

Article

Development of a Screening-Based Machine Learning Algorithm for PV Power Contribution Estimation

Yong-Joo Jeon  and Yun-Hyuk Choi * 

Department of Electrical Engineering, Daegu Catholic University, Gyeongsan 38430, Republic of Korea;
jyjoo1212@naver.com

* Correspondence: yhchoi@cu.ac.kr

Abstract

The growing penetration of distributed energy resources in distribution networks has highlighted the need to analyze power contribution characteristics under different operating and interconnection conditions. In particular, photovoltaic systems exhibit varying load-level power contribution patterns depending on their output levels and interconnection configurations, and machine-learning-based approaches have recently been investigated to efficiently estimate these contribution characteristics. However, conventional full-scale machine learning approaches require large datasets covering numerous PV operating and interconnection scenarios, resulting in substantial computational burden during dataset generation and model training. To address this issue, this paper proposes a screening-based machine learning algorithm for load-level PV power contribution estimation. A PV Impact-Range Index is developed by jointly considering active power contribution, voltage sensitivity, and electrical proximity to identify PV interconnection scenarios with meaningful contribution characteristics. Based on this index, non-informative scenarios are excluded, and a screening-scale dataset is constructed to train a computationally efficient learning model. Simulation results on a 19-bus distribution test system show that the proposed algorithm achieves prediction accuracy comparable to the full-scale learning approach while significantly reducing dataset generation and model training time. These results demonstrate that the proposed algorithm provides an efficient and practical solution for PV power contribution estimation in distribution networks.

Keywords: PV power contribution estimation; screening-based machine learning; random forest; distribution networks; distributed energy resources



Academic Editor: Daniela Proto

Received: 10 June 2026

Revised: 21 July 2026

Accepted: 24 July 2026

Published: 26 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

As global electricity demand continues to rise, power systems worldwide are seeking strategies to ensure a reliable and secure electricity supply [1]. This increase is largely driven by rapid industrial electrification, growing data center- and AI-related computing loads, and the widespread adoption of electric vehicles. According to a report by the U.S. Department of Energy, electricity usage by data centers is projected to increase substantially in the coming years, reflecting the growing power demand from digital infrastructure and the broader increase in electricity consumption across multiple sectors [2]. Consequently, electricity systems that rely heavily on fossil-fuel-based generation face increasing environmental and economic constraints due to fuel price volatility and stringent carbon emission regulations [3]. In this context, renewable energy resources have emerged as a key solution for reducing carbon emissions and enhancing energy self-sufficiency [4].

According to the IEA's Global Energy Review 2025, global electricity demand increased by 4.3% in 2024, driven primarily by increasing electrification and the rapid growth of digital infrastructure. In parallel, renewable energy capacity expanded rapidly, with solar photovoltaic (PV) and wind power accounting for approximately 70% of new global capacity additions, indicating a clear shift away from centralized fossil-fuel-based generation [5]. Among various distributed energy resources (DERs), PV systems are among the most widely deployed DERs connected at the distribution network level because of their modular structure, scalability, and location-dependent installation characteristics. Therefore, this paper focuses on PV systems and their load-level power contribution characteristics under different output levels and interconnection conditions.

The widespread deployment of DERs has consequently accelerated the transition toward distributed generation architectures that are predominantly interconnected at the distribution network level. As emphasized by the International Renewable Energy Agency (IRENA), the large-scale integration of variable renewables introduces new operational requirements for distribution systems, including enhanced flexibility, advanced grid management, and improved coordination of distributed assets [6]. These developments highlight a structural transformation of the power sector and further underscore the increasing operational importance of distribution networks in accommodating high levels of renewable penetration. Recent studies have also shown that the reliable integration of PV-based resources requires advanced distribution system optimization strategies, particularly under high PV penetration conditions [7].

DERs are predominantly integrated into distribution networks, and the increasing share of DER-based power injected into these networks introduces a variety of operational challenges. These challenges include voltage fluctuations, reverse power flow, line overloading, increased fault current levels, and difficulties in maintaining protection coordination, all of which can adversely affect system reliability and stability [8]. In particular, PV output varies significantly depending on solar irradiance, weather conditions, and operating environments, further complicating real-time operation and long-term planning of distribution systems. Under such conditions, it becomes increasingly important to quantitatively assess how the power generated by PV systems is distributed among individual loads and how this load-level power contribution affects overall system operation. Accurate estimation of PV power contributions is therefore a critical requirement for achieving stable and efficient distribution network operation under high PV penetration levels [9].

Against this background, artificial intelligence and machine learning techniques have been actively investigated to address operational challenges in distribution networks with high DER and PV penetration levels [10,11]. These approaches can learn nonlinear relationships between system operating conditions and DER output variations without explicitly formulating complex network models. Recent studies have applied machine learning techniques to various distribution network applications, including smart-meter-data analysis, fault localization, distribution system management, protection, and disturbance management [12–15].

In addition to data-driven approaches, analytical power allocation and power tracing methods have been used to quantify how power injected by generation sources is distributed to individual loads [16]. These methods can provide physically interpretable source-to-load contribution results and are useful for evaluating DER or PV power contribution under a given operating condition. However, when PV output levels, interconnection locations, or network operating conditions change, power-flow and power-tracing calculations must be repeated for each case. Therefore, the computational burden can increase significantly when numerous PV output and interconnection scenarios are considered.

To reduce computational burden, screening, feature selection, and scenario reduction concepts have been used to eliminate redundant variables or non-informative cases in power system studies. Practical DER interconnection studies have also emphasized the need for efficient evaluation procedures because increasing interconnection requests can expand the number of system impact assessment cases [17]. However, existing approaches mainly focus on general system impact assessment, representative scenario selection, or operational classification, rather than identifying the effective influence range of PV power contribution at the load level. Therefore, a screening criterion specifically designed for load-level PV active power contribution estimation is required.

In this paper, a screening-based machine learning framework is proposed to efficiently estimate load-level PV active power contributions in distribution networks. The objective of the proposed method is to predict how the active power generated by an additionally integrated PV system is allocated to individual loads under different PV output levels and interconnection configurations. The proposed framework uses power-flow and power-tracing results as reference data and applies a screening procedure to identify impact-relevant buses and interconnection cases before machine learning training. The machine learning model receives the PV active power output and interconnection location variables as input and returns the estimated load-level PV active power contribution expressed in kW.

The proposed screening process is distinguished from general feature selection or scenario reduction approaches because it is directly linked to load-level PV active power contribution estimation. In the full-scale approach, all candidate PV output levels and interconnection configurations are simulated, and power-flow and power-tracing calculations are repeatedly performed to obtain the PV contribution at each load. Although this approach provides detailed reference results, the number of simulation cases increases rapidly as the number of candidate interconnection buses and PV output levels increases. In contrast, the proposed method first identifies the effective impact range of the additional PV system and retains only the buses and interconnection cases that produce meaningful PV contribution effects. These selected cases are then used to construct the screening-scale dataset. The learning model uses the PV output level and interconnection location variables as input and estimates the load-level active power contribution supplied by the additional PV system in kW. Therefore, the proposed method is not a general data reduction technique; rather, it is a contribution-oriented screening framework for estimating how additional PV power is allocated to individual loads.

The main contributions of this paper are summarized as follows. First, a load-level PV active power contribution estimation framework is developed by considering both PV output levels and interconnection configurations. Second, a PV Impact-Range Index is introduced to identify the effective source-to-load impact range of the additional PV system by selecting buses and interconnection cases that have meaningful PV contribution effects. Third, the selected impact-relevant cases are used to construct a screening-scale training dataset, through which the machine learning model estimates the load-level active power contribution of the additional PV system without relying on exhaustive full-scale simulations. Finally, the proposed method is validated by comparing its prediction accuracy and computational efficiency with those of the full-scale learning approach and PF-based power tracing results. The overall workflow of the proposed algorithm is summarized in Figure 1.

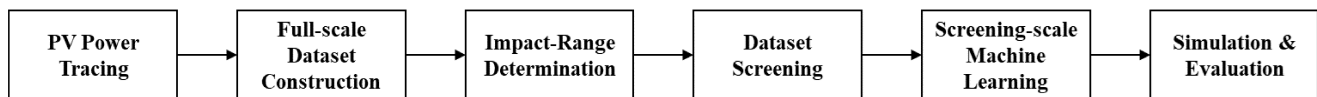


Figure 1. Overall Workflow of the Proposed Framework.

The proposed algorithm consists of a sequential process that integrates simulation-based data generation, screening-based dataset reduction, and machine-learning-based estimation. First, a simplified 19-bus distribution test system representing a practical Korean distribution line (DL) configuration is established, and a full-scale reference dataset is constructed by generating multiple simulation scenarios with varying PV output levels and interconnection configurations [18]. Based on this dataset, a screening procedure is applied to identify network elements and interconnection cases that exhibit significant PV power contribution characteristics. These selected cases are then used to construct a screening-scale training dataset. Subsequently, the screening-scale dataset is employed for machine learning training, and the resulting predictions are evaluated through comparison with those obtained from the full-scale learning method in terms of prediction accuracy and computational efficiency.

2. Methodology

2.1. Machine Learning Dataset Construction

The full-scale dataset is constructed as a reference dataset for evaluating the proposed screening-based PV contribution estimation method. This dataset is designed to include various PV output levels and interconnection configurations, thereby representing diverse load-level PV power contribution characteristics under different simulation conditions. Since the full-scale dataset is generated through repeated power-flow and power-tracing calculations for all candidate PV conditions considered in this paper, it provides a comprehensive basis for comparison with the proposed screening-based approach. The full-scale dataset serves two primary purposes in this paper. First, it provides reference data for validating the prediction accuracy of the proposed screening-based learning approach. Second, it serves as the initial data pool from which impact-relevant buses and interconnection cases are identified and extracted during the dataset screening stage.

2.1.1. DER Power Tracing Method

The DER power tracing method adopted in this paper is based on the proportional sharing rule (PSR), which enables quantitative analysis of load-level DER power contributions at individual buses. PSR assumes that the total incoming power at each bus is allocated to downstream branches in proportion to their respective power-flow magnitudes, thereby allowing the active power originating from DERs to be traced throughout the network [19].

To apply power tracing, detailed information on the locations and capacities of DER sources, load demands, and the electrical connectivity of buses and lines must be available. In addition, line power-flow results are required and are obtained through power-flow analysis before performing the tracing procedure.

$$P_i = \sum_{j \in \alpha_i^{(n)}} |P_{j-i}| + P_{DER_i} \quad (1)$$

Equation (1) calculates the total power injection at each bus by combining the active power flowing into the bus through connected lines and the active power generated by the DER installed at the same bus. The term P_{j-i} represents the active power flow from adjacent bus j to bus i , and the summation term represents the total incoming active power

at bus i . Here, i denotes the index of the bus being evaluated, and j refers to the adjacent buses electrically connected to bus i . The set $\alpha_i^{(n)}$ indicates the group of neighboring buses linked to bus i through the n -th branch. The second term, P_{DERi} , corresponds to the active power generated by the DER installed at bus i .

$$[A_p]^{-1} = \begin{cases} 1 & \text{for } i = j \\ -\frac{P_{j-i}}{P_j} & \text{for } j \in \alpha_i^{(n)} \\ 0 & \text{for otherwise} \end{cases} \quad (2)$$

In Equation (2), the coefficient matrix represents the proportional sharing relationship among buses based on the active power-flow direction. The diagonal elements of the matrix are set to one, indicating the self-contribution term of each bus. The off-diagonal elements are defined only when two buses are electrically connected, and active power is transferred between them. In this case, the off-diagonal coefficient is calculated as the ratio of the branch active power flow to the total incoming active power of the receiving bus according to the proportional sharing rule. If there is no direct electrical connection or no active power transfer between two buses, the corresponding matrix element is set to zero. Therefore, the coefficient matrix expresses how injected power is propagated through the network according to the electrical connectivity and power-flow distribution.

$$P_i = \sum_{i=1}^n \left([A_p]_{ik}^{-1} \times P_{DERi} \right) \quad (3)$$

Equation (3) computes the amount of DER-generated power that ultimately reaches each bus i , based on the contribution allocation defined in Equation (2). This formulation enables a quantitative evaluation of how injected power propagates through the system.

$$P_{Distribution} = \sum_{i=1}^n \left(\frac{P_{DERi}}{P_i} \times [A_p]_{ik}^{-1} \times P_{Loadi} \right) \quad (4)$$

Equation (4) represents the distribution results from the DER perspective, determining how much power from each generating unit is ultimately allocated to individual loads. By computing the DER-supplied power at each load bus, the final contribution outcome can be systematically derived. Although the tracing formulation can be applied to general DERs, this paper applies it to quantify the load-level contribution of the additionally integrated PV system.

In this paper, the power tracing method is used as a physically interpretable reference method to generate training data and validate the prediction results of the machine learning model. Although power tracing can directly calculate PV contributions for a given case, its repeated application requires power-flow analysis and recalculation of tracing coefficients whenever the PV output level or interconnection configuration changes. Therefore, the proposed machine learning model is trained using the power tracing results to rapidly estimate load-level PV contributions under different PV output levels and interconnection configurations, thereby reducing the computational burden of repeated scenario evaluations.

2.1.2. Full-Scale Dataset Construction

To construct the full-scale dataset, the power tracing procedure described in the previous subsection is repeatedly applied under various PV output levels and interconnection conditions. Multiple simulation cases are generated by systematically varying the output level of the additional PV system and its interconnection configuration with the existing

network. For each simulation case, power-flow analysis is performed first, followed by power tracing to obtain the corresponding PV contribution results. The overall procedure for full-scale dataset construction is illustrated in Figure 2.

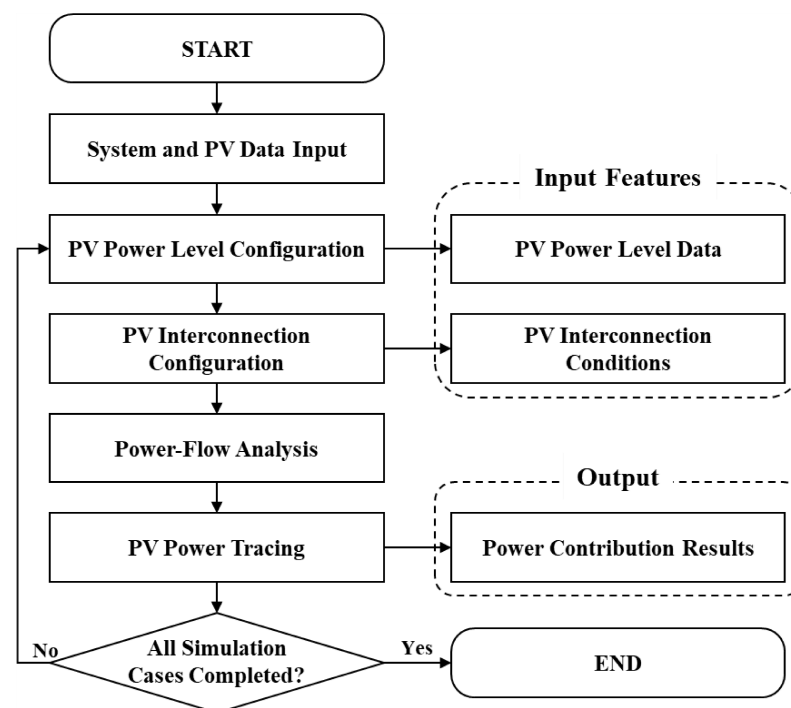


Figure 2. Full-Scale Dataset Construction Procedure for PV Power Contribution Estimation.

Following the procedure illustrated in Figure 2, each simulation case is represented by an input feature vector that explicitly reflects the output level of the additional PV system and its interconnection configuration with the existing network. Accordingly, the full-scale input feature vector is defined as

$$\mathbf{x}^{\text{full-scale}} = [P_{PV}, c_1, c_2, \dots, c_N] \quad (5)$$

where P_{PV} denotes the active power output of the additionally integrated PV system, and $c_n (n = 1, 2, \dots, N)$ is a binary connectivity variable indicating whether the additional PV system is connected to the n -th bus, with $c_n = 1$ if the PV system is connected to bus n and $c_n = 0$ otherwise. Therefore, the input feature vector is composed of the PV output condition represented by P_{PV} and the interconnection location information represented by the binary connectivity vector $[c_1, c_2, \dots, c_N]$. This input structure allows the machine learning model to learn how both the PV output level and the interconnection location affect the load-level PV contribution pattern.

Based on this input feature representation, the output component of the full-scale dataset is defined as

$$\mathbf{y}^{\text{full-scale}} = [\mathbf{p}^{\text{EX}_1}, \mathbf{p}^{\text{EX}_2}, \dots, \mathbf{p}^{\text{EX}_k}, \mathbf{p}^{\text{PV}}] \quad (6)$$

$$\begin{cases} \mathbf{p}^{\text{EX}_k} = [P_1^{\text{EX}_k}, P_2^{\text{EX}_k}, \dots, P_{N_L}^{\text{EX}_k}] \\ \mathbf{p}^{\text{PV}} = [P_1^{\text{PV}}, P_2^{\text{PV}}, \dots, P_{N_L}^{\text{PV}}] \end{cases}$$

Here, $\mathbf{p}^{\text{EX}_k}$ denotes the active power contribution from the k -th existing power source to each load, and $\mathbf{p}^{\text{PV}}$ represents the load-level power contribution of the additionally integrated PV system. Each element of the output vector is obtained from the power tracing results and quantitatively describes the amount of active power contributed by each source

to a specific load. In the proposed estimation model, the primary output of interest is $\mathbf{P}^{\text{PV}}$, where each element P_i^{PV} represents the active power contribution supplied by the additional PV system to load bus i . The unit of P_i^{PV} is kW. In the result figures, the PV contribution is additionally expressed as a percentage of the corresponding load demand to facilitate interpretation of the load-level contribution.

2.1.3. Dataset Screening Procedures

The dataset screening process is designed to reduce the number of repeated simulations for different interconnection cases by identifying the effective impact range of an additional PV system prior to machine learning training. Rather than exhaustively generating simulation data for all buses and interconnection configurations, the proposed approach focuses on selecting a subset of network elements that are significantly affected by PV power injection, thereby preserving the dominant PV contribution characteristics while reducing the computational burden. An illustrative example of the proposed dataset screening concept is presented in Figure 3.

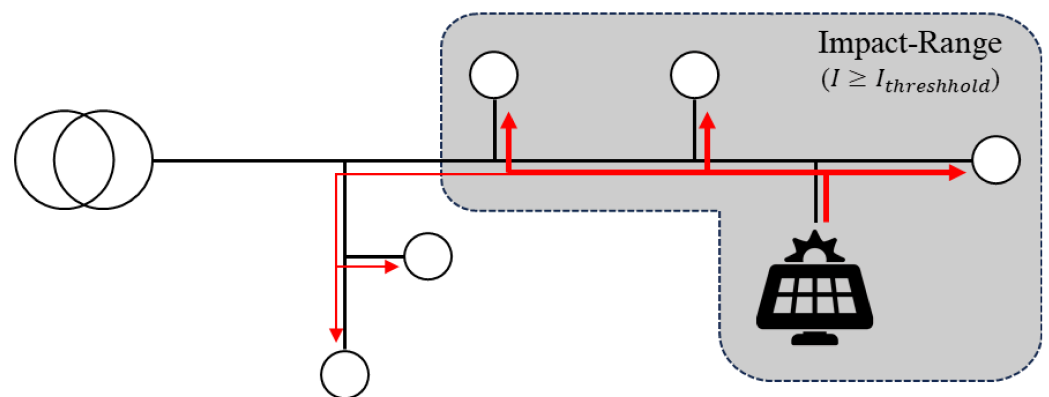


Figure 3. Conceptual illustration of PV impact-range identification. Red arrows indicate PV power delivery paths, solid lines represent feeder connections, the dashed boundary denotes the impact range, and circle sizes indicate relative PV contribution magnitudes.

Figure 3 conceptually illustrates how the power generated by an additional PV system is delivered through the distribution network and results in different levels of power contribution at individual load buses. As the injected PV power flows through the network, it is redistributed according to the electrical connectivity and power-flow paths, leading to non-uniform PV contribution levels across load buses. Load buses electrically closer to the PV interconnection point receive a relatively larger portion of PV-supplied power, whereas buses located farther away along indirect or weakly coupled paths exhibit progressively reduced contribution levels. Based on these bus-level differences in PV contribution magnitude, only buses exhibiting contribution levels above a certain threshold are defined as belonging to the effective impact range of the PV system. In Figure 3, the shaded region represents the set of buses where PV-generated power reaches the loads with sufficient magnitude to produce a meaningful operational contribution, while buses outside this region experience negligible PV contributions and can therefore be excluded during the screening process. This example demonstrates that PV contributions are not uniformly distributed across the system but are governed by the power delivery process itself, enabling a systematic reduction in interconnection-based simulation cases for dataset construction.

$$\text{Index}_i = \omega_P \left| \frac{\partial P_i}{\partial P_{PV}} \right| + \omega_V \left| \frac{\partial |V_i|}{\partial P_{PV}} \right| + \omega_d (1 - d_i^{\text{norm}}) \quad (7)$$

In Equation (7), Index_i represents a quantitative measure used to identify the effective impact range of the additional PV system on bus i . The first term quantifies the sensitivity of active power contribution redistribution with respect to variations in PV output, reflecting how changes in PV injection affect the power-flow distribution within the network. The second term represents voltage sensitivity, capturing the response of the bus voltage magnitude to PV output variations. The third term incorporates the normalized electrical distance between the PV interconnection point and bus i , such that buses electrically closer to the PV system are assigned higher index values.

The weighting coefficients ω_P , ω_V , and ω_d determine the relative importance of active power contribution, voltage sensitivity, and electrical proximity, respectively. Therefore, the PV Impact-Range Index can be adjusted depending on which physical factor is emphasized in the screening process. The screening threshold is then used to determine whether each bus is included in the effective impact range of the additional PV system.

Since the weighting coefficients and threshold directly affect the size of the screened dataset and the resulting prediction accuracy, their values should be selected carefully. In this paper, these screening parameters are determined through the sensitivity analysis presented in Section 3.4.1. Through this formulation, the PV Impact-Range Index provides a physically interpretable criterion for identifying the buses that have meaningful PV contribution effects.

2.2. Machine Learning Training Procedure

2.2.1. Random Forest Regressor

A supervised learning model is trained using the screening-scale dataset to estimate load-level PV power contributions. In this paper, a Random Forest Regressor is adopted as the learning model because the PV contribution estimation problem involves nonlinear and topology-dependent relationships between PV output, interconnection location variables, and load-level contribution results. In addition, the input feature vector includes both continuous PV output values and binary interconnection variables, and Random Forest can handle such mixed feature structures with relatively limited preprocessing. The overall learning structure of the Random Forest model employed in this work is illustrated in Figure 4.

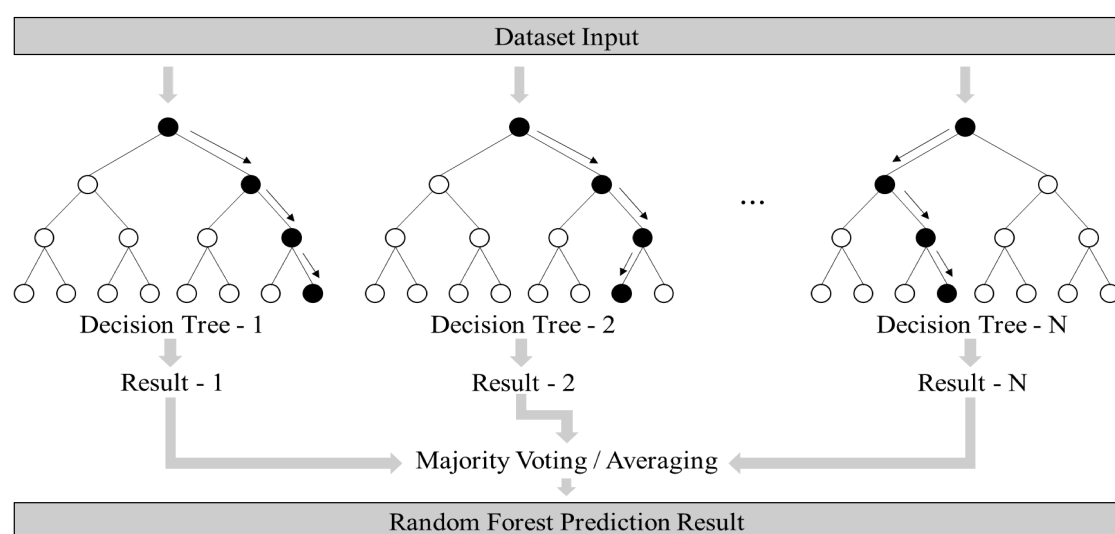


Figure 4. Learning Structure of the random forest prediction model. Black and white circles represent selected and non-selected nodes in each decision tree, respectively.

As shown in the figure, the input dataset representing PV operating conditions and network interconnection configurations is distributed across multiple decision trees. Each tree is trained independently using different subsets of the input samples and feature space, enabling diverse representations of PV-induced power contribution patterns within the distribution network. The prediction process of an individual decision tree can be expressed by Equation (8).

$$h(x) = \sum_{m=1}^M c_m \cdot 1(x \in R_m) \quad (8)$$

In this formulation, the input space is partitioned into multiple regions, and a constant prediction value is assigned to each region. When an input sample x belongs to a specific region R_m , the corresponding coefficient c_m is selected as the output of the decision tree.

The final output of the Random Forest model is obtained by aggregating the predictions of all individual trees, as defined in Equation (9).

$$\hat{y}(x) = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (9)$$

For a given input sample, each decision tree produces an individual prediction of the PV power contribution to each load. These predictions are then aggregated through ensemble averaging to generate the final output of the Random Forest model. Through this ensemble-based learning process, the model effectively captures the nonlinear and topology-dependent relationships between PV injection conditions and load-level power contribution characteristics. By training the Random Forest using the screening-scale dataset, reliable estimation performance is achieved with significantly reduced computational complexity compared to full-scale learning.

2.2.2. PV Contribution Estimation Method

The proposed PV contribution estimation method evaluates the load-level power contribution of an additionally integrated PV system in an efficient and interpretable manner. Unlike conventional full-scale approaches that analyze power contributions from all generation sources, the proposed method focuses exclusively on the redistribution of power supplied by the additional PV. This formulation is based on the observation that PV-induced power redistribution does not uniformly affect the entire network but is instead concentrated along specific interconnection paths and localized regions.

To exploit this characteristic, the proposed method first identifies the effective impact range of the additional PV using an index-based screening process. By evaluating the sensitivity of system variables to PV injection under single-bus interconnection conditions, only buses and interconnection cases with meaningful PV contributions are retained for further analysis. As a result, the analysis is reformulated to operate on a reduced set of impact-relevant simulations, while preserving the dominant characteristics of PV-induced power contribution. The overall procedure of the proposed PV contribution estimation method is illustrated in Figure 5.

The proposed method consists of three main phases, as shown in Figure 5. In the first phase, the additional PV system is iteratively interconnected to each candidate bus under predefined PV output conditions. For each interconnection case, power-flow analysis is first performed, and the PV Impact-Range Index is then calculated to quantify the localized impact of the additional PV on each bus. This index-based evaluation is repeated for all candidate buses and buses whose index values exceed the predefined threshold are selected as the effective PV impact range.

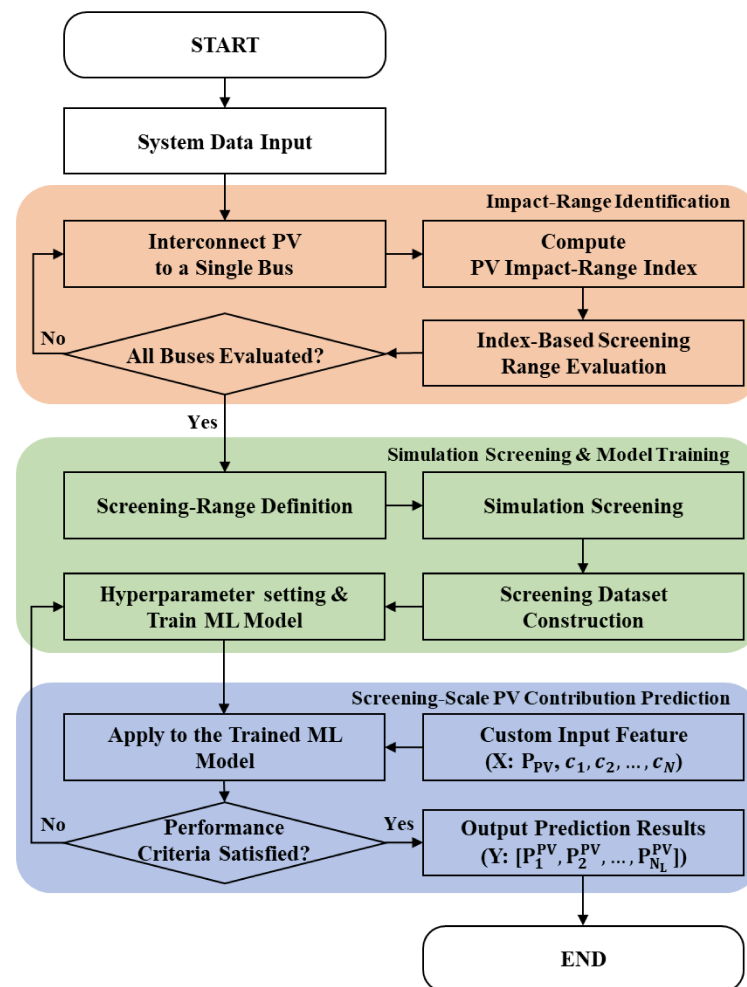


Figure 5. Overall Procedure of the Proposed PV Contribution Estimation Method.

In the second phase, simulation screening is performed based on the identified impact range. Only the impact-relevant buses and interconnection cases are retained, while low-impact cases are excluded from further dataset construction. The selected cases are then used to construct the screening-scale dataset. This dataset consists of input features representing the PV output level and interconnection location variables, along with output values representing the load-level PV active power contribution obtained from the corresponding power tracing results.

In the final phase, the screening-scale dataset is used to train the Random Forest Regressor for PV contribution estimation. Given a user-defined input feature vector, the trained model predicts the load-level active power contribution supplied by the additional PV system. The prediction result is represented as a vector that describes how the power generated by the additional PV is allocated among individual loads. The predicted results are then evaluated using predefined performance criteria to verify whether the screening-scale learning method preserves the dominant PV contribution characteristics without requiring exhaustive full-scale simulations.

3. Simulation

3.1. Simulation System Overview

To estimate the power contribution of an additional PV integrated into distribution system operation, a 19-bus radial distribution network is adopted as the test system [20]. The test system was constructed as a simplified equivalent model derived from the practical

Pyoseon DL in the Republic of Korea [21]. Figure 6 shows the practical configuration of the Pyoseon DL.

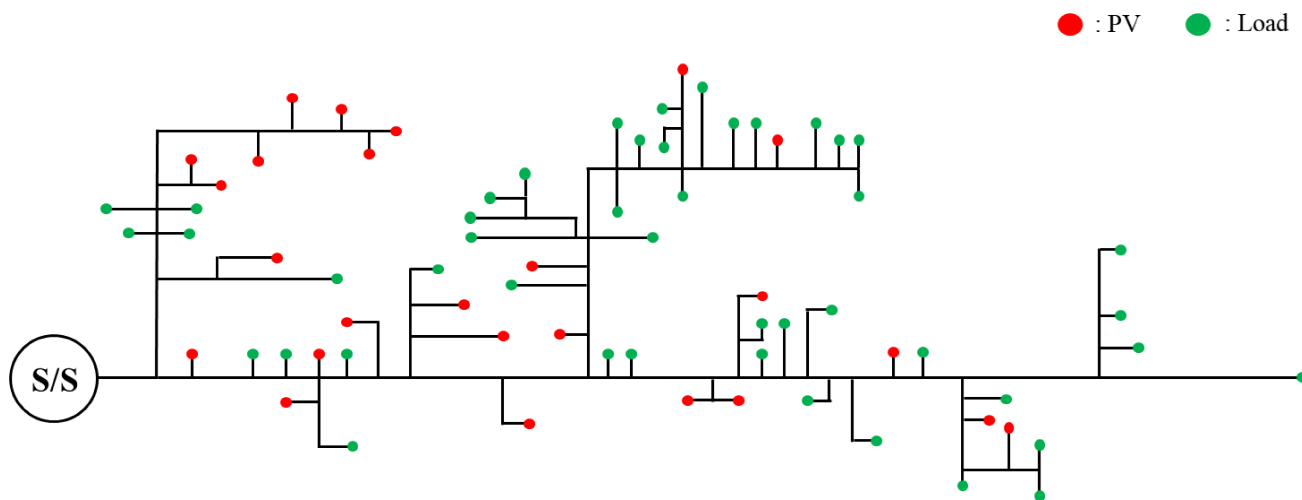


Figure 6. Practical Configuration of the Pyoseon DL.

As shown in Figure 6, the original Pyoseon DL includes multiple lateral branches, load points, PV interconnection points, detailed connection nodes, and intermediate line nodes. The purpose of the simplification was not to reproduce every detailed component of the practical feeder, but to construct a representative equivalent model for evaluating the proposed PV contribution estimation algorithm. The simplification was performed as an equivalent modeling process for constructing a representative test system, rather than as an automated topology reduction algorithm. In the equivalent model, the main radial power-flow paths, representative load locations, and the PV-dominant section were retained, while detailed connection nodes, intermediate line nodes, and minor lateral sections were aggregated or omitted. Table 1 summarizes the configuration of the simplified 19-bus equivalent model derived from the practical Pyoseon DL.

Table 1. Configuration of the Practical Pyoseon DL and Simplified 19-Bus Equivalent Model.

Item	Practical Pyoseon DL	Simplified 19-Bus Equivalent Model	Equivalent Representation
Network topology	291 buses	19 buses	Aggregated into a simplified equivalent model
Line model	262 sections, 39.69 km	18 sections, 21.9 km	
Load model	46 load elements, 11,300 kW	11 load buses, 5500 kW	
PV model	25 PV units, 4375.2 kW	1 PV unit, 2000 kW	

As summarized in Table 1, the original 291-bus feeder was represented by a 19-bus radial equivalent model. Since the simplified system is radial, the 19 retained buses are connected by 18 equivalent line sections. The total line length was represented as 21.9 km in the equivalent model because detailed intermediate sections and minor lateral branches were not represented individually, while the main feeder and major branching paths were preserved.

For the load model, the 46 load elements in the original feeder were represented by 11 representative load buses. The total active power demand used in the simulation was set to 5500 kW, corresponding to approximately 48.7% of the original feeder load.

For the PV model, the 25 existing PV units were represented by one representative PV unit connected to bus 015. The original PV output condition was calculated as 40%

of the total rated PV capacity, or 4375.2 kW. In the simplified model, this condition was represented by a 2000 kW PV unit, corresponding to approximately 45.7% of the original PV output condition.

Figure 7 presents the simplified 19-bus equivalent test system used for the simulation. During the equivalent modeling process, the detailed loads and PV units in the original feeder were aggregated according to the main feeder and major branching sections. Each aggregated section was classified according to its net active power characteristic. Sections in which the local load demand was larger than the existing PV output were represented as load-dominant sections, whereas the section in which the existing PV output was dominant was represented by a representative PV unit.



Figure 7. Simplified 19-bus equivalent test system derived from the Pyoseon DL. Red and green circles indicate PV and load buses, respectively, while black circles denote buses without PV or load connections.

As shown in Figure 7, the simplified 19-bus system consists of 11 representative load buses and one existing PV unit. Under the simulation condition, the total load demand is 5500 kW and the existing PV output is 2000 kW, resulting in a substation-supplied net load of 3500 kW. Although the simplified model does not reproduce every detailed component of the practical feeder, it preserves the main radial topology, representative load locations, PV-dominant section, and dominant source-to-load contribution paths required for load-level PV contribution analysis.

3.2. Simulation Scenarios

The simulation scenarios used to analyze the influence of PV integration in this paper are configured and summarized in Table 2. The scenarios are classified into three categories according to the interconnection relationship between the additional PV and the existing network. To ensure a fair comparison of PV-induced effects under different generation capacities, each scenario is evaluated under two operating cases distinguished by different output ratios relative to the net load.

Table 2. Simulation Scenarios for PV Power Contribution Analysis.

Scenario	Interconnection Type	Case	PV Output Level
Scenario 1	Single-Point	A	30% of Net Load
		B	70% of Net Load
Scenario 2	Multi-Point	A	30% of Net Load
		B	70% of Net Load
Scenario 3	Complex Multi-Point	A	30% of Net Load
		B	70% of Net Load

In general, network operating conditions and PV generation output can affect load-level PV contribution results by changing the power-flow distribution and the electrical range over which PV-supplied power is delivered. In this paper, the network topology and load condition are kept unchanged, while the output level and interconnection structure of the additional PV system are varied as the main scenario variables. This scenario configuration enables a consistent analysis of how PV generation output and interconnection complexity affect load-level PV contribution characteristics.

The simulation scenarios are designed to investigate the effect of PV interconnection complexity on power contribution characteristics in the distribution network. Scenario 1 considers a single-point interconnection in which the additional PV is electrically coupled to only one location in the system, representing a low-complexity integration structure. Scenario 2 extends this configuration to a multi-point interconnection, allowing the PV to interact with the network through multiple connection paths and thereby increasing power-flow interactions. Scenario 3 further examines a complex multi-point interconnection, representing the highest level of interconnection complexity considered in this paper.

Across all scenarios, two operating cases are evaluated. Case A corresponds to a PV output level of 30% of the net load, equivalent to 1000 kW, while Case B corresponds to 70% of the net load, or 2500 kW. Applying identical output conditions across all scenarios enables a consistent assessment of how PV output levels interact with different interconnection configurations and affect overall system performance.

3.3. Simulation Results

To clearly illustrate the load-level power contribution characteristics of the additional PV system, the simulation results are presented using per-bus contribution visualizations. In this visualization, each load bus is represented in a circular format, and the shaded portion indicates the PV-supplied power, while the newly integrated PV unit and its connection paths are highlighted using a red dashed line. The quantitative prediction accuracy of the proposed screening-scale machine learning algorithm is separately evaluated in Section 3.4 using the power-flow-based (PF-based) power tracing results as the reference values.

3.3.1. Scenario 1 Results

Scenario 1 evaluates PV-induced power contribution characteristics under a single-point interconnection configuration. The contribution results for Scenario 1, including both Case A and Case B, are presented in Figure 8.

For Case A, the PV system predominantly supplies power to buses electrically close to the interconnection point, with Buses 006, 007, and 008 exhibiting contribution levels exceeding 95%, indicating that the injected PV power is largely allocated within the downstream feeder segment. Limited contributions are also observed at adjacent buses such as Buses 011 and 012, confirming a localized influence even under a single-point interconnection.

For Case B, a similar spatial pattern is maintained, with the PV system supplying the majority of load demand at Buses 005–008 and adjacent Buses 011 and 012, while Bus 004 receives a moderate contribution. In contrast, buses electrically farther from the interconnection point show negligible PV contributions, indicating that PV-supplied power remains spatially concentrated within a specific feeder region, with contribution levels varying according to operating conditions.

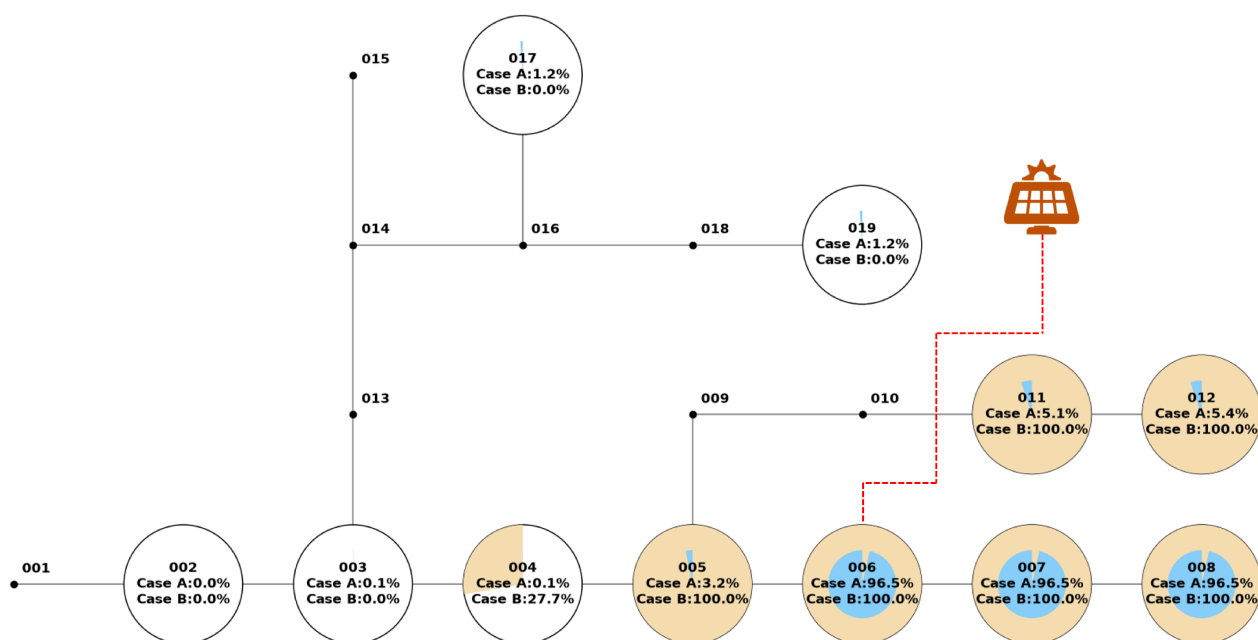


Figure 8. Load-Level PV Power Contribution Results in Scenario 1. Inner blue and outer yellow portions indicate PV contribution levels for Case A and Case B, respectively. White circles represent load buses, black dots denote non-load buses, and the red dashed line indicates the additional PV interconnection path.

3.3.2. Scenario 2 Results

Scenario 2 examines PV power contribution characteristics under a multi-point interconnection configuration. The contribution results for both Case A and Case B are jointly presented in Figure 9.

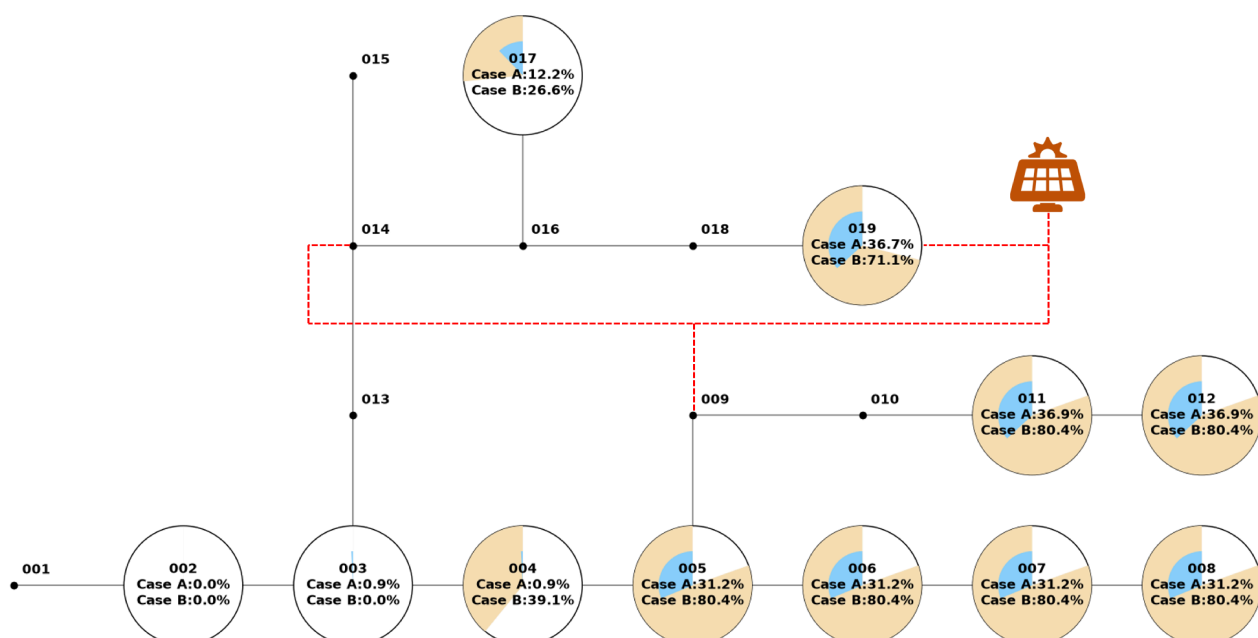


Figure 9. Load-Level PV Power Contribution Results in Scenario 2. Inner blue and outer yellow portions indicate PV contribution levels for Case A and Case B, respectively. White circles represent load buses, black dots denote non-load buses, and the red dashed line indicates the additional PV interconnection path.

Under Case A, PV-supplied power is distributed across multiple feeder branches. Several buses, including Buses 005–008, 011–012, 017, and 019, receive approximately 31–37% of their load demand from the PV system, indicating that the PV contribution extends beyond a single downstream segment.

When the PV output level increases to Case B, the PV contribution becomes more pronounced. Buses 005–008, 011–012, and 019 receive approximately 70–80% of their load demand from the PV system, while Bus 017 shows a moderate contribution of about 27%, confirming a broader allocation of PV-supplied power under the multi-point interconnection structure.

3.3.3. Scenario 3 Results

Scenario 3 evaluates PV power contribution characteristics under a complex multi-point interconnection configuration, representing the highest level of interconnection complexity considered in this paper. The contribution results for both Case A and Case B are jointly presented in Figure 10.

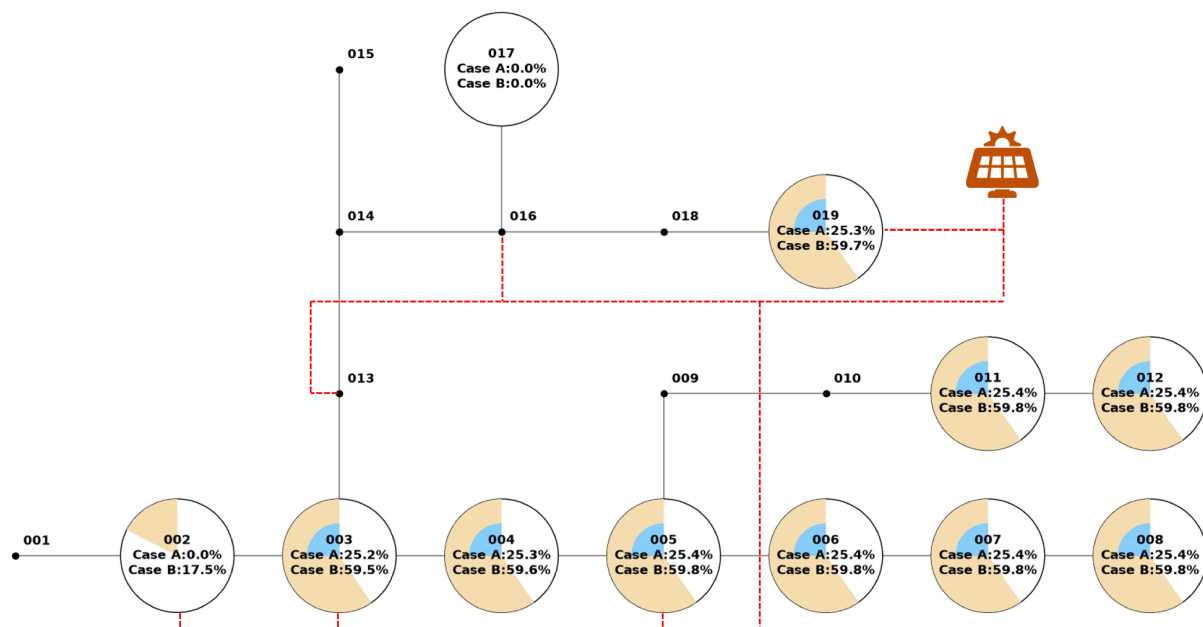


Figure 10. Load-Level PV Power Contribution Results in Scenario 3. Inner blue and outer yellow portions indicate PV contribution levels for Case A and Case B, respectively. White circles represent load buses, black dots denote non-load buses, and the red dashed line indicates the additional PV interconnection path.

As shown in Figure 10, under Case A, the PV contribution is widely distributed across the feeder. A relatively uniform contribution level is observed at multiple downstream and lateral branch buses, including Buses 003–008, 011–012, and 019, indicating that PV-supplied power is shared across a broad interconnected area rather than being concentrated near a specific location.

Under Case B, the PV contribution increases across most downstream and lateral branch buses. Buses 003–008, 011–012, and 019 receive a substantial portion of their load demand from the PV system, and a limited contribution is also observed at Bus 002, indicating that the contribution range extends further upstream as the PV output increases. In contrast, Bus 017 shows no PV contribution because its demand is already supplied by the existing PV connected at Bus 015.

These results indicate that the complex multi-point interconnection configuration enables the widest distribution of PV-supplied power among the three scenarios, while the contribution pattern is still affected by existing local generation within the feeder.

The significance of these results is that the proposed method quantitatively identifies not only the loads directly supplied by the additional PV system but also the effective spatial range over which the PV contribution is distributed under different output levels and interconnection configurations. In a radial distribution network, the concentration of PV-supplied power around electrically close load buses is physically reasonable because active power contribution follows the network topology and power-flow paths. However, the results also show that the contribution range can expand to multiple feeder branches or upstream buses depending on the interconnection structure and PV output level. Therefore, the proposed method can be used to evaluate candidate PV interconnection locations, identify load buses significantly affected by newly integrated PV systems, and support distribution network operation and planning by providing fast load-level PV contribution estimates.

3.4. Performance Evaluation

The performance of the proposed screening-scale machine learning algorithm was evaluated from three perspectives: screening parameter sensitivity, prediction accuracy, and computational efficiency. To improve the reproducibility of the machine learning procedure, the hyperparameter settings of the Random Forest Regressor are summarized in Table 3. The same hyperparameter settings were applied to both the full-scale and screening-scale learning models to ensure a fair comparison. No additional hyperparameter optimization was performed separately for each dataset.

Table 3. Hyperparameter Settings of the Random Forest Regressor.

Hyperparameter	Value
Number of decision trees	500
Maximum depth	None
Bootstrap sampling	Enabled
Number of jobs	1
Criterion	Squared error
Random seed	42

3.4.1. Sensitivity Analysis of Screening Parameters

To justify the weighting coefficients and screening threshold used in the PV Impact-Range Index, a sensitivity analysis was conducted by configuring six parameter cases, as summarized in Table 4.

Table 4. Sensitivity Cases for Screening Parameter Selection.

Case	Weighting Condition	Threshold	Purpose
1	$\omega_P = 0.33, \omega_V = 0.33, \omega_d = 0.33$	0.1	Equal-weight condition
2	$\omega_P = 0.1, \omega_V = 0.8, \omega_d = 0.1$	0.1	Voltage-dominant condition
3	$\omega_P = 0.1, \omega_V = 0.1, \omega_d = 0.8$	0.1	Distance-dominant condition
4	$\omega_P = 0.8, \omega_V = 0.1, \omega_d = 0.1$	0.1	Power-dominant condition
5	$\omega_P = 0.8, \omega_V = 0.1, \omega_d = 0.1$	0.3	Higher threshold condition
6	$\omega_P = 0.8, \omega_V = 0.1, \omega_d = 0.1$	0.5	Stronger threshold condition

The sensitivity cases were designed to evaluate the effects of both the weighting condition and the screening threshold on prediction accuracy and computational efficiency. Cases

1–3 were configured with the same threshold value of 0.1, but with different weighting conditions, including equal-weight, voltage-dominant, and distance-dominant settings. Case 4 represents the power-contribution-dominant condition with the same threshold value of 0.1. Cases 5 and 6 were then configured by increasing the threshold to 0.3 and 0.5 under the power-contribution-dominant weighting condition.

The prediction accuracy results for the six sensitivity cases are shown in Figure 11. Both RMSE and R^2 score were compared to evaluate the effect of the weighting condition and screening threshold on the estimation performance.

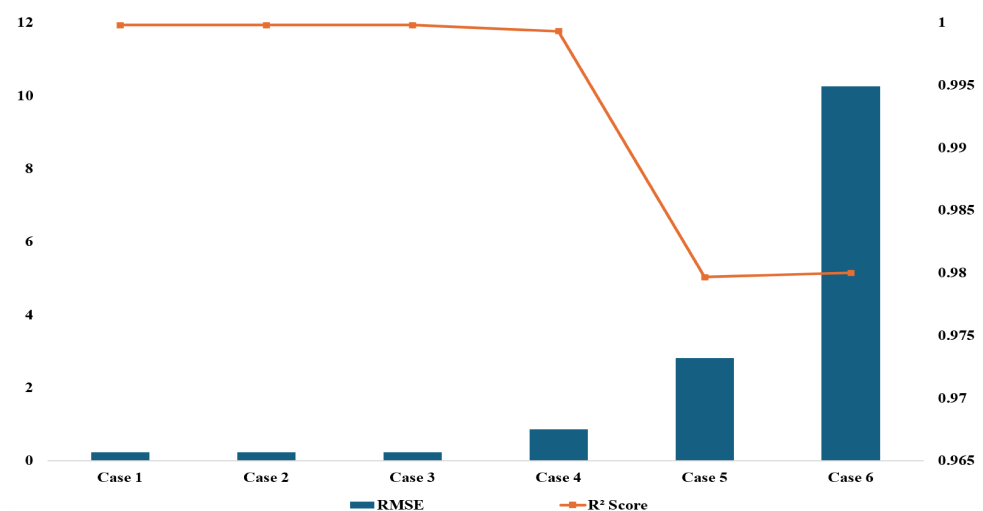


Figure 11. Case-wise RMSE and R^2 score comparison according to screening sensitivity case.

As shown in Figure 11, Cases 1–3 show the lowest RMSE value and an R^2 score of 0.9998, because these conditions retain most of the network elements and produce results close to the full-scale case. Case 4 shows a slightly higher RMSE of 0.8605, but its R^2 score remains very high at 0.9993. This indicates that the power-contribution-dominant condition preserves the overall PV contribution pattern while reducing the dataset scale. In contrast, Cases 5 and 6 show larger RMSE values as the threshold increases, while their R^2 scores are lower than those of Cases 1–4. This indicates that excessive screening can remove meaningful PV contribution information and degrade prediction performance.

The dataset generation time for each sensitivity case is compared in Figure 12 to evaluate the computational efficiency of the screening conditions.

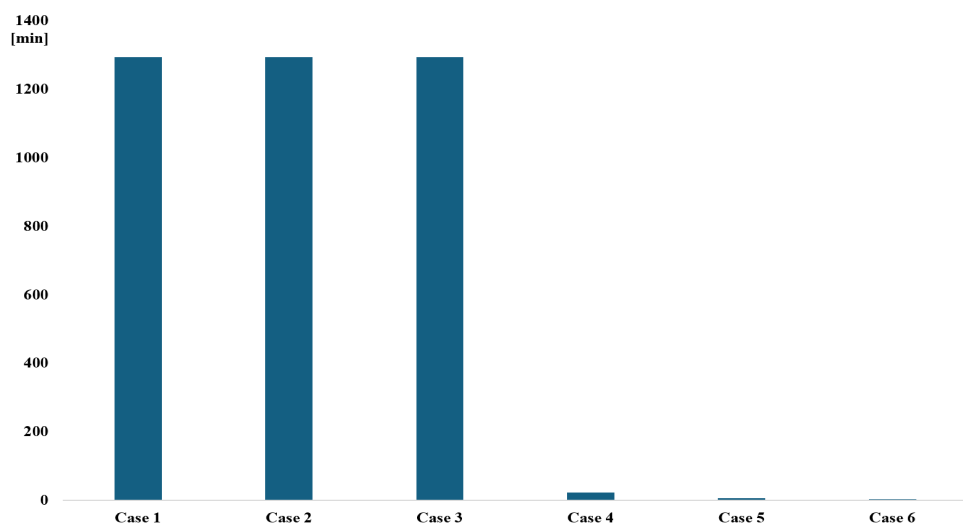


Figure 12. Case-wise dataset generation time for screening parameter selection.

As shown in Figure 12, Cases 1–3 require the same dataset generation time as the full-scale case, indicating that these weighting conditions do not provide meaningful screening reduction. Case 4 reduces the dataset generation time from 21 h 34 min 15 s to 21 min 34 s, corresponding to an approximately 60-fold reduction. Although Cases 5 and 6 further reduce the computational time, their prediction accuracy decreases, as shown in Figure 11.

Based on these results, Case 4 was selected as the final screening condition. Cases 1–3 provide the highest prediction accuracy, but they do not reduce the dataset generation time. Cases 5 and 6 provide shorter dataset generation times, but their prediction errors increase due to excessive screening. In contrast, Case 4 maintains a high R^2 score of 0.9993 while significantly reducing the dataset generation time. Therefore, Case 4 provides the best balance between prediction accuracy and computational efficiency. The weighting coefficients and threshold were set to $\omega_P = 0.8$, $\omega_V = 0.1$, $\omega_d = 0.1$, and threshold = 0.1. The selected parameter setting was used for the subsequent prediction accuracy and computational efficiency evaluations.

3.4.2. Prediction Accuracy Evaluation

To quantitatively evaluate the prediction accuracy of the proposed screening-scale machine learning algorithm, the estimation results are compared with two baseline methods: the PF-based power tracing method and the full-scale machine learning model. The PF-based power tracing method is used as the physics-based baseline and reference solution because it directly calculates load-level PV contribution through repeated power-flow and power-tracing calculations. The full-scale machine learning model is used as the data-driven baseline because it is trained using the complete dataset without the proposed screening process. In this comparison, the normalized absolute error (NAE) is defined as the ratio of the total absolute prediction error to the total PF-based PV contribution value, thereby mitigating the distortion that can occur when very small reference values are included in the error calculation. The prediction accuracy comparison results for each simulation scenario and operating case are summarized in Table 5.

Table 5. Comparison of Normalized Absolute Error Based on PF-Based Power Tracing Results.

Scenario	Case	Full-Scale ML NAE [%]	Screening-Scale ML NAE [%]
Scenario 1	A	14.3195	6.0588
	B	0.0034	0.0029
Scenario 2	A	0.0213	1.5851
	B	0.4078	3.8451
Scenario 3	A	0.0084	0.0558
	B	0.0151	0.0089

As summarized in Table 5, the proposed screening-scale machine learning model maintains low NAE values across all simulation scenarios. In Scenario 1, Case A shows a relatively high NAE in the full-scale ML model. This result is mainly attributed to the low PV output level and the localized contribution pattern under the single-point interconnection condition. In Case A, the PV contribution is concentrated around a limited downstream feeder segment, and several contribution values are relatively small. Therefore, even small absolute prediction deviations can result in a larger normalized error. In addition, the feature-importance analysis showed that the interconnection location variables associated with the local downstream buses in Scenario 1 had relatively lower importance than PMPP and several upstream or PV-adjacent bus variables. This indicates that the model was less sensitive to localized contribution variations under the low-output condition. In contrast, under Case B, the higher PV output produces a clearer and more dominant

contribution pattern, allowing the model to estimate the PV contribution with a much lower NAE.

In Scenario 2, the screening-scale model shows relatively larger prediction errors than the full-scale model, particularly in Case B. This increase is mainly attributed to localized prediction deviations at specific load buses rather than to the overall interconnection complexity itself. Under the multi-point interconnection condition, the PV contribution to certain loads becomes sensitive to branch-level power-flow redistribution characteristics, and part of this localized interaction information may be underrepresented in the screened dataset. As a result, relatively larger deviations are observed at several sensitive load buses, leading to an increase in the NAE value. Nevertheless, the screening-scale model maintains an NAE value below 4%, indicating that the dominant PV contribution distribution characteristics are still well preserved.

In Scenario 3, both the full-scale and screening-scale models exhibit very small prediction errors. This result confirms that the proposed screening-scale model can accurately reproduce the widely distributed PV contribution pattern even under the complex multi-point interconnection configuration. In particular, the relatively uniform PV contribution distribution observed in Scenario 3 enables the screening-scale model to capture the overall spatial distribution trend without significant degradation in prediction performance. Overall, these results demonstrate that the proposed screening-based learning approach can preserve the dominant PV contribution distribution characteristics while substantially reducing the dataset scale required for model training.

3.4.3. Computational Efficiency Evaluation

After selecting Case 4 as the final screening condition, the computational efficiency of the resulting screening-scale model was compared with that of the full-scale learning approach. A quantitative comparison between the screening-scale and full-scale machine learning approaches is summarized in Table 6.

Table 6. Performance Comparison of Screening-Scale and Full-Scale Machine Learning Models.

Performance	Screening-Scale ML Model	Full-Scale ML Model
Dataset Generation Time	21 m 34 s	21 h 34 m 15 s
Model Training Time	2 m 21 s	13 m 4 s
Root Mean Square Error	0.8605	0.2406
R ² Score	0.9993	0.9998

In terms of dataset generation time, the screening-scale approach requires only 21 min and 34 s, whereas the full-scale approach requires 21 h, 34 min, and 15 s. This corresponds to an approximately 60-fold reduction in dataset construction time, clearly demonstrating the effectiveness of the proposed index-based screening strategy. By restricting dataset generation to impact-relevant system elements, the proposed approach substantially reduces the number of repetitive power-flow and power-tracing simulations that dominate the computational cost of conventional full-scale analysis.

A similar reduction is observed in model training time. The screening-scale learning requires only 2 min and 21 s, compared with 13 min and 4 s for the full-scale learning. These results indicate that the reduction in dataset size not only accelerates dataset generation but also alleviates the computational burden of the learning process, while retaining the essential information required to characterize PV-induced power contribution behavior.

With respect to prediction accuracy, the screening-scale model exhibits a higher RMSE than the full-scale model. However, considering that the predicted load-level PV power

contribution values are generally on the order of several hundred kW, the resulting deviation remains relatively small. In addition, both approaches maintain high coefficients of determination, indicating that the overall distribution trends and spatial patterns of PV power contributions across loads are well preserved despite the increase in absolute error. These results demonstrate that the proposed screening-scale approach achieves a substantial improvement in computational efficiency while maintaining sufficient prediction accuracy for practical PV power contribution estimation.

4. Conclusions

High penetration of PV systems introduces increasing complexity in distribution system operation, making it challenging to accurately assess load-level PV power contributions using conventional simulation-based approaches due to excessive computational burden. To address this challenge, this paper proposes a screening-based machine learning algorithm for efficient PV power contribution estimation in distribution networks.

The proposed approach is motivated by the observation that PV-induced power redistribution does not occur uniformly across the network but is instead concentrated along specific interconnection paths and localized regions. A PV Impact-Range Index was introduced to quantify the effective influence range of an additional PV system by jointly considering power sensitivity, voltage sensitivity, and electrical distance. Based on this index, impact-relevant buses were systematically identified under single-bus interconnection conditions, enabling effective screening of simulation cases prior to dataset construction. This process significantly reduced repetitive power-flow and power-tracing simulations while preserving the dominant characteristics of PV-induced power contribution patterns.

Using the screened dataset, a Random Forest-based supervised learning model was developed to predict the load-level power contributions of an additional PV system from its output levels and interconnection configurations. Simulation results on a 19-bus test system demonstrated that the proposed screening-scale learning approach successfully reproduced the spatial distribution trends of PV power contributions obtained from full-scale learning and PF-based power tracing, while achieving substantial reductions in dataset generation and model training time. These load-level contribution estimates can be used to assess candidate PV interconnection locations, identify loads significantly affected by additional PV integration, and support distribution network operation and planning.

The proposed method assumes that network topology and system parameters, including bus-line connectivity, feeder configuration, line impedance, load locations, PV interconnection points, load demand, and DER capacity, are available and sufficiently accurate. Therefore, the current validation is limited to the topology, load condition, PV output range, and interconnection configurations represented in the simulation dataset. If these data are incomplete, inaccurate, or outdated, the accuracy of the power-flow analysis, power-tracing results, and screening outcome may be affected.

Future work will extend the proposed algorithm to scenarios involving multiple simultaneously interconnected PV systems and time-varying operating conditions. In addition, the applicability of the proposed algorithm will be further investigated using larger-scale distribution networks and real measurement data, as well as through comparisons with alternative machine learning models to enhance robustness and generality. Furthermore, the proposed framework may be extended to other types of DERs to evaluate its applicability beyond PV-based contribution estimation.

Author Contributions: Conceptualization, Y.-J.J. and Y.-H.C.; data curation, Y.-J.J.; formal analysis, Y.-J.J.; investigation, Y.-J.J.; methodology, Y.-J.J.; project administration, Y.-H.C.; resources, Y.-H.C.; software, Y.-J.J.; supervision, Y.-H.C.; validation, Y.-J.J. and Y.-H.C.; visualization, Y.-J.J.; writing—original draft preparation, Y.-J.J.; writing—review and editing, Y.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Climate, Energy & Environment (MCEE) of the Republic of Korea (No. RS-2026-25537902).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. International Energy Agency. *Global Energy Review 2025*; International Energy Agency: Paris, France, 2025.
2. International Energy Agency. *Electricity 2025: Analysis and Forecast to 2027*; International Energy Agency: Paris, France, 2025.
3. Shehabi, A.; Smith, S.J.; Hubbard, A.; Newkirk, A.; Lei, N.; Siddik, M.A.B.; Holecek, B.; Koomey, J.; Masanet, E.; Sartor, D. 2024 *United States Data Center Energy Usage Report*; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2024.
4. Poudineh, R. *Global Electricity Demand: What's Driving Growth and Why it Matters?* OIES Energy Comment; Oxford Institute for Energy Studies: Oxford, UK, 2025.
5. Graham, E.; Fulghum, N.; Altieri, K. *Global Electricity Review 2025*; Ember: London, UK, 2025.
6. International Renewable Energy Agency. *Innovation Landscape for a Renewable-Powered Future: Solutions to Integrate Variable Renewables*; International Renewable Energy Agency: Abu Dhabi, United Arab Emirates, 2019.
7. Choobdari, M.; Moghaddam, M.S.; Davarzani, R.; Azarfar, A.; Hoseinpour, H. Robust distribution networks reconfiguration considering the improvement of network resilience considering renewable energy resources. *Sci. Rep.* **2024**, *14*, 23041. [[CrossRef](#)] [[PubMed](#)]
8. Astapov, V.; Shabbir, N.; Rosin, A.; Kütt, L.; Maask, V.; Tiismus, H. Review of technical solutions addressing voltage and operational challenges in a distribution grid with high penetration of intermittent RES. *Energy Rep.* **2025**, *14*, 1738–1760. [[CrossRef](#)]
9. Zabihi, A.; Badesa, L.; Hernandez, A. Evaluation of voltage unbalance metrics in distribution networks with high DER penetration. *arXiv* **2025**, arXiv:2505.23435v1.
10. Zhang, X.; Wu, Z.; Sun, Q.; Gu, W.; Zheng, S.; Zhao, J. Application and progress of artificial intelligence technology in the field of distribution network voltage control: A review. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114282. [[CrossRef](#)]
11. Shingade, G.; Upadhyay, N.V.; Pandya, H.S.; Bhatt, K.K.; Sukre, P.; Ghatge, V.V. Applications of artificial intelligence and machine learning in power system operation and control: A comprehensive review. *Commun. Appl. Nonlinear Anal.* **2025**, *32*, 2007–2022. [[CrossRef](#)]
12. Dokur, E.; Erdogan, N.; Sengor, I.; Yuzgec, U.; Hayes, B.P. Near real-time machine learning framework in distribution networks with low-carbon technologies using smart meter data. *Appl. Energy* **2025**, *384*, 125433. [[CrossRef](#)]
13. Lee, S.-H.; Jeon, Y.-J.; Choi, Y.-H. Development of a machine-learning-based fault localization method under complex conditions in distribution power systems. *Trans. Korean Inst. Electr. Eng.* **2025**, *74*, 2181–2187. [[CrossRef](#)]
14. Markovic, M.; Bossart, M.; Hodge, B.-M. Machine learning for modern power distribution systems: Progress and perspectives. *J. Renew. Sustain. Energy* **2023**, *15*, 032301. [[CrossRef](#)]
15. Oelhaf, J.; Kordowich, G.; Pashaei, M.; Bergler, C.; Maier, A.; Jäger, J.; Bayer, S. A scoping review of machine learning applications in power system protection and disturbance management. *arXiv* **2025**, arXiv:2509.09053v1.
16. Vardi, J.; Avi-Itzhak, B. *Electric Energy Generation; Economics, Reliability and Rates*; MIT Press: Cambridge, MA, USA, 1981; pp. 75–94.
17. U.S. Department of Energy (DOE). *Interconnection Innovation e-Xchange (i2X): DER Interconnection Roadmap (2025)*; U.S. Department of Energy: Washington, DC, USA, 2025.
18. Jeon, Y.-J.; Lee, S.-H.; Choi, Y.-H. Development of robust DER contribution estimation model using machine learning under incomplete data. *Trans. Korean Inst. Electr. Eng.* **2025**, *74*, 1645–1651. [[CrossRef](#)]
19. Choi, Y.-H.; Bae, M.-S.; Lee, B. Reactive power traceable system based vulnerable areas detection for reactive power. *Trans. Korean Inst. Electr. Eng.* **2015**, *64*, 1145–1153. [[CrossRef](#)]

20. Lee, S.-H.; Jeon, Y.-J.; Gwon, D.; Choi, Y.-H. Machine learning enhances fault localization in network distribution systems. In *Proceedings of the 2025 International Conference on Power Engineering and Electrical Technology (PEET), Shiga, Japan, 22–24 October 2025*; IEEE: New York, NY, USA, 2025; pp. 1–6.
21. Hyun, S.; Kim, W.; Choi, Y.; Kim, T.; Kim, G.; Park, J. Development of VPP impact analysis simulator in distribution system. In *Proceedings of the 2025 International Conference on Power Engineering and Electrical Technology (PEET), Shiga, Japan, 22–24 October 2025*; IEEE: New York, NY, USA, 2025; pp. 1–6.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.