Detection and Localization of Electric Vehicles in Low Voltage Network

Manuel P. S. Nunes [a], João F.P. Fernandes [b], Paulo J. C. Branco [b]

[a] Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[b] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

Abstract: In the next few years, an increase of electric vehicles (EVs) on the road is expected. This increase comes with several challenges in terms of the charging of EVs and there must be a straight cooperation between the automotive and the energy industries. To allow for a large-scale EV integration into the power grid, Distribution System Operators must ensure that the demand in the areas where there is a large proliferation of EVs is met. This thesis proposes a classification algorithm to detect, in real-time, EV charging events using ENEIDA® EWS DTVI-g smart sensor installed in the secondary of the distribution substation. By sampling the active (P) and reactive (Q) power values and crossing this information with the one from the EV charging station network - MOBI.E® - some pattern recognition was conducted, using linear regressions and Gaussian membership functions, resulting in the development of a supervised machine learning algorithm to detect EV charging events. This algorithm takes into consideration the P and Q relationship as well as the EV charging duration. The feasibility limits of the proposed classifier were tested in terms of the load magnitude of the EV charger line as well as the charging power mode. In addition, a statistical study on the behaviour of the EV charger users was conducted, using Gaussian Mixture Model, to obtain a realistic model that allows the creation of an EV charging profile.

Keywords — **Electric vehicles, Detection of charging, Machine learning algorithm, Low Voltage network**

I. INTRODUCTION

According to the *International Energy Agency*, in 2018 the global electric vehicle (EV) fleet exceeded 5.1 million units, almost doubling the number of new EV registrations from the previous year. The total number of light-duty vehicle chargers amounted to 5.2 million, 540 000 of which are publicly accessible and 157 000 are fast chargers for buses. For the year 2030, a global stock of 140 million EVs is expected [1].

With the objective of integrating this growing number of EVs into the power grid, there are several challenges for the Distribution System Operator (DSO) namely the identification of areas where there is proliferation of EVs, real-time detection of EV charging events and forecast their demand in the short, medium and long term [1]. One of the objectives proposed by this paper is the creation of a classification algorithm to detect EV charging events using Low Voltage (LV) network parameters. In this manner, using ENEIDA® EWS DTVI-g smart sensor placed in the Secondary of the Distribution Substation (SDS), the LV network parameters were sampled, namely the active (P) and reactive (Q) power. By crossing this data with the one from MOBI.E®, responsible for managing the Portuguese EV charger network, pattern recognition algorithms were used, namely linear regressions together with Gaussian functions, creating a supervised machine learning algorithm to detect EV charging events [2], [3] . In addition, the modelling of the users charging behaviour of the used EV charger is accomplished so that it is possible to realistically create EV charging profiles, using Gaussian Mixture Models (GMM) [4].

As previously introduced, ENEIDA® EWS DTVI-g smart sensor was used to extract LV network data. By measuring, at maximum, 6 SDS feeders and 3 lines per feeder in each 5-minutes window, 4 voltage (v) and current (i) measurements are made, of 1 second each at a frequency $f=4096Hz$, spaced equidistantly in time. For each set of 4 measurements of the parameters, the average is calculated and sent to the server, with U_{max}, U_{min}, I_{max} and I_{min} obtained as well. All physical phenomenon whose time duration occurs at a scale smaller than $75\ s$ ($= 5min/4$) is out of reach. For this study, the distribution station used is in Avenida Calouste Gulbenkian, Coimbra (coordinates 40.12507, -8.24485) and data from feeder 6 and line 2 was used. In order to sample EV charging validation data, the MOBI.E® website was used because it has publicly accessible real time data of the EV chargers that belong to its network [3]. The used EV charging station is in Alameda Armando Gonçalves, Coimbra, Portugal (coordinates 40.215254, -8.413305) and comprises of a $3.7\ kW$, type 2, slow charging station. The state of the charger was obtained with a sampling time of 5 minutes, and the area around the EV charger can be described as being mainly industrial.

Furthermore, the days that comprise our data lake are: 5-12 and 14-31 of March, 1-23 and 27-30 of April, 1-31 of May, 1-30 of June, 1-26 of July and 25 of September. This work will be conducted using data sampled in these 141 days exclusively. The days in between are not available provided that EV charging validation data could not be sampled.

This research is the product of a R&D project with a collaboration between Instituto Superior Técnico and Eneida Wireless and Sensors, S.A. in the ambit of *Portugal2020*.

II. TEMPORAL ANALYSIS OF P AND Q

Initially, data from the smart sensor was compared with the one from the MOBI.E® operator, responsible for managing the charging station (Figure 1, where *P* is the active power and *Q* is the reactive power). This way, it is possible to determine the exact time corresponding to an EV charging and realize which sampled LV parameters are more influenced by it. As it is possible to observe in figure 1, the most relevant parameter is the active power.

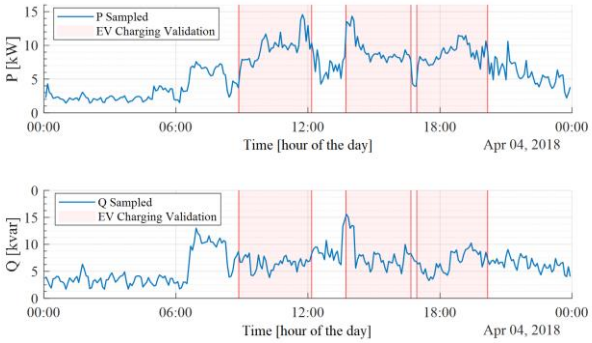

Figure 1: Active power, P, and reactive power, Q, of feeder 6, line 2 in blue and respective EV charging validation in red on 04/04/2018.

The active and reactive power were the chosen features, given that whenever an EV is charging (with slow charging power mode), the main features are the active power increase in demand, with a power factor of 0.98, inductive [5]. To obtain the data patterns that best characterize EV charging events, two types of data aggregations were conducted and tested:

- **Weekday aggregation of P-Q samples**, grouped in *Mondays*, *Tuesdays*, *Wednesdays*, *Thursdays*, *Fridays*, *Saturdays*, *Sundays*;
- **Workday aggregation of P-Q samples**, grouping working days in one set and weekends in another set

Initially, the weekday aggregation was tested in which an average value of samples per timestamp was computed. Then, all sampled P-Q points were plotted and finally, a time-frame differentiation was performed.

III. AVERAGE VALUE OF P AND Q – WEEKDAY AGGREGATION

After collecting the data and overlapping the measurements corresponding to each day of the week, the average value was computed. This average value was obtained making, at each timestamp, a summation of all values (P and Q) and divide them by the number of days. Figure 2 refers to Wednesdays, respectively 7, 14, 21 and 28 of March, 4, 11 and 18 of April, 2, 9, 16, 23 and 30 of May, 6, 13, 20 and 27 of June, 4, 11, 18 and 25 of July, 2018. The blue line represents the values of active (P) and reactive power (Q) in each day, and the red line corresponds to the average of those points in each timestamp.

After that, the same procedure was applied to the same data, however, every time there was an active EV charging, P and Q values were put to zero during the full duration of the charging and the average was computed with the remaining days. This average corresponds to the secondary of the distribution substation line load, without EV charges, as it is possible to observe in Figure 3, in green.

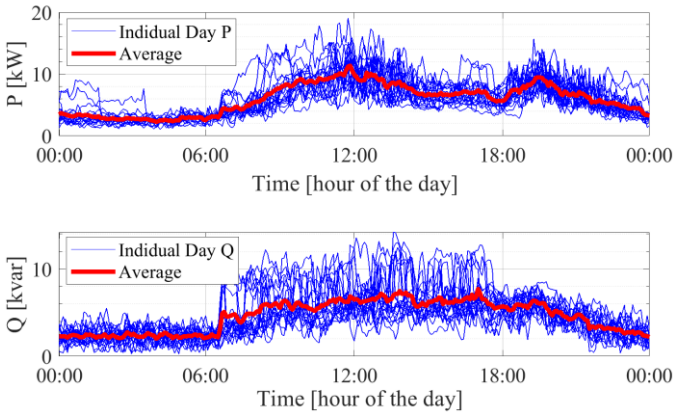

Figure 2: Overlapping of P and Q values (in blue) of each Wednesday and respective average values (in red).

The objective of this analysis is to obtain the average evolution of P and Q over time, with and without EV charging associated. From Figure 3, it is possible to verify the influence of EV charging in the load profile.

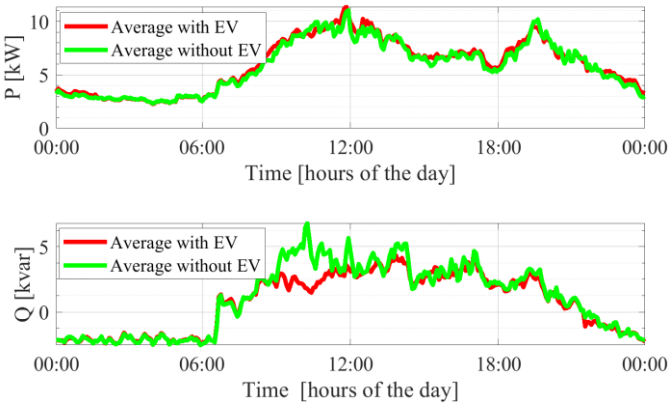

Figure 3: Overlapping of average values of P and Q values of each Wednesday, with (red) and without EV charging (green) periods.

The sum of residuals of the P graph (sum of the y-difference between average w/ EVs and average w/o EVs) from figure 3 is 49.39 and of the Q graph in the same figure is -38.90. These results reveal the lack of data in terms of line load without EV charging, making it not possible to extract a feasible average value.

IV. P-Q PLOT ANALYSIS – WEEKDAY AGGREGATION

In this section, two weekday aggregation analyses were performed based on the hypothesis that in the same weekday, the EV charger users behaviour might be similar, revealing patterns in the active and reactive power, likely to be detected.

The used tool to conduct these analyses was the P-Q scatter plot alongside the origin constrained and unconstrained linear regressions of the data. Two different data classes were computed: class "EV is Charging" corresponding to P-Q points sampled when an EV was connected to the EV charger (in red); class "EV is Not Charging" corresponding to P-Q points sampled when an EV was not connected to the EV charger (in blue). To avoid an overly fitted P-Q plot, forbidding the visualization of patterns, the used dataset was divided in roughly half – between 5 of March and 7 of May, 2018.

A. P-Q analysis based on all extracted samples – weekday separated

A P-Q analysis is now conducted considering all extracted values from the same weekday. It is now possible to visualize the two classes, although not completely isolated from one another. These two classes samples are characterized mostly by different values of active power. According to bibliography [5], when an EV is charging, the active power increases and the reactive one remains almost constant (*PF=0.98*). A linear fitting of each data class was done (constrained to the origin), and it is possible to point out the slope difference between each line. Thus, this became one of the possible parameters for EV charging estimation. Figure 4 is the P-Q plot representation of both classes and the respective origin constrained linear fitting, on Wednesday's.

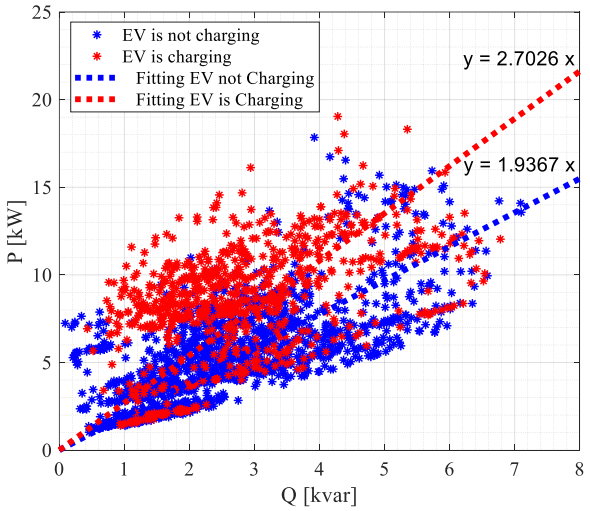

Figure 4: P-Q analysis containing all values with (red) and without (blue) EV charging and with the respective data fitting.

The angle ϕ of the fitting lines of both classes, presented in the P-Q graphs is presented for each weekday, with the purpose of demonstrating results consistency. The results corresponding to each cluster ϕ of the fitting lines, its difference and the norm of the residuals (**normres**) [6] are present in Table 1.

Table 1: Angle in degrees for the lines obtained after the fitting of the P-Q points with and without EV charging, the difference between both and the norm of the residuals for each day of the week.

	MON	TUE	WED	THU	FRI	SAT	SUN
ϕ (º) w/ EV's	68.5	68.4	69.7	70.8	70.8	71.4	72.7
ϕ (º) w/o EV's	59.0	61.6	62.7	62.1	61.7	65.1	62.5
Δϕ (º)	9.5	6.9	7.0	8.8	9.1	6.3	10.1
normres w/ EV's	110.8	108.6	97.5	94.2	111.0	71.3	71.0
normres w/o EV's	52.4	53.9	80.3	58.7	80.0	77.4	48.0

The mean value of the angle difference is 8.22º and the standard deviation is 1.48º. The total norm of the residuals is 254.8 in the case "With EVs" and 173.9 in the one "Without EV's". These quantities are displayed for comparison purposes.

B. P-Q analysis with dual tariff - weekday separated

Now, a separation of data was made by time schedules, namely between the timeframes 00h00 - 07h00 and 07h00 - 24h00. These timeframes represent different periods in which electricity is charged differently according to the consumption and this billing method is exercised in different countries, namely Portugal, by EDP [7]. The results presented in Figure 5, represent the P-Q samples of the previously mentioned Wednesdays but this time, when an EV charging is occurring, they are plotted according to the time frame, in naval and yellow respectively.

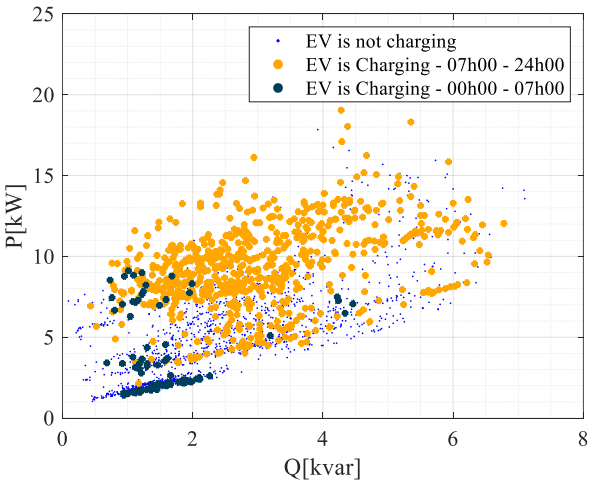

Figure 5: P-Q analysis of Wednesday's data, separated per time schedules. In dotted blue, the samples without EV charging and in naval and yellow, the samples corresponding to an EV charge.

In figure 5, there is a higher number of "EV is Charging" samples from 07h00-24h00 (88%) than 00h00-07h00 (12%). Moreover, from 00h00-07h00, most of the Q values are below *2 kvar* (90% of samples), however with variable P values. The samples with *Q<2 kvar* and *P<3 kW* are, henceforth, called *phantom charges* and they correspond the P-Q points, sampled when an EV battery was fully charged, but were still connected to the EV charger.

The predominance of P - Q samples, at the time of EV charging occurrences and within certain regions, can be explained considering some aspects, namely:

- Proximity to residential and/or industrial areas, since they influence the coupled loads in that line;
- Number of EV charging stations nearby;
- Number of sockets associated with each EV charging station;
- EV charging habits of regular users of that station.

V. P-Q Analysis – Working Days Aggregation

As previously stated, another study was performed, this time aggregating all working days and weekends and not just the same weekdays.

A. P-Q analysis based on all extracted samples – workday aggregated

In a similar way as it was done in section **IV**-*B*, the P-Q plot of all data points during working days was done and the line fitting, constrained to the origin, of both data classes ("EV is Charging" and "EV is Not Charging") was realized and represented in Figure 6.

For this data, the resulting ϕ angle of both linear fittings is of 69.54º and 61.38º for classes "EV is Charging" and "EV is not Charging" respectively with a difference between them of 8.14º. The norm of the residual is 236.05 and 149.94 respectively.

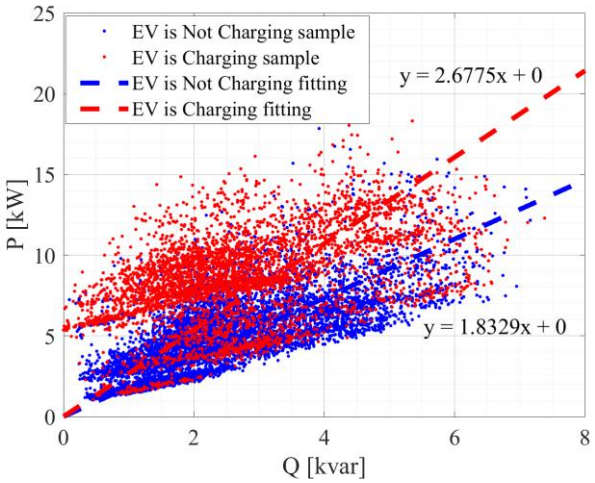

Figure 6: P-Q analysis containing all working days samples with (red) and without (blue) EV charging and the respective data fitting constrained to the origin.

In order to successfully build an EV charging classifier from the P-Q power sampling from the secondary of the substation, the chosen data aggregation was the one including all samples aggregated by workday/weekend, taking into consideration, the ϕ angle difference between fitting lines, as well as the norm of the residuals that present similar values. Moreover, it is needed 5X more data to sample the weekday aggregation than the working day aggregation, to have the same number of useful samples needed to build the EV charging classifier.

Knowing which data aggregation to use (workday aggregation), an unconstrained linear fitting (linear regression) [8] was computed, resulting in a better modelling of each class, with a lower error when compared with the origin constrained one. The Figure 7 is representative of that fitting and the norm of the residuals is 160.4 and 141.1, respectively using classes "EV is Charging" and "EV is Not Charging".

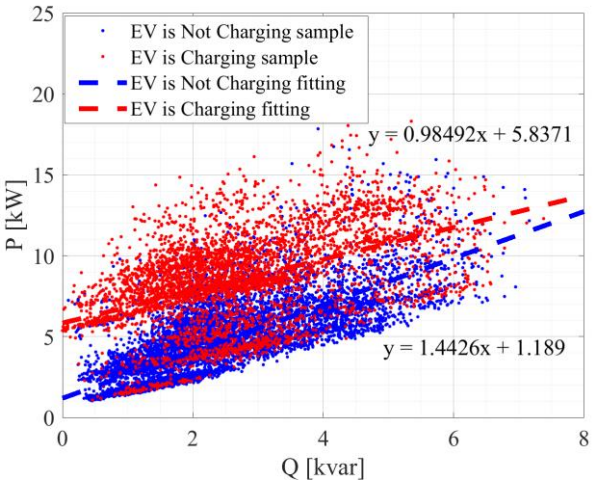

Figure 7: P-Q analysis containing all working day samples with (red) and without (blue) EV charging and with the respective unconstrained data fitting.

VI. Gaussian Distribution Function

With the final objective of obtaining a comparative likelihood between EV charging and non-charging, the Gaussian distribution curve of both classes was computed. The mean (μ) is given by the unconstrained linear regression, presented in the previous section, and standard deviation (σ) obtained by the cartesian distance between each data point (*P*-*Q* pair) and the correspondent clusters linear regression. Moreover, the *Gaussian Membership Function* (GMF) [9] was used, establishing two relative likelihood values for each *P*-*Q* point, associated with each class. The comparison between these two likelihoods results in a binary prediction of EV charging, constituting the baseline solution.

For each class, the distance between a sampled point (Q_i, P_i) and the fitting line of all data points $(ax + by + c = 0)$ is given by Equation 1, with i being the index of the point.

$$d_i = \frac{\left|aQ_i + bP_i + c\right|}{\sqrt{\left(a^2 + b^2\right)}} \quad (1)$$

The standard deviation σ_{gmf} is then given by Equation 2, being n the total number of *P*-*Q* points used in the "training" of the model.

$$\sigma_{gmf} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \quad (2)$$

The Gaussian membership function is defined by Equation 3, with d being the distance from a *P-Q* point to the correspondent class fitting line and $\mu = 0$:

$$f_{gmf}(d) = e^{\frac{-(d-\mu)^2}{2\sigma_{gmf}^2}} \quad (3)$$

The Figure 8 is representative of the GMF computed using the classes previously presented.

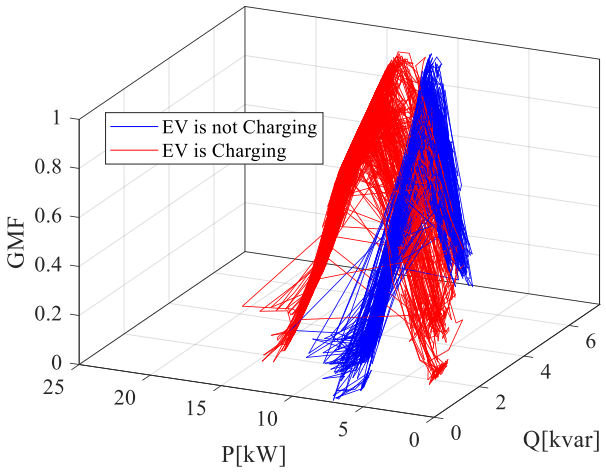

Figure 8 : Gaussian Membership Function for classes "EV is Not Charging" in blue and "EV is Charging" in red, using unconstrained linear data fitting as mean value.

Having the previously explained model "trained", it is possible to obtain a likelihood associated with each class having a *P-Q* point. The baseline solution of the proposed classifier is obtained by direct comparison of the likelihood associated with the class "EV is Charging" and "EV is Not Charging".

VII. Optimizations – Different Solutions

In order to effectively classify an extracted sample as belonging to the class "EV is Charging" or "EV is Not Charging", different algorithmic optimizations were formulated. They not only consider a purely analytical correlation of *P* and *Q*, as it is done with the baseline solution, but also a contextualized reasoning, taking into consideration EV charging characteristics. This includes the temporal evolution of consecutive samples and a threshold establishment between likelihood values to validate a prediction.

The performed optimizations are:

A. *Without temporal filter, binary – baseline solution;*

B. *With temporal filter, binary;*

C. *Without temporal filter, weighted;*

D. *With temporal filter, weighted*

In this matter, given an observation $x = \{P, Q\} \in \mathbb{R}^2$ and wanting to predict its class $y = \Omega, \Omega = \{Y, N\}$, being Y the positive prediction and N the negative one, each classifier optimization function is denoted as $\hat{y}_j = f_j(x) \in \Omega$, with $j = \{A, B, C, D\}$ [8]. Moreover, the GMF likelihood associated with the class "EV is Charging" is denoted as L_{gmf}^{C} and the likelihood associated with the class "EV is Not Charging" is denoted as L_{gmf}^{NC}.

All these optimizations are conducted in the testing part of the classifier, posterior to the training of the GMFs associated with each class,

A. *Without Temporal Filter, Binary – Baseline Solution*

This optimization comprehends the baseline solution with which the classifier determines whether a sample is labelled as belonging to the class "EV is Charging" (Y) or "EV is not Charging" (N). For every *P-Q* sample of the testing set, a likelihood in each of the two GMFs is computed and if the one associated with the class "EV is Charging" is greater or equal than the one associated with class "EV is Not Charging", the classification is positive. Otherwise it is negative.

$$\hat{y}_A = f_A(x) = \begin{cases} Y, & if\ L_{gmf}^{C} \geq L_{gmf}^{NC} \\ N, & otherwise \end{cases} \quad (4)$$

This solution is irrespective of GMF likelihood of each class, taking only into consideration the direct comparison between them.

B. *With Temporal Filter, Binary*

In this solution, a chronological analysis of the extracted samples is taken into consideration when performing the classification. The reasoning behind this solution concerns the actual users behaviour of that EV charger. The Figure 9 represents a histogram of the EV charging duration during the workdays of approximately the 5 months correspondent to the full dataset.

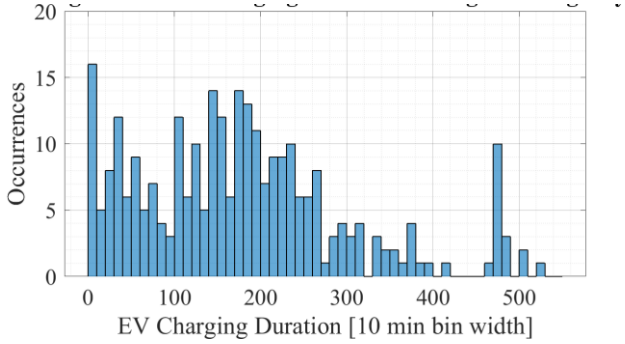

Figure 94 : Histogram of EV charging duration during workdays with 10 minutes bin width.

Observing Figure 9, most of the EV charges have a duration between 0 and 10 minutes. By individual inspection on the EV charges with less than 10 minutes and between 10 and 20 minutes, both the extracted *P* and *Q* evolution show very little variation, with almost no impact on the grid. With this information, a temporal filter was computed in which only consecutive samples classified as "EV is Charging", according to the baseline solution and with a duration bigger than 20 minutes are to be validated as such, represented in Equation 5. Furthermore, a new temporal filter is also applied in which there is a minimum time interval of at least 5 minutes in between EV

charges, validating classification as “EV is Not Charging” class, represented in Equation 6. This second temporal filter ($\hat{y}_{B2}$) proceeds the first one ($\hat{y}_{B1}$) and the main advantage of this solution is the minimization of isolated predictions that have no feasible meaning.

$$\hat{y}_{B1} = f_{B1}(x) = \begin{cases} Y, if\ [L^{C}_{gmf} \geq L^{NC}_{gmf}]\ AND\ \Delta t_{[L^{C}_{gmf} \geq L^{NC}_{gmf}]} > 20\ min \\ N, \quad otherwise \end{cases} \quad (5)$$

$$\hat{y}_{B2} = f_{B2}(x) = \begin{cases} N, if\ [L^{NC}_{gmf} > L^{C}_{gmf}]\ AND\ \Delta t_{[L^{NC}_{gmf} > L^{C}_{gmf}]} > 5\ min \\ Y, \quad otherwise \end{cases} \quad (6)$$

C. Without Temporal Filter, Weighted

The solution C is also an incremental optimization on the baseline solution nonetheless with a threshold (T), in terms of the difference between GMF likelihoods, to validate a prediction. With the baseline solution, an implicit threshold of $T = 0$ was assumed. In this solution, a positive prediction is obtained if $L^{C}_{gmf} \geq L^{NC}_{gmf} + T\ , T\ \in]0{,}1[$.

To obtain an accurate T value, the truly predicted samples and falsely predicted ones had to be taken into consideration. In another words, T was chosen so that the true positive and true negative predictions and are maximized and false positive and false negative predictions minimized. To accomplish that, a histogram comparing the true positive predictions with the positive ones [10] was computed and it is presented in Figure 10.

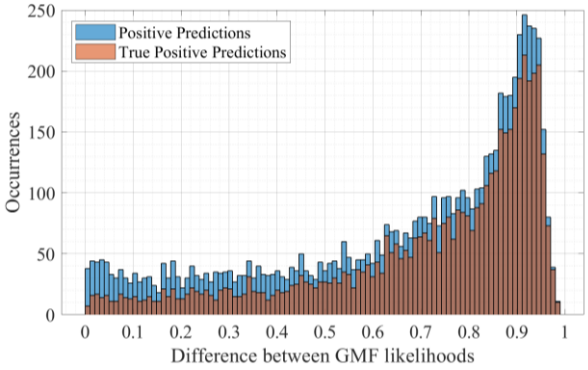

Figure 10: Histogram representing the number of positive and true positive predictions with different likelihood difference between GMFs of classes ”EV is Charging” and ”EV is Not Charging”.

Based on Figure 10, the chosen threshold T was 0.1 given that the total percentage of true positive predictions out of the positive ones when $T < 0.1$ is of 36.9% while between 0.1 and 1 is of 76.8%. In this manner, the expression that represents this solution is represented in Equation 7.

$$\hat{y}_{C} = f_{C}(x) = \begin{cases} Y, if\ L^{C}_{gmf} \geq \left(L^{NC}_{gmf} + 0.1\right) \\ N, \quad otherwise \end{cases} \quad (7)$$

D. With Temporal Filter, Weighted

The last solution results from a confluence of the three previously presented optimizations, using the baseline solution with both the temporal filter, explained in VII. *B* and the threshold value explained in VII. *C*.

The expression of this solution is presented in Equation 8 and 9.

$$\hat{y}_{D1} = f_{D1}(x) = \begin{cases} Y, if\ [L^{C}_{gmf} \geq (L^{NC}_{gmf} + 0.1)]\ \ AND \\ \qquad \Delta t_{[L^{C}_{gmf} \geq (L^{NC}_{gmf} + 0.1)]} > 20\ min \\ N, \quad otherwise \end{cases} \quad (8)$$

$$\hat{y}_{D2} = f_{D2}(x) = \begin{cases} N, if\ \left[\left(L^{NC}_{gmf} + 0.1\right) > L^{C}_{gmf}\right]\ AND \\ \qquad \Delta t_{[(L^{NC}_{gmf} + 0.1) > L^{C}_{gmf}]} > 5\ min \\ Y, \quad otherwise \end{cases} \quad (9)$$

VIII. Results

This section comprises the results of the proposed classifier and its subsequent optimizations. These results encompass the performance assessment using the confusion matrix metric [10] and its respective performance indices alongside a 5-fold cross validation [11] to ensure impartiality in subsection **VIII**-*A*. Furthermore, the feasibility limits of this classifier are tested in terms of different levels of *background* power load, in subsection **VIII**-*B*, as well as EV charging power load, in subsection **VIII**-*C*.

A. Overall Results regarding the Classifier Performance

To have a realistic assessment of the performance of this classifier solutions, the confusion matrix [10] was computed, and some efficiency indices extracted. These indices comprise the overall *accuracy* and *Miss-Classification Rate* (MCR) as well as the *True Positive Rate* (TPR), *False Positive Rate* (FPR), *specificity* and *precision*. Furthermore, to increase the impartiality of the assessment of the classifier performance a stratified 5-fold cross-validation [11] was done using the full dataset and the results are presented in Figure 11. Moreover, it is worth considering the prevalence level of that EV charger of 30.8%, indicating the usage percentage of that EV charger with the used *data lake*.

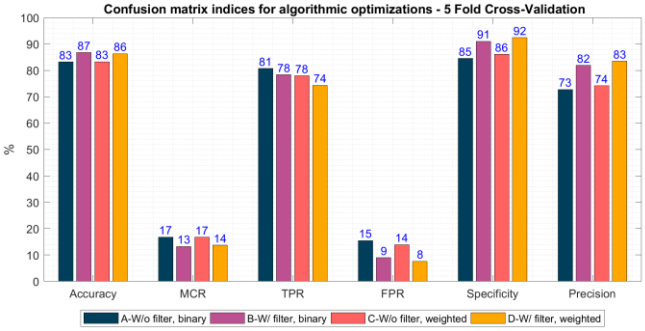

Figure 11: Representation of the confusion matrix indices to evaluate the classifier performance of each of the optimizations A-D, being MCR the miss-classification rate, TPR the true positive rate and FPR the false positive rate, using a 5-fold cross validation.

In Figure 11, the first thing to notice is the *accuracy* and *miss-classification rate*, respectively between 83% - 87% and 13% - 17% across solutions *A* to *D*. These results give an overall perspective on the classifier performance. However, they can be misleading, hence the need for the other indices. Moreover, knowing that the *prevalence* level is around 31%, there is significantly more data concerning the class "EV is Not Charging". This information might be related with the difference between the TPR and *specificity* - how often does the classifier predicted that an EV charging event was not occurring when, in fact, it was not occurring, respectively 74% - 81% and 85% - 92% across solutions. In addition, the *precision* – percentage of correct positive predictions - ranges from 73% - 83% across solutions.

In conclusion, the four presented classifier solutions have different performances and they can be employed according to the DSO specification. In the case that the TPR performance is more important, at the expense of a worse FPR, the solution A is more suitable. On the other hand, if the minimization of the FPR is more important than the maximization of the TPR, the solution B is the best one. In addition, if the specificity and precision is preferable when performing the classification, solution D is more adequate. In the end, the B solution is the one that best maximizes the TPR and minimizes the FPR.

B. Feasibility Limits of Proposed Classifier – Background Power Load Variation

In the feeder and line where the smart meter is installed, the associated load is not exclusive of the EV charger in study. The total power load excluding the one associated with the EV charging is, henceforth, denominated as *background* load. This analysis will assess the classifier performance when subjected to different levels of background power load magnitude, illustrating the feasibility of the proposed classifier when the load increases/decreases.

To accomplish this assessment, the following assumptions were taken into consideration:

- An EV charging is characterized by unitary power factor;
- The P of an EV charging is constant and equal to 3.7 kW during the full duration of charge.

In detail, the classifier feasibility is determined when the background *P* and *Q* load varies linearly by a multiplication factor K. To accomplish that, whenever a charge was occurring, the P load was subtracted 3.7kW, remaining the original *background* load (P_1^{bk} *and* Q_1^{bk}). Then a multiplication of this quantities is performed by $K, K \in [0.1, 5]$ resulting in a new *background* load $[P_k^{BK} = P_1^{BK} \times K\ (kW)\ AND\ Q_k^{BK} = Q_1^{BK} \times K\ (kvar)]$. Then, 3.7kW are added whenever an EV charging occurred. With these modified data, the classification model is trained and furthermore tested, with the *accuracy*, TPR and *specificity* results presented in Figure 12.

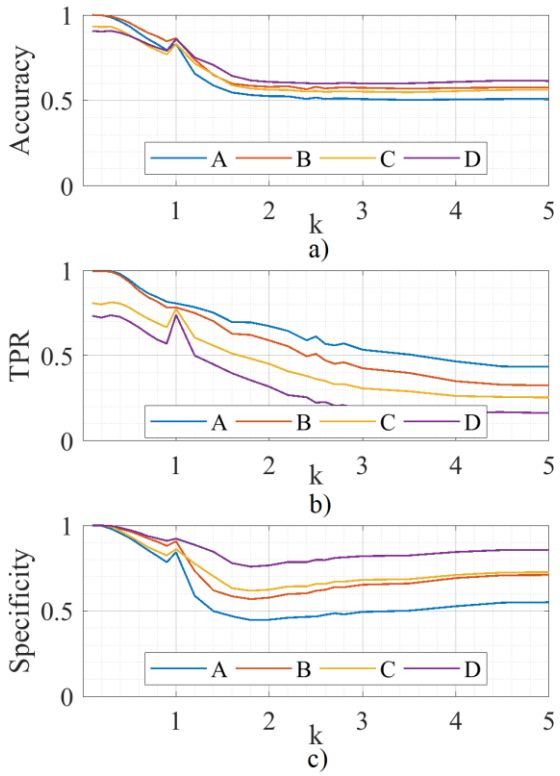

Figure 12: *Accuracy* (a)), TPR (b)) and s*pecificity* (c)) of each solution A to D, for a *K* variation of background P and Q load.

Looking at Figure 12, the *accuracy* of the classifier decreases with an increase in the background load ($K > 1$), which also means that the MCR increases giving that they are complementary. Supporting this result, the TPR decreases with a *K* factor increase and the *specificity* drops. Now, when the background load is lower than the original ($K < 1$), the classifier presents a better performance. With a decrease in K, the TPR and the *specificity* greatly improve, resulting in an overall *accuracy* performance increase.

These results are somewhat expected giving that, when the background load is low, the load corresponding to the EV charging is more easily disaggregated than when the background load is high, in which a 3.7 kW load variation is less significant.

C. Feasibility Limits of Proposed Classifier – EV Charging Load Variation

Up until this point, the research was conducted using the actual *P* demand of the used EV charger – *3.7 kW* – corresponding to *Slow Charging* mode [12]. In this subsection, different *P* charging values (P_k^{CH}) are tested to understand the feasibility of this classifier when it is applied to charging stations with different power modes. For instance, instead of a *3.7 kW* charging station, what would be the classifier behaviour when there is a *15 kW* charging station, under the same *background* load.

To do that and following a similar reasoning as the one from the previous subsection, a $K \in [0.1, 5]$ multiplication factor is used to change the P of EV charging, simulating different charging

modes. Likewise, when an EV charging is occurring, *3.7 kW* are subtracted from the original *P* load, which are then multiplied by *K* ensuring different EV charging active power magnitude $[P_k^{CH} = 3.7 \times K\ (kW)]$. Then, P_k^{CH} is further added to the background load and the classification model is further trained and tested.

Moreover, there are two assumptions when performing this analysis, like the ones from the previous subsection:

- An EV charging is characterized by unitary power factor;
- The P of an EV charging is constant and equal to $(3.7 \times K)\ kW$.

The *accuracy*, TPR and *specificity* results of the classifier performance with different EV charging power magnitudes is presented in Figure 13, being K the multiplication factor.

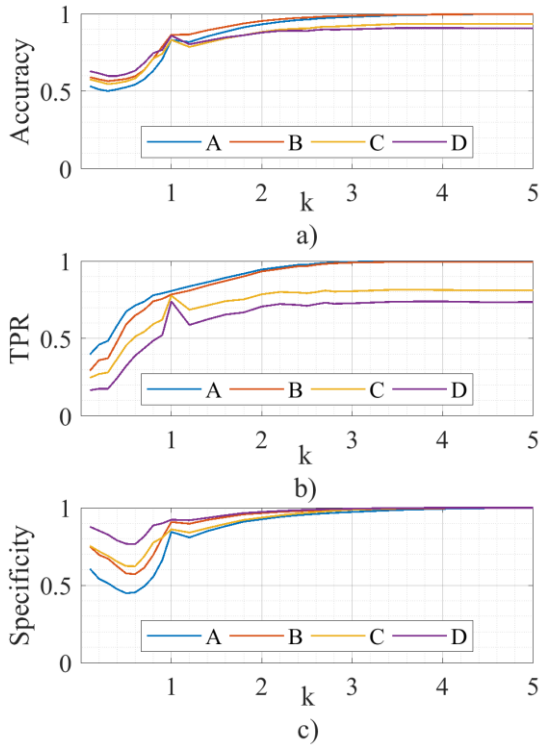

Figure 13: *Accuracy* (a)), TPR (b)) and *specificity* (c)) of each solution A to D, for a *K* variation of the EV charging load.

It is possible to conclude that, with an increase in the EV charging load ($K > 1$), both the TPR and the *specificity* increase, resulting in an overall *accuracy* increase of the classifier performance. Moreover, since the FPR is complementary of the *specificity*, it decreases with a *k* factor increase. These results are once more expected, giving that with an increase in the EV charging power, the dissociation of P – Q samples when building and testing the model is greater, resulting in a better classification performance.

On the other hand, with a decrease in *K*, the *accuracy*, TPR and *specificity* of the classifier drop initially and even presents a slight inflexion point. The performance drop, when the *K* value decreases from $K = 1$ and until it starts to increase again, can be explained with a decrease in significant EV charging P demand from the grid when compared with the background load. The increase in performance, as K is approaching zero, results from the imposed assumptions, namely the one in which the EV charging *P* is constant and equal to $(3.7 \times K)\ kW$, which is only true when the State of Charge (SoC) of the vehicle is less than 85%, in the *slow charging* mode, as well as the unitary power factor imposition [5].

IX. EV CHARGING PROFILE DESIGN

In this chapter, a behavioural model of the users of the EV charging station in study is presented. These statistical models can further extend the previous existing knowledge and can be used either to create, complement or validate an existing dataset as well to estimate the LV grid behaviour when one, or multiple EV charging loads are added.

The used features, based on the histograms, were the following: Number of EV charging events per day (section **IX.** *B*), EV charge starting time of the day (section **IX.** *C*) and duration of EV charges (section **IX.** *D*). Moreover, the *Gaussian Mixture Model* (GMM) [4], [13] algorithm was used to obtain some of the PDF of the models (section **IX.** *A*). Given that the SoC of the vehicle cannot be obtained, an assumption had to be taken into consideration, in which the EV charging *P* is constant and equal to *3.7 kW* throughout the full charge, with unitary power factor. Moreover, this analysis focuses on the working days of the week.

Having the *probability density function* (PDF) models of the previously stated features, the following steps were taken to create an EV charging profile:

1. Random selection of number of charges per day constrained to be greater or equal to zero charges. For each charging event follow the steps (2-5);
2. Random selection of start time of the day, constrained to being greater than 00:00h and less than 24:00h;
3. Random selection of duration of charge, constrained to being greater than zero;
4. In the case where there is more than one charge per day, if after establishing the duration of an EV charge, there is an overlap, a new duration random sample is extracted. Repeat step 3 until there is no EV charging overlap.
5. Having the time periods in which EV charges are occurring, increase active power demand by *3.7 kW*.

A. *Gaussian Mixture Model*

This probabilistic model was based on some bibliography, namely [4] and [13] and it is composed of a finite sum of Gaussian PDFs (L), each with respective weight (w_i), mean (μ_i) and variance (σ_i^2). The integral of the PDF over the sampled space has to be one, the mixture weights must so that $0 \leq w_i \leq 1$ and their sum must be unitary ($\sum_{i=1}^{L} w_i = 1$). In this matter, $f_Y(y)$ represents the GMM which is created by the sum of *L* individual weighted Gaussian components, whose random variable Y is defined in Equation 10.

$$f_Y(y) = \sum_{i=1}^{L} w_i \, f_{N(\mu_i,\sigma_i^2)}(y) \quad (10)$$

The i^{th} GMM component PDF is presented in Equation 11 with random variable Y mean (μ_Y) and variance (σ_Y^2) presented in Equations 12 and 13 respectively.

$$f_N(\mu_i, \sigma_i^2, x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}} \quad (11)$$

$$\mu_Y = \sum_{i=1}^{L} w_i \, \mu_i \quad (12)$$

$$\sigma_Y^2 = \sum_{i=1}^{L} w_i \, [\sigma_i^2 + (\mu_i - \mu_Y)^2] \quad (13)$$

To obtain the necessary parameters used in the GMM, namely each i^{th} individual component weight (w_i) and standard deviation (σ_i), the *Expected Maximization* (EM) algorithm was used [4], [8], [13]. This algorithm input is the desired number of Gaussian components L as well as each individual Gaussian parameters, $\Gamma = \{\gamma : \gamma = \{w_i, \mu_i, \sigma_i, \sum i\}_{i=1}^{L}\}$, and it iteratively computes the expectation (E) of the log-likelihood of the complete data using the current estimate for the parameters Γ, followed by a computation of those same parameters, maximizing (M) the previous log-likelihood. This procedure is executed until convergence is achieved and with Γ values randomly initialized [13].

B. *Number of EV charges per day*

To model this feature, the GMM algorithm was not necessary, given the histogram shape, possible to see in Figure 14. A Gaussian distribution fitted the histogram, resorting to the *Maximum Likelihood Estimation* for the mean with the estimate value of the standard deviation parameter being the square root of the unbiased estimate of the variance [8]. In Figure 14 it is presented both the histogram of the number of EV charges per day, as well as the correspondent Gaussian PDF.

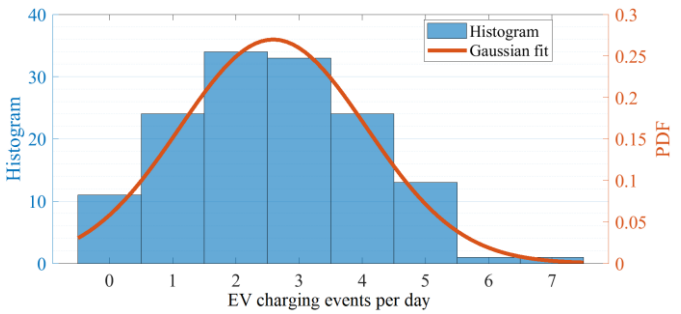

Figure 14: Histogram, in blue, and Gaussian PDF, in orange, of EV charging occurrences per day, on working days of the week.

Hence, the Gaussian PDF function is presented in Equation 11, being $i = 1$, with an estimated average of 2.58 and standard deviation of 1.48. When creating the EV charging profile, a round of the random sample from this PDF must be conducted.

C. *Beginning of EV Charge - Time of the Day*

Looking at the histogram that represents this feature in Figure 15, it is possible to infer that its shape is a rather complex one and it is not easily represented by any of the traditional distributions, hence the need for the GMM [4]. Figure 15 also contains the GMM fit of this distribution with $L = 2$.

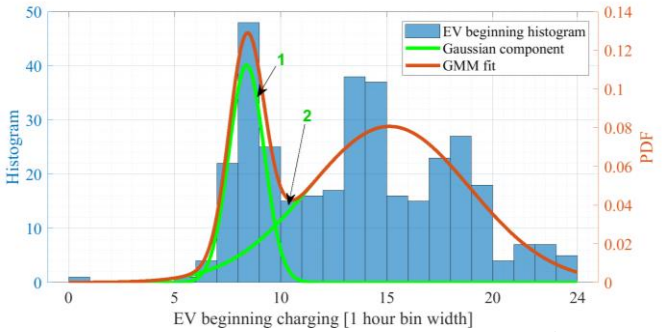

Figure 15: Histogram, in blue, and GMM PDF, in orange, of beginning of EV charging, for workdays, throughout the day. Individual Gaussians that compose the GMM in green.

The most likely time periods to begin an EV charging, according to Figure 16, are between 08h00 and 09:00h and 13h00 and 15:00h. Given the industrial area near the EV charger, these periods of time represent the driver's arrival at work in the morning and after lunch. Another time period with great number of occurrences is between 18h00 and 19:00h, possibly indicating the existence of a residential area nearby, in which EV charging begins when its users arrive home. Concerning the EM algorithm, it converged with an error $\epsilon <$ 1%, and the final Γ parameters are presented in Table 2.

Table 2: Γ parameters of beginning of EV charging GMM components.

GMM i^{th} Component	w_i	μ_i	σ_i
1	0.23	8.40	0.83
2	0.77	15.14	3.79

D. *EV Charging Duration*

In this subsection, a PDF model of the EV charging duration is computed using the GMM. In Figure 16 it is represented the histogram of EV charging duration in minutes, without the bin corresponding to 0-10 minutes, for the reasons stated in section **VII.** *B*, as well as the corresponding GMM fit.

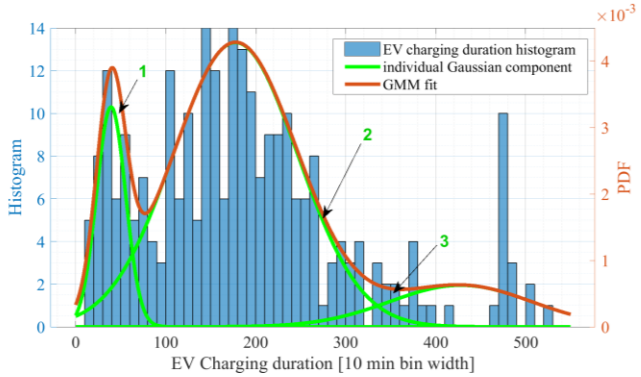

Figure 16: Histogram, in blue, and GMM PDF, in orange, of EV charging duration during workdays. Individual Gaussians that compose the GMM in green.

This GMM has $L = 3$ different components and it constitutes a statistical model to obtain EV charging duration random samples. The EM algorithm converged with an error $\epsilon < 1\%$. From Figure 16 it is possible to establish three main EV charging duration periods, represented by individual Gaussian components that constitute the GMM, in green. The final Γ parameters – weighting (w_i), mean (μ_i) and variance (σ_i^2) of each GMM component - are presented in Table 3.

Table 3: Γ parameters of EV charging duration GMM components.

$GMM\ i^{th}\ Component$	w_i	μ_i	σ_i
1	0.13	39.32	16.16
2	0.74	177.53	69.15
3	0.12	427.78	78.75

X. CONCLUSIONS

The current work presents a methodology for the EV charging identification and classification in a LV distribution system. It uses a smart sensor installed at the secondary of a distribution substation and, by assessing the active and reactive power in each feeder and line, EV charging key features are extracted, making a prediction possible. The current work was done using ENEIDA® DTVI smart sensor installed in the city centre of Coimbra. Additionally, data provided by the MOBI.E® operator website, responsible for managing the EV charging station, was obtained to identify the charging events.

Different approaches for the EV charging identification were studied: 1) from temporal curves of active, P, and reactive, Q, power; 2) using the average values aggregated by weekdays only; 3) from overall P-Q curves aggregated by working days/weekends; 4) from overall P-Q curves aggregated by working days, making a differentiation according to the time of the day. From the previous approaches, the ones providing more accurate information about the EV charging events were the P-Q curves made from an aggregation of all working days/weekends, without time frame differentiation.

With the P-Q curves, it was possible to identify two classes (one without EV charging and other with EV charging events) defined by similar reactive power values but different active power ones. During an EV charging event, the reactive power remains almost the same, but the active power increases steeply. With those P-Q curves, an unconstrained linear fitting line was computed and from that, two Gaussian curves were obtained that probabilistic label a P-Q point with two likelihoods: one of belonging to the class "EV is Charging" and the other belonging to the class "EV is Not Charging". The next step was to define different classification methods that would optimize the EV charging detection. These methods are: A) Without temporal filter, binary – the baseline solution in which a direct comparison is done between the Gaussian likelihoods; B) With temporal filter, binary – an additional temporal filter to validate predictions is applied; C) Without temporal filter, weighted – a threshold application to validate predictions is applied; D) With temporal filter, weighted – both a temporal filter as well as a threshold value are used to validate predictions. The results range from 83% to 87% in classification *accuracy*, across these four optimizations.

Furthermore, two different tests were performed in which the feasibility of this classification model is tested. The first comprise of a performance assessment with different levels of *background* load magnitude, in which the greater the background load, the less an EV charging power is relevant, leading to worse predictions. On the other hand, when the background load is smaller, the classification performance improves. The second assessment verified the behaviour of the classifier when used with different charging power modes. As a result, in this charger location, instead of a *slow charge* power mode charger, if a higher power mode is installed, the classifier performance increases.

Additionally, some statistical analyses were conducted on the behaviour of that EV charger users and a model that allows the realistic creation of an EV charging profile computed and presented. These analyses comprise the number of EV charges per day, the beginning of EV charges throughout the day and their duration, and the *Gaussian Mixture Model* algorithm was used.

References

[1] IEA(2019), "Global EV Outlook 2019," Paris.

[2] "ENEIDA.IO (c)." [Online]. Available: https://eneida.io. [Accessed: 15-Mar-2019].

[3] "MOBI.E - Mobilidade Elétrica (c)." [Online]. Available: https://www.mobie.pt. [Accessed: 15-Mar-2019].

[4] J. Quirós-Tortós, A. N. Espinosa, L. F. Ochoa, and T. Butler, "Statistical Representation of EV Charging: Real Data Analysis and Applications," in *2018 Power Systems Computation Conference (PSCC)*, 2018, pp. 1–7.

[5] J. Quirós-Tortós, L. F. Ochoa, and B. Lees, "A statistical analysis of EV charging behavior in the UK," in *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, 2015, pp. 445–449.

[6] R. W. Farebrother, *L 1 -Norm and L inf -Norm Estimation*. Springer, Berlin, Heidelberg, 2013.

[7] "Opções horárias," *EDP Comercial*. [Online]. Available: https://www.edp.pt/particulares/apoio-cliente/perguntas-frequentes/tarifarios/o-que-preciso-de-saber-para-contratar/o-que-e-a-opcao-horaria-e-qual-a-melhor-para-mim/faq-4823. [Accessed: 08-Oct-2019].

[8] J. S. Marques, *Reconhecimento de Padrões: Métodos Estatisticos e Neuronais*, 2nd ed. IST Press, 2005.

[9] T. J. Ross, *Fuzzy Logic with Engineering Applications*, Third Edit. Copyright (c) 2010 John Wiley and Sons, Ltd, 2010.

[10] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, p. 209.

[11] H. Refaeilzadeh, Payam and Tang, Lei and Liu, *Cross-Validation*. 2009.

[12] M. C. Falvo, D. Sbordone, I. S. Bayram, and M. Devetsikiotis, "EV Charging Stations and Modes : International Standards," in *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*.

[13] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 29–37, Feb. 2010.