Transfer Learning for PMU-Based Online Inertia Monitoring under Domain Shift

Hangdong AN
School of Electrical and Computer Engineering
The University of Sydney
Sydney, Australia
hangdong.an@sydney.edu.au

Jin MA
School of Electrical and Computer Engineering
The University of Sydney
Sydney, Australia
j.ma@sydney.edu.au

Thomas Chaffey
School of Electrical and Computer Engineering
The University of Sydney
Sydney, Australia
thomas.chaffey@sydney.edu.au

***Abstract*—System inertia varies with operating conditions and network reconfigurations, which makes measurement-based estimators noise-sensitive and purely data-driven models brittle under topology change. Recent reviews also highlight two practical bottlenecks for inertia monitoring: scarcity of labeled data and degradation under distribution shift. This work presents a scenario-aware transfer-learning framework for phasor measurement unit (PMU)-based online inertia estimation. A long-term recurrent convolutional network is pretrained on a source system and, at deployment, adapted per scenario using a validation-selected route among parameter fine-tuning, feature alignment, and supervised contrastive adaptation. A lightweight regressor is few-shot tuned with limited target labels, optionally augmented by confidence-screened pseudo-labels, while standardised preprocessing and channel adapters align heterogeneous PMU layouts to a fixed-width encoder. The design builds on prior LRCN-style inertia estimation from PMU data and established domain-adaptation principles. Experiments on IEEE-39 and IEEE-118 under representative operating and topology changes, including N-1/N-2 outages, motor-load dynamics, and wind-turbine replacement, show consistently lower error than from-scratch training under matched label budgets across the studied scenarios, indicating label-efficient adaptation for online inertia monitoring.**

***Index Terms*—Machine learning, Inertia estimation, Transfer learning, Domain adaptation, Supervised contrastive learning**

I. Introduction

With the rapid integration of inverter-based wind and solar generation, system inertia has become increasingly time-varying, resulting in higher rates of change of frequency (RoCoF), narrower frequency margins, and degraded performance of traditional protection and control schemes. Accurate, near-real-time estimation of equivalent system inertia is therefore essential to maintain frequency stability and enable secure operation.Conventional inertia estimation methods are commonly grouped into measurement-based and model-based categories. Measurement-based estimators infer inertia from frequency or RoCoF responses during large disturbances [1], from small-signal injections [2], or from ambient fluctuations using covariance or ARMAX modeling [3]. Model-based estimators fit swing-equation formulations to system responses [4]–[6].These approaches depend on accurate parameters and linearised assumptions, so parameter uncertainty, unmodelled load dynamics, and network reconfigurations often degrade their robustness under changing operating conditions.

Data-driven estimators address part of these limitations by learning nonlinear mappings from PMU measurements to inertia. Deep neural architectures, including CNN–LSTM, graph neural networks (GNNs), and physics-informed model have reported low errors on benchmark systems [7]–[9]. In practice, two hurdles persist: substantial labelled datasets are needed to reach those accuracies, and accuracy degrades under topology or operating changes, because PMU measurement distributions and early-time frequency responses remain system- and scenario-dependent. Diggikar *et al.* [10] further argue that transfer learning and few-shot adaptation will be needed for practical inertia monitoring, but stop at outlining this as a research direction rather than providing an implemented framework. In this context, the present work takes PMU-based learning models as its starting point and focuses on improving their label-efficiency and robustness under distribution shift, rather than revisiting numerical comparisons with classical analytic estimators. Transfer learning (TL) provides a means to reuse pretrained knowledge from a source domain in a related but data-scarce target domain. Prior successes include voltage stability prediction [11], hierarchical load forecasting [12], and cross-region demand prediction [13]. However, for inertia estimation, TL has so far only been identified as a promising direction [10]: there is still no end-to-end framework that (i) explicitly treats PMU-based inertia estimation as a cross-domain problem with topology and operating-point shifts, and (ii) systematically compares different adaptation routes under matched label budgets. This leaves a gap between the conceptual call for transfer/few-shot learning and practical deployment in real grids.

To bridge this gap, this study presents an end-to-end TL framework for online inertia estimation from PMU frequency/rate of change of frequency (RoCoF) traces. Three adaptation routes are considered—parameter fine-tuning (small shift), feature alignment (moderate), and supervised contrastive adaptation (large)—with the final choice selected by validation under a fixed target-label budget. A lightweight regression head is then few-shot fine-tuned on limited target labels, with uncertainty-aware pseudo-labelling applied when appropriate.

979-8-3315-8138-1

Inputs across systems are aligned through standardised preprocessing and channel masking.

1) Introduces transfer learning into PMU-based inertia estimation as a framework for handling cross-scenario/domain shifts, enabling models to be adapted across varying operating conditions and network topologies.
2) Develops a validation-selected transfer framework (parameter/feature/representation adaptation) that maintains a fixed label budget and reuses a single encoder across scenarios.
3) Provides a unified evaluation on IEEE-39/118 with N–1/N–2 outages, motor dynamics, and wind replacement, illustrating the label efficiency and robustness of the proposed transfer framework.

The remainder of this paper is organised as follows. Section II formulates the inertia estimation problem and reviews swing-equation fundamentals. Section III presents the proposed three-stage transfer-learning framework. Section IV provides case studies and performance analysis. Section V concludes and discusses future research directions.

II. Problem Formulation

A. Problem Statement

The objective is to learn a mapping from short-window PMU measurements to a *scenario-level* equivalent inertia constant. Given an N-bus system with PMUs at buses $i = 1, \ldots, N$, each providing local frequency $f_i(t)$ and RoCoF $\dot{f}_i(t)$, a scalar label H_{eq} is predicted from a 1 s window aligned to the event onset:

$$\widehat{H}_{\text{eq}} = \mathcal{F}\big(\mathbf{F}, \dot{\mathbf{F}}\big), \quad \mathbf{F} = [\tilde{f}_i(t)]_{i,t},\ \dot{\mathbf{F}} = [\tilde{\dot{f}}_i(t)]_{i,t}, \tag{1}$$

where $\tilde{f}_i$ and $\tilde{\dot{f}}_i$ denote filtered and standardised signals. Small active-power injections are applied during data generation to elicit informative transients but are not required as model inputs at deployment. The estimator output $\widehat{H}_{\text{eq}}$ targets the scenario-level label H_{eq} defined in (5).

B. Second-Order Swing Equation

During the initial post-disturbance interval, where inertia dominates, damping and primary control are neglected and the undamped swing approximation is used:

$$\frac{2H_{\text{sys}}}{\omega_s}\,\frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \;=\; \Delta P(t), \tag{2}$$

where $\delta(t)$ is the rotor-angle deviation, $\omega_s = 2\pi f_0$ is the synchronous angular frequency, f_0 is the nominal frequency, and H_{sys} is the per-unit inertia constant (seconds). Let $\omega(t) = \dot{\delta}(t) = 2\pi f(t)$ and define the per-unit speed deviation $\Delta\omega_{\text{pu}}(t) = \big(\omega(t) - \omega_s\big)/\omega_s$. With per-unit power imbalance $\Delta P(t)$ on the system base, the swing equation gives $2H_{\text{sys}} \frac{\mathrm{d}}{\mathrm{d}t}\Delta\omega_{\text{pu}}(t) = \Delta P(t)$. Noting $\frac{\mathrm{d}}{\mathrm{d}t}\Delta\omega_{\text{pu}}(t) = \frac{1}{f_0}\dot{f}(t)$, the following relation is obtained:

$$2\,H_{\text{sys}}\,\dot{f}(t) = \Delta P(t)\,f_0, \tag{3}$$

$$\Longrightarrow \quad H_{\text{sys}} = \frac{\Delta P(t)\,f_0}{2\,\dot{f}(t)}. \tag{4}$$

All powers are in per unit on a common system MVA base; H_{sys} is expressed in seconds.

C. Network-Level Equivalent Inertia

For a multi-machine system on the system base, the aggregate (static) equivalent inertia is

$$H_{\text{eq}} = \frac{\displaystyle\sum_{k=1}^{N_g} H_k\,S_k}{\displaystyle\sum_{k=1}^{N_g} S_k}, \tag{5}$$

where H_k is the inertia constant and S_k the MVA rating of generator k. If motor-load inertia is modelled explicitly, it is included as separate terms in the sum. For a coherent system, $H_{\text{sys}} \approx H_{\text{eq}}$ on the system base.

D. Assumptions and Practical Considerations

The aggregate H_{eq} in (5) is adopted as the *scenario-level* supervised label, representing the static equivalent inertia on the system base for a given operating point. Short-window PMU responses $(\mathbf{F}, \dot{\mathbf{F}})$ serve as features from which the estimator $\mathcal{F}(\cdot)$ predicts this constant.

The following modelling assumptions are used:

- **Small-signal regime:** Frequency deviations $\Delta f_i = f_i - f_0$ and injected power perturbations are sufficiently small for linearised swing dynamics to be valid.
- **Per-unit representation and static label:** All quantities are expressed on a common per-unit system base, and the aggregated inertia H_{eq} from (5) is treated as constant within each scenario and used as the reference label for training.

This formulation bridges the physical interpretation of inertia with the data-driven estimation problem, providing a consistent foundation for the scenario-aware transfer-learning framework introduced in Section III.

III. Methodology

A. Overview

The framework comprises three stages (Fig. 1): (i) source-domain pretraining of an LRCN on IEEE-39; (ii) metric-guided assessment of distribution shift and scenario-aware adaptation; and (iii) few-shot fine-tuning in the target domain under fixed label budgets. maximum mean discrepancy (MMD), Wasserstein distance (WD), and CORrelation ALignment (CORAL) act as soft diagnostics to *rank* adaptation candidates; the final route per scenario is selected by target-side validation with an identical label budget across routes.

979-8-3315-8138-1

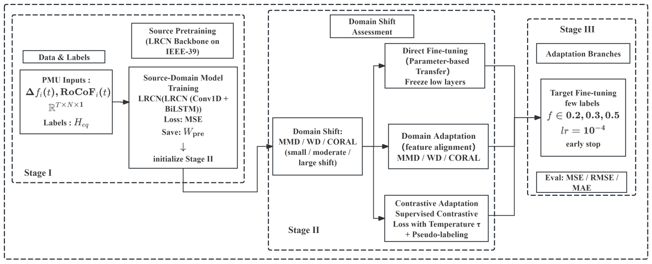

Fig. 1: Three-stage scenario-aware transfer pipeline: pre-training, route ranking (MMD/WD/CORAL), and validation-selected few-shot adaptation.

B. Base Estimator and Objective

Following [14], an LRCN backbone models temporal and cross-channel structure in PMU data. Let $X_{\text{raw}} \in \mathbb{R}^{T\times D_{\text{dom}}}$ denote standardized PMU inputs in a domain, where $D_{\text{dom}} = 2N_{\text{PMU}}$. A lightweight *input–feature unification* adapter g_ϕ maps X_{raw} to the fixed width $D^\star$ expected by the pretrained backbone:

$$\tilde{X} \;=\; g_\phi(X_{\text{raw}}) \in \mathbb{R}^{T\times D^\star}, \qquad D^\star \triangleq 2N_{\text{PMU}}^{(\text{source})}. \tag{6}$$

The source loss is the mean-squared error

$$\mathcal{L}_{\text{src}} = \frac{1}{N_s}\sum_{j=1}^{N_s}\big(\widehat{H}_{\text{eq}}^{(j)} - H_{\text{eq}}^{(j)}\big)^2, \tag{7}$$

optimized by Adam (10^{-3}, batch $= 32$, early stopping). The encoder comprises three 1D-CNN blocks (filters $32/64/128$, kernels $5/5/3$; ReLU; MaxPool$\times$2), two LSTMs (128 units, dropout 0.2), and a linear head:

$$z = \text{LSTM}\big(\text{CNN}(\tilde{X})\big), \qquad \widehat{H}_{\text{eq}} = W z + b. \tag{8}$$

Standardized channels and the adapter g_ϕ ensure a fixed-width encoder across heterogeneous PMU layouts with minimal trainable overhead.

C. Scenario-Aware Domain Adaptation

1) Distribution Shift Evaluation: To quantify source–target similarity, three complementary metrics are used: MMD [15], Wasserstein Distance (WD) [16], and CORAL [17]. All channels are z-scored. MMD uses a Gaussian kernel with the median heuristic (unbiased estimator); WD is Sinkhorn divergence (ε=0.01, 100 iters); CORAL is the Frobenius norm between standardized covariances. In line with adaptation theory [18], [19], these metrics serve as *relative* indicators of shift severity rather than hard thresholds; absolute magnitudes depend on feature scaling and scenario design. Metrics only prioritize which adaptation modes are tried first; the final route is selected by the lowest target-side validation error under fixed label budgets. Pre-adaptation shift magnitudes are summarised in Table I.

TABLE I: Pre-adaptation Distribution-Shift Metrics

Scenario	**Max MMD**	**Avg WD**	**CORAL**
N-1 line (13–14)	0.4654	0.0054	4.27
N-2 lines (4–5, 13–14)	0.4166	0.0059	32.32
Motor dynamics	0.4816	0.0051	1.76
IEEE-118 network	0.8715	0.0526	3.06×10^5
Wind-turbine replacement	1.0006	40.5488	3.76×10^{13}

2) Adaptation Strategies: Depending on the metric-informed scheduling (soft prior), three adaptation strategies are attempted in order of priority; budgets are allocated proportionally across modes, and the final choice is made based on validation performance.

a) Parameter-Based Transfer (Small Shift).: For minor shifts (e.g., N–1 line faults), low-level LRCN layers are frozen while upper layers are fine-tuned using a small learning rate (10^{-4}). This retains general feature representations while adjusting task-specific mappings.

b) Feature-Space Transfer (Moderate Shift).: For moderate shifts or (T, F) mismatch, transfer in the source feature space: center-crop/pad to T_{src} and project channels to F_{src} via a per-time 1×1 adapter,

$$\begin{aligned}\tilde{X} &= g_\phi(\mathcal{T}(X)) = \mathcal{T}(X)W_\phi + \mathbf{1}\,b_\phi^\top, \\ z &= e_\theta(\tilde{X}), \qquad \widehat{H} = h_\psi(z).\end{aligned} \tag{9}$$

Train g_ϕ and h_ψ with MSE; optionally unfreeze a few upper encoder layers by validation.

c) Representation-Based Contrastive Adaptation (Large Shift).: A supervised normalized temperature-scaled cross-entropy (NT-Xent) on a normalized projection is combined with an auxiliary regression on the encoder feature:

$$\mathcal{L}_{\text{pre}} = \lambda_{\text{con}}\mathcal{L}_{\text{con}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}. \tag{10}$$

Positives satisfy $|H_{\text{eq}}^{(i)} - H_{\text{eq}}^{(j)}| \le \eta$; two stochastic views (noise, small time shift, channel dropout) form pairs with cosine similarity at temperature τ. High-confidence pseudo-labels are selected by kNN in the embedding space (lowest qth-percentile mean distance) and appended to the labeled set. The adapted weights initialize Stage III.

D. Unified Training Procedure

The overall training workflow is summarised in Algorithm 1. Stage I pretrains the source-domain model; Stage II evaluates the distribution shift and selects the appropriate adaptation mode; and Stage III performs few-shot fine-tuning on the target domain using labelled fractions $f \in \{0.2, 0.3, 0.5\}$. Hyperparameters (learning rate, dropout, frozen layers, head units) are tuned via Bayesian optimisation, with fine-tuning performed at 10^{-4} learning rate and early stopping.

979-8-3315-8138-1

Algorithm 1 Unified Transfer Learning with Parameter/Feature/Representation Adaptation

Require: Source (X_s, y_s); target scenarios $\mathcal{S}$; label fractions $f \in \{0.2, 0.3, 0.5\}$; temperature τ; pseudo-label confidence $p^\star$
Ensure: Per-scenario models and metrics (MSE, RMSE, MAE)
Stage I: Source Pretraining Train LRCN on (X_s, y_s) with $\mathcal{L}_{\mathrm{src}}$; save W_{pre}.
for all $S \in \mathcal{S}$ **do** ▷ Stage II: Scenario-Aware Adaptation
    Compute MMD/WD/CORAL; rank candidate modes.
    Train each candidate on the same labeled subset of S (identical budget); pick the route with the lowest validation MSE; output $W_{\mathrm{adapt}}(S)$.
end for
for all $S \in \mathcal{S}$ **do** ▷ Stage III: Few-shot Fine-tuning
    for all $f \in \{0.2, 0.3, 0.5\}$ **do**
        Split (X_S, y_S) (train fraction f).
        Initialize with $W_{\mathrm{adapt}}(S)$; attach regression head.
        Fine-tune with $\mathcal{L}_{\mathrm{src}}$ (lr=10^{-4}, early stopping); evaluate on the held-out set.
    end for
end for

E. Baselines and Implementation Details

Baselines are (i) from-scratch per target and (ii) direct transfer when an identity/trivial adapter exists. Unless noted, Adam, early stopping (val-MSE, patience 20), and gradient clipping 1.0 are used.

TABLE II: Implementation knobs (condensed).

Item	Setting
Search ranges	lr 10^{-6}–10^{-2}; dropout 0.1–0.7
Transfer knobs	Frozen layers 0–L; head $\{16, 32, 64, 128, 256\}$
Runtime control	Adam; early stop (pat.=20); clip=1.0

IV. Case Studies

A. Systems and Scenarios

Time-domain simulations were run on IEEE 39 and a modified IEEE 118 (172-bus). Each sample is a 1 s trace at 200 Hz. IEEE 39 serves as the source domain, while five target scenarios are considered: (i) N–1 line outage (13–14, IEEE 39); (ii) N–2 line outages (4–5 & 13–14, IEEE 39); (iii) synchronous-motor addition at bus 4 (IEEE 39); (iv) wind-turbine replacement at bus 34 (IEEE 39); and (v) a 172-bus variant of IEEE 118 with full load-injection tests. The source IEEE 39 dataset comprises 36 171 one-second windows. Each of the IEEE 39 target scenarios (i)–(iv) contains 585 windows, and the modified IEEE 118 scenario in (v) contains 2 040 windows.

B. Data Collection

Nonlinear dynamics are simulated in PowerFactory. PMUs are installed at all buses and record synchronised $f_i(t)$ and RoCoF $\dot{f}_i(t)$. Each sample is a 1 s window (200 samples at 200 Hz) aligned to the disturbance onset. In both the source and target domains, operating points are chosen such that the aggregated inertia H_{eq} spans 3–8 s; in the source IEEE 39 system this range is sampled every 0.5 s with small active-power perturbations at selected buses, whereas in each target scenario it is sampled every 1 s with three perturbation magnitudes in the 0.01%–0.1% total-load range injected at every bus to generate distinct responses. Preprocessing applies per-channel z-scoring, short-window filtering, and additive zero-mean Gaussian noise (45 dB SNR); all quantities are expressed on a common per-unit base.The scenario-level equivalent inertia H_{eq} in (5) is used as the supervised label, fixed within each scenario but varying across operating points. To reflect practical constraints where target-side labels are scarce, we fix three target label budgets $f \in \{20\%, 30\%, 50\%$ that represent low/medium/high availability. For each scenario, the labeled subset is obtained by stratified sampling over (bus, perturbation magnitude) with a fixed random seed; the remaining samples form the test set. These budgets are kept identical across adaptation strategies to enable a fair comparison of label efficiency. Events are generated at all buses by design, so no bus-level cherry-picking is introduced by the labeling process.

C. Adaptation-Strategy Ablation (Stage II)

The route with the lowest validation MSE is fixed for Stage III. To instantiate the scenario-aware policy under a fixed label budget (default f=20%), three adaptation candidates are trained per scenario and compared on the held-out validation set. Table III reports mean squared error (MSE), root-mean-squared error (RMSE), and mean absolute error (MAE); these metrics are used throughout the paper.

TABLE III: Ablation by scenario and adaptation strategy (MSE/RMSE/MAE); "N/A" denotes input dimensionality mismatch.

Scenario	adaptation strategy	MSE	RMSE	MAE
N-1 line	Parameter	**0.0057**	**0.0752**	**0.0520**
	Feature	0.0151	0.1228	0.0893
	Representation	0.2177	0.4666	0.3719
N-2 line	Parameter	**0.0677**	**0.2603**	**0.1719**
	Feature	0.2725	0.5220	0.3534
	Representation	1.1663	1.0800	0.8142
Motor	Parameter	**0.0112**	**0.1060**	**0.0764**
	Feature	0.0303	0.1740	0.1354
	Representation	0.1773	0.4210	0.3275
Wind-Turbine	Parameter	2.0583	1.4347	1.2376
	Feature	2.2794	1.5098	1.2677
	Representation	**1.9952**	**1.4125**	**1.2037**
Modified IEEE-118	Parameter	N/A	N/A	N/A
	Feature	**0.0815**	**0.2856**	**0.1842**
	Representation	0.2412	0.4911	0.3942

D. Evaluation Protocol and Baselines

To ensure comparability across routes, the same labeled subsets per scenario are formed by stratified sampling over (bus, perturbation magnitude) at fixed budgets $f \in \{20\%, 30\%, 50\%\}$. Three approaches were compared: (i) from-scratch LRCN training per target scenario, (ii) the proposed transfer framework with scenario-aware adaptation, and (iii) direct transfer of the source-pretrained model without any target training.

979-8-3315-8138-1

TABLE IV: Comparison of inertia estimation performance. Direct transfer evaluates the source-pretrained model on the target without training

Scenario	Fraction	Direct transfer			From scratch			Transfer learning		
		MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE
N-1 line	20%	0.1330	0.3647	0.2599	0.2978	0.5457	0.4375	**0.0795**	**0.2821**	**0.1961**
	30%	0.1321	0.3635	0.2582	0.2128	0.4613	0.3573	**0.0625**	**0.2500**	**0.1797**
	50%	0.1198	0.3461	0.2471	0.3418	0.5846	0.4097	**0.0386**	**0.1966**	**0.1439**
N-2 line	20%	3.7201	1.9287	1.3226	0.4325	0.6576	0.4744	**0.0677**	**0.2603**	**0.1719**
	30%	3.7773	1.9435	1.3340	0.2529	0.5029	0.3578	**0.0615**	**0.2480**	**0.1479**
	50%	3.8589	1.9644	1.3474	0.1546	0.3932	0.2654	**0.0745**	**0.2731**	**0.1367**
Motor	20%	0.0470	0.2168	0.1747	0.6313	0.7946	0.6062	**0.0112**	**0.1060**	**0.0764**
	30%	**0.0464**	**0.2154**	0.1730	0.8038	0.8965	0.6848	0.0543	0.2997	**0.1532**
	50%	**0.0437**	**0.2091**	0.1699	0.7419	0.8613	0.6577	0.0487	0.2934	**0.1471**
Wind-Turbine	20%	22.0404	4.6947	4.4199	8.9543	2.9923	2.6310	**1.9800**	**1.4071**	**1.2060**
	30%	22.0544	4.6962	4.4244	2.2913	1.5137	1.2857	**2.0047**	**1.4159**	**1.2168**
	50%	21.3419	4.6197	4.3218	2.7559	1.6601	1.4085	**2.0039**	**1.4156**	**1.2111**
Modified IEEE-118	20%	N/A	N/A	N/A	0.1192	0.3453	0.2515	**0.0815**	**0.2856**	**0.1842**
	30%	N/A	N/A	N/A	0.0924	0.3040	0.2435	**0.0394**	**0.1986**	**0.1093**
	50%	N/A	N/A	N/A	0.1405	0.3748	0.2788	**0.0257**	**0.1605**	**0.0821**

Across five scenarios and three label budgets, the transfer approach consistently outperforms both from-scratch training and direct transfer. Across the 15 scenario–budget combinations, transfer learning attains the lowest MAE in all cases and the lowest $\mathrm{MSE}/\mathrm{RMSE}$ in 13 cases; the only exceptions are Motor–30% and Motor–50%, where direct transfer yields slightly lower $\mathrm{MSE}/\mathrm{RMSE}$ but higher MAE. Gains are pronounced in low-label regimes: the median MSE reduction relative to from-scratch is 77.9% at 20% labels, and remains high at 70.6% (30%) and 81.7% (50%). Under the largest distribution shift (wind-turbine replacement), RMSE drops from 2.9923 to 1.4071 (a 53% reduction) at 20% labels, with consistent MAE improvements across all fractions. These results support the scenario-aware adaptation policy: little or no adaptation suffices under minimal shift (Motor), whereas representation-level adaptation is beneficial as the shift increases. The strategy-wise pattern aligns with the ablation: parameter fine-tuning dominates in the small-shift N-1 case; feature-space transfer is preferred under moderate operating changes (N-2, motor); representation-level contrastive adaptation is superior for large, non-Gaussian shifts (wind-turbine and cross-system IEEE-118). These results indicate that (i) adaptation is necessary—direct transfer is frequently suboptimal and sometimes infeasible—and (ii) a validation-selected route yields label-efficient accuracy across shifts while preserving a fixed-width encoder for heterogeneous PMU layouts.

V. Conclusions and Future Work

This paper examined a scenario-aware transfer-learning framework for PMU-based inertia estimation under operating and topology changes. Across five representative scenarios and three label budgets on IEEE-39/118, the transfer models generally achieved lower errors and better label efficiency than from-scratch training under identical budgets. A simple pattern emerges: parameter fine-tuning tends to suffice under small shifts, feature alignment is preferable for moderate shifts, and supervised-contrastive adaptation is most effective when distributions depart substantially from the source. These results suggest that transfer learning is a practical tool for label-efficient inertia estimation and can support deployment with a fixed-width encoder across heterogeneous PMU layouts. This study uses simulated PMU traces under fixed label budgets and does not quantify deployment costs. Future work will (i) validate on field PMU data with realistic timing errors, dropouts, and noise, (ii) report deployment-oriented metrics and analyze sparse or misaligned PMU placements, and (iii) explore conditional/physics-guided augmentation together with automatic route selection and basic uncertainty estimation. Gains are typical but scenario-dependent; operational validation will determine practical impact.

References

[1] P. M. Anderson and M. Mirheydar, “A low-order system frequency response model,” *IEEE Trans. Power Syst.*, vol. 5, no. 3, pp. 720–729, 1990.

[2] E. O. Kontis, I. D. Pasiopoulou, D. A. Kirykos, T. A. Papadopoulos, and G. K. Papagiannis, “Estimation of power system inertia: A comparative assessment of measurement-based techniques,” *Elect. Power Syst. Res.*, vol. 196, p. 107250, 2021.

[3] L. Lugnani, D. Dotta, C. Lackner, and J. Chow, “ARMAX-based method for inertial constant estimation of generation units using synchrophasors,” *Elect. Power Syst. Res.*, vol. 180, p. 106097, 2020.

[4] D. P. Chassin, Z. Huang, M. K. Donnelly, C. Hassler, E. Ramirez, and C. Ray, “Estimation of WECC system inertia using observed frequency transients,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1190–1192, 2005.

[5] K. Tuttelberg, J. Kilter, D. Wilson, and K. Uhlen, “Estimation of power system inertia from ambient wide area measurements,” *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7249–7257, 2018.

[6] P. Wall and V. Terzija, “Simultaneous estimation of the time of disturbance and inertia in power systems,” *IEEE Trans. Power Del.*, vol. 29, no. 4, pp. 2018–2031, 2014.

[7] D. A. Kisinga, P. Makolo, and P. Trodden, “A neural network-based online estimation of stochastic inertia in a power system,” in *Proc. IEEE PES/IAS PowerAfrica*, 2022, pp. 1–5.

[8] E. S. N. R. Paidi, H. Marzooghi, J. Yu, and V. Terzija, “Development and validation of artificial neural network-based tools for forecasting of power system inertia with wind farms penetration,” *IEEE Syst. J.*, vol. 14, no. 4, pp. 4978–4989, 2020.

[9] M. Tuo and X. Li, “Machine learning assisted inertia estimation using ambient measurements,” *IEEE Trans. Ind. Appl.*, vol. 59, no. 4, pp. 4893–4903, 2023.

[10] S. Diggikar, A. Patil, S. S. Katkar, and K. Samad, “Machine learning-based inertia estimation in power systems: A review of methods and challenges,” *Energy Informatics*, vol. 8, no. 1, p. 57, 2025.

[11] Y. Li, S. Zhang, Y. Li, J. Cao, and S. Jia, “PMU measurements-based short-term voltage stability assessment of power systems via deep transfer learning,” *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[12] A. Antoniadis, S. Gaucher, and Y. Goude, “Hierarchical transfer learning with applications to electricity load forecasting,” *Int. J. Forecast.*, vol. 40, no. 2, pp. 641–660, 2024.

[13] A. M. Tzortzis, S. Pelekis, E. Spiliotis, E. Karakolis, S. Mouzakitis, J. Psarras, and D. Askounis, “Transfer learning for day-ahead load forecasting: A case study on European national electricity demand time series,” *Mathematics*, vol. 12, no. 1, p. 19, 2023.

[14] M. Tuo and X. Li, “Long-term recurrent convolutional network-based inertia estimation using ambient measurements,” in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting (IAS)*, 2022, pp. 1–6.

[15] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” in *Proc. NIPS*, 2012.

[16] B. Damodaran, B. Kellenberger, R. Flamary, D. Tuia, and N. Courty, “DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation,” in *Proc. ECCV*, 2018.

[17] B. Sun and K. Saenko, “Deep CORAL: Correlation alignment for deep domain adaptation,” in *Proc. ECCV Workshops*, 2016.

[18] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, “Analysis of representations for domain adaptation,” *Adv. Neural Inf. Process. Syst.*, vol. 19, 2006.

[19] M. Long, H. Zhu, J. Wang, and M. I. Jordan, “Deep transfer learning with joint adaptation networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 2208–2217.

979-8-3315-8138-1