

Leakage-Resistant and Uncertainty-Aware Machine Learning for Carrier-Grade Router Power Prediction Using Real Traffic Telemetry

Clinton Amponsah

`clinton.amponsah001@gmail.com`

University of Energy and Natural Resources

Kyiewu Bernard

University of Energy and Natural Resources

Linda Bessa-Simmons

University of Energy and Natural Resources

Article

Keywords: Green networking, router power prediction, machine learning, Network Digital Twin, zero-touch management

Posted Date: October 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10578529/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

LEAKAGE-RESISTANT AND UNCERTAINTY-AWARE MACHINE LEARNING FOR CARRIER-GRADE ROUTER POWER PREDICTION USING REAL TRAFFIC TELEMETRY

Abstract—Energy-aware network management requires reliable models that translate operational traffic conditions into equipment-level power estimates. This study developed leakage-resistant machine-learning models for predicting instantaneous power consumption in two carrier-grade router families using 23,705 real telemetry observations collected at approximately five-second intervals. The dataset covered traffic-load and packet-length variation and multiple concurrent-flow conditions. Throughput utilisation, packet size, logarithmic packet rate and a load–packet-rate interaction were used as predictors. Upward traffic ramps supported grouped cross-validation and model selection, whereas downward ramps formed independent temporal test sets. Polynomial ridge regression performed best for Router RA, achieving an MAE of 5.043 W, RMSE of 7.151 W and R^2 of 0.645. Linear regression performed best for Router RB, with an MAE of 1.383 W, RMSE of 1.876 W and R^2 of 0.823. Moving-block bootstrap confidence intervals and split-conformal prediction intervals quantified sampling and predictive uncertainty. Feature-set ablation confirmed traffic utilisation as the dominant predictor, while packet-rate transformations provided modest gains. The findings support lightweight, router-specific power models for Network Digital Twin and zero-touch green-networking applications, while distinguishing prediction accuracy from demonstrated network-level energy savings.

Keywords: Green networking, router power prediction, machine learning, Network Digital Twin, zero-touch management

1. Introduction

The energy efficiency of communication networks has become a systems-level design requirement rather than an equipment-level consideration. Contemporary routers, switches, virtualized functions and control platforms must sustain increasing traffic volumes while operators face pressure to reduce electricity consumption, operating expenditure and associated emissions. ITU-T L.1310 consequently emphasizes repeatable equipment-level measurement and comparable energy-efficiency metrics because optimization decisions are credible only when power consumption is quantified under defined traffic conditions [1]. Recent reviews similarly demonstrate that energy-aware routing must balance power reduction against congestion, reliability and delay rather than minimizing energy in isolation [2]. In software-defined data-center networks, traffic-priority and load-aware path selection have established the importance of jointly considering route energy and resource utilization [3], while energy-efficient virtual-network-function placement confirms that computational location can materially affect network lifetime and service feasibility [4].

Artificial intelligence provides a mechanism for learning nonlinear relationships between network operation and equipment power consumption. However, its practical value depends on whether

predictions remain reliable beyond the observations used during training. Energy-efficient network-slicing research increasingly treats AI as a cross-layer optimization instrument while identifying model cost, hardware heterogeneity and lifecycle coordination as unresolved concerns [5]. Real-world mobile-traffic studies have also shown that machine-learning forecasts can guide resource activation and reduce avoidable energy consumption [6]. Nevertheless, numerous green-networking studies depend on simulated power functions, randomly partition temporally correlated measurements or report point accuracy without uncertainty. These practices can overestimate generalization and are unsuitable for closed-loop control, where inaccurate predictions may trigger unnecessary rerouting, activate additional equipment or violate service requirements.

This problem is directly aligned with zero-touch network and service management. The ETSI ZSM landscape positions automation, analytics and closed-loop operation as foundations for autonomous network management [7], while intent-driven management translates high-level operational objectives into monitored and enforceable actions [8]. A Network Digital Twin provides an intermediate environment in which candidate decisions can be evaluated before physical enforcement. ITU-T Y.3090 defines a digital twin network as a virtual representation connected to the physical network through data, models and interfaces [9]. Digital-twin research for 6G identifies telemetry fidelity and behavioral modeling as prerequisites for dependable optimization [10]. Practical 5G methodologies further demonstrate how learned models may be embedded in digital twins to evaluate alternative configurations [11]. However, a systematic review identified persistent limitations in model validation, reproducibility and operational integration [12]. ETSI therefore recommends evaluating machine-learning decisions within Network Digital Twin sandbox environments before recommendations affect the physical network [13]. For green traffic engineering, a fundamental requirement is an accurate equipment-specific power function. Let u_t represent throughput utilization, s_t packet size and r_t packet rate. Instantaneous router power was estimated as

$$\hat{P}_t = f_\theta(u_t, s_t, \log(1 + r_t), u_t \log(1 + r_t)) \quad (1)$$

where f_θ denotes a model selected through leakage-resistant grouped validation. This formulation recognizes that equal bit rates may impose different packet-processing workloads. Recent machine-learning surveys emphasize that telemetry-driven network management requires deployment-aware validation, appropriate feature selection and transparent performance assessment [14]. The ITU-T data-domain framework similarly identifies timely, accurate and governed physical-network data as essential for maintaining a useful digital twin [15]. This study uses the 2025 open dataset of measured traffic and power telemetry from physical carrier-grade routers rather than a synthetic power model [16]. The five-second measurements cover two router families under controlled traffic-load, packet-size and multiple-flow conditions, providing a suitable basis for hardware-specific prediction and uncertainty analysis.

The study makes the following contributions:

- i. It develops separate lightweight power-prediction models for two carrier-grade router families using measured utilization, packet-size and packet-rate telemetry.
- ii. It introduces a leakage-resistant design that separates complete traffic-ramp cycles and groups repeated operating states during cross-validation.
- iii. It quantifies sampling and predictive uncertainty using moving-block bootstrap confidence intervals and split-conformal prediction intervals.
- iv. It evaluates the incremental contribution of packet-rate transformations through feature-set ablation, supporting parsimonious deployment in Network Digital Twin and zero-touch management environments.

2. Related Studies

2.1 Telemetry-Driven Machine Learning for Network Energy Estimation

Recent research has increasingly treated network telemetry as a basis for predicting equipment behaviour and supporting automated energy management. Recursive machine-learning models have been developed to estimate router metrics within a network digital replica, demonstrating that node-level behaviour can be learned from operational measurements [17]. Similarly, traffic records collected from operational 4G+ and 5G infrastructure have been used to evaluate machine-learning forecasting for energy-efficient resource activation [6]. These findings established the value of real network traces, although the primary prediction target was traffic demand rather than equipment power. Predicted throughput has also been used to guide small-cell ON/OFF decisions, illustrating how forecasts may support energy control while preserving user performance [18]. Adaptive energy management has further been extended to user equipment through an AI-driven discontinuous-reception mechanism trained with real wireless traces [19]. At the infrastructure level, traffic and power telemetry from physical carrier-grade routers has provided an important basis for equipment-specific modelling under controlled load, packet-size and flow conditions [16]. More broadly, machine-learning resource allocation requires representative data, deployment-aware validation and appropriate model complexity [14]. Collectively, these studies support telemetry-driven energy management; however, limited attention has been given to temporally separated evaluation, hardware-specific model selection and calibrated uncertainty in router power prediction.

2.2 Energy-Aware Routing and Resource-Control Mechanisms

Energy-aware networking research has progressed from static routing costs toward adaptive decisions based on traffic, device state and service requirements. Energy-aware routing must balance residual resources, delivery reliability, delay and network lifetime rather than optimise power consumption independently [2]. Cloud-based IoT communication energy has been reduced by minimising routing hops and improving subchannel selection, although the formulation focused on constrained IoT environments [20]. Within software-defined data-centre networks, traffic

priority, load conditions and energy costs have been combined to prevent congestion while limiting the number of active resources [3]. Virtual-network-function placement also influences energy use because processing location affects forwarding distance, resource activation and service feasibility [4]. Network-slice energy efficiency is likewise inherently cross-layer, involving radio, transport, computing and orchestration decisions [5]. More recently, deep learning has been incorporated into SDN routing so that controllers can use network-state and energy information when selecting paths [21]. Although these studies demonstrate the potential of intelligent routing, many evaluate energy savings using assumed or simulated equipment power functions. Consequently, routing decisions may appear efficient without establishing whether the underlying device-level power estimates generalise across hardware and unseen traffic cycles. This limitation motivates measured, router-specific power modelling before prediction outputs are embedded in green traffic-engineering controllers.

2.3 Network Digital Twins and Trustworthy Predictive Automation

Network Digital Twins provide an architectural context in which predictive models can support monitoring, simulation and controlled optimisation. A modelling and deployment methodology for 5G digital twins has emphasised synchronisation between physical measurements and virtual network representations [11]. Existing Network Digital Twin studies have nevertheless varied considerably in validation practice, fidelity and operational maturity [12]. This gap has been addressed by integrating network simulation with digital-twin functionality, thereby enabling reusable what-if analysis for IoT network configuration [22]. Recent surveys further identify telemetry acquisition, model updating, edge–cloud coordination and physical-network control as central components of dependable twins [23], [24]. The discussion has also extended to generative-AI-enabled Network Digital Twins, while highlighting risks associated with synthetic fidelity, transfer to physical systems and control reliability [25]. These concerns imply that point predictions alone are insufficient for autonomous energy management. Conformal prediction provides a model-agnostic approach for producing statistically calibrated predictive sets [26], while conformal methods have also been adapted to temporally dependent multivariate forecasting [27]. Nevertheless, uncertainty-aware validation remains uncommon in network-energy modelling. The present study therefore connects measured router telemetry, leakage-resistant temporal evaluation and conformal intervals to provide a more defensible predictive component for Network Digital Twin and zero-touch green-networking applications.

3. Methodology

3.1 Research Design

The study employed a quantitative experimental machine-learning design to estimate the instantaneous power consumption of carrier-grade network routers from operational traffic telemetry. Four measurement files representing router families RA and RB were analysed under two controlled operating conditions: simultaneous variation in traffic load and packet length, and multiple concurrent traffic flows. Measurements were recorded at approximately five-second

intervals as traffic utilisation increased progressively from 0% to 100% and subsequently decreased to 0%. The initial dataset comprised 23,707 observations. Two distinct records from the RA multiple-flow experiment were excluded because one lacked the measured-power response and another lacked the packet-rate predictor, producing a final analytical sample of 23,705 valid observations. Instantaneous router power consumption, measured in watts, constituted the response variable, while throughput utilisation, packet size in bytes and packet rate served as the principal predictor variables. This design supported controlled comparison of alternative regression models and independent evaluation across temporally separated traffic ramps.

3.2 Data Preparation and Feature Engineering

All files were inspected for missing values, duplicate observations and physically implausible measurements. Records were considered invalid when throughput utilisation was outside the interval 0–100%, packet size was non-positive, packet rate was negative or measured power was non-positive. No retained observation violated these conditions. Throughput utilisation was represented as a fraction $u_t \in [0, 1]$, packet size was denoted by s_t , and packet rate by r_t . Packet rate was logarithmically transformed to reduce its strong positive skew:

$$r_t^* = \log(1 + r_t). \quad (2)$$

An interaction term was subsequently constructed to represent the combined influence of traffic load and packet-processing intensity:

$$z_t = u_t \log(1 + r_t). \quad (3)$$

The complete predictor vector was therefore defined as

$$x_t = [u_t, s_t, \log(1 + r_t), u_t \log(1 + r_t)]^\top \quad (4)$$

Timestamp, elapsed time, power lags and rolling target statistics were excluded. Their exclusion prevented the models from memorising the programmed traffic sequence or indirectly accessing previous target values. Separate models were developed for RA and RB because the two devices operated at substantially different power scales and exhibited different load–power relationships.

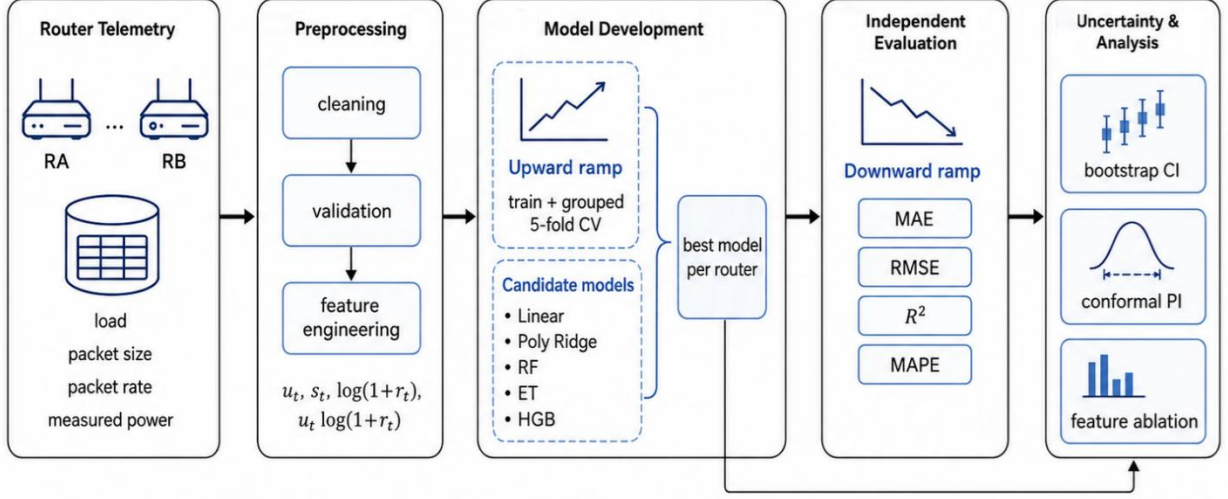


Fig. 1. Leakage-Resistant and Uncertainty-Aware Methodological Framework for Carrier-Grade Router Power Prediction Using Real Traffic Telemetry

The overall methodological workflow is presented in Fig. 1. The framework summarizes the progression from real router telemetry acquisition and preprocessing to leakage-resistant model development, independent temporal evaluation, uncertainty quantification and feature-set ablation.

3.3 Leakage-Resistant Experimental Partitioning

A conventional random row-level split was not used because adjacent five-second measurements represented highly correlated observations from the same operating state. Random allocation could therefore place nearly identical records in both the training and testing subsets, producing overly optimistic estimates of predictive performance. Each experimental trace comprised an upward traffic ramp followed by a downward ramp. The midpoint of the maximum-load plateau was used to separate the two temporal cycles. The upward ramp provided 5,921 RA observations and 5,918 RB observations for model development, whereas the subsequent downward ramp contained 5,930 RA observations and 5,936 RB observations for held-out temporal evaluation. Model selection within the upward-ramp data employed five-fold grouped cross-validation. Groups were defined by experiment type, throughput-utilisation level and packet size, ensuring that repeated observations from an identical operating condition remained within the same fold. The downward-ramp observations were excluded from model fitting, hyperparameter tuning and prospective feature-set selection. The final models and predictor configurations were fixed before evaluation on the downward-ramp data. Because both ramps originated from the same experimental traces and contained overlapping operating conditions, the reported results represent generalisation to a held-out temporal ramp within the same experiments rather than external deployment generalisation.

3.4 Candidate Models and Model Selection

Six regression approaches were compared independently for each router. A mean-value dummy regressor was included as a non-informative baseline. The predictive models comprised linear regression, second-order polynomial ridge regression, Random Forest, Extra Trees and Histogram Gradient Boosting. The polynomial ridge model expanded the predictor vector to include squared terms and pairwise interactions. Its coefficients were estimated by minimizing

$$\hat{\beta} = \arg \min_{\beta} \left[\sum_{i=1}^n (P_i - \phi(x_i)^T \beta)^2 + \lambda \|\beta\|_2^2 \right], \quad (5)$$

where P_i is measured power, $\phi(x_i)$ is the second-order feature expansion and λ is the ridge penalty. The penalty was set to 10^{-6} . Random Forest and Extra Trees were trained with 300 trees, a minimum of three observations per terminal leaf and all available features considered at each split. Histogram Gradient Boosting used 300 boosting iterations, a learning rate of 0.05, 15 terminal leaves, a minimum leaf size of 30 and an L_2 regularisation coefficient of 0.1. A fixed random seed of 42 was applied for reproducibility. The model with the lowest grouped cross-validation root mean squared error was selected for each router. The selected estimator was then fitted to the complete upward-ramp dataset and evaluated once on the untouched downward-ramp observations.

3.5 Performance Evaluation

Predictive performance was assessed using mean absolute error, root mean squared error, coefficient of determination and mean absolute percentage error:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i - \hat{P}_i|, \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2}, \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - \hat{P}_i)^2}{\sum_{i=1}^n (P_i - \bar{P})^2}, \quad (8)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{P_i - \hat{P}_i}{P_i} \right|. \quad (9)$$

MAE and RMSE were expressed in watts. RMSE assigned greater weight to large prediction errors, while R^2 quantified the proportion of independent-test power variation explained by the model. Mean prediction bias and the 95th percentile of absolute error were also calculated to

identify systematic deviation and tail-error behaviour. Performance was reported both for the complete test partition and separately for the two experimental traffic conditions.

3.6 Confidence Intervals and Predictive Uncertainty

Sampling uncertainty was assessed through a moving-block bootstrap with 500 repetitions. Within each experimental file, consecutive blocks of 60 observations were resampled with replacement. Given the approximately five-second sampling interval, each block preserved about five minutes of local temporal dependence. The 2.5th and 97.5th percentiles of the bootstrap distributions formed the 95% confidence intervals for MAE, RMSE, R^2 , MAPE and mean bias. Prediction uncertainty was quantified using grouped split-conformal inference. Absolute out-of-fold residuals from the selected model were calculated on the upward-ramp data:

$$e_i = |P_i - \hat{P}_i^{\text{OOF}}|. \quad (10)$$

For a nominal coverage level of $1 - \alpha = 0.90$, the finite-sample corrected 90th percentile of these residuals was denoted by $q_{0.90}$. The prediction interval for a new observation was then defined as

$$C_{0.90}(x) = [\hat{P}(x) - q_{0.90}, \hat{P}(x) + q_{0.90}] \quad (11)$$

Interval quality was evaluated using empirical coverage and mean interval width on the independent downward-ramp test observations.

3.7 Feature-Set Ablation and Scope of Analysis

A post hoc feature-set ablation was conducted after completion of the primary independent temporal evaluation. The analysis compared five predictor configurations: traffic load only, packet size only, the original load-and-packet-size combination, the enhanced feature set without the interaction term and the complete enhanced feature set. The ablation used the downward-ramp observations solely to describe the relative contribution of the available telemetry variables. It was not used to revise the selected models, alter their hyperparameters or determine the principal performance results. Accordingly, the ablation findings were interpreted as diagnostic rather than confirmatory evidence. The headline independent-test results remained those obtained from the models and feature configurations selected exclusively through grouped cross-validation on the upward-ramp data. A stronger external design would conduct feature ablation within a separate validation partition and preserve an unseen iteration, day or router trace for a single final assessment. The analysis therefore evaluated router-level power prediction under held-out temporal conditions but did not establish external generalisation across independently collected datasets or operational deployments.

4. Results

4.1 Dataset Integrity and Measured Power Characteristics

The analysis retained 23,705 valid observations from four experimental files covering two carrier-grade router models, RA and RB. Router RA contributed 11,851 observations, while RB contributed 11,854 observations. The experiments represented two operating conditions: simultaneous variation in traffic load and packet length, and multiple concurrent traffic flows. Telemetry was recorded at approximately five-second intervals over approximately nine hours per experiment. Two distinct observations were removed from the RA multiple-flow file: one because the measured-power target was missing and another because the packet-rate predictor was unavailable. No retained record violated the predefined physical-range checks. As reported in Table 1, RA operated at a substantially higher absolute power level than RB. Mean RA consumption ranged from 720.63 to 721.80 W, whereas mean RB consumption ranged from 76.03 to 76.68 W. This hardware-dependent difference justified developing separate prediction models for the two routers rather than pooling their measurements within a single estimator.

Table 1. Dataset Characteristics and Measured Router Power

Router	Experimental condition	Observations	Duration (h)	Packet size (B)	Mean power (W)	SD (W)	Power range (W)	Idle power (W)	Full-load power (W)	Dynamic increase (W)
RA	Multiple traffic flows	6,583	9.180	669.1	721.80	10.39	692.00 – 747.00	705.16	732.95	27.79
RA	Traffic-load and packet-length variation	5,268	9.082	160–1,500	720.63	13.86	690.00 – 766.00	702.86	731.05	28.19
RB	Multiple traffic flows	6,585	9.236	669.1	76.03	4.26	63.26–89.41	69.92	82.26	12.34
RB	Traffic-load and packet-length variation	5,269	9.157	160–1,500	76.68	4.69	64.20–92.20	69.92	83.32	13.40

The load–power relationship for RA is illustrated in Fig. 2. Mean power increased from approximately 703–705 W at zero traffic to 731–733 W at full load. This represented an increase of only 3.94%–4.01% above idle consumption, indicating that RA’s total consumption was dominated by a large fixed-power component. Nevertheless, the approximately 28 W dynamic range was sufficiently systematic for traffic-driven power modeling. The two experimental curves remained close at intermediate loads but diverged slightly above 70% utilization, where the multiple-flow experiment produced marginally higher power.

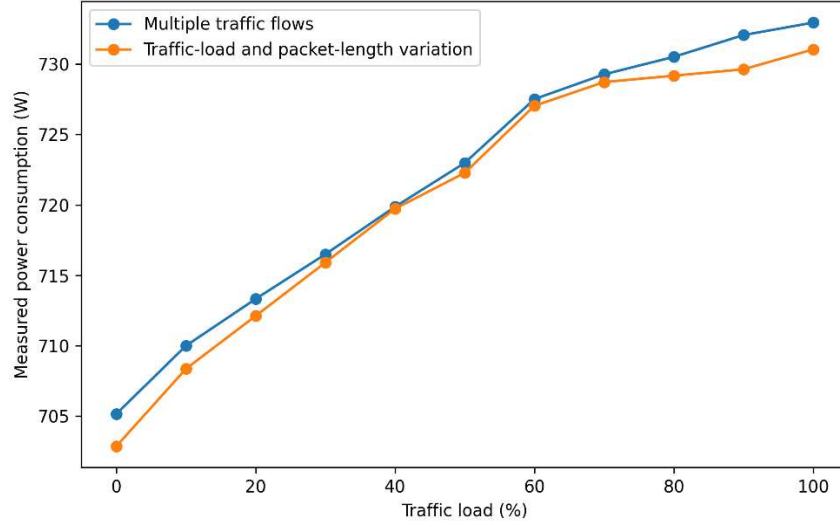


Fig. 2. RA Measured Power Response to Traffic Load

RB exhibited a more pronounced relative response to traffic load. As shown in Fig. 3, power increased from approximately 69.92 W at idle to 82.26 W under multiple-flow traffic and 83.32 W under varying packet lengths. These increases corresponded to 17.65% and 19.16% above idle consumption, respectively. The load–power curves were nearly linear, although the packet-length experiment increasingly exceeded the multiple-flow curve at high utilization. This pattern indicates that RB’s forwarding power was influenced not only by transmitted load but also by the packet-processing intensity associated with different packet sizes.

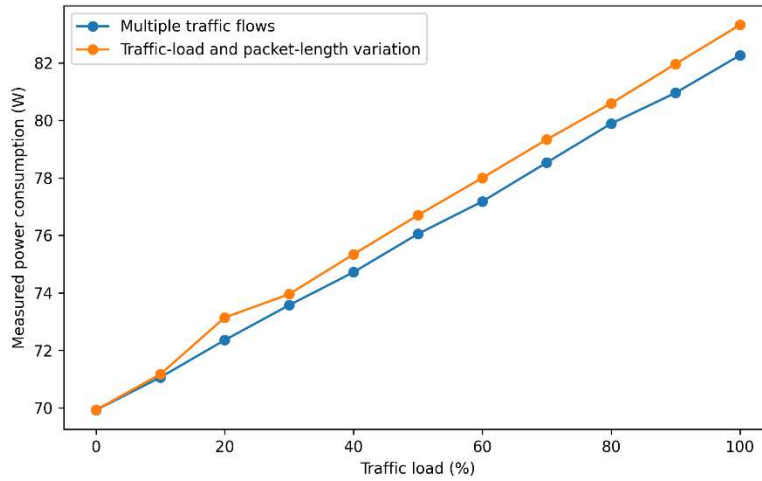


Fig. 3. RB Measured Power Response to Traffic Load

4.2 Leakage-Resistant Model Selection

Model development was performed separately for RA and RB. Only observations from the upward traffic ramp were used for training and model selection, while the subsequent downward ramp was reserved for independent temporal evaluation. Five-fold grouped cross-validation withheld complete experiment–load–packet-size operating points. Consequently, repeated five-second measurements from the same operating condition could not occur simultaneously in the training

and validation portions of a fold. The grouped cross-validation results are presented in Table 2. For RA, second-order polynomial ridge regression achieved the lowest RMSE of 6.552 W and the highest (R^2) of 0.708. Extra Trees followed closely with an RMSE of 6.605 W. The remaining machine-learning models produced RMSE values between 6.944 and 6.973 W. All trained models substantially outperformed the mean baseline, which produced an RMSE of 12.522 W and a negative (R^2).

Table 2. Grouped Cross-Validation Performance of Candidate Models

Model	RA RMSE (W)	RA (R^2)	RB RMSE (W)	RB (R^2)
Polynomial ridge, degree 2	6.552	0.708	1.908	0.819
Extra Trees	6.605	0.704	1.919	0.816
Histogram Gradient Boosting	6.944	0.673	1.898	0.821
Linear Regression	6.957	0.671	1.881	0.824
Random Forest	6.973	0.670	1.941	0.812
Mean baseline	12.522	−0.065	4.586	−0.048

For RB, linear regression achieved the strongest grouped validation performance, with an RMSE of 1.881 W and an (R^2) of 0.824. Although Histogram Gradient Boosting produced a marginally lower MAE, its RMSE was higher at 1.898 W. The linear model was therefore retained because it provided the lowest squared-error loss and the strongest explained variance. The comparison in Fig. 4 shows that model performance differences were relatively small among the trained estimators but substantial relative to the mean baseline. The findings support a router-specific modeling strategy: RA required nonlinear interaction terms, whereas RB was represented adequately by a lower-complexity linear relationship.

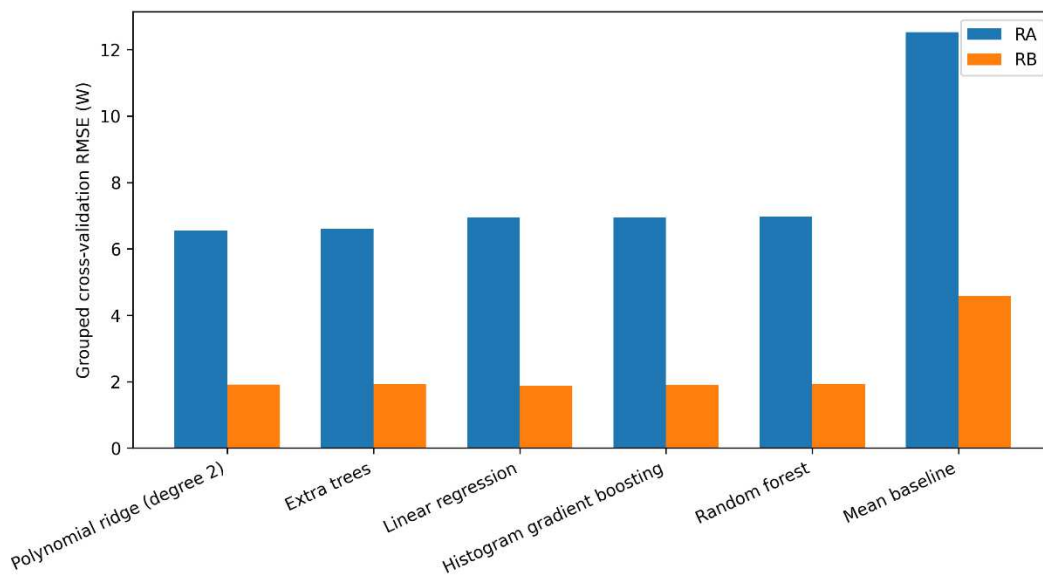


Fig. 4. Grouped Cross-Validation Comparison of Candidate Models

4.3 Independent Temporal-Cycle Performance

The selected models were subsequently evaluated on the untouched downward-ramp observations. As summarized in Table 3, the RA polynomial ridge model achieved an overall MAE of 5.043 W, RMSE of 7.151 W, MAPE of 0.698%, and (R^2) of 0.645. Its 95th-percentile absolute error was 13.954 W. Performance differed between experimental regimes. Under multiple-flow traffic, MAE was 3.954 W and (R^2) was 0.772. Under joint traffic-load and packet-length variation, MAE increased to 6.408 W and (R^2) declined to 0.554. The larger error under varying packet sizes demonstrates that instantaneous utilization and packet statistics did not fully explain RA's short-term power variability.

Table 3. Independent Downward-Ramp Prediction Performance

Router	Evaluation condition	Test observations	MAE (W)	RMSE (W)	(R^2)	MAPE (%)	Mean bias (W)	P95 absolute error (W)
RA	Overall	5,930	5.043	7.151	0.645	0.698	0.325	13.954
RA	Multiple traffic flows	3,297	3.954	4.946	0.772	0.548	0.389	9.722
RA	Load and packet-length variation	2,633	6.408	9.195	0.554	0.886	0.246	19.350
RB	Overall	5,936	1.383	1.876	0.823	1.812	-0.032	3.925
RB	Multiple traffic flows	3,302	1.278	1.695	0.840	1.688	0.164	3.482
RB	Load and packet-length variation	2,634	1.515	2.081	0.803	1.966	-0.278	4.416

The RA measured-versus-predicted distribution in Fig. 5 confirms that predictions captured the central load-dependent power trend but compressed part of the observed variability toward the conditional mean. Predictions were concentrated largely between approximately 700 and 743 W, whereas measured values extended from 690 to 766 W. The model therefore represented normal operating levels more accurately than extreme short-duration power deviations. This behavior is consistent with the moderate independent-test (R^2) of 0.645 and the higher errors recorded during packet-length variation.

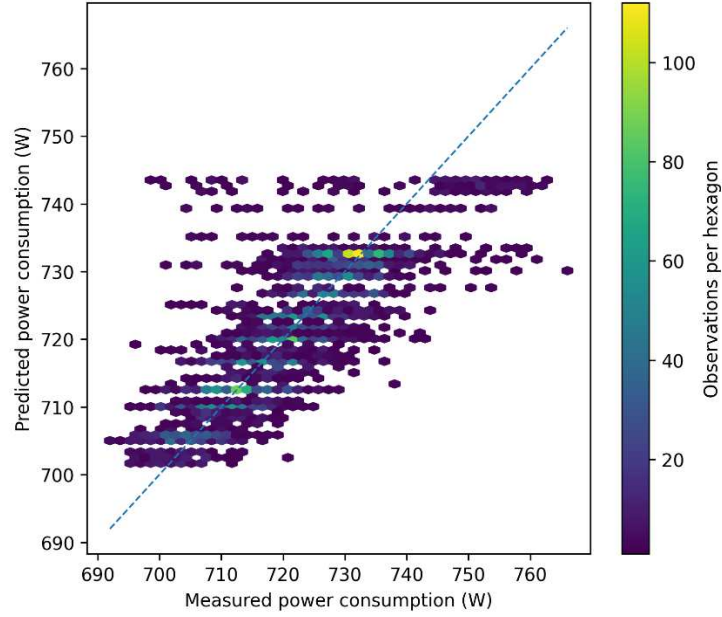


Fig. 5. RA Measured Versus Predicted Power on the Independent Test Cycle

The RB predictions were more tightly aligned with the reference diagonal in Fig. 6. The independent model achieved an MAE of 1.383 W, RMSE of 1.876 W, and (R^2) of 0.823. Its mean bias was only -0.032 W, indicating that cumulative predictions were not materially shifted toward either overestimation or underestimation. Although the predicted values formed operating-state bands, their centers followed the measured power levels closely across the approximately 64–92 W range.

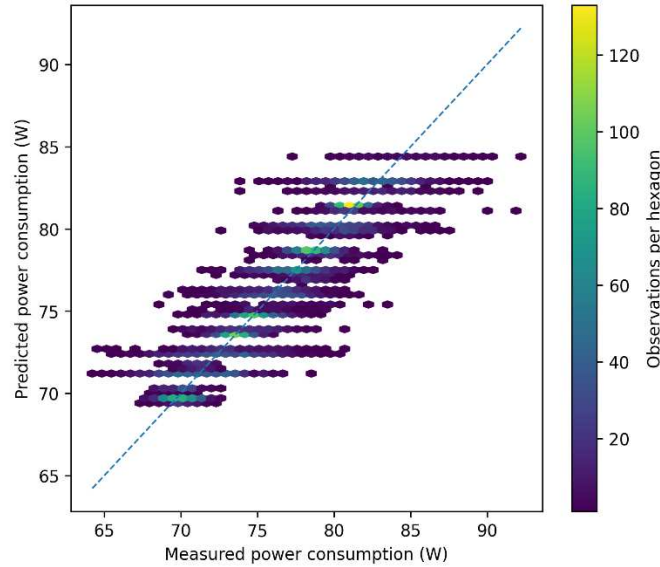


Fig. 6. RB Measured Versus Predicted Power on the Independent Test Cycle

The temporal behavior of the RA model under multiple-flow traffic is shown in Fig. 7. The measured trace contained rapid fluctuations of several watts around each operating level, while the

predicted trace represented the smoother state-dependent mean. The model reproduced the downward changes in power associated with decreasing traffic load but did not attempt to reproduce high-frequency measurement noise. This distinction explains why the model attained a relatively low MAPE of 0.548% in the multiple-flow experiment despite visible point-level fluctuations.

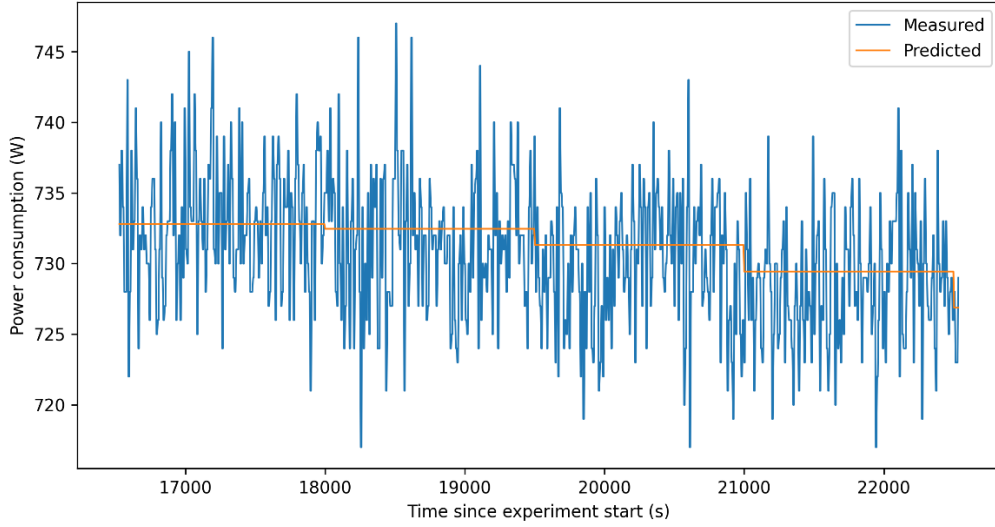


Fig. 7. RA Measured and Predicted Multiple-Flow Power Trace

A clearer state-tracking pattern is visible for RB in Fig. 8. The predicted trace followed successive power reductions from approximately 82.8 W to 81.5 W, 80.0 W, and below 79 W as the downward traffic ramp progressed. Measured values fluctuated around these predicted plateaus, but the model retained the major temporal transitions. This behavior is consistent with the stronger multiple-flow performance of MAE 1.278 W and (R^2) 0.840.

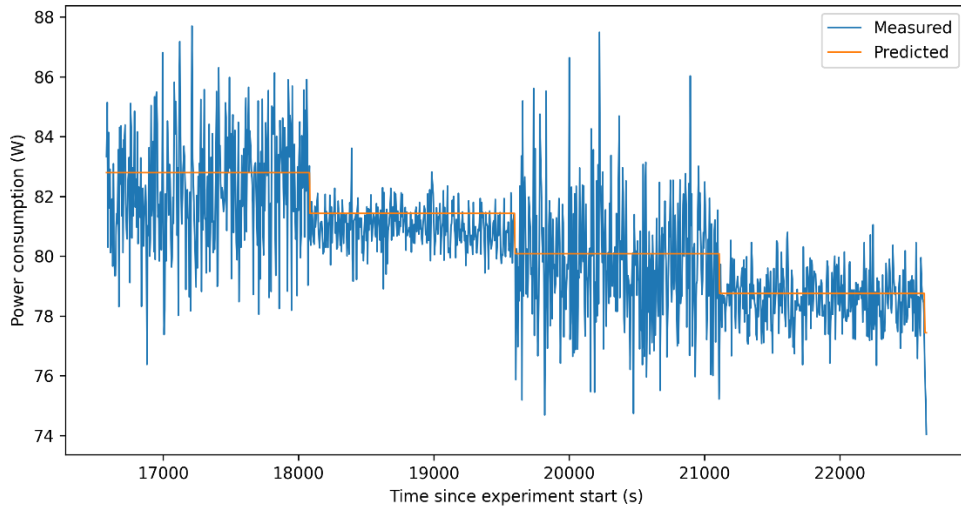


Fig. 8. RB Measured and Predicted Multiple-Flow Power Trace

The stability of the overall results was assessed using a 500-repetition moving-block bootstrap that preserved approximately five minutes of local temporal dependence. As reported in Table 4, the 95% confidence interval for RA MAE was 4.606–5.441 W, while its RMSE interval was 6.285–7.832 W. For RB, the corresponding intervals were 1.290–1.468 W for MAE and 1.764–1.990 W for RMSE. The confidence intervals remained sufficiently narrow to support the observed performance difference between the two routers.

Table 4. Bootstrap Confidence Intervals and Conformal Uncertainty

Router	MAE, 95% CI (W)	RMSE, 95% CI (W)	(R^2), 95% CI	Nominal interval coverage (%)	Empirical coverage (%)	Mean interval width (W)
RA	4.606–5.441	6.285–7.832	0.573–0.700	90.0	88.55	19.58
RB	1.290–1.468	1.764–1.990	0.784–0.845	90.0	90.52	6.34

4.4 Feature-Set Contribution and Predictive Uncertainty

Feature-set ablation was conducted to determine whether the expanded packet-rate representation provided meaningful predictive value beyond the variables used in the original dataset design. As shown in Table 5, traffic load alone produced an MAE of 5.698 W for RA and 1.410 W for RB. Packet size alone performed poorly, with MAEs of 9.418 and 3.747 W, respectively. These results establish utilization as the principal single predictor of router power.

Table 5. Independent-Test Feature-Set Ablation

Router	Feature set	MAE (W)	RMSE (W)	(R^2)
RA	Load only	5.698	7.798	0.578
RA	Packet size only	9.418	11.682	0.053
RA	Official two-feature set	5.084	7.106	0.650
RA	Enhanced without interaction	5.040	7.109	0.649
RA	Enhanced full set	5.043	7.151	0.645
RB	Load only	1.410	1.918	0.815
RB	Packet size only	3.747	4.453	0.001
RB	Official two-feature set	1.411	1.914	0.816
RB	Enhanced without interaction	1.411	1.913	0.816
RB	Enhanced full set	1.383	1.876	0.823

For RA, combining load and packet size reduced MAE from 5.698 to 5.084 W. Adding transformed packet-rate variables produced only a marginal further reduction to approximately 5.04 W. For RB, the enhanced full feature set reduced MAE from 1.411 to 1.383 W and increased (R^2) from 0.816 to 0.823. The ablation comparison in Fig. 9 therefore shows that packet-rate information refined prediction but did not replace traffic utilization as the dominant explanatory variable. The limited difference between the official and enhanced feature sets also suggests that increased feature complexity should be justified by measurable generalization gains rather than training accuracy alone.

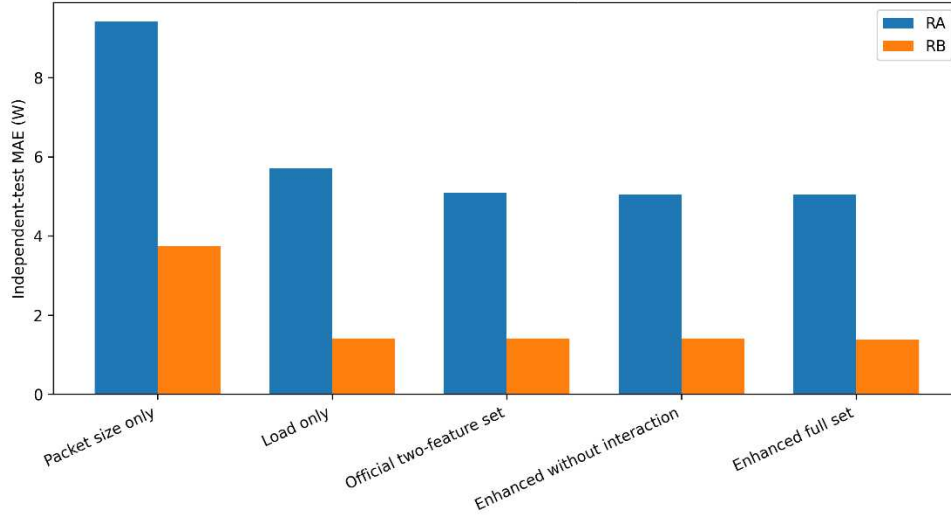


Fig. 9. Independent-Test Ablation of Network Telemetry Feature Sets

Predictive uncertainty was quantified using grouped split-conformal intervals with nominal 90% coverage. RA achieved empirical coverage of 88.55% with a half-width of 9.79 W and a mean total interval width of 19.58 W. The slight undercoverage reflects the greater variability observed during RA packet-length changes. RB achieved 90.52% empirical coverage with a half-width of 3.17 W and a mean interval width of 6.34 W, closely matching the nominal coverage target.

The representative RB interval in Fig. 10 demonstrates that the measured power trace remained inside the calibrated uncertainty band for most observations, including the transition between load plateaus. The interval also widened the interpretation of point predictions: a proposed traffic-engineering action should not be treated as energy beneficial when its expected saving is smaller than the associated predictive interval. This uncertainty-aware decision rule is particularly important where route changes introduce migration, control-plane or additional-hop overhead.

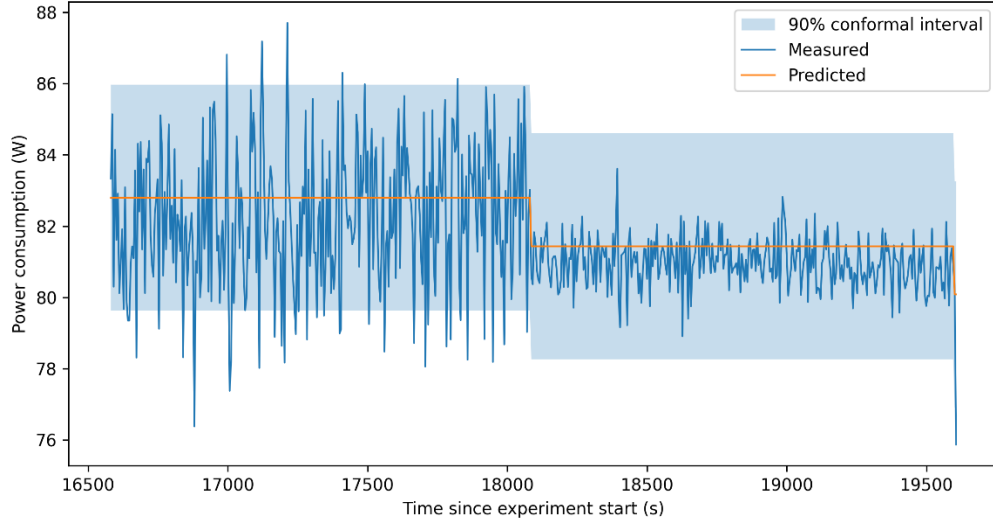


Fig. 10. Representative RB 90% Conformal Prediction Interval

5. Discussion

The findings demonstrate that router power consumption can be estimated effectively from a compact set of traffic variables, although predictive behaviour is strongly hardware dependent. Router RA operated at a substantially higher power level and exhibited a relatively small dynamic increase above idle consumption, whereas Router RB showed a lower absolute power level but a larger proportional response to traffic load. This distinction explains why a single pooled model would have been inappropriate. The result supports the argument that Network Digital Twin models should preserve equipment-specific characteristics rather than assume a universal power function across heterogeneous devices [11], [12]. It is also consistent with evidence that node-level network behaviour can be learned more reliably when device-specific operational patterns are represented explicitly [17]. The superior performance of polynomial ridge regression for RA and linear regression for RB indicates that model complexity should be determined empirically rather than assumed. RA required nonlinear interaction terms to represent the relationship among utilisation, packet size and packet-processing intensity, whereas RB was sufficiently described by a linear structure. This finding complements the view that machine-learning models for network resource management should balance predictive accuracy, computational cost and deployment practicality [14]. The relatively modest gains obtained from packet-rate transformations further confirm that throughput utilisation was the dominant explanatory variable, although packet size and packet rate refined prediction under more variable forwarding conditions. Similar dependence on traffic intensity has been reported in telemetry-driven mobile-network energy studies [6], [19].

The independent downward-ramp evaluation produced stronger explained-variance performance for RB than RA. RA's lower R^2 and wider error range indicate that its instantaneous power contained variability not fully explained by the available telemetry, particularly when packet length changed repeatedly. By contrast, RB displayed a more stable load–power relationship. These differences reinforce the value of leakage-resistant temporal evaluation. Randomly mixing neighbouring five-second observations would likely have overstated generalisation because repeated measurements from identical operating states are highly correlated. The present design

therefore provides a more credible estimate of generalisation to a held-out temporal ramp than conventional row-level partitioning. The bootstrap and conformal analyses add an important reliability dimension. Empirical interval coverage closely approached the nominal 90% target, particularly for RB, showing that uncertainty can be quantified alongside point prediction [26], [27]. This is operationally significant because energy-aware routing should not react to predicted savings smaller than model uncertainty. Consequently, the results support router-specific power models as components of zero-touch and digital-twin systems, but they do not alone establish network-wide energy savings. Such claims require additional evaluation of routing, interface consolidation, sleep states, congestion and latency trade-offs [3], [5].

6. Limitations

This study has several limitations. The analysis used four author-released experimental files covering only two carrier-grade router families and two controlled traffic conditions, which limits broader hardware and operational generalization. The available telemetry excluded routing paths, interface states, temperatures, protocol mix, congestion and control-plane activity. Consequently, the models estimate instantaneous equipment power but cannot establish end-to-end network energy savings. RA uncertainty coverage also remained slightly below the nominal target directly.

7. Conclusion and Future Work

This study demonstrated that carrier-grade router power consumption can be predicted from a compact set of traffic measurements when model development controls temporal leakage and hardware heterogeneity. Using 23,705 observations, polynomial ridge regression provided the strongest generalization for Router RA, whereas linear regression was most suitable for Router RB. Independent temporal evaluation confirmed that model structure should be selected separately for each equipment family rather than imposed uniformly. Traffic utilization was the dominant predictor, while packet-size and packet-rate transformations produced smaller but measurable refinements. Moving-block bootstrap intervals showed that the reported error estimates were stable under temporal dependence, and split-conformal intervals provided a useful representation of predictive uncertainty. These findings support lightweight power models as components of Network Digital Twins and zero-touch energy-management systems.

Future work should extend the analysis to router vendors, capacities, protocol mixes and long-duration traces containing diurnal and failure conditions. Transfer learning and online adaptation should reduce recalibration when models are deployed on unseen hardware. Further experiments should incorporate temperature, interface-level counters, queue occupancy and control-plane activity. Most importantly, the models should be embedded in an SDN testbed where routing, interface consolidation and sleep-state decisions are evaluated jointly against energy, latency, congestion, reliability and switching overhead.

References

- [1] International Telecommunication Union. (2024). *Energy efficiency metrics and measurement methods for telecommunication equipment* (Recommendation ITU-T L.1310, 09/2024). ITU Telecommunication Standardization Sector.
- [2] Poornima, M. R., Vimala, H. S., & Shreyas, J. (2023). Holistic survey on energy-aware routing techniques for IoT applications. *Journal of Network and Computer Applications*, 213, Article 103584. <https://doi.org/10.1016/j.jnca.2023.103584>
- [3] Pathan, M. N., Muntaha, M., Sharmin, S., Saha, S., Uddin, M. A., Nur, F. N., & Aryal, S. (2024). Priority-based energy- and load-aware routing algorithms for SDN-enabled data center networks. *Computer Networks*, 240, Article 110166. <https://doi.org/10.1016/j.comnet.2023.110166>
- [4] Tchinda, A. P., Shala, B., Lehmann, A., Ghita, B., Walker, D., & Trick, U. (2024). Energy-efficient placement of virtual network functions in a wireless mesh network. *IEEE Access*, 12, 64807–64822. <https://doi.org/10.1109/ACCESS.2024.3394907>
- [5] Donatti, A. W., Machado, M. C., Lopez Martinez, M. A., Antunes, S. R. S., Souza, E. C. F., Correa, S. L., Ferreto, T. C., Suruagy, J. A., Martins, J. S. B., & Carvalho, T. C. (2025). Energy efficiency in network slicing: Survey and taxonomy. *IEEE Access*, 13, 134570–134589. <https://doi.org/10.1109/ACCESS.2025.3590365>
- [6] Kolackova, A., Sevgican, S., Fatih Ulu, M., Sadreddini, J., Masek, P., Hosek, J., Jerabek, J., & Tugcu, T. (2024). Exploring potential of ML-aided mobile traffic prediction for energy-efficient optimisation of network resources using real-world dataset. *IEEE Access*, 12, 93606–93622. <https://doi.org/10.1109/ACCESS.2024.3421633>
- [7] European Telecommunications Standards Institute. (2022). *Zero-touch network and service management: Landscape* (ETSI GR ZSM 004 V2.1.1). ETSI.
- [8] European Telecommunications Standards Institute. (2023). *Zero-touch network and service management: Closed-loop automation—Part 3: Advanced topics* (ETSI GR ZSM 009-3 V1.1.1). ETSI.
- [9] International Telecommunication Union. (2022). *Digital twin network: Requirements and architecture* (Recommendation ITU-T Y.3090, 02/2022). ITU Telecommunication Standardization Sector.
- [10] Kuruvatti, N. P., Habibi, M. A., Partani, S., Han, B., Fellan, A., & Schotten, H. D. (2022). Empowering 6G communication systems with digital twin technology: A comprehensive survey. *IEEE Access*, 10, 112158–112186. <https://doi.org/10.1109/ACCESS.2022.3215493>
- [11] Sanz Rodrigo, M., Rivera, D., Moreno, J. I., Alvarez-Campana, M., & López, D. R. (2023). Digital twins for 5G networks: A modelling and deployment methodology. *IEEE Access*, 11, 38112–38126. <https://doi.org/10.1109/ACCESS.2023.3267548>

- [12] Verdecchia, R., Scommegna, L., Picano, B., Becattini, M., & Vicario, E. (2024). Network Digital Twins: A systematic review. *IEEE Access*, 12, 145400–145416. <https://doi.org/10.1109/ACCESS.2024.3453034>
- [13] European Telecommunications Standards Institute. (2024). *Zero-touch network and service management: Network digital twin* (ETSI GR ZSM 015 V1.1.1). ETSI.
- [14] Pivoto, D. G. S., de Figueiredo, F. A. P., Cavdar, C., Tejerina, G. R. de L., & Mendes, L. L. (2026). A comprehensive survey of machine learning applied to resource allocation in wireless communications. *IEEE Communications Surveys & Tutorials*, 28, 1986–2053. <https://doi.org/10.1109/COMST.2025.3552370>
- [15] International Telecommunication Union. (2025). *Digital twin networks: Framework and functional requirements of the data domain in the network digital twin layer* (Recommendation ITU-T Y.3093, 08/2025). ITU Telecommunication Standardization Sector.
- [16] Lentisco, C. M., De la Osa Mostazo, D., Karamchandani Batra, A., González Sanchez, D., de la Cal García, L., Bellido Triana, L., Mozo Velasco, B. A., Pastor Perales, A., & López, D. R. (2025). *Power consumption and network traffic telemetry datasets of carrier-grade network equipment for smart energy-aware zero-touch traffic engineering* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.17282065>
- [17] Hattori, K., Korikawa, T., Takasaki, C., & Oowada, H. (2023). Recursive router metrics prediction using machine learning-based node modelling for network digital replica. *IEEE Access*, 11, 138638–138654. <https://doi.org/10.1109/ACCESS.2023.3340696>
- [18] Natarajan, J., & Rebekka, B. (2023). Machine learning based small cell ON/OFF for energy efficiency in 5G heterogeneous networks. *Wireless Personal Communications*, 130, 2367–2383. <https://doi.org/10.1007/s11277-023-10383-7>
- [19] Lazrek, H., El Ferindi, H., Zouiten, M., & Moumen, A. (2025). Enhancing energy efficiency in 5G networks through AI-driven dynamic discontinuous reception. *Discover Computing*, 28, Article 245. <https://doi.org/10.1007/s10791-025-09765-1>
- [20] Al-Kadhim, H. M., & Al-Raweshidy, H. S. (2023). Energy-efficient traffic in cloud-based IoT. *IEEE Sensors Journal*, 23(22), 28035–28043. <https://doi.org/10.1109/JSEN.2023.3323805>
- [21] Lloret, J., Romero, O., & Leon, A. (2026). Energy-efficient SDN routing with deep learning for data center networks. *Sustainable Computing: Informatics and Systems*, 51, Article 101368. <https://doi.org/10.1016/j.suscom.2026.101368>
- [22] Si-Mohammed, S., Bardou, A., Begin, T., Guérin Lassous, I., & Vicat-Blanc, P. (2024). NS+NDT: Smart integration of network simulation in Network Digital Twin, application to IoT networks. *Future Generation Computer Systems*, 157, 124–144. <https://doi.org/10.1016/j.future.2024.03.038>

- [23] Poorzare, R., Kanellopoulos, D. N., Sharma, V. K., Dalapati, P., & Waldhorst, O. P. (2025). Network Digital Twin toward networking, telecommunications, and traffic engineering: A survey. *IEEE Access*, 13, 16489–16538. <https://doi.org/10.1109/ACCESS.2025.3531947>
- [24] Raza, S. M., Minerva, R., Crespi, N., Alvi, M., Herath, M., & Dutta, H. (2025). A comprehensive survey of Network Digital Twin architecture, capabilities, challenges, and requirements for the edge–cloud continuum. *Computer Communications*, 236, Article 108144. <https://doi.org/10.1016/j.comcom.2025.108144>
- [25] Li, T., Long, Q., Chai, H., Zhang, S., Jiang, F., Liu, H., Huang, W., Jin, D., & Li, Y. (2025). Generative AI empowered network digital twins: Architecture, technologies, and applications. *ACM Computing Surveys*, 57(6), Article 157, 1–43. <https://doi.org/10.1145/3711682>
- [26] Zhou, X., Chen, B., Gui, Y., & Cheng, L. (2025). Conformal prediction: A data perspective. *ACM Computing Surveys*, 58(2), 1–37. <https://doi.org/10.1145/3736575>
- [27] Schlembach, F., Smirnov, E., Koprinska, I., & Winands, M. H. M. (2025). Conformal multistep-ahead multivariate time-series forecasting. *Machine Learning*, 114(7), Article 165. <https://doi.org/10.1007/s10994-024-06722-9>

Authors:

1. Clinton Amponsah
ORCID: 0009-0006-8414-9794
Department of Computer and Electrical Engineering, University of Energy and Natural Resources, Sunyani, Ghana
Email: clinton.amponsah001@gmail.com
2. Bernard Kyiewu
ORCID: 0009-0004-4806-0077
Department of Computer and Electrical Engineering, University of Energy and Natural Resources, Sunyani, Ghana
Email: kyiewubernard18@gmail.com
3. Linda Bessa-Simmons
Department of Computer and Electrical Engineering, University of Energy and Natural Resources, Sunyani, Ghana
Email: bessa-simons.linda.stu@uenr.edu.gh

Corresponding Author: Clinton Amponsah

Email: clinton.amponsah001@gmail.com

DECLARATIONS**Conflict of Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval

This study did not involve human participants, animal subjects, or any sensitive personal data. Therefore, ethical approval was not required.

Author Contribution

C.A. conceived and designed the study, performed the data curation and preprocessing, developed the machine-learning models, conducted the formal analysis and interpretation of results, and wrote the original manuscript draft. B.K. assisted with model implementation, performed the uncertainty quantification including moving-block bootstrap and split-conformal prediction intervals, and contributed to the validation and visualisation of results. L.B.-S. contributed to the literature review, assisted with feature-set ablation and experimental design, and reviewed and edited the manuscript for intellectual content. All authors read and approved the final manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The telemetry data supporting the findings of this study were collected from operational carrier-grade routers and are publicly available at Zenodo: <https://zenodo.org/records/17282065>.