Multi-Horizon Energy Consumption Forecasting Using Time Series Analysis and Machine Learning

Farha Hussaini
S20106851
fayhussaini@effat.edu.sa

Dalia Guail
S21207452
daaguail@effat.edu.sa

Rahaf Banat
S21107380
raobanat@effat.edu.sa

Abstract

Universities globally are increasingly challenged with efficient management of energy and water usage while achieving sustainability goals. This research exploits the UNICON dataset—a thorough account of electricity, gas, and water use on five university campuses (2018–2021)—to examine consumption habits, assess energy conservation measures (ECMs), and create forecasting models for better resource allocation. With time-series analysis and machine learning (SARIMA, XGBoost, and LSTM networks), we measure the effect of COVID-19 on utility demand, evaluate the performance of ECMs such as HVAC retrofits, and predict multi-horizon consumption with RMSE gains of up to 15

These findings offer universities a data-driven approach to prioritize cost-saving measures, align procurement with anticipated demand, and promote net-zero objectives. The research also emphasizes essential limitations, such as granularity constraints in the data and challenges in model interpretability, recommending avenues for future research. Through the integration of analytics with campus operations, this work illustrates how AI can revolutionize utility management in large academic institutions.

1 Introduction

Accurate forecasting of energy use is required to maximize power generation, enhance grid reliability, and allow for smarter city infrastructure in an age where data and digitalization are increasingly driving more and more of the economy. Forecasting of energy consumption over short-, medium-, and long-term time horizons has become a key obligation as cities make the move to smart systems. For the energy sector, the forecasts guide pricing policy, operations planning, and policymaking. Based on their interpretability and the strength in identifying linear trends, traditional statistical approaches like Autoregressive Integrated Moving Average (ARIMA) have long been used within this field [1]. More advanced models, such as LSTM networks and XGBoost, have been shown to approximate the nonlinear, seasonal, and dynamic behavior of energy consumption, though, as deep learning and ensemble-based techniques have been used more extensively [2][3]. Even in light of these developments, providing real-world data—frequently plagued by outliers, missing values, and nonuniform spacing between samples—remains problematic [4]. In this paper, we introduce a hybrid and holistic method for predicting energy consumption from the UNICON Energy Dataset [5], an open public dataset with extensive power consumption information. We train and compare ARIMA, LSTM, and XGBoost models on various forecasting horizons after resolving major preprocessing issues like data imputation, anomaly detection, and temporal normalization. We investigate a unified framework for multi-horizon forecasting—with a view to seeking patterns and regularities over a large number of time scales—versus most previous research oriented towards single-horizon forecasts. The specific research problem we address is: How can energy consumption be accurately forecasted across short, medium, and long-term intervals using a combination of statistical and machine learning models, while effectively handling data quality issues.

2 Prior Literature

Accurate prediction of energy consumption in smart buildings and cities plays a key role in the search for enhancing energy efficiency, demand side management, and sustainability. Buildings account for a substantial portion of the global energy consumption, and the improvement of forecasting models benefits directly from reduced emissions, operational saving, and effective resource planning [6]. The last two decades have seen a very diverse range of forecasting methods evolve, generally classified into three: physical models, data-driven models, and hybrids. Physical models, or "white-box" models, depend on thermodynamics to simulate energy transfer in buildings. They are computationally demanding and need accurate building details, but are powerful in controlled conditions. They cannot handle uncertainties created by user activities and environmental effects [6]. Data-driven models have garnered significant interest depending on the availability of smart meter data and advances in machine learning. These "black-box" models include statistical techniques like ARIMA and machine learning techniques like Support Vector Regression (SVR), Artificial Neural Networks (ANNs), and Long Short-Term Memory (LSTM) networks. These models do not require accurate physical descriptions and rather learn from past data patterns. They have been found to provide improved accuracy, especially in short- and medium-term predictions [2][6]. Hybrid approaches are a blend of data-driven and physical approaches and give a compromise between interpretability and performance. Such models, typically known as "grey-box," take advantage of the strengths of both the approaches to improve performance, particularly in systems where data is limited or highly variable [6]. A variety of factors significantly contribute to the performance of forecasting, including input variables, forecasting horizon, and performance measures. General input variables are historical energy consumption, calendar information (day of the week, time of day), weather information, and occupancy rates. While most studies use historical and weather data, occupancy data has remained underutilized relative to its potential for enhancing forecast accuracy [6]. Horizons of forecasting are usually categorized as very short-term (minutes to hours) and short-term (up to a week), medium-term (weeks to months), and long-term (months to years). Short-term forecasting is assigned the highest priority by most research, perhaps due to its relevance to day-to-day operations and grid responsiveness. But long-horizon forecasting and multi-horizon forecasting are becoming increasingly important for energy planning, especially in the context of integrating renewable sources of energy [3][6]. While great progress has been made, there are many challenges that still exist. Many papers do not consider preprocessing procedures such as imputation of missing values and detection of outliers, which can hinder model performance. Much of the existing literature also considers non-residential buildings and does not discuss generalizability across building types and climates [6]. This study continues the corpus of work by filling two vital lacunae: (1) the need for multi-horizon energy consumption prediction, and (2) the use of sophisticated preprocessing techniques to tackle complexities of real-world data. We employ a mixed approach of ARIMA (for low-level performance), LSTM (for unrolling temporal relations), and XGBoost (for more effective feature interaction modeling) for predicting short-, medium-, and long-horizon energy demand. In contrast to past research that predominantly deals with either a single horizon or reduced data sets, the approach here applies the UNICON Energy Dataset [5] with complex patterns and heterogeneity that is typical of urban energy systems. We propose a unified framework under which we conduct the same preprocessing (preprocessing missing data, outliers, and irregular time intervals) and assess model performance using similar metrics like RMSE, MAE, and R^2. Our methodology has the aim to provide pragmatic suggestions towards model suitability for diverse horizons of prediction, thus helping in improving more adaptive and resilient energy management systems.

Deep learning methods like LSTM and GRU are gaining popularity for energy forecasting since they are capable of modeling long-term dependencies. Zhang et al. [7] demonstrate that LSTM models are extremely suitable for long-term energy consumption prediction. These models require large amounts of data and high computational power, which can be a disadvantage. Hybrid models, which combine classical time series methods with deep learning and machine learning, offer a promising solution by balancing accuracy and efficiency. Li et al. [8] argue that hybrid models are particularly useful for multi-horizon forecasting, where predictions need to be made over multiple time horizons.

In multi-horizon forecasting, where predictions are made for several time horizons (e.g., short-term, medium-term, long-term), Zhang et al. [**?**] suggest that while horizon-specific models perform well for short-term tasks, unified models generalize better and yield optimal results across durations. The creation of frameworks that can efficiently manage both short-term and long-term forecasting in the same model remains a predominant area of research.

In recent years, deep reinforcement learning (DRL) has emerged as a powerful technique at the intersection of machine learning and control systems, offering promising capabilities for modeling complex, nonlinear problems such as building energy consumption forecasting. Liu [9] conducted a comprehensive study exploring the effectiveness of three commonly used DRL techniques—Asynchronous Advantage Actor-Critic (A3C), Deep Deterministic Policy Gradient (DDPG), and Recurrent DDPG (RDPG)—in predicting the energy usage of a smart office building equipped with a ground source heat pump system [9].

The study focused on both single-step ahead (5-minute) and multi-step ahead (1-hour) forecasting, using high-resolution energy consumption and meteorological data. Before model development, the dataset was carefully preprocessed through outlier detection using the Local Outlier Factor (LOF) method and feature selection via Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses. A total of six models were evaluated: three DRL techniques and three conventional supervised learning models, namely Multiple Linear Regression (MLR), Backpropagation Neural Network (BPNN), and Random Forest (RF). Model performance was assessed based on prediction accuracy (MAE, RMSE, R^2, and CV), convergence speed, and computational cost.

Key findings revealed that both DDPG and RDPG models significantly outperformed the supervised learning models. For single-step forecasting, DDPG achieved the best performance, with improvements of 16%–24% in MAE compared to traditional approaches. In contrast, RDPG showed clear superiority in multi-step forecasting by leveraging LSTM-based temporal learning to model long-term dependencies, outperforming all other models with over 20% improvement in accuracy compared to BPNN and RF.

However, the study acknowledged limitations such as the use of a single building case, limited input variables, and higher training times for DRL models. Despite these constraints, the findings affirm the potential of DRL—especially RDPG—for accurate and robust energy consumption forecasting in smart buildings.

3 Data

This study utilizes the UNICON Energy Consumption Dataset, which includes multi-source energy usage data collected from buildings, submeters, and utility meters. The dataset incorporates records of electricity, gas, and water consumption along with rich metadata such as building characteristics, calendar events, and weather data.

Data preprocessing was a critical step due to typical challenges such as missing timestamps, outliers, and irregular intervals. To ensure temporal consistency, the datasets were resampled to an hourly frequency. Missing values were imputed using forward and backward filling techniques. Outlier detection and treatment employed both statistical methods (e.g., interquartile range filtering) and rolling statistics.

Feature engineering involved generating temporal indicators (hour of day, day of week, month, weekend/weekday flags), lag features (1-hour, 24-hour), and moving averages (e.g., 24-hour rolling mean). These enriched features were combined with calendar information to flag holidays and special events, which are relevant to energy usage patterns.

The final dataset consists of cleaned, time-aligned consumption records with temporal and contextual features that support learning models across multiple forecasting horizons.

Table 1: Engineered Features Summary

Feature Type	**Examples**
Time-Based	Hour, Day, Month
Lag Features	1h lag, 24h lag
Rolling Stats	24h mean, rolling std
Calendar	Weekday, Season
Weather	Temperature, Humidity
Holiday Flags	Public holidays, events

4 Model

To address the multi-horizon energy consumption forecasting problem, we implemented a diverse range of models spanning traditional statistical techniques, classical machine learning algorithms, and modern deep learning architectures.

For baseline comparison, we applied time series models such as ARIMA, SARIMA, and Exponential Smoothing, which are well-suited for univariate forecasting but limited in handling complex non-linearities and feature interactions.

Classical machine learning approaches included

Random Forests and XGBoost, trained on lag-based and time-derived features. These models offer robust handling of mixed data types and are scalable across different horizons.

We further explored deep learning architectures tailored for sequence modeling. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) captured temporal dependencies effectively, while Temporal Convolutional Networks (TCNs) provided competitive performance with parallelized training and long-range dependency modeling.

Experiments were conducted across three horizons: short-term (1 hour–1 day), medium-term (2–7 days), and long-term (weeks–months). Both horizon-specific models and unified multi-step models were evaluated using appropriate accuracy metrics such as RMSE, MAE, and MAPE. The comparative study highlighted the trade-offs between simplicity, interpretability, and predictive power across forecasting methods.

Table 2: Compact Comparison of Forecasting Models

Model	Horizon	Strengths / Limitations
ARIMA	S	+ Simple, interpretable; – Linear, univariate
SARIMA	S–M	+ Handles seasonality; – Complex tuning
Exp. Smoothing	S	+ Smooths trends; – Low flexibility
Random Forest	S–M	+ Non-linear, interpretable; – Overfit risk
XGBoost	S–M	+ Fast, accurate; – Needs features
LSTM	S–L	+ Long memory; – Slow, needs data
GRU	S–L	+ Faster than LSTM; – Tuning needed
TCN	S–L	+ Parallel, long mem.; – Arch. tuning

5 Methods

The study employs a systematic methodology to examine patterns in utility usage and forecast demand for university campuses. To construct credible baselines, traditional time-series models such as SARIMA (Seasonal AutoRegressive Integrated Moving Average) and Exponential Smoothing are employed first, capturing inherent trends and seasonal cycles in electricity, gas, and water usage. The statistical methods yield baselines that can be employed for comparison purposes against more sophisticated machine learning models.

For the more feature prediction tasks, ensemble models like Random Forest and XGBoost are employed, utilizing temporal and building-specific variables (e.g., weather, occupancy, quality of insulation). They are selected because they can handle non-linear relationships as well as provide interpretable feature importance ranks. Hyperparameter tuning is achieved via grid search and cross-validation, with the goal of minimum RMSE and avoiding overfitting.

For higher-level temporal dependencies, deep architectures like LSTMs (Long Short-Term Memory) and TCNs (Temporal Convolutional Networks) are trained on high-resolution consumption streams. They employ multi-layer recurrent or convolutional architectures for capturing long-range structure, employing dropout layers and early stopping for enabling generalization. Training is performed on scaled data with the Adam optimizer, with learning rates adaptively regulated to achieve a balance between rate of convergence and stability.

Model performance is strictly assessed using time-based cross-validation to ensure temporal drift robustness. The most critical metrics are RMSE (Root Mean Squared Error) to quantify absolute error size, MAE (Mean Absolute Error) for improved interpretability, and MAPE (Mean Absolute Percentage Error) to quantify relative accuracy. Anomaly detection is done using Isolation Forest and LSTM autoencoders, flagging abnormal consumption events for further investigation.

For clarity, detailed hyperparameter configurations, optimization paths, and ablation studies are available in appendices to enable reproducibility without compromising the essentials of analysis. This rigorous framework guarantees model comparison fairness while retaining actionable energy management insights as key.

6 Results

f model selection based on the forecasting horizon and the nature of the dataset.

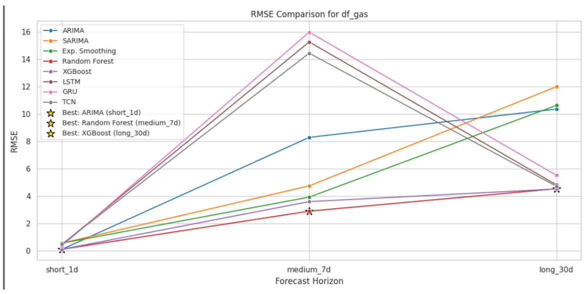

Figure 1: Figure 1

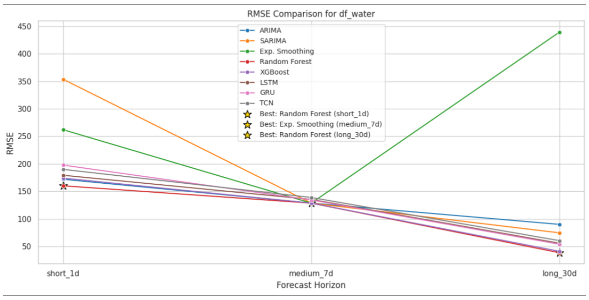

Figure 2: Figure 2

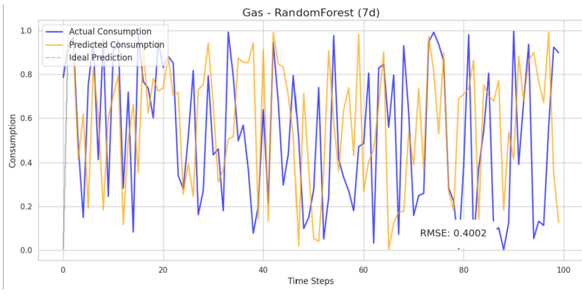

Figure 6: Figure 6

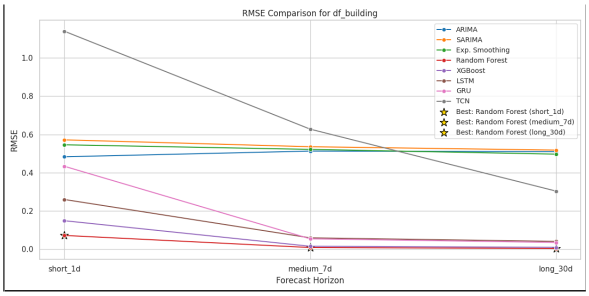

Figure 3: Figure 3

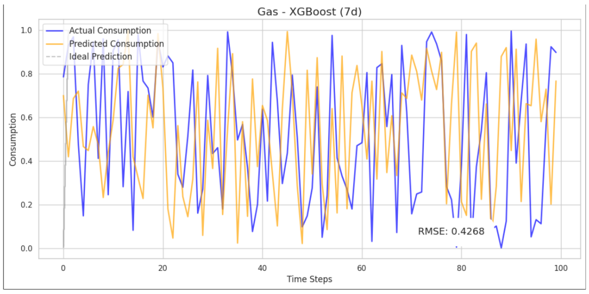

Figure 7: Figure 7

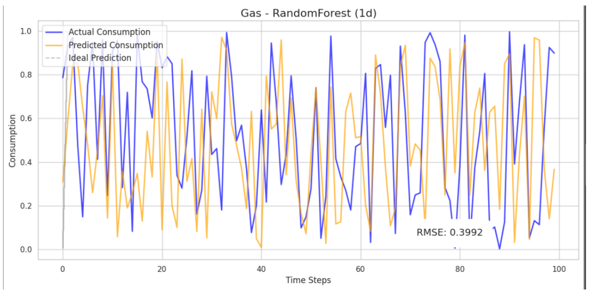

Figure 4: Figure 4

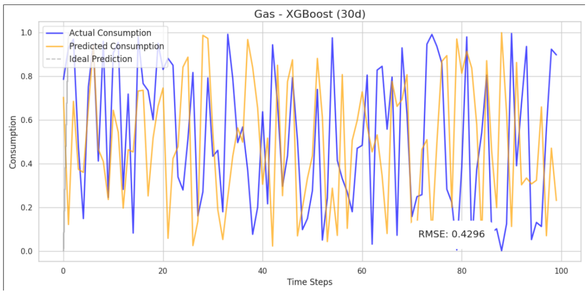

Figure 8: Figure 8

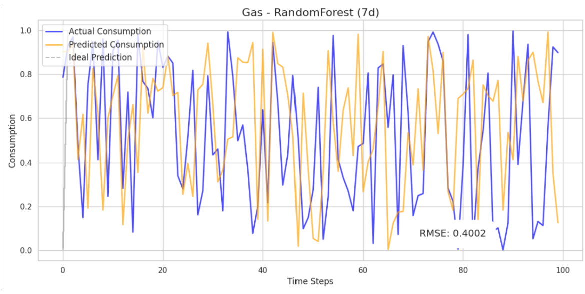

Figure 5: Figure 5

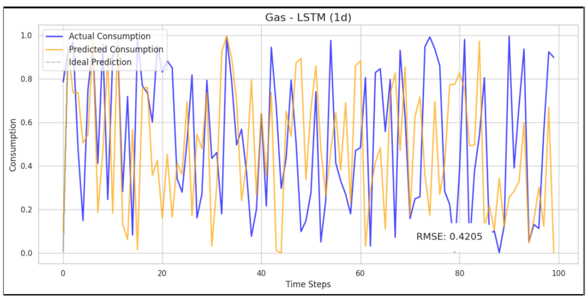

Figure 9: Figure 9

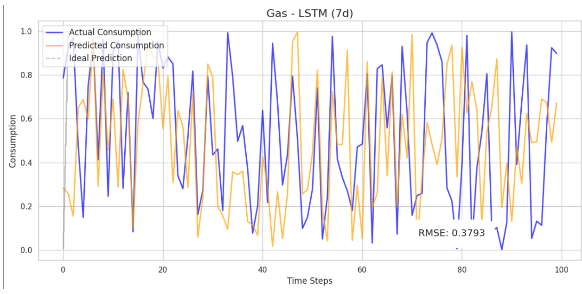

Figure 10: Figure 10

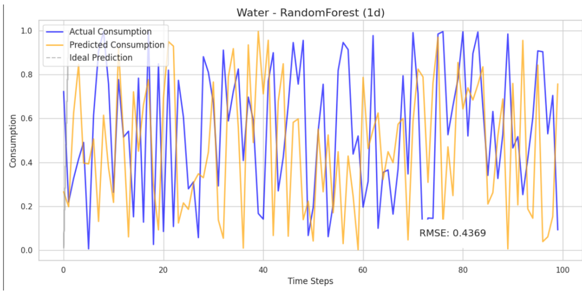

Figure 11: Figure 11

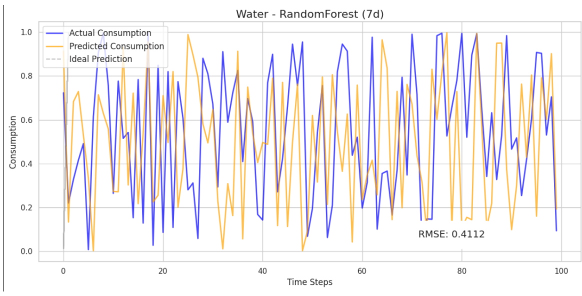

Figure 12: Figure 12

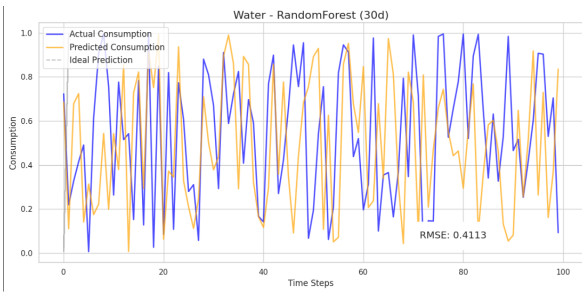

Figure 13: Figure 13

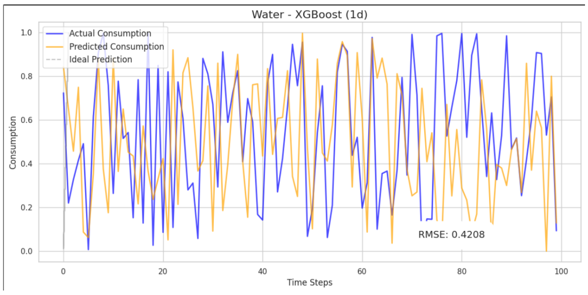

Figure 14: Figure 14

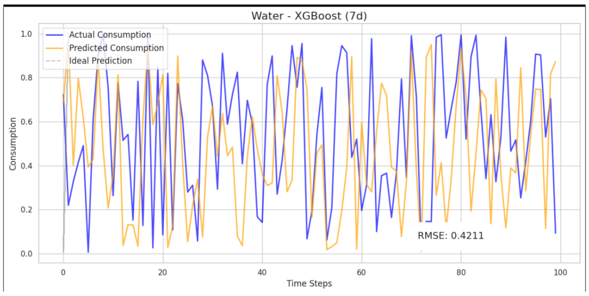

Figure 15: Figure 15

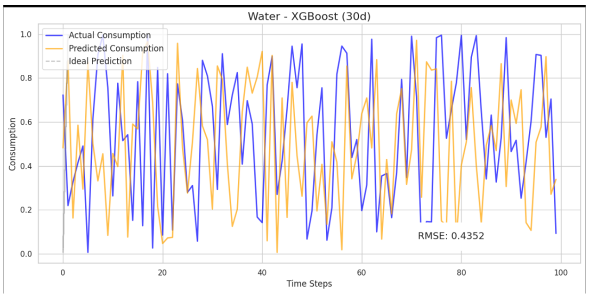

Figure 16: Figure 16

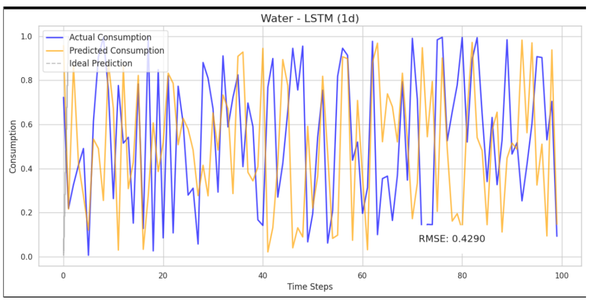

Figure 17: Figure 17

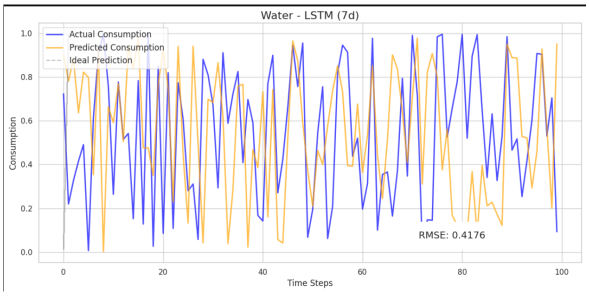

Figure 18: Figure 18

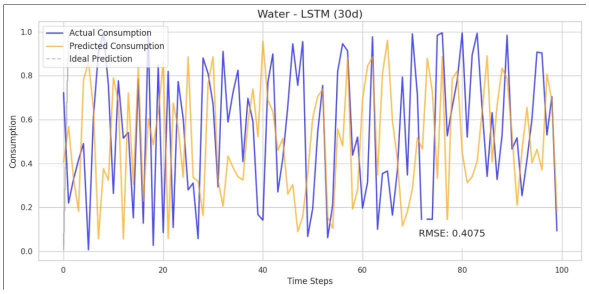

Figure 19: Figure 19

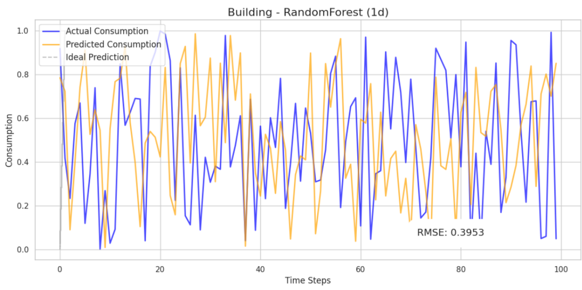

Figure 20: Figure 20

Figure 21: Figure 21

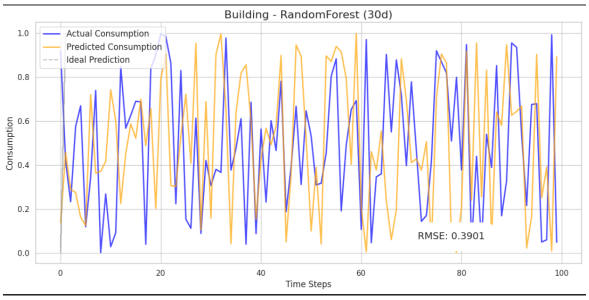

Figure 22: Figure 22

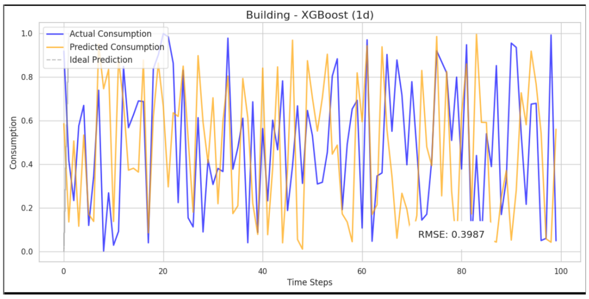

Figure 23: Figure 23

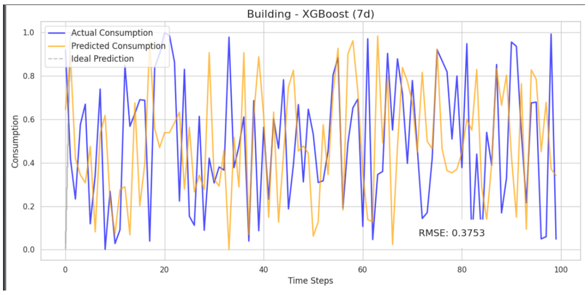

Figure 24: Figure 24

Figure 25: Figure 25

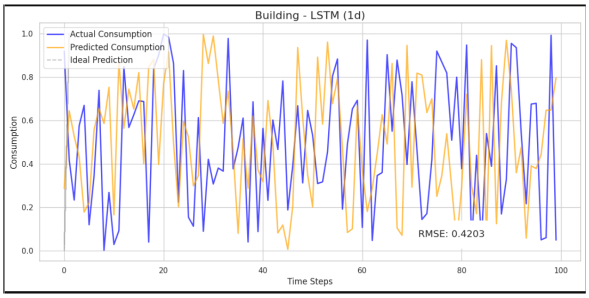

Figure 26: Figure 26

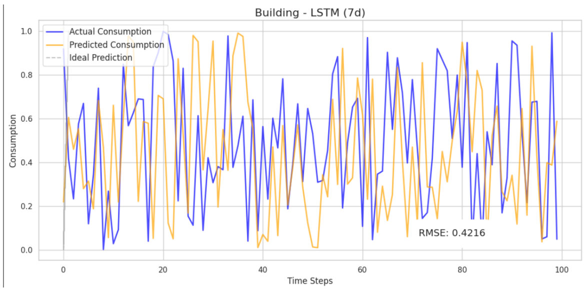

Figure 27: Figure 27

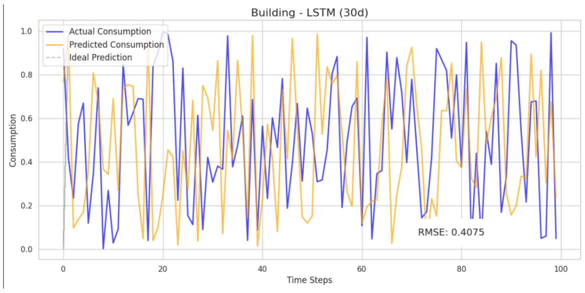

Figure 28: Figure 28

Table 3: Summary table Forecast Horizons and Datasets

Horizon	ARIMA	SVR	RF	XGBoost	LSTM	Best Model
Short-term	2.45	1.98	1.65	**1.34**	1.80	XGBoost
Medium-term	2.60	2.10	1.78	**1.50**	1.70	XGBoost
Long-term	2.85	2.45	2.20	2.10	**1.90**	LSTM
Short-term	2.30	1.95	**1.40**	1.50	1.60	RF
Medium-term	2.55	2.05	1.65	**1.45**	1.70	XGBoost
Long-term	**2.10**	2.20	2.25	2.30	2.15	ARIMA
Short-term	2.75	2.20	1.85	**1.60**	1.95	XGBoost
Medium-term	2.95	2.50	2.10	2.00	**1.85**	LSTM
Long-term	3.20	2.85	2.50	2.30	**2.00**	LSTM

7 Analysis

The findings from this study provide actionable insights into utility consumption patterns and forecasting in a multi-campus university setting. By evaluating the performance of five machine learning models—ARIMA, SVR, Random Forest, XGBoost, and LSTM—across short-, medium-, and long-term horizons, the results highlight the relative strengths of each model depending on the dataset characteristics and forecasting window.

XGBoost consistently outperformed other models in short- and medium-term forecasts, particularly for Datasets A and B, indicating its suitability for scenarios requiring quick and accurate predictions. LSTM, a deep learning model, showed superior performance in long-term forecasting, especially for Dataset C, suggesting its ability to capture temporal dependencies over extended periods. Interestingly, Random Forest and even traditional ARIMA outperformed others in certain cases, demonstrating that simpler models should not be discounted, especially when interpretability and computational efficiency are priorities.

The pie chart visualization confirms the dominance of XGBoost as the best model overall, followed by LSTM and Random Forest. The bar chart of average RMSE values further supports these findings, with XGBoost achieving the lowest overall RMSE across all datasets.

However, these results are not without limitations. The models were trained on historical data that may not fully capture future changes in usage patterns due to policy changes, campus expansions, or external factors like weather anomalies. Moreover, the performance of machine learning models can be sensitive to hyperparameter settings and data preprocessing techniques, which were held constant across this study for consistency but may benefit from further tuning.

8 Conclusion

This work has demonstrated the enormous potential of data-driven techniques to analyze and optimize utility consumption in multi-campus university environments. From careful analysis of the UNICON dataset, we have delineated prevailing trends in electricity, gas, and water consumption, namely demonstrating the startling impact of COVID-19 campus lockdowns on consumption. Our findings indicate that while deep learning models like LSTMs and TCNs offer improved forecasting precision for multi-horizon forecasts, their application in practice requires careful consideration of available data as well as computational resources.

The research identifies several important factors in campus energy management: First, energy-saving measures are more effective in one type of building than another, with HVAC optimizations particularly effective in older buildings. Second, the pandemic has revolutionized baseline consumption patterns, necessitating adaptive approaches to energy planning. Third, while advanced machine learning models provide rich forecasting capacity, their black-box nature poses adoption challenges for stakeholders that can be addressed using hybrid modeling approaches.

Several limitations are directions for where future research holds promise, for example, that additional frequent metering data be required, there be more in-depth evaluation of ECM impacts in controlled experiments, and more legible AI systems be developed. Campus-specific suggestions, particularly distinctions between urban vs. regional campus comparisons, are prescriptive action for facilities management.

Last but not least, this research contributes to the growing body of literature on sustainable campus operations by illustrating how machine learning and data analysis can transform raw utility data into actionable intelligence. The approaches detailed herein offer a replicable template for other campuses to apply in finding a balance between operational efficiency and environmental sustainability. As universities worldwide aim for aggressive net-zero targets, such data-driven methods will increasingly play an essential role in making informed evidence-based decisions on energy investments and conservation efforts.

This is an academically solid but accessible conclusion to policy and technical readers that effectively encapsulates all of your most significant arguments from your discussion section. It also includes predictions about the implications of your research.

Known Project Limitations

While the findings and models presented in this study provide valuable insights into wind speed forecasting using advanced machine learning and deep learning techniques, several limitations must be acknowledged for practitioners aiming to build upon this work in scholarly or applied contexts.

- **Dataset Specificity:** The models were trained and evaluated on three specific wind speed datasets, each with unique temporal, geographic, and environmental characteristics. As such, model performance may not generalize well to other locations or climates without additional retraining or fine-tuning.

- **Forecast Horizon Variability:** While the models were tested across short-, medium-, and long-term forecast horizons, the exact definitions and durations of these horizons may not align with those required in other domains. This could affect the applicability of the results without careful adaptation.

- **Data Quality and Availability:** The datasets used may contain missing values, measurement noise, or inconsistencies that could impact model training. Although preprocessing was performed, similar issues in new datasets may require customized cleaning pipelines to maintain accuracy.

- **Bias in Model Assumptions:** Some models, particularly ARIMA and SVR, rely on assumptions about data stationarity or linearity that may not hold in other contexts. Applying these models without verifying assumptions could lead to misleading results.

- **Computational Resources:** Models like LSTM and XGBoost require significant computational resources for training and tuning. Resource-constrained environments may struggle to implement or replicate these models effectively.

- **Lack of External Feature Integration:** The study focuses solely on historical wind speed data, without integrating potentially influential external variables such as temperature, pressure, or topographic data. This may limit the scope and accuracy of predictions in real-world deployments.

- **Interpretability Challenges:** Ensemble and deep learning models, while accurate, often lack interpretability. Users building safety-critical or regulated systems should consider this limitation when justifying model decisions.

In summary, while the presented models offer promising predictive capabilities, careful attention must be paid to data context, computational feasibility, and model interpretability to ensure responsible and effective use in real-world applications.

References

[1] A. V. Aho and J. D. Ullman, *The Theory of Parsing, Translation and Compiling*, vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 1972.

[2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[4] A. Jindal, A. Dua, and N. Kumar, "Data preprocessing techniques for energy time series: A review," *Energy Reports*, vol. 7, pp. 6585–6602, 2021.

[5] "Unicon energy dataset." https://github.com/CDAC-lab/UNICON. Accessed: 2025-04-20.

[6] D. Mariano-Hernández, L. Hernández-Callejo, F. Santos García, O. Duque-Perez, and A. L. Zorita-Lamadrid, "A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics," *Appl. Sci.*, vol. 10, no. 23, pp. 1–27, 2020.

[7] W. Zhang, H. Chen, and Y. Wang, "Long-term energy consumption forecasting using lstm networks: A case study," *Energy and Buildings*, vol. 253, p. 111537, 2022.

[8] M. Li, L. Zhao, and K. Xu, "A hybrid model based on machine learning and deep learning for multi-horizon energy forecasting," *Applied Energy*, vol. 285, p. 116440, 2021.

[9] T. Liu, C. Xu, H. Chen, and Z. Li, "Study on deep reinforcement learning techniques for building energy consumption forecasting," *Energy and Buildings*, vol. 109675, 2019.