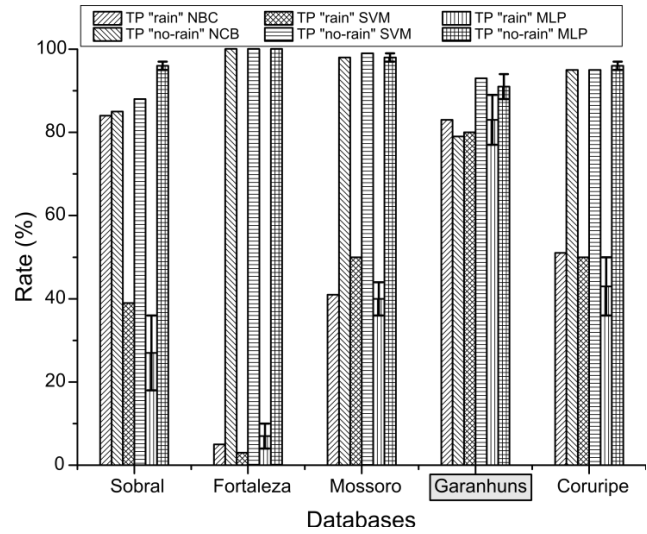


Figure 8. True positive rates for all nodes with Garanhuns data set training.

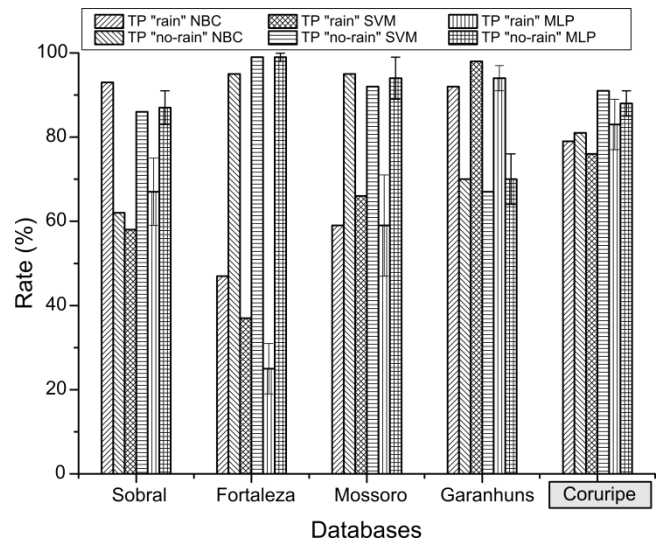


Figure 9. True positive rates for all nodes with Coruripe data set training.

C. Preliminary Evaluation on the ADECI data

The ADECI data does not include the information about the amount of rain, so, a visual analysis was made in order to verify results. In this preliminary investigation, the ANN MLP classifier was tested on the ADECI databases.

Fig. 10 shows the result of the ANN MLP classifier trained with the data from Coruripe INMET station and applied to the São Miguel dos Campos sensor system. The result of the ANN MLP is a binary neuron indicating class *rain* (one) and *no-rain* (zero). As can be seen in Fig. 10, the rain pattern was successfully recognized in some data subsets. The visual analysis of the rain pattern matches with the previous patterns in Fig. 2. Fig. 11 shows the result of the same ANN MLP applied in the Mossoró sensor system.

There are some possible rain events not successfully recognized. These events are marked in Figs. 10 and 11.

However, for every rain detected, the visual analysis of temperature and humidity suggests a rain event.

If a rain is not properly detected, as indicated in Figs. 10 and 11, the risk level will not be reset. If the risk level before the rain event was high enough to require schedule maintenance, this maintenance will happen, even with the insulator rain wash, causing an unnecessary spending by the electric company. But some rain events were detected, and in these cases the maintenance schedule could be reprogrammed with this new information. It is not possible to quantify the rain detection efficiency but visually it is possible to verify that 6 out of the 9 rain events in Fig. 10 were properly detected.

Furthermore, during a rain event, naturally there is an increase in the activity rate, mainly in the N1 range as can be seen in the last rain event, marked in Fig. 10. This activity increase causes an increment in the risk level leading to wrong interpretations. With the proper rain detection strategy, the activity increase can be related to the rain event and the risk level is not increased.

Regarding the rain events that were not recognized in the data of Figs.10 and 11, the general visual analysis suggests that the false negative rain rate was higher for the Mossoró records relative to those of São Miguel dos Campos. The efficiency decrease observed in the Mossoró station suggests that it decreases with distance, indicating that one single model cannot be used to analyze all network nodes.

D. Long term ADECI data Analysis

After the training generalization described in Section VI-C, the classifiers were applied on long term ADECI data from Mossoró and São Miguel dos Campos. For the Mossoró ADECI data analysis, the ANN MLP classifiers trained on Mossoró and Coruripe INMET stations were used.

Fig. 12 shows only seven rain patterns recognized when the MLP trained on INMET Mossoró was applied on the Mossoró ADECI data. Fig. 13 presents the results obtained by the MLP trained on Coruripe INMET and applied on the ADECI data of Mossoró. The first 20 days are typically not rainy days and the pollution deposition was registered as activities on ranges N1 to N4. A small electric activity variation, which could be explained by a rain event, was observed. Only the MLP trained on Coruripe station was able to detect this possible event. Between days 25 and 275 there is no significant electric activity registered on ranges N1 to N4. This is because many of the rain events, shown in Fig. 13, clean the insulators.

In the last 100 days of the experiment, strong rain events are detected by both MLP results of Figs. 12 and 13. It is easy to reinforce the results shown in Figs. 7 and 9 that the best station (at this moment) for detecting rain events occurring at the Mossoró node is the Coruripe INMET station. The Mossoró INMET station data has 574 *no-rain* examples against only 70 *rain* examples; while the Coruripe INMET station data has 422 *no-rain* examples and 211 *rain* examples. In other words, the Coruripe INMET station is a more balanced database.

The severe weather conditions near Mossoró with high temperatures and low relative humidity damaged the

humidity sensors. Even after its replacement on day 240, approximately six month later the sensor was damaged again. Even with the malfunction of the humidity sensors the ANN MLP classifier was able to detect the rain pattern recorded by the temperature sensor in Mossoró. It means that the daily pattern recorded by the temperature sensor alone contains enough information to allow inferring the occurrence of rain events. For further experiments, a feature selection can be used to reduce the amount of data presented to the classifier.

Loss of data transmission were expected below 2% as specified by the satellite link providers [4]. But sequential losses were observed for periods longer than a few days. The causes of these large losses remain unknown, but strong rain events can jam the satellite transmitting signals producing these outages. In spite of these losses, future sensor system firmware updates will reduce these effects with the incorporation of data delivery checks and requests for package retransmissions.

Both ANN MLP classifiers trained on Mossoró and Coruripe INMET stations were applied on the long term ADECI data of São Miguel dos Campos. Results are presented in Figs. 14 and 15 for a period almost two years long. After the initial humidity sensor calibration on the first 70 days the sensor behaves very well for almost 600 days. After this period of continuous work the humidity sensor appears to exhibit malfunction yielding unreliable records.

Again, periods of no data transmission were observed, some of which occurring simultaneously with the periods of data loss verified in the Mossoró sensor system. Given that the distance between the two stations is approximately 600 km, and the cities have significant differences in climate, these transmission losses probably occurred due to an overall outage of the satellite link.

During the first 50 days of experiment, few rain events were observed, as indicated in Figs. 14 and 15. This period corresponds to the dry season and the absence of rainy days is reinforced by the activities registered by the sensor network on ranges N1 to N4.

The period between days 50 and 350 corresponds to the rainy season and a large number of rain events were observed, as indicated in the plots of Figs. 14 and 15. Again the absence of PD activities during this period reinforces the presence of many rainy events as correctly detected by the ANN MLP classifier.

Approximately one year after the sensor installation, a new dry season initiated near day 350. The electric activity on the polluted insulators increases again (ranges N1 to N4) and few rain events were detected. The classifier trained on Mossoró INMET station detected only two rain events between days 400 and 450, against four rain events detected by the classifier trained on Coruripe INMET station. The two last rain events, observed only on Fig. 15, are the most important difference of Figs. 14 and 15. These two rains match with the activity decrease, meaning that these rains washed the insulators naturally preventing partial discharges to grow to dangerous values.

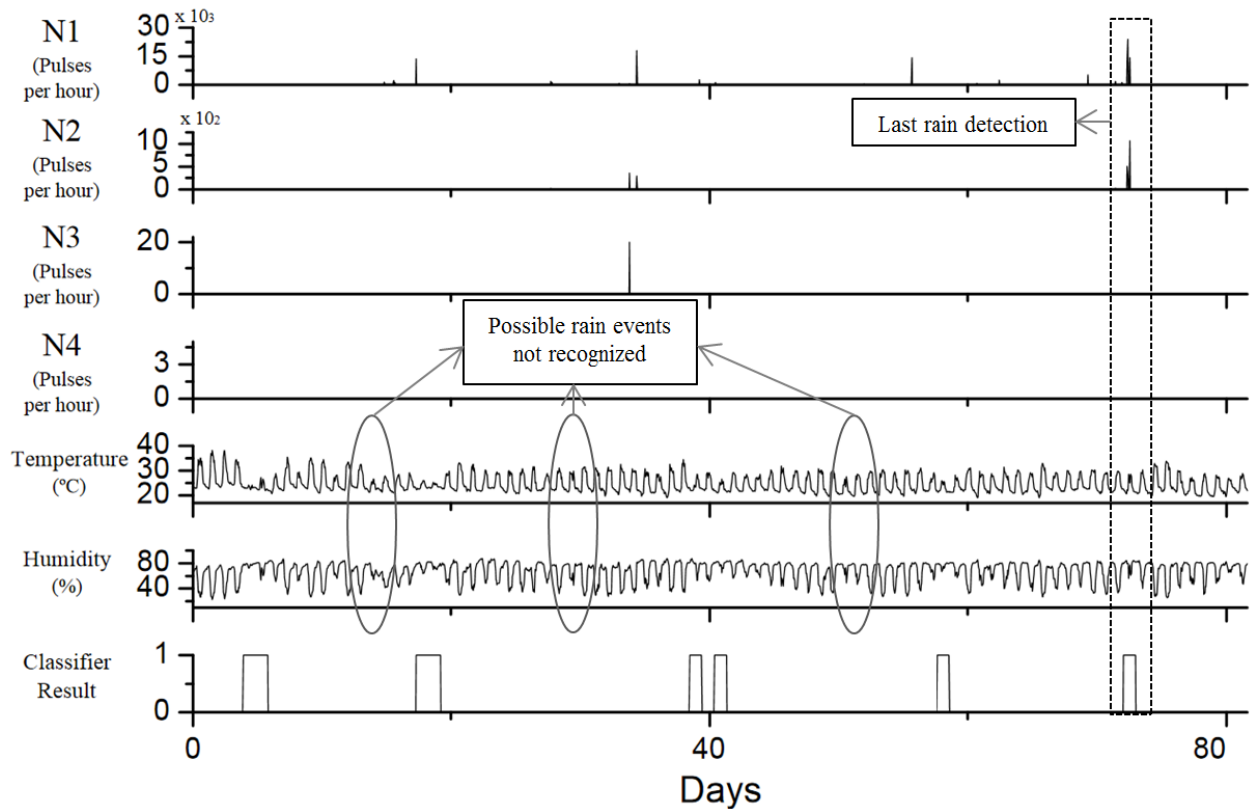


Figure 10. ADECI registers from *São Miguel dos Campos* sensor system and rain detection by ANN MLP classifier for 80 days.

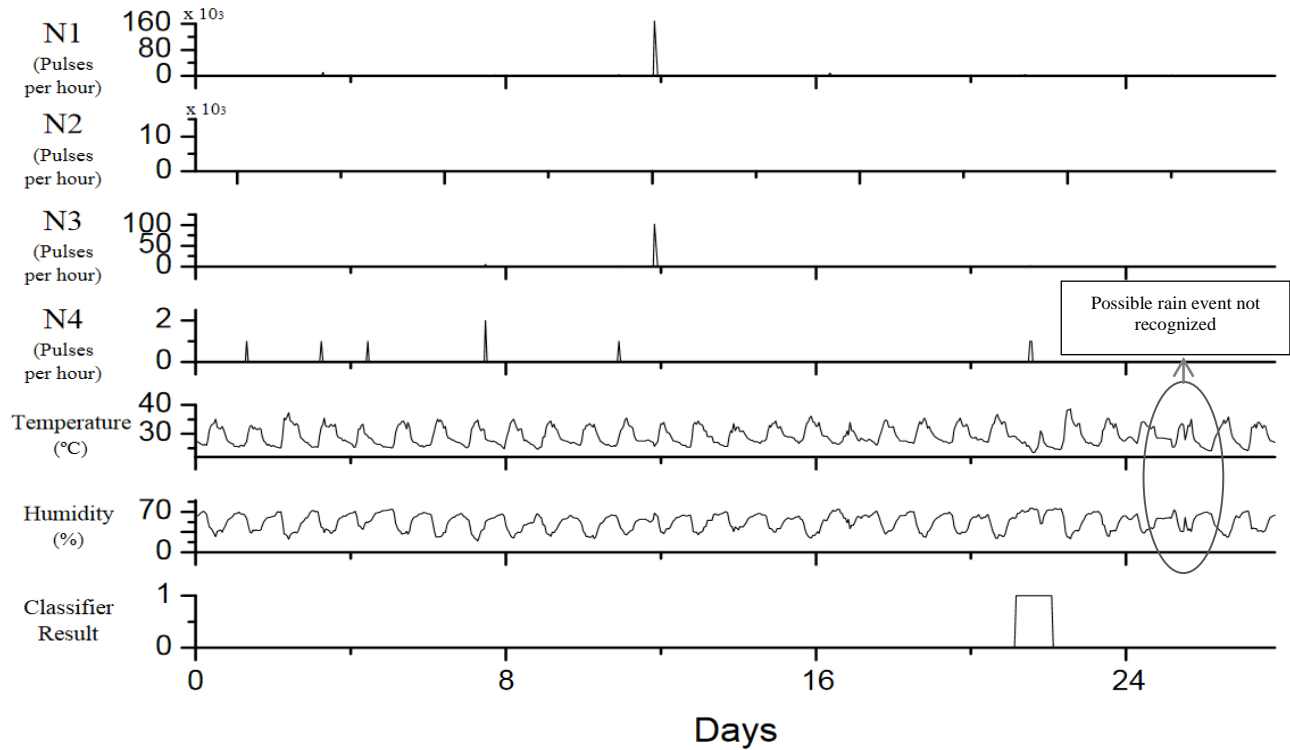


Figure 11. ADECI registers from *Mossoró* sensor system and rain detection by ANN MLP classifier for 28 days.

Comparing Figs. 14 and 15, it is possible to infer that the data of Fig.15 exhibit a better correlation between the detected rain events and the activities recorded on ranges N1 to N4. This visual analysis on the ADECI data of São Miguel dos Campos confirms the more efficient training data set of Coruripe INMET station as expected by results of Figs. 7 and 9.

These initial analyses on the ADECI data from São Miguel dos Campos and Mossoró visually shows that a balanced database for training the MLP classifier is more important that the climate differences. Due to that, some algorithms to balance the input data for the classifiers must be used in future works.

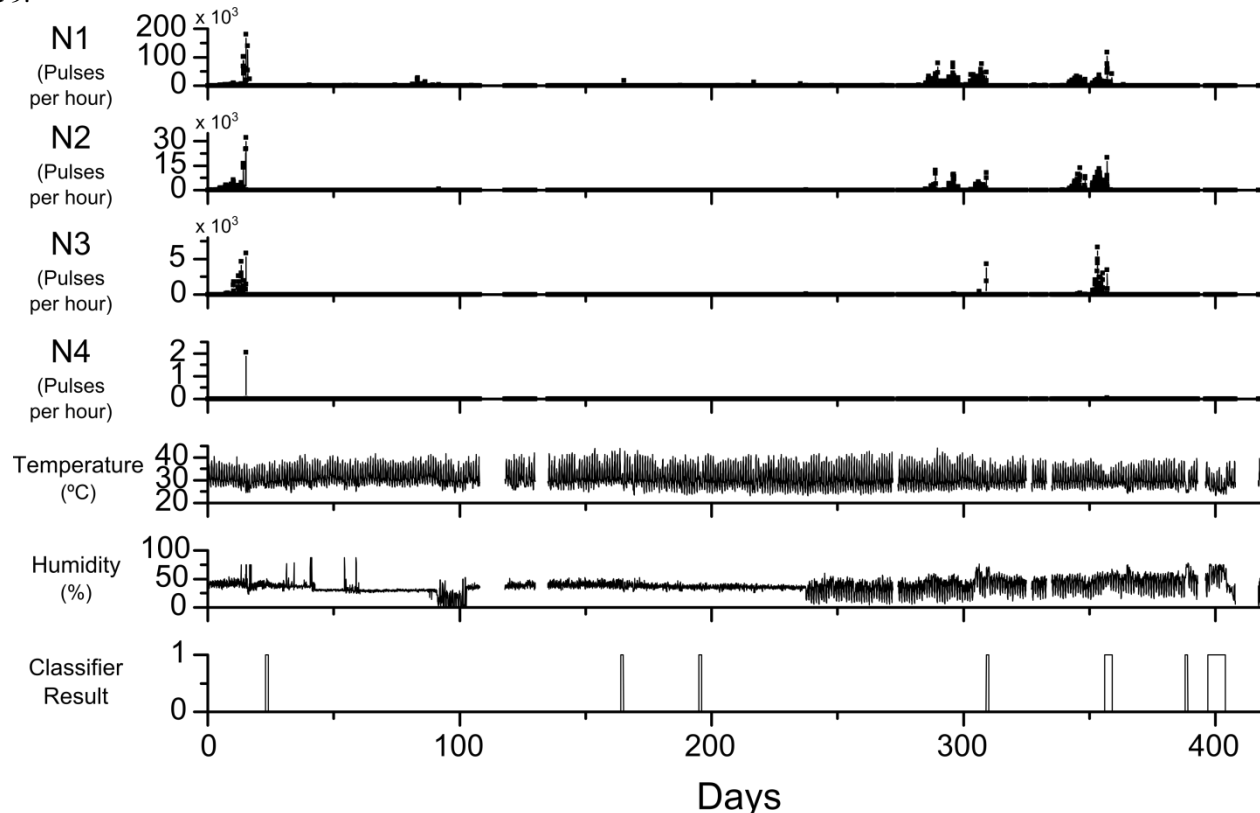


Figure 12. Application of ANN MLP classifier trained and applied on *Mossoró* INMET and ADECI data, respectively.

VII. CONCLUSION AND FUTURE WORK

This work presented a successful attempt to detect rain from the analysis of relative humidity and temperature data obtained from Chesf's sensor network. Results show that it is possible to detect rain events and use them to improve the flashover risk classification.

The initial tests were performed on the reliable data from INMET meteorological stations in the Northeast Region of Brazil. Three techniques employed, namely, Naïve Bayes Classifier, ANN MLP and SVM, presented acceptable results when tested on data from the same database. However, when the classifiers were trained with data from one station and applied to a distinct station, only the SVM and ANN MLP classifiers presented acceptable results. Given the different climates between the station sites, the generalization ability of the classifier is an important feature. Since SVM and ANN MLP presented similar results, only the ANN MLP was used as the classifier to be used with de ADECI data.

In the initial tests the ANN MLP trained with the São Miguel dos Campos INMET station was applied in data sets from the ADECI database. Data gathered by ADECI from two sensor locations in the network stations were used to evaluate the ANN MLP. The rain pattern was successfully recognized in this database, with some false negatives. Long term studies were performed on the ADECI data recorded by the Mossoró and São Miguel dos Campos sensor systems. The two closest INMET stations chosen for direct and cross tests confirmed that the best option is to perform training on a database from a nearby INMET station.

The results of this work can help improve the maintenance schedule system of the electric utility company. Without a rain detection attribute, when a rain event occurs, the sudden PD activity increase can give a false indication of risk of flashover. With the addition of the rain detection attribute, this effect will not be taken into account and after the rain event, the predicted risk of flashover can be reset because the insulators were washed.

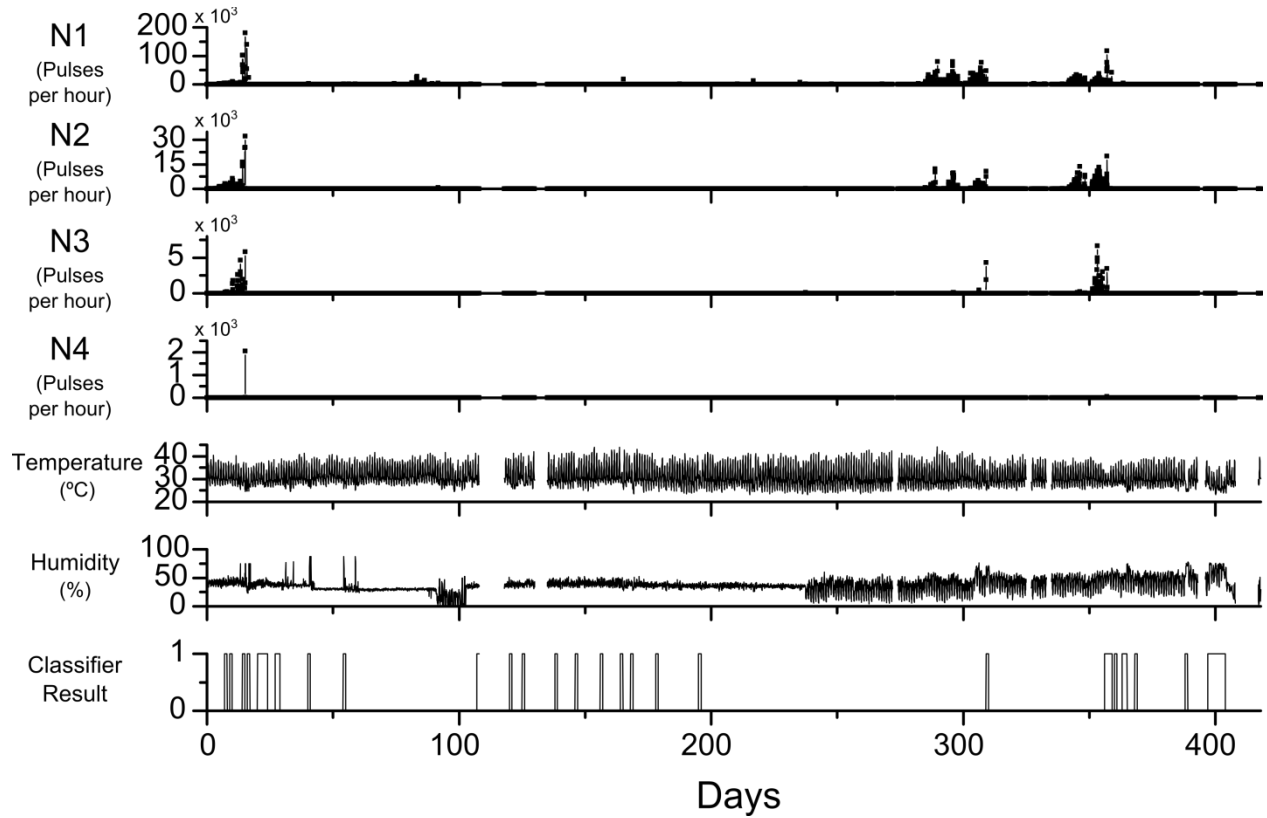


Figure 13. Application of ANN MLP classifier trained on *Coruripe* INMET and applied on *Mossoró* ADECI data.

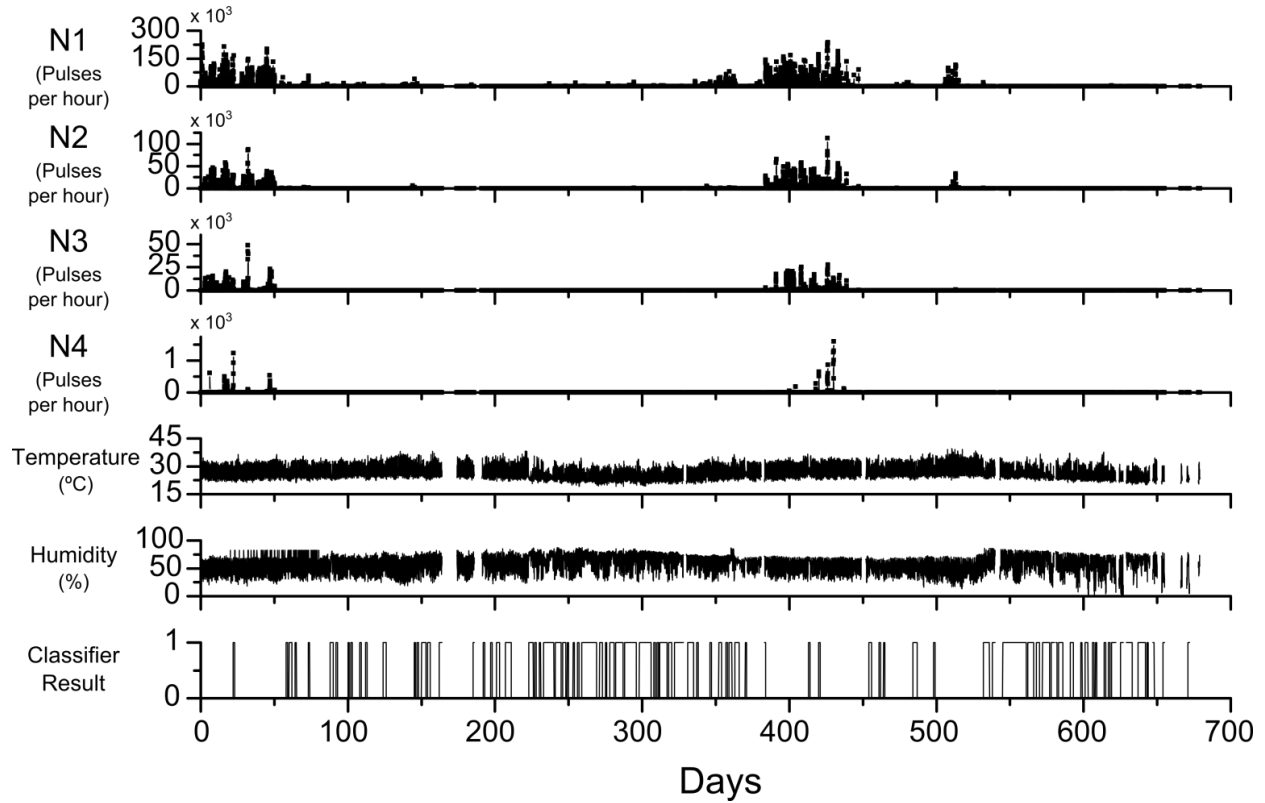


Figure 14. ANN MLP classifier trained on *Mossoró* INMET and applied on *São Miguel dos Campos* ADECI data.

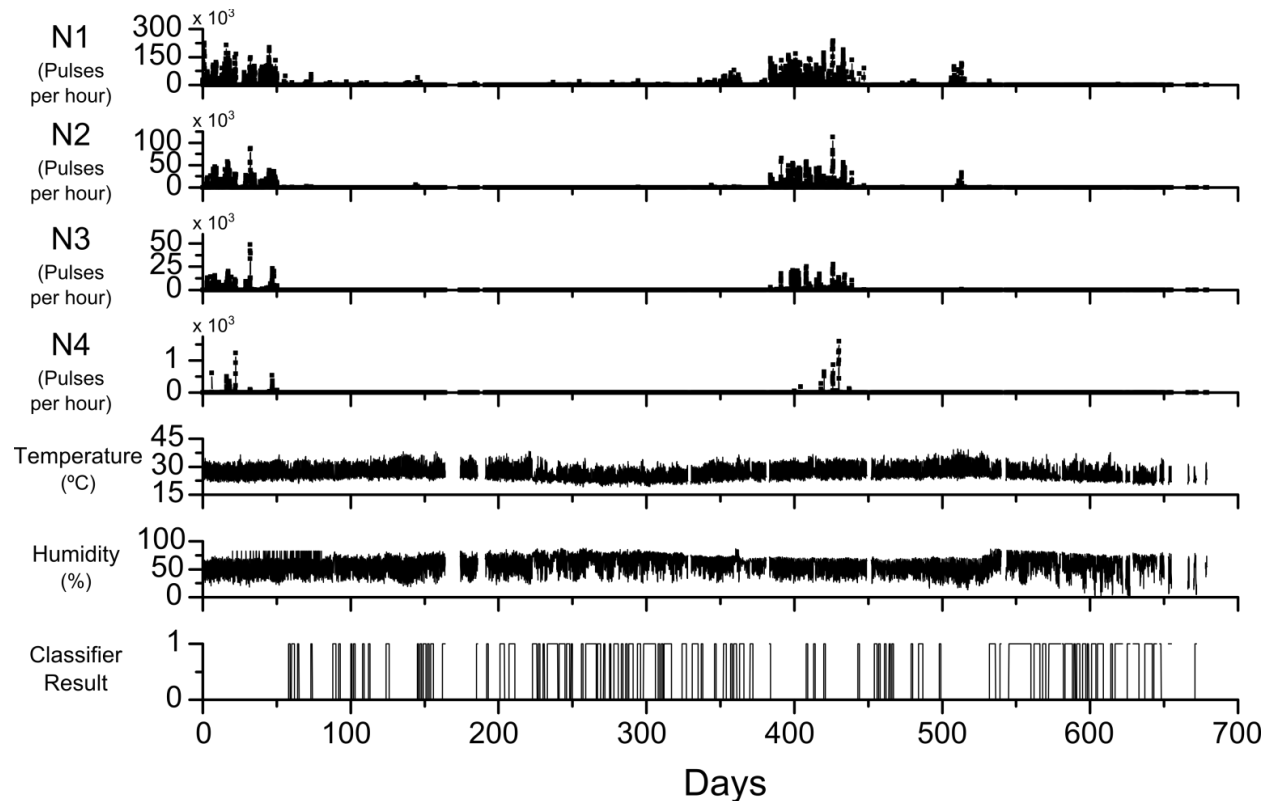


Figure 15. ANN MLP classifier trained on *Coruripe* INMET and applied on *São Miguel dos Campos* ADECI data.

Future work on the use of machine learning methods for rain detection will be directed to the evaluation of the threshold for rain precipitation to allow classifying or not a given day as rainy. Data sets from different locations will also be used in order to test the influence of climate characteristics on the proposed approach for rain detection and define the geographical boundary within which the same model can be applied.

Another improvement on the flashover risk prediction system is to use more than one classifier in order to prevent false positive and negative results.

ACKNOWLEDGMENT

The authors thank CHESF.

REFERENCES

- [1] L. H. S. Silva, S. C. Oliveira, and E. Fontana, "Evaluation of Machine Learning Methods in a Rain Detection System for Partial Discharge Data Analysis," *INTELLI 2013: The Second International Conference on Intelligent Systems and Applications*, pp. 176–183, 2013.
- [2] M. G. Danikas, "The definitions used for partial discharge phenomena," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 1075–1081, 1993.
- [3] E. O. Abdelaziz, M. Javoronkov, C. Abdeliziz, G. Fethi, and B. Zohra, "Prevention of the interruptions due to the phenomena of the electric insulators pollution," *Control, Communication and Signal Processing, 2004. First International Symposium on*, pp. 493–497, 2004.
- [4] E. Fontana, J. F. Martins-Filho, S. C. Oliveira, F. J. M. M. Cavalcanti, R. A. Lima, G. O. Cavalcanti, T. L. Prata, and R. B. Lima, "Sensor Network for Monitoring the State of Pollution of High-Voltage Insulators Via Satellite," *IEEE Transactions on Power Delivery*, vol. 27, no. 2, pp. 953–962, Apr. 2012.
- [5] E. Thalassinakis and C. G. Karagiannopoulos, "Measurements and interpretations concerning leakage currents on polluted high voltage insulators," vol. 421, pp. 421–426, 2003.
- [6] E. Fontana, S. Oliveira, F. J. M. M. Cavalcanti, R. Lima, J. f. Martins-Filho, and E. Meneses-Pacheco, "Novel Sensor System for Leakage Current Detection on Insulator Strings of Overhead Transmission Lines," in *2006 IEEE PES Power Systems Conference and Exposition*, 2006, vol. 21, no. 4, pp. 2255–2255.
- [7] H. O. de Lima, S. C. Oliveira, and E. Fontana, "Flashover risk prediction on polluted insulators strings of high voltage transmission lines," in *2011 11th International Conference on Intelligent Systems Design and Applications*, 2011, pp. 397–401.
- [8] K. N. Choi, "Omni-directional rain sensor utilizing scattered light reflection by water particle on automotive windshield glass," in *2011 IEEE SENSORS Proceedings*, 2011, pp. 1728–1731.
- [9] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley-Interscience; 2 edition (October 2000), 2000, p. 654.
- [10] S. Haykin, *Neural Networks: A Comprehensive Foundation*. PTR Upper Saddle River, NJ, USA; Prentice Hall, 1994.
- [11] A. Christmann and I. Steinwart, *Support Vector Machines*, Springer. New York, NY: Springer New York, 2008.
- [12] G. K. Kanji, *100 Statistical Tests*, 3rd editio. London: SAGE Publications Ltd, 2006, p. 256.