

ORIGINAL RESEARCH

Open access

Integrated Machine Learning and Reinforcement Learning for Optimized Energy Management in Green University Campuses

Helena Costa^{1*}, Rui Martins¹, Bruno Lopes², Ana Ferreira¹, Pedro Sousa²

Abstract

The rising priority placed on energy conservation, together with the adoption of renewable sources, has grown essential to sustainable practices across today's university campuses. This research introduces an intelligent green energy management framework (SGEMS) that integrates machine learning predictions with reinforcement learning (RL) to enhance energy utilization and photovoltaic output on an eco-conscious campus. Leveraging historical records from three campus structures, the study developed a forecasting engine capable of projecting near-term electricity demand and solar yield. When paired with RL, the XGBoost method achieved markedly better accuracy in estimating consumption and production levels than competing algorithms, with root mean square errors (RMSEs) of 14.72, mean absolute errors (MAEs) of 12.00, and mean absolute percentage errors (MAPE) of 2.18%. The work also describes an accessible, browser-based dashboard that enables live energy oversight and supports timely decision-making, empowering facility managers to anticipate deficits and regulate consumption more intelligently. Overall, the presented methodology provides an adaptable blueprint that enables campuses to reduce dependence on utility grids, improve efficiency, and serve as a reference model for subsequent eco-friendly campus programs.

Keywords Reinforcement learning, Smart green energy management, Machine learning, Energy consumption forecasting, Solar energy prediction

*Correspondence:

Helena Costa

helena.costa@up.pt

¹ Department of Intelligent Energy Systems, Faculty of Engineering, University of Porto, Porto, Portugal

² Department of Grid Optimization Technologies, Faculty of Science and Technology, University of Lisbon, Lisbon, Portugal

Introduction

Mounting alarm over global warming, combined with steadily rising electricity requirements, has drawn increasing focus to energy-smart infrastructure, especially at higher-education sites that support a wide array of energy-intensive operations. University grounds require substantial electrical power to support teaching, scientific research, student accommodation, and extracurricular activities. Achieving effective control of this demand while simultaneously curbing carbon emissions makes deployment of a smart green energy management system (SGEMS) exceptionally valuable.

Among available clean-energy options, photovoltaic systems stand out as particularly attractive for shrinking the environmental impact of campus operations [1]. Research reported in Lazariou *et al.* [2], Moura *et al.* [3], and Kourgiozou *et al.* [4] consistently underscores the gains from integrating renewable generation into campus-wide energy strategies, thereby advancing autonomy and ecological responsibility. As a zero-emission, naturally replenished resource, solar electricity offers strong potential for broad-scale campus adoption [4]. Still, its output fluctuates with weather and time of

day, creating difficulties for reliable supply continuity [5] and reinforcing the value of an advanced SGEMS capable of sophisticated energy balancing.

Data-driven algorithms under the machine-learning umbrella have earned widespread recognition for their success in anticipating electricity loads [6] and renewable energy output, drawing on archived measurements, atmospheric variables, and occupant patterns [7, 8]. The ability to forecast future demand—known as electric load forecasting—represents a cornerstone capability of modern energy platforms; it enables forward-looking resource allocation and helps eliminate avoidable waste. Machine-learning-based predictors can deliver precise estimates of upcoming consumption and generation trends, laying the groundwork for forward-thinking energy governance [9].

Complementing conventional predictive models, reinforcement learning (RL) brings powerful advantages to energy optimization by learning adaptive policies that respond fluidly to changing circumstances [10]. Through trial-and-error interaction with the environment, RL controllers can dynamically reconfigure energy assets and adjust demand-side elements in harmony with instantaneous supply–demand conditions [11]. In university settings, such methods have proven adept at automatically orchestrating resource assignment, boosting operational performance while safeguarding uninterrupted supply [12].

This article presents an integrated smart green energy management architecture that combines load prediction and solar generation forecasting via a machine-learning and reinforcement-learning pipeline. Chief goals include accurate demand estimation, maximum exploitation of on-site renewables, and proactive mitigation of supply shortfalls through optimized use of photovoltaic resources. Implementation of this strategy is projected to raise overall efficiency, reduce reliance on the external grid, and ensure dependable, low-impact energy provisioning for daily campus operations.

Related Works and Our Main Contributions

Here, we survey representative earlier efforts that apply sophisticated machine-learning tools to intelligent electricity oversight. We then summarise the primary innovations contributed by the comprehensive campus-oriented green energy management solution developed in the current study.

Accurate short-horizon electric load estimation constitutes a foundational requirement for any high-performing smart power management platform [13]. Earlier investigations assessed an array of techniques—including multi-layer perceptron (MLP), convolution neural network (CNN), long short-term memory (LSTM), gated recurrent unit (GRU), support vector machine regression (SVR), adaptive network-based fuzzy inference system (ANFIS), and various hybrid designs (e.g., LSTM + CNN, CNN + LSTM)—for forecasting electricity market transactions in South Korea [14]. Among them, the one-dimensional CNN variant produced the most accurate results, demonstrating the strength of deep architectures for improving forecast quality in power systems. Separately, researchers in Yazici *et al.* [15] introduced a one-dimensional CNN augmented with video pixel networks (VPNs) to achieve 24-hour-ahead load predictions based on Istanbul's historical consumption records. Another line of work evaluated numerous estimators for monthly demand projection at Dubai International Airport, ultimately identifying the light gradient boosting (LightGBM) ensemble as the top performer [16]. Nevertheless, most of these contributions relied on standalone predictive models, which typically exhibit limited flexibility compared to multi-component frameworks that strategically blend diverse learning paradigms. Key characteristics of the reviewed short-term load-forecasting studies are compiled in Table 1.

Table 1. Details of existing works on short-term electric load prediction.

Research focus	Reference	Methods evaluated	Best performing model and key metrics	Data source
Forecasting electric power transactions	[14]	MLP, CNN, LSTM, GRU, SVR, ANFIS,	The one-dimensional CNN delivered the top results:	Electric power records supplied by the Korea Meteorological

		LSTM + CNN, CNN + LSTM	RMSE = 67.35, MAPE = 3.513%	Administration and Korea Power Exchange
24-hour-ahead short-term electricity load forecasting	[15]	One-dimensional CNN with VPNS, LSTM, GRU	The one-dimensional CNN enhanced by VPNS recorded the lowest average MAPE of 2.21%	Hourly electricity consumption and temperature measurements collected in Istanbul, 2015–2017
Monthly electricity consumption forecasting	[16]	Multiple machine learning algorithms	LightGBM achieved the strongest accuracy with an RMSE of 0.0334	Monthly electricity usage figures for Dubai International Airport, May 2017 – December 2019

On the other hand, a major challenge in power distribution systems involves forecasting potential electricity shortages [17]. To address this issue, researchers in Eskandarpour and Khodaei [18] developed a logistic regression model incorporating a second-order polynomial to predict failures in grid components caused by intense hurricanes. In Jaech et al. [17], a bidirectional recurrent neural network (RNN) with an attention layer was proposed for real-time estimation of outage durations in distribution networks, leveraging past maintenance and repair records. Furthermore, support vector machine (SVM) and random forest regression (RFR) techniques were applied, combined with Monte Carlo simulations, to assess the reliability of power systems [19]. However, these prior efforts largely focus on outage prediction in isolation, without embedding these predictions into broader energy management frameworks, thereby limiting their practical value for holistic energy optimization. A summary of these related studies on electric shortage forecasting appears in **Table 2**.

Table 2. Overview of prior research on electric shortage prediction.

Research topic	Approach used	Highest performance achieved	Data utilized
Forecasting possible outages of power grid components [18]	Logistic regression incorporating a second-order function	The logistic regression approach obtained an F1-score of 0.9027	Artificial dataset consisting of 1000 samples generated under a normal distribution
Real-time estimation of outage duration in distribution systems [17]	Recurrent Neural Network (RNN) model integrated with a bidirectional mechanism	The developed RNN model recorded an RMSE of 4.25 and a negative log-likelihood of 2.66	Fifteen years of historical outage records supplied by Seattle City Light
Reliability assessment for improved electric power systems [19]	Support Vector Machine (SVM) and Random Forest (RF) techniques were assessed using the Monte Carlo method	The Random Forest (RF) technique produced result variations within 3% of the reference value	24-node IEEE RTS-96 system

Despite the two critical challenges outlined above, renewable energy sources such as solar power provide a sustainable foundation for intelligent power management. Currently, machine learning and deep learning algorithms are widely used to identify patterns and forecast solar energy output [20]. In Mahmud et al. [21], several machine learning techniques—including linear regression (LR), decision tree regression (DTR), SVR, RFR, LSTM, and MLP regression—were tested for near-term solar power forecasting, with RFR achieving the highest performance. In Liu et al. [22], a streamlined LSTM architecture was applied to incorporate solar and meteorological variables and predict solar generation one day in advance. Separately, an LSTM-RNN configuration was constructed for monthly photovoltaic (PV) output estimation [23]. The key aspects of these earlier investigations into solar power forecasting are compiled in **Table 3**.

Table 3. Overview of prior research on solar power prediction.

Study objective	Technique applied	Reported results	Data origin
Short-term solar power forecasting [21]	LR, DTR, SVR, RFR, LSTM, and MLP regression models	The RFR model produced a mean absolute error (MAE) of 0.0098 and an R^2 value of 0.9919	Data gathered from Alice Springs, Australia
Medium-term solar power forecasting [22]	Simplified LSTM model	The suggested LSTM model achieved an RMSE of 0.512	Data obtained from PV devices, including inverters, power meters, and weather station sensors from PV systems located in Chiang Rai, Thailand, and Kaohsiung, Taiwan
Long-term solar power forecasting [23]	LSTM-RNN approach	The LSTM-RNN approach recorded a mean absolute percentage error (MAPE) of 0.805%	Power output records from operational PV plants, estimated solar irradiation for each site, and regional weather data in South Korea

In contrast to the studies mentioned earlier, the core innovations of the present work can be outlined as follows:

- This research introduces a smart green energy management system that combines electricity load demand and solar production, drawing on real measurements gathered from three structures at the Valahia University of Târgoviște (UVTgv) campus in Romania.
- A novel hybrid learning framework is presented, merging a predictive machine learning component with reinforcement learning to support the smart green energy management system.
- An intuitive web-based platform has been developed, allowing users to register connected devices, leverage backend analytics, and obtain predictive insights for effective smart green energy oversight.

The Proposed Smart Green Energy Management System

In this study, we present a unified methodology for near-term forecasting of both power demand and solar output. The structure and components of the proposed smart green energy management system are detailed below.

Data acquisition

For this investigation, solar generation, irradiance, and meteorological information were sourced from the E-LAND project [24]. The primary datasets originate from three buildings located on the UVTgv campus in Romania [25], supplemented by residential community records from Australia [26]. Meteorological observations were additionally obtained via Visual Crossing [27]. Within the framework of the E-LAND initiative [24], three separate solar photovoltaic (PV) installations were deployed across distinct buildings on the UVTgv eco-campus for continuous data capture. All datasets utilized here are verified, publicly accessible, and provided free of charge by the contributing entities.

From the E-LAND project [24], power demand readings were acquired for the three UVTgv campus buildings, denoted as $Pb1(i)$, $Pb2(i)$, and $Pb3(i)$ at each time step i , where $i = 1, 2, \dots, I$. The aggregate consumption across the three buildings at time i , denoted (i) , served as the primary target variable.

$$P(i) \stackrel{\text{def}}{=} P_{b_1}(i) + P_{b_2}(i) + P_{b_3}(i). \quad (1)$$

Likewise, solar output data were collected from the same three UVTgv buildings, represented as $Lb1(i)$, $Lb2(i)$, and $Lb3(i)$ for each time instance i . The total solar generation across the buildings at time i is expressed as (i) .

$$L(i) \stackrel{\text{def}}{=} L_{b_1}(i) + L_{b_2}(i) + L_{b_3}(i). \quad (2)$$

The six selected meteorological variables encompass temperature $T(i)$, humidity $H(i)$, dew point $D(i)$, precipitation $K(i)$, cloud cover $C(i)$, and wind speed $W(i)$ at each time step i , sourced from Visual Crossing [27]. Irradiance measurements from the E-LAND project [24] supplied beam irradiance (i) , diffuse irradiance (i) , and reflected irradiance (i) on tilted surfaces as additional input features for the model.

Feature engineering

Once the relevant features were gathered, initial analysis involved box plots to detect anomalies in the unprocessed datasets. Box plots depicting power usage across the three buildings, along with those for temperature (i) , humidity (i) , and dew point (i) from the weather variables, are presented in Figures 1 and 2. As shown in Figure 1, consumption patterns across all three buildings exhibit right-skewed distributions with prominent outliers. These outliers were cross-checked against historical peak consumption levels and retained, as they correspond to legitimate high-demand events. Comparable box plots were generated for the remaining features to identify and address any anomalies systematically.

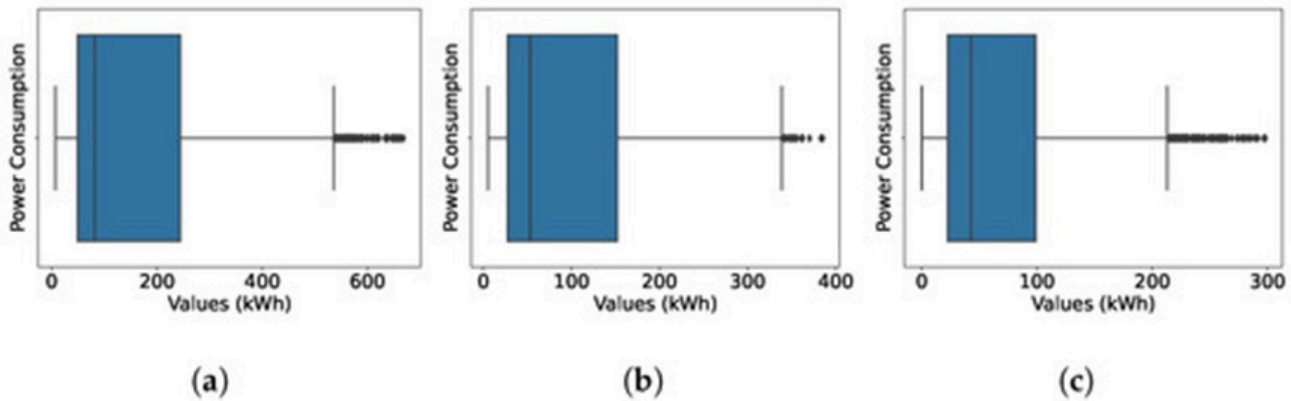


Figure 1. Box plots illustrating power consumption across the three UVTgv buildings: (a) Building 1 power consumption, (b) Building 2 power consumption, and (c) Building 3 power consumption.

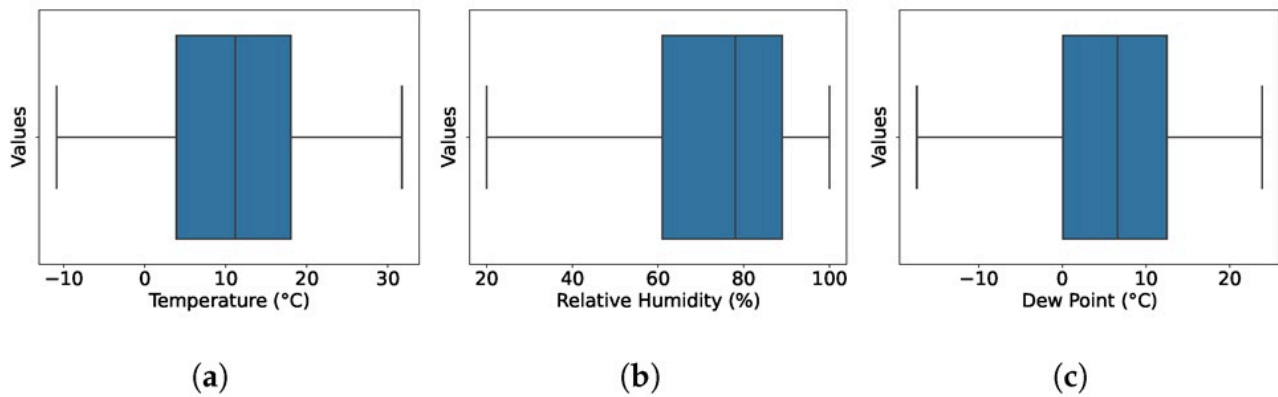


Figure 2. Box plots for selected weather parameters from the collected data: (a) Temperature (*i*), (b) Humidity (*i*), and (c) Dew point (*i*).

Subsequently, the dataset was converted into graphical formats to reveal any latent structures or recurring behaviors. Following the seasonal decomposition approach outlined in Rojo *et al.* [28], the time series was decomposed into three components: trend, seasonality, and residuals. This separation facilitates a deeper examination of the intrinsic dynamics and fluctuations within the series. The trend element illustrates the long-term movement, revealing whether the observations show an overall upward, downward, or steady trajectory over extended periods. The seasonal element captures predictable, repeating oscillations that align with fixed periods—such as daily, weekly, or monthly rhythms—and, here, specifically reflects hourly cycles. Lastly, the residual element (or noise term) accounts for the remaining unexplained variation, incorporating irregular shifts, anomalies, and random disturbances inherent in the raw data.

The decomposed results—including the original series, trend, seasonal, and residual components—for power consumption during January 2019 are displayed in Figure 3. As depicted in Figure 3, the trend in power usage in January 2019 rises throughout the daytime hours and falls at night. Moreover, the seasonal component reveals a pronounced and consistent cyclical pattern embedded in the series. The residual component demonstrates homoscedasticity, indicating that the residual spread remains uniform across the observed timeframe.

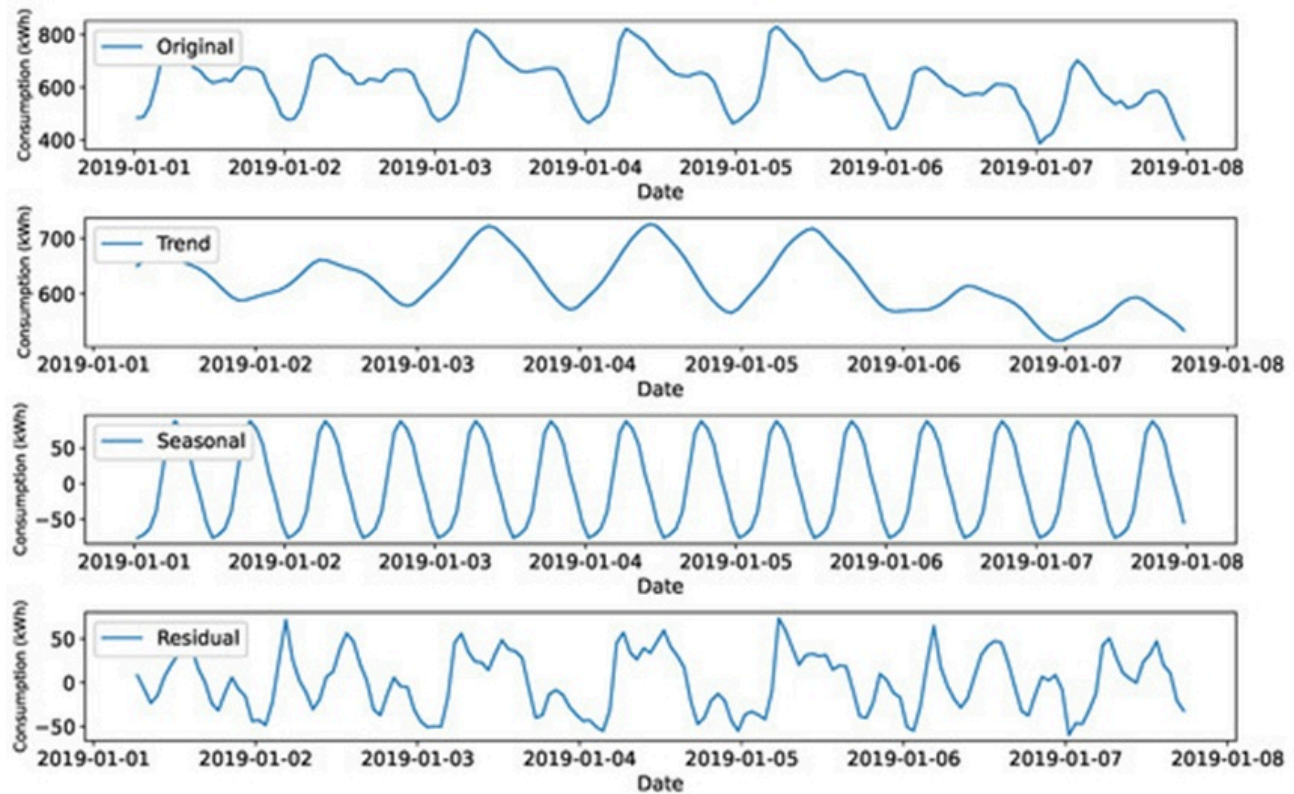


Figure 3. Outcomes of the seasonal decomposition applied to power consumption data for January 2019.

An additional preprocessing step involved reclassifying the numerical cloud cover values, denoted as (i), into discrete categories. The categorization used the following ranges: 0%–20% as clear, 20%–40% as mostly clear, 40%–60% as partly cloudy, 60%–80% as mostly cloudy, and 80%–100% as overcast. During the subsequent modeling stage, these categories will be encoded into binary indicator variables through one-hot encoding.

Machine learning models

The primary objective of this research is to construct machine learning and deep learning architectures capable of accurately forecasting both power demand and solar output. The input features comprise hourly measurements of solar generation, electricity consumption, solar irradiance, and various meteorological variables, all aligned to support the study's forecasting aims. This complete dataset is partitioned into separate training and evaluation subsets. The training portion is used to fit the models, whereas the testing portion assesses their generalization ability. Given the sequential, time-ordered nature of the data, random partitioning is avoided to prevent lookahead bias; instead, chronological splitting is applied. Hyperparameter optimization was conducted via grid search paired with 5-fold cross-validation on the training data. Once optimal hyperparameters were identified, the definitive model was retrained on the full training set using those tuned settings. Predictions were then generated on the held-out testing set, followed by a thorough performance assessment. The overall structure and processing pipeline of the modeling framework are illustrated in [Figure 4](#). A selection of state-of-the-art algorithms was utilized for this purpose, encompassing LSTM [29], CNN [30], RF regressor [31], SVR [32], and XGBoost [33].

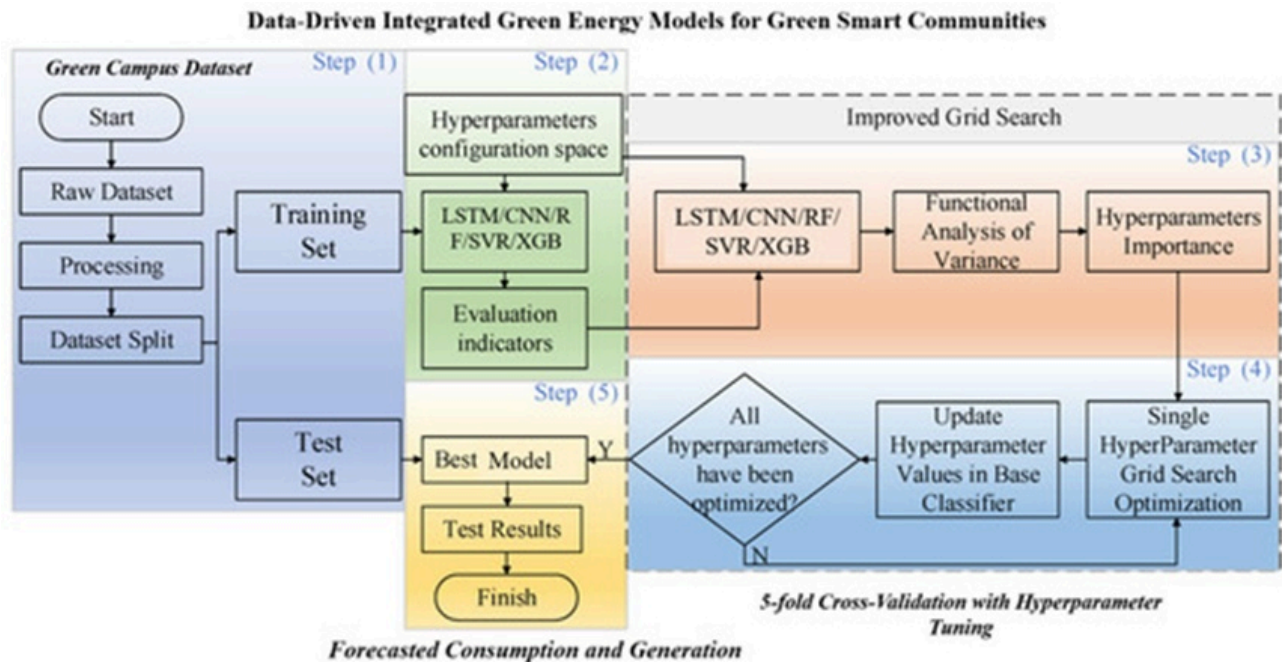


Figure 4. Schematic overview and data processing flow of the machine learning models employed.

LSTM represents a specialized variant of recurrent neural networks designed to process sequential inputs while preserving critical temporal dependencies across extended series [29]. Its particular effectiveness in capturing chronological dependencies renders it especially suitable for applications such as modeling energy usage trends within sustainable energy contexts.

CNN constitutes a powerful neural architecture originally developed for grid-like structured inputs, including imagery and time-ordered sequences [30]. The core principle of CNNs is to extract intricate hierarchical features via successive layers of convolution and pooling operations. Convolution applies learnable filters to the input, computing element-wise multiplications across localized regions of the feature maps. Post-convolution, a nonlinear activation function such as ReLU [34] is typically introduced to enhance the model's expressive capacity. Pooling layers then condense spatial or temporal information by aggregating feature presence within defined windows. After progressing through several convolutional and pooling stages, the architecture concludes with dense, fully connected layers that integrate the learned representations to produce final outputs or predictions.

“Rather than depending on a lone decision tree, the Random Forest regressor constructs an ensemble of numerous trees to forecast continuous outcomes [31]. Every tree within the ensemble is grown on a unique random subsample of the dataset, along with a randomly chosen subset of features at each split. This intentional diversity substantially mitigates overfitting. After all trees are fully trained, their individual predictions are aggregated through simple averaging to produce the final output. These characteristics have established Random Forest as one of the most reliable and frequently used approaches for regression problems in machine learning.

Support Vector Regression (SVR) serves as a powerful tool for modeling nonlinear relationships in datasets [32]. It adapts the Support Vector Machine framework to regression tasks involving continuous targets — for example, estimating future energy generation or demand. The central aim of SVR is to determine a function (represented by a hyperplane in feature space) that keeps as many data points as possible within a specified tolerance margin around the predictions while minimizing deviations beyond that margin.

XGBoost is an advanced, highly optimized, and distributed implementation of gradient-boosted decision trees (GBDT) [33]. Its distinctive advantage lies in the systematic way it corrects previous prediction errors in successive boosting

rounds. At its core, XGBoost prioritizes well-defined objective functions, precise evaluation of model improvements, intelligent use of tree-based learners, and strong regularization mechanisms that discourage overly complex models.

Integrated machine learning model with reinforcement learning

Figure 5 presents the schematic overview of the combined machine learning and reinforcement learning architecture developed for the proposed Smart Grid Energy Management System (SGEMS). Embedding reinforcement learning within the power shortage forecasting pipeline enables the identification of optimal control decisions, thereby markedly improving prediction reliability. Moreover, the system incorporates electricity price forecasting to enable intelligent trading strategies; these price predictions interact closely with the shortage estimates generated by the reinforcement learning module, ultimately providing end-users with more reliable and decision-ready information.

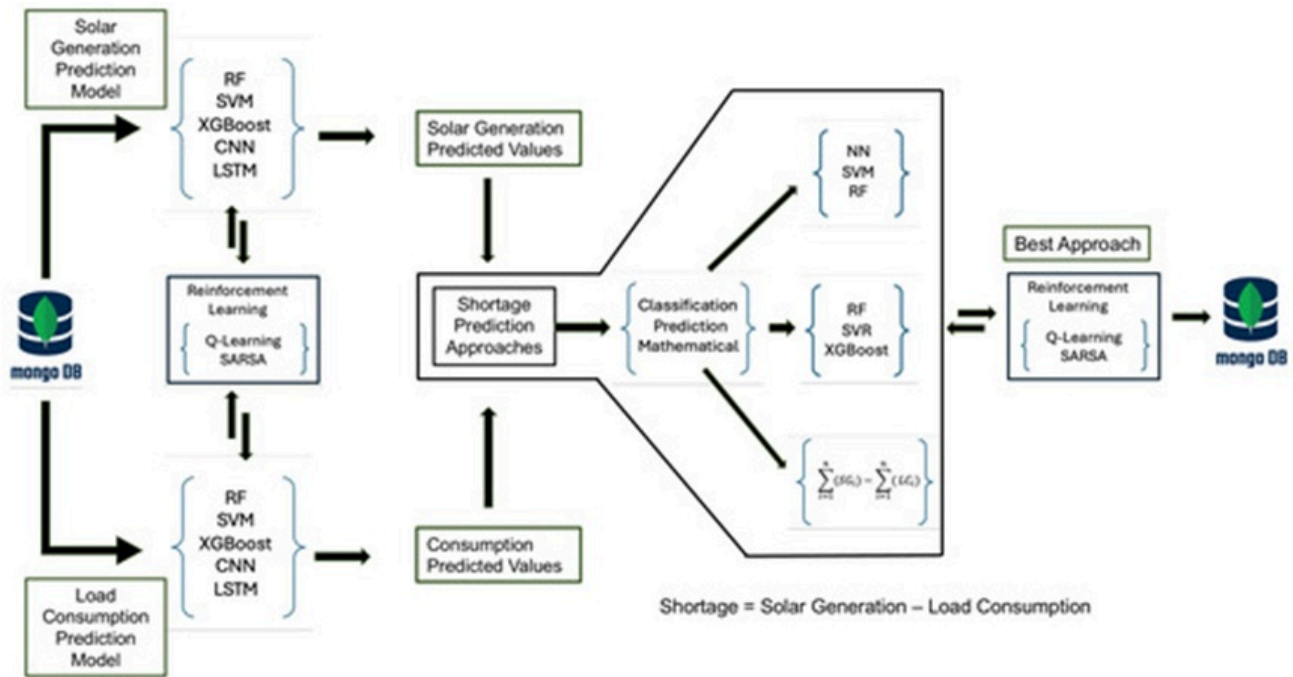


Figure 5. Block diagram illustrating the integrated model.

To further strengthen the proposed framework, the reinforcement learning procedure detailed in Sutton [35] is combined with a previously trained machine learning predictor. Following the formulation in Sutton [35], the reinforcement learning task is structured as a Markov Decision Process (MDP). The state space $\mathcal{S}$ is composed of temporally indexed historical records of electricity consumption and photovoltaic generation, formally defined as:

$$\mathcal{S} \stackrel{\text{def}}{=} \{s_i | i = 1, 2, \dots, I\}, \quad (3)$$

Where every state s_i encapsulates the relevant feature vector of consumption and solar output observed at time index i . The corresponding action space $\mathcal{A}$ contains feasible shortage forecast values and is given by

$$\mathcal{A} \stackrel{\text{def}}{=} \{a_i | i = 1, 2, \dots, I\}, \quad (4)$$

With $a_i \triangleq \hat{y}_i$ signifying the action that corresponds to the machine learning model's forecast of consumption or solar generation at that timestep. Upon execution of action a_i , the environment returns the ground-truth consumption/generation value (y_i) together with a scalar reward $R(i)$ computed as

$$R(i) \stackrel{\text{def}}{=} -|\hat{y}_i - y_i|, \quad (5)$$

Where $\hat{y}_i$ stands for the predicted consumption or generation value at time t . In pursuit of an optimal policy π^* that associates states with the most advantageous actions, the agent implements classical Q-learning and applies the well-known update formula reported in Sutton [35]:

$$Q(s_i, a_i) \leftarrow Q(s_i, a_i) + \alpha \left(R(i+1) + \gamma \max_{a \in A} Q(s_{i+1}, a) - Q(s_i, a_i) \right), \quad (6)$$

In which α denotes the learning rate, γ is the discount factor, $R(i+1)$ is the immediate reward following action a_i , and $\max_{a \in A} Q(s_{i+1}, a)$ gives the best estimated action value in the successor state s_{i+1} . The ultimate objective of the agent is to maximize the expected long-term cumulative reward by converging toward an optimal policy π^* that consistently yields superior prediction accuracy. This optimal policy can be expressed as

$$\pi^*(s) \stackrel{\text{def}}{=} \underset{a}{\operatorname{argmax}} Q(s, a). \quad (7)$$

The overall strategy, therefore, merges an initial high-quality forecast generated by a supervised machine learning model with subsequent policy-based refinement via reinforcement learning, progressively increasing cumulative reward through ongoing trial-and-error interaction with the environment.

Web interface

Once training concluded, the model variant exhibiting the highest accuracy on the held-out test and validation sets was selected for deployment. Forecasts produced by this chosen model are delivered to end users through an accessible, intuitive web-based dashboard. The user-facing component was implemented with ReactJS, while the server-side logic relies on NodeJS together with Flask, as depicted in [Figure 6](#). Upon visiting the platform in any modern browser, users encounter a clean dashboard that summarizes predicted energy production, anticipated consumption levels, and the estimated risk of shortages. Equipped with these projections, consumers can decide whether to sell surplus energy, purchase additional supply, or store electricity for later use.

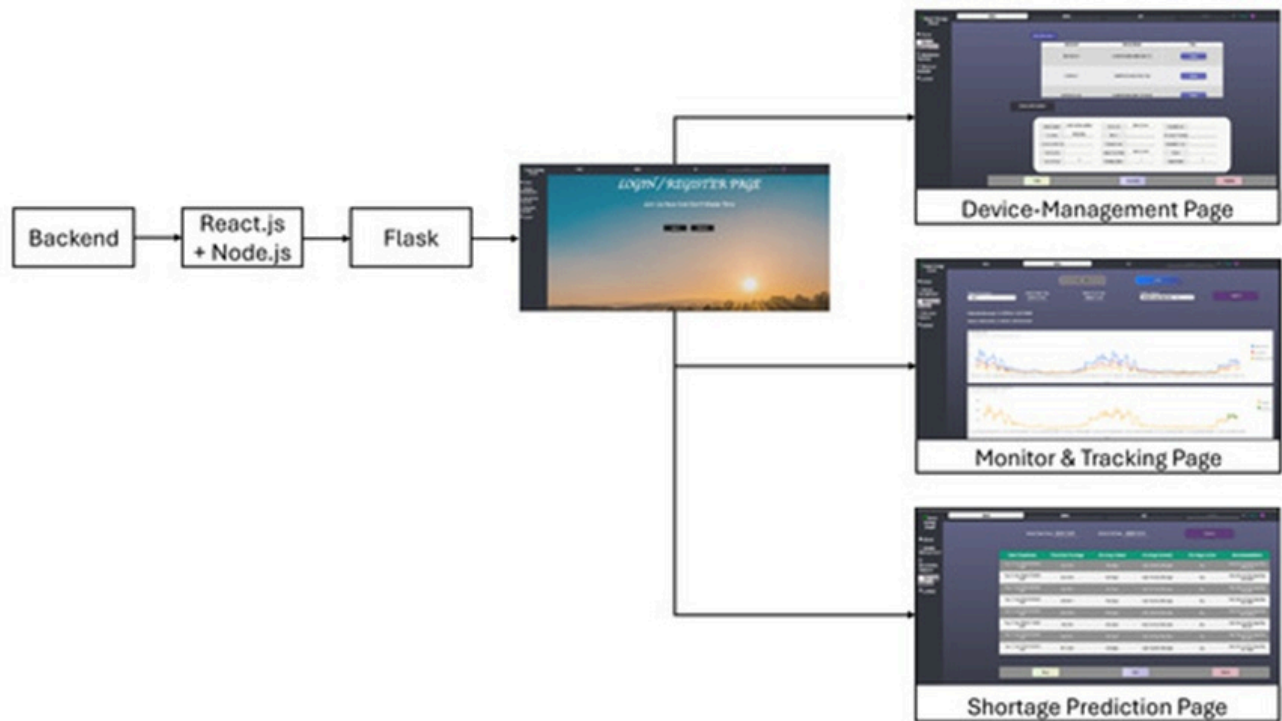


Figure 6. The proposed web interface for the SGEMS.

Simulation Results

For this study, data were gathered from three structures situated on the UVTgv campus [36], as shown in **Figure 7**, positioned in central Romania. The region experiences intense rainfall, strong winds, and pronounced seasonal temperature fluctuations, making it particularly suitable for our investigation given the wide variety of meteorological and solar irradiance variables available.



Figure 7. A photograph of the UVTgv campus [36].

The collected dataset spans 2019 to 2020 and comprises 17,521 data points and 20 input attributes. The training portion was sized adequately to discern broad underlying trends in the data without becoming overly responsive to random fluctuations, thus reducing the likelihood of overfitting. We partitioned the entire dataset roughly as 80% for training, 10% for validation, and 10% for performance testing. This division corresponds to roughly 18 months for training, 3 months for validation, and 3 months for testing. The training subset served to construct both traditional machine learning regressors—including RF, SVR, and XGBoost—and deep learning architectures such as LSTM and CNN. Model effectiveness was assessed using standard error measures: MAE, MAPE, and RMSE, as defined in [Eqs. 8–10](#).

$$\text{MAE} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n |y(i) - \hat{y}(i)|, \quad (8)$$

$$\text{MAPE} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \left| \frac{y(i) - \hat{y}(i)}{y(i)} \right|, \quad (9)$$

$$\text{RMSE} \stackrel{\text{def}}{=} \sqrt{\frac{1}{n} \sum_{i=1}^n (y(i) - \hat{y}(i))^2}, \quad (10)$$

A comprehensive overview of model performance in forecasting power consumption is presented in [Figure 8a](#) and detailed in [Table 4](#). Based on the findings presented in [Table 4](#) and [Figure 8a](#), the XGBoost algorithm emerged as the top performer for short-term energy demand forecasting. The best-performing configuration for XGBoost included a learning rate of 0.05, a maximum tree depth of 5, and an ensemble of 100 trees. This tuned model was fitted to the training data and then assessed on the unseen test portion. On the test set, it achieved an RMSE of 14.657, with a 95% confidence interval of [13.383, 15.932], indicating the typical scale of prediction errors relative to observed consumption.

The MAE was 11.67, with a 95% confidence interval of [11.03, 12.31], capturing average absolute errors, whereas the MAPE was 2.14%, with a 95% confidence interval of [2.02%, 2.26%], providing a percentage-based gauge of precision. In addition, **Table 5** reports the training durations for all consumption forecasting approaches when executed on hardware comprising a 13th-generation Intel i7 processor and an NVIDIA GeForce RTX 4060 GPU. Under these consistent conditions, the XGBoost approach exhibited markedly lower training times than the competing machine learning techniques.

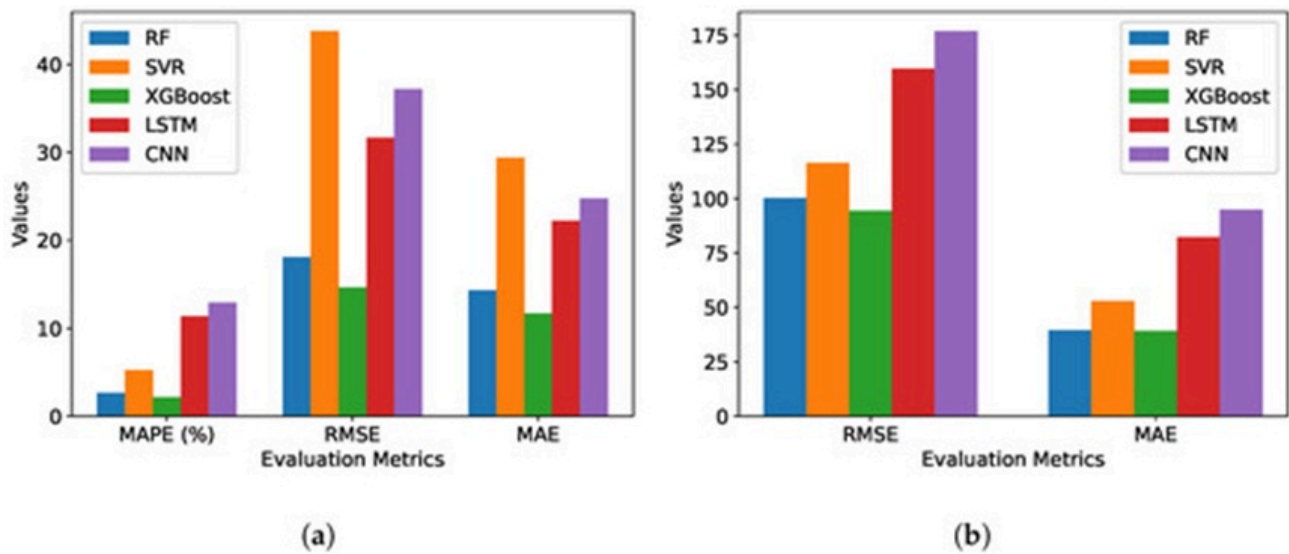


Figure 8. Performance comparison of models for forecasting power consumption and solar generation: (a) Model comparison using three metrics (MAPE, RMSE, and MAE) for power consumption, and (b) Model comparison using two metrics (RMSE and MAE) for solar generation.

Table 4. Performance comparison of power consumption forecasting models.

Model	MAPE	RMSE	MAE
RF	2.65%	18.076	14.33
SVR	5.23%	43.775	29.39
XGBoost	2.14%	14.657	11.67
LSTM	11.39%	31.672	22.20
CNN	12.91%	37.217	24.76

Table 5. Training durations of machine learning models for power consumption forecasting.

Model	Training time (s)
RF	3147.6
SVR	1158.55
XGBoost	19.55
LSTM	86.10
CNN	123.35

Predicting solar output is equally essential for assessing potential power shortages within the community. The comparative analysis and evaluation results for solar generation forecasting are presented in **Table 6** and illustrated in **Figure 8b**. As indicated by **Table 6** and **Figure 8b**, XGBoost achieved the strongest results, recording the smallest RMSE of 92.83 (95% confidence interval [79.19, 106.47]) and an MAE of 38.91 (95% confidence interval [32.83, 45.00]), confirming its superior ability to model solar production dynamics. The RF model performed nearly as well, delivering competitive RMSE and MAE figures that outperformed those of SVR, LSTM, and CNN. In contrast, SVR showed higher RMSE and MAE, suggesting greater vulnerability to short-term solar variability in this green campus setting. Both LSTM and CNN exhibited comparatively larger errors, highlighting difficulties encountered when deploying deep learning methods to this particular forecasting challenge.

Table 6. Performance comparison of solar generation forecasting models.

Model	RMSE	MAE
RF	100.28	39.54
SVR	116.44	52.84
XGBoost	92.83	38.91
LSTM	159.58	82.17
CNN	176.84	94.97

To enable reliable campus-wide predictions of energy demand and supply, we incorporated the high-performing XGBoost model—chosen for its outstanding accuracy—into the reinforcement learning framework. This hybrid system aggregates measurements from all three buildings to estimate overall campus energy consumption and photovoltaic output. Following the creation of this unified model, separate predictive models were developed for each building, providing finer-grained insight into building-specific energy behaviors and revealing distinct consumption and generation characteristics unique to each structure. Outcomes for power consumption forecasting are compiled in **Table 7**. The results in **Table 7** reveal that standalone models for Buildings 1, 2, and 3 yielded MAPE values of 12.42%, 7.26%, and 7.89%, respectively, whereas the integrated campus-level model attained a considerably improved MAPE of 2.18%. The corresponding RMSE and MAE figures corroborate this trend, underscoring greater predictive precision when leveraging pooled data across the entire campus rather than treating each building in isolation.

Table 7. Power consumption forecasting results from individual building models and the proposed integrated campus model.

Model description	Mean absolute percentage error (MAPE)	Root mean square error (RMSE)	Mean absolute error (MAE)
Separate model applied to building 1	12.42%	34.43	32.06
Separate model applied to building 2	7.26%	14.96	12.46
Separate model applied to building 3	7.89%	10.12	8.57
Combined (integrated) model	2.18%	14.72	12.00

Turning to solar generation forecasting, we similarly incorporated the high-performing XGBoost algorithm into the reinforcement learning framework. The outcomes for solar generation prediction are presented in **Table 8**. As shown in **Table 8**, the integrated model—which aggregates solar generation data from the entire campus—achieved an RMSE of 75.45 and an MAE of 33.12. These error metrics are lower than those obtained with the standalone model for Building 2, yet higher than those for Buildings 1 and 3. This observation implies that Building 2 is subject to factors that make its solar output harder to predict, possibly including panel orientation, partial shading, or site-specific microclimatic conditions. By pooling campus-wide data, the integrated approach delivers greater overall stability in forecasts, even if it falls short of the peak accuracy seen in the strongest individual building models. These findings point to the value of a deeper analysis of structural, environmental, or locational factors that impact solar performance at Building 2. At the same time, the more consistent results from the other buildings suggest relatively favorable conditions for photovoltaic generation.

Table 8. Results for solar generation prediction from the individual building models and the proposed integrated campus model.

Model description	Root mean square error (RMSE)	Mean absolute error (MAE)
A separate model was implemented for building 1	53.15	26.32
A separate model was implemented for building 2	140.80	72.58
A separate model was implemented for building 3	63.88	31.53
Unified (integrated) model	75.45	33.12

In conclusion, the markedly better predictive performance of the integrated model in estimating energy consumption underscores the advantage of combining consumption and generation records from all campus buildings to produce more dependable forecasts of overall energy requirements for the green campus. Nevertheless, the varying levels of accuracy in solar generation predictions highlight the critical role of site-specific characteristics in determining solar energy potential. Such observations offer valuable guidance for long-term energy strategy development and the rollout of efficiency-focused measures, potentially establishing a reference point for forthcoming green campus initiatives globally.

Conclusion

This research introduced a smart green energy management system (SGEMS) that integrates machine learning and reinforcement learning to forecast energy consumption and solar generation on a green campus. The findings demonstrate that the hybrid model—built around XGBoost combined with reinforcement learning—substantially enhances the precision of campus-level energy consumption predictions when compared with models trained on individual buildings alone. These improvements provide essential support for effective energy management planning and the adoption of sustainability-oriented practices. Beyond optimizing energy use, the proposed system establishes a valuable precedent for future green campus initiatives by demonstrating the practical benefits of integrating sophisticated learning algorithms into sustainable energy systems. That said, the methodology's performance remains dependent on the quantity and reliability of available historical records and may vary across different campus layouts. Subsequent research should prioritize improvements in computational speed, broadening the dataset to encompass a wider range of environments, and integrating additional variables, such as occupant patterns, to further strengthen both the environmental impact and the system's real-world applicability.

Abbreviations

SGEMS: smart green energy management system

RL: reinforcement learning

MLP: multi-layer perceptron

CNN: convolution neural network

LSTM: long short-term memory

GRU: gated recurrent unit

ANFIS: adaptive network-based fuzzy inference system

VPNs: video pixel networks

LightGBM: light gradient boosting

RNN: recurrent neural network

SVM: support vector machine

RF: random forest

PV: photovoltaic

UVTgv: Valahia University of Târgoviște

GBDT: gradient-boosted decision tree.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 12 Mar 2022

Revised: 09 May 2022

Accepted: 18 Jun 2022

Published online: 20 September 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Genta C, Favaro S, Sonetti G, Barioglio C, Lombardi P. Envisioning green solutions for reducing the ecological footprint of a university campus. *Int J Sustain High Educ*. 2019;20:423–40.
- Lazaroiu AC, Panait C, Seritan G, Popescu CL, Roscia M. Maximizing renewable energy and storage integration in university campuses. *Renew Energy*. 2024;230:120871.
- Moura P, Moreno JI, López López G, Alvarez-Campana M. IoT platform for energy sustainability in university campuses. *Sensors*. 2021;21:357.
- Kourgiouzou V, Commin A, Dowson M, Rovas D, Mumovic D. Scalable pathways to net zero carbon in the UK higher education sector: a systematic review of smart energy systems in university campuses. *Renew Sustain Energy Rev*. 2021;147:111234.
- Soomar AM, Hakeem A, Messaoudi M, Musznicki P, Iqbal A, Czapp S. Solar photovoltaic energy optimization and challenges. *Front Energy Res*. 2022;10:879985.
- García-Martín E, Rodrigues CF, Riley G, Grahn H. Estimation of energy consumption in machine learning. *J Parallel Distrib Comput*. 2019;134:75–88.
- Ntsaluba S, Nwulu N, Ntsaluba K. Deep learning based building load prediction for residences of an academic institutions. In: *Proceedings of the International Conference on Electrical, Computer and Energy Technologies (ICECET)*. Cape Town, South Africa; 2021:1–8.
- Akhtar S, Sujod MZB, Rizvi SSH. An intelligent data-driven approach for electrical energy load management using machine learning algorithms. *Energies*. 2022;15:5742.
- Shahcheraghian A, Ilinca A. Advanced machine learning techniques for energy consumption analysis and optimization at UBC campus: correlations with meteorological variables. *Energies*. 2024;17:4714.
- Mason K, Grijalva S. A review of reinforcement learning for autonomous building energy management. *Comput Electr Eng*. 2019;78:300–12.
- Yu L, Qin S, Zhang M, Shen C, Jiang T, Guan X. A review of deep reinforcement learning for smart building energy management. *IEEE Internet Things J*. 2021;8:12046–63.
- Shaqour A, Hagishima A. Systematic review on deep reinforcement learning-based energy management for different building types. *Energies*. 2022;15:8663.
- Khan ZA, Ullah A, Haq IU, Hamdy M, Mauro GM, Muhammad K, Hijji M, Baik SW. Efficient short-term electricity load forecasting for effective energy management. *Sustain Energy Technol Assess*. 2022;53:102337.
- Bak G, Bae Y. Predicting the amount of electric power transaction using deep learning methods. *Energies*. 2020;13:6649.
- Yazici I, Beyca OF, Delen D. Deep-learning-based short-term electricity load forecasting: a real case application. *Eng Appl Artif Intell*. 2022;109:104645.

Hosny M, Waraga OA, Talib MA, Abdallah M. Simulation of electricity consumption data using multiple artificial intelligence models and cross validation techniques. *Data Brief*. 2023;51:109718.

Jaech A, Zhang B, Ostendorf M, Kirschen DS. Real-time prediction of the duration of distribution system outages. *IEEE Trans Power Syst*. 2019;34:773–81.

Eskandarpour R, Khodaei A. Machine learning based power grid outage prediction in response to extreme events. *IEEE Trans Power Syst*. 2017;32:3315–6.

Boyarkin DA, Krupenev DS, Iakubovsky DV. Prediction of the power shortage in the electric power system by means of regression analysis by machine learning methods. *E3S Web of Conferences*. 2019;114:03003.

Mehta Y, Xu R, Lim B, Wu J, Gao J. A review for green energy machine learning and AI services. *Energies*. 2023;16:5718.

Mahmud K, Azam S, Karim A, Zobaed S, Shanmugam B, Mathur D. Machine learning based PV power generation forecasting in Alice Springs. *IEEE Access*. 2021;9:46117–28.

Liu CH, Gu JC, Yang MT. A simplified LSTM neural networks for one day-ahead solar power forecasting. *IEEE Access*. 2021;9:17174–95.

Jung Y, Jung J, Kim B, Han S. Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea. *J Clean Prod*. 2020;250:119476.

E-LAND. E-LAND Horizon 2020. <https://elandh2020.eu/> (accessed on 13 November 2023).

Zenodo. University Campus Buildings Electrical and Thermal Demand and Generation. <https://zenodo.org/records/7371390> (accessed on 13 November 2023).

Australian Government Department of Climate Change, Energy, the Environment and Water. Smart-Grid Smart-City Customer Trial Data. <https://data.gov.au/data/dataset/smart-grid-smart-city-customer-trial-data> (accessed on 13 November 2023).

V C. Corporation. Weather Data and Weather API | Visual Crossing. <https://www.visualcrossing.com/> (accessed on 13 November 2023).

Rojo J, Rivero R, Romero-Morte J, Fernández-González F, Pérez-Badia R. Modeling pollen time series using seasonal-trend decomposition procedure based on LOESS smoothing. *Int J Biometeorol*. 2017;61:335–48.

Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735–80.

Venkatesan R, Li B. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. Boca Raton (FL): CRC Press; 2017.

Ho, T.K. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, Canada. 1995(1):278–82.

Cortes C, Vapnik V. Support-vector networks. *Mach. Learn*. 1995(20):273–97.

Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; San Francisco, CA, USA. 2016.p. 785–94.

Agarap A.F. Deep learning using rectified linear units. *arXiv* 2018, arXiv:1803.08375.

Sutton RS. Reinforcement learning: An introduction. In *A Bradford Book*; MIT Press: Cambridge, MA, USA, 2018.

Valahia University of Târgoviște. <https://www.masterstudies.com/institutions/valahia-university-of-targoviste> (accessed on 13 November 2023).