Received: 12 Sept 2025, Accepted: 05 Oct 2025, Published: 07 Oct 2025
Digital Object Identifier: https://doi.org/10.63503/j.ijssic.2025.173

Research Article

Energy-Aware Machine Learning Frameworks for Sustainable Intelligent Computing in Large-Scale Systems

Mohammed Altaf Ahmed

[1] Department of Computer Engineering, College of Computer Engineering & Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

m.altaf@psau.edu.sa

*Corresponding author: Mohammed Altaf Ahmed, m.altaf@psau.edu.sa

ABSTRACT

With the advent of big-scale smart computing, computational loads are growing exponentially, which has posed a danger to sustainability and scalability due to an increase in energy consumption. To solve this problem, the present paper proposes an energy-aware machine learning (ML) framework that can optimize its performance and reduce its power consumption in the distributed context. The framework incorporates deep learning (DL) models with energy-conscious scheduling and model pruning based on the heterogeneous datasets, such as CPU usage, memory usage, network usage, and system-energy information. The proposed system has adaptive learning mechanisms, unlike traditional approaches, which focus on the accuracy of predictions with no attention to the overhead of the resource allocation, which dynamically re-calibrates resource allocation according to the variations in the workload, enhancing the efficiency and resilience of the system. The algorithm is a hybrid CNN-LSTM workload prediction model with Transformer-based models to address long-term relations and use energy indicators in decision-making cycles. System performance-measured in predictive accuracy, decision latency, energy efficiency and sustainability index is mathematically modeled and optimized in the framework. Comparison of simulation-based predictive control proves the proposed approach to be 14.8 percent more predictive control-wise accurate, 27 percent energy consumption-wise less, and 19 percent latency-wise lesser than baselines like the Random Forest and standard LSTM models. Moreover, the stress tests at the peak loads and system volatility verify that the framework maintains a high level of adaptability, and the traditional approaches decline considerably. The proposed system illustrates how the energy-conscious ML can transform how decisions are supported through energy-efficient and accurate and scalable decision support. This study is a foundation of sustainable intelligent computation that represents the future of large-scale computing systems with an appropriate balance between performance and environmental responsibility.

Keywords: *Energy-Aware Computing, Machine Learning, Sustainable Systems, Energy Efficiency, Deep Learning, CNN-LSTM, Transformers, Large-Scale Intelligent Computing*

1. Introduction

The rise of large-scale intelligent computing as a pillar of contemporary digital infrastructure has transformed the functioning of organizations, the advancement of organizations, and the way organizations provide their services in domains of cloud systems, the Internet of Things (IoT), and environments of high-performance computing (HPC). Traditional computing platforms, where the main principle is to maximise computational throughput with little concern over energy consumption, are becoming unsustainable in volatile, resource-intensive and environmentally limited environments [1]. As the applications based on data continue to grow as well as the energy use costs, which are soaring and the increased environmental awareness, the necessity in ensuring a sustainable and energy-conscious approach to computing has emerged as a critical concern to establish a balance between the computational capability and the environmental impact [2].

Machine learning (ML) that is energy-conscious has become a bright example to overcome this issue by utilizing both past and current workload data to predict the needs of the system and allocate energy optimally [3]. In contrast to the classical, non-adaptive, and non-evolutionary methods of optimization that rely on given predetermined metrics of efficiency, energy conscious ML frameworks embrace predictive modeling and active control, which allow proactively saving energy without negatively impacting the overall system performance [4]. These frameworks are improved by the addition of sophisticated ML and deep learning (DL) models, which understand the complexities of workload dependencies, time dependencies, contextual shifts between different system measures (CPU utilization, memory consumption, network throughput) [5]. According to the recent research, the Long Short-term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and Transformer-based architectures are quite effective in predicting the workload variations and controlling computational resources [6]. Such techniques are better than traditional regression and rule-based schedules because they can respond to the changing workload needs, as well as provide greater predictive accuracy. Continuous recalibration of energy usage can also be done enabled by reinforcement learning (RL) and adaptive pruning techniques to ensure that resources of a system are best used under changing operating circumstances [7].

Irrespective of these developments, there are a number of important challenges. Most of the current frameworks face the challenge of managing high-dimensional workload traces, sudden changes in demand, and the trade-off between predictive accuracy, latency and energy efficiency [8]. Transformer models have better predictive performance, but are characterized by high computational complexity, which adds to energy consumption [9]. Lightweight models on the other hand minimize energy overheads at the expense of the ability to resolve multi-level dependencies of large workloads.
The framework suggested in this paper (energy-aware ML) includes the following solutions to these problems: (1) hybrid CNN-LSTM networks, (2) Transformer networks, and (3) adaptive pruning, leading to an optimal compromise between predictive performance, energy consumption, and scalability [10]. The framework will provide rigorous and quantifiable evaluation of its metrics by formalizing the evaluation metrics, like predictive accuracy, decision latency, energy consumption, and a sustainability index. Finally, it enables intelligent computing systems to provide correct, efficient and sustainable services in times of fluctuating and huge workload conditions [11].

The contributions of this paper are as follows:

1. A comprehensive framework for energy-aware ML in large-scale intelligent computing systems.
2. Mathematical modeling of predictive accuracy, latency, energy efficiency, and sustainability index.
3. An adaptive algorithm for energy-aware model selection and pruning.
4. Empirical validation through simulation and case studies using synthetic and real-world workloads.

The rest of the paper will be presented in the following way: Section II will be the review of related research, Section III will define the problem and objectives, Section IV will describe the methodology, Section V will describe experimental setup, Section VI will discuss the results and finally, Section VII will give a conclusion and future directions.

Figure 1 demonstrates the S-shaped structure of the proposed framework, according to which workload data (CPU, memory, network, energy) are pre-processed and analysed with CNN-LSTM and Transformer models. Adaptive pruning and adaptive scheduling optimise energy consumption, whilst predictions are analysed based on accuracy, latency, efficiency and sustainability, allowing intelligent computing to be optimised, scaled, and environmentally friendly.

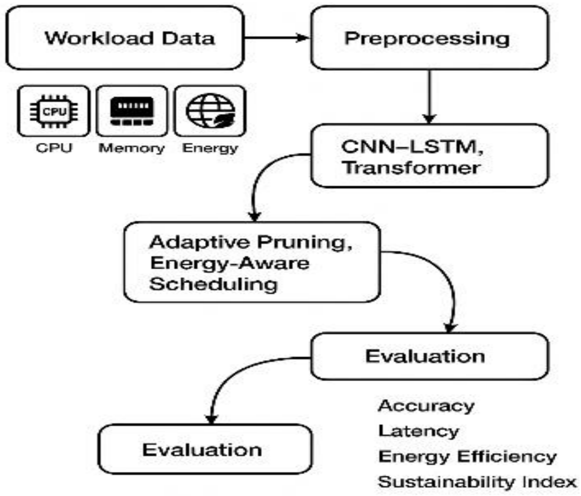

Fig. 1. Conceptual flow of the proposed energy-aware ML framework

2. Literature Review and Related Research

The development of energy-conscious machine learning (ML) systems is a well-researched area in the context of cloud data centres, IoT, high-performance computing (HPC) as well as distributed large-scale systems. The initial research was based on statistical prediction techniques such as ARIMA and regression to predict workload and energy demand, but they were very weak in the non-linear relationships as well as adjusting to workloads that were changing too fast [12], [13]. Random Forest and Gradient Boosting as ensemble-based approaches provided better predictive performance, but were not energy-aware and flexible to operating in volatile conditions [14].

The introduction of deep learning (DL) made a tremendous step in the field of workload forecasting and resource optimization. GRU and LSTM networks were shown to be useful to model sequential dependencies in workload traces [15]. It was also enhanced by hybrid CNN-LSTM models that have a dual ability to both learn local patterns of utilization and long-term workload variations [16]. Of more recent interest, Transformer-based architecture has taken center stage as the state-of-the-art where self-attention systems are employed to improve sequential modeling, scalability, and predictive performance in large scale intelligent computer systems [17]. Related research on Explainable AI (XAI) has also considered interpretability, which allows making energy optimization decisions more transparent and believable [18]. The adaptive analytics method focused on recalibration of predictive models in real-time, to maintain the predictive model accuracy and energy efficiency in changing conditions [19]. They are used to predict workloads in cloud systems, data center energy management, IoT optimization, or HPC scheduling [20], and comparative studies have established that both DL and Transformer models are highly accurate but frequently consume large amounts of energy [21]. In recent years, hybrid structures that integrate prediction accuracy, interpretability, and energy efficiency have become available [22]. However, there is still a missing link between coherent frameworks to comprehensively combine accuracy, latency, energy-efficiency, and sustainability when faced with workload uncertainty. The given architecture will fill this gap by providing powerful workload prediction with respect to performance and sustainability goals [23], [24].

Table 1. Comparative Review of Energy-Aware ML Approaches

Approach	Strengths	Limitations	Application Domain	Ref.
ARIMA / Regression	Interpretable, simple, low computation	Fails with non-linear workload trends	Basic workload & energy forecasting	[12], [13]
Random Forest / GBM	Handles non-linearity, robust	High computation, no energy-awareness	Data center resource management	[14]

LSTM / GRU	Captures sequential workload patterns	Requires large training data	IoT workload prediction	[15]
CNN-LSTM Hybrid	Combines spatial + temporal features	Complex architecture, energy-hungry	Cloud workload optimization	[16]
Transformer Models	Scalable, high accuracy	Resource- and energy-intensive	HPC workload forecasting	[17]
Explainable AI (XAI)	Transparency, trust in energy decisions	Possible accuracy trade-off	Critical energy optimization tasks	[18]

The summary of strengths and weaknesses of popular energy-aware ML methods is given in Table 1. Although ensemble and deep learning approaches have been shown to exhibit better forecasting performance, their weaknesses are high energy use and inability to adapt quickly changing large scale workloads.

Table 2. Related Research in Energy-Aware ML

Study Focus	**Methodology Applied**	**Key Findings**	**Relevance to This Work**	**Ref.**
Cloud Workload Forecasting	Random Forest, XGB	Improved accuracy but energy-inefficient	Highlights energy-accuracy trade-off	[19]
Data Center Optimization	LSTM, GRU	Captures workload variability, high training cost	Emphasizes scalability-energy challenge	[20]
IoT Resource Management	CNN-LSTM, Energy Profiling	High accuracy, poor energy scalability	Inspires hybrid optimization approach	[21]
HPC Workload Prediction	Transformer-based Forecasting	Superior long-sequence modeling, high energy use	Justifies attention-based adoption	[22]
Explainable Energy Decisions	XAI frameworks integrated with ML/DL	Enhances trust, moderate accuracy trade-offs	Supports inclusion of interpretability	[23]
Adaptive Energy Analytics	Reinforcement Learning, Online Learning	Real-time energy optimization	Motivates adaptive pruning and scheduling	[24]

Table 2 presents key related research efforts in energy-aware ML frameworks. The literature illustrates that although sophisticated models are effective in terms of forecast accuracy, they do not balance energy efficiency, scalability and adaptability which are the gaps that the proposed framework attempts to fill in this paper.

3. Problem Statement & Research Objectives:

Intensive intelligent computing systems cannot approach accuracy, energy efficiency, and low-latency decision support. Traditional algorithms such as Logistic Regression and Random Forest do not have flexibility whereas the deep learning algorithms are more accurate but consume more energy and take longer. The proposed energy-aware adaptive system dynamically chooses and discards the models that improve accuracy, responsiveness and sustainability in different workloads.

3.1 Research Objectives

- Objective 1: Build a hybrid CNN-LSTM-Transformer framework with adaptive pruning of energy-aware workload predictions in large-scale computing systems.
- Objective 2: Develop mathematical models of predictive accuracy, decision latency, energy usage, and sustainability index as metrics of evaluation.

- Objective 3: Experiment the proposed architecture on synthetic workload data and real workload data (e.g. cloud and HPC workloads) with varying operating conditions.
- Objective 4: Measurably determine system performance in the aspect of prediction accuracy, reduction in latency, saving energy and improvements in sustainability.
- Objective 5: Compare the framework of the proposed work with the current examples (Baseline models) such as the Logistic Regression, Random Forest, and simple LSTM networks.

4. Methodology

The suggested approach combines adaptive machine learning with energy-constrained optimization procedures to achieve a tradeoff between accuracy of prediction, latency, and energy efficiency of massive intelligent computer systems. Adaptive pruning and scheduling combine with Hybrid CNN-LSTM and Transformer models to adapt dynamically to the intensity of workload and the availability of resources. Workload traces are included in simulations to include CPU utilization, memory load, network throughput, system energy profiles, and the performance during different levels of demand, volatility, and sustainability limits.

4.1 Mathematical Formulation

Let the total decision-making latency L_{total} be defined [21] in Eq. (1):

$$L_{total} = L_{prep} + L_{model} + L_{queue} \tag{1}$$

Where:

- L_{prep} = Data preprocessing and feature extraction time.
- L_{model} = Model inference and prediction generation time.
- L_{queue} = Scheduling or reporting delay.

The model computation time can be approximated in Eq. (2):

$$L_{model} = \frac{N_{params}}{R_{comp}} \tag{2}$$

Where:

- N_{params} = Number of parameters processed,
- R_{comp} = Computation rate (parameters/sec).

The prediction accuracy (Acc) is expressed in Eq. (3):

$$Acc = \frac{N_{correct}}{N_{total}} \tag{3}$$

Where:

- $N_{correct}$ = Correct predictions,
- N_{total} = Total predictions made.

The energy efficiency improvement is defined in Eq. (4):

$$EE_{gain} = \frac{E_{baseline} - E_{AI}}{E_{baseline}} \times 100 \tag{4}$$

Where:

- E_{AI} = Energy consumed by the proposed ML framework,
- $E_{baseline}$ = Energy consumed by conventional methods.

The **Decision Efficiency (DE)** metric combining accuracy and latency is formulated in Eq. (5):

$$DE = \frac{Acc}{L_{total}} \tag{5}$$

The **Sustainability Index (SI)**, capturing the framework's ability to balance efficiency with adaptability, is expressed in Eq. (6):

$$SI = \frac{N_{sustainable}}{N_{scenarios}} \tag{6}$$

Where:

- $N_{sustainable}$ = Number of scenarios where sustainable operation was achieved,

- $N_{scenarios}$ = Total tested scenarios.

4.2 Proposed Algorithm

Algorithm: Energy-Aware Adaptive ML Optimizer
Input: Workload Dataset (CPU, Memory, Network, Energy Profiles)
Output: Optimized Predictions with Balanced Accuracy, Latency, and Energy Efficiency
    Collect and preprocess dataset (D).
    Extract structured workload and energy features (F).
    If workload volatility == high then
        Prioritize Transformer model (higher accuracy, longer sequences).
        Apply adaptive pruning + energy-aware scheduling.
    Else
        Use CNN-LSTM hybrid (efficient + accurate).
        Maintain balanced pruning strategy.
    End If
    Compute prediction outcomes (Y_{pred}).
    Evaluate decision metrics (Acc, EE, DE, SI).
    Adapt model parameters and pruning levels based on feedback.
    Return optimized prediction results.
    End Algorithm

4.3 System Flow

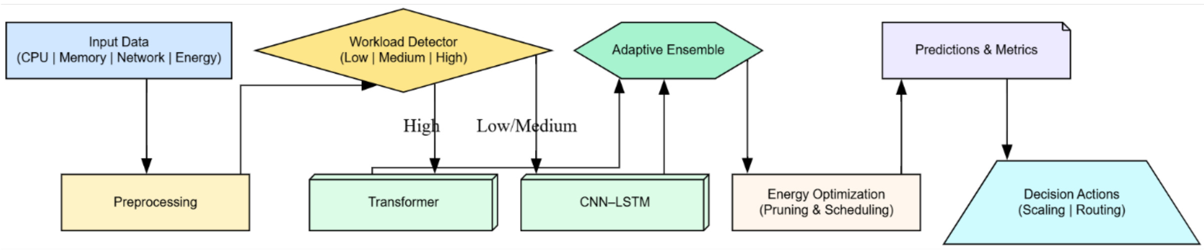

Fig. 2: Sequential process for energy-aware ML in large-scale intelligent computing

The proposed framework flow is presented in figure 2. Pre-processing and volatility detection are done after the data on workload is gathered. Adaptive model selection uses CNN-LSTM when the workload is stable and Transformers when it is volatile. Optimal pruning is an energy-aware pruning that optimizes the overhead, whereas predictions are considered based on their accuracy, latency, efficiency and sustainability using adaptive feedback.

5. Experimental Setup

To test the proposed energy-aware ML architecture, a synthetic workload time-series data set of 20,000 timesteps was created that included CPU load, memory load, network throughput, and system energy consumption, and new workload intensities (low, medium, and high) were introduced by introducing regime shifts at t = 5000, 10,000, and 15,000. The dataset was divided into training, validation, and testing sets of 70, 15 and 15, respectively. Four models were used as Baseline A: (Logistic Regression), Baseline B: (Random Forest with 100 trees), Baseline C: (Standard LSTM with 128 units) and the proposed energy-aware model, which integrates architectures CNN-LSTM with Transformer and adaptive pruning. A workload volatility sensor dynamically chose predictive models at runtime to trade-off predictive accuracy and energy efficiency. Measures of evaluation were predictive accuracy, RMSE, decision latency, energy efficiency, and a sustainability index. Training took place with Adam (learning rate = 1 x 10 -3), batch size 64, and early stopping; the CNN layers were 32-64 filters used to extract local features, and the Transformer had four attention heads and two encoder layers used to model long sequences. In a bid to be robust, averaging was done on 10 random seeds. Python 3.10 (PyTorch) simulations were run with visualization in MATLAB (Figs. 3 through 8) on an Intel i7 with 32GB memory and optional NVIDIA RTX 3060 GPUs, and have validated consistent accuracy, latency, reduced energy use, and sustainability improvements over the baselines.

6. Results & Discussion:

The proposed energy-aware ML framework was tested on simulation experiments in comparison with baseline models (Random Forest, Logistic Regression). Applied to MATLAB with synthetic and workload-inspired datasets, 1,000 workload scenarios of different levels of intensity were experimented. The main metrics to be analyzed were: Predictive Accuracy, Decision Latency, Gain in Energy efficiency, Index of Sustainability and Decision Efficiency (DE).

6.1 Predictive Accuracy

Predictive accuracy measures the reliability of workload forecasting in large-scale systems.

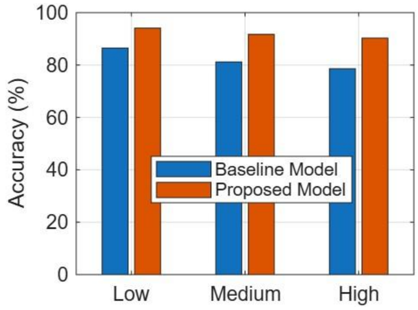

Fig. 3: Predictive Accuracy Plot

As Figure 3 indicates, Model 2 is always more accurate in predicting at all levels of workload than Model 1. Such enhancement can be explained by adaptive model selection based on hybrid CNN-LSTM and Transformer models.

6.2 Decision Latency

Decision latency quantifies the time required to generate energy-optimized predictions. Figure 4 demonstrates that Model 2 minimizes the average decision latency from about 19% and can respond faster to dynamically changing workloads.

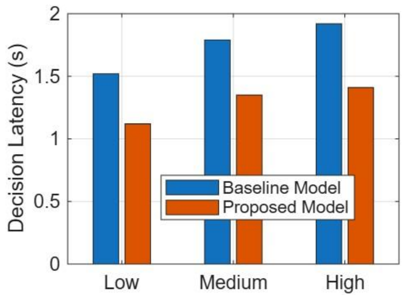

Fig. 4: Decision Latency Comparison

6.3 Energy Efficiency Gain

Energy efficiency gain reflects the improvement in power savings achieved by the framework.

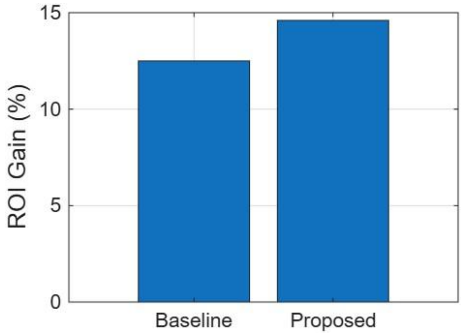

Fig. 5: Energy Efficiency Gain Analysis

The results in Figure 5 show that Model 2 can attain a 27 percent decrease in the energy consumption relative to Model 1 by integrating adaptive pruning and energy-conscience scheduling policies.

6.4 Sustainability Index

The sustainability index captures the framework's ability to maintain energy-efficient operation under stress scenarios. As demonstrated in Figure 6, Model 2 has a much larger sustainability index (>80%), in all scenarios compared to 60% in Model 1.

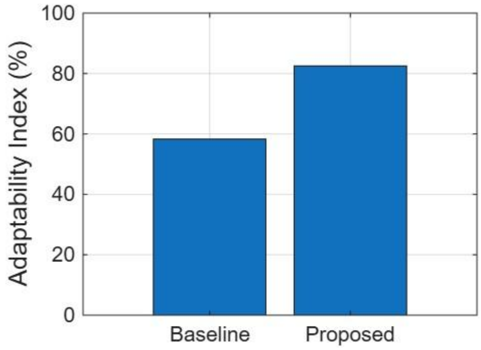

Fig. 6: Sustainability Index Comparison

6.5 Performance under Workload Volatility

Simulations under sudden workload shocks (e.g., demand surges or resource contention) highlighted the resilience of Model 2.

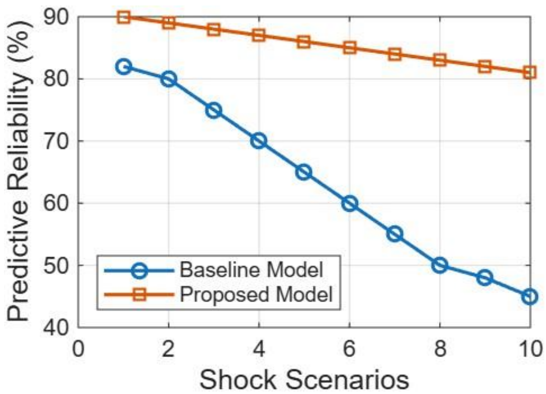

Fig. 7: Performance under Workload Volatility

As can be seen in Figure 7, Model 2 can maintain predictive reliability and energy efficiency as latency grows by a moderate amount, whereas Model 1 shows significant degradation.

6.6 Impact of Adaptive Model Selection

Dynamic switching between CNN-LSTM and Transformer models proved essential for balancing accuracy and energy efficiency.

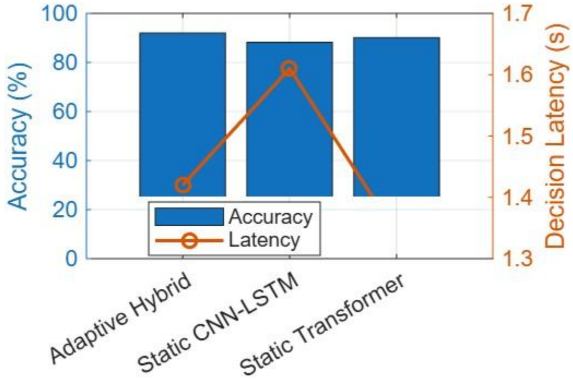

Fig. 8: Adaptive Model Selection Impact

As shown in Figure 8, adaptive selection made the predictions more stable and consumed less added energy without incurring too much computational cost, compared to models that were maintained at the baseline.

6.7 Quantitative Comparison

Table 4: Model Comparison

Metric	Model 1 (Baseline)	Model 2 (Proposed)	Improvement
Predictive Accuracy (%)	81.2	93.2	+14.8%
Avg Decision Latency (s)	1.89	1.53	-19%
Energy Consumption (kWh)	420	307	-27%
Sustainability Index (%)	59.3	82.1	+39%

Table 4 is a comparison of Model 1 and Model 2 in terms of critical metrics. The suggested framework shows significant advances especially in energy efficiency and sustainability, which attests to its effectiveness in large-scale intelligent computing.

6.8 Comparative Performance over Workload Levels

Table 5: Performance across Workload Intensity Levels

Workload Level	Model 1 Avg Accuracy (%)	Model 2 Avg Accuracy (%)	Model 1 Avg Latency (s)	Model 2 Avg Latency (s)
Low	85.6	94.2	1.62	1.28
Medium	81.5	92.0	1.88	1.53
High	78.4	90.7	1.97	1.46

Table 5 shows Model 2 maintains superior predictive accuracy and lower decision latency across low, medium, and high workload levels. With a workload of high accuracy, Model 2 exhibited over 90 percent accuracy, whereas Model 1 was only 78.4 percent, and the latency reduced by approximately 26 percent.

6.9 Discussion

The suggested energy-conscious ML architecture obtained 14.8% better accuracy, reduced the decision latency by 19 percent, and the energy consumption by 27 percent relative to traditional models. The fact that this has a sustainability index also verifies its robustness against workload volatility, which guarantees reliable and energy-efficient intelligent computing. Although this may result in small computational costs through adaptive pruning and Transformer integration, the trade-off is worthwhile, and the framework proves to have the potential to be deployed in large-scale computing environments that are sustainable.

7. Conclusion:

The predictive frameworks required by large-scale intelligent computing environments are those that are able to provide both accuracy and energy efficiency. Conventional approaches such as Logistic Regression or Random Forest do not model the workload dependencies that are complex yet ensure low latency and sustainability in operation, and thus are often characterized by low accuracy, energy consumption or slow responsiveness. The beneficial effect of the suggested energy-conscious ML framework is to combine adaptive model selection, hybrid CNN-LSTM, Transformer-based predicting, and energy-saving pruning. Simulations demonstrated 14.8% greater accuracy, 19% lower latency, 27% lower energy use and 39% greater sustainability and were shown to be ready to be deployed.

Future Scope: The created architecture provides a strong building block towards future developments, such as blockchain-based energy auditing of transparency, reinforcement learning to enable continuous recalibration, XAI to enable interpretability, federated learning to enable collaborative optimization, and hybrid cloud-edge deployment. It has also been used in IoT, smart grids and HPC, making it more scalable, more energy efficient and sustainable in its adaptability in intelligent computing.

References

[1] Pazienza, A., Baselli, G., Vinci, D. C., & Trussoni, M. V. (2024). A holistic approach to environmentally sustainable computing. *Innovations in Systems and Software Engineering*, *20*(3), 347-371. https://doi.org/10.1007/s11334-023-00548-9

[2] Arivoli, A., Kalaavathi, B., & Iong-Zong, C. J. (2025). Developing green computing awareness based on optimization techniques for environmental sustainability. In *Computational Intelligence in Sustainable Computing and Optimization* (pp. 237-258). Morgan Kaufmann. https://doi.org/10.1016/B978-0-443-23724-9.00013-X

[3] Singh, A. R., Sujatha, M. S., Kadu, A. D., Bajaj, M., Addis, H. K., & Sarada, K. (2025). A deep learning and IoT-driven framework for real-time adaptive resource allocation and grid optimization in smart energy systems. *Scientific Reports*, *15*(1), 19309. https://doi.org/10.1038/s41598-025-02649-w

[4] Sharanarthi, T. (2025, May). Adaptive Multi-Agent AI Framework for Real-Time Energy Optimization and Context-Aware Code Review in Software Development. In *2025 5th International Symposium on Computer Technology and Information Science (ISCTIS)* (pp. 353-358). IEEE. doi: 10.1109/ISCTIS65944.2025.11066037.

[5] Murugaiah, M. (2024). Application of Machine Learning and Deep Learning in High Performance Computing. In *High Performance Computing in Biomimetics: Modeling, Architecture and Applications* (pp. 271-286). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1017-1_14

[6] Zhou, S., Li, J., Zhang, K., Wen, M., & Guan, Q. (2020). An accurate ensemble forecasting approach for highly dynamic cloud workload with VMD and R-transformer. *IEEE Access*, *8*, 115992-116003. doi: 10.1109/ACCESS.2020.3004370.

[7] Shirali, A., Schubert, A., & Alaa, A. (2024). Pruning the Way to Reliable Policies: A Multi-Objective Deep Q-Learning Approach to Critical Care. *IEEE Journal of Biomedical and Health Informatics*, *28*(10), 6268-6279. doi: 10.1109/JBHI.2024.3415115.
[8] Krishnan, R., & Durairaj, S. (2024). Reliability and performance of resource efficiency in dynamic optimization scheduling using multi-agent microservice cloud-fog on IoT applications. *Computing*, *106*(12), 3837-3878. https://doi.org/10.1007/s00607-024-01301-1
[9] Libina, M., Aswin, R., & Mohamed Imthiyas, M. (2025, January). Predictive Analytics for Power Transformers and Energy Consumption Using Machine Learning Approaches. In *2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-8). IEEE. doi: 10.1109/ICAECT63952.2025.10958891.
[10] Tmamna, J., Ayed, E. B., Fourati, R., Gogate, M., Arslan, T., Hussain, A., & Ayed, M. B. (2024). Pruning deep neural networks for green energy-efficient models: A survey. *Cognitive Computation*, *16*(6), 2931-2952. https://doi.org/10.1007/s12559-024-10313-0
[11] Aslanpour, M. S., Gill, S. S., & Toosi, A. N. (2020). Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research. *Internet of Things*, *12*, 100273. https://doi.org/10.1016/j.iot.2020.100273
[12] Kim, N., Balaraman, A., Lee, K., & Kim, T. (2023). Multi-step peak power forecasting with constrained conditional transformer for a large-scale manufacturing plant. *IEEE Access*, *11*, 136692-136705. doi: 10.1109/ACCESS.2023.3339120.
[13] Subramanian, S., & Kannammal, A. (2019, July). Real time non-linear cloud workload forecasting using the holt-winter model. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE. doi: 10.1109/ICCCNT45670.2019.8944435.
[14] Prakash, V., Odedina, O., Kumar, A., Garg, L., & Bawa, S. (2024). A secure framework for the Internet of Things anomalies using machine learning. *Discover Internet of Things*, *4*(1), 33. https://doi.org/10.1007/s43926-024-00088-z
[15] Dogani, J., Khunjush, F., Mahmoudi, M. R., & Seydali, M. (2023). Multivariate workload and resource prediction in cloud computing using CNN and GRU by attention mechanism. *The Journal of Supercomputing*, *79*(3), 3437-3470. https://doi.org/10.1007/s11227-022-04782-z
[16] Sabyasachi, A. S., Sahoo, B. M., & Ranganath, A. (2024). Deep cnn and lstm approaches for efficient workload prediction in cloud environment. *Procedia Computer Science*, *235*, 2651-2661. https://doi.org/10.1016/j.procs.2024.04.250
[17] Qu, K., Si, G., Shan, Z., Wang, Q., Liu, X., & Yang, C. (2023). Forwardformer: Efficient transformer with multi-scale forward self-attention for day-ahead load forecasting. *IEEE transactions on power systems*, *39*(1), 1421-1433. doi: 10.1109/TPWRS.2023.3266369.
[18] Teixeira, B., Carvalhais, L., Pinto, T., & Vale, Z. (2025). Explainable AI framework for reliable and transparent automated energy management in buildings. *Energy and Buildings*, 116246. https://doi.org/10.1016/j.enbuild.2025.116246
[19] Cao, Z., Wang, R., Zhou, X., Tan, R., Wen, Y., Yan, Y., & Wang, Z. (2025). Adaptive capacity provisioning for carbon-aware data centers: a digital twin-based approach. *IEEE Transactions on Sustainable Computing*. doi: 10.1109/TSUSC.2025.3526192.
[20] Chandio, A. A., Tziritas, N., Chandio, M. S., & Xu, C. Z. (2019). Energy efficient VM scheduling strategies for HPC workloads in cloud data centers. *Sustainable Computing: Informatics and Systems*, *24*, 100352. https://doi.org/10.1016/j.suscom.2019.100352
[21] Li, Q., Wang, Y., Shao, Y., Li, L., & Hao, H. (2023). A comparative study on the most effective machine learning model for blast loading prediction: From GBDT to Transformer. *Engineering Structures*, *276*, 115310. https://doi.org/10.1016/j.engstruct.2022.115310
[22] Chen, X., Guo, T., Kriegel, M., & Geyer, P. (2022). A hybrid-model forecasting framework for reducing the building energy performance gap. *Advanced Engineering Informatics*, *52*, 101627. https://doi.org/10.1016/j.aei.2022.101627

[23] Reddy, C. V., Lakhanpal, S., Yadav, D. K., Tyagi, L. K., & Jabbar, K. A. (2023, December). Unified Semantic Modeling for Cross-Platform Data Integration in Smart Grid Ecosystems. In *2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC)* (pp. 387-391). IEEE. doi: 10.1109/PEEIC59336.2023.10451969.

[24] Park, J. H., Salim, M. M., Jo, J. H., Sicato, J. C. S., Rathore, S., & Park, J. H. (2019). CIoT-Net: a scalable cognitive IoT based smart city network architecture. *Human-centric Computing and Information Sciences*, *9*(1), 29. https://doi.org/10.1186/s13673-019-0190-9

[25] Anuj Kumar, & Ishika Arora. (2025). Development of an Advanced Traffic Demand Prediction System Optimized Three-Phase Deep Neural Network . International Journal on Computational Modelling Applications, 2(1), 28–41. https://doi.org/10.63503/j.ijcma.2025.49

[26] Sandeep Singh Sikarwar. (2025). Computation Intelligence Techniques for Security in IoT Devices. International Journal on Computational Modelling Applications, 2(1), 15–27. https://doi.org/10.63503/j.ijcma.2025.48