IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 April 2025, accepted 10 May 2025, date of publication 15 May 2025, date of current version 22 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3570682

RESEARCH ARTICLE

An Ensemble-Based Classifier to Determine the Harmonic Contribution Responsibility Between Utility's Grid and Microgrids

MATHEUS C. MELLA[1], RICARDO A. S. FERNANDES[2], (Senior Member, IEEE), AND DANIEL BARBOSA[3], (Senior Member, IEEE)

[1]Graduate Program in Electrical Engineering, Federal University of São Carlos, São Carlos 13565-905, Brazil
[2]Department of Electrical and Computer Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos 13566–590, Brazil
[3]Department of Electrical Engineering, Federal University of Bahia, Salvador 40110-909, Brazil

Corresponding author: Ricardo A. S. Fernandes (ricardo.asf@usp.br)

This work was supported in part by São Paulo Research Foundation (FAPESP) under Grant 2023/00182-3, and in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant 406453/2021-7.

ABSTRACT The integration of nonlinear loads and distributed energy resources in microgrids (MG) has increased the levels of harmonic distortions in distribution feeders. Identifying the responsibility for the harmonic distortions measured at the MG's point of common coupling (PCC) is essential for managing the power quality. For this purpose, the present paper aims to determine whether the responsibility predominates from the utility's grid or the MG itself. Thus, an ensemble-based approach was proposed, comparing the behaviour of *Random Forest* and *XGBoost* classifiers to identify the side with the predominant harmonic contribution. In order to validate the proposed approach, scenarios were simulated in the IEEE 34-node test feeder using the ATP/ATPDraw software, considering the harmonic sources located at different nodes of the feeder. Three-phase voltage and current signals were acquired at MG's PCC to train and validate the ensembles. However, a feature extraction stage was performed to obtain time and frequency features used as inputs to the predictive models. The results demonstrated that the proposed approach effectively identified the contribution side, achieving an average F_1-score above 99% for both ensembles, representing a feasible solution for determining the harmonic distortion responsibilities at the MG's PCC.

INDEX TERMS Harmonic contribution, harmonic distortion, machine learning, power distribution systems, power quality.

I. INTRODUCTION

The growing use of nonlinear loads and distributed energy resources (DERs) connected to power distribution systems introduces new challenges in terms of power quality (PQ) [1], [2]. In power distribution systems, the massive presence of harmonic sources leads to the degradation of the PQ delivered to consumers [3], [4].

This concern is particularly significant due to the increasing deployment of inverter-based distributed generators, which is especially prevalent in the context of microgrids (MG). In this sense, MGs are potential regions of a power distribution system to have a greater density of harmonic sources, as they represent a conglomerate with distributed generators (usually based on inverters) that aim to fully or partially supply the local demand formed mainly by nonlinear loads [5]. Despite this high density of harmonic sources, the MGs are a trend, driven by the potential to establish more resilient and reliable grids [6].

The challenges imposed by harmonic distortions caused by the utility- or MG-side are notable from the aforementioned scenario. Although regulatory agencies, through standards and/or recommendations [7], [8], aim to define acceptable levels of harmonic distortions in distribution grids, there is a lack of research regarding the harmonic contribution responsibility, i.e., the determination of the predominant side

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Bellan.

in terms of harmonic distortion measured at the point of common coupling (PCC). It is important to mention that a PCC represents a location (node) common to different consumers. In the present paper, it is also a common node to connect an MG.

Currently, research has sought to quantify the contribution of harmonic distortions at points of common coupling (PCC) between the medium voltage network (utility-side) and a consumer (commonly an industrial plant). Classical approaches for identifying dominant harmonic sources are generally based on harmonic power [9] and/or harmonic impedance [10], [11], [12], [13]. It is worth mentioning that most of these approaches have the disadvantage of requiring the estimation of the networks' harmonic impedance, which is not a trivial task. This way, machine learning-based approaches have been proposed. Among them, it can be highlighted those based on neural networks [14], [15], [16], fuzzy inference systems [17], [18], decision trees [18] and clustering algorithms [19].

Mazumdar et al. [14] proposed an approach based on Multilayer Perceptron (MLP) neural networks to identify the load model, evaluating the impact of changes in the harmonic source impedance to analyze its contribution. Experimental tests were performed, and the approach could predict the total harmonic distortion (THD) on the consumer side, considering only currents measured at the PCC.

In [15], the authors employed a self-organizing feature map (SOFM) network to classify linear and nonlinear loads. The Wavelet transform was used to extract features from V-I curves. The proposed SOFM was compared with Support Vector Machines (SVM), demonstrating a faster training stage and accurate results.

An exact radial basis function neural network (ERBFNN) was proposed by [16] to identify the harmonic contribution in comparison with other neural networks (such as MLP and radial basis function). All the neural networks were trained to estimate the THD of current on the utility- and consumer-side. The proposed ERBFNN obtained accurate results and demonstrated less computational cost.

A Fuzzy inference system was proposed by [17] to determine the harmonic contribution at the PCC. The authors considered measurements acquired at the PCC before and after a capacitor bank switching. This way, a set of Fuzzy systems was adjusted and used to estimate the contributions of odd harmonics. The results presented mean percentage errors ranging from 1% to 10%.

Fernandes et al. [18] proposed an approach based on decision trees (DTs) and neural-fuzzy systems combined with a feature extraction stage. The DTs were used to identify the contribution side (utility or consumer), while the neural-fuzzy systems were employed to estimate the contribution. This approach reached 99% in identifying the contribution side and mean square error between $1.1e^{-2}$ and $3.0e^{-9}$ in estimating the contribution of each harmonic order.

A density-based spatial clustering of applications with noise (DBSCAN) was used by [19] to identify harmonic impedance changes. However, before clustering, the proposed approach considered a filtering stage based on cross-approximation entropy. The filtering was applied to the current and voltage signals measured at the PCC. From the clusters, it was possible to calculate the harmonic contribution for each of them.

According to the previously mentioned context, this paper proposes an ensemble-based classifier to identify the harmonic distortion responsibility at the MG's PCC. The proposed approach employs a feature engineering stage to extract time and frequency domain features from voltage and current signals. Next, Random Forest and XGBoost classifiers were trained and validated to identify the harmonic distortion responsibility, i.e., the contribution side.

To the best of our knowledge, no work has analyzed the harmonic responsibilities at the MG's PCC, being this the main contribution of the paper. Moreover, the proposed approach also advances the state-of-the-art due to its independence in relation to the *a priori* information from the distribution feeder parameters. Therefore, the only use of three-phase voltages and currents measured at the MG's PCC is sufficient to determine the contribution side. In addition, the proposed approach considered a predictive model for each system's phase, making it applicable even in unbalanced networks. These characteristics make the proposed approach a versatile and effective solution for managing the harmonic distortions in MGs.

The remainder of the paper is organized as follows. Section II presents the modeling and simulation aspects of the power distribution feeder and the harmonic sources considered in this paper. The proposed approach for determining the harmonic contribution at MG's PCC is detailed in section III. Results and discussions are presented in section IV. Finally, the section V draws the conclusions.

II. DISTRIBUTION TEST FEEDER AND HARMONIC SOURCES MODELING AND SIMULATION

This section covers the modeling aspects of the distribution test feeder and harmonic sources, which were implemented in the ATP/ATPDraw software [20] for subsequent simulation. Feeder topology and loading were maintained following the specifications presented in [21].

A. IEEE 34-NODES TEST FEEDER

The IEEE 34-node test feeder was used for the analysis, as it is considered long and lightly loaded, allowing harmonic sources to be duly located.

In Fig. 1, the single-line diagram of the aforementioned feeder is shown. The MG was defined from node 858, which contains a small hydraulic facility (SHF) of 1,860 kVA located at node 846 and a photovoltaic (PV) system of 300 kWp that uses inverters to inject power into the grid, connected to node 836.

This feeder is based on a real system in Arizona, USA. It operates with a nominal frequency of 60 Hz,

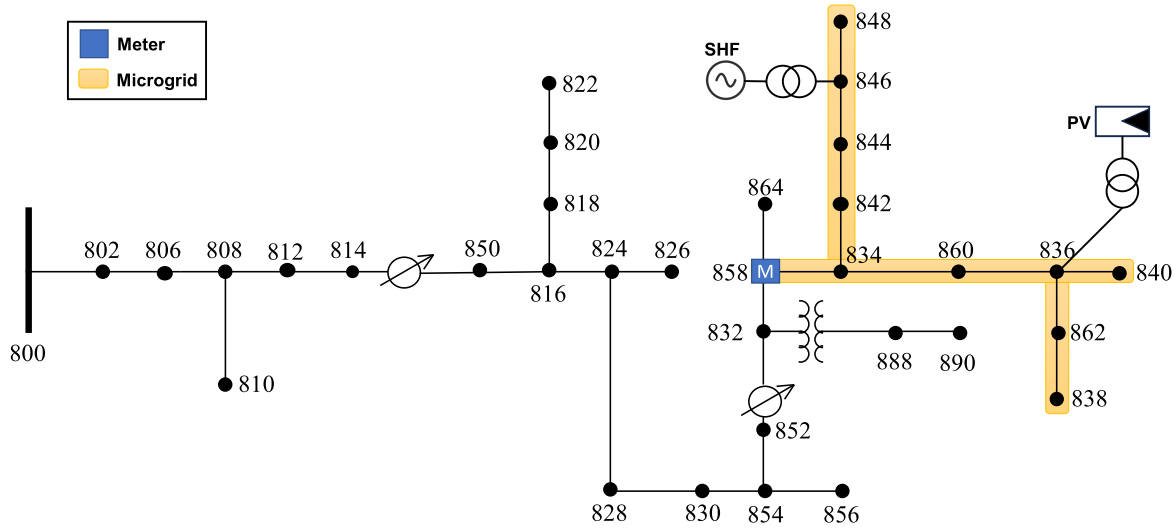

FIGURE 1. IEEE 34-nodes test feeder with the MG region highlighted.

distribution voltage of 24.9 kV, and 12 MVA of distributed and concentrated loads (unbalanced system).

B. HARMONIC SOURCES

From the distribution feeder modeling, harmonic sources (nonlinear loads) were introduced at several three-phase nodes, both on the utility-side (main grid) and the MG-side, in a singular manner. This procedure allows for the identification of the side responsible for the harmonic distortion, which is used as the output variable in the training dataset.

It is important to highlight that different harmonic profiles were analyzed, covering 6- and 12-pulse rectifiers, static frequency converter (SFC), DC motor drive, and thyristor controlled-reactor (TCR). These harmonic sources were modeled according to the harmonic current signatures presented in [22] and [23]. Their characteristics are detailed in Table 1.

TABLE 1. Characteristics of harmonic sources.

Parameters	6-pulse	12-pulse	SFC	DC motor	TCR
3rd (p.u.)	0.015	0.002	0.000	0.012	0.138
5th (p.u.)	0.220	0.006	0.170	0.336	0.051
7th (p.u.)	0.150	0.003	0.101	0.016	0.026
9th (p.u.)	0.000	0.000	0.000	0.000	0.016
11th (p.u.)	0.102	0.062	0.061	0.087	0.011
13th (p.u.)	0.084	0.045	0.044	0.012	0.008
15th (p.u.)	0.000	0.000	0.000	0.000	0.006
$P_{3\phi}$ **(MW)**	0.100	0.300	0.500	1.100	0.724
$Q_{3\phi}$ **(MVAr)**	0.050	0.150	0.200	0.000	0.828
$S_{3\phi}$ **(MVA)**	0.112	0.336	0.539	1.100	1.100

Although the PV system was properly modeled with its inverter to allow both connection and power dispatch, it was not considered as a harmonic source to be identified. This exception was considered because it corresponds to an element of the MG, which was previously dimensioned, planned, and authorized by the utility.

C. SIMULATION SETUP

The simulations considered three operating conditions before the insertion of harmonic sources: (i) energy import mode – where the MG consumes energy from the utility's grid; (ii) energy export mode – where the MG supplies energy to the utility's grid; and (iii) neutral mode – where there is no energy exchange between the MG's generation and the utility's grid.

For each operating condition, simulations were performed to assess the effectiveness of the proposed approach, considering the location of each harmonic source across the 25 available three-phase nodes in the entire system. Of these, 16 three-phase nodes are located on the utility-side (800, 802, 806, 808, 812, 814, 850, 816, 824, 828, 830, 854, 852, 832, 888 and 890), while the remaining 9 are situated on the MG-side (834, 836, 840, 842, 844, 846, 848, 860, and 862). In this study, only one harmonic source was included per simulation, either in the utility- or MG-side.

Additionally, variations of up to ±30% in the apparent power ($S_{3\phi}$) of the harmonic sources were considered for each scenario. Thus, each harmonic source was simulated 50 times for each node. This approach ensured a broad diversity of scenarios, covering a wide range of possible operating conditions.

Finally, three-phase voltage and current measurements were acquired at the MG's PCC. This process resulted in a dataset composed of 18,750 signals. This comprehensive approach allowed the generation of a robust dataset suitable for training and validating the ensembles.

III. ENSEMBLE-BASED APPROACH TO DETERMINE THE HARMONIC CONTRIBUTION

As previously mentioned, the approach outlined in this paper focused on identifying the side of the grid with the predominant harmonic contribution. It is essential to emphasize that,

although it was validated using the IEEE 34-node test feeder, the underlying philosophy of this approach applies to any medium voltage radial feeder. After simulating the scenarios with the presence of harmonic sources, as it is an approach based on supervised machine learning models, there was a need to train them to identify the harmonic contribution side. The steps that compose the proposed approach are shown in the flowchart of Fig.2.

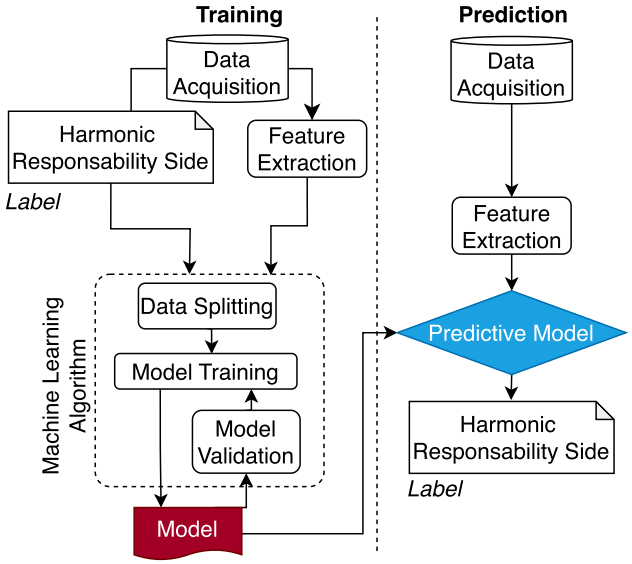

FIGURE 2. Flowchart of the proposed approach.

A. DATA ACQUISITION

This stage involves the acquisition of three-phase voltage and current measurements at the MG's PCC. The PQ meter was configured to perform signal acquisitions with a sampling rate of 7,680 Hz, which corresponds to 128 samples per cycle for the nominal frequency of 60 Hz.

B. FEATURE EXTRACTION

In the context of pattern recognition, the feature extraction stage plays a fundamental role, as it is responsible for preserving relevant information and reducing the complexity of the dataset. In the present paper, the features were firstly extracted in the time domain, considering the calculations of root mean square (F_1), kurtosis (F_2), skewness (F_3), Rényi entropy (F_4), crest factor (F_5), and form factor (F_6):

$$F_1 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}, \tag{1}$$

$$F_2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - mean)^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - mean)^2\right]^2}, \tag{2}$$

$$F_3 = \frac{\sum_{i=1}^{N}(x_i - mean)^3}{(N-1)\sigma^3}, \tag{3}$$

$$F_4 = \frac{1}{1-\alpha}\log\sum_{i=1}^{N}\log(x_i^{\alpha}), \tag{4}$$

$$F_5 = \frac{max(x)}{F_1}, \tag{5}$$

$$F_6 = \frac{F_1}{\frac{\sum_{i=1}^{N} x_i}{N}}, \tag{6}$$

where x is the instantaneous signal measured at the MG's PCC; N represents the signal length; and α is a parameter empirically adjusted to 0.4.

In addition, the THD (F_7) and the magnitudes of odd harmonics from the 3rd to the 13th order (F_8 to F_{13}) were extracted in the frequency domain, considering the Discrete Fourier Transform calculation:

$$H_k = \sum_{i=1}^{N} x_i \cdot \exp\left(-\frac{i2\pi}{N}ki\right), \tag{7}$$

where each harmonic H_k at order k can be obtained with its magnitude and phase angle. Thus, the THD is calculated as:

$$F_7 = \sqrt{\frac{\sum_{j=3}^{13} h_j^2}{h_1}}, \tag{8}$$

being h_j the magnitude of harmonic j, ranging from 3 to 13. All these features were extracted for both voltage and current signals, resulting in a total of 26 features used as inputs.

Since the test feeder is unbalanced, the feature extraction stage was carried out to each system's phase (A, B or C), as illustrated in Fig. 3.

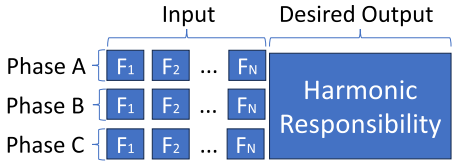

FIGURE 3. Composition of a feature vector for each system's phase.

It is worth noting that, in addition to the features extracted from the signals (input variables), the desired output of the predictive models (target variable) defines the responsibility for harmonic distortion at the MG's PCC.

C. PREDICTIVE MODELS

The models selected and analyzed for the proposed approach were Random Forest and XGBoost, being both ensembles and widely used in classification tasks.

1) RANDOM FOREST

A random forest is constructed from an ensemble of DTs, each created from a random sample of the original dataset. This sample is obtained through the bootstrap process, where observations from the dataset are randomly selected with replacement, creating multiple training sub-samples. This process ensures diversity in the datasets used to train each DT in the forest [24].

During the construction of each DT, a random sample of features is selected at each split node rather than considering all available features. Thus, each DT is responsible for an individual prediction. After training all DTs in the forest, the final set of predictions is obtained through an ensemble process. In random forest, the ensemble method used is voting. Given an input instance, each DT in the forest votes for an output class. Thus, the class with the most votes is selected as the final prediction for that instance. This voting process ensures that the predicted class has the highest number of votes among all the DTs, aiming to reduce the impact of incorrect individual predictions and increase the robustness of the model as a whole [25].

In terms of classification, this technique can handle large and complex datasets without intensive preprocessing. Furthermore, it is less sensitive to outliers and noise than other machine learning algorithms, as the voting process selects the most frequent class among individual DT predictions. This voting approach contributes to increasing the stability and accuracy of final predictions.

2) XGBOOST

The XGBoost (eXtreme Gradient Boosting) classifier represents an advance of the Random Forest algorithm, enhancing predictive performance through gradient boosting techniques [26]. Unlike Random Forest, which constructs independent DTs, the XGBoost builds DTs sequentially. Each new DT aims to correct the errors of its predecessors by focusing on residuals, i.e., the differences between predicted and actual values.

XGBoost incorporates regularization to manage model complexity and mitigate the risk of overfitting, while also efficiently handle sparse data and enable parallel processing. After training, the predictions from all DTs are combined to produce the final output. Thus, each DT's response is weighted according to its importance.

3) TRAINING AND VALIDATION

The dataset was divided into training (70% of the data) and validation (30% of the remaining data) subsets. The training subset plays an important role in the model's learning process, allowing it to assimilate patterns and relationships between input variables and the desired output. In this sense, an adequate training stage allows the model to generalize these patterns to unobserved data, ensuring the model's validity in different scenarios. During the training process, the cycle begins with the initialization of the chosen model. Thus, the model is tuned iteratively using the training subset, optimizing its ability to make accurate predictions.

As previously stated, the test feeder is unbalanced, resulting in feature vectors for each system's phase. This way, it was trained and validated one predictive model per phase, as presented in Fig. 4.

As can be seen, after training these models, the class prediction stage occurred, in which the values of the validation subset were used (excluding the output labels) to make new predictions. Thus, three predictions were obtained (one per phase) and combined into a voting method to select the final predicted class. Subsequently, the final predictions were compared with the true labels to define the overall performance of the proposed approach.

4) MODEL'S FINE-TUNING

In order to improve the predictive accuracy of the models, fine-tuning process was conducted to identify the optimal hyperparameters, with a focus on maximizing the F_1-score.

For the Random Forest classifier, the following hyperparameters were adjusted:

- Number of DTs – from 100 to 150;
- Maximum depth – from 5 to 15;
- Min. number of samples in split nodes – from 4 to 20;
- Min. number of samples in a leaf node – from 4 to 20.

Similarly, for the XGBoost classifier, key hyperparameters were adjusted as follows:

- Number of boosting rounds – from 100 to 150;
- Maximum depth – from 3 to 15;
- Learning rate – from 10^{-3} to 0.1.
- Fraction of samples used to adjust DTs – from 0.5 to 0.7;
- Fraction of features used for each DT – from 0.05 to 0.5;
- Minimum loss reduction – from 0 to 5;
- L1 regularization term – from 0 to 10;
- L2 regularization term – from 0 to 10.

The fine-tuning process is essential for achieving the better model performance, ensuring that the selected parameters effectively enhanced the F_1-score during validation.

5) PERFORMANCE EVALUATION METRICS

The proposed approach was evaluated using commonly employed metrics, such as: accuracy (*acc*), precision, recall, and F_1-score. In addition, the confusion matrix was also used to demonstrate the models' behavior. These metrics are critical to determining the quality and effectiveness of the models against the validation data. They are based on values of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) predictions.

Accuracy (*acc*) measures the proportion of correct predictions to the total number of predictions made by a model [27]. In this sense, the greater the accuracy, the better the model performance.

$$acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{9}$$

Precision evaluates the proportion of TP to the total instances classified as positive [27]. It is useful to avoid FP.

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{10}$$

Recall measures the proportion of predicted TP relating to the total TP instances [27]. It is important when it is critical

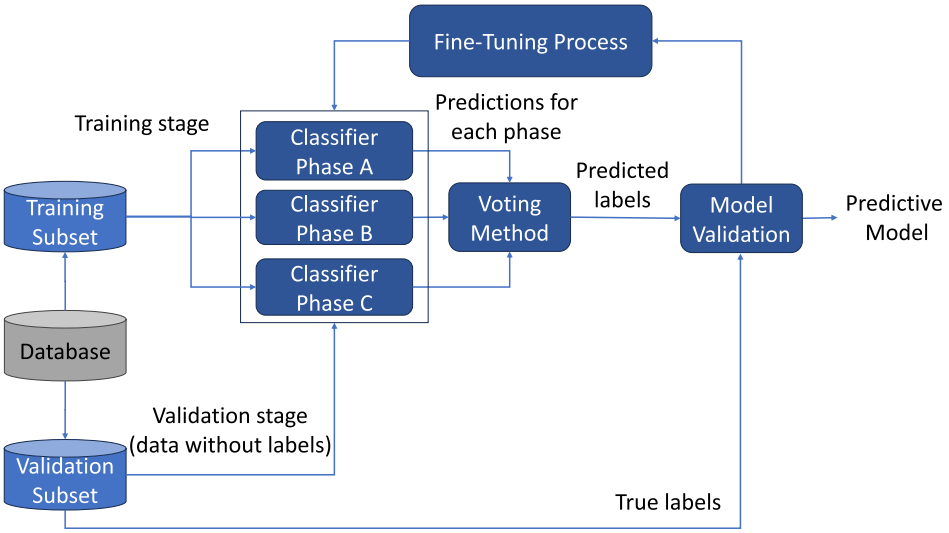

FIGURE 4. Training and validation flow of the predictive models.

to identify all positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{11}$$

F_1-score is a metric that combines precision and recall into a single score, considering both FP and FN [27]. The higher the F_1-score, the better the balance between precision and recall, minimizing the bias of the majority class in unbalanced datasets.

$$F_1\text{-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}. \tag{12}$$

The confusion matrix summarizes the performance of the classification model, showing the number of TP, TN, FP, and FN [27]. It is useful to understand where the model presents correct and incorrect predictions.

IV. RESULTS AND DISCUSSIONS

This section presents the performances of Random Forest and XGBoost classifiers in predicting the harmonic contribution responsibility side. A binary classification was used, considering the label "0" for scenarios where the harmonic source is located on the utility side. In contrast, the label "1" was used to designate the location of the harmonic source on the MG-side.

A. OVERALL CLASSIFICATION

As previously mentioned, fine-tuning was considered to obtain the optimal hyperparameters for both classifiers. Thus, these hyperparameters can be verified in Table 2.

The overall performances of each classifier, including accuracy, precision, recall, and F_1-score for each harmonic source, are shown in Tables 3 and 4.

As the dataset is unbalanced, i.e., the class labeled as "0" is a majority to the class "1", the F_1-score is a more representative metric as the results are not biased. The results demonstrated the robustness of both the Random Forest and XGBoost classifiers in predicting the harmonic contribution side. In summary, the Random Forest achieved an average F_1-Score of 99.05%, while XGBoost reached 99.41%.

TABLE 2. Best hyperparameters for Random Forest and XGBoost.

Classifier	Hyperparameter	Value
Random Forest	*Number of DTs*	12
	Maximum depth	15
	Min. number of samples in split nodes	7
	Min. number of samples in a leaf node	4
XGBoost	*Number of boosting rounds*	150
	Maximum depth	9
	Learning rate	0.1000
	Fraction of samples to adjust DTs	0.6967
	Fraction of features for each DT	0.3815
	Mininum loss reduction	0.1967
	L1 regularization	0.1909
	L2 regularization	0.1331

TABLE 3. Random forest performance by harmonic source.

Harmonic Source	Accuracy	Precision	Recall	F_1-score
6-pulse	99.38%	98.78%	99.51%	**99.15%**
12-pulse	99.20%	98.54%	99.26%	**98.90%**
SFC	99.11%	99.01%	98.53%	**98.77%**
DC Motor	99.56%	99.51%	99.26%	**99.39%**
TCR	99.29%	98.54%	99.51%	**99.02%**
Average	**99.31%**	**98.88%**	**99.22%**	**99.05%**

TABLE 4. XGBoost performance by harmonic source.

Harmonic Source	Accuracy	Precision	Recall	F_1-score
6-pulse	99.73%	99.51%	99.75%	**99.63%**
12-pulse	99.47%	98.79%	99.75%	**99.27%**
SFC	99.47%	99.02%	99.51%	**99.27%**
DC Motor	99.64%	99.75%	99.26%	**99.51%**
TCR	99.47%	99.02%	99.51%	**99.27%**
Average	**99.57%**	**99.27%**	**99.56%**	**99.41%**

In this approach, XGBoost performed slightly better. This difference can be attributed to the boosting process, which refines errors from previous iterations, capturing more complex patterns.

In addition to the previously presented performance metrics, confusion matrices were generated for each harmonic source and classifier, as shown in Fig. 5 and 6, which display the results for Random Forest and XGBoost, respectively.

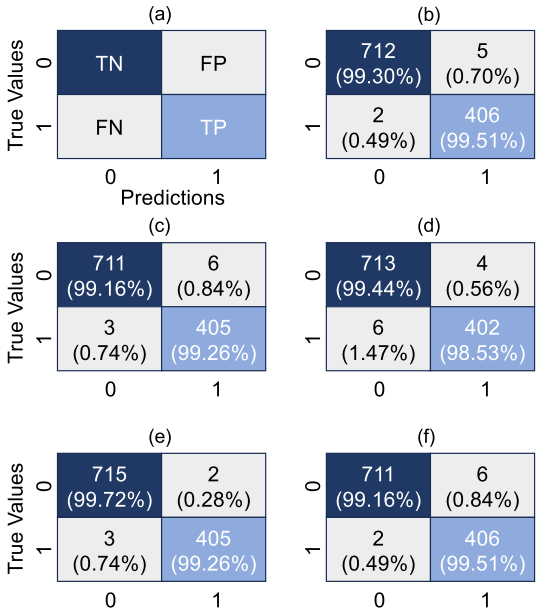

FIGURE 5. Confusion matrices obtained by Random Forest: (a) interpretation of a generic confusion matrix; (b) 6-pulse rectifier; (c) 12-pulse rectifier; (d) SFC; (e) DC motor drive; and (f) TCR.

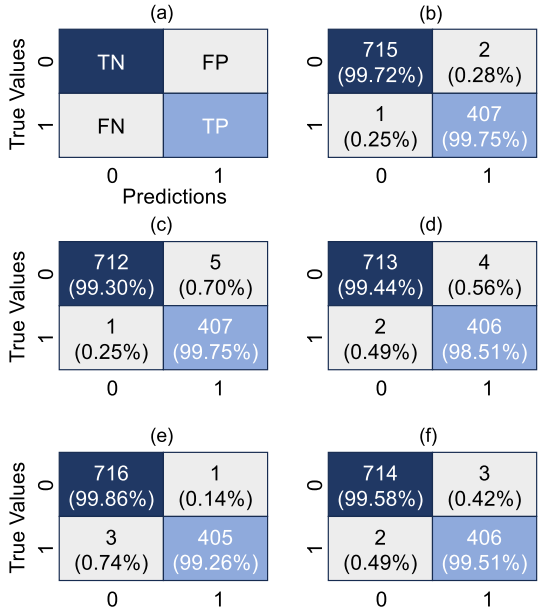

FIGURE 6. Confusion matrices obtained by XGBoost: (a) interpretation of a generic confusion matrix; (b) 6-pulse rectifier; (c) 12-pulse rectifier; (d) SFC; (e) DC motor drive; and (f) TCR.

The confusion matrix is a valuable tool for evaluating classifier's performance, presenting the values of *TP*, *TN*, *FP*, and *FN*. Analysis of both figures reveals a high rate of *TP* and *TN*, indicating the models' ability to correctly identify when the harmonic source is located on the utility- or MG-side, respectively.

B. HARMONIC SOURCE DISTANCE ANALYSIS

Given the overall results, we expanded our analysis to relate the classification performance (F_1-score) to the distance between the harmonic source and the MG's PCC, as depicted in Fig. 7.

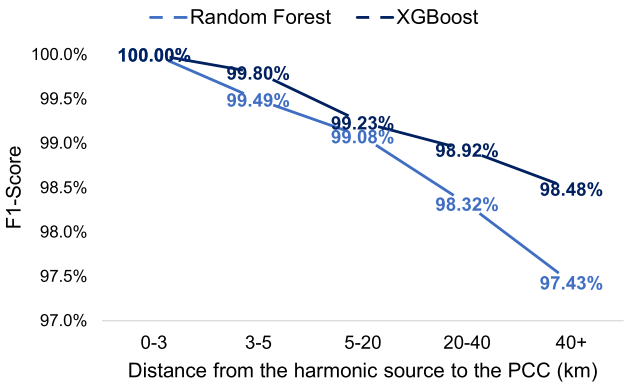

FIGURE 7. Classifiers' average performances by distance between the harmonic source and the MG's PCC.

The analysis revealed that both the Random Forest and XGBoost models experienced a decrease in F_1-score as the distance between the harmonic source and the MG's PCC increased. These results are expected since the propagation of harmonic distortion depends on the equivalent impedance seen by the PCC, as presented in [28].

C. MULTIPLE HARMONIC SOURCES ANALYSIS

The pre-trained Random Forest and XGBoost models were evaluated under scenarios of multiple harmonic sources (being one on the utility-side and another on the MG-side) in order to verify their generalization.

A total of 675 simulations involving multiple harmonic sources were performed under three operating conditions, that are: energy import mode; energy export mode; and neutral mode. The harmonic sources were the same presented in Table 1.

Due to the presence of multiple harmonic sources, it was assumed that the harmonic contribution side is the one on which the harmonic source with the highest apparent power is located.

The performance of the proposed models in scenarios with multiple harmonic sources is presented in Table 5, along with the corresponding confusion matrices in Fig. 8. Both models maintained good results, with F_1-scores exceeding 93%.

TABLE 5. Random forest and XGBoost performance with multiple harmonic sources.

Classifier	Accuracy	Precision	Recall	F_1-score
Random Forest	92.44%	96.35%	91.67%	**93.95%**
XGBoost	94.22%	97.81%	93.06%	**95.37%**

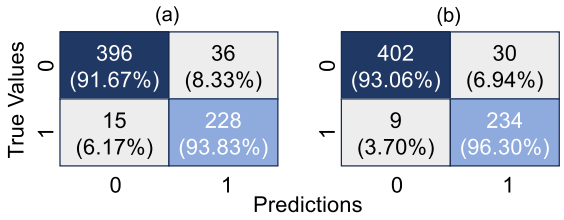

FIGURE 8. Confusion matrices obtained by (a) Random Forest and (b) XGBoost.

These outcomes indicate that the Random Forest and XGBoost models exhibited good adaptability, even though they were trained exclusively on single harmonic source scenarios. This reinforces the ability of ensemble-based models to generalize and perform well under varying system conditions.

Although the pre-trained models have proven to be robust for classifying scenarios with multiple harmonic sources, some field applications may require retraining. However, an alternative that would require less data and computational burden would be a transfer learning strategy through fine-tuning.

D. COMPARATIVE ANALYSIS WITH THE STATE-OF-THE-ART

In the literature, several studies have proposed approaches to identify the harmonic contribution at the consumers' PCC. In this context, the present paper introduced a machine learning-based approach, employing Random Forest and XGBoost classifiers to address this challenge with high efficiency and scalability.

Traditional approaches, such as those based on the active power direction method [9], [10], [11], [13], are appealing due to their simplicity, real-time capabilities, and extensive practical validation. However, they rely on accurate knowledge of the harmonic impedances of both the utility and the consumer/microgrid sides, and their performance degrades when multiple harmonic sources are present. Our approach does not require *a priori* knowledge regarding the harmonic impedance to accurately identify the contribution side, even in the presence of multiple harmonic sources.

Compared to the ANN-based approaches, such as those proposed by [14], [15], and [16], our ensemble-based classifiers offers lower computational complexity, faster training times, and improved accuracy, even in the presence of data variability. While ANNs can reach good results, they are generally more sensitive to noise and require extensive parameter tuning and frequent retraining as grid conditions evolve, limiting their applicability in dynamic scenarios. Additionally, these approaches commonly require high computational cost, making them difficult to be used in real-time, mainly to make decisions embedded in hardware. Contrarily, our approach employs a simple and robust binary classification strategy, enabling faster inference, reduced implementation complexity for embedded hardware, and lower computational burden, making it practical for deployment. In addition, it was evaluated under a wide range of simulated conditions, demonstrating greater robustness and adaptability to various scenarios and types of harmonic sources.

In terms of machine learning, there are still approaches based on Fuzzy inference systems [17], [18], which are very limited by the number of input features. It is also worth mentioning that adjustments to membership functions and decision rules must be made based on the knowledge of an expert in the problem domain. However, it is not guaranteed that there will be a power quality engineering at every utility with the necessary experience to perform such an adjustment. On the other hand, as our approach was based on ensembles of decision trees, it guarantees the interpretability, resulting in decision rules that do not require adjustments. The decision rules are automatically determined during the training process, without the need for an expert in the problem domain.

Overall, the proposed approach combines high accuracy with low computational burden, offering a practical and scalable solution for a variety of operational scenarios. Its balance between performance, simplicity, and adaptability makes it especially suitable for smart distribution grids.

V. CONCLUSION

This paper presented an ensemble-based approach, comparing the behavior of Random Forest and XGBoost to identify the side with the predominant harmonic contribution observed at the MG's PCC. Simulations under different operating conditions were conducted to validate the proposed approach. In this sense, the approach was validated and achieved high performances, measured in terms of accuracy, precision, recall and F_1-score. This way, both classifiers demonstrated their robustness, as the average F_1-scores were about 99.05% and 99.41% for Random Forest and XGBoost, respectively. Given these results and the diversity of simulated scenarios, the proposed approach represents a feasible and alternative solution for identifying the side with the predominant harmonic contribution.

As future work, the authors intend to focus on estimating the magnitude and phase angles of the harmonic sources together with the contribution side. In addition, it is important to evaluate the presence of multiple harmonic sources simultaneously in more depth.

ACKNOWLEDGMENT

The authors would like to thank the University of São Paulo and the Federal University of São Carlos for the facilities provided.

REFERENCES

[1] S. Elphick, V. Gosbell, V. Smith, S. Perera, P. Ciufo, and G. Drury, "Methods for harmonic analysis and reporting in future grid applications," *IEEE Trans. Power Del.*, vol. 32, no. 2, pp. 989–995, Apr. 2017.

[2] A. Taghvaie, T. Warnakulasuriya, D. Kumar, F. Zare, R. Sharma, and D. M. Vilathgamuwa, "A comprehensive review of harmonic issues and estimation techniques in power system networks based on traditional and artificial intelligence/machine learning," *IEEE Access*, vol. 11, pp. 31417–31442, 2023.

[3] R. A. S. Fernandes, M. Oleskovicz, and I. N. da Silva, "Harmonic source location and identification in radial distribution feeders: An approach based on particle swarm optimization algorithm," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3171–3179, May 2022.
[4] C. Wang, F. Xu, Q. Zhou, H. Zheng, Q. Shu, L. Fan, H. Li, and Z. Liu, "Estimating the harmonic impacts of multiple harmonic-producing loads considering network impedance variations," *IEEE Trans. Power Del.*, vol. 38, no. 6, pp. 3950–3958, Dec. 2023.
[5] M. H. Saeed, W. Fangzong, B. A. Kalwar, and S. Iqbal, "A review on microgrids' challenges & perspectives," *IEEE Access*, vol. 9, pp. 166502–166517, 2021.
[6] A. A. Alkahtani, S. T. Y. Alfalahi, A. A. Athamneh, A. Q. Al-Shetwi, M. B. Mansor, M. A. Hannan, and V. G. Agelidis, "Power quality in microgrids including supraharmonics: Issues, standards, and mitigations," *IEEE Access*, vol. 8, pp. 127104–127122, 2020.
[7] *IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems*, IEEE Standard 519-2014, 2014.
[8] *Electromagnetic Compatibility (EMC)-part 3-6: Limits - Assessment of Emission Limits for the Connection of Distorting Installations To Mv, Hv and Ehv Power Systems*, Standard IEC/TR 61000-3-6, 2008.
[9] F. Xu, H. Yang, J. Zhao, Z. Wang, and Y. Liu, "Study on constraints for harmonic source determination using active power direction," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 2683–2692, Dec. 2018.
[10] W. Xu and Y. Liu, "A method for determining customer and utility harmonic contributions at the point of common coupling," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 804–811, Apr. 2000.
[11] W. Xu, E. E. Ahmed, X. Zhang, and X. Liu, "Measurement of network harmonic impedances: Practical implementation issues and their solutions," *IEEE Trans. Power Del.*, vol. 17, no. 1, pp. 210–216, Jan. 2002.
[12] C. Li, W. Xu, and T. Tayjasanant, "A' critical impedance'-based method for identifying harmonic sources," *IEEE Trans. Power Del.*, vol. 19, no. 2, pp. 671–678, Feb. 2004.
[13] B. Wang, G. Ma, J. Xiong, H. Zhang, L. Zhang, and Z. Li, "Several sufficient conditions for harmonic source identification in power systems," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 3105–3113, Dec. 2018.
[14] J. Mazumdar, R. G. Harley, F. Lambert, G. K. Venayagamoorthy, and M. L. Page, "Intelligent tool for determining the true harmonic current contribution of a customer in a power distribution network," *IEEE Trans. Ind. Appl.*, vol. 2, no. 5, pp. 664–671, Oct. 2006.
[15] C.-H. Huang and C.-H. Lin, "Multiple harmonic-source classification using a self-organization feature map network with voltage–current wavelet transformation patterns," *Appl. Math. Model.*, vol. 39, no. 19, pp. 5849–5861, Oct. 2015.
[16] A. S. S. Murugan and V. S. Kumar, "Determining true harmonic contributions of sources using neural network," *Neurocomputing*, vol. 173, pp. 72–80, Jan. 2016.
[17] C. B. S. Silva, I. N. D. Silva, J. V. P. Aravechia, and R. A. S. Fernandes, "A fuzzy-based approach for harmonic contribution determination at points of common coupling," *IEEE Eindhoven PowerTech*, vol. 15, no. 6, pp. 1–6, Jun. 2015.
[18] R. A. S. Fernandes, D. Barbosa, A. N. Montagnoli, and M. Suetake, "Determining the responsibility sharing of harmonic distortion: An approach based on decision trees and neural-fuzzy systems," in *Proc. IEEE PES Innov. Smart Grid Technol. Latin Amer. (ISGT-LA)*, Nov. 2023, pp. 225–229.
[19] H. Sha, F. Mei, C. Zhang, Y. Pan, J. Zheng, and T. Li, "Multi-harmonic sources harmonic contribution determination based on data filtering and cluster analysis," *IEEE Access*, vol. 7, pp. 85276–85285, 2019.
[20] L. Prikler and H. K. Høidalen, "ATPDRAW version 5.6 for windows 9x/NT/2000/XP/vista users' manual," Norwegian Univ. Sci. Technol., Trondheim, Norway, Tech. Rep. 1.0, 2009.
[21] K. P. Schneider, B. A. Mather, B. C. Pal, C.-W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.
[22] W.-M. Lin, C.-H. Lin, K.-P. Tu, and C.-H. Wu, "Multiple harmonic source detection and equipment identification with cascade correlation network," *IEEE Trans. Power Del.*, vol. 20, no. 3, pp. 2166–2173, Jul. 2005.
[23] C.-H. Lin and C.-H. Wang, "Adaptive wavelet networks for power-quality detection and discrimination in a power system," *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1106–1113, Jul. 2006.
[24] J. Liu, H. Song, H. Sun, and H. Zhao, "High-precision identification of power quality disturbances under strong noise environment based on FastICA and random forest," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 377–387, Jan. 2021.
[25] R. Wardoyo, A. Musdholifah, G. Angga Pradipta, and I. N. Hariyasa Sanjaya, "Weighted majority voting by statistical performance analysis on ensemble multiclassifier," in *Proc. 5th Int. Conf. Informat. Comput. (ICIC)*, Nov. 2020, pp. 1–8.
[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[27] D. L. Olson, D. Delen, D. L. Olson, and D. Delen, "Performance evaluation for predictive modeling," in *Advanced Data Mining Techniques*. Berlin, Germany: Springer, 2008, pp. 137–147.
[28] A. Baggini and Z. Hanzelka, *Voltage Current Harmon*. Hoboken, NJ, USA: Wiley, 2008, ch. 7, pp. 187–261.

MATHEUS C. MELLA received the B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of São Carlos, São Carlos, Brazil, in 2022 and 2025, respectively. He provides machine learning and data science services to companies. His research interests include signal processing, machine learning, and smart grids.

RICARDO A. S. FERNANDES (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Educational Foundation of Barretos, Barretos, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, São Carlos, Brazil, in 2009 and 2011, respectively. From 2015 to 2017, he was a Visiting Professor with the Polytechnic Institute of Porto. He is currently an Assistant Professor with the University of São Paulo. His research interests include embedded systems, signal processing, machine learning, and smart grids.

DANIEL BARBOSA (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Federal University of Bahia, in 2005, and the master's and Ph.D. degrees in electrical engineering from São Carlos School of Engineering, University of São Paulo. He is currently a Professor with the Federal University of Bahia. His research interests include distributed generation, power quality, and power system protection.