Explainable Machine Learning for Energy Consumption Prediction: A Case Study in the Gumi Industrial Complex

Vivian Ukamaka Ihekoronye, Wisdom Chizitere Njoku, Esmot Ara Tuli, Jae Min Lee, Dong-Seong Kim
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
Department of Electrical and Electronics Engineering, Federal University of Technology, Owerri, Nigeria
Kumoh National Institute of Technology Gumi, South Korea
(ihekoronyevivian, chiziterewisdom24) @gmail.com,(esmot, ljmpaul, dskim)@kumoh.ac.kr

***Abstract*—This paper investigates and compares the performance of various predictive models for energy consumption forecasting in the Gumi Industrial Complex. A 2024 survey highlighted energy efficiency initiatives such as equipment upgrades (19.5%) and energy monitoring systems (4.3%), emphasizing the need for accurate and interpretable predictive technologies in this sector. Hence, we employed the SHAP explainer to provide insights into the predictions of the optimal model. Simulation results demonstrate the superior performance of the Random Forest (RF) model, achieving a prediction error of 0.253. Also, the model's forecasts offer actionable insights for optimizing energy usage and support decision-making toward self-sufficiency targets.**

***Index Terms*—Energy consumption, Explainable AI, Predictive models**

I. Introduction

Minimizing energy consumption and reducing carbon footprints are crucial for a cleaner environment, particularly in industrialized nations like South Korea. Rapid urbanization and industrialization have significantly increased energy consumption, with the industrial sector responsible for over 60% of South Korea's energy use [1]. Moreover, a recent survey of companies in the Gumi Industrial Complex highlights the need for energy efficiency initiatives, such as replacing obsolete equipment and integrating intelligent energy systems capable of accurately predicting and monitoring consumption. Thereby reducing emissions and supporting energy self-sufficiency to promote sustainability.

Machine learning (ML) models, such as Random Forest (RF) and Linear Regression (LR) [2], along with deep learning (DL) models like Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, are widely used for energy forecasting due to their high accuracy [3]. Although these models show strong predictive performances offering insights for reducing consumption and emissions, they are often considered 'black box' models due to their complex internal operations, which make it challenging to interpret how specific inputs influence the predictions. The emergence of explainable AI (XAI) techniques such as SHapley Additive exPlanations (SHAP) is used to address the interpretability challenges of complex ML and DL models. By attributing importance scores to input features, SHAP provides insights into how each feature contributes to model predictions, enabling more transparent and trustworthy decision-making.

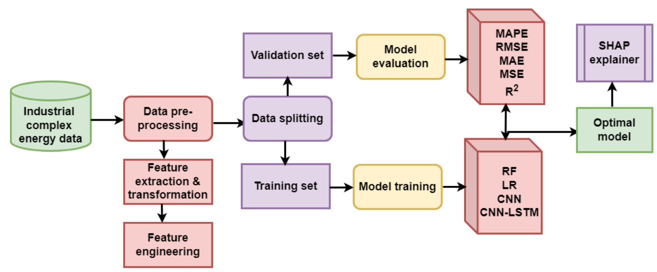

Fig. 1. Proposed Methodology

Therefore, this study leverages XAI to evaluate the performance of two ML and two DL models for predicting energy consumption trends using data from the Korea Electric Power Corporation. The results provide actionable insights into key features influencing the optimal model's predictions, supporting strategies to enhance energy sufficiency.

II. Methodology

The proposed methodology implemented in this work is highlighted in Fig.1. The dataset utilized consists of hourly time series data from January 2022 to July 2023, capturing energy consumption and solar power generation for 53 companies in the Gumi Industrial Complex, containing 727,584 samples with features such as company, date, hour, and consumption. Due to sparsity and periods of zero consumption in the data, a log transformation was applied to stabilize variance and improve stationarity. Temporal features (day, week, month, year), lag features (Lag-1, Lag-24, Lag-168), and moving averages were generated to capture dependencies and highlight trends. These features form the predictor variables (X), while log-transformed consumption was used as the target variable (y).

The data was split into training and validation sets (80/20), and two ML models (RF, LR) and two DL models (CNN, CNN-

TABLE I
PERFORMANCE COMPARISON OF THE ML AND DL MODELS

MODEL	MAE	MSE	MAPE	R2	RMSE
RF	0.253	0.319	0.788	0.897	0.564
LR	1.380	3.028	2.427	0.021	1.740
CNN	0.482	0.665	0.894	0.723	0.816
CNN-LSTM	0.440	0.647	0.872	0.749	0.804

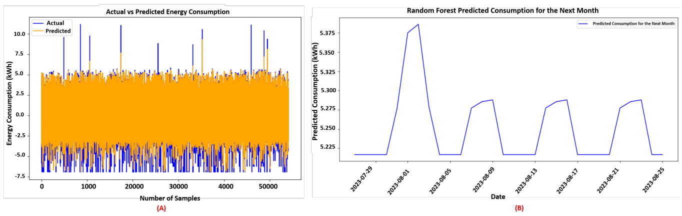

Fig. 2. Prediction results of RF model (A) Predicted vs. actual energy consumption (B) Prediction for the next month

LSTM) were trained using hyperparameters obtained from grid search cross-validation. Model performance was evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) Finally, the optimal model was analyzed with SHAP (XAI) to interpret the key predictive features and offer insights for optimizing energy usage.

III. EXPERIMENTAL SETUP AND RESULT ANALYSIS

All experiments were performed using the Scikit-learn Python library on a system with an Intel(R) Core(TM) i5-8500 CPU @ 3.00GHz and 16GB of RAM. The evaluation results in Table I highlight the superior performance of the RF model, which outperforms the other models across all metrics. With an R^2 of 0.897 and an MSE of 0.319, the RF model explains approximately 89.7% of the variance in energy consumption while maintaining a low average prediction error. Thus, making it a reliable choice for forecasting energy consumption. In contrast, the LR model had the lowest performance, while the CNN-LSTM model ranked second best. Also, as shown in Fig. 2(A), the actual vs. predicted graph demonstrates that the RF model effectively captures the general trend of energy consumption and accurately predicts consumption patterns for the majority of the samples. The forecast in Fig. 2(B) reveals recurring periods of higher and lower consumption, with spikes at the end of July and August, indicating increased production and energy demand in the industrial complex. The lower energy usage on other days in August suggests that industries could maximize solar power utilization during high-demand periods to reduce reliance on the national grid.

The RF model's forecast for the next year, shown in Fig. 3(A), reveals a seasonal energy consumption pattern in the Gumi industrial complex, likely due to operational schedules or production cycles. The forecast indicates higher demand from September to November 2023, offering insights for management to plan for sufficient energy supply during peak

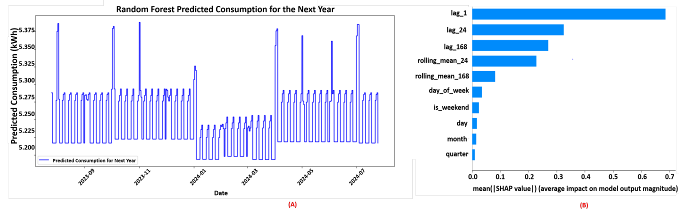

Fig. 3. (A) Prediction output for the next year (B) SHAP feature importance plot of RF model

periods. In contrast, the stable lower consumption from January to March reflects the minimum demand level, providing an opportunity to optimize energy use and storage for improved efficiency. Lastly, the SHAP explainer is used to enhance the understanding and transparency of the model's predictions. In Fig. 3(B), the most prominent features contributing to the RF model's predictions are the lag features, which capture past consumption values from the previous hour, 24 hours, and a week ago. Understanding the importance of these features can help management make informed decisions on future energy planning based on historical consumption patterns.

IV. CONCLUSION

In this study, we leveraged the predictive performance of the Random Forest (RF) model to forecast energy consumption in the Gumi Industrial Complex accurately. Also, by employing the SHAP explainer for transparent interpretation, we highlighted key features influencing the model's predictions. This integration of RF and SHAP offered valuable insights and strategies for optimizing energy consumption to support self-sufficiency goals.

ACKNOWLEDGMENT

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-RS-2020-II201612, 25%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 25%). This research used data provided by the Energy Self-Sustaining Infrastructure Development Project, funded by Korea Industrial Complex Corporation (KICOX, 25%), and by the MSIT, Korea, under the ITRC support program(IITP-2024-RS-2024-00438430, 25%) supervised by the IITP.

REFERENCES

[1] K. E. E. I. M. S. Korea), "Distribution of final energy consumption in south korea in 2022, by sector," *Korea Energy Economics Institute*, 2024.

[2] S. Kapp, J.-K. Choi, and T. Hong, "Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters," *Renewable and Sustainable Energy Reviews*, vol. 172, p. 113045, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032122009261

[3] R. Morcillo-Jimenez, J. Mesa, J. Gómez-Romero, M. A. Vila, and M. J. Martin-Bautista, "Deep learning for prediction of energy consumption: an applied use case in an office building," *Applied Intelligence*, vol. 54, 2024. [Online]. Available: https://doi.org/10.1007/s10489-024-05451-9