Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

Smart Energy Monitor: A Java-Based Application with Artificial Intelligence for Household Electricity Consumption Prediction

*Mounia Stitou[1], Mohamed Ben Jouad[1]

[1] *Department of Computer Engineering, ENSI Tangier*
Tangier, Morocco
*mounia.stitou@ensi.ma
mohamed.benjouad@ensi.ma

***Abstract*— EnerWise is a household energy management prototype that combines appliance monitoring, cost estimation, recommendations, and one-hour-ahead energy consumption forecasting. A customized six-month synthetic dataset was developed for six household appliances. Following an initial Random Forest experiment (MAE = 0.1927 kWh, MSE = 0.1262, and R^2 = 0.3825), the temporal consistency of the dataset and the input features were improved. The resulting model achieved an R^2 of 0.5397. Hyperparameter tuning using RandomizedSearchCV and TimeSeriesSplit further improved the performance, achieving an MAE of 0.1502kWh, an MSE of 0.0797, and an R^2 of 0.5828. RNN, LSTM, and SSM models were also evaluated. The selected model was integrated into a Vue.js–Spring Boot–H2–FastAPI architecture.**

***Keywords*— household energy consumption, energy forecasting, Random Forest, RNN, LSTM, SSM, hyperparameter tuning, energy efficiency.**

I. INTRODUCTION

Short-term residential energy forecasting is challenging because of behavioral variability and rapid changes in appliance operating states. Recent studies have highlighted the value of appliance-level modeling [1], recurrent neural networks [2], and feature engineering [3]. EnerWise aims to connect these approaches within an application that can provide actionable information to the user.

The contribution of this work is an end-to-end pipeline comprising the generation of an appliance-oriented dataset, exploratory data analysis, comparison of four model families, investigation of temporal memory, hyperparameter optimization, and integration of the best-performing model into a Java application.

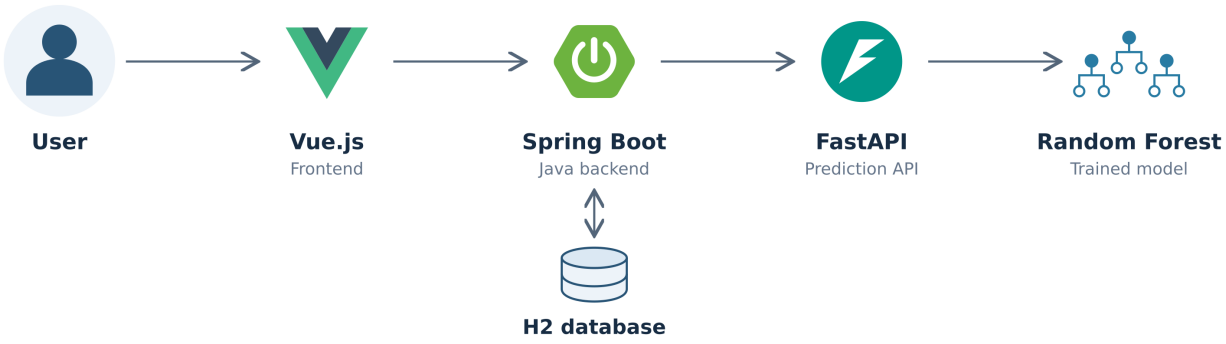


Fig. 1. General architecture of the EnerWise prototype.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

II. STATE OF THE ART AND ARCHITECTURAL DIFFERENCES

The models considered in this study do not process sequential information in the same way. Elman's RNN [4] feeds the previous hidden state back into the network to build a dynamic memory, but it may suffer from vanishing gradients over long dependencies. LSTM [5] introduces gating mechanisms and a memory cell to preserve information more effectively. State Space Models (SSMs) describe the evolution of a latent state, while Mamba [6] introduces selective SSMs whose parameters depend on the input. Mamba is cited here as a recent development in SSM architectures but was not implemented in EnerWise.

Model	Architectural principle	Main advantage	Limitation in this study
Random Forest	Ensemble of independent trees	Robust for tabular variables	No native memory
RNN	Recurrent hidden state	Models temporal order	Long dependencies are difficult
LSTM	Memory cell + gates	Improved long-term memory	More parameters
SSM	Dynamic latent state	Structured sequences and memory	Simple version tested
Mamba [6]	Input-dependent selective SSM	Selection and linear-time computation	Not tested here

TABLE I. DIFFERENCES BETWEEN THE STUDIED MODELS

A. Positioning with Respect to Similar Studies

Wang et al. [1] proposed a bottom-up approach based on appliance characteristics. Forootani et al. [3] showed that machine learning enhanced through feature engineering can remain competitive with deep learning. HouseEEC [7], evaluated on more than 4.8 million real-world samples for day-ahead forecasting, achieved $R^2 = 0.90$, while its Random Forest model achieved $R^2 = 0.79$ using enriched features. These values are not directly comparable with ours because the datasets, forecasting horizons, and experimental protocols differ.

Study	Data / horizon	Approach	R^2
HouseEEC [7]	Real data, day-ahead	Deep learning	0,90
HouseEEC [7]	Real data, day-ahead	Random Forest	0,79
EnerWise	Synthetic, +1 h	RF before tuning	0,5397
EnerWise	Synthetic, +1 h	RF after tuning	0,5828

TABLE II. INDICATIVE POSITIONING WITH RESPECT TO THE LITERATURE

The comparison is therefore intended to provide an indicative reference and, more importantly, to highlight the difference in experimental maturity. EnerWise still requires validation using real-world, multi-household data before any direct comparison with the state of the art can be made.

From an architectural perspective, the literature highlights a trade-off between explicit feature representation and automatic learning of dependencies. Random Forest directly exploits engineered variables, which facilitates interpretation of appliance influence. In contrast, RNN and LSTM learn an internal representation of the sequence. LSTM was specifically introduced to preserve information over longer intervals [5]. SSMs represent a sequence through a latent state whose evolution is parameterized, whereas Mamba makes this evolution selective according to the input [6].

The choice of model therefore depends on the amount of available data, the length of temporal dependencies, and the nature of the input variables. In our case, the presence of explicit appliance-related rules and the moderate dataset size favor a tabular learning method. This observation is consistent with Forootani et al. [3], who emphasize the importance of identifying dominant features before increasing model complexity.

Finally, the forecasting horizon must be considered in any comparison. EnerWise targets one-hour-ahead forecasting for a single household, whereas HouseEEC [7] performs day-ahead forecasting for 300 households using several years of data and weather variables. The literature results are therefore used as methodological reference points rather than as evidence of direct superiority or inferiority.

Copyright © 2026
ISSN: 2961-6611

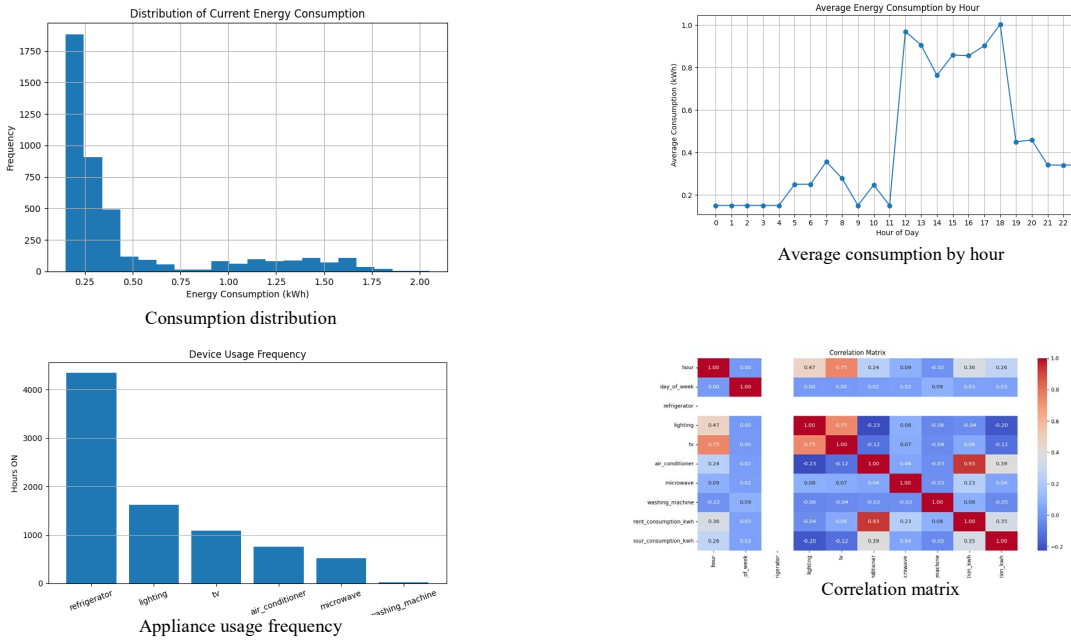
Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

III. DATA AND METHODOLOGY

A. Dataset and Exploratory Data Analysis

The dataset, specifically designed for EnerWise, covers the period from January 1 to June 30, 2025, at an hourly resolution. After creating the next_hour_consumption_kwh target, 4,343 observations are available. The modeled appliances are a refrigerator (150 W), lighting (100 W), television (120 W), air conditioner (1,500 W), microwave oven (1,200 W), and washing machine (800 W).

The EDA shows that most hours have low energy consumption, with peaks associated with the simulated appliance usage periods and a strong association between air-conditioner operation and current consumption. The initial error analysis revealed excessively random transitions; the data-generation process was therefore revised to improve temporal consistency.



Consumption distribution — Average consumption by hour — Appliance usage frequency — Correlation matrix

Fig. 2. Exploratory analysis of the EnerWise dataset.

B. Model Preparation

Random Forest uses temporal variables, the operating states of the six appliances, current consumption, previous-hour consumption, and the cyclical components hour_sin/hour_cos. The train/test split is chronological (80/20). RNN, LSTM, and SSM use six-hour sequences with nine variables, resulting in 3,469 training sequences and 868 test sequences.

For Random Forest, two cyclical variables, hour_sin and hour_cos, encode the 24-hour periodicity without introducing an artificial discontinuity between 23:00 and 00:00. Previous-hour consumption is added by shifting the time series. Incomplete rows are removed after this operation. The target remains exclusively the next-hour consumption, thereby preventing future information from being used at prediction time.

A chronological split was preferred over random shuffling in order to preserve the temporal structure of the problem. For sequential networks, each sample contains six consecutive time steps. MAE, MSE, and R^2 were selected as evaluation metrics: MAE expresses the average prediction error directly in kWh, MSE penalizes large errors more strongly, and R^2 measures the proportion of explained variance. Together, these metrics distinguish average accuracy, sensitivity to peaks, and explanatory performance.

The first experimental phase also served as a diagnostic step. The initial R^2 of 0.3825 was not considered a final result: error analysis showed that some appliance activations were too independent from one hour to the next. The generation process was therefore revised to produce more persistent and temporally coherent usage patterns. This iteration is an important part of the methodology because it directly links the EDA to model improvement.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

IV. RESULTS, HYPERPARAMETER TUNING, AND ANALYSIS

Model	MAE (kWh)	MSE	R^2
Random Forest	0,1556	0,0880	0,5397
RNN	0,1969	0,0995	0,4799
LSTM	0,1947	0,1016	0,4692
Simple SSM	0,1946	0,1034	0,4597

TABLE III. PERFORMANCE BEFORE HYPERPARAMETER TUNING

Random Forest provides the best performance among the initial configurations. RNN and LSTM learn temporal dependencies, but their largest errors occur during abrupt load changes. The simple SSM remains close to the recurrent models in terms of performance.

A. Random Forest Hyperparameter Tuning

To improve the initial hyperparameter selection, RandomizedSearchCV was applied to 24 configurations using four-fold TimeSeriesSplit validation, exclusively on the training portion, with MAE as the optimization criterion. The explored parameters included n_estimators, max_depth, min_samples_split, min_samples_leaf, and max_features.

Hyperparameter	Value
n_estimators	300
max_depth	16
min_samples_split	10
min_samples_leaf	2
max_features	log2

TABLE IV. BEST CONFIGURATION AFTER HYPERPARAMETER TUNING

The best mean cross-validation MAE is 0.1561 kWh. On the separately preserved test set, the optimized model achieves MAE = 0.1502 kWh, MSE = 0.0797, and R^2 = 0.5828, corresponding to an absolute R^2 improvement of 0.0431 over the final untuned Random Forest model.

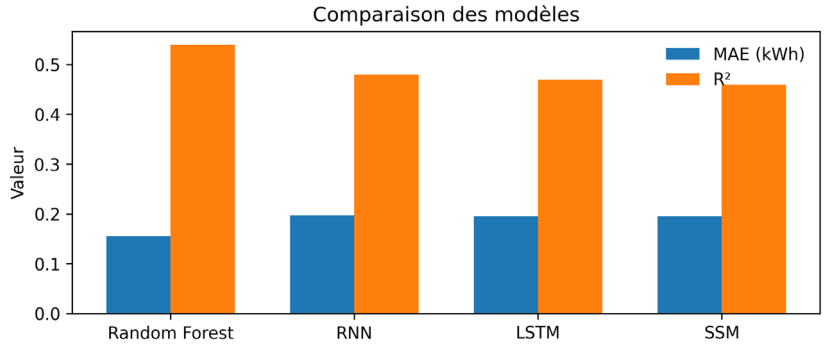


Fig. 3. Model comparison before hyperparameter tuning.

B. Feature and Memory Influence

The feature-influence experiment yields R^2 = 0.3444 using current consumption alone, 0.3171 using current consumption + t−1, and 0.5482 using the complete feature set. Regarding temporal memory, R^2 increases from 0.1443 (1 h) to 0.2313 (3 h), 0.3084 (6 h), and 0.3305 (12 h). Appliance and temporal context therefore provide more useful information than isolated consumption history.

Hyperparameter tuning provides an additional improvement without modifying the test set. The search is performed exclusively on the training data to prevent the test set from influencing hyperparameter selection. TimeSeriesSplit preserves the temporal order of observations during cross-validation. The best configuration uses 300 trees, limits the maximum depth to 16, sets min_samples_split = 10 and min_samples_leaf = 2, and selects log2 for max_features. The increase in R^2 from 0.5397 to 0.5828 is moderate but consistent with the simultaneous reductions in MAE and MSE. It indicates that the initial model was not fully optimized, without suggesting that tuning radically transformed the results. A more exhaustive search could still be conducted, but it should be applied fairly to the other models to preserve the validity of the comparison.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

V. DISCUSSION AND INTEGRATION INTO ENERWISE

A. Discussion

The results confirm the importance of feature engineering highlighted in [3]. The stronger Random Forest performance is consistent with the tabular and explicitly structured nature of the dataset. However, it does not support a general conclusion that Random Forest is superior to RNN, LSTM, or SSM, since their input representations differ and the synthetic dataset remains limited.

Hyperparameter tuning improves the model without modifying the test set, further supporting the selection of Random Forest for the prototype. However, the gap with HouseEEC [7] emphasizes the need for validation using real-world data from multiple households and seasons. The work of Fekri et al. [2] also demonstrates the relevance of adaptive recurrent models when consumption behavior changes over time.

B. Application Integration

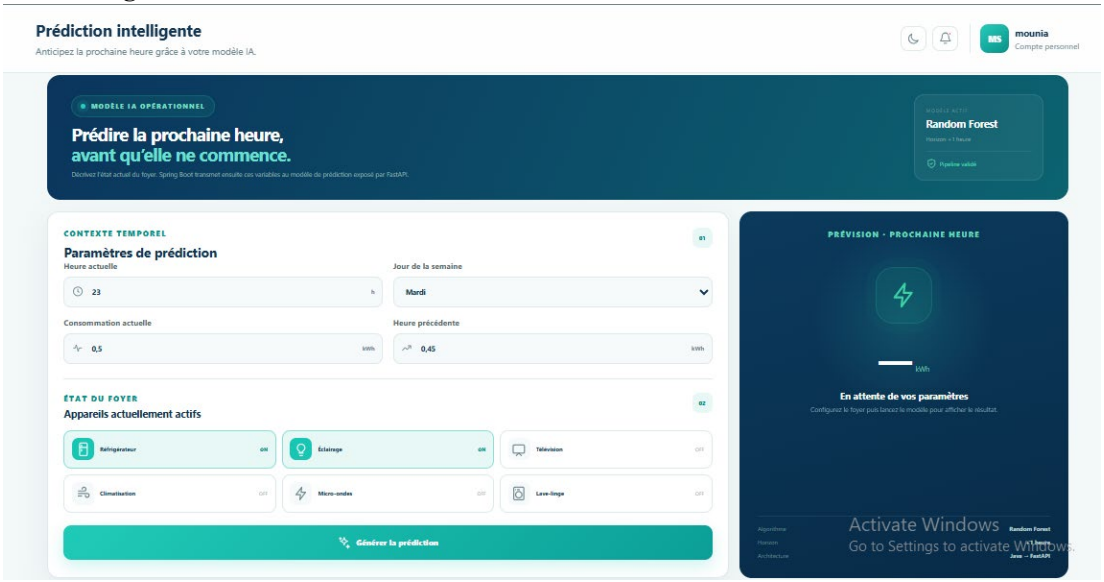


Fig. 4. Intelligent forecasting interface integrated into EnerWise.

Vue.js provides the user interface; Spring Boot centralizes business logic and H2 access; FastAPI exposes the Python model. The prediction request contains the hour, day, current and previous consumption values, and appliance states. The AI service computes the derived variables and returns the predicted consumption for the next hour.

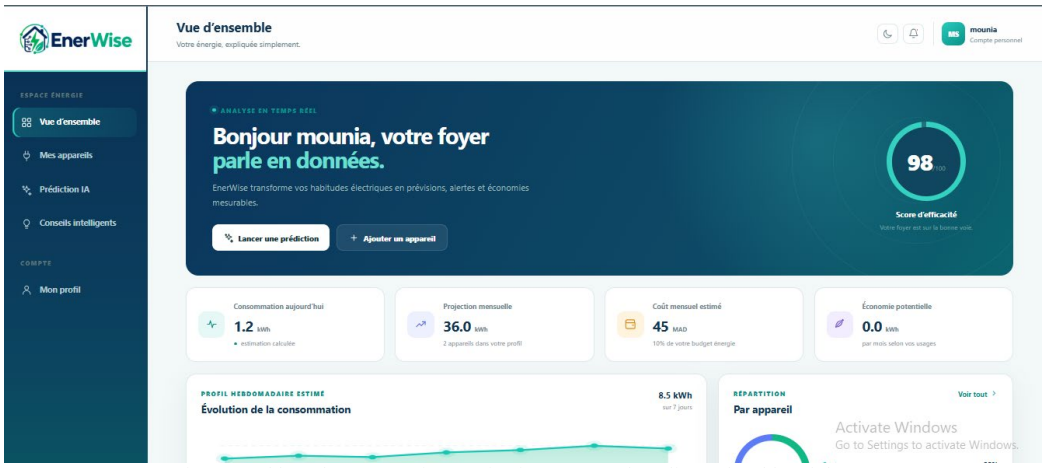


Fig. 5. Dashboard: consumption, projection, cost, and appliance-level breakdown.

The user therefore has a single environment in which to register appliances, monitor estimates, obtain a forecast, and consult recommendations. This integration represents the main application-oriented contribution of the work.

Copyright © 2026
ISSN: 2961-6611

Proceedings of Book Series PBS -Vol 42, pp.1-6 **Advances in Energy Transition and Environmental Impacts**

VI. LIMITATIONS, FUTURE WORK, AND CONCLUSION

A. Limitations and Future Work

The main limitation is the synthetic nature of the dataset. The rules simulate plausible usage patterns but do not capture the full range of human, weather-related, tariff-related, or seasonal variability. The comparison between model families is also imperfect because Random Forest and the sequential models do not use exactly the same input representation.

Future work will focus on real smart-meter data, multi-household evaluation, harmonized model inputs, fair hyperparameter tuning for each model, the inclusion of weather variables, and the investigation of multiple forecasting horizons. Mamba [6] may also be evaluated in future work as a selective SSM, but it is not presented as a result of the current version.

B. Conclusion

EnerWise demonstrates the feasibility of a complete residential energy forecasting pipeline. Improving the dataset increased the Random Forest R^2 from 0.3825 to 0.5397, while hyperparameter tuning further increased it to 0.5828 with an MAE of 0.1502 kWh. The analyses show that appliance-oriented features and temporal context are key factors in prediction performance. The Vue.js–Spring Boot–H2–FastAPI integration transforms the model into a directly usable application feature. Validation using real-world data is now the main priority.

REFERENCES

[1] S. Wang, X. Deng, H. Chen, Q. Shi and D. Xu, “A bottom-up short-term residential load forecasting approach based on appliance characteristic analysis and multi-task learning,” Electric Power Systems Research, vol. 196, art. 107233, 2021.

[2] M. N. Fekri, H. Patel, K. Grolinger and V. Sharma, “Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network,” Applied Energy, vol. 282, art. 116177, 2021.

[3] A. Forootani, M. Rastegar and A. Sami, “Short-term individual residential load forecasting using an enhanced machine learning-based approach based on a feature engineering framework: A comparative study with deep learning methods,” Electric Power Systems Research, vol. 210, art. 108119, 2022.

[4] J. L. Elman, “Finding Structure in Time,” Cognitive Science, vol. 14, no. 2, pp. 179–211, 1990, doi: 10.1207/s15516709cog1402_1.

[5] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[6] A. Gu and T. Dao, “Mamba: Linear-Time Sequence Modeling with Selective State Spaces,” arXiv:2312.00752, 2023.

[7] I. Kiprijanovska, S. Stankoski, I. Ilievski, S. Jovanovski, M. Gams and H. Gjoreski, “HousEEC: Day-Ahead Household Electrical Energy Consumption Forecasting Using Deep Learning,” Energies, vol. 13, no. 10, art. 2672, 2020, doi: 10.3390/en13102672.

[8] H. Zang and al., “Residential load forecasting based on LSTM fusing self-attention mechanism with pooling,” Energy, vol. 229, art. 120682, 2021.

Copyright © 2026
ISSN: 2961-6611