Title
Pattern similarity-based phase identification problem formulation for medium voltage distribution networks

Journal
International Journal of Electrical Power & Energy Systems, 176

ISSN
0142-0615

Authors
Seong, Jungmin
Hwang, Jin Sol
Hussain, Shahid
et al.

Publication Date
2026-03-01

DOI
10.1016/j.ijepes.2026.111756

Peer reviewed

Electrical Power and Energy Systems 176 (2026) 111756

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Pattern similarity-based phase identification problem formulation for medium voltage distribution networks

Jungmin Seong [a,1], Jin Sol Hwang [b,1], Shahid Hussain [c], Miguel Heleno [d], Yun-Su Kim [b,*]

[a] Enel X, Seoul, 04520, Republic of Korea

[b] Department of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, 61005, Republic of Korea

[c] Department of Business & Accounting, Atlantic Technological University (ATU), Dublin Road, Galway, H91 T8NW, Ireland

[d] Lawrence Berkeley National Laboratory, Berkeley, 94729, CA, USA

ARTICLE INFO

Keywords:
DER curtailment
Empirical phase data
Modified Kirchhoff's Current Law
Permutation matrix
Phase identification

ABSTRACT

In this work, we propose an enhanced phase identification method to support distributed energy resources management in a medium voltage network. The problem is formulated with the objective function of minimizing the sum of the modified Kirchhoff's Current Law at the branch node and power-current pattern similarity. This combined objective function is then linearized by introducing auxiliary variables and subsequently resolved through Mixed-Integer Linear Programming. A permutation matrix is employed to establish a correlation between the magnitudes of electrical currents across distinct phases, achieved through manipulation based on empirical data. Subsequently, by comparing this matrix with the observed magnitudes of currents, the approach facilitates the deduction and identification of phase information inherent in the electrical signals. Consequently, the permutation matrix functions as decision variables to identify the phases of currents for wye-connected loads, delta-connected matrices for delta-connected loads, and percentage matrices for aggregated multi-phase loads. The proposed method is applied to a dataset obtained from real power consumption with one-hour resolution, and validation is conducted through several case studies considering scenarios with and without power pattern similarity on IEEE 13, IEEE 34, and IEEE 123 test feeders, followed by comparative analysis against the correlation coefficient and Euclidean based methods. This demonstrates the robustness and practicality of the proposed method by exhibiting superior accuracy compared to the state-of-the-art method across diverse environments.

1. Introduction

In the conventional unidirectional power flow paradigm, the primary role of Distributed System Operators (DSOs) was to efficiently deliver energy to consumers [1]. However, the advent of renewable energy technologies has prompted a shift towards bidirectional power flow configurations, incorporating Distributed Energy Resources (DERs). This transition has introduced several challenges, including network congestion, voltage instability, and impacts on overall grid operations [2]. In addition, the integration of DERs into the power system is constrained by the thermal capacities of transmission and distribution infrastructure [3]. These limitations necessitate the curtailment of certain DERs to maintain smooth system operation. The thermal constraints arise from physical factors such as transmission lines and distribution transformers, which have predefined maximum operating temperatures [4]. When the penetration of DERs exceeds these limits, equipment can overheat, leading to reliability issues. Consequently, DSOs must implement phase control strategies to mitigate network congestion and address potential economic repercussions [5]. The need to manage each phase is crucial for maintaining power flow equilibrium, regulating voltage levels, and ensuring grid stability amidst varying DER outputs and demand fluctuations [6]. To mitigate network congestion and ensure smooth power flow operations, DSOs incentivize DER owners to curtail some of their DER's power [7]. However, in cases where a DSO requests a DER owner to reduce power output, but the DER is not actually connected to the specified phase, the anticipated reduction in congestion will not materialize. Despite the lack of impact on congestion, DER owners are still entitled to compensation due to the interconnected nature of the three-phase power system, resulting in DSOs having to pay without receiving any corresponding benefit [8]. This issue is amplified in systems with

* Corresponding author.
E-mail addresses: jungmin.seong@enel.com (J. Seong), j.s.hwang@gm.gist.ac.kr (J.S. Hwang), shahid.hussain@atu.ie (S. Hussain), MiguelHeleno@lbl.gov (M. Heleno), yunsukim@gist.ac.kr (Y. Kim).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijepes.2026.111756
Received 26 June 2025; Received in revised form 2 February 2026; Accepted 5 March 2026
Available online 7 March 2026

bidirectional power flows, such as Vehicle-to-Grid (V2G) or Energy Storage Systems (ESS), which introduce additional complexities in accurately identifying and managing power flow dynamics across the grid [9]. In addition to those complexities, the practical distribution network data itself has its own inherent problem, which is that the field data does not provide phase classification among different line sections [10]. That is, phase 'a' in one line section does not always match with phase 'a' in the other line sections. The inability to identify phases causes poor network models, which makes distribution system operation more difficult.

To address these challenges and mitigate economic losses incurred by DSOs, numerous studies in the literature have explored methods for controlling individual phases (such as 3-phase unbalance mitigation and 3-phase power flow) [11]. Conventional approach for identifying the phase of single-phase customers in the Low Voltage (LV) network typically involves utilizing the Advanced Meter Infrastructure (AMI) [12]. Given the rapid expansion of AMI, various methods have been explored, which can be broadly categorized into two groups. (1) The first category involves leveraging the correlations or similarities in consumption behavior or voltage fluctuations between individual consumers and the overall network behavior observed at the substation [13]. (2) The second category, grounded in the concept of energy conservation, involves analyzing the energy flow within the network. This approach assumes that the energy supplied from the distribution transformer or primary feeder is equivalent to the combined energy consumption and meter error. By considering power consumption and distribution factors, this method identifies the phase of individual customers based on the overall energy balance within the LV network [14].

The studies in [15,16] focused on the first category, measuring pattern similarity using the cross-correlation coefficient. They determined the phase of the corresponding substations by maximizing this coefficient. Correlation characteristics among consumers and LV buses of distribution transformer were used in the identification process in [17]. Moreover, various unsupervised machine learning methods have been employed for clustering phases based on similarity profiles. For example, in [18], K-means clustering algorithms were used to group phases based on cosine or Euclidean distance mechanisms. Additionally, studies such as [19,20] utilized spectral clustering to compute eigenvectors representing voltage magnitudes and subsequently clustered them in reduced dimensions. Identifying phase change events employing spectral clustering was also proposed in [21], using AMI voltage data. In [22], the Spectral and Saliency Analysis Algorithm (SSA) was employed, utilizing active power to extract unique time series features that distinguish consumers and correlate them with substation power patterns. Furthermore, [23] applied K-means clustering by extracting unique consumer patterns using a high-pass filter. The research presented in [24] delved into the energy conservation category and introduced a continuous relation method. This approach involved the development of an indicator matrix to identify the increasing number of measurements attributed to the growing number of consumers. On the other hand, the study outlined in [25] proposed a technique that combined a connection matrix (i.e., linking child nodes with parent nodes) with Principal Component Analysis (PCA). These methodologies aimed to identify single-phase consumers, with some extending their scope to encompass a combination of multiple phases, as evidenced in studies such as [26,27]. In [28], an optimization formulation was developed to estimate the adjacency matrix by utilizing graph signal processing techniques. While these studies predominantly focus on the LV network, access to medium voltage (MV) information is essential as the MV networks involve higher voltages and more complex configurations compared to LV networks, necessitating a deeper understanding of phase characteristics and distribution within the network [29].

In other studies, voltage control techniques utilizing reactive power [30], active power drooping strategies [31], and modified damping control approaches [32], have been investigated to optimize power flow, stabilize voltage levels, and enhance grid performance, ultimately reducing economic burdens on DSOs. However, many of these studies operate under the assumption of accurate phase information available to the DSOs. In real-world scenarios, in most countries, distribution systems frequently undergo maintenance activities, such as changes in the connectivity of distribution lines, equipment replacement, and adjustments in network configurations [33]. Moreover, several other factors, such as equipment failures, grid expansions, or integration of new renewable energy sources, can further contribute to the dynamic nature of the distribution network [34]. Therefore, despite efforts by DSOs to maintain accurate phase information, the evolving nature of distribution systems makes it challenging to consistently uphold the precision of phase data in many countries [10]. Consequently, precise phase identification is essential for enabling DSOs to effectively manage power distribution among different phases, especially when controlling DER power output, thus optimizing grid operations, minimizing power losses, and enhancing overall system reliability [35]. This precision in identifying phases not only helps reduce operational costs but also ensures the overall stability and efficiency of the power system. The absence of accurate network information maintenance and the failure to incorporate MV data for phase identification highlight a knowledge gap that necessitates comprehensive data management strategies encompassing both LV and MV networks, ensuring the accuracy and reliability of phase identification processes across all voltage levels.

To bridge this knowledge gap, this paper proposes an enhanced phase identification approach by employing pattern similarity and optimization modeling with Mixed Integer Linear Programming (MILP) in the MV network. In the power system, the DSO relies on voltage magnitude, current magnitude, and power factor data obtained from the Feeder Remote Terminal Unit (FRTU) for the MV network and AMI for the LV network. However, the inherent errors in the voltage magnitude, stemming from the dynamic nature of the power system, restrict the primary usable data to the current magnitude [36]. To represent all possible combinations of phases simultaneously, this work offers the following main contributions.

- We proposed an enhanced phase identification method based on pattern similarity and MILP to identify both single-phase and multi-phase configurations in the MV network. The problem is formulated with the objective function of minimizing the sum of the modified Kirchhoff's Current Law (KCL) at the branch nodes and the similarity of power-current patterns. The combined objective function is then linearized by introducing auxiliary variables and is subsequently resolved using MILP.
- A permutation matrix is employed to establish correlations between the magnitudes of electrical currents across distinct phases in the complex MV networks. This matrix is manipulated based on empirical data to accurately reflect real-world scenarios. Subsequently, through comparisons with observed magnitudes of currents, the proposed approach is able to deduce and identify the phases of currents for wye-connected loads, delta-connected loads, and aggregated multi-phase loads in the electrical signals, thus providing a robust foundation for phase identification in MV networks.
- The proposed method utilizes real power consumption data, providing one-hour resolution spanning 168 time steps. The dataset includes distinct load patterns for weekdays and weekends, facilitating an analysis of consumption behaviors based on temporal factors. The simulation encompassed various case studies conducted on IEEE 13, IEEE 34, and IEEE 123 test feeders. The validity of the results is verified through rigorous accuracy assessments under different conditions, including ideal scenarios, with and without the similarity function, varying pattern similarities, and different penetration ratios of summation nodes at 30%, 50%, and 100% (e.g., all nodes), followed by comparative analysis against correlation coefficient and Euclidean distance based methods.

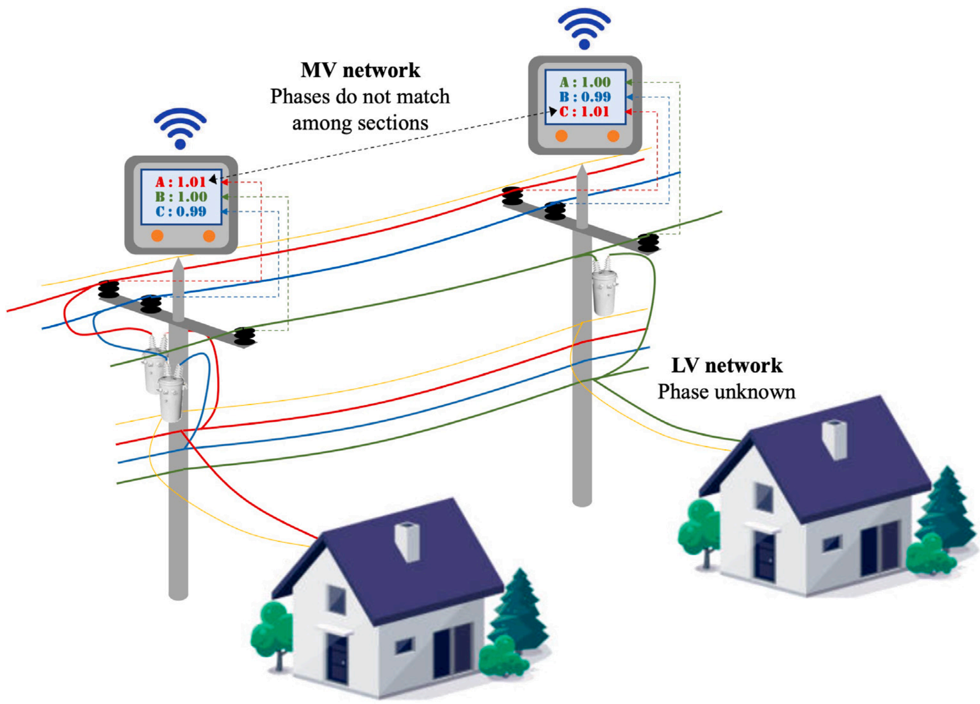

Fig. 1. An illustration of three phase MV and single-phase LV highlighting possibilities for the single-phase load and MV networks.

The remainder of this paper is structured as follows: Section 2 presents the formulation of the optimization problem and resolves it through MILP. Section 3 outlines the simulation environment setup, while Section 4 offers a discussion on simulation results. Finally, Section 5 concludes the paper, highlighting possible future directions.

2. The proposed MILP based approach for phase identification

Considering the intricacies inherent in the MV-network and the challenges posed by the lack of precise phase information availability, this section provides an in-depth exploration of the development of a MILP-based approach for phase identification in the MV-network.

2.1. Problem description

To define the problem, we first examine the representation of the MV-network, which entails connecting phases with adjacent conductors. Additionally, we incorporate single-phase LV network loads, which connect to one of the three phases, as illustrated in Fig. 1. It is evident from Fig. 1 that the process of identifying the phase of a single-phase load involves selecting one out of the three available phases (e.g., A, B, C). Conversely, phase identification within an MV network entails connecting the three conductors of adjacent buses in various permutations (e.g., ABC, ACB, ⋯, CAB). Therefore, for single-phase loads, there are three possibilities, while for MV networks, there are six possibilities. Notably, the six combinations in MV networks can be conceptualized as a permutation problem (${}_3P_3$) of arranging three elements. Consequently, phase identification in this paper is achieved by leveraging the magnitudes of currents in conjunction with a permutation matrix.

However, the delta-connected load configuration introduces inherent complexity due to the distribution of current, wherein each load is connected between two phases. This configuration leads to shared currents among phases, posing challenges for modeling using a permutation matrix. For example, in a delta configuration, the current flowing through one phase (such as phase A) serves as input for two distinct loads connected to phases AB and CA [37]. This shared current characteristic disrupts the straightforward one-to-one correspondence between phases and loads, which is fundamental to the application of permutation matrices. Moreover, in scenarios where AMI data is lacking, only the sum of power for multi-phase loads may be measured from the FRTU. In such cases, it is challenging to determine whether the load is a wye or delta connection.

2.2. Problem formulation

The problem formulation involves constructing permutation matrices for the current with both wye-connected and delta-connected loads. Given the challenge of discerning whether the load is wye or delta-connected when only summation data is available, a percentage load matrix is computed. This matrix allows for the manipulation of the data to accurately identify the corresponding currents and their respective phases, as detailed in the subsequent subsections.

2.2.1. Permutation matrix for current and wye-connected load

To obtain the phase information of the current and the wye-connected load, current magnitude and active power consumptions are computed first, followed by the development of a permutation matrix. The current magnitude $C^t_{calc,i}$ for each ith node such that $i \in N$ at time t is computed using the permutation matrix X_i and the collected current data M^t_i, which represents the electrical properties of the network at time t, as expressed by (1). Likewise, the active power consumption $P^t_{calc,i}$ for each node $i \in \mathcal{N}$ at time t is computed using the permutation matrix X_i and the collected power data P^t_i, which represents the power consumption characteristics of the network at time t, as expressed by (2).

$$\boldsymbol{C}^t_{calc,i} = \boldsymbol{X}_i \cdot \boldsymbol{M}^t_i \quad \forall i \in \mathcal{N} \tag{1}$$

$$\boldsymbol{P}^t_{calc,i} = \boldsymbol{X}_i \cdot \boldsymbol{P}^t_i \quad \forall i \in \mathcal{N}_Y \tag{2}$$

To comprehend the phase information for the current and wye-connected load from the collected data, several permutations need to be considered. A permutation matrix [38] rearranges the order of rows or columns, with the number of possible arrangements depending on the size of the matrix; for instance, for a size of $n \times n$, it has a total of $n!$ possible arrangements. Given the problem description (Fig. 1), for the wye-connected load, there are six possible combinations for the corresponding arrangements of the three phases (A, B, and C) in the dataset, representing different scenarios of how the currents may flow through the phases. To facilitate this arrangement process, a permutation matrix of 3 × 3 is established wherein its elements represent the possible combinations of phases, as presented in (3). This matrix is then manipulated through inner product operations with the collected data to determine the phase information. Furthermore, (4)–(5) ensure that for each row u, there is exactly one column v that can have a value of 1, and for each column v, there is exactly one row u that can have a value of 1, ensuring a one-to-one association between the rows and columns. Eq. (6) restricts the values of the decision variable $x_{u,v} \in [0,1]$, where it can only take on the values of 0 or 1.

$$\boldsymbol{X}_i = \begin{bmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \\ x_{21} & x_{21} & x_{22} \end{bmatrix} \tag{3}$$

$$\sum_u x_{u,v} = 1 \quad \forall u \tag{4}$$

$$\sum_v x_{u,v} = 1 \quad \forall v \tag{5}$$

$$x_{u,v} \in [0,\, 1] \tag{6}$$

Subsequently, the dot product operation results in values that correspond to phases a, b, and c in the order starting from the first row of the resulting matrix. Therefore, if the diagonal $diag(X_i) = [1,1,1]$, it indicates that the measured data was collected in the correct phase order, accurately representing the magnitudes of electrical currents for phases a, b, and c, respectively.

2.2.2. Permutation matrix for current and delta connected load

In the case of delta-connected loads, to obtain the phase information, we compute the active power $P^t_{calc,i}$ for each node $i \in \mathcal{N}_\Delta$, which represents delta-connected loads, at time t through the inner dot product of the permutation matrix D_i and the collected power data P^t_i, which characterizes the power consumption properties of the network at time t, in a similar manner as computed in (2). The resulting $P^t_{calc,i}$ provides an estimate of the active power consumption for each delta-connected load node i at time the specified time, as presented in (7). The permutation matrix D_i for the ith node, where each element $d_{u,v}$ represents the association between the rows and columns for the delta-connected load is represented in (8). The non-negativity of the elements in the permutation matrix D_i is guaranteed by (9) and (10) ensures that for each row u of D_i, the sum of its elements is equal to 1, indicating that the power must be injected from each phase in proportion to the load consumed by the law of energy conservation.

However, unlike a wye-connected load, where the current flows through each phase individually, in a delta-connected load, the current flows through one phase and is then divided into two loads. Therefore, in (11), the constraint ensures that the maximum value of the injection current occurs when all currents flowing through the two loads are combined (v-connection), indicating the maximum possible current injection for the delta-connected load configuration. This restriction ensures that the total power supplied from a single phase (e.g., phase A) can be received by a maximum of two loads (e.g., AB, CA loads), thereby preventing the supplied power from exceeding the capacity of the phase.

$$\boldsymbol{P}^t_{calc,i} = \boldsymbol{D}_i \cdot \boldsymbol{P}^t_i \quad \forall i \in \mathcal{N}_\Delta \tag{7}$$

$$\boldsymbol{D}_i = \begin{bmatrix} d_{00} & d_{01} & d_{02} \\ d_{10} & d_{11} & d_{12} \\ d_{20} & d_{21} & d_{22} \end{bmatrix} \tag{8}$$

$$d_{u,v} \geq 0 \quad \forall u, v \tag{9}$$

$$\sum_u d_{u,v} = 1 \quad \forall u \tag{10}$$

$$\sum_v d_{u,v} \leq 2 \quad \forall v \tag{11}$$

2.2.3. The percentage matrix for unknown load

In scenarios where the load data represents the summation of power, it remains ambiguous whether the load is wye-connected or delta-connected. In the case of a wye-connected load, the power consumption in each phase is determined using (12). However, for a delta-connected load, this value represents the power injected into each phase. This distinction arises from the fact that in a delta connection, the injected power is distributed between two loads. Unlike the previous formulation, which utilized vectors, in this case, as it is a scalar value, the phase of the load is identified through multiplication. The load percentage matrix Y^t_i, depicted in (13), consists of scalar elements y^t_u each representing the percentage of power consumption or injection for phase u, at time t. The non-negativity of the elements in Y^t_i is ensured through (14), while (15) ensures that the sum of the percentage matrix Y^t_i is equal to 1 for all the phases u at time t. This represents that the total power consumption or injection is correctly distributed among the phases, maintaining energy conservation principles.

$$\boldsymbol{P}^t_{calc,i} = \boldsymbol{Y}^t_i \times P^t_i \quad \forall i \in \mathcal{N}_{load} \tag{12}$$

$$\boldsymbol{Y}^t_i = \begin{bmatrix} y^t_0 \\ y^t_1 \\ y^t_2 \end{bmatrix} \tag{13}$$

$$y^t_u \geq 0 \quad \forall u \tag{14}$$

$$\sum_u y^t_u = 1 \quad \forall u \tag{15}$$

It should be noted that if one of the components y^t_u in Y^t_i equals 1, it indicates that the load is a single-phase load, suggesting that the power consumption or injection is concentrated in a single phase. However, if multiple components of Y^t_i have non-zero values, it suggests that the load is a multi-phase load, where power consumption or injection is distributed across multiple phases of the electrical system.

2.3. Objective function based on modified KCL and power pattern similarity

To formulate the objective function, we employ a modified version of KCL along with considerations for power pattern similarity to ensure electrical balance within the network and enhance phase identification accuracy. The KCL is based on the principle of conservation of charge, stating that the sum of currents entering a junction (i.e., node) in an electrical circuit must equal the sum of currents leaving the junction, thereby accounting for the conservation of electric charge, as expressed by (16) [39].

$$\sum_i I_i = 0 \tag{16}$$

where I represents the currents entering or leaving the ith node. While KCL is based on ideal scenarios, in realistic scenarios, determining the outflow of current directly is challenging, as measurements are often taken on the upstream side of the nodes [40]. Additionally, the current flowing through the load must be considered when a load is connected to a conductor. To address these challenges, this work modifies KCL, as presented in (17), aiming to minimize the sum of the injection currents at each junction along with the currents flowing through connected loads. This modification ensures that the overall current balance is maintained within the network, incorporating both measured currents and the presence of loads. To ensure that the optimization process minimizes the discrepancies between the calculated and actual current and power-voltage ratio, we define the function presented in (18). This function computes the values to be minimized in the objective function, considering various scenarios for leaf nodes that are load and non-load nodes, as well as for branch nodes. To ensure that the calculated values reflect the net current and power contributions of each node after considering the flow of current and power to neighboring nodes, adjustments to the calculated current and power values for each node i based on the current and power flows in the network edges ($\mathcal{E}$) are required, as computed in (19) and (20).

$$\min_{\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\boldsymbol{D}_i} \sum_{t\in T}\sum_{q\in Q}\left|f(\mathbf{X}_i,\mathbf{Y}_i^t,\mathbf{D}_i)\right| \tag{17}$$

$$f(\mathbf{X}_i,\mathbf{Y}_i^t,\boldsymbol{D}_i) = \begin{cases} C^t_{calc,i,q} & \forall i\in\mathcal{N}_{leaf}-(\mathcal{N}_{leaf}\cap\mathcal{N}_{load}) \\ I^t_{calc,i,q} - \frac{P^t_{calc,i,q}}{V^t_{i,q}} & \forall i\in(\mathcal{N}_{branch}\cap\mathcal{N}_{load}) \\ C^t_{calc,i,q} - \frac{P^t_{calc,i,q}}{V^t_{i,q}} & \forall i\in(\mathcal{N}_{leaf}\cap\mathcal{N}_{load}) \end{cases} \tag{18}$$

$$I^t_{calc,i,q} = C^t_{calc,i,q} - \sum_j C^t_{calc,i,q} \quad \forall(i\to j)\in\mathcal{E} \tag{19}$$

$$P^t_{calc,i,q} = P^t_{calc,i,q} - \sum_j P^t_{calc,i,q} \quad \forall(i\to j)\in\mathcal{E} \tag{20}$$

To achieve the objective of minimizing the current and power-voltage ratio, Eq. (17) relies on voltage data, which is unfortunately unreliable in distribution networks, indicating that (17) cannot be directly applied. However, leveraging the principle of balanced three-phase systems [41], wherein current magnitudes correlate with the power consumption of each phase, an assumption is made that the injected current magnitude at a node is proportional to the power consumed by each phase, given nearly uniform three-phase voltages. This assumption allows us to deduce that the ratio of a specific phase's current magnitude to the total injection current at a node mirrors the proportion of that phase's power consumption to the total consumed power at the node, as expressed in (21) and (22).

$$\frac{I^t_{calc,i,q}}{\sum_{j,q} I^t_{calc,i,q}} - \frac{P^t_{calc,i,q}}{\sum_q P^t_{calc,i,q}} \quad \forall i\in(\mathcal{N}_{branch}\cap\mathcal{N}_{load}) \tag{21}$$

$$\frac{C^t_{calc,i,q}}{\sum_q C^t_{calc,i,q}} - \frac{P^t_{calc,i,q}}{\sum_q P^t_{calc,i,q}} \quad \forall i\in(\mathcal{N}_{leaf}\cap\mathcal{N}_{load}) \tag{22}$$

In scenarios where a conductor is devoid of load, the patterns of current flow among adjacent conductors often exhibit similarity, particularly when these conductors are interconnected within the same phase. However, when a load is attached to a conductor, the consumption pattern of the load and the corresponding current pattern injected into it tend to manifest similarities. This characteristic becomes pivotal for discerning the phase similarity, as depicted in Fig. 2. It is noteworthy that when expressing these similarities in fractional form, as indicated in (21), the resulting values might become minimal, posing challenges in practical application. To mitigate this issue, a formalization method outlined in [22] has been devised to modify the objective function (17) and reformulate (18), as presented in (23) and (24). This process ensures that even the smallest values are taken into account for enhancing the precision and credibility of phase identification in complex MV networks.

$$\min_{\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\boldsymbol{D}_i} \sum_{t\in T}\sum_{q\in Q}\left|g(\mathbf{X}_i,\mathbf{Y}_i^t,\mathbf{D}_i)\right| \tag{23}$$

$$g(\mathbf{X}_i,\mathbf{Y}_i^t,\mathbf{D}_i^t) = \begin{cases} C^{\Delta t}_{calc,i,q} & \forall i\in\mathcal{N}_{leaf}-(\mathcal{N}_{leaf}\cap\mathcal{N}_{load}) \\ I^{\Delta t}_{calc,i,q} - P^{\Delta t}_{calc,i,q} & \forall i\in(\mathcal{N}_{branch}\cap\mathcal{N}_{load}) \\ C^{\Delta t}_{calc,i,q} - P^{\Delta t}_{calc,i,q} & \forall i\in(\mathcal{N}_{leaf}\cap\mathcal{N}_{load}) \end{cases} \tag{24}$$

where Δ represents the time interval, such that $C^{\Delta t}_{calc,i,q}$ means $C^t_{calc,i,q}$ - $C^{t-1}_{calc,i,q}$. The objective function developed in (23) is subject to several constraints that must be fulfilled to resolve the objective function.

Table 1
Load types classified by phases in the test feeder.

Feeder	Total load	Single-phase load	Multi-phase load
IEEE 13	9	5	4
IEEE 34	25	12	13
IEEE 123	114	109	5
modified IEEE 123	114	8	106

According to the law of conservation of energy, the power at the substation equals the sum of the power consumed by the load and the losses. Therefore, when the losses are neglected, the power at the substation is always greater than or equal to the consumed power. Eq. (25) ensures that the power at the substation ($P^t_{sub,q}$) is greater than or equal to the sum of the calculated power consumption of all nodes ($P^t_{calc,i,q}$) for each phase q and time t. While (26) represents the constraint on the maximum current that can be injected through each load node ($I^t_{calc,i,q}$) for each phase q and time t. It ensures that the calculated current does not exceed the maximum allowable current, which is determined by the power consumed ($P^t_{calc,i,q}$), the minimum voltage (V_{min}), and the minimum power factor (PF_{min}).

$$P^t_{sub,q} \ge \sum_i P^t_{calc,i,q} \quad \forall i,q,t \tag{25}$$

$$I^t_{calc,i,q} \le \frac{P^t_{calc,i,q}}{V_{min} * PF_{min}} \quad \forall i,q,t \tag{26}$$

2.4. Linearization of the objective function

The objective function formulated in (23) entails both discrete (i.e., integer) and continuous decision variables, resulting in an MILP problem. The use of MILP solvers is advantageous for optimization problems characterized by linear objective functions and constraints, where decision variables can assume discrete values alongside continuous values [42]. However, when faced with non-linear functions within such problems, linearization techniques become essential to render them compatible with MILP solvers [43]. Several linearization methods exist, each tailored to specific problem contexts. On the other hand, linearization of products and ratios using auxiliary variables offers versatility, enabling the handling of diverse constraint types within the optimization framework [44]. The choice of linearization technique depends on the problem characteristics, such as the nature of the nonlinearity and the desired balance between computational efficiency and solution accuracy. In line with the inherent characteristics of the objective functions delineated in this study, we employ the auxiliary variables linearization technique to approximate the nonlinear components present in (17) and (23) in order to resolve them through the MILP solver [45], and is presented in (27).

$$\begin{aligned} \min_{\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\boldsymbol{D}_i} \quad & \sum_t(\mathbf{Z}_i^t+\mathbf{U}_i^t) \\ \text{s.t.} \quad & \boldsymbol{Z}_i^t \ge f(\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\mathbf{D}_i) \\ & \boldsymbol{Z}_i^t \ge -f(\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\mathbf{D}_i) \\ & \boldsymbol{U}_i^t \ge g(\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\mathbf{D}_i) \\ & \boldsymbol{U}_i^t \ge -g(\boldsymbol{X}_i,\boldsymbol{Y}_i^t,\mathbf{D}_i) \\ & (19)\text{–}(22),\ (25)\text{–}(26) \end{aligned} \tag{27}$$

3. Simulation environment and assumptions

3.1. Simulation environment

The proposed pattern similarity-based MILP phase identification approach for MV networks is validated by considering various scenarios, including single-phase loads and multi-loads, as presented in Algorithm 1. This validation is conducted using benchmark test feeders such as the IEEE 13, 34, 123, and a modified version of the IEEE 123, as

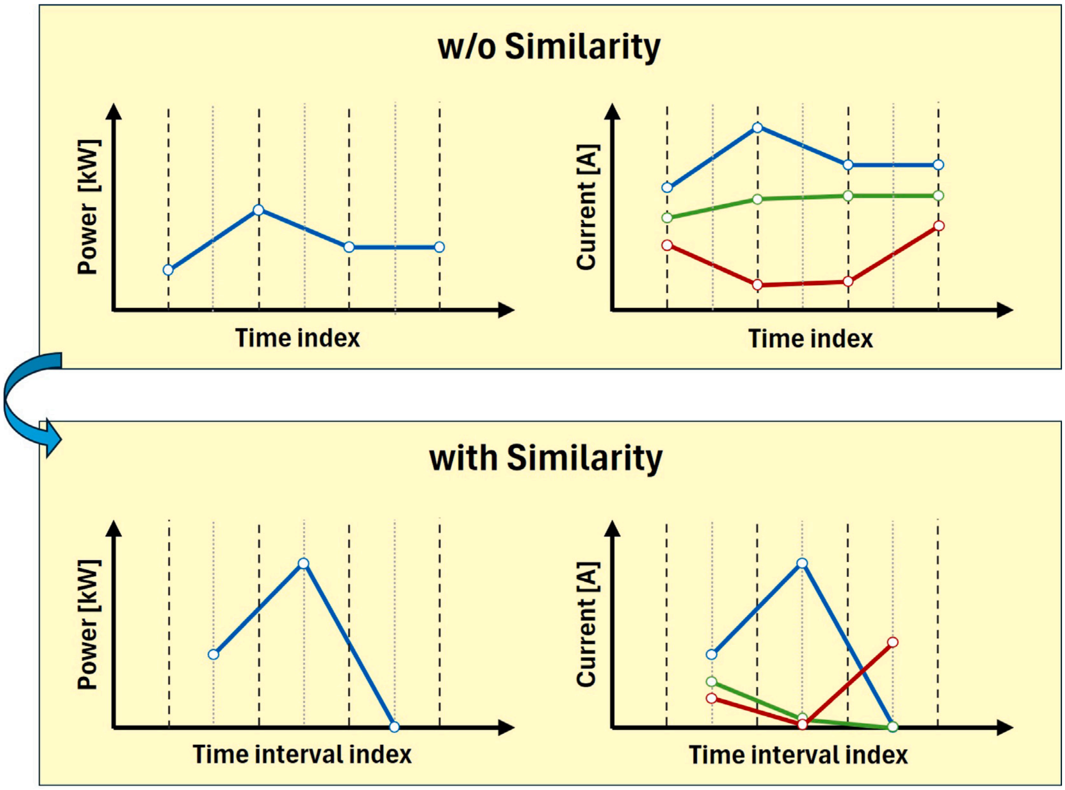

Fig. 2. An illustration of the pattern similarity function.

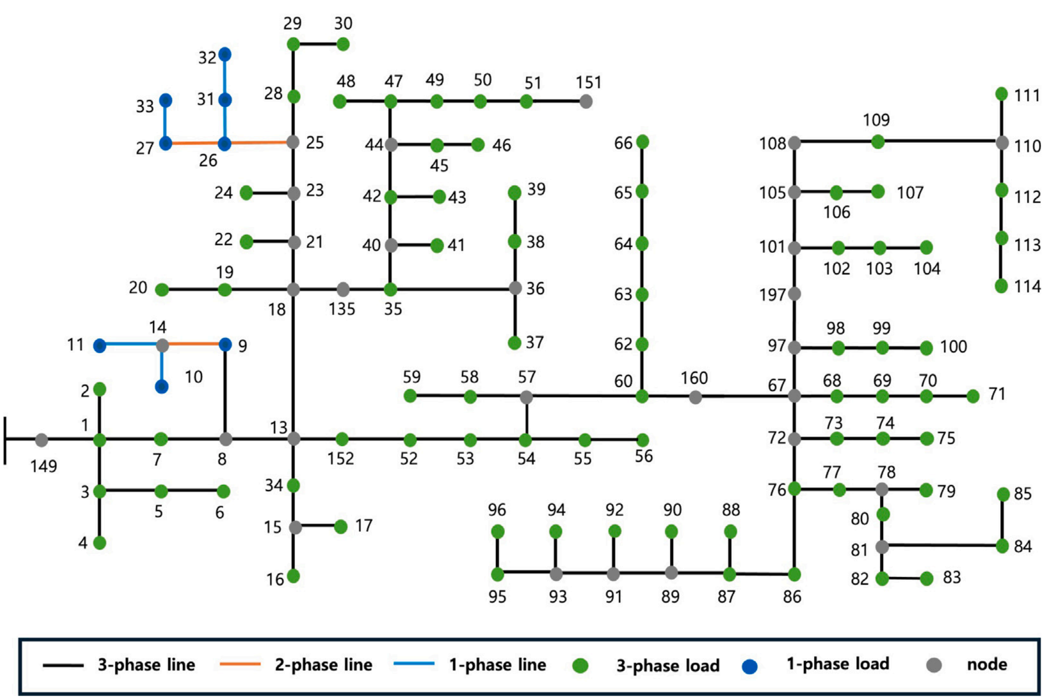

Fig. 3. Modified IEEE 123 test feeder highlights single-phase and multi-phase loads.

depicted in Fig. 3. These test feeders represent diverse and realistic network configurations, enabling an evaluation of the proposed approach's effectiveness and robustness across different network topologies and load scenarios. Information about each test feeder is shown in Table 1. The dataset utilized in this study was sourced from [46], offering one-hour resolution data spanning across 168 time steps. Notably, the dataset comprises distinct load patterns for weekdays and weekends, enabling an analysis of varying consumption behaviors based on temporal factors. To ensure realistic simulation conditions, the peak active power for each test feeder was set to 80% of their respective substation capacities. The magnitude data were acquired through power flow analysis conducted using the OpenDSS software [47]. Moreover, to analyze the influence of load patterns on accuracy, the load patterns were categorized into two distinct cases for comparative analysis: similar and

Algorithm 1 MILP based Phase Identification in Medium Voltage Networks

Input : Network topology, Parameters (PF_{min}, V_{min}), Measurements ($\boldsymbol{M}_i^t$, $\boldsymbol{P}_i^t$, $\boldsymbol{P}_{sub}^t$)
Output : $\mathbf{X}_i$ of $\mathcal{N}$

Set all the in each node using $\mathcal{N}$ and $\mathcal{E}$
while $(i \le |\mathcal{N}|)$ **do**
  Initialize $X[i]$, $D[i]$
  Validate Constraints via (4)–(6), (9)–(11)
  while $(t \le$ to $|\mathcal{T}|)$ **do**
    Initialize $Y[i][t]$
    Validate constraints via (14)–(15)
    $C_{calc}[i][t] \leftarrow X[i] \times M[i][t]$
    if $(i \in \mathcal{N}_Y)$ **then**
      $P_{calc}[i][t] \leftarrow X[i] \times P[i][t]$
    else if $(i \in \mathcal{N}_\Delta)$ **then**
      $P_{calc}[i][t] \leftarrow D[i] \times P[i][t]$
    else
      $P_{calc}[i][t] \leftarrow Y[i][t] \times P[i][t]$
    $t \leftarrow t + 1$
  $i \leftarrow i + 1$
while $(i \le |\mathcal{N}|)$ **do**
  while $(t \le$ to $|\mathcal{T}|)$ **do**
    for $(i, j \in \mathcal{E})$ **do**
      for $(j \in \mathcal{E})$ **do**
        $I_{calc}[i][t] \leftarrow C_{calc}[i][t] - C_{calc}[j][t]$
        $P_{calc}[i][t] \leftarrow P_{calc}[i][t] - P_{calc}[j][t]$
        $j \leftarrow j + 1$
    Formulate physical constraints of the power system via (25)–(26)
    if $(t \ge 1)$ **then**
      Compute $\Delta C_{calc}[i][t]$, $\Delta P_{calc}[i][t]$, $\Delta I_{calc}[i][t]$
    $t \leftarrow t + 1$
  $i \leftarrow i + 1$
Generate $f(\mathbf{X}_i, \mathbf{Y}_i^t, \mathbf{D}_i)$, $g(\mathbf{X}_i, \mathbf{Y}_i^t, \mathbf{D}_i)$
Add Auxiliary matrix $\mathbf{Z}_i^t$ and $\mathbf{U}_i^t$ for linearization.
Minimize$(Z_i^t + U_i^t)$
Matching substation's $\mathbf{X}_i$ with $\mathbf{X}_i$ of $\mathcal{N}$

non-similar. The determination of load pattern similarity was facilitated by computing the Euclidean distance metric, as referenced in [48], and presented in (28).

$$d_e = \sqrt{\sum_{i,t,q'} (P_{i,q}^t - P_{i,q'}^t)^2} \quad q, q' \in Q, q \neq q', t \in T \tag{28}$$

The measurement error is assumed to adhere to the three-sigma rule of the zero-mean Gaussian distribution. For the current magnitude, a 1% meter error is considered, following the distribution N(0, 0.3). Moreover, it should be noted that considering meter errors may result in the validation of constraints leading to potentially infeasible solutions. Consequently, to avoid infeasibility arising from the validation of constraints under meter errors, the problem was solved without imposing constraints.

The analysis was further extended to include a scenario involving multi-phase summation loads, in addition to the analysis of individual loads, thereby providing an understanding of the transition from individual to summation load scenarios. In the IEEE 13, 34, and 123 test feeders, all the lesser multi-phase loads were aggregated into a single summed-up load data. Conversely, in the modified IEEE 123 test feeder, 30% and 50% of the randomly selected 79 multi-phase loads were consolidated into three-phase aggregated loads. The power consumed by each phase of the summation load is determined by the load percentage matrix. Furthermore, the proposed method was implemented using Gurobi version 11, and the accuracy was evaluated according to the criteria outlined in [49].

3.2. Assumptions on the data collection and measurement mechanism

To deal with the complexities of the phase identification problem in the MV network concerning current allocation to various loads, several assumptions are made:

- *Steady-State Conditions and Constant Phase Angles:* All data collection occurs under steady-state conditions to ensure that the system operates in a stable state without transient effects, simplifying the analysis and providing consistent data for phase identification. Moreover, it is assumed that the phase angles of currents remain nearly constant within the same phase.
- *Measurement Location:* Data are collected on the upstream side of nodes to capture the total current flowing through the network, thereby facilitating accurate load identification.
- *Current Threshold and Phase Separation:* To distinguish between different phases, loads, and mitigate the risk of misidentification, a threshold of 1 A (1 Ampere) is considered.
- *Treatment of Neutral Line Current:* While KCL traditionally incorporates neutral line currents in three-phase power systems, it is assumed that the magnitude of the neutral current is negligible in the context of the studied MV networks. Therefore, simplification is made to KCL, applying it exclusively to the a.b.c. phases that allow for a streamlined analysis without significant loss of accuracy, as it aligns with the common observation that neutral currents in MV networks are minimal compared to phase currents.

4. Simulation results

This section discusses the results in three categories: the first category corresponds to the validation of the objective function to understand how it enhances the results; the second category is dedicated to exploring various scenarios and their corresponding outcomes, while the third category deals with comparing the results against state-of-the-art phase identification methods.

4.1. Validation of objective function

It is evident that the objective function developed in (27) is based on the combination of two components: the modified KCL and the power pattern similarity. Consequently, it is important to comprehend how the objective function improves the identification accuracy by focusing on the role of the individual components and their contributions.

To verify this, we demonstrate the use of the power pattern similarity function in Fig. 4 to reduce redundant solutions in the aggregated three-phase load. Following this, Fig. 5 illustrates the contribution of each component, along with the current and load pattern, both with and without the power similarity function. This figure shows that the proposed solution emphasizes the presence of duplicate solutions for individual phases and, thereby, even with variations in the current phase, the outcome of the modified KCL objective function remains consistent based on the load distribution. The figure shows that in the early time steps, around 0 to 25, the current without similarity fluctuates between 0.3 and 0.4, consistently underestimating the reference power, which is closer to 0.6 to 0.8. With the similarity function applied, the current follows the reference power much more closely, reaching about 0.7 to 0.8. This makes the error in the early time steps much smaller, unlike the case without the similarity function, where the current clearly struggles to capture the load changes. Following the pattern in the middle time steps, around 75 to 100, the difference between the two approaches becomes evident. Without the similarity function, the current drops to about 0.4 to 0.6, while the reference power rises above 0.8 to 0.9. This leads to a noticeable gap, showing

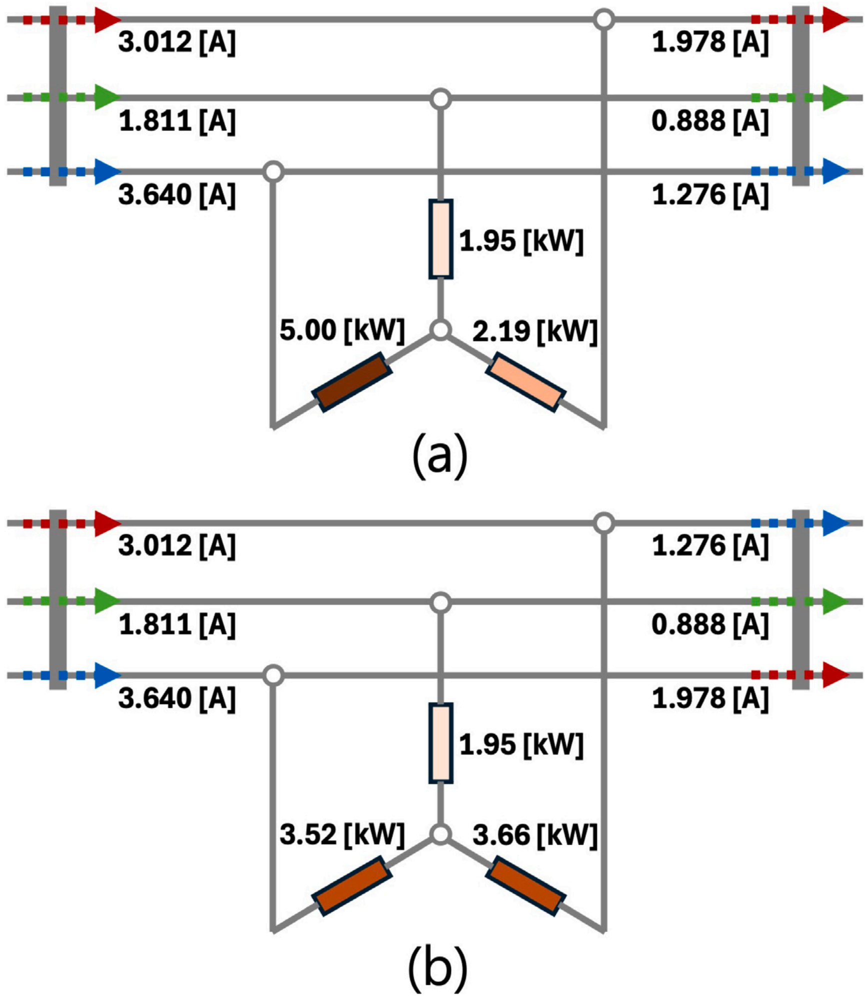

Fig. 4. Current and load flow highlighting duplicate solutions of objective functions. (a). With power pattern similarity, (b). Without power pattern similarity.

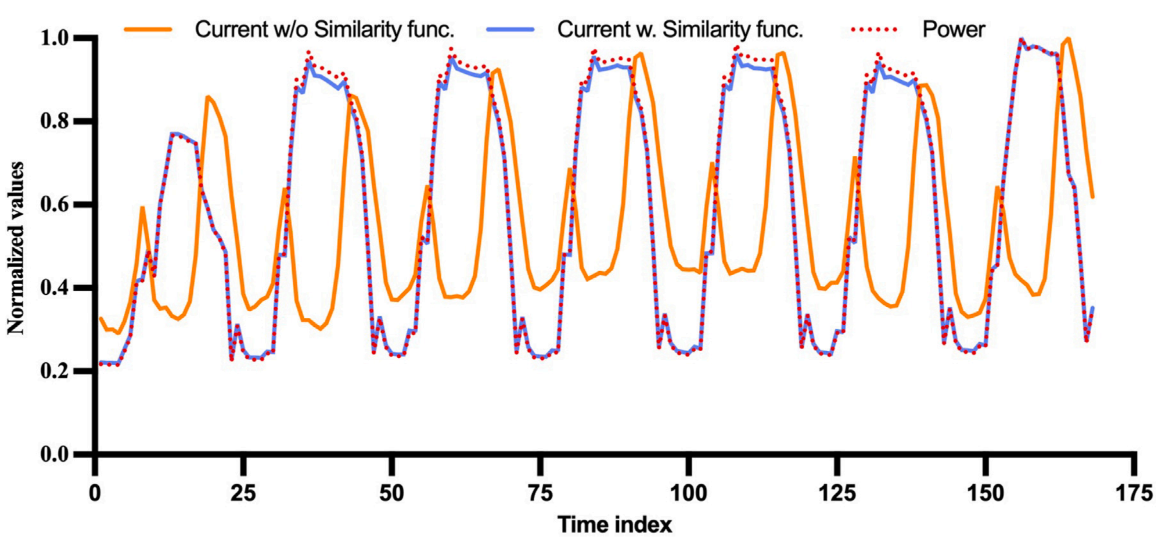

Fig. 5. Load and current patterns for phase *a* with and without the similarity function.

that the current is underestimating the actual trend. In contrast, with the similarity function, the current tracks the power curve closely, maintaining itself within 0.85 to 0.95, which demonstrates a much better match during periods of high demand. Likewise, towards the end of the time index, around 150 to 175, the same pattern holds. The current without similarity lags behind, remaining around 0.4 to 0.6, even though the reference power climbs close to 1.0. With the similarity function, however, the current and power curves almost overlap, both rising to about 0.95 to 1.0 before gradually declining. This shows that the similarity-based objective not only improves accuracy but also

Table 2
Accuracy for converted summation power nodes.

Feeder	Accuracy (%)	
	with similarity function	without similarity function
IEEE 13	100	92.96
IEEE 34	100	68.84
IEEE 123	99.1	91.6
modified IEEE 123	100	94.4

Table 3
Accuracy of various load patterns under ideal conditions.

Feeder	Pattern similarity (kWh)	Accuracy (%)
IEEE 13	728.3	100
	269	100
IEEE 34	810.1	100
	474.2	96.6
IEEE 123	698.7	100
	291.9	100
modified IEEE 123	2806.1	100
	723.2	100

maintains it consistently across the entire time horizon. The accuracy corresponding to the ideal condition (i.e., incorporating two objective functions) using (27) was compared to the performance achieved by excluding (23) from the ideal condition, as shown in Table 2. The table indicates that in the absence of a power pattern similarity function, the accuracy decreases.

This analysis validates that the pivotal role of the power pattern similarity function, in conjunction with KCL, enhances the accuracy of phase identification in MV networks.

4.2. Results of the various scenarios

The phase identification accuracy of the proposed MILP approach is further evaluated in various scenarios that resemble realistic cases. The first scenario corresponds to the benchmark where all load information is known, and the results are presented in Table 3. It is noteworthy that, despite significant variations in the load pattern for each phase, the proposed approach maintains a high accuracy of 100% for the IEEE 13, 123 test feeders, and the modified IEEE 123 test feeder. However, in cases where the load patterns are similar for each phase, a slight decrease of 3.6% in accuracy can be observed only in the 34-test feeder. This decrease is attributed to the high number of delta-connected 2-phase load instances in the 34-test feeder, causing identical current flow in both phases. Consequently, an increased presence of similar current magnitudes in the feeder led to reduced accuracy.

The next scenario corresponds to the aggregated load, where the accuracy results for various load patterns aggregated into single power nodes are compared with the ideal conditions (calculated in Table 2) and are shown in Table 4. In Table 4, where a multi-phase load is converted into an unknown load, it is confirmed that in most feeders, the higher the value of pattern similarity (i.e., the more unique the power pattern consumed by each phase within the conductor), the higher the accuracy. Interestingly, in the IEEE 13 test feeder, accuracy remained consistent regardless of pattern similarity when compared to the ideal condition. However, phase identification encountered challenges in nodes with delta-connected 2-phase loads due to errors in injected currents for each phase. Specifically, under ideal conditions, delta-connected loads were excluded from the phase identification targets because the injected currents in each phase were identical.

Nevertheless, errors in the meters led to discrepancies that increased the difference between the magnitudes of the two currents. As a result, the third scenario takes meter errors into account to evaluate the accuracy of the proposed MILP, as illustrated in Table 5. It can be seen that in all feeders except IEEE 123, accuracy suffered due to the presence

Table 4
Accuracy of various converted load patterns to summation node.

Feeder	Summation node	Pattern similarity (kWh)	Accuracy (%)
IEEE 13	All	728.3	100
		269	92.5
IEEE 34	All	810.1	100
		474.2	96.6
IEEE 123	All	698.7	99.1
		291.9	99.1
modified IEEE 123	30%	2806.1	100
		723.2	99.4
	50%	2806.1	100
		723.2	97.6

Table 5
Accuracy of various load patterns with meter error.

Feeder	Pattern similarity (kWh)	Accuracy (%)
IEEE 13	728.3	92.5
	269	92.5
IEEE 34	810.1	96.6
	474.2	96.6
IEEE 123	698.7	100
	291.9	100
modified IEEE 123	2806.1	98.81
	723.2	91.96

of meter erroneous data. It is noteworthy that, in a single-phase load, the injected current is most affected by the power consumption. If the power consumption exceeds the meter's error, the algorithm can identify the phase, even when there is a meter error. The high accuracy with the IEEE 123 feeder, regardless of the load pattern and meter errors, is primarily because single-phase loads make up the majority of the total load.

4.3. Comparative analyses with state-of-the-art methods

Following the objective function validation and different load scenarios, the analyses are extended with a comparative study against state-of-the-art methods. The comparative study employs the Euclidean distance based method [50] and correlation coefficient based method [51] for identifying relationships between branches and child nodes in the phase identification problem. In addition to accuracy performance, the comparative analysis is extended with incorporating True Positive (TP), True Negative (TN), F1 score, Precision, and Recall measurements. The corresponding results of each method are presented in Fig. 6. The results reveal that the proposed method has dominant results, achieving an accuracy of 98.81%, a recall of 99.00%, an F1 score of 98.80%, a true positive rate of 98.81%, and a true negative rate of 99.00 %, respectively. This implies that the proposed MILP method is able to improve the accuracy by 31.65% and 36.38%, recall by 32.00% and 37.00 %, F1 score by 31.80% and 36.60%, true positive rate by 31.81% and 36.81%, and true negative rate by 33.00% and 23.00%, when compared to the correlation coefficient based method (accuracy 67.16%, recall 67.00%, F1-score 67.00%, true positive rate 67.00%, and true negative rate 66.00%) and Euclidean distance based method (accuracy 62.43%, recall 62.00%, F1-score 62.20%, true positive rate 62.00%, and true negative rate 76.00%), respectively.

The enhanced accuracy of the proposed method compared to state-of-the-art approaches can be attributed to two fundamental theoretical factors: the enforcement of physical constraints and the complementary role of pattern similarity. First, unlike purely data-driven methods that rely solely on statistical correlations, the proposed method explicitly integrates the Modified Kirchhoff's Current Law (KCL) as a hard constraint. This mathematical modeling ensures that any identified solution strictly adheres to the physical laws of power flow,

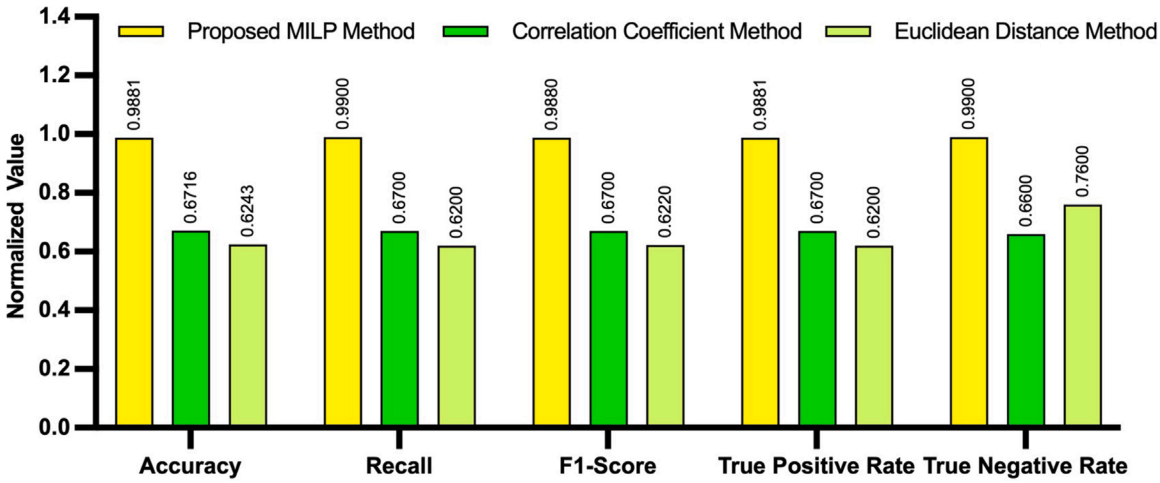

Fig. 6. Comparative study against cutting-edge phase identification methods.

effectively eliminating statistically plausible but physically impossible phase configurations that other methods fail to reject. Furthermore, while these physical constraints define the valid solution space, the problem can still remain underdetermined in complex topologies. In such cases, the Power Pattern Similarity Function acts as a decisive criterion, guiding the optimization solver to select the unique, correct solution by leveraging temporal data consistency. Consequently, the proposed method achieves robust performance by combining the necessary condition of physical energy balance with the sufficient condition of pattern similarity.

4.4. Strengths and limitations

The primary strength of the proposed method lies in its high accuracy and robustness, which is validated across multiple IEEE test feeders and under various challenging scenarios. The comparative analysis in Section 4.3 demonstrates that our approach significantly outperforms conventional methods based on correlation coefficients and Euclidean distance, achieving an accuracy of 98.81%, which represents an improvement of over 31%. This superior performance is attributable to several key innovations:

- Hybrid Objective Function: Unlike methods that rely solely on pattern similarity, our approach integrates a physics-based constraint (the modified Kirchhoff's Current Law) with data-driven pattern similarity. This hybrid formulation creates a more constrained and realistic optimization problem, allowing the model to effectively reduce redundant solutions and improve accuracy, especially for aggregated three-phase loads.
- Comprehensive Load Modeling: The method is explicitly designed for the complexities of MV networks. By employing permutation, delta-connected, and percentage matrices, it can successfully identify phases for wye-connected, delta-connected, and unknown aggregated multi-phase loads. This is a significant advancement over many existing techniques that are primarily focused on identifying single-phase customers in LV networks.
- Resilience to Realistic Conditions: The method maintains high performance even when faced with real-world data challenges. It was validated using real power consumption data with distinct weekday and weekend patterns. Furthermore, it demonstrates resilience under varying load pattern similarities and in the presence of meter errors, which are common issues in practical distribution systems.

Despite its strengths, the proposed method has several limitations that warrant discussion and offer avenues for future research:

- Computational Complexity: The formulation as a MILP problem, while powerful, can be computationally intensive. Although the method performed well on test feeders with up to 123 nodes, its scalability to significantly larger, real-world distribution networks with thousands of nodes could pose a challenge in terms of processing time. However, given that phase identification is inherently an offline planning task rather than a real-time operation, strict computational speed is considered less critical than accuracy in this context. Therefore, a detailed analysis of solution time relative to node scale was excluded from the scope of this study. Future work could explore decomposition techniques or computationally efficient heuristics to address this.
- Dependence on Data Quality: The accuracy of the method is inherently linked to the quality and resolution of the input data. As shown in Table 5, the presence of meter errors can degrade performance, although the method remains relatively robust. The current study utilized one-hour resolution data; while effective, higher-resolution data could potentially improve accuracy but would also increase the computational burden.
- Underlying Assumptions: The model is based on several simplifying assumptions outlined in Section 3.2. It assumes steady-state conditions, negligible neutral line currents, and nearly constant phase angles. In distribution networks with high levels of imbalance or significant penetration of intermittent DERs, these assumptions may not perfectly hold, potentially impacting the accuracy of the modified KCL component of the objective function.
- Performance with Highly Similar Patterns: While generally robust, the method's accuracy slightly decreases when load patterns across different phases are highly similar, as observed in the IEEE 34 test feeder results. Because the objective function partially relies on differentiating power-current patterns, a lack of distinctiveness among phases can make the identification task more challenging.

5. Conclusion

This paper presented an enhanced method for phase identification in medium voltage networks. The approach integrated an objective function aimed at minimizing the sum of the modified Kirchhoff's Current Law with power-current pattern similarity. The proposed approach was applied to a dataset of real power consumption with one-hour resolution data spanning 168 time steps, including distinct load patterns for weekdays and weekends. The simulation encompassed various case studies conducted on IEEE 13, IEEE 34, and IEEE 123 test feeders. The validation of the objective function underscored the significant role played by the power pattern similarity function and Kirchhoff's Current Law in enhancing phase identification accuracy, particularly by mitigating redundant solutions in aggregated three-phase loads. Moreover, the results obtained under various scenarios demonstrated the efficiency of the proposed pattern similarity-based mixed integer linear programming approach, maintaining high accuracy across different test feeders, even in scenarios with varying load patterns that mimic real-world situations. This was followed by a comparative study against correlation coefficient and Euclidean based methods, where the results showed an improvement of 31.65% and 36.38% in accuracy compared to these state-of-the-art methods. These results demonstrate the superior performance and high practical applicability of the proposed phase identification method within medium voltage distribution networks.

CRediT authorship contribution statement

Jungmin Seong: Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Jin Sol Hwang:** Writing – review & editing. **Shahid Hussain:** Writing – review & editing, Writing – original draft, Visualization, Validation. **Miguel Heleno:** Supervision. **Yun-Su Kim:** Supervision, Funding acquisition.

Declaration of competing interest

The authors have declared no conflict of interest.

Acknowledgments

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. RS-2024-00394225) and also supported by the Gwangju Institute of Science and Technology (GIST) research fund (Future-leading Specialized Research Project, 2025).

Data availability

All data used in this study are publicly available and cited within the article's references.

References

[1] Huang N, Li B, Wang Y, Cai G, Zhang L. Co-optimization and benefit allocation mechanism of DSO-VPP under the secondary flexibility market based on SPP. Int J Electr Power Energy Syst 2025;171:111044.
[2] Hu J, Liu X, Shahidehpour M, Xia S. Optimal operation of energy hubs with large-scale distributed energy resources for distribution network congestion management. IEEE Trans Sustain Energy 2021;12:1755–65.
[3] Ochoa LF, Harrison GP. Minimizing energy losses: Optimal accommodation and smart operation of renewable distributed generation. IEEE Trans Power Syst 2010;26:198–205.
[4] Jamali-Abnavi A, Hashemi-Dezaki H, Ahmadi A, Mahdavimanesh E, Tavakoli M-J. Harmonic-based thermal analysis of electric arc furnace's power cables considering even current harmonics, forced convection, operational scheduling, and environmental conditions. Int J Therm Sci 2021;170:107135.
[5] Lind L, Cossent R, Chaves-Ávila JP, Gómez San Román T. Transmission and distribution coordination in power systems with high shares of distributed energy resources providing balancing and congestion management services. Wiley Interdiscip Rev: Energy Environ 2019;8:e357.
[6] Muhanji SO, Muzhikyan A, Farid AM. Distributed control for distributed energy resources: Long-term challenges and lessons learned. IEEE Access 2018;6:32737–53.
[7] Ruester S, Schwenen S, Batlle C, Pérez-Arriaga I. From distribution networks to smart distribution systems: Rethinking the regulation of European electricity DSOs. Util Policy 2014;31:229–37.
[8] Silva R, Alves E, Ferreira R, Villar J, Gouveia C. Characterization of TSO and DSO grid system services and tso-dso basic coordination mechanisms in the current decarbonization context. Energies 2021;14:4451.
[9] Shyni R, Kowsalya M. HESS-based microgrid control techniques empowered by artificial intelligence: A systematic review of grid-connected and standalone systems. J Energy Storage 2024;84:111012.
[10] Park J-S, Kim J-H, Noh S-G, Cho W-Y, Kim H-K, Kim BH, et al. Distribution system operator (DSO) functions in South Korea—part I: Development, practical obstacles, and future work. Int Trans Electr Energy Syst 2024;2024:3353622.
[11] Bozalakov D, Laveyne J, Desmet J, Vandevelde L. Overvoltage and voltage unbalance mitigation in areas with high penetration of renewable energy resources by using the modified three-phase damping control strategy. Electr Power Syst Res 2019;168:283–94.
[12] Kabalcı E, Kabalcı Y, Siano P. Design and implementation of a smart metering infrastructure for low voltage microgrids. Int J Electr Power Energy Syst 2022;134:107375.
[13] García S, Mora-Merchán JM, Larios DF, Personal E, Parejo A, León C. Phase topology identification in low-voltage distribution networks: A Bayesian approach. Int J Electr Power Energy Syst 2023;144:108525.
[14] Hoogsteyn A, Vanin M, Koirala A, Van Hertem D. Low voltage customer phase identification methods based on smart meter data. Electr Power Syst Res 2022;212:108524.
[15] Pezeshki H, Wolfs P. Correlation based method for phase identification in a three phase LV distribution network. In: 2012 22nd australasian universities power engineering conference. IEEE; 2012, p. 1–7.
[16] Luan W, Peng J, Maras M, Lo J, Harapnuk B. Smart meter data analytics for distribution network connectivity verification. IEEE Trans Smart Grid 2015;6:1964–71.
[17] Zhou L, Li Q, Zhang Y, Chen J, Yi Y, Liu S. Consumer phase identification under incomplete data condition with dimensional calibration. Int J Electr Power Energy Syst 2021;129:106851.
[18] Arya V, Mitra R, Mueller R, Storey H, Labut G, Esser J, et al. Voltage analytics to infer customer phase. In: IEEE PES innovative smart grid technologies, Europe. IEEE; 2014, p. 1–6.
[19] Blakely L, Reno MJ, Feng W-c. Spectral clustering for customer phase identification using AMI voltage timeseries. In: 2019 IEEE power and energy conference at Illinois. IEEE; 2019, p. 1–7.
[20] Liu S, Cui X, Lin Z, Lian Z, Lin Z, Wen F, et al. Practical method for mitigating three-phase unbalance based on data-driven user phase identification. IEEE Trans Power Syst 2020;35:1653–6.
[21] Peña BD, Blakely L, Reno MJ. Online data-driven detection of phase changes in evolving distribution systems. In: 2023 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2023, p. 1–5.
[22] Xu M, Li R, Li F. Phase identification with incomplete data. IEEE Trans Smart Grid 2016;9:2777–85.
[23] Hosseini ZS, Khodaei A, Paaso A. Machine learning-enabled distribution network phase identification. IEEE Trans Power Syst 2020;36:842–50.
[24] Arya V, Seetharam D, Kalyanaraman S, Dontas K, Pavlovski C, Hoy S, et al. Phase identification in smart grids. In: 2011 IEEE international conference on smart grid communications. IEEE; 2011, p. 25–30.
[25] Pappu SJ, Bhatt N, Pasumarthy R, Rajeswaran A. Identifying topology of low voltage distribution networks based on smart meter data. IEEE Trans Smart Grid 2017;9:5113–22.
[26] Wang W, Yu N. Maximum marginal likelihood estimation of phase connections in power distribution systems. IEEE Trans Power Syst 2020;35:3906–17.
[27] Dilek M, Broadwater RP, Sequin R. Phase prediction in distribution systems. In: 2002 IEEE power engineering society winter meeting. conference proceedings (cat. no. 02CH37309), vol. 2, IEEE; 2002, p. 985–90.
[28] Dahale S, Pahwa A, Natarajan B. Phase identification in unobservable distribution systems. IEEE Trans Power Deliv 2023;38:3067–75.
[29] Hubana T, Saric M, Avdaković S. Approach for identification and classification of hifs in medium voltage distribution networks. IET Gener Transm Distrib 2018;12:1145–52.
[30] Eid A, Abdel-Akher M. Voltage control of unbalanced three-phase networks using reactive power capability of distributed single-phase pv generators. Int Trans Electr Energy Syst 2017;27:e2394.
[31] Baker K, Bernstein A, Dall'Anese E, Zhao C. Network-cognizant voltage droop control for distribution grids. IEEE Trans Power Syst 2017;33:2098–108.
[32] Bozalakov DV, Vandoorn TL, Meersman B, Papagiannis GK, Chrysochos AI, Vandevelde L. Damping-based droop control strategy allowing an increased penetration of renewable energy resources in low-voltage grids. IEEE Trans Power Deliv 2016;31:1447–55.
[33] Liu X, Liu Y, Liu J, Xiang Y, Yuan X. Optimal planning of ac-dc hybrid transmission and distributed energy resource system: Review and prospects. CSEE J Power Energy Syst 2019;5:409–22.
[34] Mlilo N, Brown J, Ahfock T. Impact of intermittent renewable energy generation penetration on the power system networks—A review. Technol Econ Smart Grids Sustain Energy 2021;6:25.
[35] Liu D, Giraldo JS, Palensky P, Vergara PP. A siamese neural network model for phase identification in distribution networks. Int J Electr Power Energy Syst 2025;169:110718.
[36] Kim H-S, Choi M-S, Lee S-J. A study on power flow method of radial distribution system using a measured data from FRTU in distribution automation system. In: 2015 KIEE summer conference. 2015, p. 286–7.
[37] Zhao C, Dall-Anese E, Low S. Optimal power flow in multiphase radial networks with delta connections: Preprint. 2017, IREP 2017 Bulk Power Systems Dynamics and Control Symposium ; Conference date: 27-08-2017 Through 01-09-2017.

[38] Schrijver A. On the history of combinatorial optimization (till 1960). In: Aardal K, Nemhauser G, Weismantel R, editors. Discrete optimization. Handbooks in operations research and management science, vol. 12, Elsevier; 2005, p. 1–68. http://dx.doi.org/10.1016/S0927-0507(05)12001-5, URL: https://www.sciencedirect.com/science/article/pii/S0927050705120015.

[39] Razavi B. Fundamentals of microelectronics. John Wiley & Sons; 2021.

[40] Brandao DI, Guillardi H, Morales-Paredes HK, Marafao FP, Pomilio JA. Optimized compensation of unwanted current terms by ac power converters under generic voltage conditions. IEEE Trans Ind Electron 2016;63:7743–53.

[41] Pandey A, Jereminov M, Wagner MR, Bromberg DM, Hug G, Pileggi L. Robust power flow and three-phase power flow analyses. IEEE Trans Power Syst 2018;34:616–26.

[42] Salo A, Andelmin J, Oliveira F. Decision programming for mixed-integer multi-stage optimization under uncertainty. European J Oper Res 2022;299:550–65.

[43] Mikolajková M, Saxén H, Pettersson F. Linearization of an minlp model and its application to gas distribution optimization. Energy 2018;146:156–68.

[44] Wirtz M, Hahn M, Schreiber T, Müller D. Design optimization of multi-energy systems using mixed-integer linear programming: Which model complexity and level of detail is sufficient? Energy Convers Manage 2021;240:114249.

[45] McCarl BA, Spreen TH. Applied mathematical programming using algebraic systems. Cambridge, MA; 1997.

[46] Ong S, Clark N. Commercial and residential hourly load profiles for all TMY3 locations in the United States. Open Energy Data Initiative (OEDI), National Renewable Energy Laboratory; 2014, http://dx.doi.org/10.25984/1788456, URL: https://data.openei.org/submissions/153. [Accessed 01 October 2025].

[47] Guerra G, Martinez-Velasco JA. Evaluation of matpower and opendss load flow calculations in power systems using parallel computing. J Eng 2017;2017:195–204.

[48] Javed A, Lee BS, Rizzo DM. A benchmark study on time series clustering. Mach Learn Appl 2020;1:100001.

[49] Therrien F, Blakely L, Reno MJ. Assessment of measurement-based phase identification methods. IEEE Open Access J Power Energy 2021;8:128–37.

[50] Serra J, Arcos JL. An empirical evaluation of similarity measures for time series classification. Knowl-Based Syst 2014;67:305–14.

[51] Ye J, Xiao C, Esteves RM, Rong C. Time series similarity evaluation based on spearman's correlation coefficients and distance measures. In: Cloud computing and big data: second international conference, cloudCom-Asia 2015, Huangshan, China, June 17-19, 2015, revised selected papers 2. Springer; 2015, p. 319–31.