A Data-Driven Energy Consumption Prediction Model for 5G Base Stations: Addressing Static and Dynamic Power Components

Anonymous authors
Paper under double-blind review

Abstract

The rapid deployment of 5G networks has intensified concerns about energy consumption in mobile communication systems. Unlike previous generations, 5G base stations (BSs) exhibit significant power draw even under zero traffic conditions, with static power accounting for $30 \sim 40\%$ of total energy consumption. This paper proposes a novel data-driven framework that decouples total base station energy consumption into static and dynamic components, enabling more precise energy optimization. For static consumption modeling, we introduce a hybrid ResNet-XGBoost architecture that processes configuration parameters including bandwidth, antenna elements, transmit power, carrier count, and tilt angle. For dynamic consumption, we implement a Tabular Probabilistic Function Network (TabPFN) to capture the nonlinear relationship between resource utilization and energy demand. Experimental results using real-world data from a provincial Chinese telecom operator demonstrate that our model achieves a 15.5% reduction in Mean Absolute Error (MAE) and an R^2 of 0.91 compared to conventional approaches.

1 Introduction

The fifth-generation (5G) wireless technology has ushered in an era of unprecedented data rates, low latency, and massive connectivity, supporting a wide range of applications from enhanced mobile broadband to mission-critical Internet of Things (IoT). However, these advancements come with substantial energy consumption requirements that pose significant challenges for network operators. Compared to 4G systems, 5G base stations typically consume up to three times more power due to increased component complexity, higher bandwidth operations, and extensive antenna configurations (Jianbin et al., 2023). This elevated energy demand not only increases operational expenditures for telecom operators but also contributes substantially to their carbon footprint, conflicting with global sustainability initiatives.

A distinctive characteristic of 5G base station energy consumption is the substantial static power component that persists even when no users are actively being served. Studies indicate that between $30 \sim 40\%$ of total base station power is consumed in idle states, primarily attributable to static configurations of radio units, cooling systems, synchronization mechanisms, and essential signaling overhead (Wu et al., 2015). This static energy drain represents a significant optimization opportunity, as traditional energy management approaches primarily focus on dynamic power components related to traffic load while overlooking the substantial fixed costs inherent in current 5G architectures.

Existing energy models often employ aggregated approaches that combine static and dynamic power components, limiting their effectiveness in identifying specific optimization opportunities. While several machine learning approaches have been proposed for base station energy prediction, including multiple linear regression models and deep reinforcement learning for dynamic sleep strategies (Liu et al., 2018), these methods typically lack the architectural specificity to accurately decouple and quantify the static energy component.

Furthermore, conventional models struggle to capture the complex, nonlinear relationships between base station configuration parameters and their resulting energy impacts, particularly under varying operational conditions.

This paper presents a comprehensive data-driven framework that explicitly addresses these limitations through three key contributions

- Decomposition of total base station energy consumption into distinct static (load-independent) and dynamic (load-dependent) components, enabling more targeted optimization strategies.
- Development of a hybrid ResNet-XGBoost model for static energy prediction that leverages residual learning and gradient boosting to accurately estimate power consumption from configuration parameters.
- Implementation of a Tabular Probabilistic Function Network (TabPFN) for dynamic energy modeling that captures the complex relationship between resource utilization metrics and energy demand.
- Validation using real-world operational data from a provincial Chinese telecom operator, demonstrating significant improvements in prediction accuracy.

The remainder of this paper is organized as follows: Section 2 reviews related work in 5G energy modeling. Section 3 details our methodology for energy decomposition and model development. Section 4 presents experimental results and performance comparisons, and Section 5 concludes with findings.

2 Related Work

The escalating energy consumption of 5G base stations has spurred significant research interest, leading to diverse modeling and optimization approaches. Existing efforts can be broadly categorized into AI-driven predictive modeling, multi-domain dynamic optimization, and the integration of energy storage with demand response mechanisms.

A substantial body of research has focused on applying artificial intelligence (AI) to predict and manage BS energy usage. Early work by Ringwald and Larsson (Ringwald & Larsson, 2023) employed multiple linear regression algorithms to establish relationships between wireless utilization metrics and energy consumption, achieving high goodness-of-fit ($R^2 > 0.99$) in specific scenarios. While effective for particular configurations, these linear models struggle to capture the complex nonlinear interactions between multiple parameters in heterogeneous network environments. Li et al. (Li et al., 2021) developed an energy management model that coordinates communication equipment with supporting infrastructure, formulating the problem as a mixed-integer linear program (MILP) and applying Benders decomposition for large-scale optimization. Their work highlighted the importance of considering both communication and ancillary equipment, which collectively determine total base station energy consumption. Several studies employ deep learning architectures for load forecasting. For instance, Liao et al. (Jianbin et al., 2023) proposed a hybrid DCNN-LSTM model to predict the utilization of Physical Resource Blocks (PRBs), which in turn informs dynamic energy-saving strategies for 5G BSs. This approach highlights the value of temporal and spatial feature extraction from network traffic data. Similarly, an LSTM-based network (Parvathareddy et al., 2025) was integrated with energy storage systems to optimize power load scheduling for 5G BSs in open-pit mines, demonstrating effectiveness in reducing energy consumption. From a security-conscious perspective, a hybrid AI framework combining Explainable AI, CNNs, and RNNs has been proposed to balance energy consumption with cybersecurity in 5G networks (Mishra, 2023), achieving high anomaly detection accuracy alongside significant energy savings.

Despite these advancements, current energy modeling approaches face several limitations. First, most models treat base station energy consumption as a monolithic entity rather than decomposing it into static and dynamic components, missing opportunities for targeted optimizations of the substantial static power component. Second, many proposed algorithms require extensive computational resources that may limit their practical deployment in real-time energy management systems. Finally, validation is often conducted in controlled environments or with limited datasets, raising questions about generalizability across diverse network configurations and operating conditions.

Our work addresses these limitations through a specialized model architecture that explicitly decouples static and dynamic energy components, employs efficient hybrid modeling techniques, and validates using comprehensive operational data from commercial networks.

3 Methodology

Traditional base station energy models typically aggregate total consumption without distinguishing between different operational components

$$E_{\text{total}}^{\text{traditional}} = f(S_{\text{config}}, D_{\text{traffic}}), \tag{1}$$

where f represents a mapping function, S_{config} denotes network configuration parameters and D_{traffic} represents traffic data. While such models can reflect the energy carrying capacity at the device configuration level, they fundamentally overlook the inherent energy overhead associated with essential signaling functions including PDCCH, DMRS, SSB, and CSI-RS transmissions.

To address this limitation, we decompose total base station energy consumption into two distinct components

$$E_{\text{total}} = E_{\text{static}} + E_{\text{dynamic}}. \tag{2}$$

This decomposition enables more precise characterization of energy patterns and targeted optimization strategies for each component.

3.1 Static energy consumption modeling

The static energy component represents the inherent power draw that persists even under zero traffic conditions, constituting $30 \sim 40\%$ of total base station consumption (Piovesan et al., 2022). This component comprises several sub-elements

$$E_{\text{static}} = E_{\text{rf}} + E_{\text{cool}} + E_{\text{sync}} + E_{\text{sig}}, \tag{3}$$

where E_{rf} represents radio frequency circuit energy consumption, E_{cool} denotes cooling system operations, E_{sync} accounts for clock synchronization, and E_{sig} encompasses periodic transmission overhead for control channels, synchronization signals, and reference signals.

To identify the most influential configuration parameters affecting static energy consumption, we applied a backdoor adjustment algorithm (Pearl, 2009) for feature selection

$$S_{\text{config}} = \{p_i \in \boldsymbol{P} \mid \text{ATE}\left(p_i \to E_{\text{static}}\right) > \tau\}, \tag{4}$$

where the Average Treatment Effect (ATE) is defined as

$$\text{ATE} = \mathbb{E}\left[E_{\text{static}} \mid \text{do}\left(p_i = p_i^{\text{high}}\right)\right] - \mathbb{E}\left[E_{\text{static}} \mid \text{do}\left(p_i = p_i^{\text{low}}\right)\right]. \tag{5}$$

Through randomized experiments estimating ATE for various configuration parameters with $\tau = 1.5\sigma_{E_{\text{static}}}$, we identified the most influential parameters for static energy modeling

$$S_{\text{config}} = \{B_{\text{band}}, N_{\text{ant}}, P_{\text{t}}, N_{\text{cc}}, \theta_{\text{tilt}}\}, \tag{6}$$

where B_{band} represents frequency band bandwidth, N_{ant} denotes the number of antenna elements, P_{t} indicates transmit power, N_{cc} signifies carrier count, θ_{tilt} represents antenna tilt angle.

Analysis of actual base station energy data reveals complex nonlinear relationships between configuration parameters and static energy consumption. Specifically, we observed a sublinear relationship between antenna element count and energy consumption, and a segmented linear relationship with frequency bandwidth. To capture these complex patterns, we developed a hybrid ResNet-XGBoost architecture that combines the representational power of deep residual networks with the predictive efficiency of gradient boosting.

The model input is the set of five influential configuration parameters identified through causal discovery $S_{\text{config}} = \{B_{\text{band}}, N_{\text{ant}}, P_{\text{t}}, N_{\text{cc}}, \theta_{\text{tilt}}\}$. A critical pre-processing step is the transformation of these low-dimensional scalar values into a structured tensor format compatible with the convolutional layers of the ResNet architecture. This is achieved through a feature replication and spatial structuring technique. First, each of the five parameters is normalized to a $[0, 1]$ range using min-max normalization to ensure stable training. The normalized 5-dimensional vector is then replicated 13 times to create a 65-element extended vector, which is subsequently truncated to 64 elements. This 64-element vector is finally reshaped into an 8×8 matrix, forming a single-channel ”feature image” or tensor with dimensions $1 \times 8 \times 8$. This structured representation provides a spatial canvas for the subsequent convolutional layers to operate on, enabling them to learn local patterns and complex interactions between the replicated features.

The core feature extractor is a modified 34-layer ResNet architecture adapted for the $1\times8\times8$ input. The initial layer is a 2D convolution with 64 filters of size 3×3 and a stride of 1, followed by batch normalization and a ReLU activation. This is followed by four consecutive residual blocks, each containing two 3×3 convolutional layers with batch normalization and a skip connection. The mathematical formulation of a residual block is

$$\boldsymbol{z}^{(\ell)} = \text{BN}\left(\boldsymbol{W}_2^{(\ell)} \cdot \text{ReLU}\left(\text{BN}\left(\boldsymbol{W}_1^{(\ell)}\boldsymbol{y}^{(\ell-1)}\right)\right)\right), \tag{7}$$

$$\boldsymbol{y}^{(\ell)} = \text{ReLU}(\boldsymbol{z}^{(\ell)} + \boldsymbol{y}^{(\ell-1)}), \tag{8}$$

where $\boldsymbol{W}_1^{(\ell)}$, $\boldsymbol{W}_2^{(\ell)} \in \mathbb{R}^{64\times64}$ represent fully connected layer weights, and BN denotes batch normalization.

To enhance the feature representation, we integrate a Squeeze-and-Excitation (SE) block after the final residual block. This module performs dynamic channel-wise feature recalibration. The ”Squeeze” operation applies global average pooling to the 64 feature maps, converting them to a 64-dimensional vector. The ”Excitation” operation then passes this vector through two fully connected layers (with a ReLU and a sigmoid activation, respectively), learning a per-channel weighting factor. Finally, the ”Scale” operation multiplies each original feature map by its calculated weight, effectively allowing the network to automatically emphasize informative features and suppress less useful ones. The output of the ResNet-SE backbone is a 64-dimensional feature embedding, which is a highly non-linear and discriminative representation of the original five configuration parameters.

The XGBoost component then processes the feature representations extracted by ResNet to generate final energy predictions. XGBoost’s gradient boosting framework employs second-order Taylor expansion for loss function approximation and incorporates regularization terms ($L1/L2$) to prevent overfitting. Through split gain analysis, we identified power amplifier efficiency and clock gating cycles as particularly influential parameters for static energy consumption.

The 64-dimensional feature embedding from the ResNet is then fed as input to an XGBoost regressor for the final energy consumption prediction. XGBoost’s gradient boosting framework employs second-order Taylor expansion for loss function approximation and incorporates L1/L2 regularization terms to prevent overfitting. We utilize 100 decision trees with a maximum depth of 6 and a learning rate of 0.1. Through split gain analysis within the trained XGBoost model, we can interpret the relative importance of different features in the embedding space, providing insights into which configuration aspects most significantly impact static energy consumption.

This hybrid design leverages the strengths of both components: the ResNet excels at learning a powerful, non-linear feature representation from the structured input, while XGBoost provides a highly effective and efficient mechanism for the final regression task on the resulting features.

Table 1: ResNet-XGBoost Hybrid Model Architecture

Component	Layers	Key Features
Input Layer	5 neurons	Configuration parameters
ResNet Backbone	4 residual blocks	Skip connections, batch normalization
SE Attention Module	Squeeze, excitation, scale	Channel-wise attention weights
XGBoost Regressor	100 trees	Max depth=6, learning rate=0.1

3.2 Dynamic business energy modeling

The dynamic energy component characterizes power fluctuations triggered by user activity, comprising two sub-components

$$E_{\text{dynamic}} = E_{\text{traffic}} + E_{\text{user}}, \tag{9}$$

where E_{traffic} represents energy for actual data transmission, and E_{user} denotes energy for connection maintenance and mobility management.

For dynamic energy modeling, we implemented a Tabular Probabilistic Function Network (TabPFN), a novel transformer-based model that leverages Bayesian inference and in-context learning to capture the complex, nonlinear relationships between resource utilization metrics and energy demand (Hollmann et al., 2022). Unlike traditional regression approaches which require dataset-specific training, TabPFN employs a pre-trained prior over neural network weights, enabling it to perform approximate Bayesian inference in a single forward pass. This architecture is particularly advantageous for our application, as it exhibits strong generalization on small-sample datasets (typically up to 10,000 instances) commonly encountered in operational telecom data analytics.

Our TabPFN implementation is based on the publicly available codebase (Hollmann et al., 2022). The core architecture is a Transformer encoder with bidirectional attention, which allows the model to process interactions across both samples (rows) and features (columns) of the input tabular data. This is crucial for capturing the complex interdependencies between PRB utilization, traffic volume, downlink rate, RRC connection count, and handover success rate in determining energy consumption.

The model was pre-trained on a massive, synthetically generated corpus of millions of diverse datasets, created using Structural Causal Models (SCMs). This large-scale pre-training instills a robust prior over tabular data functions, allowing it to generalize zero-shot to our real-world 5G energy prediction task without requiring gradient-based fine-tuning. The hyperparameters of the model were fixed to their default pre-trained values as recommended in (Hollmann et al., 2022), which include a default number of 100 estimators and a prior scaling factor optimized for broad tabular regression tasks. This zero-shot approach inherently addresses hyperparameter sensitivity, as the model is designed to be robust across diverse data characteristics without task-specific tuning.

During inference, we employ the model in its zero-shot learning mode. The entire training dataset (features and corresponding energy consumption labels) is presented to the model as a context within a single forward pass. The model then generates predictions for the test samples based on this in-context information. The output is formulated as a Gaussian mixture model, providing both the prediction and its epistemic uncertainty

$$p(y|x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y \mid \mu_k(x), \sigma_k^2(x)), \tag{10}$$

where $\mu_k(x)$ and $\sigma_k^2(x)$ are dynamically generated by the neural network, and π_k represents mixture weights.

This probabilistic output is particularly valuable for network planning, as it allows operators to assess the confidence of energy consumption forecasts. To adapt the pre-trained model to the specific numerical scale and characteristics of our 5G energy dataset, we applied standard feature scaling (standardization) before feeding the data into the model. No further optimization of the model's internal weights was performed, leveraging its core capability as a Prior-Data Fitted Network (PFN).

4 Experiments and Results

We validated our proposed model using operational data collected from a provincial Chinese telecommunications operator encompassing over 200 5G base stations with diverse configurations and traffic patterns. The dataset included 12 feature dimensions and corresponding measured energy values recorded at 15-minute intervals over a 2-month period.

The feature set included both configuration parameters (bandwidth, antenna count, transmission power, carrier aggregation settings, tilt angle) and dynamic operational metrics (PRB utilization, user count, traffic volume, connection status). The dataset was partitioned using a 7:3 ratio for training and testing, with temporal cross-validation to prevent data leakage. To address class imbalance in configuration patterns, we applied Synthetic Minority Over-sampling Technique to generate synthetic examples for underrepresented configuration combinations.

Our experimental setup implemented LightGBM (Ke et al., 2017) and XGBoost (Chen & Guestrin, 2016) for comparison. All models were evaluated using 5-fold cross-validation with consistent data splits to ensure fair comparison. Implementation was conducted in Python using PyTorch for deep learning components and Scikit-learn for traditional machine learning models.

We employed multiple metrics to comprehensively evaluate model performance, see Table 2 and Figure 1.

- Mean Absolute Error (MAE) : MAE $= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
- Mean Square Error (MSE): MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- Coefficient of Determination (R^2): $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Our proposed ResNet-XGBoost-TabPFN hybrid model demonstrated superior performance across all evaluation metrics, achieving a 15.5% reduction in MAE compared to the next best approach (XGBoost). The model's strong R^2 value of 0.91 indicates excellent explanatory power for the variability in base station energy consumption.

Table 2: Model Performance Comparison on Test Dataset

Model	MAE	MSE	R^2 score
LightGBM	1.21	4.71	0.88
XGBoost	1.16	4.51	0.88
Proposed Model (Ours)	0.98	4.37	0.91

To quantify the contribution of individual model components, we conducted a comprehensive ablation study, see Table 3. The ablation results confirm the importance of each architectural component. The SE attention mechanism contributed significantly to feature discrimination, with its removal resulting in a 12.2% performance degradation. Similarly, the hybrid ResNet-XGBoost design demonstrated clear advantages over either component in isolation, validating our integrated approach.

5 Conclusion

This paper has presented a novel data-driven approach to 5G base station energy consumption prediction that explicitly models static and dynamic components separately. Our hybrid ResNet-XGBoost architecture for static energy prediction effectively captures the

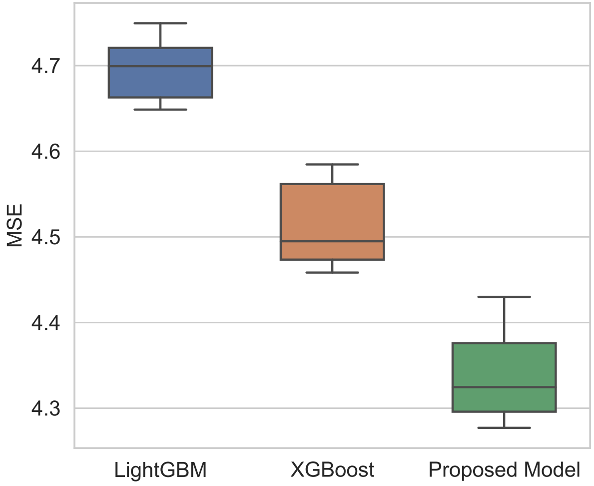

Figure 1: Model Performance Comparison of MAE.

Table 3: Ablation Study Results

Model	MAE	MSE	R^2 score	Performance Change
Complete Model	0.98	4.37	0.91	baseline
Without SE Attention	1.10	4.42	0.89	-12.2%
Without XGBoost	1.12	4.43	0.88	-14.3%

complex relationships between configuration parameters and base energy draw, while the TabPFN model for dynamic energy accurately characterizes traffic-dependent power consumption.

Validation using real-world operational data demonstrates significant improvements over existing approaches, with a 15.5% reduction in MAE and R^2 of 0.91. By providing more accurate and interpretable energy predictions, our model enables network operators to make informed decisions that balance energy efficiency with service quality, contributing to more sustainable 5G network deployment and operation.

References

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794, 2016.

Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. Tabpfn: A transformer that solves small tabular classification problems in a second. arXiv preprint arXiv:2207.01848, 2022.

WANG Jianbin, WANG Shuchun, LIAO Shangjin, and SHI Shuyuan. Research on 5g base station energy saving system based on dcnn-lstm load prediction algorithm. Telecommunications Science, 39(4), 2023.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30, 2017.

Kun Li, Xiaomeng Ai, Jiakun Fang, Bo Zhou, Lingling Le, and Jinyu Wen. Coordination of macro base stations for 5g network with user clustering. Sensors, 21(16):5501, 2021.

Jingchu Liu, Bhaskar Krishnamachari, Sheng Zhou, and Zhisheng Niu. Deepnap: Data-driven base station sleeping operations through deep reinforcement learning. IEEE Internet of Things Journal, 5(6):4273–4282, 2018.

Shailendra Mishra. Artificial intelligence assisted enhanced energy efficient model for device-to-device communication in 5g networks. Human-Centric Intelligent Systems, 3(4):425–440, 2023.

Sravani Parvathareddy, Abid Yahya, Lilian Amuhaya, Ravi Samikannu, and Raymond Sogna Suglo. Transforming mining energy optimization: a review of machine learning techniques and challenges. Frontiers in Energy Research, 13:1569716, 2025.

Judea Pearl. Causal inference in statistics: An overview. 2009.

Nicola Piovesan, David López-Pérez, Antonio De Domenico, Xinli Geng, Harvey Bao, and Mérouane Debbah. Machine learning and analytical power consumption models for 5g base stations. IEEE Communications Magazine, 60(10):56–62, 2022.

Daniel Ringwald and Daniel Larsson. Energy consumption modelling for 5g radio base stations with machine learning. 2023.

Jingjin Wu, Yujing Zhang, Moshe Zukerman, and Edward Kai-Ning Yung. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. IEEE communications surveys & tutorials, 17(2):803–826, 2015.