AI-BASED ENERGY CONSUMPTION FORECASTING AND SUSTAINABLE OPTIMIZATION: A REVIEW OF INTELLIGENT FORECASTING MODELS AND GREEN AI APPROACHES

A. K. Utepbergenova

Assistant, University of Innovation Technologies

ABSTRACT

The rapid growth of urbanization, industrialization, digital technologies, and connected devices has resulted in increasing energy demand and greater complexity in modern energy systems. Accurate energy consumption forecasting has therefore become an important component of efficient energy management, smart-grid operation, renewable-energy integration, and demand-response optimization. Artificial intelligence (AI) and machine learning (ML) have emerged as powerful approaches for modeling the nonlinear and dynamic relationships between energy consumption and influencing factors such as historical demand, weather conditions, seasonal patterns, occupancy, socioeconomic characteristics, and renewable-energy availability. Existing research demonstrates the application of artificial neural networks, support vector machines, decision trees, regression models, convolutional neural networks, recurrent neural networks, LSTM architectures, attention mechanisms, federated learning, and hybrid deep-learning models for energy forecasting.

At the same time, the increasing computational requirements of AI create a paradox: AI can contribute to energy efficiency while its own development and deployment may consume substantial amounts of electricity. Green Artificial Intelligence (Green AI) addresses this problem by emphasizing energy-efficient algorithms, model compression, transfer learning, optimized hardware, edge

computing, renewable-powered data centers, and carbon-aware computation. This review synthesizes these two perspectives and discusses the transition from AI for energy efficiency toward energy-efficient AI. The paper analyzes forecasting methodologies, data sources, applications, challenges, and emerging research directions. Finally, an integrated framework combining accurate energy forecasting with Green AI principles is proposed as a promising direction for sustainable intelligent energy-management systems.

Keywords: Artificial Intelligence, Energy Consumption Forecasting, Machine Learning, Deep Learning, Smart Grid, Green AI, Energy Efficiency, Renewable Energy, Federated Learning, Sustainable Energy Management.

INTRODUCTION

Energy is a fundamental component of economic development, industrial productivity, transportation, infrastructure, and quality of life. Increasing population, urbanization, industrial development, and the expansion of digital technologies have significantly increased the complexity of energy demand. Consequently, energy providers and consumers require reliable methods for predicting future consumption and optimizing energy resources.

Traditional forecasting methods often face difficulties when energy demand exhibits nonlinear, dynamic, and highly variable behavior. AI and ML provide an alternative because they can learn relationships from historical data and incorporate multiple influencing variables. The reviewed literature identifies historical consumption, meteorological conditions, seasonal patterns, socioeconomic indicators, occupancy, and other contextual variables as important inputs for AI-based energy forecasting.

The role of AI in energy management extends beyond forecasting. AI can support short-term load forecasting, demand-response optimization, renewable-

energy prediction, power-system control, energy-market applications, and resource planning.

However, a new challenge has emerged. While AI can reduce energy waste through better prediction and optimization, advanced AI models themselves require significant computational resources. Large-scale deep-learning models depend on GPUs, TPUs, cloud infrastructure, and data centers, resulting in considerable electricity consumption and associated environmental impacts.

This creates an important research question:

How can AI be used to reduce energy consumption while simultaneously reducing the energy required to develop and operate AI systems?

This review addresses this question by connecting two research directions: **AI for energy management** and **Green AI for sustainable computing**.

SCOPE AND RESEARCH PERSPECTIVE

The reviewed research on intelligent energy systems can be organized into two interconnected levels:

Level 1: AI for Energy

At this level, AI is used to predict and optimize the energy system itself. Key operational domain tasks include:

- Electricity consumption forecasting
- Short-term load forecasting
- Residential and building energy prediction
- Renewable-energy forecasting
- Demand-response optimization
- Smart-grid management and energy-resource planning

Various forecasting architectures are deployed here, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Sequence-to-Sequence (Seq2Seq), and hybrid structures.

Level 2: Green AI

- At this level, the objective is to reduce the energy and environmental cost of AI computation itself. Important approaches include:
- Model compression, pruning, quantization, and knowledge distillation
- Efficient neural-network architectures and transfer learning
- Low-power hardware and edge computing
- Renewable-powered data centers and carbon-aware scheduling

The combination of these two perspectives creates a broader holistic concept: **Sustainable AI-Based Energy Management**.

METHODOLOGY AND AI MODELS FOR ENERGY FORECASTING

Traditional Machine Learning

Traditional ML approaches remain important because they can provide relatively simple, interpretable, and computationally efficient forecasting solutions. Decision trees, regression models, Gaussian-process regression, ANN, SVM, and fuzzy models represent key computational-intelligence approaches. ANN, in particular, represents a substantial proportion of energy-forecasting models applied across short- and long-term horizons. Custom linear regression models also demonstrate effectiveness in target industrial energy sectors (e.g., power-generation prediction). The core advantage of traditional ML lies in its low computational complexity, though limited capability in handling highly nonlinear temporal dependencies remains a trade-off.

Deep Learning Architectures

Deep learning handles complex temporal and nonlinear patterns inherent to energy usage data:

- **CNN:** Extracts local patterns and temporal structures from repetitive energy-consumption sequences.
- **LSTM:** Specifically designed to capture sequential time-series dependencies in energy usage over extended time horizons.
- **CNN-LSTM:** Combines feature extraction from spatial/local patterns (CNN) with temporal modeling (LSTM). For instance, a hybrid CNN-LSTM architecture trained on hourly electricity and weather data achieved $R^2 = 0.9244$ on unseen test datasets.
- **Attention Mechanisms:** Dynamically weight critical input sequence frames, capturing transient energy spikes and unexpected operational shifts better than standalone structures.

HYBRID AND DISTRIBUTED AI APPROACHES

Hybrid Models

Modern energy systems increasingly move away from single algorithms toward specialized hybrid models:

- CNN + LSTM / CNN + LSTM + Attention
- Econometric models + ANN (e.g., Canonical Cointegrating Regression combined with ANN for long-term demand)
- RNN + Seq2Seq
- Machine Learning + Internet of Things (IoT) integrations

Federated Learning and Edge AI

Energy data collected by smart grids and smart meters is highly distributed and privacy-sensitive. **Federated Learning (FL)** enables decentralized model training directly on local edge devices, sending only encrypted parameter updates to central servers rather than raw consumption data.

A representative implementation utilizes a lightweight four-layer feedforward neural network trained locally with K-Means clustering for user profiling. This decentralized approach yields forecasting accuracy comparable to centralized LSTM models while dramatically decreasing network communication bandwidth and central computation energy.

DATA SOURCES AND INTEGRATED APPLICATIONS

Input Data Categorization

Accurate energy forecasting relies heavily on input feature diversity:

Data Type	Examples	Role in Forecasting
Historical Energy Data	Hourly/daily electricity consumption	Primary forecasting signal
Calendar Data	Hour, day of week, season, holidays	Captures periodic cyclical behavior
Building Data	Occupancy rates, HVAC/equipment usage	Building-level load prediction
Socioeconomic Data	Population growth, GDP, inflation	Long-term macro demand forecasting
IoT Data	Smart meters, smart plugs, ambient sensors	Real-time demand estimation
Renewable Data	Solar irradiance, wind turbine output	Grid integration and balancing

Grid Data	Real-time load, dynamic pricing, network state	Demand-response and market optimization

Applications Beyond Forecasting

Predictions serve direct decision-making processes across:

1. **Demand Response (DR):** Shifting loads during peak periods to stabilize the grid.
2. **Renewable Energy Integration:** Predicting generation intermittency to balance supply and storage.
3. **Smart Grids & Buildings:** Utilizing models like XGBoost, MLP, CNN, and Seq2Seq for real-time HVAC control, resource allocation, and fault detection.

THE GREEN AI PERSPECTIVE AND TECHNIQUES

The Energy Cost of AI Computation

AI systems consume energy across their entire operational lifecycle:

data preprocessing, model training,

hyperparameter tuning, testing, real-time inference, and infrastructure cooling. Green AI shifts emphasis from pure accuracy toward computational energy efficiency across four systematic levels: Green AI shifts research emphasis from pure accuracy gains to computational efficiency and environmental impact.

LEVELS OF GREEN AI OPTIMIZATION

1. Algorithm Level	**2. Model Reuse Level**	**3. Hardware Level**	**4. Infrastructure Level**
Model Pruning	Transfer Learning	Low-power Chips	Green Data Centers

Quantization	Pretrained Models	SpecializedTPUs/NPUs	Carbon Scheduling
Knowledge Distill	Fine-tuning	Edge Processors	Edge Computing
Lightweight Nets	Domain Adaptation	Energy-aware ASICs	Renewable Power

COMPARATIVE ANALYSIS OF APPROACHES

Approach	Main Purpose	Main Advantage	Main Limitation
Linear Regression	Energy prediction	Simple and computationally light	Cannot capture complex nonlinearities
Decision Trees	Load prediction	High interpretability	Overfitting; needs ensemble techniques
ANN	General forecasting	Handles nonlinear relationships	High training time and compute cost
SVM	Complex non-linear forecasting	High generalization capability	Slow training on large-scale datasets
LSTM	Sequential time-series	Captures long-term temporal trends	High computational and memory cost
CNN-LSTM	Spatial-temporal modeling	Comprehensive feature extraction	Complex architecture design

Attention Mechanisms	Event-driven load trends	Focuses on critical load spikes	Additional computational overhead
Federated Learning	Distributed smart-grids	Preserves data privacy; decentralized	Communication coordination overhead
Edge AI	Localized real-time inference	Low latency; reduced transmission cost	Limited edge hardware resources
Model Compression	Green AI efficiency	Lowers compute footprint	Potential slight drop in accuracy
Transfer Learning	Green AI model training	Drastically reduces training time	Requires suitable domain source models
Renewable-Powered AI	Low-carbon computing	Direct carbon emission reduction	Geographic and weather dependency

CONCLUSION AND FUTURE DIRECTIONS

This review synthesized two complementary fields: **AI for energy management** and **Green AI for computational sustainability**.

While AI models like CNN, LSTM, attention mechanisms, and hybrids provide exceptional precision in predicting energy loads, their compute demands must be balanced. The primary conclusion of this review is that future intelligent energy systems must optimize two interconnected objectives simultaneously: **the efficiency of the energy grid and the energy efficiency of the AI model itself.**

Proposed Framework: Sustainable Intelligent Energy Management

The most promising research direction lies in an integrated framework featuring:

1. **Hybrid Forecasting Core:** Lightweight CNN-LSTM or Attention architectures optimized for precision.
2. **Decentralized Deployment:** Federated Learning coupled with Edge AI to protect user privacy and lower bandwidth usage.
3. **Green AI Optimizations:** Model pruning, quantization, and transfer learning applied prior to deployment.
4. **Carbon-Aware Execution:** Aligning heavy training/inference schedules with real-time renewable energy availability.
5. **Multi-Objective Optimization:** Balancing accuracy, execution cost, compute energy, and carbon footprint.

REFERENCES

1. M. Nagraj, D. S., K. BS, L. R., and S. Patil, "Survey on AI-Powered Energy Consumption Forecasting and Prediction," *International Journal of Research Publication and Reviews*, vol. 6, no. 9, pp. 1164–1167, 2025.
2. E. K. Iriakuma, O. N. Ukela, and A. P. Uche, "Forecasting Energy Consumption with AI: A Review for Sustainable Energy Management," *International Journal of Latest Technology in Engineering, Management & Applied Science*, vol. XIII, no. III, pp. 14–21, 2024.
3. N. Anand, H. Pandey, and H. Gulati, "Green Artificial Intelligence: Reducing Energy Consumption in AI Systems," *International Journal of Creative and Open Research in Engineering and Management*, vol. 2, no. 7, pp. 1–9, 2026.