Physics-Driven Graph Attention Network for Microgrids AC-OPF Considering Distributed Generator Availability

Amirhossein Adib
Power Systems Group
Dept. of Electronic Engineering
Royal Holloway University of London
London, United Kingdom
Amir.Adib.2023@live.rhul.ac.uk

Onyema S. Nduka
Power Systems Group
Dept. of Electronic Engineering
Royal Holloway University of London
London, United Kingdom
Onyema.Nduka@rhul.ac.uk

***Abstract*—Data-driven surrogates can solve the AC Optimal Power Flow (AC-OPF) problem far faster than interior-point solvers, but two capabilities are usually missing for distribution-grid use: a self-reported measure of predictive reliability and an evaluation that reflects the changing availability of distributed generators (DGs) in a real microgrid (MG).**

This paper proposes a physics-informed Graph Attention Network (GAT) surrogate that is trained on 24-hour load and renewable datasets, encodes branch admittance into edge-conditioned message passing, and is trained with a mean squared error objective and a single active-power balance penalty. At inference, Monte Carlo (MC) dropout supplies a per-scenario predictive standard deviation at no additional training cost. On an IEEE 33-bus feeder with dispatchable DGs, real solar and wind profiles, and per-scenario stochastic DG availability, the method is compared against the interior-point solver, a parameter-matched multilayer perceptron (MLP), and a physics-free ablation using several statistical tests. Results show that the proposed model lowers active-power dispatch error by 56% relative to the MLP while running about $763\times$ faster than the base solver, and that its predicted uncertainty stays well-calibrated and informative under high renewable penetration and out of distribution load.

***Index Terms*—AC optimal power flow, optimisation, graph attention network, physics-informed learning, Monte Carlo, uncertainty, microgrid, renewable energy.**

I. Introduction

The complexity of modern power systems has grown with the increase of renewable energy sources (RESs), which inject inherently stochastic power, and with the rise of microgrids (MGs) as a structural response to that uncertainty. The Optimal Power Flow (OPF) problem sits at the centre of this operational picture, governing economic dispatch under physical and security constraints [1]. In AC form, the non-linearity of the power-flow (PF) equations renders OPF non-convex and computationally heavy [2], and the interior-point methods that solve it struggle to meet the real-time deadlines that distribution-scale operation increasingly demands.

Machine learning surrogates have emerged as an acceleration strategy [3]: trained on historical or simulated AC-OPF labels, they map operating conditions to dispatch decisions without invoking the solver at inference time. Two strategies are common in the literature. Direct surrogates regress the OPF solution from load and renewable inputs and routinely achieve one-to-two orders of magnitude speedup over conventional optimisation [4], [5], while indirect methods accelerate the solver itself, such as predicting active constraint sets or warm-start points [6], [7]. Within the direct family, end-to-end networks may be trained purely for accuracy or augmented with physical structure by embedding the PF equations either as soft penalties in the loss or as hard output projections, trading a modest accuracy cost for improved feasibility [6].

Soft-penalty physics-informed networks stay end-to-end differentiable but give no guarantee of constraint satisfaction, while hard-projection methods restore feasibility only by re-invoking the PF equations at inference, at a cost that grows super-linearly with system size and erodes the speedup that motivates a surrogate [8], [9]. Such learning-based OPF proxies sit within a broader family of feed-forward, convolutional, and graph-based methods [4], [6], complemented by work on feasibility guarantees [10], representative dataset sampling [11], [12], and transferable formulations [13]. Outside the data-driven realm, renewable-aware OPF has also been studied by directly incorporating RES generation profiles into a classical islanded MG AC-OPF formulation [14]; these models typically assume a fixed network topology and return point predictions with no measure of their own reliability.

This paper is motivated by three specific observations. First, on small fixed-topology feeders, multilayer perceptrons (MLPs) are already a strong baseline, while [15] reports that a physics-informed graph network's advantage appears only under topological uncertainty, i.e., when the grid configuration varies stochastically across scenarios. Second, deterministic surrogates provide no measure of their own reliability, especially at high renewable penetration, where forecast error is large and a confidently wrong dispatch is most costly. Third, in radial distribution MGs, a physically realistic topological uncertainty is not line outages (every line is a bridge), but rather stochastic distributed generator (DG) connection status, representing one of the key uncertainties in sustainable operations [16], which is always feasible and non-disconnecting.

Building on these observations, this paper's contributions are: (i) A physics-informed Graph Attention Network (GAT) surrogate that processes 24-h trajectories via a per-bus bidirectional Gated Recurrent Unit (GRU) and encodes branch series admittance directly into edge-conditioned message passing. (ii)

An inference-time Monte Carlo (MC) dropout (K=50) that provides a per-scenario predictive standard deviation at no additional training cost; the intervals are well-calibrated, track realised error, and remain reliable under out-of-distribution (OOD) load and high penetration, where deterministic baselines report no confidence. (iii) An IEEE 33-bus MG benchmark integrating per-scenario stochastic DG connection status into both dataset and model; it stress-tests the surrogate under high renewable penetration and up to 20% overload OOD load, conditions absent from fixed-topology, single-snapshot studies. (iv) An ablation without the balance penalty demonstrates that the architecture's inductive biases drive accuracy, while the fixed active-power balance penalty ensures feasibility under stress rather than accuracy.

Other portions of the paper are structured as follows: Section II formalises the AC-OPF problem and the graph representation. Section III describes the proposed architecture, the physics-informed loss, and the MC-dropout uncertainty head. Section IV details the test-system construction, dataset generation, baselines, and evaluation protocol. Section V reports the results, statistical tests, and uncertainty calibration. Section VI concludes.

II. Problem Formulation and Graph Representation

A. AC-OPF

The AC-OPF problem minimises the generation cost subject to nodal power balance, generator capability, voltage box, and line thermal constraints:

$$\min_{P_{g,i},Q_{g,i},V_i,\theta_i} \quad \sum_{i\in\mathcal{G}} \left(c_2^{(i)} P_{g,i}^2 + c_1^{(i)} P_{g,i} + c_0^{(i)}\right) \tag{1}$$

s.t.

$$P_{g,i} - P_{d,i} = \sum_j V_i V_j \left(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\right) \quad \forall i \in \mathcal{N} \tag{2}$$

$$Q_{g,i} - Q_{d,i} = \sum_j V_i V_j \left(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\right) \quad \forall i \in \mathcal{N} \tag{3}$$

$$\begin{aligned} P_{g,i}^{\min} \le P_{g,i} \le P_{g,i}^{\max} \\ Q_{g,i}^{\min} \le Q_{g,i} \le Q_{g,i}^{\max} \\ V_i^{\min} \le V_i \le V_i^{\max} \end{aligned} \tag{4}$$

$$|S_{ij}| \le S_{ij}^{\max} \tag{5}$$

$$\theta_{\text{slack}} = 0 \tag{6}$$

Here $\mathcal{G}$ is the set of generator buses, $\mathcal{N}$ the set of all buses, $\theta_{ij} = \theta_i - \theta_j$, and $G_{ij} + jB_{ij}$ are the elements of the complex bus admittance matrix $\mathbf{Y}_{\text{bus}}$. The decision variables are the generator active and reactive setpoints $(P_{g,i}, Q_{g,i})$ and the bus voltage magnitudes and angles (V_i, θ_i); the limits $V_i^{\min/\max}$, $P_{g,i}^{\max}$ and $Q_{g,i}^{\max}$ are fixed parameters. At non-generator buses $P_{g,i} = Q_{g,i} = 0$. The slack bus represents the grid connection (point of common coupling) and acts as the network reference, with its voltage angle fixed at zero. The net active load $P_{d,i}$ absorbs renewable injection modelled as negative load (Section IV-B), so the formulation captures the full renewable-rich MG operating point.

B. Graph Representation

Following the graph abstraction adopted for OPF in [15], the distribution feeder is modelled as a graph $\mathcal{G}_{\text{raph}} = (\mathcal{N}, \mathcal{E})$ with buses as nodes and branches as edges. Each node carries a static feature vector of nominal (base-case) quantities and grid limits, augmented here with a DG-availability flag:

$$\begin{aligned} \mathbf{x}_i = \big[& P_{d,i}^{\text{base}}, Q_{d,i}^{\text{base}}, V_i^{\min}, V_i^{\max}, \\ & \alpha_i^{\text{gen}}, P_{g,i}^{\max}, Q_{g,i}^{\max}, \alpha_i^{\text{DG}} \big], \end{aligned} \tag{7}$$

where $\alpha_i^{\text{gen}} \in \{0,1\}$ is a static flag marking the five buses that host a generator (the substation slack and the four dispatchable DGs), and the eighth channel $\alpha_i^{\text{DG}} \in \{0,1\}$ is the per-scenario availability status of the dispatchable DG at bus i: it is 1 when the DG is online and 0 when it is offline (enforced via GEN_STATUS in the underlying solve); this channel is present only when DG availability varies. The per-scenario operating point is supplied separately through the temporal encoder (Section III-A), which processes the 24-hour (P_d^{net}, Q_d) trajectory. Each edge carries the branch impedance $[R_{ij}, X_{ij}]$, which the model maps to a four-dimensional conditioning vector:

$$\mathbf{e}_{ij} = \big[\log(1{+}R_{ij}),\ \log(1{+}X_{ij}),\ g_{ij},\ b_{ij}\big], \tag{8}$$

where $g_{ij} + jb_{ij} = 1/(R_{ij} + jX_{ij})$ is the series admittance of the branch. These terms appear directly in $\mathbf{Y}_{\text{bus}}$ and thus in the power-balance equations (3). The logarithmic compression of R_{ij}, X_{ij} and the differing numeric ranges of g_{ij} and b_{ij} are handled by standard input normalisation before the attention layers, keeping the edge channels compatible for stable training.

III. Methodology

A. Architecture

The proposed model is built around four edge-conditioned GATv2 layers [17], which extend graph attention by making the attention scores a function of both node and edge features. The network processes the distribution feeder in three stages as illustrated in Fig. 1.

(i) Temporal encoding: Each bus receives a 24-hour (P_d, Q_d) window through a shared bidirectional GRU (hidden size 48). The final forward and backward hidden states form a 96-dimensional per-bus temporal signature that captures daily load behaviour.

(ii) Graph propagation: The temporal signature is concatenated with the static node features (7), per-scenario load summaries, and a two-dimensional system-wide demand context. Four edge-conditioned GATv2 layers operate on a symmetrised bidirectional graph with self-loops. Within each block, a learnable global attention-pooled readout computes a system summary, which is gated by $\sigma(\mathbf{U}\mathbf{h}_i)$ and broadcast to every node, providing an $\mathcal{O}(1)$ path for the system-wide power-balance signal that purely local message passing cannot

supply. Edge conditioning uses the four-dimensional vector of (8), which encodes the series admittance that appears directly in $\mathbf{Y}_{\text{bus}}$.

(iii) Decoding: A shared per-bus linear head produces (V_m, V_a) from the concatenation of the final node embedding and the global readout. Separate per-generator heads produce (P_g, Q_g) from the same concatenation, allowing the slack bus and the dispatchable DGs to be resolved at their natural scales. Outputs are clipped to physical limits, and the slack-bus angle is set to zero post-inference. The architecture achieves the reported accuracy without a post-processing feasibility stage.

B. Training Loss

The model is trained to minimise four prediction errors equally, including voltage magnitude, voltage angle, active power, and reactive power, together with a small penalty that encourages the dispatch to satisfy the AC PF equations:

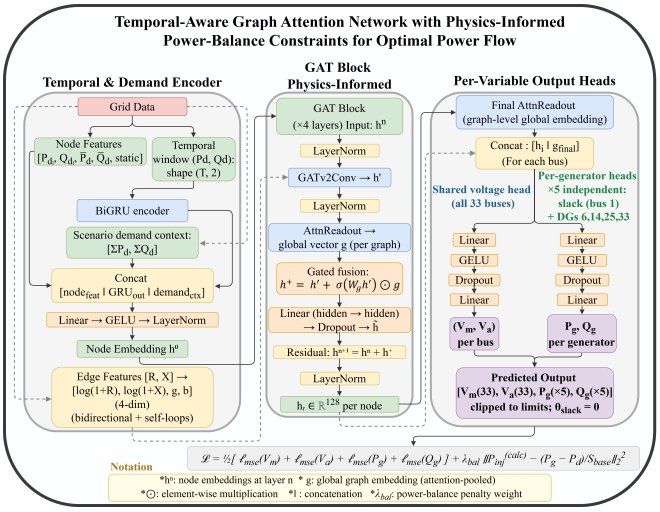

Fig. 1. Proposed physics-informed temporal architecture.

$$\mathcal{L} = \underbrace{\tfrac{1}{2}\big(\ell_{\text{MSE}}(V_m) + \ell_{\text{MSE}}(V_a) + \ell_{\text{MSE}}(P_g) + \ell_{\text{MSE}}(Q_g)\big)}_{\text{data fidelity}} + \lambda_{\text{bal}} \underbrace{\left\| \mathbf{P}_{\text{inj}}^{(\text{calc})} - \frac{\mathbf{P}_g - \mathbf{P}_d}{S_{\text{base}}} \right\|_2^2}_{\text{AC power-balance penalty}}, \quad (9)$$

where the factor $\frac{1}{2}$ scales the summed regression terms relative to the balance penalty. Each ℓ_{MSE} is computed on z-scored targets (per-generator for P_g, Q_g) so that all output quantities contribute equally regardless of their physical units. The right-most term is the active-power balance residual, in which $\mathbf{P}_{\text{inj}}^{(\text{calc})}$ is the nodal active injection obtained by inserting the predicted voltages into the network admittance equations, and the mismatch with the predicted net injection is penalised with a small weight $\lambda_{\text{bal}} = 0.02$. No terms for voltage limits, reactive power, or thermal ratings are included, and the training uses fixed weights throughout. In the operating scenario studied, the edge features and the bidirectional message passing provide enough physical structure to keep the outputs within bounds without additional penalties.

C. Monte-Carlo Dropout for Uncertainty Quantification

Deterministic point predictions are produced with dropout disabled during a single forward pass. To quantify predictive uncertainty, MC dropout ($p = 0.10$) is re-enabled at inference within both the GAT convolutions and the decoder heads. By executing $K = 50$ stochastic forward passes using the trained weights, the per-output standard deviation $\hat{\boldsymbol{\sigma}}$ is computed and used as an approximation to the posterior predictive distribution [18]. This MC procedure is reserved strictly for uncertainty quantification, while all reported point accuracies correspond to the deterministic forward pass, because averaging the stochastic passes was found to degrade the point estimate.

IV. Experimental Setup

A. Test System and Its Construction

The benchmark extends the IEEE 33-bus radial feeder (12.66 kV, 10 MVA base) [19] into the MG as shown in Fig. 2. Four dispatchable DGs are sized at 1.0 MW/$\pm$0.7 MVAr, except at bus 25 (1.2 MW/$\pm$0.8 MVAr); these sizes and locations are chosen to give a non-trivial dispatch, and reactive limits are imposed without an explicit power-factor bound. Six renewable units (three PV at 0.3-0.4 MW, three wind at 0.4 MW) are modelled as uncontrollable negative loads (Section IV-B). To ensure a non-trivial optimisation landscape that rigorously tests the surrogate's structural constraints, three modifications are applied to the base case:

(i) Operational limits: Voltage bounds are tightened to $[0.93, 1.07]$ pu. Branch thermal ratings are reduced to $2.0/4.0$ MVA (radial/substation) to induce binding congestion; otherwise, line loading rarely exceeds 17%, trivialising the OPF for pure Mean Squared Error (MSE).

(ii) Import restrictions: The substation slack is bounded at $[0, 5]$ MW and ± 4 MVAr. Enforcing $P^{\min}$=0 MW prevents grid export, avoiding a degenerate optimum where cheap DGs trivially pin at maximum output across all scenarios.

(iii) Economic trade-off: Quadratic generation costs (c_2, c_1) are set higher for the substation $(0.5, 30.0)$ than the DGs $(0.1, 20.0$ with c_0=0), forcing an active trade-off between local dispatch and import.

B. Dataset Generation

Temporal load, solar, and wind profiles are sourced from the Open Power System Data (OPSD) repository (German subset, 1-h resolution) [20]. Series are linearly interpolated, normalised to $[0, 1]$, and sequentially partitioned into train (70%), validation (15%), and test (15%) blocks. A 24-hour guard gap separates blocks to strictly prevent temporal leakage. A dataset of $10,000$ scenarios is generated as follows:

i) Nodal Injections: Base loads are spatially diversified using Gaussian jitter ($\mu = 1, \sigma = 0.15$). Renewable generation follows the OPSD shapes scaled by bus capacity. The preceding 24-hour trajectory of net active load ($P_d^{\text{net}} = P_d - P_{\text{PV}} - P_{\text{wind}}$) and reactive load ($Q_d$) forms the dynamic input tensor.

ii) Ground-Truth Labels: The AC-OPF problem is solved for each scenario using PYPOWER's interior-point method.

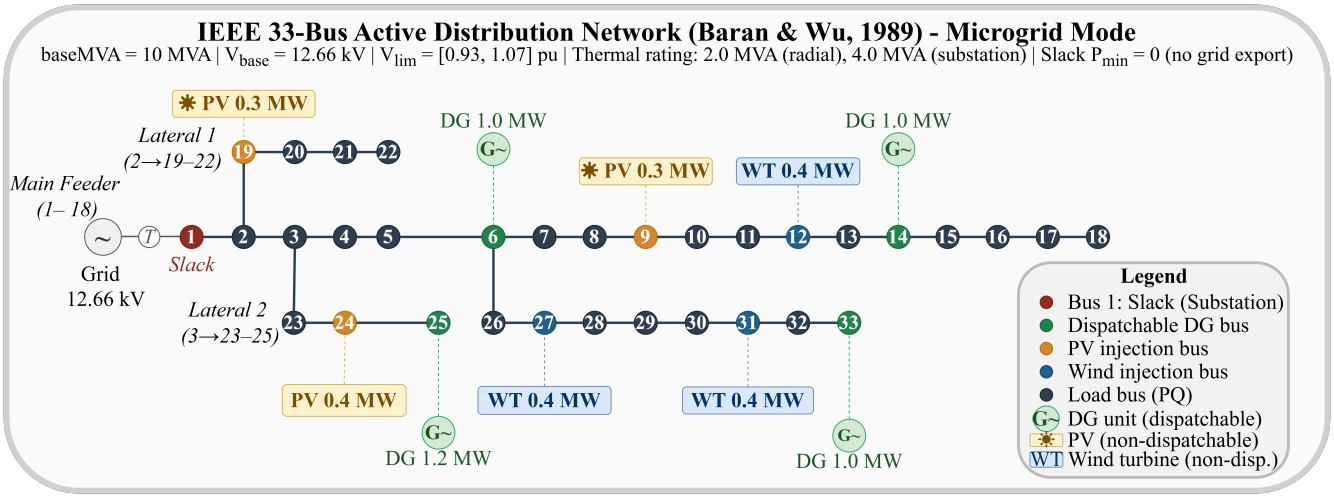

Fig. 2. IEEE 33-bus radial microgrid (Baran & Wu, 1989) with four dispatchable DGs (buses 6, 14, 25, 33), non-dispatchable PV/WT modeled as negative load, and a no-export slack substation at bus 1.

Non-converged scenarios are strictly discarded to ensure a feasible ground-truth distribution.

iii) Feature Scaling: Static bus features are max-scaled to $[0, 1]$ using training-set statistics. To account for the wide capacity disparity between the substation and local DGs, generation targets (P_g, Q_g) are individually standardised using per-generator statistics, ensuring assets are learned at their native scales.

C. DG Availability Variation

A defining novelty of this benchmark is the introduction of per-scenario topological variation via DG dropout. Each dispatchable DG (excluding the substation) is independently taken offline with probability $p_{\text{drop}} = 0.15$, provided at least one DG remains active to service the feeder. This physical unavailability is strictly enforced in the underlying AC-OPF solver (GEN_STATUS=0), zeroing the corresponding generation targets. Crucially, the neural network is informed of this dynamic state via a binary availability flag exposed in node-feature channel 8.

This mechanism allows the MG's operational topology to vary drastically across scenarios without ever structurally severing the radial graph. This challenges the surrogate's adaptability, distinguishing it from classical $N-1$ line contingencies in meshed transmission networks [15], where individual branches can typically be removed without causing isolated islands.

D. Baselines and Training Configuration

All neural models are trained for up to 300 epochs with a batch size of 64, cosine learning-rate schedules, and best-validation checkpointing. The proposed physics-informed GAT (AdamW, max LR 2×10^{-3}, weight decay 2×10^{-4}, gradient clipping) is benchmarked against three references: (1) the interior-point AC-OPF solver of PYPOWER, which provides the ground-truth labels and defines the optimality ceiling; (2) a parameter-matched, fully connected MLP baseline (≈ 0.64 million parameters versus the GAT's ≈ 0.66 million) processing flattened temporal and static features, trained with pure MSE via Adam (LR 10^{-3}); and (3) a physics-free ablation of the exact GAT architecture trained with pure MSE (λ_{bal}=0), which isolates the contribution of the physics penalty from the graph-structural gains.

V. Results

A. Accuracy and Optimality

Table I shows that while MLPs perform well under static, deterministic feeder conditions, their efficacy degrades significantly under grid uncertainty, specifically stochastic DG availability. Under these conditions the proposed model, which runs about $763\times$ faster than the interior-point solver, decisively outperforms the MLP baseline. It achieves an active-power generation root-mean-square error (RMSE) of 5.44 kW and a voltage-magnitude RMSE of 1.39×10^{-3} pu, representing 56% and 47% reductions over the MLP, respectively. Crucially, the physics-free ablation exhibits near-identical point accuracy (RMSE 5.66 kW) under the same stochastic conditions. This indicates that the graph inductive bias drives tracking performance, while the physics loss serves a distinct role by enforcing AC-feasibility under stress (Section V-E).

Economically, the proposed model's mean optimality gap is tightly bounded within $\pm 0.1\%$ of the PYPOWER reference (averaging -0.08%). The marginal negative sign reflects sub-0.1% residual AC infeasibility during raw neural-network inference rather than genuine economic super-optimality, a distinction quantified by the active-power balance residual $|\Delta P|$ in Table I.

B. Statistical Significance

An omnibus Friedman test yields a test statistic of $\chi^2 = 1381.3$ ($p = 3.4 \times 10^{-299}$) with a Nemenyi critical difference of $CD = 0.158$, confirming significant global divergence among the models. Post-hoc pairwise comparisons (Table II) show that the proposed GAT is separated from the MLP baseline, with a large effect size ($d = -0.76$) and a bootstrap confidence interval (CI) excluding zero, indicating a

TABLE I
BENCHMARK RESULTS ON THE IEEE 33-BUS MICROGRID TEST SET.

Method	RMSE P_g (kW)	RMSE V_m (10^{-3} pu)	Gap (%)	$\lvert\Delta P\rvert$ (MW)	Time (ms)	Speedup
PYPOWER	—	—	0.00	< 0.001	756.00	1×
MLP	12.46	2.61	+0.26	0.009	0.44	1737×
GAT (no phys.)*	5.66	1.41	+0.01	0.016	0.98	771×
GAT (prop.)	**5.44**	**1.39**	**−0.08**	**0.015**	**0.99**	**763×**

*The physics-free ablation is the proposed model with the balance penalty disabled (λ_{bal}=0), not a separate architecture.

statistically reliable error reduction. The comparison between the proposed model and its physics-free ablation isolates the empirical effect of the active-power balance penalty on tracking accuracy. As shown in Table II, the resulting ΔRMSE is practically negligible, with the 95% bootstrap CI spanning the origin ($[-0.33, +0.17]$ kW) and a trivial effect size ($d = +0.08$). This statistical parity confirms that enforcing physical compliance introduces no meaningful degradation to the model's baseline tracking accuracy.

TABLE II
PAIRWISE WILCOXON SIGNED-RANK TESTS FOR ACTIVE POWER ERROR

Baseline	p_{Holm}	Cohen's d	ΔRMSE [95% CI] (kW)
MLP	4.7×10^{-138}	−0.76	−5.63 [−7.10, −4.45]
GAT (no phys.)	2.8×10^{-5}	+0.08	−0.08 [−0.33, +0.17]

C. Uncertainty Calibration

Table III evaluates the predictive uncertainty generated via MC dropout ($K = 50$). Here the "+MC" suffix denotes MC dropout applied at inference to the already-trained model (Section III-C); it is not a separately trained network and does not alter the point prediction, but only supplies the predictive standard deviation. Point accuracy is therefore identical to the deterministic model of Table I. While the MLP baseline achieves valid coverage for active (P_g) and reactive (Q_g) power, it produces wide bounds with more than twice the Mean Prediction Interval Width (MPIW) of the proposed method. Conversely, for voltage magnitudes (V_m) the MLP is overconfident, failing to meet the 95% target prediction interval coverage probability (PICP) (PICP$_{V_m} = 0.884$). In contrast, the proposed method safely meets the 95% nominal target across all output quantities (PICP ≥ 0.980) and delivers substantially sharper intervals. This confirms that the graph inductive bias not only improves point accuracy but also yields superior statistical calibration.

Crucially, this predicted uncertainty is informative of localised model reliability. For P_g, the per-scenario predicted standard deviation correlates with the realised absolute error ($r = 0.54$, 95% CI $[0.49, 0.60]$). A Mann-Whitney U test confirms that high-uncertainty predictions incur significantly larger tracking errors (3.99 kW vs. 1.77 kW, $p = 1.6 \times 10^{-37}$). This separation shows that the model's predicted variance $\hat{\sigma}^2$ can be used by operators as a proxy for per-scenario dispatch confidence.

Furthermore, as shown in Table IV, a sensitivity study over $K \in [10, 100]$ MC-dropout passes confirms that coverage (PICP) and interval sharpness stabilise by K=30. Because performance varies negligibly beyond this point, the reported K=50 is a robust computational trade-off within the converged state rather than a hyper-tuned choice.

TABLE III
UNCERTAINTY CALIBRATION METRICS FOR TARGET PICP OF 0.95 (LOWER IS BETTER).

Target	Method	PICP	MPIW	2σ-cov
P_g (kW)	MLP + MC	0.996	63.9	0.996
	GAT + MC	**0.992**	**28.3**	**0.992**
V_m (pu)	MLP + MC	0.884	0.0073	0.888
	GAT + MC	**0.989**	**0.0075**	**0.990**
Q_g (kVAr)	MLP + MC	0.985	48.1	0.985
	GAT + MC	**0.980**	**21.8**	**0.981**

TABLE IV
MC-DROPOUT SENSITIVITY TO FORWARD PASSES FOR ACTIVE POWER.

K	PICP$_{P_g}$	MPIW$_{P_g}$ (kW)	2σ-cov
10	0.985	26.48	0.985
20	0.991	27.68	0.992
30	0.991	27.92	0.991
50	**0.992**	**28.35**	**0.992**
100	0.993	28.50	0.993

D. Renewable Penetration

Table V stratifies test-set performance across renewable-penetration quartiles (R_{pen}). While the dispatch error grows monotonically for both models as renewable injection volatility rises, the proposed method maintains a decisive margin over the MLP. In the highest cohort ($R_{\text{pen}} = 38\%$), it achieves a 59% reduction in mean absolute error (MAE) relative to the MLP, demonstrating the robustness of the graph inductive bias under extreme generation stochasticity.

The proposed model also flags its own tracking degradation. As detailed in Table V, the mean predicted standard deviation ($\bar{\sigma}$) scales monotonically with rising dispatch difficulty, and the correlation between predicted uncertainty and realised error, $r(\sigma, e)$, peaks in the highest penetration quartile, confirming that the variance generated by the model functions as a penetration-adaptive proxy for confidence precisely when the grid is most stressed.

TABLE V
PERFORMANCE METRICS STRATIFIED BY RENEWABLE PENETRATION (R_{pen}).

		MAE (kW)		GAT+MC Uncertainty		
Bin	R_{pen} (%)	MLP	GAT	$\bar{\sigma}$ (kW)	$r(\sigma, e)$	2σ-cov
1	7.2	4.89	1.92	6.06	0.157	0.995
2	15.3	5.07	2.09	6.28	0.119	0.998
3	24.5	6.68	2.62	7.02	0.162	0.994
4	38.0	9.68	3.93	9.50	0.200	0.983

E. OOD and Uncertainty Stress

Table VI shows model performance under OOD stress, governed by systemic load multipliers μ. As loading escalates beyond nominal capacity, tracking errors rise, but the model adaptively widens its predictive uncertainty bounds to compensate. The empirical 2σ prediction-interval coverage remains highly reliable (≥ 0.966) even under a 20% overload. The error-variance correlation $r(\sigma, e)$ is weakest in-distribution (0.199 at $\mu = 1.00$) and rises in both OOD directions, reaching its strongest value (0.300) under the heaviest overload. For context, the PYPOWER solver itself struggles under these conditions, with AC-OPF convergence rates dropping to 97.9% and 92.8% at $\mu = 1.10$ and 1.20, respectively.

The value of the physics-informed loss is sharpest under stress. We flag a sample as infeasible when its largest nodal active-power balance residual exceeds 1 MW. Without the penalty, the ablation crosses this threshold on 0.23% of samples at $\mu = 1.20$, whereas the proposed method ($\lambda_{\text{bal}} = 0.02$) retains strict active-power balance compliance across the entire loading envelope. Both GAT variants exhibit a marginal reactive-power (Q-balance) exceedance on 0.23% of extreme-overload samples, an expected consequence of penalising only active-power residuals while relying on the data-driven decoder for reactive power. Neither method incurs any voltage, generator-capacity, or thermal line violation at any multiplier.

TABLE VI
UNCERTAINTY CALIBRATION UNDER OOD SYSTEMIC LOAD MULTIPLIERS

μ	$\bar{\sigma}$ (kW)	MAE_{P_g} (kW)	2σ-cov [95% CI]	$r(\sigma, e)$
0.80	7.89	2.83	0.990 [0.987, 0.993]	0.255
1.00	7.22	2.64	0.992	0.199
1.10	7.48	3.10	0.990 [0.986, 0.992]	0.225
1.20	8.14	4.64	0.966 [0.960, 0.971]	0.300

VI. CONCLUSION

This paper presented a physics-informed GAT surrogate for AC-OPF, uniquely driven by sequential 24-hour real-world profiles, and evaluated under dynamic DG availability in a renewable-rich MG. The findings demonstrate that the graph inductive bias drives significant accuracy gains over flat neural baselines, while the active-power balance penalty ensures strict physical feasibility under OOD stress by preventing the threshold violations seen in physics-free models. Beyond providing orders of magnitude acceleration over conventional interior-point solvers, a key strength of the proposed framework is its built-in uncertainty quantification. Through inference-time MC dropout, the model yields reliable, penetration-adaptive confidence intervals that strongly correlate with actual dispatch errors. These bounds remain well-calibrated even under severe load overloads, allowing operators to treat the surrogate's variance as a trustworthy confidence signal rather than relying on opaque point predictions. Future work will extend this framework to coupled AC/DC MG uncertainties and develop selective prediction policies that automatically defer low-confidence inferences to classical solvers.

DATA AVAILABILITY

The IEEE 33-bus microgrid model will be available upon request.

ACKNOWLEDGMENT

This work is supported by Royal Holloway University of London.

REFERENCES

[1] A. Adib and O. S. Nduka, "Extreme weather impacts on microgrid components: A critical review establishing data-driven methods as the definitive path forward," in *2025 13th International Conference on Smart Grid (icSmartGrid)*. IEEE, 2025, pp. 737–742.
[2] M. S. Hasan, M. M.-U.-T. Chowdhury, and S. Kamalasadan, "Optimal power flow in electric transmission system: Comparative study and analysis of modeling methods," *IEEE Transactions on Industry Applications*, 2025.
[3] A. Adib and O. S. Nduka, "Hyperparameter optimization techniques for enhanced machine learning energy forecasting: A comparative analysis," in *2025 13th International Conference on Smart Grid (icSmartGrid)*. IEEE, 2025, pp. 220–224.
[4] Y. Jia, Y. Zhou, and Y. Wang, "Scalable graph-based opf learning," *Available at SSRN 5735326*.
[5] Z. Hu and H. Zhang, "A transparent neural network solver for optimal power flow problem," *IEEE Transactions on Power Systems*, 2025.
[6] A. Babiker, S. S. Ahmad, I. Ahmed, M. Khalid, M. A. Abido, and F. S. Al-Ismail, "Optimal power flow: A review of state-of-the-art techniques and future perspectives," *IEEE Access*, 2025.
[7] A. Rosemberg, M. Klamkin, and P. Van Hentenryck, "Differentiable optimization for deep learning-enhanced dc approximation of ac optimal power flow," *arXiv preprint arXiv:2504.01970*, 2025.
[8] W. Huang, K. Gong, Z. Jiang, X. Zhang, and L. Zheng, "A decentralized unsupervised learning approach for ac-opf based on multi-branch deep neural network," *IEEE Access*, vol. 13, pp. 206 932–206 947, 2025.
[9] M. Gao, J. Yu, Z. Yang, and J. Zhao, "A physics-guided graph convolution neural network for optimal power flow," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 380–390, 2023.
[10] D. Ajeyemi, Y. Chen, A. Colot, J. Cortés, and E. Dall'Anese, "Learning to pursue ac optimal power flow solutions with feasibility guarantees," *IEEE Open Journal of Control Systems*, 2026.
[11] S. Bouchkati, L. Kampmann, L. Böttcher, and A. Ulbig, "Feasible operating region-based sampling for representative opf datasets," in *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*. IEEE, 2025, pp. 1–5.
[12] Y. Yu, O. S. Nduka, F. U. Nazir, and B. C. Pal, "A three-stage stochastic framework for smart electric vehicle charging," *IEEE Access*, vol. 11, pp. 655–666, 2022.
[13] Y. Zhu, Y. Zhou, W. Wei, P. Li, and W. Huang, "Gnns' generalization improvement for large-scale power system analysis based on physics-informed self-supervised pre-training," *IEEE Transactions on Power Systems*, 2025.
[14] S. F. Zandrazavi, M. Pashaei, N. B. Arias, J. C. Lopez, T. B. Soeiro, and M. Shafie-Khah, "Ac-opf for droop-based islanded microgrids with voltage-and frequency-dependent loads," *IEEE Transactions on Industry Applications*, 2026.
[15] Y. Zhang, Q. Hou, and Y. Ding, "Physics-informed and constraint-satisfied graph attention network for solving optimal power flow," in *2026 IEEE PES International Meeting (PES IM)*, 2026, pp. 1–5.
[16] O. Nduka and A. Adib, "Sustainability-oriented framework for design, planning and operations of isolated microgrids in the era of uncertainties," in *2026 14th International Conference on Smart Grid (icSmartGrid)*, 2026, accepted for Presentation.
[17] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, 2021.
[18] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.
[19] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.
[20] Open Power System Data, "Data package time series, version 2020-10-06," https://doi.org/10.25832/time_series/2020-10-06, 2020.