HARMONIC SOURCES IDENTIFICATION ON THE POWER GRID WITH SOLAR POWER PENETRATION USING MACHINE LEARNING BASED ON FEATURE EXTRACTION

Cuong Chi Dang, Viet Quoc Huynh*, Khai Phuc Nguyen

Ho Chi Minh City University of Technology

Vietnam National University Ho Chi Minh City, Viet Nam

*Email: *hqviet@hcmut.edu.vn*

Received: 24 January 2025; Revised: 17 February 2025; Accepted: 24 February 2025

ABSTRACT

This paper aims to identify and analyze the sources of harmonic distortion in power grids with solar power penetration. The increasing integration of solar power into the power grids presents challenges in maintaining power quality and ensuring system stability. Among these challenges, harmonic distortion caused by the nonlinear operation of photovoltaic (PV) systems, particularly inverters, has become a critical issue. A comprehensive methodology is proposed, using single-point measurement methods combined with machine learning based on feature extraction to accurately locate and quantify harmonic sources at PCC (point of common coupling). The features of the 3-phase current and 3-phase voltage waveform at PCC are saved in image and numerical data by MATLAB/Simulink data generation model. Applying a machine learning (ML) model as a Random Forest Classification (RFC) model will extract the features based on numerical data of waveforms. With current and voltage input at PCC under the single-point method, the RFC model will identify what these waveforms are the same as the waveform of the datasets. From that, classify the characteristics of current and voltage waveform at PCC with the direction, phase angle, amplitude, and order of the harmonics, as well as the voltage amplitude and phase angle of the power grid and solar power source. The results help identify the responsibilities of the involved parties and enhance the effectiveness of harmonic quality management.

Keywords: Harmonics, solar power penetration, power quality, PCC, machine learning, feature extraction.

1. INTRODUCTION

Identifying harmonic sources in distribution grids with high penetration of distributed energy resources is crucial for maintaining power quality and ensuring the stable operation of power systems (PS). As "Net zero" becomes a global priority in energy development, Vietnam has implemented policies such as Power Development Plan 8 [1] and Decree 135/2024/ND-CP [2]. While the growth of renewable energy presents significant opportunities, it also introduces challenges, particularly regarding power quality. Harmonics in the grid can increase power losses, reduce efficiency and equipment lifespan, trigger resonance, overload components, and impair the operation of protection and measurement devices. These issues directly impact the reliability, quality, and stability of the power system [3, 4].

Figure 1. Point of common coupling (PCC)

To limit harmonics generated by distributed power sources/PV sources/loads at PCCs (shown in Figure 1), international standards are established, of which the most common are IEEE 519:2014 (harmonic level limits for PS), and IEC 61000-3-6:2008. In Vietnam, there are also standards such as TCVN 7909-3-6:2020 (equivalent to IEC 61000-3-6:2008), Circular 39/2015/TT-BCT for the distribution grid, Circular 30/2019/TT-BCT [5-8].

Identifying harmonic sources at PCC is an important issue for power companies and customers. Since the 1990s, many methods have been proposed [9-11], divided into two main groups: multipoint measurement and single-point measurement [12]. Multipoint measurement methods provide accurate information on harmonic states but are difficult to implement and expensive due to complex equipment requirements. In medium and low-voltage grids, single-point measurement is more popular despite difficulties in accurately determining grid impedance. Single-point measurement methods are divided into three types, based on active power flow direction (APFD), reactive power flow (RPF), and current-voltage ratio (C-VR) [13]. Some prominent methods include: the power direction method, based on the phase angle difference between current and voltage but the accuracy is not high [14]; the harmonic current vector (HVM) method, which determines the proportion of supplier and customer contributions based on impedance, is easy to implement but requires system information [15, 16]; and the RLC method, an improvement of HVM, which models customer impedance as a parallel RLC circuit, requiring only current and voltage measurements at the PCC to give approximate results [16]. However, these methods have limitations and have not been approved by international organizations such as IEEE, CIGRE, or IEC to fully meet practical and commercial requirements [13].

In Vietnam, most research projects focus on reducing harmonics caused by nonlinear loads and solar power sources. Therefore, research to determine harmonic sources especially when the solar power penetration on the power grid in Vietnam needs to be considered and studied comprehensively, in the context of Decree 135/2024/ND-CP was approved.

Consequently, identifying harmonics sources using a single-point measurement method and machine learning in the direction of extracting waveform features at PCC is proposed. The numerical data of voltage and current waveforms generated from the MATLAB/Simulink model are used as training datasets for the RFC model to identify the source of harmonics. The characteristics of current and voltage at PCC with possible cases will be saved in two forms: numerical data of the waveform and its images. In addition, this method can be used as a sample model for model development with other voltage levels in distribution grids. The accuracy of this prediction model depends on the amount of data generated and trained. The results achieved in this paper will contribute to determining the source of harmonics at PCC.

2. METHODOLOGY

2.1. Harmonics

According to [5, 14, 15], harmonics in PS are sinusoidal periodic waves with frequencies that are integer multiples of the fundamental frequency of the system and are classified according to criteria such as: by harmonic order including even order (even multiple of the fundamental frequency) and odd order (odd multiple of the fundamental frequency); by harmonic direction from the grid, from the customer source or both directions to PCC; by the level of influence including low-order harmonics (under 15^{th}) and high-order harmonics (above 15^{th}); by origin including linear and nonlinear harmonics; by waveform and impact including continuous and discontinuous harmonics.

Based on international standards, harmonics are measured based on Total Demand Distortion – TDD and Total Harmonic Distortion for voltage - THD_v, according to formula 1 and formula 2 respectively [5, 6]. Additionally, to determine the origin and impact of each individual harmonic component, the individual harmonic distortion coefficient - IHD can be used. TDD is defined as the ratio of the total harmonic distortion to the maximum value of the fundamental load current, following as

$$TDD = \frac{\sqrt{I_2^2+I_3^2+I_4^2+\cdots+I_n^2}}{I_{Lmax}} \times 100\% \quad (1)$$

Moreover, THD_v is the index of the synthesis of harmonic components of the voltage compared to the fundamental frequency component, reflecting the level of distortion of the voltage caused by harmonics, which is calculated as

$$THD_v = \frac{\sqrt{V_2^2+V_3^2+V_4^2+\cdots+V_n^2}}{V_1} \times 100\% \quad (2)$$

In formulas 1 and 2, I_{Lmax} is the maximum demand load current (fundamental frequency component) at PCC; $I_2, I_3, \ldots I_n$ are currents of harmonic component; V_1 is the voltage of the fundamental frequency component; $V_2, V_3, \ldots V_n$ are voltages of harmonic components.

2.2. Recommendations for harmonic limits

According to IEEE 519:2014 [5], the recommendations for harmonic limits include: THD_v does not exceed 8% for voltage levels under $1kV$, and 5% for voltage levels between $1kV$ and $69kV$) in the distribution system. TDD depends on the short-circuit power level at PCC, individual current of harmonic component, and voltage level, usually not exceeding 20% in many cases.

In Vietnam, under the conditions of the nominal voltage limit of the distribution grid (110 kV, 35 kV, 22 kV, 15 kV, 10 kV, 6 kV, and 0.38 kV), the allowable operating voltage deviation is ± 5% at the connection point with the electricity customer [8]. According to regulations [8], the total harmonic voltage distortion THD_v should not exceed 8%, 5%, and 3% for low voltage, medium voltage, and voltages greater than or equal to 110 kV, respectively. For current harmonics, there is a distinction in the requirements between power plant elements and electrical loads. TDD of current harmonics is stricter for power plants than for electrical loads. For example, for medium voltage, the TDD requirement for electrical loads is 8%, while the requirement for power plants is 5%.

2.3. Data generation model

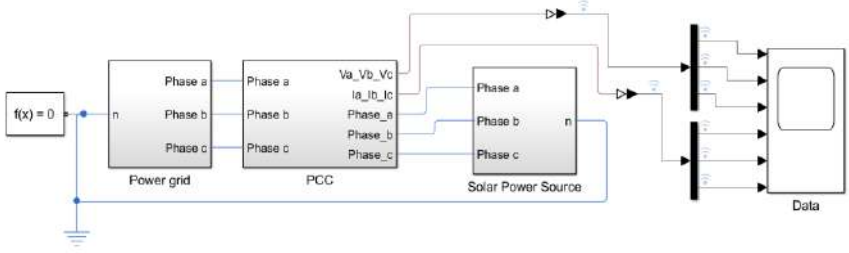


Figure 2. Model for generating voltage and current waveform data at PCC in MATLAB/Simulink

A model for determining harmonic origin by single-point measurement method based on waveform feature extraction from the machine learning model is proposed. This model includes a MATLAB/Simulink data generation model and a classification model. In Figure 2, the data generation model is used to generate voltage and current waveform data (both images and numerical data) at PCC in MATLAB/Simulink, which consists of a 3-phase grid source ($V_G = \pm 5\%\, pu, f = 50Hz$), a solar power source (AC side of the inverter, $V_{PV} = \pm 5\%\, pu, f = 50Hz$), and a 3-phase voltage and current measurement block at PCC. Voltage and current waveforms for phases a, b, and c (in the time series) are saved as a numerical dataset for a machine learning model to predict harmonic sources.

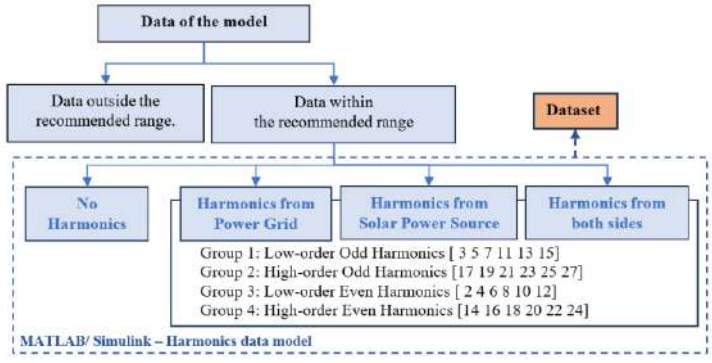


Figure 3. Harmonic waveform data structure from MATLAB/Simulink Model

The harmonic groups (shown in Figure 3) include: group of Low-order Odd Harmonics – LO ($3^{rd}, 5^{th}, 7^{th}, 11^{th}, 13^{th}, 15^{th}$), group of High-order Odd Harmonics – HO ($17^{th}, 19^{th}, 21^{st}, 23^{rd}, 25^{th}, 27^{th}$), group of Low-order Even Harmonics – LE

($2^{nd}, 4^{th}, 6^{th}, 8^{th}, 10^{th}, 12^{th}$), and group of High-order Even Harmonics – HE ($14^{th}, 16^{th}, 18^{th}, 20^{th}, 22^{th}, 24^{th}$).

The data generation model generates a dataset within the allowable limits: the operating voltage at each source is always maintained within ±5% of the nominal voltage [0.95-1.05] pu (per unit); the amplitude of each harmonic in the harmonic groups (HM_G- from the grid, and HM_{PV}- from the solar power source) is assumed to be 0, 0.01, 0.02 and 0.03 on each side; the phase angle of harmonic groups (HP_G- from the grid, and HP_{PV}- from the solar power source) is assumed to be 0, 90, and 180 degrees (individual harmonics in the groups have the same amplitude and phase angle).

Table 1. Harmonic source from both source sides (grid and solar power source)

Group of Harmonics	The voltage on the grid side: V_G (pu)	The voltage on the solar power source side: V_{PV} (pu)	Phase Angle of harmonics group (HP_G - HP_{PV})	The amplitude of the harmonic group (HM_G- HM_{PV})	The amount of data
LO	0.95	0.95	(0-0), (0-90), (90-0), (0-180), (180-0)	0.01, 0.02 or 0.03	2205
HO	0.97	0.97			2205
LE	0.99	0.99			2205
HE	1	1			2205
	1.01	1.01			
	1.03	1.03			
	1.05	1.05			

With the data generation model, there are 4 simulation scenarios with 12397 data. Case 1: No harmonic source with 49 data. Case 2: Harmonic source from the grid side with 1764 data. Case 3: Harmonic source from the solar power source of the customer side with 1764 data. Case 4: Harmonic source from both sides with a total of 8820 data, detailed in Table 1. Each data will have 2 images: a 3-phase voltage waveform and a 3-phase current waveform. Image and numerical data will be labeled according to the following name structure like the example: $(LE|HE|LO|HO)\_V_G\_V_{PV}\_HM_G\_HM_{PV}\_HP_G\_HP_{PV}$

2.4. Feature Extraction method

The characteristics of the 3-phase voltage and current waveforms are clearly described in the first cycles from the time of measurement. In the frequency domain, according to the Nyquist theorem, with 27^{th} harmonic order ($f_{27} = 27f_0 = 27 \times 50 = 1350\,Hz$), the sampling frequency is $f_s \geq 2f_{27} = 2700Hz$. When analyzing up to 27^{th} harmonic order, the data is saved with the first 10 cycles (each cycle 0.02s). The numerical data will be saved with 541 points in 10 cycles for each waveform in a time series.

Two extraction methods are applied to each classification machine learning model: the Convolutional Neural Network (CNN) to extract features on images, and the Random Forest Classification (RFC) model to extract features on numerical data.

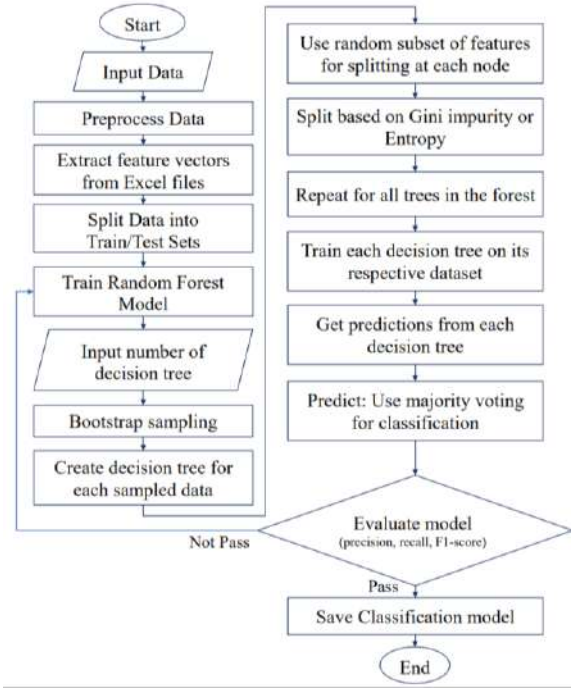


Figure 4. The flowchart of the Random Forest Classification model for feature extraction based on numerical data

The RFC [17] model for numerical data works by generating multiple decision trees from different data samples, its flowchart is shown in Figure 4. The data is split into training and testing sets. Each decision tree is trained using a random subset of the data and features, with splits based on Gini impurity or entropy. After training, each tree makes a prediction, and the final result is determined by majority voting. The model is evaluated using metrics like precision, recall, and F1-score [18] and then saved for future classification tasks. The objective of the classification model is to classify numerical data that exhibit characteristics similar to those of numerical data of waveforms in the training dataset, thereby drawing conclusions.

3. DATA GENERATION SCENARIO

3.1. No harmonic source

In the ideal case, there are no harmonic sources from either side, no voltage difference between the grid source and the solar power source ($V_G = V_{PV} = 1pu$), and no load as shown in Figure 5a. The 3-phase voltage waveforms have amplitudes equal to the nominal voltage, with a 120-degree phase shift between the three phases, and the current waveform is nullified.

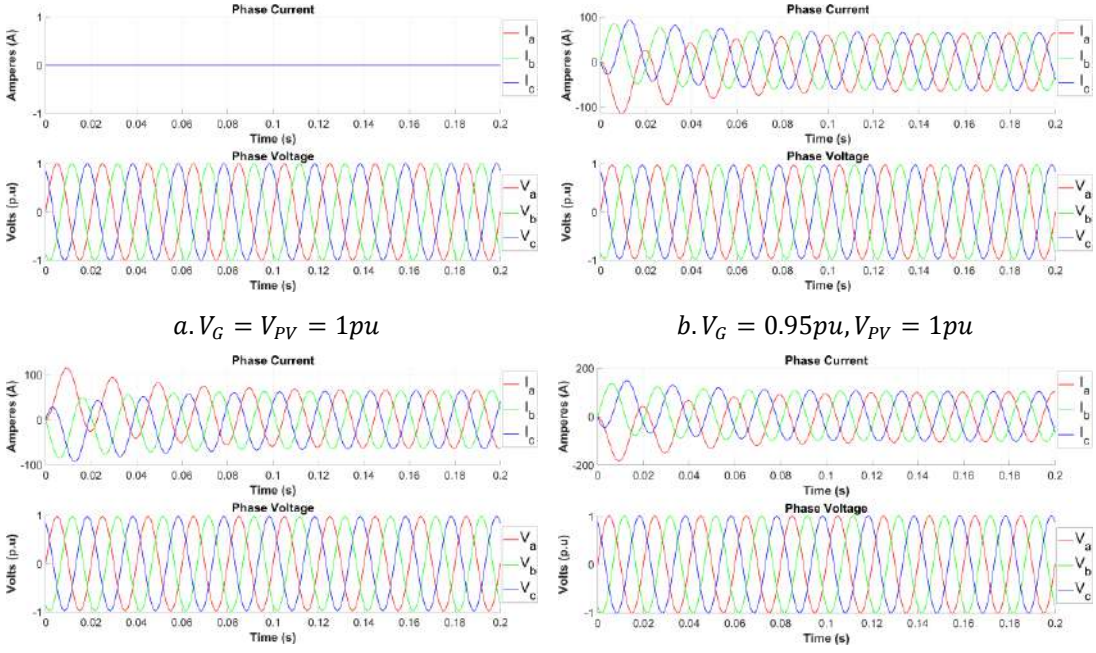


Figure 5. The current and voltage waveforms (3-phase) at PCC

The voltage waveform is always kept sinusoidal with amplitude within the allowable limit and the three-phase sequence is a-b-c phase in all cases. Moreover, the current waveform is sinusoidal, but its amplitude changes (Figure 5b, d) and phase sequence changes (Figure 5c, d) when there is a voltage difference between the two sides. Then, the phase sequence of the current is phase a-b-c when $V_G < V_{PV}$ (Figure 5b, d), and the phase sequence is phase c-a-b when $V_G > V_{PV}$ (Figure 5c).

3.2. Harmonic sources from the grid

In all cases of harmonic sources, the voltage waveform is still maintained sinusoidal with amplitude within the allowable limit and the three-phase sequence is a-b-c phase. Under the effect of the harmonic source (LE, LO, HE, or HO), the sinusoidal shape of the voltage is mainly affected at the peaks causing a multi-peak phenomenon but its amplitude is still within the allowable range, it can be also seen from Figure 6.

In the case of harmonic sources from the grid, when there is no voltage difference from both sides ($V_G = V_{PV} = 1pu$), the harmonic source (LE, LO, HE, or HO) affects the current waveform in terms of both

amplitude and phase sequence. When the harmonic amplitude increases while the phase angle remains constant relative to the system phase angle, the current amplitude increases without altering the phase sequence (Figure 6a, c). Conversely, when the harmonic amplitude remains constant but the phase angle changes, the current amplitude remains unchanged, whereas the phase sequence shifts (Figure 6c, e).

Additionally, when ($V_G = V_{PV} = 1pu, LE$), it can be also seen from Figure 6a, c, e that the 3-phase current waveform exhibits a sawtooth shape, with sharp peaks appearing at half-cycle and full-cycle positions. In addition, when the harmonic phase angle changes, the phase sequence of the 3-phase current is changed from phase a-b-c to phase c-a-b.

In the case of a harmonic source from the grid, when voltage deviations occur on both sides relative to the nominal voltage (V_G and V_{PV} compared to 1pu), harmonic groups (LE, LO, HE, HO) influence the current waveform in both amplitude and phase sequence. Meanwhile, the voltage waveform remains sinusoidal with small multi-peaks, and its phase sequence remains unchanged relative to the system (Figure 6b, d, f).

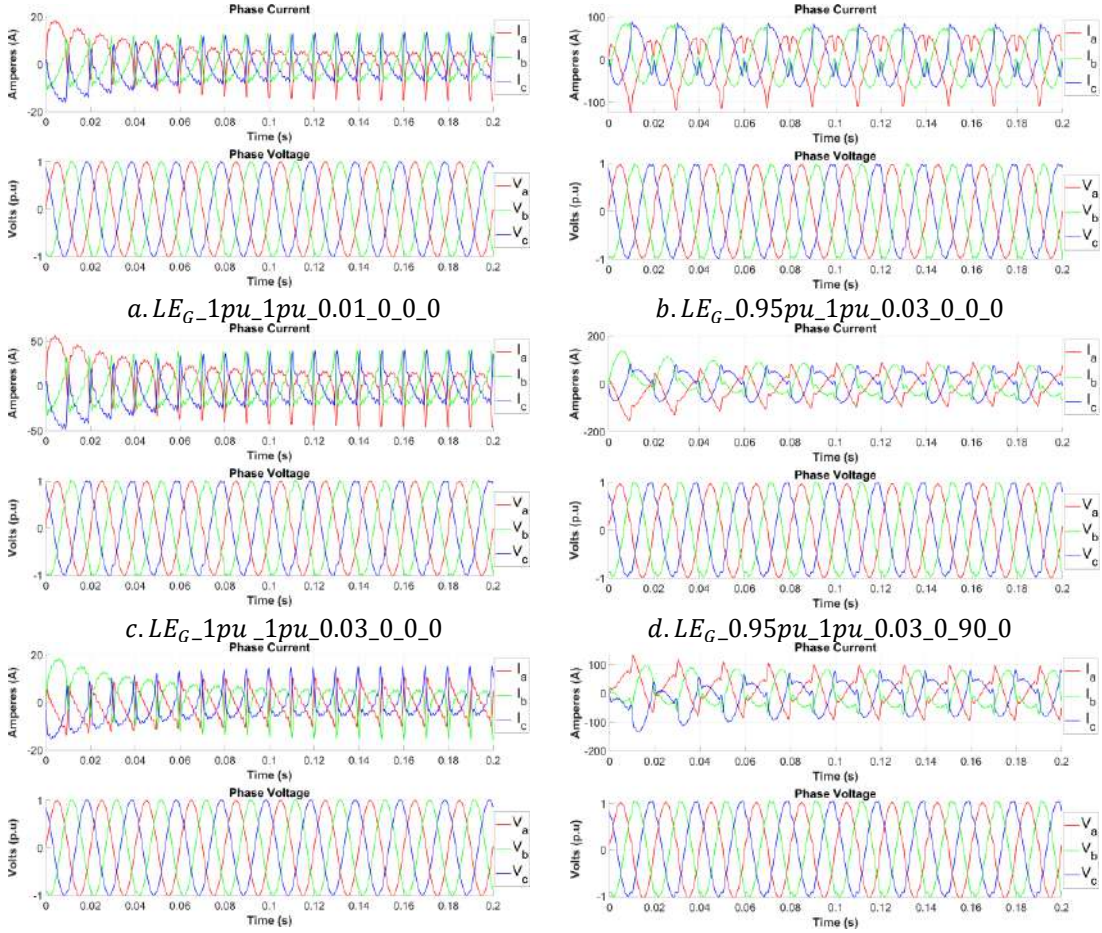


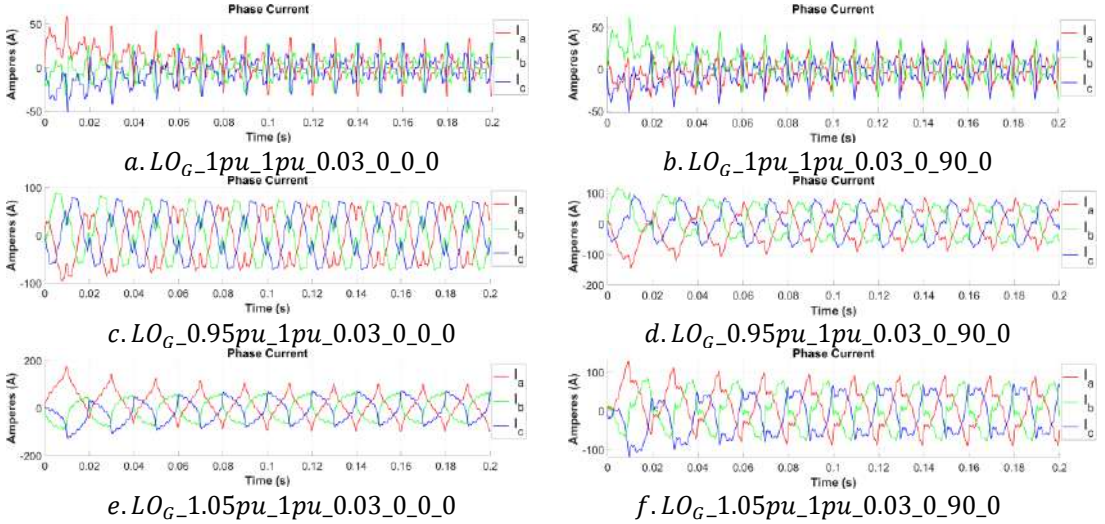
Figure 6. The current and voltage waveforms (3-phase) at PCC when LE from the grid

As the voltage deviates from the nominal value, the phase sequence gradually shifts: from phase a-b-c (if $V_G < V_{PV}$ like Figure 6b, d) to phase c-a-b (if $V_G > V_{PV}$ like Figure 6d, f), similar to the trend observed in the absence of harmonics. In Figure 6d, with $(V_G < 1pu, V_{PV} = 1pu)$, increasing the harmonic phase angle from the grid ($HP_G = 90$ instead of 0) results in a multi-peak current waveform where phase a becomes sharper, yet the current phase sequence remains unchanged.

For LO from the grid (Figure 7), when $V_G = V_{PV} = 1\,pu$, the current waveform exhibits sharp spikes at half-cycle and cycle intervals (Figure 7a, b), the larger the harmonic amplitude and the larger the current amplitude. Additionally, as the harmonic phase angle varies, the phase sequence of the

current changes (Figure 7a, b). However, when $V_{PV} = 1pu$ and $V_G < V_{PV}$ or $V_G > V_{PV}$, the change of the harmonic phase angle does not change the 3-phase current phase sequence, as shown in (Figure 7c, d) and (Figure 7e, f).

For HO, HE from the grid, it can also be seen from Figure 8a-e (HO) and Figure 8f-k (HE), the current waveform when $V_G < V_{PV}$ or $V_G > V_{PV}$, retains the sinusoidal shape, and there are disturbances on the peaks of the sine wave.

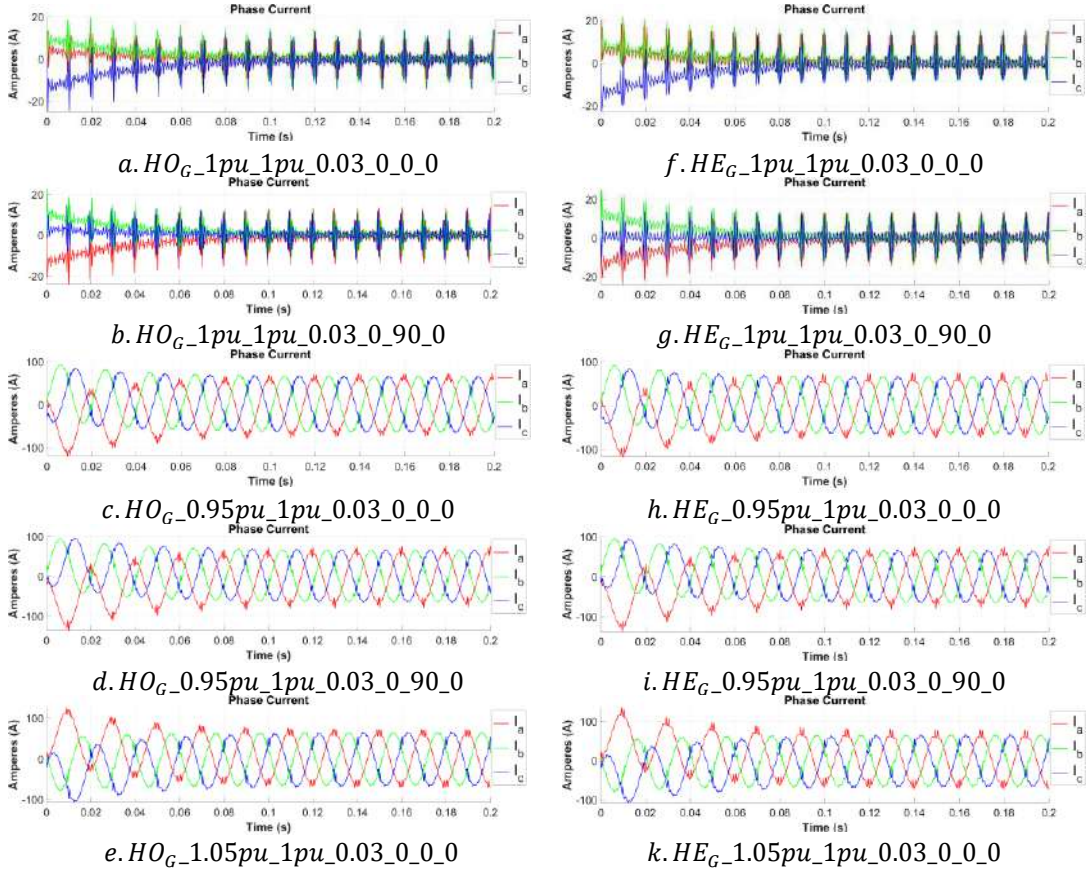


$a.LO_G\_1pu\_1pu\_0.03\_0\_0\_0$ $b.LO_G\_1pu\_1pu\_0.03\_0\_90\_0$

$c.LO_G\_0.95pu\_1pu\_0.03\_0\_0\_0$ $d.LO_G\_0.95pu\_1pu\_0.03\_0\_90\_0$

$e.LO_G\_1.05pu\_1pu\_0.03\_0\_0\_0$ $f.LO_G\_1.05pu\_1pu\_0.03\_0\_90\_0$

Figure 7. The 3-phase current waveforms at PCC when LO from the grid



$a.HO_G\_1pu\_1pu\_0.03\_0\_0\_0$ $f.HE_G\_1pu\_1pu\_0.03\_0\_0\_0$

$b.HO_G\_1pu\_1pu\_0.03\_0\_90\_0$ $g.HE_G\_1pu\_1pu\_0.03\_0\_90\_0$

$c.HO_G\_0.95pu\_1pu\_0.03\_0\_0\_0$ $h.HE_G\_0.95pu\_1pu\_0.03\_0\_0\_0$

$d.HO_G\_0.95pu\_1pu\_0.03\_0\_90\_0$ $i.HE_G\_0.95pu\_1pu\_0.03\_0\_90\_0$

$e.HO_G\_1.05pu\_1pu\_0.03\_0\_0\_0$ $k.HE_G\_1.05pu\_1pu\_0.03\_0\_0\_0$

Figure 8. The 3-phase current waveforms at PCC when HO and HE are from the grid

3.3. Harmonic sources from the solar power source of the customer

In the case of the harmonics group from the solar power source of the customer, the current amplitude increases with rising harmonic amplitude (Figure 9a, b and 9g, h). When $V_G = V_{PV} = 1pu$, changing the phase angle of harmonics (LO, LE, HO, HE) alters the current phase sequence relative to the system (Figure 9c, i). However, when $V_G > V_{PV}$ with $V_G = 1pu$ (Figure 9d, e, and 9k, l) or $V_G < V_{PV}$ (Figure 9e, m), phase angle variations affect only the waveform profile without changing the phase sequence.

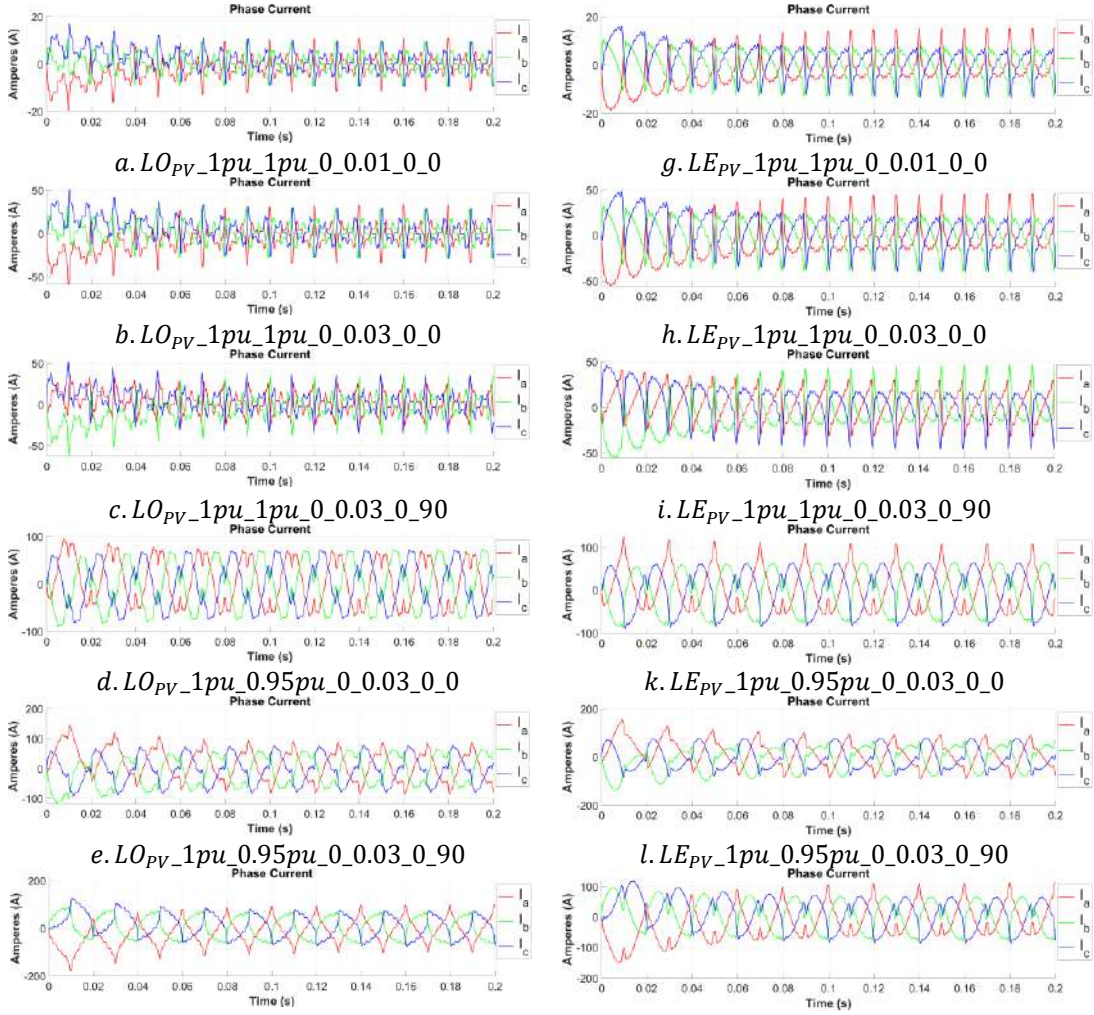


Figure 9. The 3-phase current waveforms at PCC when LO and LE from the solar power side

For HO, HE when $V_{\mathrm{G}} = 1pu$ and $V_{\mathrm{G}} < V_{\mathrm{PV}}$ or $V_{\mathrm{G}} > V_{\mathrm{PV}}$, having the following observations based on Figure 10c, d, e and 10h, i, k that: The 3-phase current waveform is sinusoidal and has disturbances at the peak amplitude, the phase sequence is c-a-b phase (if $V_{\mathrm{G}} > V_{\mathrm{PV}}$) and a-b-c phase (when $V_{\mathrm{G}} < V_{\mathrm{PV}}$).

For HE at $V_{\mathrm{G}} = V_{\mathrm{PV}} = 1pu$, an inconsistency between negative and positive amplitudes occurs (Figure 8f, 10f). If HE originates from the solar power source with a harmonic phase angle matching the system (Figure 10f), the 3-phase current shifts toward the negative edge. Conversely, if HE comes from the grid (Figure 8f), the amplitude shifts toward the positive edge. Additionally, changes in the phase angle of HE influence the 3-phase current amplitude (Figure 10f, g).

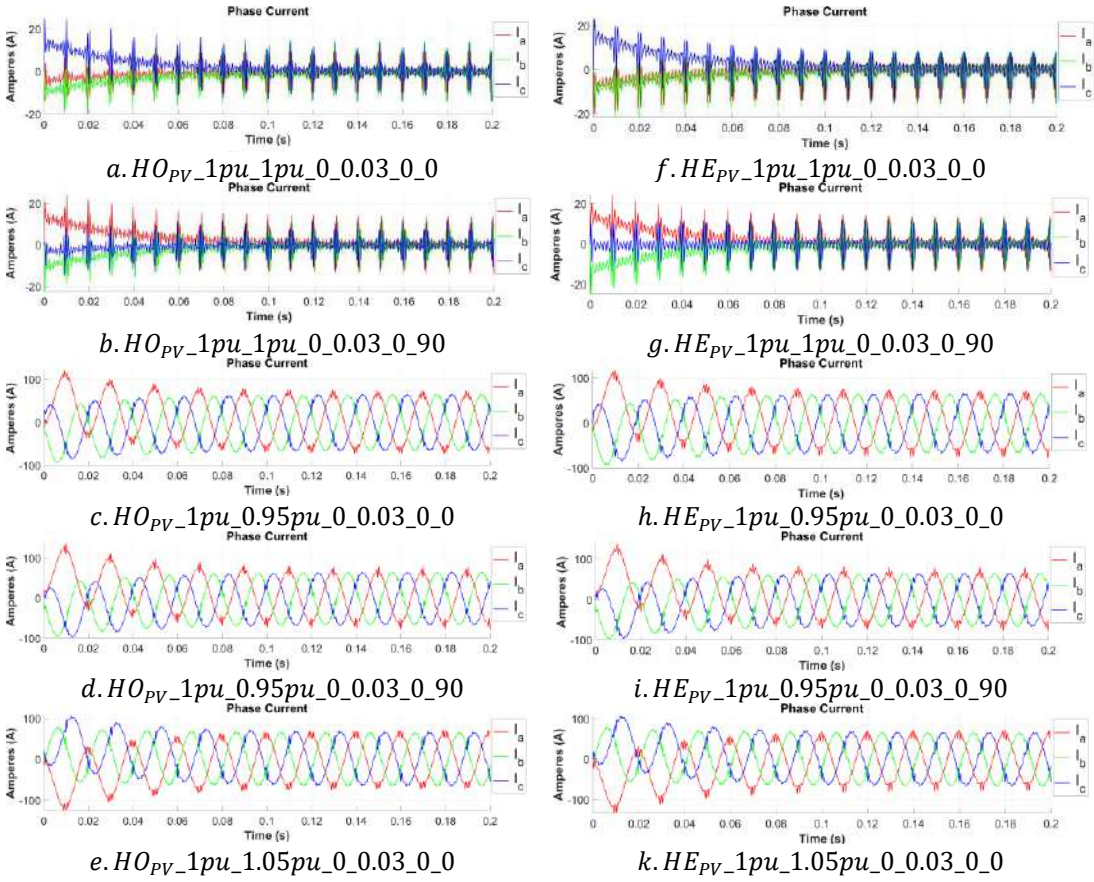
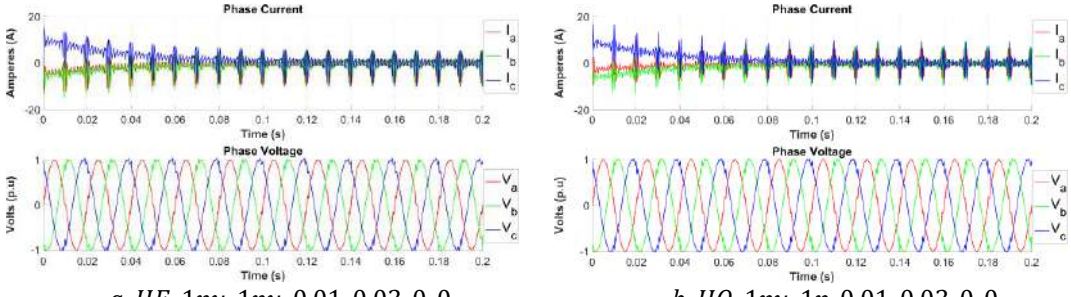


Figure 10. The 3-phase current waveforms at PCC when HO and HE from solar power source

3.4. Harmonic sources from both sides

In the case of harmonic sources from both the power grid and the solar power source, the influence of harmonics on the output 3-phase voltage and current waves is obvious, regardless of HO, HE, LO, and LE as shown in Figure 11a, b, c, d.

If there is no voltage difference between the grid source and the solar power source, the harmonics from both sides are the same in amplitude and their phase angle is equal to the phase angle of the system as shown in Figure 11e. In this case, the 3-phase voltage waveform appears multipeak but retains its sinusoidal shape. Similarly, when harmonics from both sources are identical, the 3-phase current waveform remains sinusoidal, though the voltage waveform still shows multi-peak characteristics (Figure 11f). However, when the harmonic phase angles differ, the current waveform undergoes noticeable changes (Figure 11e, g and 11f, h).



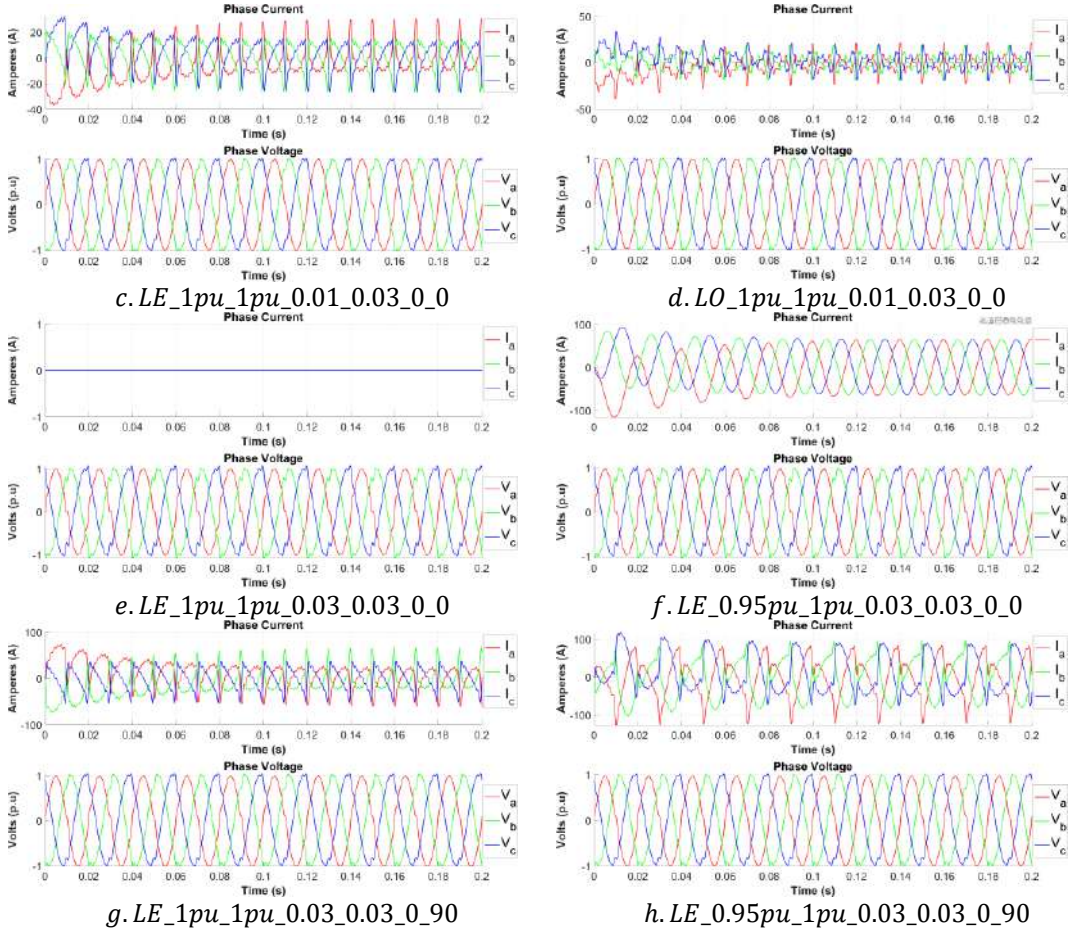


Figure 11. The 3-phase current and voltage waveforms at PCC when harmonic sources are from both sides

4. RESULTS

In conclusion, the shape of the 3-phase voltage and current waveforms at PCC will have clear effects when harmonics appear. The effects of harmonics at PCC are unique to each harmonic group, to each harmonic amplitude level, and to different harmonic phase angles. The difference in voltage levels from the two sources to PCC also contributes to the impact on the shape and phase sequence of the current. Each set of data on the 3-phase voltage and current waveforms contains characteristics, these characteristics are used to identify the source and the characteristics of the harmonics: harmonic type, amplitude, and phase angle.

The Random Forest Classification model for feature extraction based on numerical data was trained with 500 decision trees per sample, delivering exceptional performance in the classification task. The results indicate that the model achieved near-perfect accuracy, with a precision of 0.99879, a recall of 0.99879, and an F1-score of 0.99879, shown in Table 2. These metrics demonstrate not only the high predictive accuracy of the model but also its reliability and consistency in identifying and classifying data effectively.

The identification of harmonic sources using this RFC model will facilitate the accurate classification of voltage and current waveforms in each scenario at PCC when harmonic sources are present. Furthermore, in addition to identifying the source of harmonics, the model also provides insights into their amplitude characteristics, phase angle, order, and voltage status at both the grid source and the customer source up to PCC. The development of an application (as shown in Figure 12) or the integration of this model into measurement devices would be a feasible solution for identifying the sources of harmonics, ensuring power quality, and clearly determining the responsibilities of stakeholders within the power system.

Table 2. The classification report of the Random Forest Classification model

Dataset	Precision	Recall	F1-score
No harmonic sources	1	1	1
HE_G with $HP_G = 0, HP_{PV} = 0$	1	0.965517	0.982456
HE_G with $HP_G = 90$ or $180, HP_{PV} = 0$	1	1	1
HO_G with $HP_G = 0$ or 90 or $180, HP_{PV} = 0$	1	1	1
LO_G with $HP_G = 0$ or 90 or $180, HP_{PV} = 0$	1	1	1
LE_G with $HP_G = 0$ or 90 or $180, HP_{PV} = 0$	1	1	1
HE_{PV} with $HP_G = HP_{PV} = 0$	1	0.965517	0.982456
HE_{PV} with $HP_G = 0, HP_{PV} = 90$ or 180	1	1	1
HO_{PV} with $HP_G = 0$ and $HP_{PV} = 0$ or 90 or 180	1	1	1
LO_{PV} with $HP_G = 0$ and $HP_{PV} = 0$ or 90 or 180	1	1	1
LE_{PV} with $HP_G = 0$ and $HP_{PV} = 0$ or 90 or 180	1	1	1
Both sides HE with $HP_G = HP_{PV} = 0$	1	1	1
Both sides HE with $HP_G = 0, HP_{PV} = 180$	0.988636	1	0.994286
Both sides HE with $HP_G = 180, HP_{PV} = 0$	0.989796	0.989796	0.989796
Both sides HE with $HP_G = 0, HP_{PV} = 90$	1	1	1
Both sides HE with $HP_G = 90, HP_{PV} = 0$	1	1	1
Both sides HO with $HP_G = 180, HP_{PV} = 0$	0.988235	1	0.994083
Both sides LE with $HP_G = HP_{PV} = 0$	1	1	1
Both sides LO with $HP_G = HP_{PV} = 0$	1	1	1
The others	1	1	1
Accuracy	0.99879	0.99879	0.99879
Macro avg	0.999259	0.998241	0.998735

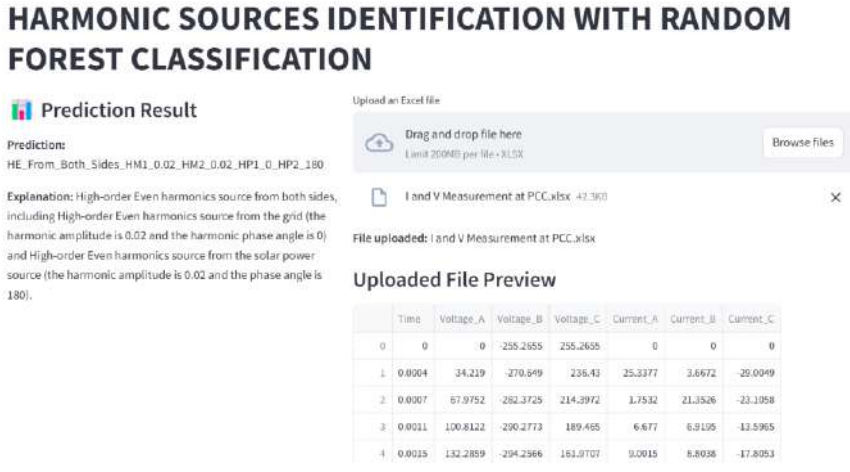


Figure 12. Harmonic sources identification with Random Forest Classification Application

5. DISCUSSION

The detection of harmonic sources in PS, particularly at PCC, remains a critical issue. According to [13], existing approaches can be categorized into three main groups: APFD, RPD, and C-VR. The first method determines the primary harmonic source based on the direction of active power flow at PCC: if the power value is positive, the customer is the main harmonic source; if negative, the power grid is the primary source. However, this method faces limitations when multiple harmonic sources

exist on both the grid and customer sides under different loading conditions. The second method relies on accurately determining the impedance of both the customer and the grid, but in practice, the impedance varies with operating conditions, making precise identification challenging. Similarly, the third method also depends on knowing the system impedance to accurately assess the contribution of each side to harmonics. Since no absolute reference point exists, accurately assigning responsibility for harmonic distortion is challenging, and determining the exact origin, amplitude, and type of harmonics remains difficult, complicating the implementation of appropriate mitigation measures. The findings in this study further highlight that harmonic order, magnitude, and phase angle significantly impact power quality at PCC under varying voltage conditions from both sides. In practice, the APFD method is more practical and cost-effective since it only requires voltage and current measurements at PCC without prior knowledge of the system impedance, though it has inherent accuracy limitations [19]. The method proposed in this paper also relies solely on measured voltage and current data at PCC, but its key advantage lies in its ability not only to determine harmonic direction but also to classify harmonic type, order, group, magnitude, and phase angle under different voltage conditions. However, for real-world applications, more detailed data is required, even though its feasibility has been demonstrated in laboratory conditions using an RFC model based on feature extraction. Notably, MATLAB/Simulink provides a powerful framework for generating highly detailed simulation data, enabling the accurate and comprehensive reproduction of real-world scenarios, thereby further reinforcing the applicability of this method.

6. CONCLUSION

The paper has indicated that harmonic groups (from the grid, from the solar power source, or both sides) impact on PCC mainly distort and destabilize the current waveform and current phase sequence, the voltage waveform remains sinusoidal and phase sequence in all cases but multi-peaks appear. Different profiles of the 3-phase current and voltage waveforms are unique features and are used for the training dataset in the RFC model.

This paper has also completed the goal of identifying the source of harmonics through the combination of single-point measurement methods, along with the application of machine learning to build a classification model. The numerical data generated from the MATLAB/Simulink model played an important role in the training process, improving the ability to identify the source of harmonics. The images are used as references for the shape of the current and voltage waveforms when harmonics affect PCC. The proposed method not only has the potential to be applied to different voltage levels in the distribution grid but also ensures high efficiency and accuracy, depending on the quantity and quality of the training data. The Random Forest Classification model produces results with high accuracy. This is a promising approach to developing technical solutions for power quality analysis and management, especially harmonics.

7. FUTURE WORK

Expand the scope of application of the model by applying it to real power grid systems with diverse configurations and voltage levels. To improve the prediction model, it is necessary to increase the amount of training data from real power grids. In addition, applying advanced deep learning algorithms or hybrid models will help improve the prediction performance and improve the accuracy of the model in handling complex data. Test the trained model and deploy it to hardware.

Acknowledgement: This research is funded by Ho Chi Minh City University of Technology – VNU-HCM under grant number To-ĐĐT-2023-05. We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology - VNU-HCM for this study.

REFERENCES

1. Prime Minister of Vietnam - Decision 500/QD-TTg on approving the National Electricity Development Planning of 2021 - 2030 and vision for 2050 (“PDP-8”), 2023.
2. Government of the Socialist Republic of Vietnam - Decree 135/2024/ND-CP on policies encouraging development of self-produced and self-consumed rooftop solar power, 2024.

3. Long T.D. - Reference on power quality. Bach Khoa Publishing House, Hanoi, 2013.
4. Das J.C. - Power system harmonics and passive filter designs. John Wiley & Sons, Hoboken, 2015.
5. IEEE Standards Association - IEEE Std 519-2014: IEEE recommended practice and requirements for harmonic control in electric power systems. IEEE 2014, 1-29.
6. IEC - IEC 61000-3-6:2008- Electromagnetic compatibility (EMC) - Part 3-6: Limits – Assessment of emission limits for the connection of distorting installations to MV, HV and EHV Power Systems, 2008.
7. Commission for the Standards, Metrology and Quality of Viet Nam - TCVN 7909-3-6:2020 (Equivalent to IEC 61000-3-6:2008)
8. Ministry of Industry and Trade of Vietnam – Circular 30/2019/TT-BCT: Amendments and supplements to certain provisions of Circular 25/2016/TT-BCT on transmission power system, and Circular 39/2015/TT-BCT on distribution power system.
9. Papic I., Matvoz D., Špelko A., …, Testa A. - A benchmark test system to evaluate methods of harmonic contribution determination. IEEE Transactions on Power Delivery **34** (1) (2019) 23-31. https://doi.org/10.1109/TPWRD.2018.2817542
10. Davis E.J., Emanuel A.E. and Pileggi D.J. - Evaluation of single-point measurements method for harmonic pollution cost allocation, IEEE Transactions on Power Delivery **15** (1) (2000) 14-18. https://doi.org/10.1109/61.847222
11. Ren Y., Liang J., Zhang S., Guo X. and Cheng Y. - Harmonic direction identification of new energy stations based on instantaneous reactive power theory and harmonic power, 2022 International Conference on Wireless Communications, Electrical Engineering and Automation (WCEEA), Indianapolis, USA (2022) 58-62. https://doi.org/10.1109/WCEEA56458.2022.00020
12. Bazina M. and Tomiša T. - Comparison of various methods for determining the direction of harmonic distortion by measuring in point of common coupling. 2014 IEEE International Energy Conference (ENERGYCON), Cavtat, Croatia (2014) 392-399. https://doi.org/10.1109/ENERGYCON.2014.6850457
13. Sinvula R., Abo-Al-Ez K.M. and Kahn M.T. - Harmonic source detection methods: A systematic literature review, IEEE Access **7** (2019) 74283-74299. https://doi.org/10.1109/ACCESS.2019.2921149
14. Bollen M.H.J. - Understanding power quality problems: Voltage sags and interruptions. New York, NY, USA: IEEE Press, 2000.
15. Arrillaga J., Smith B.C., Watson N.R., and Wood A.R. - Power system harmonic analysis. New York, NY, USA: Wiley, 1997.
16. Farhoodnea M., Mohamed A., Shareef H., Khan R.A. Jabbar - An improved method for determining contribution of utility and customer harmonic distortions in a power distribution system, International Journal on Electrical Engineering and Informatics **2** (3) (2010) 204-215. https://doi.org/10.15676/ijeei.2010.2.3.4
17. Salman H.A., Kalakech A., and Steiti A. - Random forest algorithm overview. Babylonian Journal of Machine Learning **2024** (2024) 69-79. https://doi.org/10.58496/BJML/2024/007
18. Gaye B., Zhang D., Wulamu A. - Sentiment classification for employees reviews using regression vector- stochastic gradient descent classifier (RV-SGDC). PeerJ Computer Science **7** (2021) e712. https://doi.org/10.7717/peerj-cs.712
19. Xu W. - Power direction method cannot be used for harmonic source detection, 2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134) **2** (2000) 873-876, Seattle, WA, USA. https://doi.org/10.1109/PESS.2000.867472.

TÓM TẮT

XÁC ĐỊNH NGUỒN SÓNG HÀI TRÊN LƯỚI ĐIỆN CÓ SỰ THÂM NHẬP CỦA ĐIỆN MẶT TRỜI SỬ DỤNG THUẬT TOÁN HỌC MÁY DỰA TRÊN TRÍCH XUẤT ĐẶC TRƯNG

Đặng Chí Cường, Huỳnh Quốc Việt*, Nguyễn Phúc Khải

Trường Đại học Bách khoa Thành phố Hồ Chí Minh,

Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam

*Email: *hqviet@hcmut.edu.vn*

Bài báo này nhằm mục đích xác định và phân tích nguồn gốc sóng hài gây méo dạng trong lưới điện có sự thâm nhập của điện mặt trời. Việc phát triển nhanh chóng của các nguồn điện mặt trời trên lưới điện đặt ra những thách thức trong việc duy trì chất lượng điện năng và đảm bảo tính ổn định của hệ thống điện. Trong số các thách thức thì sóng gây méo dạng do hoạt động của các thiết bị phi tuyến của hệ thống quang điện, đặc biệt là biến tần, đã trở thành một vấn đề cần lưu tâm. Một phương pháp xác định sóng hài được đề xuất, áp dụng phương pháp đo đơn điểm (single-point measurement) kết hợp với thuật toán học máy (Machine Learning - ML) dựa trên trích xuất đặc trưng (feature extraction) để xác định nguồn gốc và định lượng chính xác các nguồn sóng hài gây méo dạng tại điểm đấu nối chung (PCC) giữa lưới điện và nguồn điện mặt trời. Các đặc trưng của dạng sóng dòng điện ba pha và điện áp ba pha tại PCC được lưu lại với các hình ảnh và dữ liệu số của các dạng sóng bằng mô hình tạo dữ liệu trong MATLAB/Simulink. Áp dụng mô hình học máy như mô hình phân loại rừng ngẫu nhiên (Random Forest Classification - RFC) sẽ trích xuất các đặc trưng dựa trên dữ liệu số của các dạng sóng. Với dữ liệu đầu vào là dòng điện và điện áp của ba pha tại PCC theo phương pháp đo đơn điểm, mô hình RFC sẽ xác định những dạng sóng này giống với dạng sóng của bộ dữ liệu. Từ đó, phân loại được các đặc tính của hình dạng sóng dòng điện và điện áp tại PCC gồm nguồn gốc, góc pha, biên độ và bậc của sóng hài, cũng như biên độ và góc pha điện áp của lưới điện và của nguồn điện mặt trời. Kết quả từ nghiên cứu này sẽ giúp xác định trách nhiệm của các bên liên quan và nâng cao hiệu quả quản lý chất lượng điện năng về sóng hài.

Từ khóa: Sóng hài, sự thâm nhập điện mặt trời, chất lượng điện năng, điểm đấu nối điện chung, học máy, trích xuất đặc trưng.