

Article

An Integrated and Modular Deep Learning Framework for Distribution System State Estimation

Jorge Lara *, Mauricio Samper and Delia Graciela Colomé

Institute of Electrical Energy (IEE), National University of San Juan, National Scientific and Technical Research Council (CONICET), San Juan 5400, Argentina; msamper@iee-unsjconicet.org (M.S.); gcolome@iee-unsjconicet.org (D.G.C.)

* Correspondence: jlara@iee.unsj.edu.ar

Abstract

Modern distribution networks operate under increasingly demanding conditions, characterized by the integration of distributed energy resources, unbalanced three-phase operation, low measurement redundancy, variable topologies, and data uncertainty. In this context, distribution system state estimation (DSSE) is a key tool for operational monitoring; however, its practical deployment is often hindered by topological inconsistencies and gross measurement errors. This paper proposes an integrated and modular deep learning-based methodological framework that combines active topology identification (ATI), gross error detection (GED), error-type identification (ETI), error-location identification (ELI), measurement reconstruction and correction (MRC), and DSSE. The ATI module is formulated as a global multiclass classifier, whereas the subsequent modules are trained as topology-specific models. Compromised measurements are handled through an iterative GED–ETI–ELI–MRC loop that detects, identifies, locates, and corrects one anomalous measurement per iteration before re-evaluating the input vector. The proposed methodology was validated by using simulation-based scenarios generated in OpenDSS for a real unbalanced three-phase 240-node distribution feeder. The results show that no single architecture is dominant across all subproblems: WaveNet1D achieved the best relative performance in ATI, GED, and ETI; EncDec-CNN in ELI; NBEATS1D in MRC; and EncDec-GRU in DSSE. Additionally, WLS estimators based on both nodal voltages and branch currents failed to achieve numerical convergence on the 240-node test system under the evaluated conditions, a finding consistent with recent literature reporting analogous convergence failures in distribution networks of similar or smaller scale. Furthermore, the integrated evaluation shows that omitting ATI increases the voltage-magnitude MAE by a factor of 12.3 and the voltage-angle MAE by a factor of 8.1 with respect to the complete framework, whereas omitting only the compromised-measurement treatment increases these errors by factors of 1.8 and 1.9, respectively. The total offline computational cost was approximately 1587.9 h (66.2 GPU-days), while the online inference latency was approximately 0.45 ms per sample, making the framework compatible with AMI- and SCADA-based monitoring cycles. These findings confirm that topological consistency is the dominant factor in DSSE accuracy and that iterative measurement correction meaningfully improves estimator robustness under anomalous measurement conditions.

Academic Editors: Mateo Bašić,
Marko Vekić and Ivan Grgić

Received: 27 May 2026

Revised: 3 July 2026

Accepted: 7 July 2026

Published: 10 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 \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

Keywords: distribution system state estimation; active topology identification; gross measurement errors; measurement reconstruction; deep learning; low measurement redundancy; OpenDSS

1. Introduction

Electric distribution networks are undergoing a structural transformation driven by the large-scale integration of distributed energy resources, photovoltaic generation, energy storage systems, electric vehicles, and controllable loads, along with infrastructure digitization and the transition toward bidirectional and dynamic operating schemes [1–3]. In this context, distribution system operators require up-to-date information on the electrical state of the network, a function traditionally provided by distribution system state estimation (DSSE) [4,5]. Unlike transmission systems, distribution networks exhibit several features that make this task particularly challenging: limited measurement coverage and low redundancy [1,6,7]; inherently unbalanced three-phase operation with non-transposed lines [5,8]; high R/X ratios and radial or weakly meshed structures [4,5,9]; and variable network configurations caused by feeder reconfiguration, post-fault operation, or the flexible connection of distributed resources [10–12]. The combined effect of these characteristics often leads to numerically ill-conditioned problem formulations and makes the estimator highly sensitive to mismatches between the assumed and actual network topology, as well as to inconsistencies in the measurement set [4,8,9,13].

Data quality is a critical factor in distribution systems. Measurements collected from advanced metering infrastructure (AMI), supervisory control and data acquisition (SCADA) systems, and distribution-level phasor measurement units, such as micro-phasor measurement units (μ PMUs), may be affected by noise, systematic bias, missing data due to communication failures, time-synchronization errors, and, in some cases, gross errors associated with instrumentation failures or anomalous events during the acquisition process [1,3,14–16]. Therefore, gross error detection (GED), error-type identification (ETI), error-location identification (ELI), and measurement reconstruction and correction (MRC) are necessary stages to improve the robustness of the DSSE process [1,4,15–17]. In parallel, active topology identification (ATI) plays a central role. In distribution networks, the open/closed status of switching devices is not always available in real time or synchronized with electrical measurements. This may introduce structural errors that significantly degrade estimator performance and, in extreme cases, compromise the convergence of the iterative estimation algorithm itself [10–12].

To address these limitations, deep learning has emerged as a promising approach for handling nonlinearity, low redundancy, and uncertainty in distribution systems. Deep learning models offer universal approximation capabilities without requiring an analytical derivation of the Jacobian matrix, while shifting most of the computational burden to the offline training stage [2,4,18–20]. However, a review of the recent literature reveals three recurring limitations in data-driven approaches. First, most studies focus on a restricted subset of functions, typically treating DSSE as an isolated task or, alternatively, addressing ATI or error detection separately [10,11,15,16]. Second, many approaches rely on a single architecture to solve heterogeneous tasks, including classification, localization, reconstruction, and regression, without explicitly discussing whether the selected inductive bias is suitable for all of them [4,18]. Third, and most importantly, existing studies often do not explicitly quantify the functional contribution of each stage to the final state-estimation outcome. As a result, operationally relevant questions remain open, such as how much DSSE accuracy deteriorates when topology identification is omitted or when compromised measurements are not properly treated. Therefore, it is necessary to evaluate frameworks that organize these functions as a modular inference sequence, clearly distinguishing among topology identification, gross error detection, error-type and error-location identification, measurement reconstruction, and final state estimation, while enabling ablation studies to quantify how the omission of specific modules affects overall performance [4,19].

The primary objective of this work is to design and validate a modular deep learning-based inference framework that integrates the six functions required for distribution network monitoring ATI, GED, ETI, ELI, MRC, and DSSE—and to quantify, through an ablation study, the functional contribution of each module to the final state-estimation accuracy. The central novelty does not reside in proposing a new neural network architecture, but rather in three complementary aspects: (i) the modular functional organization that enables selecting the best-performing architecture for each subproblem; (ii) the iterative formulation of compromised-measurement treatment that avoids the combinatorial complexity of simultaneous multilabel identification; and (iii) the quantitative evidence, obtained through ablation, regarding the dominant impact of topological consistency on estimator accuracy. Furthermore, the decision to adopt a fully data-driven framework is grounded in the well-documented numerical difficulties that classical WLS estimators face under the specific operating conditions of this study, including the numerical non-convergence observed experimentally on the 240-node test system.

To address this gap, this paper proposes an integrated and modular deep learning-based framework for DSSE in unbalanced three-phase distribution networks with dynamic topologies. The methodology organizes the problem into specialized stages, namely ATI, GED, ETI, ELI, MRC, and DSSE, where each stage is solved using the model with the best relative performance according to the nature of the corresponding subproblem. The ATI module is formulated as a global multiclass classifier, whereas the subsequent modules are trained as topology-specific models, with fixed architectures and hyperparameters for each task. This strategy combines global topology identification with models specialized in the electrical relationships of each network configuration, avoiding the need for a single neural network to simultaneously learn multiple operating regimes.

The main contributions of this paper are as follows:

- An integrated and modular AI-based DSSE framework that combines ATI, GED, ETI, ELI, MRC, and DSSE for unbalanced three-phase distribution networks with dynamic topologies, positioned as a functional alternative for operating conditions under which classical WLS estimators exhibit well-documented convergence difficulties.
- A hybrid learning strategy based on a global model for ATI and topology-specific models for GED, ETI, ELI, MRC, and DSSE, with an explicit scalability analysis with respect to the number of topological configurations.
- An iterative formulation for compromised-measurement treatment, in which the GED–ETI–ELI–MRC block detects, identifies, locates, and corrects one anomalous measurement per iteration, including an analysis of failure modes under multiple interacting errors.
- A systematic comparison of deep learning architectures with different inductive biases, using Bayesian hyperparameter optimization and a normalized composite score for per-subproblem architecture selection.
- An integrated functional evaluation that quantifies the contribution of ATI and iterative measurement treatment to the final DSSE accuracy through ablation, together with a report of the offline computational cost (~1587.9 GPU-hours) and online inference latency (~0.45 ms per sample).

The proposed methodology was validated on a real unbalanced three-phase 240-node distribution feeder from the U.S. Midwest implemented in OpenDSS [21,22]. The results show that the best-performing architecture depends on the nature of the subproblem: WaveNet1D achieves the best relative performance in ATI, GED, and ETI; EncDec-CNN in ELI; NBEATS1D in MRC; and EncDec-GRU in DSSE. The integrated evaluation shows that omitting ATI increases the voltage-magnitude MAE by approximately a factor of 12 and the voltage-angle MAE by approximately a factor of 8 relative to the complete

framework, whereas omitting the treatment of compromised measurements increases these errors by approximately a factor of 1.8–1.9. These results confirm the importance of preserving topological consistency and iteratively correcting compromised measurements before the final state-estimation stage.

The remainder of this paper is organized as follows. Section 2 briefly presents the required theoretical background and positions the proposed framework within the literature. Section 3 describes the methodology, including the problem formulation, database, candidate architectures, and evaluation metrics. Section 4 presents case studies and the modular and integrated results. Section 5 discusses the main findings and limitations. Finally, Section 6 concludes the paper.

2. Theoretical Background and Related Work

DSSE aims to infer the electrical state of a distribution network from a limited and heterogeneous set of measurements. The problem is commonly formulated through the nonlinear measurement model:

$$z = h(x) + e \quad (1)$$

where z is the vector of available measurements, x is the system state vector, typically composed of nodal voltage magnitudes and phase angles for each phase, $h(\cdot)$ represents the nonlinear measurement model, and e accounts for the errors associated with the data acquisition process. In distribution networks, this model may include direct measurements of voltage, current, and power, as well as load pseudomeasurements, whose uncertainty is usually significantly higher than that of instrument-grade measurements.

To formalize the effect of topology on the estimation problem, consider that the active topology τ determines the three-phase nodal admittance matrix $Y(\tau) \in \mathbb{C}^{3N \times 3N}$ and, consequently, the structure of the measurement model $h(x, \tau)$ and its Jacobian $H(x, \tau) = \partial h / \partial x$. A change from τ_a to τ_b alters both the graph connectivity and the effective inter-node impedances, modifying the gain matrix structure:

$$G_{WLS}(\tau_a) = H^T R^{-1} H|_{\tau_a} \neq G_{WLS}(\tau_b) = H^T R^{-1} H|_{\tau_b} \quad (2)$$

This structural dependency means that an estimator configured for τ_a operates under an inconsistent physical model when the true topology is τ_b , producing a systematic error that cannot be compensated for through individual measurement treatment. The topology-induced error can be expressed as:

$$\Delta x_{topo} = x(\tau_{true}) - \hat{x}(\tau_{assumed}) \quad (3)$$

This error is structural in nature and affects the state vector globally, unlike the error induced by an individual gross measurement, whose effect depends on the corresponding Jacobian row and the weight assigned in R^{-1} . This distinction provides the conceptual foundation for the ablation result reported in Section 4.3: topology inconsistency produces degradation factors of 12.3× and 8.1×, which dominate over the 1.8× and 1.9× factors caused by untreated compromised measurements.

Within the proposed framework, every module downstream of ATI operates conditioned on the identified topology $\hat{\tau}$. Formally, the final estimate can be written as:

$$\hat{x} = f_{DSSE}(\tilde{z}^{(K)}, \hat{\tau}) \quad (4)$$

where $\tilde{z}^{(K)}$ is the measurement vector resulting from K correction iterations:

$$\tilde{z}^{(k)} = \Phi_{MRC}^{(k)}(\tilde{z}^{(k-1)}, \hat{\tau}), k = 1, \dots, K \quad (5)$$

and $\Phi_{\text{MRC}}^{(k)}$ represents the composite detection–identification–localization–correction operation at iteration k . The sequential composition of these functions implies that a diagnostic error at iteration k may propagate to subsequent an inherent trade-off of the sequential formulation, analyzed further in Section 5.

Classical approaches, such as Weighted Least Squares (WLS), Weighted Least Absolute Value (WLAV), robust estimators, and branch-current-based formulations [4,5,9] face particular difficulties in distribution systems due to unbalanced three-phase operation, low measurement redundancy, high R/X ratios, radial or weakly meshed structures, and the increasing penetration of distributed energy resources [1,2,4,5]. These characteristics can make the classical problem formulation numerically ill-conditioned: the measurement Jacobian and the WLS gain matrix may exhibit rank deficiency or a high condition number, leading to convergence issues, sensitivity to the angular reference, and a critical dependence on the weights assigned to pseudomeasurements [8,9,13].

These difficulties historically motivated the development of alternative formulations, notably the branch-current-based state estimation proposed by Baran and Kelley [9] and subsequent variants based on decoupled gain matrices [13]. More recently, they have also motivated the exploration of data-driven approaches as complements to, or replacements for, Newton–Raphson-based iterative schemes [2,4,18]. Consequently, distribution system estimators must be able to operate under heterogeneous uncertainty, variable topologies, and diverse measurement types while maintaining numerical robustness under conditions in which classical guarantees may no longer hold [1,4,8].

A central challenge is the limited measurement coverage, which requires supplementing the available information through load pseudomeasurements derived from historical profiles, statistical models, or machine learning techniques [1,6,7]. Pseudomeasurements exhibit uncertainty levels significantly higher than those of direct measurements, typically on the order of $\pm 30\%$ compared with $\pm 1\text{--}2\%$ for voltage and current measurements. This directly constrains the achievable estimation quality and motivates a differentiated treatment in the weighting model [6,7,19,20]. In addition, uncertainty in DSSE is not limited to measurement noise; it also includes parameter errors, time desynchronization across data sources, topological inconsistencies, and gross errors [4,5,14].

Topology defines the effective connectivity of the system and directly conditions the Jacobian matrix of the estimation problem. When the assumed topology does not match the actual network configuration, a structural error is introduced, altering the physical relationship among measurements, nodes, branches, and state variables. This degradation is not necessarily proportional to the apparent difference between configurations. Methods for active topology identification (ATI) can be broadly grouped into two main families: physics-based approaches, which rely on electrical constraints, residual analysis, and circuit-law verification; and data-driven approaches, which infer the operating configuration from measurement patterns, frequently using supervised learning techniques [10–12].

Traditionally, erroneous data have been handled using normalized residual analysis, hypothesis testing, and robust estimation methods [14,17,23]. In networks with low redundancy, however, these techniques lose effectiveness: the scarcity of measurements limits the effective rank of the Jacobian and makes it difficult to distinguish between individual gross errors, accumulated normal uncertainty, and actual operating changes. This limitation is widely recognized as one of the main challenges of applying the classical state-estimation paradigm to distribution networks [4,5,17]. Recent literature has shown growing interest in machine learning for GED, ETI, and ELI, often from the perspective of false data injection (FDI) detection and anomalous event identification [15,16]. However, most of these studies focus on only one of the four tasks, typically detection, leaving the explicit correction of measurements and their integration with a subsequent estimator within a unified functional scheme as a persistent gap in DSSE research [4,16,17].

Deep learning has emerged as a complement to classical estimators because of its ability to approximate high-dimensional nonlinear relationships and shift most of the computational burden to the offline training stage, enabling fast online inference [2,4,18–20]. Different architectures exploit different inductive biases, that is, structural assumptions about the nature of the data that guide the learning process [24]. One-dimensional convolutional neural networks and their dilated variants, including WaveNet, extract local and multiscale patterns by expanding the receptive field through dilated convolutions and residual connections [24–26]. Inception-type architectures capture patterns at different scales through heterogeneous filters [27]. Recurrent and sequence-to-sequence architectures, including GRU, LSTM, encoder–decoder models, and attention mechanisms, model ordered dependencies and mappings toward high-dimensional target variables [28–32]. Time-series-oriented architectures, such as SCINet and N-BEATS, provide decomposition and progressive refinement mechanisms [33,34].

Over the past several years, the application of deep learning to DSSE has expanded considerably. Graph neural networks (GNNs) have emerged as a particularly promising family, since they natively exploit the topological structure of the power network as a graph. Habib et al. [19] proposed the Deep Statistical Solver (DSS²), a GNN-based model that outperforms classical WLS in accuracy, convergence, and computational time across 14, 70, and 179-bus networks, demonstrating that WLS fails to converge on the 179-bus network under noisy conditions with pseudomeasurements. More recently, power-flow-guided GraphSAGE models [35] and physics-constrained heterogeneous graph attention networks (GATNE) [36] have confirmed WLS instability on large-scale networks and proposed GNN-based alternatives.

Transformer-based architectures have also been explored for DSSE, leveraging global attention mechanisms to model dependencies among spatially distributed measurements [37,38]. However, their application to unbalanced three-phase distribution systems with variable topologies remains nascent.

A complementary approach is offered by physics-informed neural networks (PINNs), which embed power-flow equations as constraints in the training loss. Tran et al. [20] showed that this strategy can improve the physical consistency of estimates, noting explicitly that the conventional WLS approach faces convergence issues due to insufficient measurement devices in distribution systems. Zamzam and Baker [39] proposed a physics-aware neural network for DSSE, although limited to a fixed topology.

Additionally, false data injection attack (FDIA) detection has received growing attention as a complementary DSSE task [15,16]. However, most of these studies treat detection as a standalone task without integrating it with topology identification or measurement correction within a unified pipeline.

Regarding the convergence of classical estimators, the recent literature has extensively documented the numerical difficulties of WLS in distribution networks. Langner and Abur [40] demonstrated that the conventional WLS estimator based on nodal voltages fails to converge within 20 iterations on the IEEE 123-bus feeder under a flat start, proposing a virtual-reference formulation to mitigate the angular-reference problem in unbalanced systems. Gómez-Quiles et al. [41] identified that gain-matrix ill-conditioning—caused by measurements with highly disparate weights, insufficient redundancy, high R/X ratios, and the simultaneous presence of short and long lines—can produce both convergence failures and inaccurate estimates. Dehghanpour et al. [42] documented in their survey that inaccurate pseudomeasurements can create ill-conditioned optimization formulations that prevent estimator convergence.

Because ATI, GED, ETI, ELI, MRC, and DSSE have different problem structures, namely global multiclass classification, binary classification, categorical classification, spatial localization, regression of heterogeneous variables, and state-vector regression, re-

spectively, there is no theoretical evidence suggesting the existence of a single optimal architecture for all tasks. This observation is consistent with previous findings in multitask learning [43,44]. Therefore, a systematic architecture comparison and a modular framework are necessary, allowing each stage to use the model with the best relative performance for the corresponding subproblem.

To position the proposed framework with respect to closely related literature, Table 1 summarizes six representative studies on deep learning-based DSSE and compares them with the proposed approach across five dimensions: the functional tasks covered by each proposal, namely ATI, compromised-measurement treatment, and DSSE; the treatment of unbalanced three-phase operation; the explicit consideration of variable topologies; the use of ablation studies to quantify the contribution of each module; and the test system used.

Table 1. Positioning of the proposed framework relative to representative studies on deep learning-based DSSE.

Work	Covered Tasks	Unbalanced Three-Phase Operation	Variable Topology	Ablation Study	Test System
Wang et al. [10]	ATI	No, single-phase equivalent	Yes, main focus	No	Smart meter data
Radhoush et al. [15]	DSSE + FDIA detection	Yes	No	No	IEEE 13-node system
Raghuvamsi & Teeparthi [16]	DSSE + FDIA/DoS detection and reconstruction	Yes	No	No	IEEE test networks
Habib et al. [19]	DSSE, GNN, physics-based scheme	Yes	No, fixed topology	No	Synthetic network
Tran et al. [20]	DSSE, physics-aware NN	Yes	No, fixed topology	No	IEEE test networks
Gotti et al. [12]	ATI	Limited	Yes, main focus	No	IEEE test networks
This work	ATI + GED + ETI + ELI + MRC + DSSE	Yes, native operation	Yes, multiple configurations	Yes, four functional scenarios	Real 240-node feeder [21]

The comparison highlights three fundamental differences between the proposed framework and previous studies. First, in terms of functional coverage, existing literature addresses restricted subsets of the problem, typically DSSE coupled with false data injection detection or ATI as a standalone task. In contrast, the proposed framework explicitly integrates all six functions within a single inference pipeline. Second, regarding the joint treatment of unbalanced three-phase operation and variable topologies, most studies assume a fixed topology or evaluation cases without significant imbalance, whereas the proposed framework operates natively on three-phase configurations. Third, unlike previous studies, this work includes an ablation analysis that quantifies the degradation induced by the individual and joint omission of ATI and compromised-measurement treatment, providing operationally actionable evidence on which block introduces the greatest risk when omitted.

To complement the functional comparison in Table 1, Table 2 provides a broader summary of recent data-driven DSSE studies, including whether WLS convergence was reported and the research gaps motivating the present work.

Table 2. Extended summary of recent data-driven studies on DSSE and related tasks.

Reference	Method	Tasks	System	Var. Top.	Three-Phase	WLS Reported	Key Limitation
Habib et al. [19]	GNN (DSS ²)	DSSE	IEEE 14/70/179	No	No	Non-convergent on 179-bus	Single-phase; no ATI or GED
Langner & Abur [40]	WLS + virtual ref.	DSSE	IEEE 123/8500	No	Yes	Non-convergent w/o virtual ref.	No gross errors or ATI
GraphSAGE [35]	GNN	DSSE	IEEE 14/70/179	No	No	Unstable on large networks	Single-phase; no ATI
GATNE [36]	GNN + attention	DSSE	IEEE 14/70/179	No	No	Divergent at scale	No ATI or error correction
Tran et al. [20]	PINN	DSSE	IEEE test	No	Yes	Notes convergence issues	Fixed topology
Zamzam & Baker [39]	Physics-aware NN	DSSE	IEEE 37	No	Partial	Not reported	No ATI; small network
Azimian et al. [45]	DNN	SE + topology	Various	Partial	Yes	Not reported	No error correction
Dehghanpour et al. [42]	Survey	DSSE	—	—	—	Documents challenges	Survey, not a new method
This work	Modular DL	ATI + GED + ETI + ELI + MRC + DSSE	Real 240-node [21]	Yes (12)	Yes	Non-convergent	Synthetic data

The decision to develop a fully data-driven framework is grounded in the well-documented numerical difficulties that classical WLS-based estimators exhibit under the specific operating conditions of this study. Indeed, attempts to implement WLS estimators based on both nodal voltages and branch currents on the 240-node feeder evaluated in this work did not achieve numerical convergence (Section 4.4)—a finding consistent with recent reports of WLS divergence on distribution networks of similar or smaller scale [19,35,36,40,41].

Accordingly, the proposed framework is positioned as a functional alternative for operating conditions under which classical Newton–Raphson-based estimators do not guarantee numerical convergence. The central contribution lies in the modular functional organization, per-subproblem architecture selection, and ablation-based quantification of each module’s impact on the final DSSE accuracy. It is acknowledged that more advanced classical formulations such as WLS with virtual reference [40], semidefinite programming, or Kalman-filter-based estimators may partially mitigate these convergence issues and constitute complementary baselines for future comparative studies.

Building on the gaps and challenges identified above, the following section describes the proposed framework in detail, starting with the overall pipeline architecture and proceeding through the problem formulation, database construction, candidate architectures, and evaluation criteria

3. Methodology

3.1. Overview of the AI-Based DSSE Framework

The proposed framework integrates six functional modules: active topology identification, gross measurement error detection, error-type identification, error-location identification, measurement reconstruction and correction, and distribution system state estimation. These stages are hereafter denoted as ATI, GED, ETI, ELI, MRC, and DSSE, respectively.

State estimation is not formulated as a single regression task, but rather as a modular inference pipeline. Starting from the initial vector of measurements and pseudomeasurements, $\mathbf{z}^{(0)}$, the ATI module identifies the active topology, $\hat{\tau}$, without explicitly using the open or closed status of switching devices. This design choice reflects practical operating conditions, where switch-status information may be unavailable, outdated, or not properly synchronized with the electrical measurements.

Once the active topology has been identified, the subsequent modules operate conditioned on $\hat{\tau}$. The GED module evaluates whether the measurement vector contains at least one anomalous measurement. If no gross errors are detected, the vector is directly passed to the DSSE module. Otherwise, ETI identifies the affected measurement type, ELI locates the compromised measurement, and MRC reconstructs its corrected value. The measurement vector is then updated and re-evaluated by GED.

Thus, the GED–ETI–ELI–MRC block operates as an iterative diagnosis and correction process. At each iteration, one compromised measurement is corrected, and the process is repeated until GED detects no further anomalies or a predefined maximum number of iterations is reached. Finally, the DSSE module estimates the electrical state of the network using the cleaned measurement vector, $\hat{\mathbf{z}}$, and the identified active topology.

Figure 1 summarizes the proposed methodological flow, in which ATI provides the active topology and the GED–ETI–ELI–MRC block iteratively corrects compromised measurements before the final state-estimation stage.

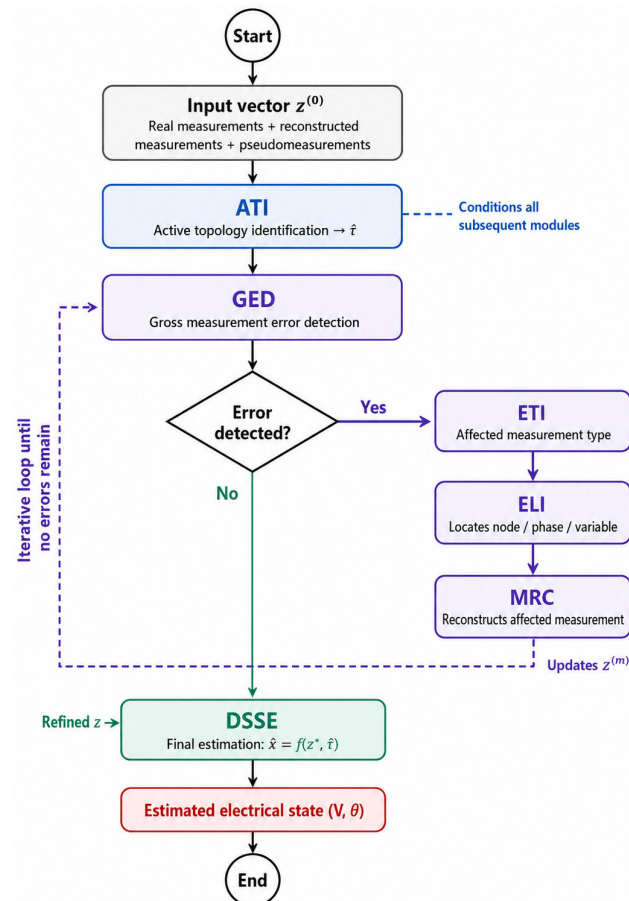


Figure 1. Functional structure of the proposed methodological framework for ATI, GED, ETI, ELI, MRC, and DSSE.

3.2. Problem Formulation

Consider an unbalanced three-phase distribution network composed of N nodes and operating under a finite set of topological configurations. For each operating scenario, the electrical state is defined by the voltage magnitudes and phase angles of each phase.

The state vector is expressed as:

$$x = [|V_1^a|, \theta_1^a, |V_1^b|, \theta_1^b, |V_1^c|, \theta_1^c, \dots, |V_N^a|, \theta_N^a, |V_N^b|, \theta_N^b, |V_N^c|, \theta_N^c]^T \quad (6)$$

where $|V_n^\phi|$ and θ_n^ϕ represent the voltage magnitude and voltage angle at node n and phase $\phi \in \{a, b, c\}$, respectively.

The vector of measurements and pseudomeasurements is constructed from nodal voltage magnitudes and angles, branch-current magnitudes and angles, active and reactive power flows, measured nodal powers, and load pseudomeasurements:

$$z = [V_i, VAng_i, I_{ft}, IAng_{ft}, PQ_{ft}, PQ_i, PQ_{psd}] \quad (7)$$

where V_i and $VAng_i$ denote nodal voltage magnitudes and angles; I_{ft} and $IAng_{ft}$ denote branch-current magnitudes and angles; PQ_{ft} denotes active and reactive power flows; PQ_i denotes measured nodal active and reactive powers; and PQ_{psd} denotes load pseudomeasurements. Under this formulation, $\mathbf{z}$ constitutes the common input to the AI-based DSSE framework.

In anomalous scenarios, between one and five compromised measurements are introduced per sample. However, error treatment is not formulated as the simultaneous identification of all compromised measurements, but rather as a sequential correction process. At each iteration, a single compromised measurement is detected, identified, located, and corrected. This avoids the combinatorial complexity of a multilabel formulation and preserves a clear supervised structure for each module.

Let $\mathbf{z}^{(0)} = \mathbf{z}$ be the initial measurement vector. For each iteration $i = 0, 1, \dots, I_{\max}$, with $I_{\max} = 5$, GED evaluates:

$$\hat{y}_{GED}^{(i)} = f_{GED}(z^{(i)}, \hat{\tau}) \quad (8)$$

where

$$\hat{y}_{GED}^{(i)} \in \{0, 1\} \quad (9)$$

If $\hat{y}_{GED}^{(i)} = 0$, no further anomalies are detected, and the cleaned measurement vector is defined as:

$$\tilde{z} = z^{(i)} \quad (10)$$

If $\hat{y}_{GED}^{(i)} = 1$, ETI identifies the affected measurement type:

$$\hat{c}^{(i)} = f_{ETI}(z^{(i)}, \hat{\tau}) \quad (11)$$

where

$$\hat{c}^{(i)} \in \{V_i, VAng_i, I_{ft}, IAng_{ft}, PQ_{ft}, PQ_i, PQ_{psd}\} \quad (12)$$

Then, ELI locates the compromised measurement:

$$\hat{l}^{(i)} = f_{ELI}(z^{(i)}, \hat{\tau}, \hat{c}^{(i)}) \quad (13)$$

The location $\hat{l}^{(i)}$ depends on the identified measurement type. For voltage measurements, it corresponds to the node and phase; for branch currents and power flows, to the element, terminal, and phase; and for nodal powers or pseudomeasurements, to the corresponding node, load, or phase.

Subsequently, MRC reconstructs the corrected value:

$$\hat{z}_l^{(i)} = f_{MRC}(z^{(i)}, \hat{t}, \hat{c}^{(i)}, \hat{l}^{(i)}) \quad (14)$$

and the measurement vector is updated as:

$$z^{(i+1)} = \text{replace}(z^{(i)}, \hat{l}^{(i)}, \hat{z}_l^{(i)}) \quad (15)$$

The process is repeated until GED detects no further anomalies or until $I_{\max} = 5$ is reached. This limit coincides with the maximum number of gross errors introduced per scenario and prevents indefinite loops caused by possible false positives or incorrect diagnoses.

This strategy consists of sequential bad-data analysis approaches, in which the most suspicious measurement is identified, corrected, or removed before the analysis is repeated. In addition, simultaneous multiple errors may lead to masking effects or incorrect identification; therefore, their joint treatment requires specific formulations [14,23,46,47].

Based on this formulation, the inputs, outputs, and functional interpretation of each module in the proposed AI-based DSSE framework are summarized in Table 3.

Table 3. Inputs, outputs, and interpretation of the modules in the AI-based DSSE framework.

Module	Input	Output	Interpretation	Task Type
ATI	$\mathbf{z}^{(0)}$	$\hat{t}$	Identified active topology	Global multiclass classification
GED	$\mathbf{z}^{(i)}, \hat{t}$	$\hat{y}_{GED}^{(i)}$	0: no anomaly; 1: anomaly detected	Topology-specific binary classification
ETI	$\mathbf{z}^{(i)}, \hat{t}$	$\hat{c}^{(i)}$	Affected measurement type	Topology-specific multiclass classification
ELI	$\mathbf{z}^{(i)}, \hat{t}, \hat{c}^{(i)}$	$\hat{l}^{(i)}$	Affected node, element, phase, terminal, or load	Conditional multiclass classification
MRC	$\mathbf{z}^{(i)}, \hat{t}, \hat{c}^{(i)}, \hat{l}^{(i)}$	$\hat{z}_l^{(i)}, \mathbf{z}^{(i+1)}$	Reconstructed value and updated vector	Topology-specific regression
DSSE	$\hat{\mathbf{z}}, \hat{t}$	$\hat{\mathbf{x}}$	Estimated network state	Topology-specific regression

The hybrid strategy, composed of a global ATI model and topology-specific models for GED, ETI, ELI, MRC, and DSSE, is justified by the fact that the topology modifies the physical relationship among measurements, errors, and state variables [38]. Therefore, once $\hat{t}$ has been identified, the subsequent modules operate through specialized instances associated with that configuration.

3.3. Uncertainty and Gross Measurement Error Modeling

To represent the uncertainty associated with the data-acquisition process, the synthetic measurements obtained from power-flow simulations were perturbed using a stochastic model differentiated according to the nature of each variable. Specifically, magnitude variables were modeled using multiplicative Gaussian noise [6,7,19,20], whereas angular variables were modeled using additive Gaussian noise [3,19,20].

For magnitude variables, including nodal voltage magnitudes V_i , branch currents I_{ft} , power flows PQ_{ft} , measured nodal powers PQ_i , and load pseudomeasurements PQ_{psd} , the perturbed measurement was obtained as:

$$\tilde{z}_k = z_k(1 + \varepsilon_k), \varepsilon_k \sim \mathcal{N}(0, \sigma_{rel,k}^2) \quad (16)$$

where z_k is the simulated value of measurement k , $\tilde{z}_k$ is the perturbed value, ε_k is a zero-mean Gaussian random variable, and $\sigma_{rel,k}$ is the relative standard deviation associated with the corresponding measurement type. This formulation makes the absolute magnitude of the perturbation proportional to the value of the electrical variable, which

is appropriate for representing relative errors in measuring instruments and pseudomeasurements.

For angular variables corresponding to nodal voltage angles $VAng_i$ and branch-current angles $IAng_{ft}$, an additive model was used:

$$\tilde{\theta} = \theta + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_{\theta,k}^2) \quad (17)$$

where θ_k is the angular value obtained from the simulation, $\tilde{\theta}_k$ is the perturbed angle, and η_k is a zero-mean additive Gaussian perturbation with standard deviation $\sigma_{\theta,k}$. This representation allows angular uncertainty to be modeled in absolute units, independently of the magnitude of the measured variable [3,19,20].

The standard deviation values were defined according to uncertainty ranges commonly used in DSSE studies, adopting the $\pm 3\sigma$ interval as an operational criterion. This interval contains approximately 99.7% of the realizations of a normal distribution. Under this criterion, the approximate maximum relative uncertainty is associated with three times the adopted standard deviation.

In this work, $\sigma_{rel} = 0.0033$ was adopted for direct measurements of nodal voltage V_i , branch current I_{ft} , and measured nodal power PQ_i , which corresponds approximately to a relative uncertainty of $\pm 1\%$ under the $\pm 3\sigma$ criterion. For power flows PQ_{ft} , $\sigma_{rel} = 0.0067$ was used, corresponding approximately to a relative uncertainty of $\pm 2\%$. For load pseudomeasurements PQ_{psd} , $\sigma_{rel} = 0.10$ was adopted, representing an approximate relative uncertainty of $\pm 30\%$, consistent with the higher uncertainty typically associated with this type of information [6,7,19,20].

Finally, for the angular variables $VAng_i$ and $IAng_{ft}$, $\sigma_{\theta} = 0.167^\circ$ was used, equivalent to an approximate angular uncertainty of $\pm 0.5^\circ$ under the $\pm 3\sigma$ criterion [3,19,20].

In addition to the measurement noise associated with normal data acquisition, the database included scenarios with gross measurement errors introduced in a controlled manner into specific variables. Unlike the Gaussian noise described above, these gross errors represent anomalous perturbations of larger magnitude, associated with sensor failures, communication errors, miscalibration, partial loss of information, or inconsistent records. Therefore, they were modeled as events distinct from ordinary instrumental noise.

In general, a compromised measurement can be expressed as:

$$z_k^{GE} = \tilde{z}_k + \Delta z_k^{GE} \quad (18)$$

where z_k^{GE} represents the measurement affected by a gross error, $\tilde{z}_k$ is the measurement previously perturbed by normal noise, and Δz_k^{GE} is the anomalous perturbation introduced. In this work, the anomalous perturbation was generated as:

$$\Delta z_k^{GE} = \alpha_k \mid \tilde{z}_k \mid, \alpha_k \sim \mathcal{N}(0, \sigma_{GE}^2), \sigma_{GE} = 0.20 \quad (19)$$

so that gross errors represented deviations proportional to the magnitude of the affected measurement. This formulation explicitly separates two sources of uncertainty: random measurement error, present in all variables, and gross error, applied only to a controlled subset of measurements.

The insertion of gross errors preserved the traceability of three fundamental elements: the affected measurement type, its location in the network, and the corresponding active topology. This information was used to construct the labels for the GED, ETI, ELI, and MRC modules. Specifically, the GED module uses a binary label indicating the presence or absence of at least one compromised measurement; the ETI module identifies the affected variable type; the ELI module locates the compromised measurement at the corresponding node, phase, or network element; and the MRC module uses the value without the gross error as the reference target for learning the reconstruction of the affected measurement.

In summary, the adopted procedure represents scenarios with different levels of measurement quality: high-accuracy direct measurements, power flows with intermediate uncertainty, pseudomeasurements with high uncertainty, and anomalous cases with gross measurement errors. This differentiation is essential for training and evaluating the modules of the proposed framework under conditions representative of monitoring in unbalanced three-phase distribution networks, where the final state-estimation accuracy depends on both topological consistency and the quality of the information used by the estimator.

3.4. Construction of the Synthetic Database

The synthetic database was constructed from electrical simulations performed in OpenDSS 10.0.0.2 (Electric Power Research Institute, Palo Alto, CA, USA), while preprocessing routines and dataset generation were implemented in Python 3.11.10 (Python Software Foundation, Wilmington, DE, USA). The procedure considered multiple operating scenarios, topological configurations, different levels of measurement uncertainty, and gross errors introduced in a controlled manner. This strategy enabled the generation of labeled input–output pairs for each module of the AI-based DSSE framework under reproducible conditions and with full traceability of the variables involved.

The overall process comprises five main stages: generation of operating scenarios from load profiles and predefined topologies; execution of power-flow simulations; perturbation of the measurements using normal noise; controlled insertion of gross errors; and construction of the input–output pairs required by each module. Figure 2 summarizes this procedure, showing the transition from electrical simulations to the generation of specific datasets for ATI, GED, ETI, ELI, MRC, and DSSE.

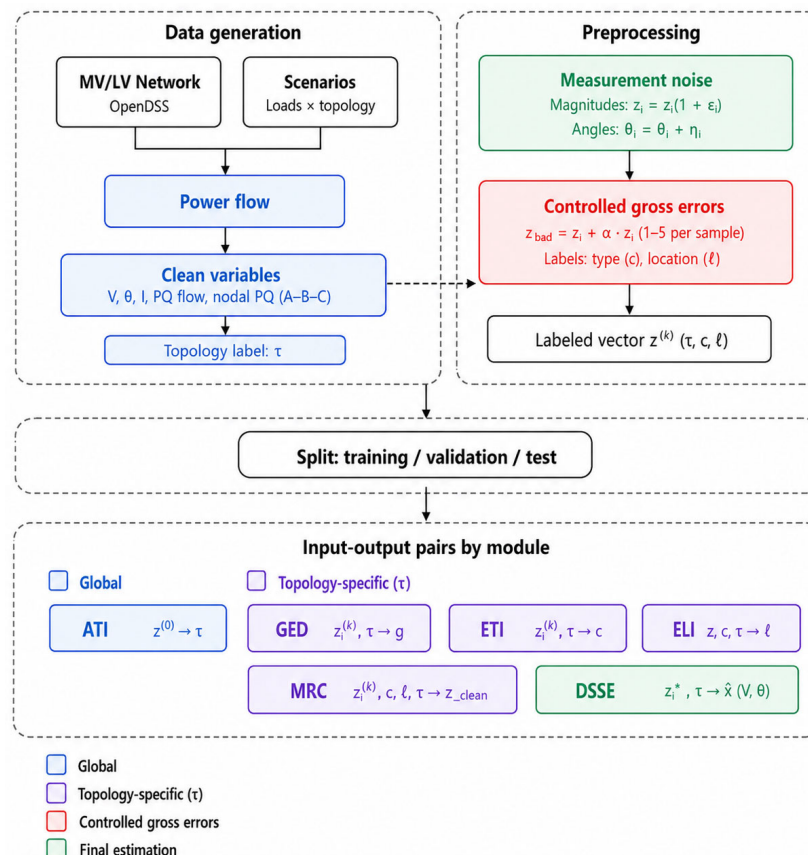


Figure 2. General procedure for constructing the synthetic database and preprocessing the cases used in the modules of the AI-based DSSE framework.

In anomalous scenarios, between one and five measurements were affected per sample. Although a sample may contain more than one gross error, the GED–ETI–ELI–MRC block is applied iteratively, correcting one measurement at a time. After each correction, the updated vector is re-evaluated by GED until no further anomalies are detected or until five iterations are reached. This formulation makes it possible to evaluate degraded scenarios with multiple errors without formulating ETI and ELI as simultaneous multilabel problems.

To maintain consistency between data generation and iterative inference, scenarios with multiple compromised measurements were transformed into supervised instances associated with a single affected measurement per iteration. Specifically, if a sample contains K gross errors, with $1 \leq K \leq 5$, up to K training instances are generated for the ETI, ELI, and MRC modules, each linked to a specific compromised measurement. In each instance, ETI uses the affected measurement type as the label, ELI uses its network location as the label, and MRC uses the clean value of that measurement as the reconstruction target. After each correction, the updated vector can be used as the input for the next iteration, following the same logic used during inference.

Thus, although an original sample may contain multiple gross errors, the ETI, ELI, and MRC modules are trained with single and fully traceable output targets. This strategy avoids a simultaneous multilabel formulation, reduces the combinatorial complexity of the learning problem, and preserves consistency with the GED–ETI–ELI–MRC loop defined in the methodology.

The datasets were constructed according to the specific objective of each module. For ATI, samples are labeled according to active topology. For GED, samples are labeled according to the presence or absence of gross errors in the evaluated vector. For ETI and ELI, anomalous scenarios are used with labels associated with the type and location of the compromised measurement at each iteration. For MRC, pairs are constructed between the compromised measurement and its clean reference value. Finally, for DSSE, the target corresponds to the state vector $\mathbf{x}$, composed of nodal voltage magnitudes and phase angles for each phase. This organization is summarized in Table 4.

Table 4. Structure of the datasets generated for the modules of the AI-based DSSE framework.

Module	Input	Target Output	Purpose
ATI	$\mathbf{z}^{(0)}$	τ	Identify the active topology
GED	$\mathbf{z}^{(i)}, \hat{\tau}$	Presence or absence of anomaly	Detect gross errors
ETI	$\mathbf{z}^{(i)}, \hat{\tau}$	$\hat{c}^{(i)}$	Identify the affected measurement type
ELI	$\mathbf{z}^{(i)}, \hat{\tau}, \hat{c}^{(i)}$	$\hat{l}^{(i)}$	Locate the compromised measurement
MRC	$\mathbf{z}^{(i)}, \hat{\tau}, \hat{c}^{(i)}, \hat{l}^{(i)}$	$\hat{z}_l^{(i)}$	Reconstruct the affected measurement
DSSE	$\hat{\mathbf{z}}, \hat{\tau}$	$\mathbf{x}$	Estimate the electrical state

As shown in Table 4, the database is not organized as a single generic dataset but rather as a structured set of labeled scenarios according to the specific function of each module. This organization preserves consistency among the measurement vector $\mathbf{z}$, the active topology τ , the GED detection labels y_{GED} , the type labels c , the location labels l , the clean reference values z_l^{clean} , and the target state vector $\mathbf{x}$.

The reference state vector was obtained from OpenDSS power-flow simulations, assuming that the electrical parameters of the model, including line impedances, transformer models, and load models, are known without error. Therefore, the proposed framework learns the mapping between measurements and the power-flow solution under the assumed network model. Uncertainty in electrical parameters is not explicitly considered in this study and is recognized as a future extension of the framework.

3.5. Candidate Architectures, Selection Criterion, and Training

Each module was evaluated using a set of candidate architectures selected for their ability to represent nonlinear relationships, capture spatial or structural patterns among variables, and approximate complex mappings. The comparison does not aim to identify a single universal model, but rather to evaluate architecture families with different inductive biases across subproblems of different natures. Table 5 summarizes the families considered and the inductive bias that motivated their inclusion.

Table 5. Candidate architecture families considered in the AI-based DSSE framework.

Family/Architecture	Modeling Principle	Motivation Within the Framework
MLP [48]	Global nonlinear approximation	Baseline model for classification and regression
CNN1D/Dilated CNN1D [24,26]	Local feature extraction; dilated convolutions	Patterns among measurement positions; enlarged receptive field
WaveNet1D [25]	Dilated convolutions with residual connections	Integration of information over a wide receptive field
Inception1D [27]	Multiscale filters	Electrical patterns at different scales
U-Net1D/EncDec-CNN [49,50]	Encoder–decoder structure with skip connections	Structured mapping, preservation of spatial information, and reconstruction of relationships among measurement positions
EncDec-GRU/Attention-RNN [28–32]	Recurrent modeling with attention	Ordered and long-range dependencies
Transformer/Informer/FED-former [37,38,51]	Global attention	Global relationships among variables
MLP-Mixer [52]	Position–feature mixing	Explicit separation of positional and feature-wise relationships
SCINet1D/NBEATS1D [33,34]	Hierarchical refinement	Reconstruction and state regression

The models were trained using separate training, validation, and test subsets. For the classification modules, cross-entropy-based loss functions were used according to the nature of each task's output. For ATI, ETI, and ELI, categorical or sparse categorical cross-entropy was used depending on the class encoding, whereas binary cross-entropy was used for GED. In this framework, ETI does not produce a simultaneous ranking of candidate measurements; instead, it identifies the affected measurement type at the current iteration. Similarly, ELI locates the compromised measurement associated with the type identified by ETI.

For the regression tasks, MRC and DSSE were trained using loss functions based on the error between estimated and reference values. In MRC, learning is performed on the measurement diagnosed at each iteration, so that the model reconstructs the clean value associated with the type and location identified by ETI and ELI. The reconstructed value is then used to update the measurement vector before a new GED evaluation. Thus, although the database may include scenarios with up to five gross errors, both training and inference of the GED–ETI–ELI–MRC block are maintained as a sequential process that handles one anomaly per iteration.

During training, EarlyStopping was applied to reduce the risk of overfitting, while retaining the weights associated with the best validation performance. Hyperparameter

optimization was carried out using Bayesian Optimization [53], with the validation set used as the selection criterion. This approach enables a more efficient exploration of architectural and training configurations than exhaustive search, especially when multiple model families are compared. A central aspect of the adopted strategy is that this optimization was performed at the module level rather than independently for each topology. Once the best configuration was identified for a given module, the architecture and its hyperparameters were kept fixed when training the topology-specific instances associated with each active topology.

This strategy separates two levels of the learning process: architectural and hyperparameter selection, performed at the module level, and electrical specialization of the model weights, performed at the topology level. In this way, performance differences among topologies can be primarily attributed to the electrical relationships inherent to each configuration, rather than to differences in architecture or hyperparameter settings. Figure 3 summarizes the overall procedure for training, validation, hyperparameter tuning, and final model selection.

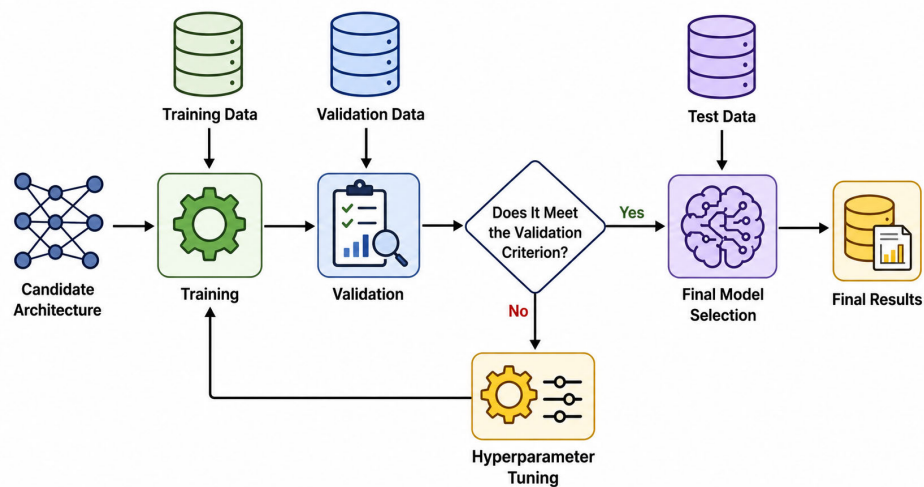


Figure 3. General procedure for training, validation, hyperparameter tuning, and final model selection.

Hyperparameter optimization was conducted via Bayesian optimization using the TPE (Tree-structured Parzen Estimator) sampler [54]. Since the early-stopping criterion does not trigger at a fixed epoch, each training run was assigned a random convergence epoch between 500 and 3000. All models were implemented in Python 3.11.10 (Python Software Foundation, Wilmington, DE, USA) with TensorFlow/Keras 2.19.0 (Google LLC, Mountain View, CA, USA) and trained on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 24 GB VRAM. Input normalization was performed via Min–Max scaling to $[0, 1]$, applied independently per topology. The complete Bayesian search space and the final selected configuration for each module are reported in Section 4.2.

3.6. Evaluation Metrics and Composite Score

To evaluate the performance of the candidate architectures, different metrics were used according to the nature of each module. For the classification modules, namely ATI, GED, ETI, and ELI, Accuracy, Precision, Recall, and F1-score were adopted. These metrics assess the overall performance of the classifier, the reliability of positive predictions, the ability to correctly identify the classes of interest, and the balance between Precision and

Recall. In addition, the complementary classification error, $E_{cls} = 1 - \text{Accuracy}$, was included to directly express the average proportion of misclassified samples.

For the regression modules, namely MRC and DSSE, MAE, RMSE, and MSE were used. MAE was adopted as the main interpretability metric because it provides a direct reading of the average error in the original evaluation scale. RMSE penalizes larger errors more strongly, whereas MSE was retained as a classical complementary reference metric.

To compare heterogeneous architectures using a common criterion, a Composite Score, denoted as CS , was defined by transforming the evaluation metrics into a common cost scale. For an architecture a , the Composite Score is defined as:

$$CS_a = \sum_{m=1}^M w_m C_{a,m} \quad (20)$$

where M is the number of metrics considered, w_m is the weight assigned to metric m , and $C_{a,m}$ is the normalized cost of architecture a for that metric. The weights satisfy:

$$\sum_{m=1}^M w_m = 1 \quad (21)$$

For classification tasks, a weight of 0.40 was assigned to the F1-score because it combines Precision and Recall into a balanced indicator, which is particularly relevant for error detection, identification, and localization modules. Accuracy, Precision, and Recall were each assigned a weight of 0.20, as they provide complementary information on overall performance, reliability of positive predictions, and the ability to correctly identify the classes of interest.

For regression tasks, the largest weights were assigned to MAE and RMSE because of their direct interpretation as reconstruction or estimation error. MAE represents the average absolute error, whereas RMSE penalizes larger errors more strongly. MSE was assigned a lower weight because it is expressed in squared units and provides information that is partially correlated with RMSE. Table 6 summarizes the weights used to compute the Composite Score by task type and the optimization direction of each metric.

Table 6. Weights used to compute the Composite Score according to the task type.

Task Type	Metric	Weight	Optimization Direction
Classification	Accuracy	0.20	Higher is better
Classification	Precision	0.20	Higher is better
Classification	Recall	0.20	Higher is better
Classification	F1-score	0.40	Higher is better
Regression	MAE	0.45	Lower is better
Regression	RMSE	0.45	Lower is better
Regression	MSE	0.10	Lower is better

As shown in Table 6, the assigned weights prioritize metrics with direct interpretability for each task type: F1-score for classification, because it balances Precision and Recall, and MAE/RMSE for regression, because they are directly related to the magnitude of the reconstruction or estimation error.

From an operational standpoint, the prioritization of the F1-score in classification tasks is motivated by the asymmetric consequences of diagnostic errors within the proposed pipeline. In GED, a false negative allows a compromised measurement to reach the estimator without correction, directly degrading DSSE accuracy. A false positive triggers an unnecessary correction that may introduce an artificial perturbation. In ETI and ELI, classification errors propagate an incorrect diagnosis to MRC. The F1-score, by simultane-

ously penalizing false positives (via Precision) and false negatives (via Recall), naturally captures this operational trade-off. For regression, MSE received a reduced weight ($w = 0.10$) because its squared scale hinders direct interpretation and its information content is partially redundant with RMSE.

To assess the sensitivity of architecture selection to the chosen weights, a perturbation analysis was performed in which each metric weight was varied by ± 0.10 relative to the nominal values in Table 6, subject to the constraint $\sum w_m = 1$. In all modules, the selected architecture (lowest CS) remained consistent under these perturbations.

To ensure comparability among metrics with different scales and optimization directions, each metric was transformed into a normalized cost within the interval $[0, 1]$. For metrics where lower values are better, such as MAE, RMSE, and MSE, the following transformation was used:

$$C_{a,m} = \frac{q_{a,m} - q_m^{\min}}{q_m^{\max} - q_m^{\min} + \epsilon} \quad (22)$$

For metrics where higher values are better, such as Accuracy, Precision, Recall, and F1-score, the following transformation was used:

$$C_{a,m} = \frac{q_m^{\max} - q_{a,m}}{q_m^{\max} - q_m^{\min} + \epsilon} \quad (23)$$

where $q_{a,m}$ represents the value of metric m for architecture a , whereas $q_m^{\min}$ and $q_m^{\max}$ are the minimum and maximum values observed among the compared architectures, respectively. The term ϵ is a small constant included to avoid division by zero.

In this way, all metrics are expressed as comparable costs, where $C_{a,m} = 0$ represents the best relative performance observed for a given metric and $C_{a,m} = 1$ represents the worst relative performance within the compared set.

The CS was computed independently for each module, comparing only architectures evaluated under the same task, dataset, and experimental conditions. Therefore, it should not be interpreted as an absolute metric comparable across modules, but rather as a relative criterion for ranking architectures within each subproblem.

Finally, the final architecture selection was performed by jointly considering the CS, the original metrics, and the complementary errors for classification tasks. A value of $CS = 0$ does not imply zero error or perfect performance; it only indicates that the corresponding architecture achieved the best relative performance among the models evaluated for a specific task.

4. Case Studies and Results Analysis

The proposed framework was validated on a real, unbalanced three-phase 240-node distribution feeder located in the Midwestern United States. The feeder was implemented in OpenDSS using the data reported in [21,22]. The evaluation is organized into two complementary levels: (i) a modular analysis aimed at comparing candidate architectures and selecting the best-performing model for each subproblem; and (ii) an integrated functional evaluation of the AI-based DSSE framework to quantify the impact of ATI and compromised-measurement treatment on the final state-estimation results.

4.1. Test System and Analyzed Scenarios

The network includes voltage, current, and power measurements represented in different formats, namely VoltPolar, CurrPolar, CurrRect, and PowersRect. The measurement vector $\mathbf{z}$ includes nodal voltage magnitudes and angles, branch-current magnitudes and angles, active and reactive power flows, nodal power measurements, and load pseudomeasurements.

A total of 12 predefined active topologies were considered. Figure 4 shows the location of the switching devices, which are denoted by cross symbols (×) in the feeder diagram. In all cases, *CB_101*, *CB_102*, and *CB_301* were assumed to be closed, whereas specific combinations of additional switches define the evaluated configurations.

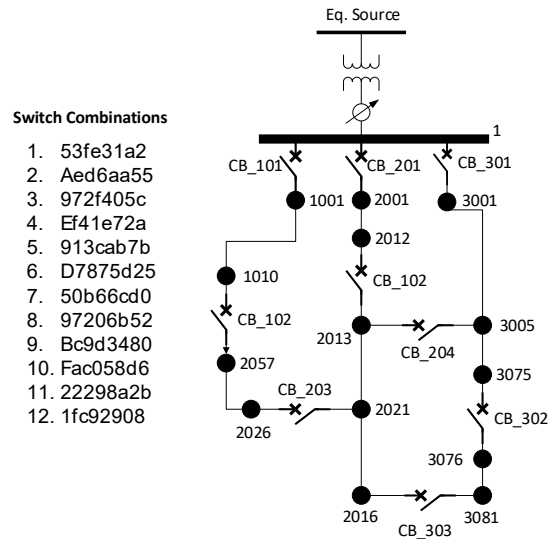


Figure 4. Location of the switching devices considered in the definition of the evaluated topological scenarios.

The number of configurations was not selected arbitrarily, but according to an operational criterion. The 12 configurations considered in this study correspond to the subset of switch-state combinations that preserve service to all nodes in the feeder, ensuring continuous radial supply without load interruption. Additional combinations within the switch-state space resulted in partial isolation of feeder sections and, consequently, loss of service at one or more nodes. These configurations fall outside the scope of a conventional state estimator because they correspond to contingency or fault-driven reconfiguration states, which require specific formulations not addressed in this work.

Therefore, the selected configurations represent the feasible operating space under normal and service-preserving reconfiguration conditions. This enables the evaluation of the models' ability to discriminate among electrically similar configurations within that operating space.

Figure 5 shows the spatial distribution and type of measurements considered, combining voltage, current, power-flow, nodal-injection, and load-pseudomeasurement information at the medium-voltage level. This configuration does not seek to solve an optimal meter-placement problem. Instead, it aims to reproduce a practical monitoring scenario with partial coverage and low redundancy.

The database was generated from hourly power-flow simulations covering the period from January to July 2017. At each time step, the 12 topologies were considered, and the required electrical variables were obtained. The uncertainty models described in Section 3.3 were applied to the simulated measurements, and gross errors were introduced in a controlled manner, affecting between one and five measurements per scenario.

The temporal organization was designed at two levels. First, January–June 2017 was used for the modular evaluation, with a stratified random split into training, validation, and test sets using a fixed seed. Second, July 2017 was reserved exclusively for the integrated functional evaluation presented in Section 4.3. This July dataset was not used for architecture selection, hyperparameter tuning, training, or modular testing.

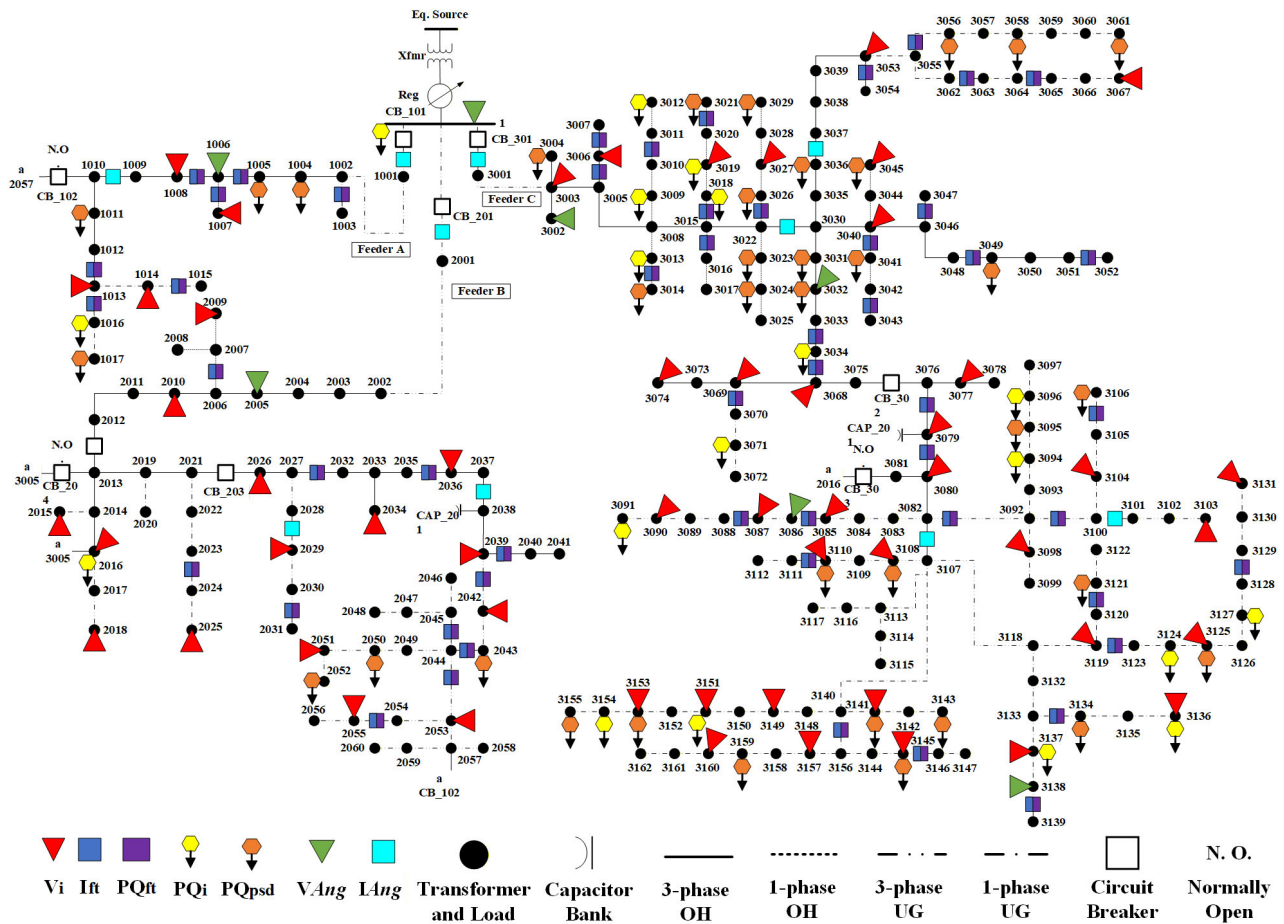


Figure 5. Spatial distribution of the measurements and pseudomeasurements used in the distribution network under study.

Although the stratified random split within January–June may introduce some correlation among subsets due to the cyclic nature of daily load profiles—a recognized issue when evaluating models on autocorrelated time series—the July dataset constitutes a strict temporal test outside the training window. Therefore, it is free from this specific effect and represents operating conditions not observed during model adjustment.

Stratification was applied by topology, presence or absence of gross errors, affected measurement type, and location. For regression tasks, stratification was performed according to the compromised measurement type to avoid biased data partitions.

4.2. Module-Level Results and Architecture Selection

Tables 7 and 8 summarize the average results of the three best-performing architectures for each classification and regression module, respectively. In Table 7, the MRC metrics should be interpreted as aggregate reconstruction indicators over heterogeneous electrical variables, whereas the DSSE metrics correspond to the global error of the state vector, which includes nodal voltage magnitudes and phase angles for each phase.

The reported CS values correspond to the original calculation over the complete set of candidate architectures evaluated in Table 4. Therefore, a value of $CS = 0$ identifies the best architecture within the full evaluated set, while intermediate values reflect relative positions within the original ranking. In all cases, the CS is used as a relative criterion within each module, whereas the original metrics provide the operational interpretation of model performance.

The high metric values observed in several modules indicate that the available electrical variables contain sufficiently discriminative patterns to solve the evaluated subproblems under the simulated conditions. However, metrics close to 1.00 should not be interpreted as a guarantee of perfect performance under general conditions, but rather as evidence of high separability within the specific experimental configuration analyzed.

Table 7. Average results of the classification modules.

Module	Architecture	Mean CS	Accuracy	Precision	Recall	F1-Score	Error 1 – Accuracy
ATI	WaveNet1D	0.000	1.0000	1.0000	1.0000	1.0000	0.00%
ATI	Inception1D	0.502	0.9997	0.9997	0.9997	0.9997	0.03%
ATI	SCINet1D	1.000	0.9994	0.9994	0.9994	0.9994	0.06%
GED	WaveNet1D	0.000	0.9537	0.9560	0.9538	0.9537	4.63%
GED	Inception1D	0.396	0.9474	0.9503	0.9475	0.9473	5.26%
GED	DilatedCNN1D	0.550	0.9449	0.9481	0.9450	0.9448	5.51%
ETI	WaveNet1D	0.000	0.9781	0.9800	0.9780	0.9785	2.19%
ETI	Inception1D	0.016	0.9639	0.9669	0.9637	0.9646	3.61%
ETI	U-Net1D	0.021	0.9575	0.9653	0.9572	0.9590	4.25%
ELI	EncDec-CNN	0.015	0.9582	0.9796	0.9464	0.9599	4.18%
ELI	U-Net1D	0.056	0.9515	0.9803	0.9402	0.9565	4.85%
ELI	SCINet1D	1.000	0.8693	0.8660	0.8185	0.8266	13.07%

Table 8. Average results of the regression modules.

Module	Architecture	Mean CS	MAE	MSE	RMSE
MRC	NBEATS1D	0.000	0.462182	1.515966	1.231246
MRC	WaveNet1D	0.002	0.475374	1.652431	1.285469
MRC	Inception1D	0.003	0.508803	1.562215	1.249886
DSSE	EncDec-GRU	0.055	0.001625	0.000011	0.003321
DSSE	NBEATS1D	0.062	0.001633	0.000011	0.003166
DSSE	Attention-RNN	0.338	0.002821	0.000032	0.005620

For the regression modules, MAE, MSE, and RMSE are reported in the corresponding evaluation scale used for each task. The Composite Score is used only as a relative comparison criterion within each module, according to the normalization procedure defined in Section 3.6. Therefore, a lower *CS* indicates better relative performance within the set of architectures evaluated for a given module, but it should not be interpreted as an absolute measure of physical error.

Figure 6 presents the comparative performance of the finalist architectures for ATI. WaveNet1D achieved *CS* = 0 and metric values equal to 1.000. Although Inception1D and SCINet1D also achieved errors below 0.1%, WaveNet1D obtained the best relative performance according to the *CS*. The central panel in Figure 6 uses a normalized visual scale within [0, 1], where 1 indicates the best relative performance for each metric, with the direction inverted for error metrics. The physical values should therefore be interpreted from Table 6.

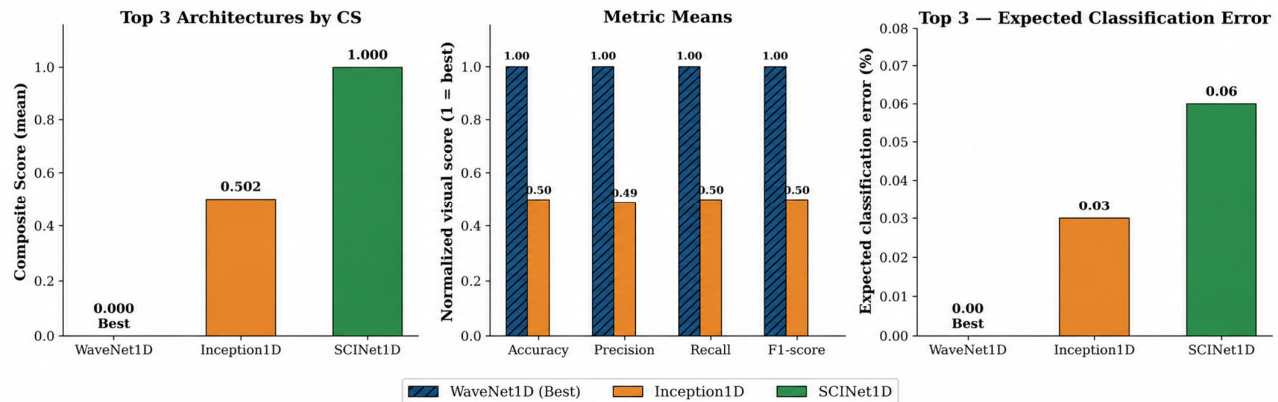


Figure 6. Comparative performance of the architectures evaluated for the ATI module.

Figure 7 presents the confusion matrix obtained with WaveNet1D on the ATI test set. The representation preserves the absolute counts per class, making it possible to verify both the absence of misclassifications and the number of evaluated samples per configuration. The 15,635 samples are concentrated on the main diagonal, with class supports ranging from 1261 to 1353. It is important to highlight that the open/closed status of the switching devices was not used as an input; therefore, the classification is based only on the observed electrical response. Nevertheless, the absence of misclassifications in the test set does not guarantee generalization to unseen topologies, structural changes, or different measurement configurations.

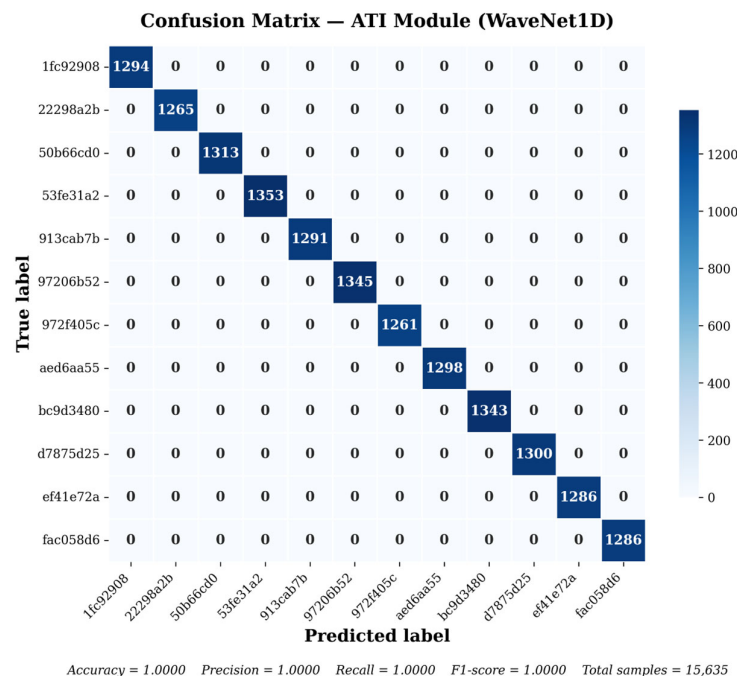


Figure 7. Confusion matrix of the ATI module using WaveNet1D.

In particular, the metric values of 1.00 obtained for ATI must be interpreted within the evaluated experimental domain: 12 known configurations that are electrically distinguishable and constructed under service-preserving operating conditions. This result indicates that, under the considered measurement placement and uncertainty levels, the

available electrical variables contain sufficiently separable patterns to identify the active topology without directly using switch-status information. However, this should not be interpreted as a guarantee of perfect performance under unseen topologies, structural modifications of the network, significant changes in meter placement, or operating conditions outside the range represented in the synthetic database. Therefore, generalization to configurations not included during training is recognized as a future extension of the framework.

For GED, WaveNet1D achieved an Accuracy of 0.9537 and an F1-score of 0.9537, outperforming Inception1D and DilatedCNN1D. This indicates an adequate ability to detect the presence of anomalous measurements, with a complementary error of 4.63%. For ETI, WaveNet1D also achieved the best performance, with an Accuracy of 0.9781 and an F1-score of 0.9785. Under the adopted iterative formulation, ETI does not operate as a simultaneous candidate-ranking mechanism; instead, it identifies the affected measurement type in the current iteration of the diagnosis and correction loop.

Similarly, ELI locates the compromised measurement associated with the type identified by ETI, allowing MRC to reconstruct a single measurement per iteration before updating the input vector and re-evaluating it through GED. The ELI module showed a higher relative difficulty, which is expected because localization requires distinguishing among nodes, phases, or elements with electrically similar responses. In this case, EncDec-CNN achieved the best relative performance, with Accuracy = 0.9582, Precision = 0.9796, Recall = 0.9464, and F1-score = 0.9599. This suggests that the encoder–decoder structure is suitable for capturing spatial relationships among measurement points.

To complement the analysis of the classification modules, Figures 8 and 9 present the confusion matrices for GED and ETI, respectively, evaluated on the test set for topology 1fc92908. This topology was selected as a representative example; the average metrics across all 12 topologies are reported in Table 7.

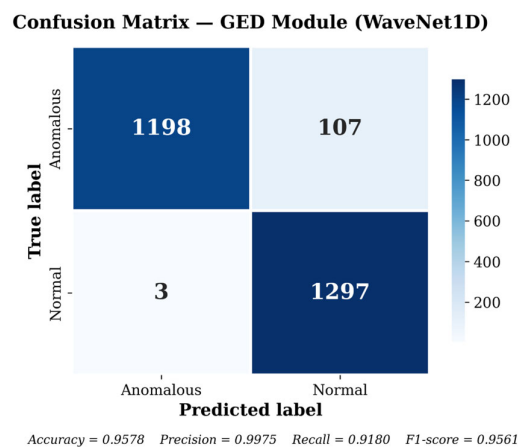


Figure 8. Confusion matrix of the GED module using WaveNet1D on topology 1fc92908.

Figure 8 shows the GED confusion matrix for binary anomaly detection using WaveNet1D on topology 1fc92908. The matrix reveals a pronounced asymmetry between error types: 107 false negatives (compromised measurements classified as normal) versus only 3 false positives (normal measurements classified as anomalous). This asymmetry confirms that the dominant risk lies in undetected gross errors passing directly to the estimator, which supports the prioritization of the F1-score in the Composite Score (Section 3.6). The topology-specific F1-score of 0.9561 is consistent with the cross-topology average of 0.9537 reported in Table 7.

Figure 9 shows the ETI confusion matrix for measurement-type identification using WaveNet1D on topology 1fc92908. The most frequent misclassifications occur between electrically correlated branch-level measurements: 72 instances of branch power flows (PQ_{ft}) misidentified as branch currents (I_{ft}), consistent with the physical coupling through the line impedance ($S = V \cdot I^*$). A secondary pattern involves confusion between nodal-level types (V_i , PQ_{psd}) and branch currents (I_{ft}), reflecting the electrical coupling between nodal voltages and branch flows in radial networks. Angular measurements ($I\text{Ang}_{ft}$, $V\text{Ang}_i$) are classified with near-perfect accuracy, which is expected given their distinct numerical range and statistical distribution. The topology-specific accuracy of 0.9774 is consistent with the cross-topology average of 0.9781 reported in Table 7.

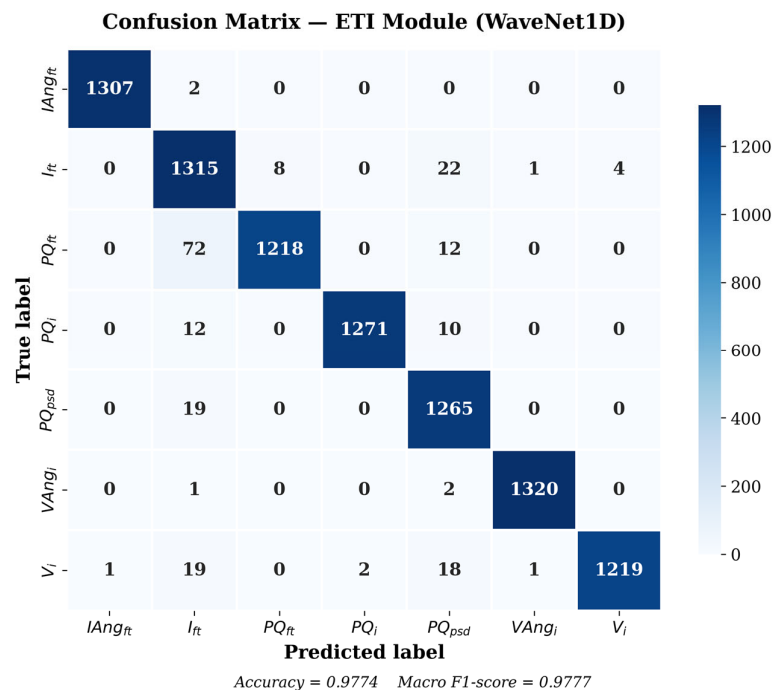


Figure 9. Confusion matrix of the ETI module using WaveNet1D on topology 1fc92908.

For the regression modules, NBEATS1D achieved the best CS and the lowest MAE in MRC, with an MAE of 0.4622. Because MRC reconstructs heterogeneous electrical variables, including voltage magnitudes, currents, powers, angles, and pseudomeasurements, the metrics for this module were computed on the normalized scale used during training. Therefore, the MRC MAE should be interpreted as an aggregate reconstruction indicator rather than as a physical error associated with a single electrical quantity. This clarification is necessary because the reconstructed variables have different units, numerical ranges, and uncertainty levels.

For DSSE, EncDec-GRU achieved the best global balance according to the CS , justifying its selection as the final architecture. Since the MAE value reported in Table 8 corresponds to a global indicator computed over the complete state vector, which includes nodal voltage magnitudes and phase angles for each phase, this result should be interpreted as an aggregate estimation error rather than as a physical error associated exclusively with the voltage magnitude in p.u. Therefore, Figure 10 should be interpreted as a relative comparison among architectures, while the final physical interpretation of performance is provided through the separate voltage-magnitude and voltage-angle metrics reported in the integrated evaluation.

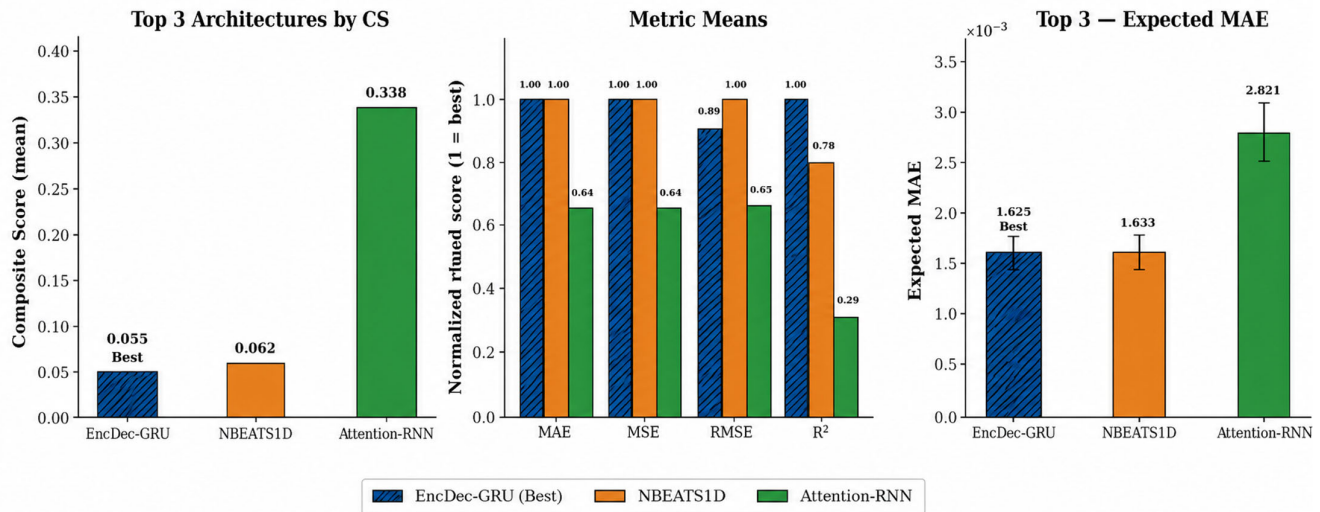


Figure 10. Comparative performance of the architectures evaluated for the state-estimation module.

It is worth noting that this level of accuracy is achieved under particularly demanding conditions, including unbalanced three-phase operation, 12 topological configurations, low measurement redundancy, and high uncertainty in pseudomeasurements. This demonstrates the ability of the proposed framework to maintain operational performance under realistic monitoring scenarios.

Because GED, ETI, ELI, MRC, and DSSE are trained using topology-specific instances, the variability of performance across configurations was also analyzed, as shown in Table 9. This analysis makes it possible to verify whether the average values represent homogeneous behavior or whether specific configurations concentrate greater difficulty. For classification modules, the best and worst cases were determined using the F1-score; for regression modules, they were determined using MAE.

Table 9. Performance variability across topologies for the topology-specific modules.

Module	Architecture	Metric	Mean	Std. Dev.	Best	Worst	Worst Topology
GED	WaveNet1D	F1-score	0.9537	3.42×10^{-3}	0.9589	0.9477	97206b52
ETI	WaveNet1D	F1-score	0.9785	3.49×10^{-3}	0.9837	0.9731	913cab7b
ELI	EncDec-CNN	F1-score	0.9599	5.90×10^{-3}	0.9679	0.9484	ef41e72a
MRC	NBEATS1D	MAE	4.62×10^{-1}	5.92×10^{-2}	4.15×10^{-1}	5.21×10^{-1}	1fc92908
DSSE	EncDec-GRU	MAE	1.62×10^{-3}	2.67×10^{-4}	1.38×10^{-3}	2.28×10^{-3}	913cab7b

The reduced standard deviations in GED and ETI indicate stable behavior across topologies. ELI shows slightly higher variability, with $\sigma = 5.90 \times 10^{-3}$, which is expected because localization is more sensitive to the active connectivity and to the electrical similarity among nodes. For MRC, the moderate dispersion, $\sigma = 5.92 \times 10^{-2}$, is consistent with the heterogeneity of the reconstructed variables.

For DSSE, the average MAE is 1.62×10^{-3} , with the worst case reaching 2.28×10^{-3} for topology 913cab7b. Overall, the topology-specific modeling strategy maintains stable performance across the evaluated configurations, without concealing severe systematic degradation in any topology. The absence of a dominant architecture across all modules confirms that ATI, GED, ETI, ELI, MRC, and DSSE impose different learning requirements, which supports the differentiated architecture selection adopted in the proposed modular framework.

Given that each module demands a distinct inductive bias, the Bayesian hyperparameter optimization was necessarily architecture-specific. Table 10 details the search space explored and the final configuration selected for each winning architecture. This optimization was performed at the module level: once the best configuration was identified for a given module, the same architecture and hyperparameters were applied to all topology-specific instances. Consequently, the performance differences across topologies reported in Table 10 can be attributed to the electrical characteristics inherent to each configuration rather than to variation in model design.

Table 10. Bayesian search space and final selected configuration for the winning architectures.

Mod.	Arch.	Bayesian Search Space	Final Configuration
ATI	WaveNet1D	Residual blocks: 12–32; dilation cycles: 3–5; pattern: [1, 2, 4, 8, 16] or [1, 2, 4, 8, 16, 32]; filters: 64, 128, 256; kernel: 2, 3, 5; dropout: 0.10–0.40; L2: 10^{-5} – 10^{-3} ; dense head: 64, 128, 256, 512 units	24 blocks in 4 cycles [1, 2, 4, 8, 16, 32], 128 filters, kernel 3, dropout 0.20, L2 = 2×10^{-4} , Batch Norm, dense 256 → 128.
GED	WaveNet1D	Blocks: 15–30; cycles: 3–5; pattern: [1, 2, 4, 8], [1, 2, 4, 8, 16] or [1, 2, 4, 8, 16, 32]; filters: 64, 128, 256; kernel: 2, 3, 5; dropout: 0.10–0.50; L2: 10^{-5} – 10^{-3} ; dense: 64, 128, 256, 512.	25 blocks in 5 cycles [1, 2, 4, 8, 16], 256 filters, kernel 2, dropout 0.30, L2 = 5×10^{-4} , Batch Norm, dense 256 → 128 → 64.
ETI	WaveNet1D	Blocks: 16–32; repetitions: 2–5; pattern: [1, 2, 4, 8, 16], [1, 2, 4, 8, 16, 32] or [1, 2, 4, 8, 16, 32, 64]; filters: 64, 128, 256; kernel: 2, 3, 5; dropout: 0.10–0.40; L2: 10^{-5} – 10^{-3} ; dense: 128, 256, 512.	28 blocks in 4 repetitions [1, 2, 4, 8, 16, 32, 64], 128 filters, kernel 3, dropout 0.25, L2 = 3×10^{-4} , Batch Norm, dense 512 → 256 → 128.
ELI	EncDec-CNN	Encoder levels: 3–5; initial filters: 32, 64, 128; ×2 progression; kernel: 3, 5, 7; strides: 1, 2; decoder levels: 3–5; upsampling: 1, 2; dropout: 0.10–0.40; L2: 10^{-5} – 10^{-3} ; Batch Norm: on/off.	Encoder 4 levels (64 → 128 → 256 → 512, kernels 5,5,3,3, strides 1,2,2,2); Decoder 4 levels (512 → 256 → 128 → 64, kernels 3, 3, 5, 5, upsampling 2, 2, 2, 1); dropout 0.10, L2 = 5×10^{-4} , Batch Norm.
MRC	NBEATS1D	Blocks: 2–6; dense layers/block: 4–12; units: 128, 256, 512, 1024; θ dim: 128, 256, 512; dropout: 0.00–0.30; L2: 10^{-5} – 10^{-3} .	4 blocks, 10 dense layers × 512 units, θ = 256, L2 = 5×10^{-4} .
DSSE	EncDec-GRU	Bidirectional encoder: 2–4 levels; units: 64, 128, 256, 512; decoder: 2–4 levels; units: 32, 64, 128, 256; dropout: 0.10–0.50; L2: 10^{-5} – 10^{-3} ; final dense: 32, 64, 128, 256.	Bidirectional encoder 3 levels (256 → 128 → 64); Decoder 3 levels (128 → 64 → 32); dropout 0.35, L2 = 10^{-3} , progressive dense 128 → 64 → 32.

4.3. Integrated Evaluation of the AI-Based DSSE Framework and Functional Contribution of Its Modules

This evaluation quantifies the functional contribution of the main modules through an ablation study. Unlike the modular analysis, its purpose is not to select architectures, but rather to assess how the omission of key stages affects the final DSSE accuracy. Four functional configurations were compared:

- Scenario 1—Complete AI-based DSSE framework: The active topology is identified through ATI, and the GED–ETI–ELI–MRC block is then applied iteratively. At each iteration, GED determines whether at least one anomalous measurement remains. If an anomaly is detected, ETI identifies the affected measurement type, ELI locates the compromised measurement, and MRC reconstructs its corrected value. The updated vector is re-evaluated until GED detects no further anomalies or the maximum number of iterations is reached. Finally, DSSE estimates the state using the cleaned measurement vector and the identified active topology.
- Scenario 2—Without compromised-measurement treatment: The GED, ETI, ELI, and MRC modules are omitted. Therefore, measurements affected by gross errors are neither diagnosed nor replaced before DSSE.

- Scenario 3—Without ATI: Active topology identification is omitted. Consequently, DSSE operates under a non-validated or incorrectly assigned topology.
- Scenario 4—Without ATI and without compromised-measurement treatment: Both active topology identification and the measurement diagnosis and correction modules are omitted. This represents the most functionally degraded case.

The degraded scenarios were constructed using proportional additive Gaussian perturbations, considering three severity levels, $\sigma_{GE} \in \{0.10, 0.15, 0.20\}$, and between one and five compromised measurements. Topological inconsistencies were modeled as substitutions through cyclic shifts of 1, 3, 6, and 9 positions with respect to the active topology.

The different number of cases across scenarios is due to the combinatorial structure of each case. Scenario 1 evaluates the complete framework over the July test window. Scenario 2 expands the evaluation across severity levels and numbers of compromised measurements. Scenario 3 expands it across topological shifts. Scenario 4 combines both spaces, namely, 4 topological shifts $\times$ 3 values of $\sigma_{GE} \times$ 5 numbers of compromised measurements, which explains its larger number of samples.

To prevent this difference in cardinality from biasing the comparison, MAE values were computed as weighted averages according to the number of samples, whereas RMSE values were computed from the weighted mean squared error. Table 11 presents the numerical summary, where V denotes voltage magnitude and θ denotes voltage angle. The degradation factors were computed with respect to Scenario 1.

Table 11. Functional contribution of the AI-based DSSE framework modules in the integrated evaluation.

Scen.	Cases	MAE_V	$RMSE_V$	MAE_θ	$RMSE_\theta$	Deg. V	Deg. θ
1	2976	1.11×10^{-3}	1.66×10^{-3}	8.41×10^{-3}	1.32×10^{-2}	1.00×	1.00×
2	5952	1.96×10^{-3}	6.06×10^{-3}	1.62×10^{-2}	7.27×10^{-2}	1.76×	1.92×
3	11,904	1.37×10^{-2}	1.93×10^{-2}	6.81×10^{-2}	9.21×10^{-2}	12.34×	8.10×
4	178,560	1.40×10^{-2}	1.97×10^{-2}	7.02×10^{-2}	9.99×10^{-2}	12.58×	8.35×

The complete framework, corresponding to Scenario 1, achieves the lowest errors in the integrated evaluation, with $MAE_V = 1.11 \times 10^{-3}$, $RMSE_V = 1.66 \times 10^{-3}$, $MAE_\theta = 8.41 \times 10^{-3}$, and $RMSE_\theta = 1.32 \times 10^{-2}$. These values define the baseline used to compute the degradation factors of the remaining scenarios.

The omission of compromised-measurement treatment, represented by Scenario 2, produces a moderate increase in error, with degradation factors of $1.76 \times$ for MAE_V and $1.92 \times$ for MAE_θ . This indicates that the GED–ETI–ELI–MRC block contributes to improving estimation robustness under gross measurement errors, although its impact is smaller than that of topology identification.

In contrast, omitting ATI, as in Scenario 3, produces a substantially larger degradation, with factors of $12.34 \times$ for MAE_V and $8.10 \times$ for MAE_θ . This result shows that an incorrect topology introduces a structural inconsistency that modifies the mapping among measurements, nodes, branches, and state variables. Therefore, topological error affects the estimation process globally, unlike a compromised measurement, whose effect tends to be more localized.

Scenario 4, in which both ATI and compromised-measurement treatment are omitted, exhibits the largest overall degradation, with factors of $12.58 \times$ for MAE_V and $8.35 \times$ for MAE_θ . However, the difference between Scenarios 3 and 4 is smaller than the difference between Scenarios 1 and 3. This confirms that, once DSSE operates under an incorrect topology, the structural error dominates over the additional effect introduced by compromised measurements.

Figure 11 provides a more detailed analysis of the scenarios in which topological discrepancies and compromised measurements are combined. For each topological shift, the mean MAE degradation factor is reported, computed as the average between voltage magnitude and voltage angle. The results show that degradation depends primarily on the topology assumed by the estimator. A one-position shift produces the largest degradation factors, whereas a six-position shift produces the smallest; three- and nine-position shifts exhibit intermediate behavior.

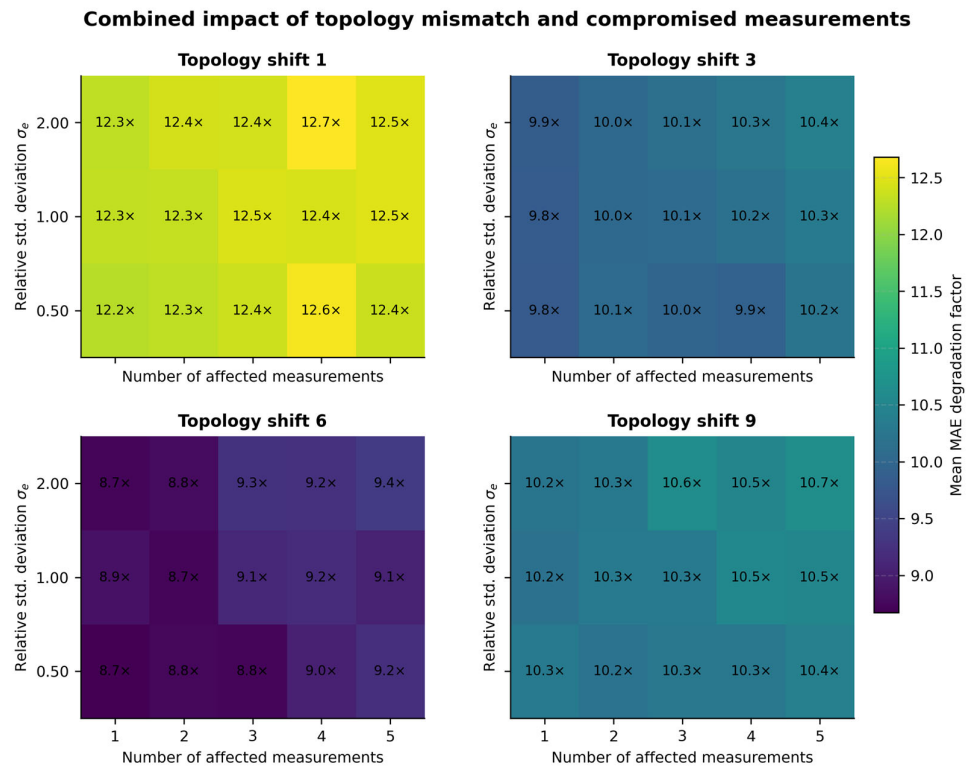


Figure 11. Combined impact of incorrect topology and compromised measurements.

This difference indicates that the severity of a topological inconsistency does not depend solely on the shift index but also on the electrical similarity between the actual and assumed topologies. Likewise, for a given shift, the variations associated with increasing σ_{GE} and the number of compromised measurements are smaller than the differences observed among topologies. This confirms that, when the topological representation is incorrect, the structural error dominates over the individual severity of the compromised measurements.

4.4. Convergence Assessment of a Classical WLS Estimator

To contextualize the performance of the proposed framework relative to classical state-estimation methods, WLS estimators based on both nodal voltages [4,5] and branch currents [9,13] were implemented on the 240-node test system, using the same measurements, pseudomeasurements, and uncertainty conditions described in Section 3.3.

In both cases, the WLS estimator failed to achieve numerical convergence. The Newton–Raphson iterative process diverged even when the convergence tolerance was relaxed from 10^{-4} to 10^{-2} p.u., and the maximum iteration count was increased from 20 to 100. Alternative initializations (flat start and warm start based on the power-flow solution) were also evaluated without achieving consistent convergence.

These findings are consistent with the convergence difficulties reported in the recent literature for distribution networks of similar or even smaller scale:

- (i) Habib et al. [19] reported that WLS never converges on the IEEE 179-bus network under noisy conditions with pseudomeasurements. Even when WLS was modified to use only voltage measurements (WLS) or when the tolerance was relaxed and iterations were increased (WLS*), convergence was not achieved on the 179-bus network.
- (ii) Langner and Abur [40] demonstrated that the conventional nodal-voltage WLS fails to converge within 20 iterations on the IEEE 123-bus feeder under flat start, identifying the angular-reference formulation as a fundamental cause in unbalanced systems.
- (iii) Gómez-Quiles et al. [41] documented that gain-matrix ill-conditioning—caused by the coexistence of measurements with highly disparate weights, insufficient redundancy, high R/X ratios, and the simultaneous presence of short and long lines—can produce both convergence failures and inaccurate estimates.
- (iv) A recent publication in this same journal [36] identified WLS iterative divergence on large-scale networks as motivation for proposing deep learning-based alternatives.

On the 240-node feeder evaluated in this work, the conditions that promote non-convergence are simultaneously present: native unbalanced three-phase operation with non-transposed lines ($Z_{aa} \neq Z_{bb} \neq Z_{cc}$, $Z_{ab} \neq Z_{bc} \neq Z_{ac}$), low measurement redundancy, pseudomeasurements with $\pm 30\%$ uncertainty ($\sigma_{\text{rel}} = 0.10$), high R/X ratios, and 12 active topological configurations. This combination produces a dense three-phase Jacobian with asymmetric impedances, a gain matrix G_{WLS} with an elevated condition number due to the disparity between direct measurements ($\sigma \approx 1\%$) and pseudomeasurements ($\sigma \approx 30\%$), and a critical dependence on the angular-reference definition.

It is noteworthy that, after an extensive literature review, no prior studies were identified reporting successful WLS-based DSSE results on this specific 240-node feeder [21,22]. Studies that have employed this test system have pursued data-driven approaches: pseudomeasurement generation [55], observability enhancement [56], and state and topology estimation via deep neural networks [45].

Therefore, the WLS non-convergence observed constitutes an experimental finding that reinforces the primary motivation of this work: the proposed deep learning framework offers a functional alternative for operating conditions with low redundancy, unbalanced three-phase operation, high pseudomeasurement uncertainty, and variable topologies—under which Newton–Raphson-based iterative methods do not guarantee numerical convergence.

4.5. Statistical Stability of Model Selection

To assess the variability inherent in deep neural network training, the selected architectures for each module were retrained with five independent random seeds (seeds: 0, 42, 123, 456, 789), keeping the optimal hyperparameters fixed. Table 12 reports the mean and standard deviation of the primary metric per module.

The coefficients of variation are below 1% for all classification modules and below 5% for the regression modules, indicating that inter-seed variability is small. Furthermore, the architecture with the lowest CSremained consistent across all five seeds for every module, confirming that the selection reported in Section 4.2 does not depend critically on any single training run.

Table 12. Statistical stability under multiple training seeds.

Module	Architecture	Metric	Mean $\pm$ Std (5 Seeds)	CV (%)
ATI	WaveNet1D	F1-score	$9.997 \times 10^{-1} \pm 3.0 \times 10^{-4}$	0.03
GED	WaveNet1D	F1-score	$9.525 \times 10^{-1} \pm 4.0 \times 10^{-3}$	0.42
ETI	WaveNet1D	F1-score	$9.770 \times 10^{-1} \pm 3.5 \times 10^{-3}$	0.36
ELI	EncDec-CNN	F1-score	$9.575 \times 10^{-1} \pm 6.0 \times 10^{-3}$	0.63
MRC	NBEATS1D	MAE	$4.640 \times 10^{-1} \pm 1.0 \times 10^{-2}$	2.16
DSSE	EncDec-GRU	MAE	$1.640 \times 10^{-3} \pm 8.0 \times 10^{-5}$	4.88

5. Discussion

The results support the central hypothesis of this work: DSSE under conditions of low redundancy, variable topologies, and potentially compromised measurements should not be formulated as a single monolithic regression task, but rather as a monitoring process composed of functionally differentiated stages. The module-dependent nature of the optimal architecture choice is evidenced by the fact that WaveNet1D was selected for ATI, GED, and ETI; EncDec-CNN for ELI; NBEATS1D for MRC; and EncDec-GRU for DSSE. This behavior confirms that the ATI, GED, ETI, ELI, MRC, and DSSE modules impose different learning requirements, since they involve topology classification, binary detection, error-type identification, spatial localization, measurement reconstruction, and electrical-state regression. Therefore, the differentiated selection of architectures makes it possible to adapt the inductive bias of each model to the specific function it performs within the AI-based DSSE pipeline.

This result is also consistent with the multitask learning literature, where it has been observed that the joint training of tasks or regimes that do not share sufficiently cooperative representations may degrade overall performance [43,44]. In this context, modularity not only facilitates the functional interpretation of the framework but also avoids forcing a single architecture to simultaneously solve subproblems with different objectives, outputs, and error structures. This is particularly relevant in distribution networks, where the relationship among measurements, errors, and state variables depends strongly on the active network connectivity.

The most relevant functional finding is that topological consistency has a dominant impact on the final DSSE accuracy. Omitting the compromised-measurement treatment produces a moderate degradation, with factors of $1.76 \times$ and $1.92 \times$ in the MAE of voltage magnitude and angle, respectively, whereas omitting ATI produces factors close to $12.3 \times$ and $8.1 \times$. This result has a direct physical explanation: a compromised measurement may locally degrade the input vector, and its impact on the estimator depends on the relative weight of the erroneous measurement within the overall measurement set. In contrast, an incorrect topology alters the structure of the measurement model $h(\cdot)$ itself and the associated Jacobian matrix, modifying the complete physical relationship among measurements, nodes, branches, and state variables [8,10–12]. This difference explains why the severity of a topological inconsistency does not depend monotonically on the number of switched elements, but rather on the electrical similarity between the actual and assumed topologies, as observed in Figure 9. This behavior is consistent with previous analyses of topology errors in state estimation [10,12]. For this reason, ATI should not be understood as an auxiliary stage, but as a structural component of the estimation pipeline that conditions the validity of the entire process. This result reinforces the perspective, present in the recent literature, that topology identification and state estimation should be treated as coupled problems rather than as independent sequential tasks [10–12].

A methodologically relevant aspect concerns the relationship between classical and data-driven approaches for DSSE. The WLS non-convergence observed in this work (Sec-

tion 4.4) should not be interpreted as a universal disqualification of classical estimators, but rather as experimental evidence that the specific operating conditions evaluated—the simultaneous combination of three-phase unbalance, low redundancy, high pseudomeasurement uncertainty, and variable topologies—lie outside the regime in which the conventional Newton–Raphson-based WLS provides robust numerical guarantees. This observation is consistent with the recent trend in the literature, where multiple authors have identified WLS non-convergence as an explicit motivation for developing deep learning-based alternatives [19,20,35,36]. In this context, classical and data-driven approaches should be understood not as mutually exclusive but as complementary: classical methods provide theoretical guarantees and interpretive transparency when numerical conditions permit, while data-driven methods offer robustness and flexibility when numerical conditions are degraded.

The strategy of combining a global model for ATI with topology-specific models for the subsequent modules offers a trade-off among physical specialization, training stability, and computational feasibility. On the one hand, it allows the downstream models to learn the electrical relationships specific to each configuration, avoiding the mixing of different topological regimes within a single representation. This consideration is consistent with the multitask learning literature, where joint training has been shown to degrade performance when tasks or regimes do not share sufficiently cooperative representations [43,44]. It also aligns with recent observations in learning-based DSSE, where the explicit incorporation of problem structure has been shown to improve generalization with respect to generic network models [19,20].

From a practical perspective, this specialization also reduces the memory burden associated with training, since it avoids forcing a single global model to simultaneously represent all combinations of topology, error diagnosis, measurement reconstruction, and state estimation. Consequently, the modular strategy enables the use of relatively smaller models, more stable training procedures, and computational requirements compatible with high-performance workstations, rather than basing this methodological choice exclusively on hardware constraints. On the other hand, the strategy increases the total number of required models. In general, the framework requires $1 + 5T$ models, where T is the number of operating topologies. For the case study evaluated ($T = 12$), this yields 61 models. It should be noted that the operationally feasible T is typically a small subset of the full combinatorial switch-state space, since many combinations result in loss of service or non-radial configurations.

For networks with a significantly larger number of feasible configurations ($T > 100$), three mitigation strategies are identified: (i) topology-conditioned global models, where a single model per module receives $\hat{\tau}$ as an auxiliary input (e.g., a one-hot vector or a learned embedding), reducing the total to six models; (ii) clustering of electrically similar topologies based on the Frobenius norm of the admittance-matrix difference or mean voltage profiles, so that a single model serves a cluster of $K \ll T$ related configurations; and (iii) transfer learning, where a model pre-trained on a base topology is fine-tuned with reduced data for similar configurations. These alternatives constitute natural extensions of the present work [2,18].

From the standpoint of computational cost, the complete process comprises two stages with distinct temporal implications. The computational workflow begins with OpenDSS simulations for the 12 topological configurations of the 240-node system. For each topology, hourly simulations were performed over a full year, yielding 8760 operating points per topology and 105,120 base scenarios in total. From these topology-specific datasets, additional samples were generated by injecting one to five simultaneous errors, yielding 525,600 error-augmented samples.

After dataset construction, an initial architecture screening was performed using ten candidate deep learning architectures: DilatedCNN1D, Inception1D, WaveNet1D, Conv-RNN, Attention-RNN, Encoder–Decoder CNN, NBEATS1D, Encoder–Decoder GRU, U-Net1D, and SCINet1D. The three best-performing architectures for each module were then selected for Bayesian hyperparameter optimization.

Using an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory, the total offline computational cost was estimated at approximately 1587.9 h, equivalent to 66.2 GPU-days. The majority of the cost was concentrated in Bayesian hyperparameter optimization, followed by final training and architecture screening. This cost is offline and incurred only once; it does not affect online operation.

In contrast, the online inference stage is negligible compared with the offline training process, with an estimated latency of approximately 0.45 ms per sample. In error-free scenarios, the pipeline executes only ATI and DSSE (two neural network inferences). In scenarios with Kgross errors, $2 + 4K$ sequential inferences are executed (ATI + Kiterations of GED + ETI + ELI + MRC + DSSE). Even in the worst case ($K = 5$, 22 inferences), the total time remains on the order of a few milliseconds.

These latencies are compatible with AMI-based monitoring cycles (minutes to one hour) and SCADA cycles (seconds to minutes). For μ PMU applications with strict real-time requirements (30–120 samples/second), the accumulated loop latency may require additional optimization such as model quantization, TensorRT compilation, or module parallelization.

During inference, only the models corresponding to the topology identified by ATI reside in memory simultaneously (five models: GED, ETI, ELI, MRC, DSSE) plus the ATI model itself—six models in total, regardless of the total number of topologies T . This is a direct advantage of the modular strategy: inference memory consumption does not scale with T .

An additional consideration is the iterative nature of the GED–ETI–ELI–MRC block. This strategy reduces the combinatorial complexity compared with a simultaneous multilabel formulation and preserves a traceable single-label supervised structure for each module. However, it also implies that diagnostic errors may propagate across iterations. Three principal failure modes are identified:

Mode 1: Propagation through incorrect type diagnosis ($\text{ETI} \rightarrow \text{ELI} \rightarrow \text{MRC}$): If ETI misidentifies the affected measurement type at iteration k , ELI will search for the location within an incorrect candidate set, and MRC may correct a healthy measurement, introducing an artificial perturbation into $\hat{z}^{(k+1)}$.

Mode 2: Confusion among electrically adjacent nodes (ELI): In networks with electrically similar nodes under the same topology and phase, ELI may localize the error at a node adjacent to the truly affected one. The cross-topology variability analysis (Table 8) confirms that ELI exhibits the highest dispersion ($\sigma = 5.90 \times 10^{-3}$ in F1-score), consistent with this susceptibility.

Mode 3: Accumulation under high noise with multiple errors: When pseudomeasurements with high uncertainty ($\sigma_{\text{rel}} = 0.10$) coexist with multiple gross errors, the separability between a measurement with a gross error and a noisy pseudomeasurement decreases, hindering GED discrimination in later iterations.

To mitigate these effects, the loop is limited to a maximum of five iterations. Even under partial correction, the iterative treatment produces a significant improvement in DSSE MAE relative to the untreated scenario (Scenario 2 vs. Scenario 1 in Table 9: degradation factors of $1.8\times$ and $1.9\times$). A natural extension of this work is the comparative evaluation of the sequential strategy against simultaneous multiple-error detection and correction formulations, such as normalized-residual ranking methods [14,23,46].

Transferring the proposed framework to real-time monitoring involves additional considerations. Regarding latency, the ~ 0.45 ms per-sample inference time is amply sufficient for AMI and SCADA cycles. For μ PMU applications, inference optimization would be required. Regarding retraining, significant network changes—new loads, permanent reconfiguration, new sensors—would require partial or full dataset regeneration and retraining of the affected modules. Regarding robustness to unmodeled conditions, phenomena such as partial communication failures, temporal desynchronization among data sources, line-parameter errors, and transient events were not explicitly modeled and should be evaluated prior to operational deployment.

Limitations

The results should be interpreted within the scope of the study:

First, validation was performed exclusively with synthetic OpenDSS-generated scenarios. This choice is driven by the specific requirements of the proposed framework: training and evaluation simultaneously require (i) synchronized three-phase unbalanced measurements, (ii) documented time-varying active topology, (iii) gross errors with known type, location, and magnitude, and (iv) the true reference state. No publicly available dataset currently satisfies all four requirements simultaneously. μ PMU-based datasets provide synchronized measurements but lack documented variable topology and labeled errors; AMI datasets provide load profiles without synchronized network measurements. For these reasons, synthetic data generation via validated simulators such as OpenDSS is the predominant strategy in the recent DL-based DSSE literature [15,16,19,20,39,45]. A viable intermediate strategy for future work would be the injection of real μ PMU measurements into the OpenDSS simulation to generate semi-synthetic scenarios with realistic noise and temporal correlations.

Second, the measurement placement was defined as a representative configuration with partial coverage and low redundancy, not as the result of an optimal meter-placement procedure.

Third, the modular metrics correspond to training processes performed with a fixed seed. The multi-seed results reported in Section 4.5 demonstrate stability; however, an exhaustive characterization would require a larger number of repetitions.

Fourth, the study evaluates 12 service-preserving configurations. More severe scenarios—contingencies, faults, feeder-section isolation, partial loss of supply—were not considered and could generate more complex electrical patterns with potential discontinuities in state variables.

Fifth, the scalability of the topology-specific approach should be assessed for $T > 100$. The alternatives discussed in this section (conditioned models, clustering, and transfer learning) have not been implemented or experimentally evaluated.

Sixth, the WLS estimators based on nodal voltages and branch currents failed to achieve numerical convergence on the evaluated test system (Section 4.4). Although this result is consistent with the literature [19,40], the evaluation was limited to classical WLS formulations. More advanced robust formulations—WLS with virtual reference [40], semidefinite programming, Kalman filters—may mitigate these issues and constitute complementary baselines.

Seventh, the framework assumes that the network electrical parameters are known without error. Parametric uncertainty constitutes an additional unmodeled source of degradation.

Eighth, the framework's ability to generalize to topologies not included in training was not evaluated and represents a relevant limitation for operational applications where unforeseen configurations may arise.

6. Conclusions

This work presented an integrated and modular deep learning framework for DSSE in unbalanced three-phase distribution networks, sequentially combining active topology identification (ATI), gross error detection (GED), error-type identification (ETI), error-location identification (ELI), measurement reconstruction and correction (MRC), and final state estimation (DSSE). The framework pairs a global ATI model with topology-specific downstream modules and treats compromised measurements through an iterative single-correction loop that avoids the combinatorial complexity of simultaneous multilabel identification while preserving a traceable supervised structure.

The results demonstrate that no single architecture dominates across all subproblems—WaveNet1D excels in ATI, GED, and ETI; EncDec-CNN in ELI; NBEATS1D in MRC; and EncDec-GRU in DSSE—confirming that module-specific architecture selection enables the framework to exploit inductive biases appropriate for each task. The ablation study reveals that topological consistency is the dominant factor in estimation accuracy: omitting ATI degrades voltage-magnitude and voltage-angle MAE by factors of 12.3× and 8.1×, respectively, whereas omitting compromised-measurement treatment produces a moderate degradation of 1.8×–1.9×.

Additionally, the non-convergence of classical WLS estimators observed on the 240-node test system—consistent with recent reports for distribution networks of similar scale—reinforces the relevance of data-driven alternatives for operating conditions characterized by high three-phase dimensionality, low redundancy, high-uncertainty pseudo-measurements, and variable topologies. The total offline computational cost was approximately 1587.9 h (66.2 GPU-days), with an online inference latency of ~0.45 ms per sample, compatible with AMI- and SCADA-based monitoring cycles.

Future research directions include: (i) validation with real or hybrid data from operating systems; (ii) expansion to post-fault topologies and feeder-section isolation; (iii) implementation of scalability strategies, particularly topology-conditioned models and transfer learning; (iv) comparison with robust classical estimators, including branch-current formulations; (v) comparative evaluation of sequential versus simultaneous error-correction strategies; (vi) incorporation of parametric uncertainty and missing data; (vii) exploration of GNN architectures; and (viii) inference optimization for real-time μ PMU monitoring.

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

Funding: This research received no external funding.

Data Availability Statement: The source distribution feeder model used in this study is available from the cited public references. The original datasets generated in this work are not publicly available because they are part of an ongoing research line and include simulation-based scenarios, pre-processing procedures, topology configurations, gross-error cases, and labeled training datasets developed specifically for this study. The main results supporting the conclusions are included in the article. Further inquiries may be directed to the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to the Institute of Electrical Energy (IEE)—CONICET—National University of San Juan, for providing the research environment and resources that supported this work. During the preparation of this work, the authors

used ChatGPT (GPT-5, OpenAI, San Francisco, CA, USA) for language editing and readability improvement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
ATI	Active Topology Identification
CNN	Convolutional Neural Network
CNN1D	One-Dimensional Convolutional Neural Network
CS	Composite Score
DilatedCNN1D	One-Dimensional Dilated Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSS ²	Deep Statistical Solver
DSSE	Distribution System State Estimation
ELI	Error-Location Identification
EncDec	Encoder–Decoder
EncDec-CNN	Encoder–Decoder Convolutional Neural Network
EncDec-GRU	Encoder–Decoder Gated Recurrent Unit
ETI	Error-Type Identification
F1-score	Harmonic Mean of Precision and Recall
FDI	False Data Injection
FDIA	False Data Injection Attack
FEDformer	Frequency Enhanced Decomposed Transformer
GATNE	Graph Attention Network with Topology-aware Node Embedding
GED	Gross Error Detection
GNN	Graph Neural Network
GPU	Graphics Processing Unit
GraphSAGE	Graph Sample and Aggregate Network
GRU	Gated Recurrent Unit
Inception1D	One-Dimensional Inception Network
Informer	Efficient Transformer for Long Sequence Forecasting
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MLP-Mixer	Multilayer Perceptron Mixer
MRC	Measurement Reconstruction and Correction
MSE	Mean Squared Error
NBEATS1D	One-Dimensional Neural Basis Expansion Analysis for Time Series
OpenDSS	Open Distribution System Simulator
PINN	Physics-Informed Neural Network
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SCADA	Supervisory Control and Data Acquisition
SCINet1D	One-Dimensional Sample Convolution and Interaction Network
TPE	Tree-structured Parzen Estimator
U-Net1D	One-Dimensional U-Shaped Convolutional Neural Network
WaveNet1D	One-Dimensional WaveNet
WLS	Weighted Least Squares
WLAV	Weighted Least Absolute Value
μPMU	Micro-Phasor Measurement Unit

References

1. Yadav, A.P.; Nutaro, J.; Park, B.; Dong, J.; Liu, B.; Yoginath, S.B.; Yin, H.; Dong, J.; Dong, Y.; Liu, Y.; et al. Review of Emerging Concepts in Distribution System State Estimation: Opportunities and Challenges. *IEEE Access* **2023**, *11*, 70503–70515. <https://doi.org/10.1109/ACCESS.2023.3292793>.
2. Matijašević, T.; Antić, T.; Capuder, T. A Systematic Review of Machine Learning Applications in the Operation of Smart Distribution Systems. *Energy Rep.* **2022**, *8*, 12379–12407. <https://doi.org/10.1016/j.egyr.2022.09.068>.
3. Kabalci, Y. A Survey on Smart Metering and Smart Grid Communication. *Renew. Sustain. Energy Rev.* **2016**, *57*, 302–318. <https://doi.org/10.1016/j.rser.2015.12.114>.
4. Raghuvamsi, Y.; Teeparthi, K. A Review on Distribution System State Estimation Uncertainty Issues Using Deep Learning Approaches. *Renew. Sustain. Energy Rev.* **2023**, *187*, 113752. <https://doi.org/10.1016/j.rser.2023.113752>.
5. Fotopoulou, M.; Petridis, S.; Karachalios, I.; Rakopoulos, D. A Review on Distribution System State Estimation Algorithms. *Appl. Sci.* **2022**, *12*, 11073. <https://doi.org/10.3390/app122111073>.
6. Muscas, C.; Pau, M.; Pegoraro, P.A.; Sulis, S. Effects of Measurements and Pseudomeasurements Correlation in Distribution System State Estimation. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 2813–2823. <https://doi.org/10.1109/TIM.2014.2318391>.
7. Singh, R.; Pal, B.C.; Vinter, R.B. Measurement Placement in Distribution System State Estimation. *IEEE Trans. Power Syst.* **2009**, *24*, 668–675. <https://doi.org/10.1109/TPWRS.2009.2016457>.
8. de Melo, V.H.P.; London, J.B.A.; Massignan, J.A.D. Distribution System State Estimation Algorithm with Improved Angular Reference Treatment. *Electr. Power Syst. Res.* **2022**, *212*, 108572. <https://doi.org/10.1016/j.epsr.2022.108572>.
9. Baran, M.E.M.; Kelley, A.W.A. A Branch-Current-Based State Estimation Method for Distribution Systems. *IEEE Trans. Power Syst.* **1995**, *10*, 483–491. <https://doi.org/10.1109/59.373974>.
10. Wang, C.; Lou, Z.; Li, M.; Zhu, C.; Jing, D. Identification of Distribution Network Topology and Line Parameter Based on Smart Meter Measurements. *Energies* **2024**, *17*, 830. <https://doi.org/10.3390/en17040830>.
11. Razmi, P.; Asl, M.G.; Canarella, G.; Emami, A.S. Topology Identification in Distribution System via Machine Learning Algorithms. *PLoS ONE* **2021**, *16*, e0252436. <https://doi.org/10.1371/journal.pone.0252436>.
12. Gotti, D.; Ledesma, P.; Amaris, H. A Fast Data-Driven Topology Identification Method for Dynamic State Estimation Applications. *Int. J. Electr. Power Energy Syst.* **2023**, *147*, 108807. <https://doi.org/10.1016/j.ijepes.2022.108807>.
13. Lin, W.M.; Teng, J.H.; Chen, S.J. A Highly Efficient Algorithm in Treating Current Measurements for the Branch-Current-Based Distribution State Estimation. *IEEE Trans. Power Deliv.* **2001**, *16*, 433–439. <https://doi.org/10.1109/61.924823>.
14. Handschin, E.; Schweppe, F.C.; Kohlas, J.; Fiechter, A. Bad Data Analysis for Power System State Estimation. *IEEE Trans. Power Appar. Syst.* **1975**, *94*, 329–337. <https://doi.org/10.1109/T-PAS.1975.31858>.
15. Radhoush, S.; Vannoy, T.; Liyanage, K.; Whitaker, B.M.; Nehrir, H. Distribution System State Estimation and False Data Injection Attack Detection with a Multi-Output Deep Neural Network. *Energies* **2023**, *16*, 2288. <https://doi.org/10.3390/EN16052288>.
16. Raghuvamsi, Y.; Teeparthi, K. Detection and Reconstruction of Measurements against False Data Injection and DoS Attacks in Distribution System State Estimation: A Deep Learning Approach. *Measurement* **2023**, *210*, 112565. <https://doi.org/10.1016/j.MEASUREMENT.2023.112565>.
17. Santhosh Kumar, C.; Rajawat, K.; Chakrabarti, S.; Pal, B.C.; Kumar, C.S.; Rajawat, K.; Chakrabarti, S.; Pal, B.C. Robust Distribution System State Estimation with Hybrid Measurements. *IET Gener. Transm. Distrib.* **2020**, *14*, 3250–3259. <https://doi.org/10.1049/iet-gtd.2020.0260>.
18. Békési, G.; Barancsik, L.; Hartmann, B. Deep Neural Network Based Distribution System State Estimation Using Hyperparameter Optimization. *Results Eng.* **2024**, *24*, 102908. <https://doi.org/10.1016/j.rineng.2024.102908>.
19. Habib, B.; Isufi, E.; van Breda, W.; Jongepier, A.; Cremer, J.L. Deep Statistical Solver for Distribution System State Estimation. *IEEE Trans. Power Syst.* **2024**, *39*, 4039–4050. <https://doi.org/10.1109/TPWRS.2023.3290358>.
20. Tran, M.-Q.Q.; Zamzam, A.S.; Nguyen, P.H. Enhancement of Distribution System State Estimation Using Pruned Physics-Aware Neural Networks. In *Proceedings of the 2021 IEEE Madrid PowerTech*; IEEE: New York, NY, USA, 2021; pp. 1–5. <https://doi.org/10.1109/PowerTech46648.2021.9494950>.
21. Bu, F.; Yuan, Y.; Wang, Z.; Dehghanpour, K.; Kimber, A. A Time-Series Distribution Test System Based on Real Utility Data. In *Proceedings of the 2019 North American Power Symposium (NAPS)*; IEEE: New York, NY, USA, 2019; pp. 1–6. <https://doi.org/10.1109/NAPS46351.2019.8999982>.
22. Dugan, R.C.; McDermott, T.E. An Open Source Platform for Collaborating on Smart Grid Research. In *Proceedings of the 2011 IEEE Power and Energy Society General Meeting*; IEEE: New York, NY, USA, 2011; pp. 1–7. <https://doi.org/10.1109/PES.2011.6039829>.

23. Cutsem, T.; Ribbens-Pavella, M.; Mili, L. Hypothesis Testing Identification: A New Method For Bad Data Analysis In Power System State Estimation. *IEEE Trans. Power Appar. Syst.* **1984**, PAS-103, 3239–3252. <https://doi.org/10.1109/TPAS.1984.318561>.
24. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**. <https://doi.org/10.48550/arXiv.1803.01271>.
25. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**. <https://doi.org/10.48550/arXiv.1609.03499>.
26. Indolia, S.; Goswami, A.K.; Mishra, S.P.; Asopa, P. Conceptual Understanding of Convolutional Neural Network—A Deep Learning Approach. *Procedia Comput. Sci.* **2018**, 132, 679–688. <https://doi.org/10.1016/j.procs.2018.05.069>.
27. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.-A.; Petitjean, F. InceptionTime: Finding AlexNet for Time Series Classification. *Data Min. Knowl. Discov.* **2020**, 34, 1936–1962. <https://doi.org/10.1007/s10618-020-00710-y>.
28. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
29. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the EMNLP 2014—2014 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1724–1734. <https://doi.org/10.3115/v1/d14-1179>.
30. Sutskever, I.; Vinyals, O.; Le, Q. V. Sequence to Sequence Learning with Neural Networks. *Adv. Neural Inf. Process. Syst.* **2014**, 4, 3104–3112.
31. Schuster, M.; Paliwal, K.K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Process.* **1997**, 45, 2673–2681. <https://doi.org/10.1109/78.650093>.
32. Luong, M.T.; Pham, H.; Manning, C.D. Effective Approaches to Attention-Based Neural Machine Translation. In *Proceedings of the EMNLP 2015—2015 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2015; pp. 1412–1421. <https://doi.org/10.18653/v1/d15-1166>.
33. Liu, M.; Zeng, A.; Chen, M.; Xu, Z.; Lai, Q.; Ma, L.; Xu, Q. SCINet: Time Series Modeling and Forecasting with Sample Convolution and Interaction. In *Proceedings of the Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation: San Diego, CA, USA, 2022; Volume 35.
34. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. In *Proceedings of the 8th International Conference on Learning Representations, ICLR 2020, Virtual, 26 April–1 May 2020*.
35. Zhai, B.; Yang, D.; Zhou, B.; Li, G. Distribution System State Estimation Based on Power Flow-Guided GraphSAGE. *Energies* **2024**, 17, 4317. <https://doi.org/10.3390/EN17174317>.
36. Qiu, Y.; Zhou, S.; Cui, S.; Sun, Q.; Liu, S.; Tang, Z.; Chai, B.; Zeng, Z. Robust Distribution System State Estimation with Physics-Constrained Heterogeneous Graph Embedding and Cross-Modal Attention. *Processes* **2025**, 13, 3073. <https://doi.org/10.3390/PR13103073>.
37. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, 35, 11106–11115. <https://doi.org/10.1609/aaai.v35i12.17325>.
38. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. In *Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022*; pp. 27268–27286.
39. Zamzam, A.S.; Sidiropoulos, N.D. Physics-Aware Neural Networks for Distribution System State Estimation. *IEEE Trans. Power Syst.* **2020**, 35, 4347–4356. <https://doi.org/10.1109/TPWRS.2020.2988352>.
40. Langner, A.L.; Abur, A. Formulation of Three-Phase State Estimation Problem Using a Virtual Reference. *IEEE Trans. Power Syst.* **2021**, 36, 214–223. <https://doi.org/10.1109/TPWRS.2020.3004076>.
41. Gómez-Quiles, C.; Romero-Ramos, E.; de la Villa-Jaén, A.; Gómez-Expósito, A. Compensated Load Flow Solutions for Distribution System State Estimation. *Energies* **2020**, 13, 3100. <https://doi.org/10.3390/en13123100>.
42. Dehghanpour, K.; Wang, Z.; Wang, J.; Yuan, Y.; Bu, F. A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems. *IEEE Trans. Smart Grid* **2019**, 10, 2312–2322. <https://doi.org/10.1109/TSG.2018.2870600>.
43. Crawshaw, M. Multi-Task Learning with Deep Neural Networks: A Survey. *arXiv* **2020**. <https://doi.org/10.48550/arXiv.2009.09796>.

44. Standley, T.; Zamir, A.; Chen, D.; Guibas, L.; Malik, J.; Savarese, S. Which Tasks Should Be Learned Together in Multi-Task Learning? In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual, 13–18 July 2020; pp. 9120–9132.
45. Azimian, B.; Sen Biswas, R.; Moshtag, S.; Pal, A.; Tong, L.; Dasarathy, G. State and Topology Estimation for Unobservable Distribution Systems Using Deep Neural Networks. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 9003514. <https://doi.org/10.1109/TIM.2022.3167722>.
46. Bretas, A.S.; Bretas, N.G.; Braunstein, S.H.; Rossoni, A.; Trevizan, R.D. Multiple Gross Errors Detection, Identification and Correction in Three-Phase Distribution Systems WLS State Estimation: A per-Phase Measurement Error Approach. *Electr. Power Syst. Res.* **2017**, *151*, 174–185. <https://doi.org/10.1016/j.epsr.2017.05.021>.
47. Xu, W.; Wang, M.; Tang, A. On State Estimation with Bad Data Detection. In *Proceedings of the 2011 50th IEEE Conference on Decision and Control and European Control Conference*; IEEE: New York, NY, USA, 2011; pp. 5989–5994. <https://doi.org/10.1109/CDC.2011.6161214>.
48. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533–536. <https://doi.org/10.1038/323533a0>.
49. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. <https://doi.org/10.1126/science.1127647>.
50. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
51. Vaswani, A.; Brain, G.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need (Transformer Architecture). In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 1–15.
52. Tolstikhin, I.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. MLP-Mixer: An All-MLP Architecture for Vision. *Adv. Neural Inf. Process. Syst.* **2021**, *29*, 24261–24272.
53. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2951–2959.
54. Ozaki, Y.; Watanabe, S.; Yanase, T. OptunaHub: A Platform for Black-Box Optimization. *arXiv* **2025**. <https://doi.org/10.48550/arXiv.2510.02798>.
55. Dehghanpour, K.; Yuan, Y.; Wang, Z.; Bu, F. A Game-Theoretic Data-Driven Approach for Pseudo-Measurement Generation in Distribution System State Estimation. *IEEE Trans. Smart Grid* **2019**, *10*, 5942–5951. <https://doi.org/10.1109/TSG.2019.2893818>.
56. Yuan, Y.; Dehghanpour, K.; Bu, F.; Wang, Z. A Multi-Timescale Data-Driven Approach to Enhance Distribution System Observability. *IEEE Trans. Power Syst.* **2019**, *34*, 3168–3177. <https://doi.org/10.1109/TPWRS.2019.2893821>.

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.