

(REVIEW ARTICLE)



TRENDS, METHODS, AND APPLICATIONS OF PREDICTIVE ENERGY ANALYTICS IN DATA CENTER CAPACITY PLANNING: A COMPREHENSIVE LITERATURE REVIEW

Afor Avwioroko *

Marshall School of Business, University of Southern California.

* Corresponding Author

Global Journal of Engineering and Technology Advances, 2026, 28(03), 312–323

Publication history: Received on 11 August 2026; revised on 27 September 2026; accepted on 29 September 2026

Article DOI: <https://doi.org/10.30574/gjeta.2026.28.3.0272>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

The rapid expansion of data-center infrastructure has intensified the need for reliable and efficient energy management, particularly as growing computational workloads place greater demands on power, cooling, and supporting infrastructure. Predictive energy analytics has emerged as a key response, using statistical and physics-based models, machine-learning and deep-learning techniques, and optimization and digital-twin approaches to predict energy behavior and support operational decision-making. However, the literature remains fragmented across these methodological streams, with limited synthesis of how they collectively support data-center capacity planning. This review asks: How do statistical, machine-learning, and digital-twin/optimization approaches to predictive energy analytics compare for data center capacity planning? A comprehensive narrative synthesis of 22 studies published between 2014 and 2026 examines evidence across the three methodological streams and their applications to data-center energy and capacity management. Five themes emerged: methodological fragmentation; digital twins as an unproven-at-scale integration layer; the dominance of PUE and cooling targets over direct capacity-planning objectives; the prevalence of single-facility validation; and the absence of calibrated uncertainty. The findings from this synthesis indicate that predictive energy analytics is substantially more developed for optimizing existing energy infrastructure than for forecasting workload-driven capacity requirements, while cross-stream integration, multi-facility validation, and uncertainty-aware prediction remain limited. Closing these gaps would move the field toward integrated, generalizable, and decision-oriented predictive frameworks capable of supporting more reliable data-center capacity-planning decisions.

Keywords: Predictive Energy Analytics; Data Center Capacity Planning; Machine Learning for Energy Forecasting; Digital Twin Energy Management; Power Usage Effectiveness (PUE)

1 INTRODUCTION

Reliable and efficient energy infrastructure is a widespread concern across the energy sector, spanning both the physical systems that support energy generation and conversion and the data-driven systems used to manage them. At the hardware level, resilient power electronics are crucial to developing robust renewable energy systems that operate under harsh environmental conditions, while the breadth of power-electronics applications and the scale of their associated market further demonstrate their importance across modern energy infrastructure (Sefa-Boateng et al.,

2025). At the data and decision-making level, AI, machine learning, and big-data analytics are increasingly integrated into energy systems to enhance operational efficiency, reduce risks, and support intelligent system management (Ighomuaye & Annankra, 2025; Oketie & Aliu, 2026). This trend has been documented across U.S. energy domains including geothermal, nuclear, and carbon capture, utilization, and storage systems, rather than data centers specifically, indicating that AI-enabled and data-driven infrastructure management is developing as a broad cross-domain trend. This broader movement toward intelligent energy management is increasingly relevant to data centers, which are recognized as critical infrastructure requiring reliable power. Alao et al. (2026) note that microgrids can improve reliability for critical infrastructure such as hospitals, data centers, and manufacturing facilities, while also documenting machine-learning-based forecasting and digital-twin methods in adjacent renewable-energy and storage microgrid systems for predictive maintenance and control. These developments provide an initial indication of the types of predictive and intelligent methods being applied to energy infrastructure. Within the data-center context itself, the increasing share of operations concentrated in large, ultra-efficient hyperscale data centers represents what Lei and Masanet (2020) describe as the “hyperscale shift”. At the same time, the development of data-center energy modeling has historically been uneven. Dayarathna et al. (2016) observed that few modeling efforts had targeted the power consumption of the entire data center as a whole, highlighting the need for approaches capable of addressing energy behavior beyond individual components or subsystems.

Predictive energy analytics has emerged as a primary response to this challenge within data centers, encompassing three broad methodological families: statistical and physics-based methods for representing physical energy behavior; machine-learning and deep-learning methods for forecasting energy consumption, PUE, and workload; and optimization and digital-twin approaches for connecting prediction with resource allocation, simulation, and operational control. Existing reviews, however, indicate that these approaches have not been consistently synthesized within a common framework. Dayarathna et al. (2016) brought statistical and machine-learning methods together within a taxonomy but explicitly excluded optimization and digital-twin approaches, treating optimization primarily as a downstream application. Panwar et al. (2022) likewise organized the field into heuristic, metaheuristic, machine-learning, and statistical families without proposing a unifying framework across them. Ferreira et al. (2026) subsequently characterized the literature as “extensive, fragmented, and often complex” and connected machine-learning forecasting with optimization, but did not incorporate the statistical/physics-based stream.

Despite this, no existing literature review has synthesized predictive energy analytics for data center capacity planning across its three main methodological streams, statistical/physics-based models, machine-learning/deep-learning forecasting, and optimization/digital-twin frameworks, into a single comparative framework. Against this background, this review asks: How do statistical, machine-learning, and digital-twin/optimization approaches to predictive energy analytics compare for data center capacity planning? To address this question, the review develops a Three-Stream Convergence Framework that provides a common structure for comparing these approaches according to their data characteristics, method type, integration scope, validation scope, uncertainty treatment, and decision applications. The framework is derived inductively from the evidence synthesized across the reviewed studies and identifies areas of methodological convergence as well as persistent limitations in validation, integration, uncertainty treatment, and direct capacity-planning applications.

1.1 Review Methodology

This study adopts a comprehensive narrative literature review to examine trends, methodological developments, and applications of predictive energy analytics in data center capacity planning. This approach was selected to enable an integrated synthesis of evidence across diverse methodological approaches, including statistical, machine-learning, optimization, and digital-twin methods. Rather than following a systematic review protocol, the review emphasizes the identification and critical synthesis of relevant research to characterize how predictive energy analytics has evolved and how these methods have been applied to data center energy and capacity-related challenges.

1.2 Search Strategy

A comprehensive literature search was conducted across Google Scholar, Scopus, IEEE Xplore, and ScienceDirect to identify studies relevant to predictive energy analytics and data center capacity planning. Searches were guided by keywords including predictive energy analytics, energy forecasting, energy consumption prediction, data center energy consumption, data center capacity planning, data center workload forecasting, machine learning for energy prediction, and energy optimization in data centers.

Retrieved studies were screened by the author for relevance to the review's scope, with emphasis on predictive, statistical, machine-learning, optimization, and digital-twin approaches to energy consumption, power demand, cooling, workload, and capacity planning. Twenty-two studies considered most directly relevant to the review were retained for analysis.

1.3 Inclusion and Exclusion Criteria

A total of 22 studies were included based on their relevance to the scope of the review. Studies were eligible if they addressed data center energy consumption, power demand, cooling, workload, or capacity planning, or, in a limited number of cases, applied directly comparable predictive or AI methods to other critical energy-infrastructure contexts. Eligible studies were also required to apply a predictive, statistical, machine-learning, optimization, or digital-twin method to the outcome of interest. Formal peer-review status was not a strict inclusion requirement. Instead, methodological relevance and technical credibility guided source selection, allowing peer-reviewed journal and conference studies to be considered alongside technically credible sources such as a Google technical white paper and an arXiv preprint. The included literature spans 2014 to 2026 and comprises primary empirical studies, technical reports, preprints, and prior literature reviews. Studies that did not satisfy the relevance and methodological criteria were excluded.

1.4 Analytical Framework

To structure the comparative synthesis, this review develops a Three-Stream Convergence Framework that organizes predictive energy analytics into three methodological streams: statistical/physics-based models, machine-learning/deep-learning approaches, and optimization/digital-twin frameworks. The framework was developed inductively from the extraction categories applied across the 22 reviewed studies, particularly data characteristics, method type, integration scope, validation scope, and uncertainty treatment. It was therefore derived from the evidence synthesized in the literature rather than adopted from an external theoretical framework.

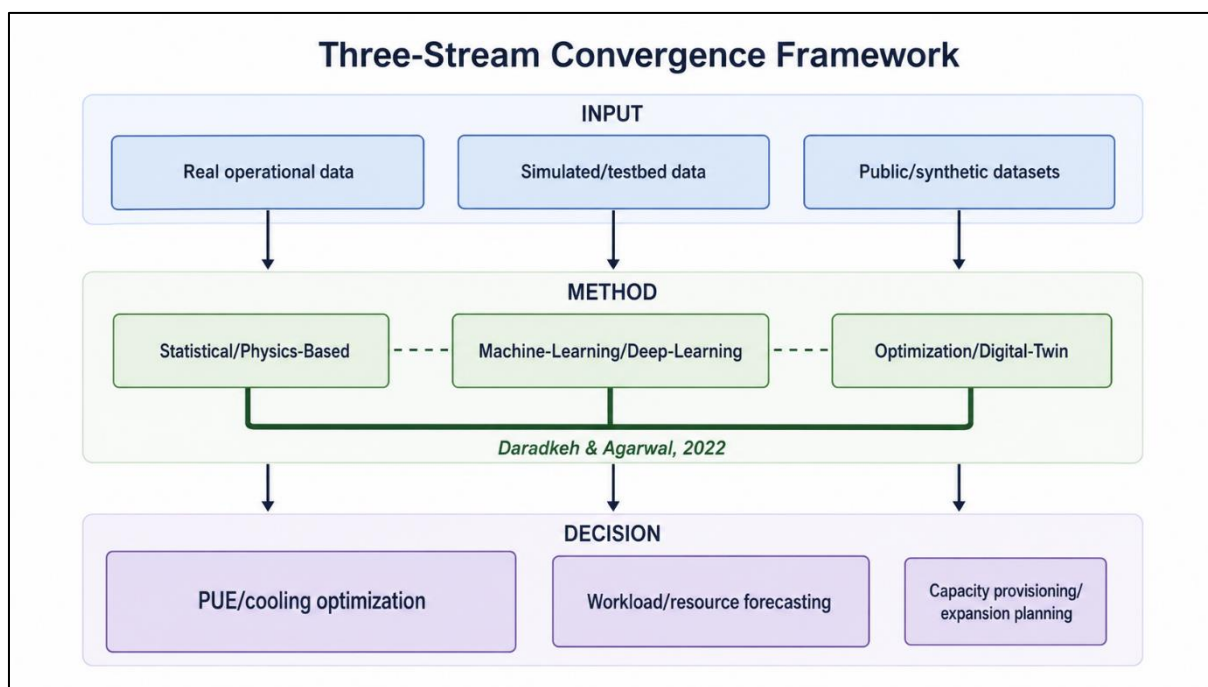


Figure 1 The resulting framework

Source: Author's construct, 2026

As illustrated in Figure 1, the framework links variations in data inputs to the three methodological streams and, subsequently, to their principal decision applications in data center energy and capacity planning. The framework highlights both the fragmentation of existing methodological approaches and the limited convergence among them.

2 TRENDS AND METHODS IN PREDICTIVE ENERGY ANALYTICS FOR DATA CENTER CAPACITY PLANNING

Predictive energy analytics in data centers has evolved from physics-based and statistical modeling toward increasingly data-driven and integrated approaches. Early research emphasized power equations, queuing models, and physics-based representations of cooling and infrastructure, as documented by Dayarathna et al. (2016). From the late 2010s onward, machine-learning and deep-learning methods became increasingly prominent for predicting energy consumption, workload, and PUE (Panwar et al., 2022, 2024; Yang et al., 2022). More recent research has increasingly combined predictive analytics with optimization and digital-twin technologies, particularly in studies published from 2022 onward (Chinnici et al., 2024; Fujii et al., 2021; Gonçalves & Santos, 2026; Zohdi, 2022). This shift extends

prediction toward integrated energy management and decision support, although the literature remains fragmented across the three methodological streams. The following sections examine these streams comparatively.

2.1 Statistical/Physics-Based Methods

Statistical and physics-based methods form the methodological foundation of predictive energy analytics in data centers, but the literature reveals a persistent trade-off between physical interpretability and comprehensive validation. Dayarathna et al. (2016) documented extensive use of hardware-level power equations, queuing models, and physics-based cooling models, while noting that accuracy was often insufficiently characterized and that whole-data-center modeling lagged behind component-level approaches. This limitation is partly addressed by Lei and Masanet (2020), whose thermodynamics-based PUE models were evaluated across 17 hyperscale facilities and achieved yearly estimates within approximately $\pm 4\%$ of observed values. More importantly, their use of calibrated 90% prediction intervals demonstrates a level of uncertainty characterization largely absent from the earlier modeling literature (Lei & Masanet, 2020). Mustaram et al. (2023) illustrates how physics-based modeling can be embedded within a digital-twin environment using real-time sensor data, identifying substantial cooling over-provisioning in a campus data center. However, its single-facility setting limits the generalizability of the findings. Thus far, the evidence suggests that physics-based approaches remain valuable where interpretability, physical constraints, and uncertainty characterization are important, but their broader capacity-planning value depends on stronger multi-facility validation and integration with operational decision processes. This limitation provides a basis for the subsequent shift toward machine-learning approaches.

2.2 Machine-Learning/Deep-Learning Methods

Machine-learning and deep-learning approaches mark a shift toward data-driven prediction, with several studies demonstrating strong predictive performance and potential for operational energy savings. Yang et al. (2022) reported substantially lower PUE prediction error using a neural network trained on nearly 900,000 observations from a large operational data center, while Gao and Jamidar (2014) achieved a low PUE prediction error using more than 184,000 real sensor observations. These findings indicate that ML can support both prediction and identification of efficiency improvements.

However, the evidence is difficult to compare across studies because of differences in data, evaluation metrics, and validation settings. Panwar et al. (2022, 2024) reported wide ranges of energy savings across ML applications while noting the substantial use of simulation environments. Gebreyesus et al. (2023) demonstrated the value of combining ML with SHAP-based feature selection, whereas Wilson et al. (2024) reported energy reductions of 12% to 18% without providing standard forecasting metrics such as RMSE, MAPE, or R^2 . Thus, although ML/DL offers greater flexibility for capturing complex energy relationships, its reported performance does not by itself establish suitability for capacity planning. This highlights the need to integrate predictive models with optimization and decision-making frameworks.

2.3 Optimization/Digital-Twin Methods

Optimization and digital-twin approaches represent a further shift from prediction toward integrated decision support, linking forecasting, simulation, optimization, and control. The evidence, however, suggests that this integration remains uneven. Zohdi (2022) demonstrated how a physics-based digital twin can be coupled with machine-learning optimization for cooling design, but the analysis relied entirely on a simulated environment without real-facility validation. In contrast, Gonçalves and dos Santos (2026) combined a digital twin with LSTM forecasting and anomaly detection in a small testbed, achieving improved forecasting performance and a reduction in PUE from 1.85 to 1.70. However, the authors explicitly caution against direct generalization to large-scale data centers.

Other studies demonstrate the potential of connecting prediction directly to operational optimization. Shaffi et al. (2026) integrated XGBoost forecasting with MILP optimization and reported improvements in renewable utilization and operating cost, while Daradkeh and Agarwal (2022) combined analytical and ML workload prediction with elastic-scaling optimization. Notably, the latter is the only study in the reviewed literature that spans all three methodological streams, providing the clearest example of convergence between statistical, ML, and optimization approaches. A broader limitation is the predominance of simulated or non-data-center settings within this stream. Ferreira et al. (2026) similarly characterize the literature as extensive but fragmented, while several digital-twin/ML studies in the literature focus on buildings rather than data centers directly (Agostinelli et al., 2021; Almadhor et al., 2025; Fujii et al., 2021; Han et al., 2024; Mahamedi et al., 2025). Thus, although optimization and digital twins offer the strongest pathway toward integrated capacity-planning decisions, limited real-world validation and cross-stream integration remain important gaps.

3 APPLICATIONS IN DATA CENTER CAPACITY PLANNING

Predictive energy analytics is applied across several areas of data center energy and resource management, but its direct contribution to capacity planning remains limited. PUE and cooling optimization is the most prominent application. Studies by Lei and Masanet (2020), Mustaram et al. (2023), Gao and Jamidar (2014), Yang et al. (2022), and Zohdi (2022), as discussed above, demonstrate how predictive and physics-based approaches can identify efficiency opportunities and optimize cooling performance.

A second application is workload and resource-utilization forecasting, where predictive models support virtual-machine consolidation, placement, migration, and scaling decisions (Panwar et al., 2022, 2024; Gebreyesus et al., 2023; Daradkeh & Agarwal, 2022). These applications provide an important link between forecasting and capacity decisions by helping align computing resources with changing workloads.

Despite this potential, explicit capacity provisioning and expansion planning is notably underrepresented. Daradkeh and Agarwal (2022) is the only study in the reviewed corpus that directly identifies data-center resource provisioning and capacity planning as an explicit objective. This scarcity is consistent with Dayarathna et al. (2016), who noted the limited development of whole-data-center modeling. The literature therefore appears considerably more mature in optimizing existing capacity than in informing decisions about future capacity requirements.

Digital twins further extend predictive analytics toward real-time monitoring and control, as demonstrated by Zohdi (2022), Mustaram et al. (2023), Gonçalves and dos Santos (2026), and Chinnici et al. (2024). Related studies by Mahamedi et al. (2025), Agostinelli et al. (2021), Han et al. (2024), Fujii et al. (2021), and Almadhor et al. (2025) apply similar digital-twin approaches to building energy systems rather than data centers, providing methodological parallels rather than direct data-center evidence. Renewable-energy integration is another emerging application, with Shaffi et al. (2026) and Ferreira et al. (2026) linking energy forecasting with optimization for renewable and grid management.

4 THEMATIC SYNTHESIS: CONVERGENCES AND DIVERGENCES

The thematic synthesis identifies five recurring patterns that cut across the methodological approaches and applications examined in the reviewed literature. These themes capture both areas of convergence and notable exceptions concerning methodological integration, digital-twin validation, capacity-planning targets, validation scope, and uncertainty treatment.

Table 1 summarizes these patterns and the studies supporting or diverging from each theme.

Table 1: Thematic Synthesis Matrix

Theme	Thesis Statement	Convergent (Supporting) Studies	Divergent Studies (Exception)
Methodological Fragmentation	The literature converges on treating statistical/physics-based, machine-learning, and optimization/digital-twin methods as separate design choices rather than complementary components, and only a small minority of studies break from this norm by demonstrating that cross-stream integration is technically feasible and yields measurable gains.	Lei and Masanet (2020); Panwar et al. (2024); Gao and Jamidar (2014); Wilson et al. (2024); Gebreyesus et al. (2023); Panwar et al. (2022); Jayaprakash et al. (2021)	Zohdi (2022); Khan et al. (2023); Shaffi et al. (2026); Daradkeh and Agarwal (2022)
Digital Twins as an Unproven-at-Scale Integration Layer	Studies agree that digital-twin architectures are the natural mechanism for coupling prediction with real-time control, but they diverge sharply on validation rigor, since nearly all confine testing to a single small, household, or simulated facility rather than production infrastructure.	Mustaram et al. (2023); Gonçalves and dos Santos (2026); Chinnici et al. (2024); Zohdi (2022); Fujii et al. (2021); Han et al. (2024); Mahamedi et al. (2025); Agostinelli et al. (2021); Almadhor et al. (2025)	No direct production-scale exception identified. The cited studies collectively demonstrate integration but remain limited by single-site, small-scale, or simulated validation.
PUE/Cooling Dominance over Capacity-Planning Targets	There is broad agreement that PUE and cooling-system energy are the most tractable and most frequently modeled targets, but this very consensus signals a divergence from the field's stated capacity-planning ambitions, since almost no study models workload-driven capacity provisioning directly.	Lei and Masanet (2020); Gao and Jamidar (2014); Yang et al. (2022); Mustaram et al. (2023); Zohdi (2022); Dayarathna et al. (2016)	Daradkeh and Agarwal (2022)
Single-Facility Validation Norm	Most studies converge on single-facility or simulated validation as sufficient grounds for publishable results, but a small minority explicitly reject this norm, arguing that single-facility validation cannot support generalizable claims.	Zohdi (2022); Gonçalves and dos Santos (2026); Mustaram et al. (2023); Chinnici et al. (2024); Khan et al. (2023); Han et al. (2024); Almadhor et al. (2025); Wilson et al. (2024); Fujii et al. (2021); Mahamedi et al. (2025)	Lei and Masanet (2020); Daradkeh and Agarwal (2022)
Absence of Calibrated Uncertainty	Nearly every reviewed study treats deterministic point forecasting as an adequate basis for energy-management decisions, a near-universal practice that two studies separately and explicitly flag as insufficient for decisions made under genuine uncertainty.	Gonçalves and dos Santos (2026); Khan et al. (2023); Yang et al. (2022); Wilson et al. (2024); Mahamedi et al. (2025); Fujii et al. (2021); Almadhor et al. (2025); Zohdi (2022); Panwar et al. (2022); Ferreira et al. (2026)	Lei and Masanet (2020)

Source: Author's construct, 2026

Table 1 summarizes the five major themes emerging from the comparative analysis, highlighting where the reviewed studies converge and where notable methodological or empirical exceptions occur. Together, these patterns reveal persistent fragmentation, limited validation at scale, an emphasis on PUE and cooling rather than capacity provisioning, and insufficient treatment of uncertainty, providing the basis for the cross-cutting synthesis that follows

4.1 Methodological Fragmentation

The literature converges on treating statistical/physics-based, machine-learning, and optimization/digital-twin methods as separate design choices rather than complementary components, and only a small minority of studies break from this norm by demonstrating that cross-stream integration is technically feasible and yields measurable gains. This fragmentation is evident across studies that remain within a single methodological stream. Lei and Masanet (2020) use thermodynamic modeling and uncertainty analysis for PUE prediction, while Gao and Jamidar (2014), Yang et al. (2022), Wilson et al. (2024), and Gebreyesus et al. (2023) primarily employ machine-learning approaches for prediction and efficiency analysis. Panwar et al. (2022) similarly organize the literature into distinct heuristic, metaheuristic, machine-learning, and statistical families, reinforcing the separation of methodological approaches.

A smaller group demonstrates the feasibility of cross-stream integration. Zohdi (2022) combines physics-based thermo-fluid modeling with machine-learning optimization within a digital-twin framework, while Khan et al. (2023) integrates statistical and machine-learning methods for energy management. Shaffi et al. (2026) further connects XGBoost-based forecasting with mixed-integer linear programming optimization. The strongest example is Daradkeh and Agarwal (2022), which integrates analytical workload modeling, deep-learning prediction, and optimization for elastic resource scaling. This convergence is nevertheless uncommon. Ferreira et al. (2026) characterize the field as “extensive, fragmented, and often complex”, while Panwar et al. (2022) similarly identify distinct methodological families with limited integration. Jayaprakash et al. (2021) reach a comparable conclusion, organizing cloud energy management into separate clustering, optimization, and machine-learning families and noting that existing optimization methods suffer from low convergence and easily trap into local optima, further reinforcing the field’s methodological fragmentation. The resulting gap is particularly important for capacity planning: prediction alone does not determine future infrastructure requirements, while optimization without reliable forecasts may rely on static or simulated assumptions. The evidence therefore suggests that integrating complementary methodological streams may offer greater value for capacity planning than treating them as competing alternatives, although real-world demonstrations of such integration remain limited.

4.2 Digital Twins as an Unproven-at-Scale Integration Layer

Studies agree that digital-twin architectures are the natural mechanism for coupling prediction with real-time control, but they diverge sharply on validation rigor, since nearly all confine testing to a single small, household, or simulated facility rather than production infrastructure.

Data-center studies illustrate this validation limitation. Mustaram et al. (2023) implemented a digital twin in a single campus data center using real-time sensor data, while Gonçalves and dos Santos (2026) evaluated a digital-twin architecture on a small controlled testbed and explicitly cautioned that its results may not generalize to large-scale data centers. Chinnici et al. (2024) similarly demonstrated a digital-twin approach using operational data from a single HPC cluster, while Zohdi (2022) evaluated a physics-based digital twin entirely through simulation of a hypothetical data-center pod. The evidence from the building domain provides stronger examples of prediction-control integration, but with similarly narrow validation. Fujii et al. (2021) used a household digital twin with an XGBoost forecasting pipeline and reported a 13.36% improvement in reference-model MSE. Han et al. (2024) combined an IoT-enabled digital-twin platform with several predictive models in a single campus building, with the best model achieving an R^2 of 0.88 and approximately 20–25% measured energy savings. Mahamedi et al. (2025) embedded a Bi-LSTM within a cloud-based digital-twin pipeline for one commercial building, while Agostinelli et al. (2021) applied a BIM/GIS digital twin across a residential district. Almadhor et al. (2025), in contrast, evaluated its digital-twin framework entirely through simulation without physical IoT sensors.

These building-domain studies are not direct evidence for data-center capacity planning, but they reveal an important pattern: digital twins currently provide a stronger demonstrated bridge between prediction and control in buildings than in data centers, while evidence from both domains remains dominated by single-site, small-scale, or simulated validation. Consequently, although digital twins offer a promising integration layer, their ability to support reliable capacity decisions at production data-center scale remains insufficiently established.

4.3 PUE/Cooling Dominance over Capacity-Planning Targets

There is broad agreement that PUE and cooling-system energy are the most tractable and most frequently modeled targets, but this very consensus signals a divergence from the field’s stated capacity-planning ambitions, since almost no study models workload-driven capacity provisioning directly.

This emphasis is evident across studies targeting PUE and cooling efficiency. Lei and Masanet (2020) model location-specific PUE and its improvement potential, while Gao and Jamidar (2014) and Yang et al. (2022) apply machine-learning models to PUE prediction using operational data. Mustaram et al. (2023) focuses on cooling performance within a campus data center, and Zohdi (2022) uses a physics-based digital twin to optimize cooling-system design.

Collectively, these studies demonstrate that energy efficiency within existing infrastructure is a well-developed and measurable application of predictive analytics.

In contrast, direct modeling of future capacity requirements is notably limited. Daradkeh and Agarwal (2022) provides the clearest exception by linking workload prediction with elastic resource scaling and optimization, thereby connecting predictive analytics to resource provisioning decisions. This distinction is also reflected in the earlier survey by Dayarathna et al. (2016), which identified the stronger development of component-level power and cooling models relative to whole-data-center modeling. This imbalance matters for capacity planning. The literature is comparatively mature in predicting and reducing the energy required by existing capacity, but provides much less evidence on predicting how much capacity will be required in the future as workloads change. Thus, PUE and cooling optimization currently dominate predictive energy analytics, while workload-driven capacity provisioning remains an underdeveloped application.

4.4 Single-Facility Validation Norm

Most studies converge on single-facility or simulated validation as sufficient grounds for publishable results, but a small minority explicitly reject this norm, arguing that single-facility validation cannot support generalizable claims.

The predominance of single-facility or simulated validation is evident across the reviewed literature. Zohdi (2022) relies entirely on simulation of a hypothetical data-center pod, while Gonçalves and dos Santos (2026), Mustaram et al. (2023), and Chinnici et al. (2024) evaluate their approaches using a small testbed, single campus facility, and single HPC cluster, respectively. Similar limitations occur in building-domain studies, including Han et al. (2024), Fujii et al. (2021), Mahamedi et al. (2025), and Almadhor et al. (2025), as well as the single-data-center study by Wilson et al. (2024). These settings can demonstrate methodological feasibility but provide limited evidence that model performance will transfer across facilities with different workloads, climates, architectures, and operating conditions.

Two studies provide notable departures from this pattern. Lei and Masanet (2020) explicitly addressed the single-facility limitation by validating their PUE models across 17 hyperscale data centers, enabling assessment across different facilities and climatic conditions. Daradkeh and Agarwal (2022) similarly evaluated their predictive and optimization framework using two independent hyperscale workload traces from Google and Alibaba. These studies provide stronger grounds for assessing generalizability than single-site experiments.

The contrast highlights a central validation weakness in the field. While single-facility and simulated studies are useful for establishing proof of concept, they cannot readily establish whether predictive relationships remain reliable across heterogeneous data-center environments. Multi-facility validation therefore represents an important methodological requirement for moving predictive energy analytics from experimental demonstrations toward robust capacity-planning applications.

4.5 Absence of Calibrated Uncertainty

Nearly every reviewed study treats deterministic point forecasting as an adequate basis for energy-management decisions, a near-universal practice that two studies separately and explicitly flag as insufficient for decisions made under genuine uncertainty.

This deterministic orientation is evident across both predictive and optimization studies. Gonçalves and dos Santos (2026), Yang et al. (2022), Wilson et al. (2024), Mahamedi et al. (2025), Fujii et al. (2021), and Almadhor et al. (2025) report point-based predictive performance, while Zohdi (2022) evaluates optimization using a single objective outcome. The broader review evidence reinforces this pattern. Panwar et al. (2022) identifies the predominance of deterministic approaches across energy-management research, while Ferreira et al. (2026) describes deterministic methods as the dominant approach in the reviewed optimization literature. Lei and Masanet (2020) provides the strongest and most explicit departure from this norm. Rather than reporting only point estimates, the study combines Sobol sensitivity analysis with bootstrapped confidence intervals to produce calibrated prediction intervals for PUE. Its predictions were therefore evaluated not only in terms of their expected values but also their uncertainty, allowing the model to characterize the range within which actual PUE values were likely to fall.

Importantly, Lei and Masanet (2020) is the only study in the entire 22-study corpus that provides calibrated prediction intervals in this manner. This makes the finding particularly significant for capacity planning, where decisions must account for variability in workload, energy demand, and operating conditions rather than rely solely on expected values. The overwhelming reliance on deterministic predictions consequently represents not merely a reporting preference but a substantive limitation in the evidence base for decision-making under uncertainty.

5 DISCUSSION

The findings show that predictive energy analytics in data centers has advanced considerably in its ability to predict and optimize energy use, but this progress has not translated proportionately into comprehensive capacity-planning applications. The literature is concentrated around three methodological streams that are often investigated separately. Cross-stream integration exists, but remains uncommon, while validation is frequently limited to individual facilities or simulated environments. The strongest evidence for generalizability and uncertainty-aware prediction comes from the relatively small number of studies that use multiple facilities and calibrated prediction intervals. Taken together, these patterns suggest that the central challenge is no longer simply improving predictive accuracy. Rather, the field needs to connect prediction, physical system representation, optimization, and uncertainty into decision frameworks that can inform how much capacity should be provisioned as workloads and operating conditions change. The findings therefore point toward a transition from energy-efficiency optimization of existing infrastructure toward integrated, uncertainty-aware predictive capacity planning.

5.1 Comparison with Prior Reviews

The present synthesis extends earlier reviews by bringing together methodological streams that have largely been examined separately. Dayarathna et al. (2016) provided an important taxonomy of data-center energy consumption modeling that incorporated both statistical and machine-learning approaches, but optimization was treated primarily as a downstream application rather than as an integrated methodological stream. Moreover, the review predates the development of the digital-twin literature examined in the present review. This paper therefore extends the foundation established by Dayarathna et al. (2016) by incorporating optimization and digital-twin approaches that have emerged more prominently in subsequent research. Panwar et al. (2022) broadened the methodological classification by organizing energy-management research into heuristic, metaheuristic, machine-learning, and statistical families. However, these families remained largely separate within the review rather than being incorporated into a common framework. The Three-Stream Convergence Framework developed in this paper addresses this gap by providing a unifying structure that connects statistical/physics-based methods, machine-learning/deep-learning approaches, and optimization/digital-twin frameworks according to their data characteristics, methodological roles, integration, validation, and decision applications.

More recently, Ferreira et al. (2026) moved toward greater integration by examining the relationship between machine-learning forecasting and optimization, explicitly characterizing the literature as fragmented. However, their synthesis does not incorporate the statistical/physics-based stream examined here. The present review therefore confirms their diagnosis of fragmentation while extending it across all three methodological streams. This broader framing is important because statistical and physics-based models provide physical interpretability and system constraints, machine learning provides flexible prediction, and optimization and digital twins provide pathways toward operational decision-making. Considering these approaches together reveals that the principal gap is not the absence of individual methods, but the limited convergence of complementary methods into validated capacity-planning frameworks.

5.2 Theoretical and Practical Implications

Theoretically, the Three-Stream Convergence Framework provides a shared vocabulary for positioning new research within the broader predictive energy analytics literature. Rather than evaluating each study primarily against methods within its own methodological stream, the framework enables researchers to consider how statistical/physics-based, machine-learning/deep-learning, and optimization/digital-twin approaches relate to one another and where integration may be possible. This cross-stream perspective can support more coherent development of the field and make methodological gaps more visible.

Practically, data-center operators seeking to adopt predictive analytics currently face a fragmented evidence base, requiring them to navigate largely separate literatures on modeling, machine learning, and optimization. The five themes identified in this study provide a practical checklist for evaluating proposed approaches before deployment. In particular, operators should consider validation scope, uncertainty treatment, integration across methodological streams, the specific decision target, and the extent to which evidence comes from real operational infrastructure rather than simulation. This can help distinguish methods that demonstrate technical performance in narrow settings from those with stronger potential for reliable capacity-planning decisions.

5.3 Strengths and Limitations of This Review

A major strength of this review is its cross-stream coverage. The study spans 2014–2026 and includes research across statistical/physics-based modeling, machine learning, and optimization/digital-twin approaches. It also draws from diverse publication venues, including Energy, Renewable Energy, IEEE Communications Surveys & Tutorials, Energies, Scientific Reports, and ACM proceedings. This breadth supports a genuinely cross-stream synthesis rather than one shaped primarily by a single disciplinary or publication venue.

This study also has several limitations. First, the study is limited to 22 studies, which constrains the breadth of evidence available for some themes, particularly data-center-specific digital-twin applications. Second, five digital-twin studies, Mahamedi et al. (2025), Agostinelli et al. (2021), Han et al. (2024), Fujii et al. (2021), and Almadhor et al. (2025), are from the building-energy domain rather than data centers. They were included as the closest available methodological analogs. However, their inclusion may make digital-twin integration appear more mature than it actually is within data-center research specifically. Finally, the search strategy and database selection used in this review may also limit the completeness of the study. These limitations should be considered when interpreting the generalizability of the thematic conclusions.

5.4 Future Research Directions

The findings of this review point to five closely related directions for future research. First, future studies should deliberately integrate statistical/physics-based, machine-learning/deep-learning, and optimization/digital-twin approaches within a single data-center framework. Daradkeh and Agarwal (2022) provides a precedent for such three-stream integration, demonstrating that these approaches can function as complementary components rather than isolated alternatives. Second, multi-facility validation should become a stronger norm. Future models should be evaluated across data centers with different operating and environmental conditions rather than relying primarily on single-facility or simulated validation. The multi-facility approach demonstrated by Lei and Masanet (2020) provides a useful precedent for establishing whether predictive relationships generalize beyond individual facilities. Third, calibrated uncertainty quantification should be incorporated into ML/DL forecasting pipelines used for capacity decisions. Rather than providing only point forecasts, future models should characterize the uncertainty surrounding predicted energy demand and related outcomes. Lei and Masanet (2020), the only study in the reviewed study to provide calibrated prediction intervals, demonstrates the feasibility of this approach. Fourth, future research should place greater emphasis on capacity-provisioning-specific prediction targets, rather than concentrating predominantly on PUE and cooling optimization. Workload-driven forecasting should be connected directly to decisions concerning resource provisioning and future capacity requirements, following the precedent established by Daradkeh and Agarwal (2022) and the higher-level modeling need identified by Dayarathna et al. (2016). Finally, data-center-specific digital-twin validation at production scale is needed. Although building-domain studies demonstrate increasingly sophisticated digital-twin and ML integration, data-center applications remain largely limited to small-scale, single-facility, or simulated settings. Future research should therefore test digital-twin frameworks on production-scale data-center infrastructure to establish whether the integration demonstrated in smaller systems can translate reliably to real operational environments.

6 CONCLUSION

Predictive energy analytics has become increasingly important for improving the efficiency and management of data-center infrastructure. However, the literature remains fragmented across statistical/physics-based models, machine-learning/deep-learning approaches, and optimization/digital-twin frameworks. Although these streams have produced substantial advances in energy prediction, PUE and cooling optimization, workload forecasting, and operational management, limited research has brought them together within a unified framework specifically oriented toward data-center capacity planning. This review addressed this gap by asking: How do statistical, machine-learning, and digital-twin/optimization approaches to predictive energy analytics compare for data center capacity planning? The Three-Stream Convergence Framework developed in this review provided the basis for comparing these methodological approaches across their data characteristics, methodological roles, integration, validation, and decision applications. The five-theme synthesis showed that methodological fragmentation remains a central characteristic of the field, with relatively few studies integrating complementary approaches. Digital twins provide a promising mechanism for connecting prediction with real-time control, but their validation remains predominantly limited to small-scale, single-facility, building-domain, or simulated environments. The synthesis also showed that predictive energy analytics is substantially more developed for PUE and cooling optimization than for workload-driven capacity provisioning, leaving future infrastructure requirements comparatively underexplored. Validation and uncertainty emerged as further limitations. Most studies rely on single-facility or simulated settings, whereas multi-facility validation remains uncommon. Similarly, deterministic point forecasts dominate the literature, with calibrated uncertainty treatment exceptionally rare. These limitations are particularly important when predictive analytics is intended to support capacity decisions, where models must remain reliable across different facilities and account for uncertainty in future energy and workload conditions. Overall, the evidence indicates that the field's next stage of development should move beyond isolated improvements in prediction or energy efficiency toward integrated, multi-facility, uncertainty-aware approaches that directly connect energy forecasting to capacity-provisioning decisions. Such development would strengthen the practical relevance of predictive energy analytics and address the fragmentation identified across the existing literature.

REFERENCES

- [1] Agostinelli, S., Cumo, F., Guidi, G., & Tomazzoli, C. (2021). Cyber-Physical Systems Improving Building Energy Management: Digital Twin and Artificial Intelligence. *Energies*, 14(8), 2338. <https://doi.org/10.3390/en14082338>
- [2] Alao, G. U., Nii-Okai, E., A. Okpoko, O., Enobakhare, B. O., Sulaimon Tomotele, A., S. Alonge, O., O. Oladunni, A., Adebayo, A., & A. Hilliman, M. (2026). Renewable Energy Integration with Energy Storage and Microgrids: Control Strategies, Polymer-Enabled Technologies, and Future Grid Trends. *International Journal of Environment and Climate Change*, 16(1), 75–91. <https://doi.org/10.9734/ijecc/2026/v16i15217>
- [3] Almadhor, A., Alsubai, S., Kryvinska, N., Ghazouani, N., Bouallegue, B., Al Hejaili, A., & Sampedro, G. A. (2025). A synergistic approach using digital twins and statistical machine learning for intelligent residential energy modelling. *Scientific Reports*, 15(1), 26088. <https://doi.org/10.1038/s41598-025-09760-y>
- [4] Chinnici, M., De Chiara, D., Antonini, M., Acampora, L., Guarnieri, G., Santomauro, G., Genovese, F., Ponti, G., & Telesca, L. (2024). Digital Twin to Modeling Data Center: An enabler for holistic approach. *Proceedings of the 11th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*, 16–24. <https://doi.org/10.1145/3696593.3696636>
- [5] Daradkeh, T., & Agarwal, A. (2022). Cloud Workload and Data Center Analytical Modeling and Optimization Using Deep Machine Learning. *Network*, 2(4), 643–669. <https://doi.org/10.3390/network2040037>
- [6] Dayarathna, M., Wen, Y., & Fan, R. (2016). Data Center Energy Consumption Modeling: A Survey. *IEEE Communications Surveys & Tutorials*, 18(1), 732–794. <https://doi.org/10.1109/COMST.2015.2481183>
- [7] Ferreira, M. S., Bispo, R., Correia, I., & Browne, D. (2026). Integrating machine learning and optimization methods for green Data Center energy management. *Renewable Energy Focus*, 58, 100867. <https://doi.org/10.1016/j.ref.2026.100867>
- [8] Fujii, T. Y., Hayashi, V. T., Arakaki, R., Ruggiero, W. V., Bulla, R., Hayashi, F. H., & Khalil, K. A. (2021). A Digital Twin Architecture Model Applied with MLOps Techniques to Improve Short-Term Energy Consumption Prediction. *Machines*, 10(1), 23. <https://doi.org/10.3390/machines10010023>
- [9] Gao, J., & Jamidar, R. (2014). Machine learning applications for data center optimization. *Google White Paper*, 21, 1–13.
- [10] Gebreyesus, Y., Dalton, D., Nixon, S., Chiara, D. D., & Chinnici, M. (2023). *Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP)*. <https://doi.org/10.3390/fi15030088>
- [11] [CORRECTED] Gebreyesus, Y., Dalton, D., Nixon, S., Chiara, D. D., & Chinnici, M. (2023). Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP). *Future Internet*, 15(3), 88. <https://doi.org/10.3390/fi15030088>
- [12] Gonçalves, R. H. de S., & Santos, W. M. dos. (2026). *A Scalable Digital Twin Framework for Energy Optimization in Data Centers* (arXiv:2605.05581). arXiv. <https://doi.org/10.48550/arXiv.2605.05581>
- [13] Han, F., Du, F., Jiao, S., & Zou, K. (2024). Predictive Analysis of a Building's Power Consumption Based on Digital Twin Platforms. *Energies*, 17(15), 3692. <https://doi.org/10.3390/en17153692>
- [14] Ighomuaye, E., & Annankra, J. A. (2025). Assessing the Risks and Opportunities of Dual-purpose CO2 Injection for Enhanced Oil Recovery (EOR) and Long-term Storage: Implications for U.S. Energy Security and Economic Growth. *International Journal of Frontline Research in Engineering and Technology*, 4(1), 001–008. <https://doi.org/10.56355/ijfret.2025.4.1.0026>
- [15] Jayaprakash, S., Nagarajan, M. D., Prado, R. P. D., Subramanian, S., & Divakarachari, P. B. (2021). A Systematic Review of Energy Management Strategies for Resource Allocation in the Cloud: Clustering, Optimization and Machine Learning. *Energies*, 14(17), 5322. <https://doi.org/10.3390/en14175322>
- [16] Khan, W., De Chiara, D., Kor, A.-L., & Chinnici, M. (2023). Advanced data analytics modeling for evidence-based data center energy management. *Physica A: Statistical Mechanics and Its Applications*, 624, 128966. <https://doi.org/10.1016/j.physa.2023.128966>
- [17] Lei, N., & Masanet, E. (2020). Statistical analysis for predicting location-specific data center PUE and its improvement potential. *Energy*, 201, 117556. <https://doi.org/10.1016/j.energy.2020.117556>
- [18] Mahamedi, E., Suliman, A., & Wonders, M. (2025). A Cloud-Based Framework for Creating Scalable Machine Learning Models Predicting Building Energy Consumption from Digital Twin Data. *Architecture*, 5(2), 29. <https://doi.org/10.3390/architecture5020029>

- [19] Mustaram, R. F., Gulo, T. S., Leksono, E., & Pradipta, J. (2023). *Audit Energi pada Data Center Kampus untuk Efisiensi Energi Berbasis Digital Twin Digital—Twin Based Energy Audit at Campus' Data Center for Energy Efficiency*. 15.
- [20] [CORRECTED] Mustaram, R. F., Gulo, T. S., Leksono, E., & Pradipta, J. (2023). Audit Energi pada Data Center Kampus untuk Efisiensi Energi Berbasis Digital Twin. *Jurnal Otomasi Kontrol dan Instrumentasi*, 15(1), 63–72. <https://doi.org/10.5614/joki.2023.15.1.6>
- [21] Oketie, W. U., & Aliu, A. (2026). A Systematic Review of Digital Energy Technologies and Data-Driven Systems for U.S. Energy Security and Efficiency. *Global Journal of Engineering and Technology Research*, 02(07). <https://doi.org/10.65150/EP-gjetr/V2E7/2026-09>
- [22] Panwar, S. S., Rauthan, M. M. S., & Barthwal, V. (2022). A systematic review on effective energy utilization management strategies in cloud data centers. *Journal of Cloud Computing*, 11(1), 95. <https://doi.org/10.1186/s13677-022-00368-5>
- [23] Panwar, S. S., Rauthan, M. M. S., Barthwal, V., Mehra, N., & Semwal, A. (2024). Machine learning approaches for efficient energy utilization in cloud data centers. *Procedia Computer Science*, 235, 1782–1792. <https://doi.org/10.1016/j.procs.2024.04.169>
- [24] Sefa-Boateng, Y., Annankra, J. A., & Yakin, Z. (2025). Exploring the potential of Wide Band Gap semiconductors in renewable energy system applications: A critical review. *World Journal of Advanced Engineering Technology and Sciences*, 14(1), 168–176. <https://doi.org/10.30574/wjaets.2025.14.1.0012>
- [25] Shaffi, S. S., Jagadeesh, S., Anantha Babu, S., & Mahesh, C. (2026). Machine learning-based optimisation of data centre energy efficiency for intelligent transitions between renewable energy and the grid. *Renewable Energy*, 265, 125618. <https://doi.org/10.1016/j.renene.2026.125618>
- [26] Wilson, J., Taylor, P., & Thomas, E. (2024). Optimizing Energy Consumption in Data Centers Using Machine Learning-Based Predictive Analytics. *International Journal of Information Engineering and Science*, 1(1), 07–10.
- [27] Yang, Z., Du, J., Lin, Y., Du, Z., Xia, L., Zhao, Q., & Guan, X. (2022). Increasing the energy efficiency of a data center based on machine learning. *Journal of Industrial Ecology*, 26(1), 323–335. <https://doi.org/10.1111/jiec.13155>
- [28] Zohdi, T. I. (2022). A digital-twin and machine-learning framework for precise heat and energy management of data-centers. *Computational Mechanics*, 69(6), 1501–1516. <https://doi.org/10.1007/s00466-022-02152-3>