

Explainable Multi-Horizon Deep Learning Framework for Household Energy Forecasting

Padmini S, Kandula Venkata Aashish, Sivarathri Venkat Kushwanth Raj

Department of Computing Technologies, SRM Institute of Science and Technology,
Kattankulathur-603203, Chennai, India

padminis@srmist.edu.in, ka7386@srmist.edu.in, vs3746@srmist.edu.in

Abstract. The growing need for intelligent energy management systems has also highlighted the importance of accurate, interpretable models for household energy consumption forecasting. This paper presents an explainable multi-horizon deep learning model for forecasting household electricity consumption at hourly, daily, and weekly time horizons. The proposed model combines optimised preprocessing, appliance-level energy decomposition, and sequence-to-sequence recurrent models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. To improve the interpretability of the proposed model, SHAP feature attribution and permutation importance analysis are used to explain the effect of historical consumption patterns and appliance-level features on the forecasting results. The proposed model forecasts multi-horizon energy consumption while providing appliance-level information. The experimental results show high forecasting accuracy across various time horizons. The proposed system improves the interpretability and trustworthiness of deep learning models in energy management systems.

Keywords: Household Energy Forecasting, Multi-Step Prediction, LSTM, GRU, SHAP, Time-Series Modelling, Explainable AI, Appliance-Level Energy Decomposition, Multi-Horizon Forecasting

1 Introduction.

As the use of smart energy management systems (SEM) continues to expand, there is a growing need for accurate forecasts of household energy consumption. Accurate forecasts of electricity consumption can enable efficient energy distribution, demand management, and cost savings. The growing use of smart meters provides us with high-resolution energy data. However, the challenge remains modelling household energy consumption due to the nonlinear, seasonal, and appliance-specific variations in the data. Traditional statistical models cannot handle time-based relationships in this data; therefore, deep learning models have been successful in forecasting time series. In particular, sequence models such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks. Deep learning models can learn long-term dependencies and nonlinear relationships from past data. However, deep learning models are considered black boxes, making it difficult to interpret how they make predictions. In energy management, a lack of interpretability may hinder the adoption of the models.

To address these challenges, the proposed project aims to develop an explainable framework for multi-step household energy forecasting. The proposed framework can predict electricity consumption at hourly, daily, and weekly levels. The framework is also designed to offer appliance-level details, providing a better understanding of energy consumption patterns. In addition, a miscellaneous load is also considered to provide a comprehensive view of total energy consumption.

The forecasting system employs sequence-to-sequence LSTM and GRU models for multi-horizon forecasting. Both models are compared in the same experimental setup to enable a fair comparison. To enhance explainability, SHAP-based feature attribution and permutation importance are employed to assess the effect of past inputs and appliance-level features on the forecasting task. This makes it possible to analyse the effect of appliances on energy use and the role of time periods in determining energy consumption. By combining forecasting accuracy with explainability, the new model promotes accurate, yet interpretable, automated smart energy management.

1.1 Main Contributions

- Developed a sequence-to-sequence deep learning model for energy consumption prediction.
- We applied two new load prediction methods to forecast combined appliance-level decomposition and miscellaneous load.
- We evaluated models by comparing LSTM against GRUs under identical experimental conditions.
- We used SHAP values and permutation importance to analyse different features and time steps.
- Model performance was evaluated at hourly, daily, and weekly granularity.
- Developed an interpretable model for intelligent energy management.

2 Literature survey

The development of deep learning techniques has shifted the focus from traditional statistical approaches to more advanced predictive models for household energy consumption. Most early energy consumption forecasts were developed using time-series models (e.g., ARIMA) that were effective for predicting stable linear loads but less effective for modelling non-linear, highly variable data (e.g., smart meter data) (1). While various improvements to statistical models were made by decomposing seasonal components from time series data and estimating trends, they were still unable to capture complex temporal dependencies (2) adequately. Multiple studies indicate that precise forecasting one day ahead is essential to maintaining the effective operation and reliability of the electric grid within electricity mandates (e.g., balancing, curtailments) (3). Machine learning prediction models, such as SVR, were developed to overcome the limitations of traditional time-series models for modelling complex nonlinear data sets (4). For electricity load forecasting tasks, ensemble learning techniques such as random forests significantly improved predictive accuracy and reduced predictive model overfitting (5). Due to deep learning methods' superior ability to automatically build hierarchical feature representations from raw data, neural network prediction models are the most successful forecasting methods in the literature (6). The Long Short-Term Memory (LSTM) network was proposed to address the vanishing gradient problem and to model long-term dependencies in sequential data effectively, making it well-suited for energy time series forecasting (7). Gated Recurrent Units (GRU) were later proposed as computationally efficient substitutes for LSTMs, simplifying parameter complexity without sacrificing performance (8). Deep neural networks were then applied to building and residential load forecasting, demonstrating substantial improvements over traditional regression models (9). Experimental studies further validated that LSTM models perform better than traditional statistical models for short-term residential energy forecasting (10). With the growing need for multi-horizon forecasting, sequence-to-sequence models with encoder-decoder architectures were proposed to facilitate multi-step-ahead forecasting simultaneously (11). Multi-step LSTM models were proposed to mitigate recursive forecasting errors and enhance stability for multi-step ahead predictions (12). Comparative studies on recurrent models confirmed the efficiency and real-time applicability of GRU models for sequential forecasting problems (13). Theoretical underpinnings of supervised sequence modelling further reinforced the role of recurrent neural networks in time-series forecasting (14). Parallel work focused on appliance-level energy modelling using Non-Intrusive Load Monitoring (NILM), allowing for the disaggregation of total energy consumption in a household into appliance-level components (15). Probabilistic modelling approaches further advanced appliance-level identification and behaviour analysis

(16). Deep learning training methodologies, including normalisation and regularisation, were shown to improve model generalisation and convergence stability (17) greatly. To address the growing need for reliable AI systems, interpretability techniques such as SHAP have been proposed to provide feature attribution explanations for machine learning models (18). Gradient-based visualisation tools, such as Grad-CAM, helped shed light on deep neural network decision-making processes (19). More recently, DeepLIFT was proposed as a robust attribution technique that compares neuron activations to reference baselines, allowing for accurate contribution analysis (20).

3 Proposed Methodology

The proposed framework uses sequence-based deep learning models to design an explainable multi-step household energy forecasting framework. The time series data is converted into supervised learning sequences by using the sliding window technique after data cleaning, Min-Max normalization, temporal aggregation, and appliance decomposition. To achieve multi-horizon forecasting at hourly, daily, and weekly time scales, both LSTM and GRU models are used. The performance of the models is evaluated using standard regression analysis metrics, and SHAP-based feature attribution and permutation importance are used for feature-level and temporal explanations.

3.1 Preprocessing of Data and Feature Engineering

The dataset used in this research is a household electricity consumption dataset obtained from smart meters. The dataset contains global active power and appliance-level energy consumption, measured at fixed time intervals. To facilitate easier analysis of the data over time, timestamp information is generated and converted to a time-indexed format. Negative and irregular values in the dataset were removed to create a "clean" dataset; any missing values were handled using linear interpolation. Finally, multi-scale forecasts were enabled by resampling the clean data to hourly, daily, and weekly values.

Normalisation Formula :

$$x' = \frac{(x - \mu)}{\sigma} \quad (1)$$

where x is the original value and x' is the normalised value. The dataset is divided into training and testing sets using a time-based split to preserve temporal order and avoid information leakage.

3.2 Appliance-Level Energy Decomposition

The dataset consists of three measurements at the appliance level: Sub_metering_1, Sub_metering_2, and Sub_metering_3. Used to represent the amount of energy being consumed by the major household appliances, and the "Other" aspect is calculated to complete the overall picture of energy consumption.

Power Decomposition Formula:

$$P_{global} = P^1 + P^2 + P^3 + P_{other} \quad (2)$$

where P_{global} is the total active power and P_1, P_2, P_3 represent the three sub-metered loads. This decomposition facilitates appliance-level analysis and enhances the interpretation of the forecasting results.

3.3 Sequence Construction and Problem Formulation

To convert time-series data into a supervised learning format (e.g., to enable the use of court-ordered methods of learning and build a supervised model of energy use), we utilise a sliding

window approach. A certain number of previous data points are collected together (as input) and used to predict the ongoing energy consumption (i.e., what the next set of consumption data will be). By using historical data to predict what will happen after each period, we have created a set of input-output pairs (or training data) for our model, which helps us develop the best possible model. As a result, when we change the duration of the (fixed-size) Input Window, we change the amount of historical data available for forecasting consumption, extending it as far back as possible in the (rolling/overlapping) historical data set. Similarly, the length of the desired prediction will determine how many future time periods will be forecast (i.e. applied for a direct multi-step forecast). Adjusting both the input window and the prediction horizon will allow us to switch from hourly forecasting in our operational models to daily or weekly (longer time frame) forecasts from the predictive models we developed. So once we develop the appropriate input-output pairs for the model, it can generate multiple future values simultaneously, thereby minimising forecast errors over long time frames as per the Fig 1.

3.4 Model Architecture

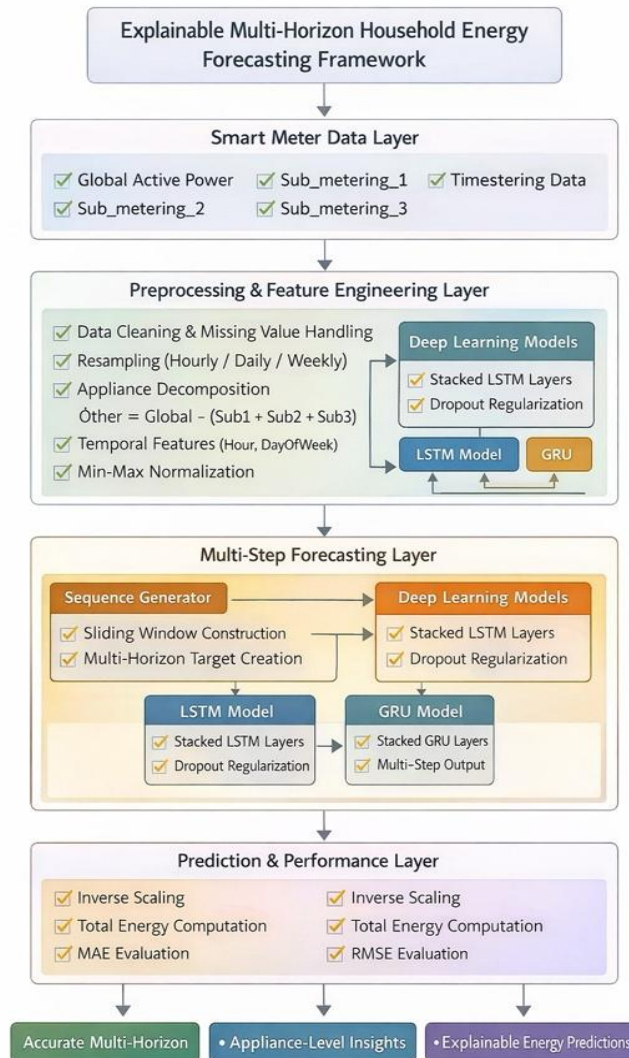


Figure 1: System Architecture of the Proposed Explainable Energy Forecasting Model

3.5 Model Training Strategy

Both LSTM and GRU models are trained using mini-batch gradient descent. The batch size and number of epochs are set using experience.

Early stopping is used to prevent overfitting by monitoring the validation loss. Learning rate reduction is also used when performance does not improve.

312

The same training parameters are used for all models.

4 Simulation Results

4.1 Experimental setup

All experiments are conducted on the household electricity consumption dataset using identical training and testing configurations. The preprocessed data is resampled to hourly, daily, and weekly resolutions to evaluate multi-scale forecasting performance. The dataset is divided into training and testing sets using a time-based split to preserve temporal dependencies. Both LSTM and GRU models are trained using the Adam optimiser with mean squared error as the loss function. Early stopping and learning rate reduction are used to limit overfitting. To evaluate how well the model works, we will use MAE and RMSE to assess total usage and individual loads for each appliance. The detailed numerical results for each experiment are presented in Tables 1-3. I used Python, Tensorflow and Sklearn on the same computer system for all conducted tests:

4.2 Comparison Evaluation

This subsection analyses how well our models perform across different time frames.

Table 1. Performance of LSTM Model at Different Time Scales (MAE)

Scale	Total MAE	Sub1 MAE	Sub2 MAE	Sub3 MAE	Other MAE
Hourly	0.5269	0.0971	0.1026	0.2556	0.2688
Daily	9.0258	1.1471	1.5677	3.3131	5.2273
Weekly	72.7755	6.4949	6.9884	26.4883	37.2072

Table 1 provides MAE for shifting hourly, daily, and weekly forecasts using LSTM Models. Forecasting accuracy decreased as the forecast horizon increased, resulting in significantly higher weekly errors than daily or hourly. Variance is highest for Sub_metering_3 and Other Load components; therefore, they have lower-than-average forecast accuracy.

Table 2: Performance of GRU Model at Different Time Scales (MAE)

Scale	Total MAE	Sub1 MAE	Sub2 MAE	Sub3 MAE	Other MAE
Hourly	0.4795	0.0982	0.1043	0.2501	0.2389
Daily	7.4330	1.2240	1.5984	2.8061	3.9877

Weekly	63.9767	5.8390	5.8622	23.7745	33.6882
--------	---------	--------	--------	---------	---------

Table 2 summarises the GRU model's performance. Compared to LSTM, the GRU model achieves lower total MAE values across all temporal resolutions, particularly at daily and weekly scales, demonstrating better generalisation capability for long-term forecasting.

Table 3: Comparison of LSTM and GRU Models Across Time Scales (MAE)

Model	Scale	Total MAE	Sub1 MAE	Sub2 MAE	Sub3 MAE	Other MAE
LSTM	Hourly	0.7189	0.1830	0.1026	0.2566	0.2688
LSTM	Daily	9.0258	1.1471	1.5677	3.3131	5.2273
LSTM	Weekly	72.7755	6.4949	6.9884	26.4883	37.2072
GRU	Hourly	0.4795	0.0982	0.1043	0.2501	0.2389
GRU	Daily	7.4330	1.2240	1.5984	2.8061	3.9877
GRU	Weekly	63.9767	5.8390	5.8622	23.7745	33.6882

Table 3 summarises the comparative performance of the LSTM and GRU models across all temporal scales.

4.3 Interpretability

The information suggests that temporal characteristics, especially hour and day of the week, significantly determine short-term energy prediction accuracy and are therefore very important for estimating daily energy use. Of the three appliance components, Sub_metering_3 and Other have a much larger positive effect on predicting energy than minor appliances do at the hourly scale.

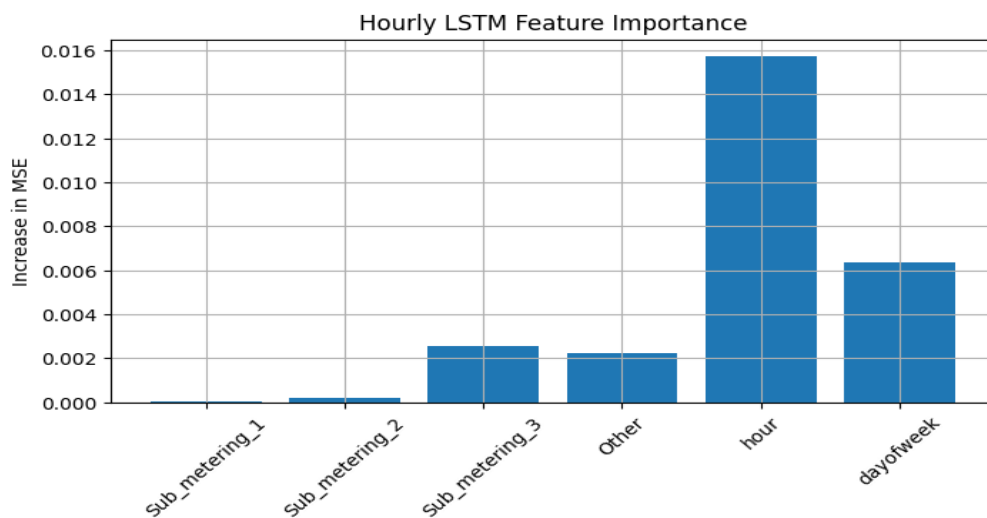


Figure 2. Hourly LSTM Permutation Feature Importance

The information suggests that temporal characteristics, especially hour and day of the week, significantly determine short-term energy prediction accuracy and are therefore very important for estimating daily energy use. Of the three appliance components, Sub_metering₃ and Other have a much larger positive effect on predicting energy than minor appliances do at the hourly scale as per Figure 2.

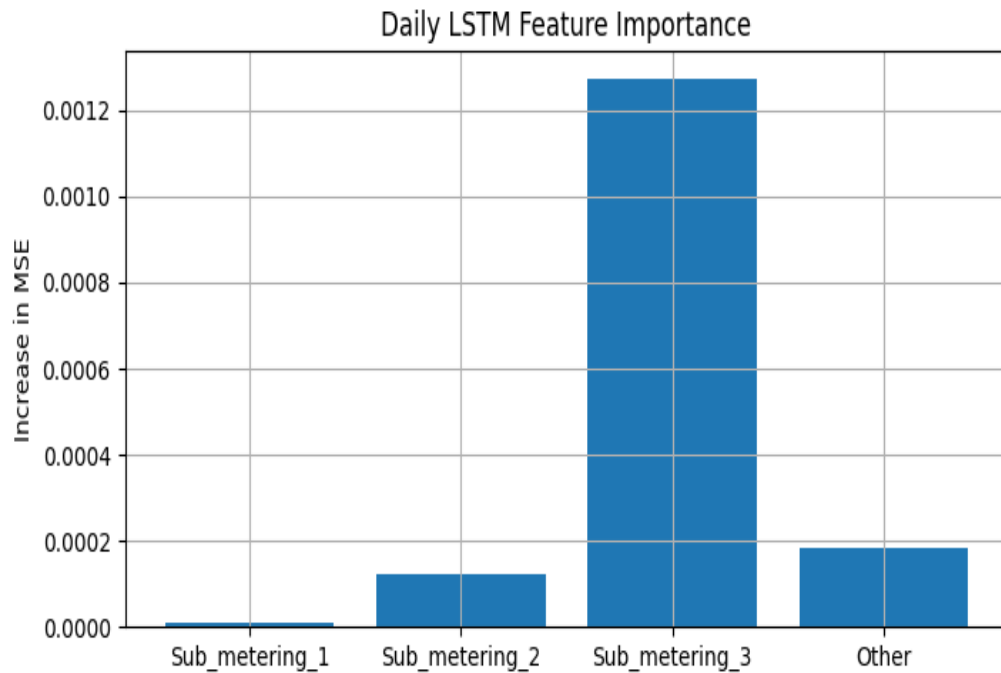


Figure 3. Daily LSTM Permutation Feature Importance

At the daily scale, Sub_metering₃ emerges as the most influential feature, followed by Other loads. Sub_metering₁ and Sub_metering₂ are comparatively less important as per Figure 3.

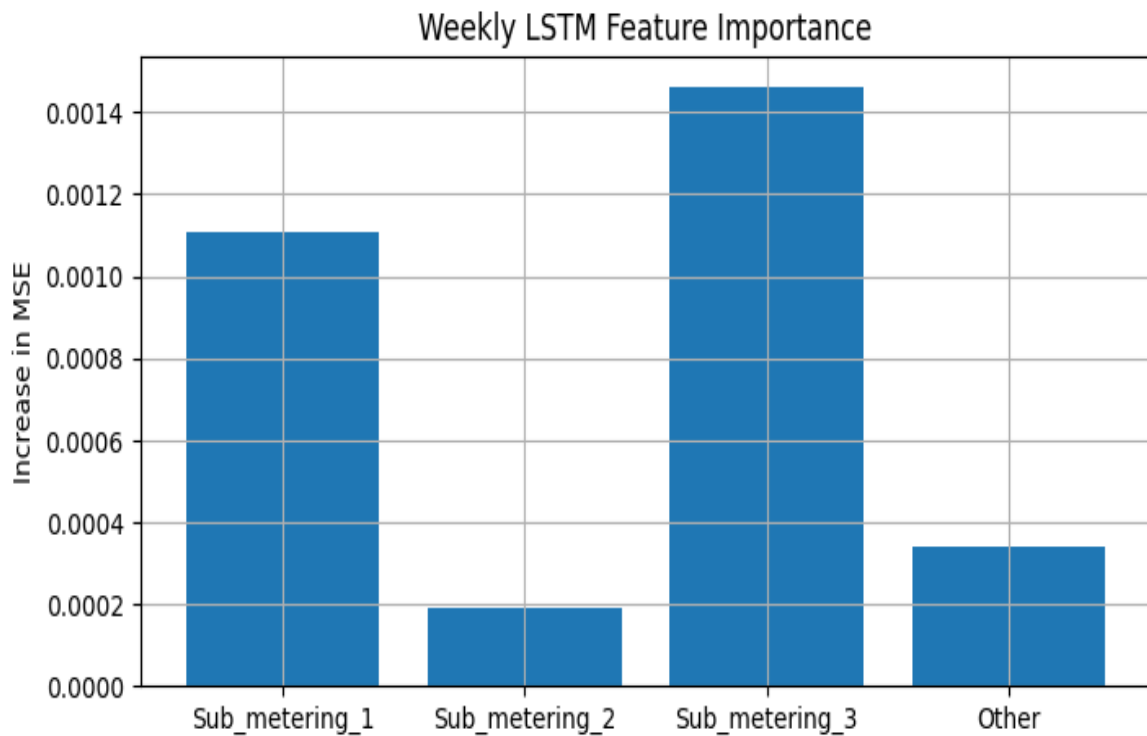


Figure 4. Weekly LSTM Permutation Feature Importance

This plot shows that, month by month, Sub_metering_3 contributes the most to the overall forecast, and Sub_metering_1 and Other also affect it to some extent. The impact of small appliances appears negligible in both short- and long-range predictions as per Figure 4.

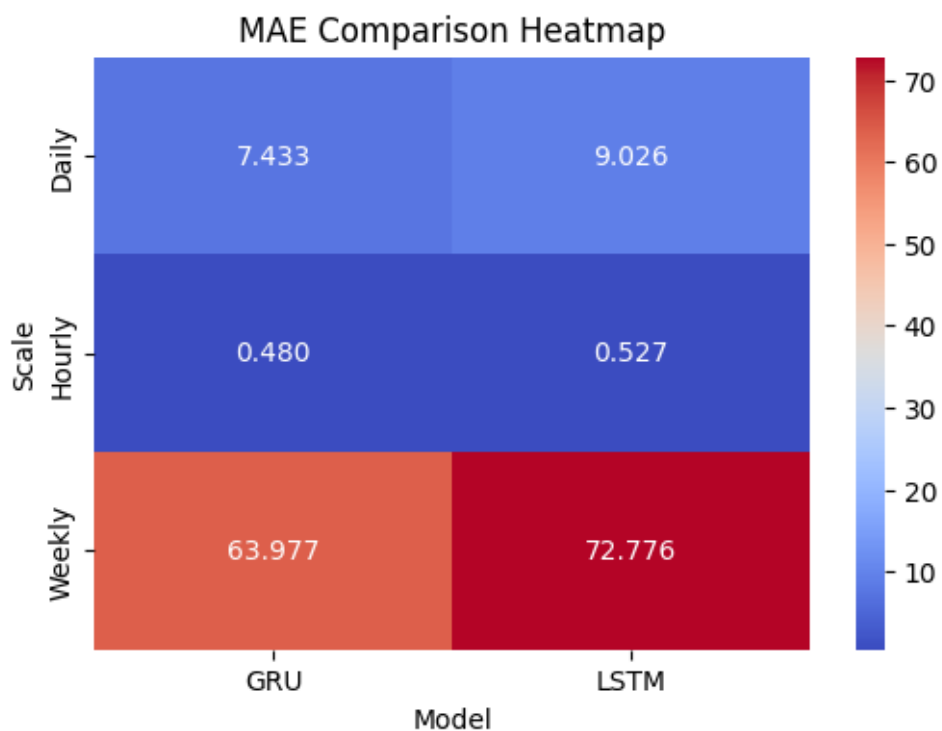


Figure 5. MAE Comparison Heatmap for LSTM and GRU Models

This figure shows the Mean Absolute Error (MAE) for LSTM and GRU across hourly, daily, and weekly time intervals. The results show that GRU consistently has lower prediction error than LSTM, regardless of whether the model is evaluated hourly, daily, or weekly as per Figure 16.

5 Conclusion

This study proposes an explainable multi-horizon deep-learning framework for forecasting household energy consumption. The proposed framework combines existing techniques (data preprocessing, appliance-level disaggregation/decomposition, and sequence-to-sequence modelling with LSTMs and GRUs) to predict household energy consumption at hourly, daily, and weekly intervals. The experimental results demonstrate that the GRU outperformed the LSTM in predicting household energy usage across all time periods. The researchers use permutation-based feature importance and SHAP analysis to improve the interpretability of their models by determining how much individual appliances and time-based features contribute to the models' predictions. Based on the results, the proposed framework is effective in providing accurate and transparent energy forecasts for intelligent energy management. Future studies will focus on adding contextual features to this framework and expanding its capability to create lightweight, real-time, adaptive forecasts. The multiple measurement periods in the proposed framework provide additional verification of its strength and adaptability, offering an option for different future time frames when generating forecasts.

The main contribution of this work is the development of scalable and reliable intelligent energy forecasting systems. The proposed system's modularity allows for easy integration into demand-side management applications and smart grid systems. An opportunity to extend this research is to add further weather inputs, room usage prediction, and social behaviour patterns, thereby improving model performance and generalisation. Further optimisations using lightweight methods will develop efficient implementations on edge-enabled smart homes.

6 References

- [1] Assaf, R., Schumann, A., 2019. Explainable deep neural networks for multivariate time series predictions. In: IJCAI. pp. 6488–6490. Barredo Arrieta, A., Gil-López, S., Laña, I., Bilbao, M.N., Del Ser, J., 2022. On the post-hoc explainability of deep echo state networks for time series forecasting, image, and video classification. *Neural Comput. Appl.* 34 (13), 10257–10277.
- [2] Bettini, C., Civitarese, G., Fiori, M., 2021. Explainable activity recognition over interpretable models. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops). IEEE, pp. 32–37.
- [3] Chadoulos, S., Koutsopoulos, I., Polyzos, G.C., 2021. One model fits all: Individualised household energy demand forecasting with a single deep learning model. In: Proceedings of the Twelfth ACM International Conference on Future Energy Systems. pp. 466–474.
- [4] Chakraborty, D., Alam, A., Chaudhuri, S., Başağaoğlu, H., Sulbaran, T., Langar, S., 2021. Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence. *Appl. Energy* 291, 116807. Crabbé, J., van der Schaar, M., Supplementary materials for explaining time series predictions with dynamic masks. Crabbé, J., Van Der Schaar, M., 2021. Explaining time series predictions with dynamic masks. In: International Conference on Machine Learning.

PMLR, pp. 2166–2177.

- [6] Dalvi-Esfahani, M., Mosharaf-Dehkordi, M., Leong, L.W., Ramayah, T., Kanaan Jebna, A.M.J., 2023. Exploring the drivers of XAI-enhanced clinical decision support systems adoption: Insights from a stimulus-organism-response perspective. *Technol. Forecast. Soc. Change* 195, 122768.
- [7] Das, D., Nishimura, Y., Vivek, R.P., Takeda, N., Fish, S.T., Ploetz, T., Chernova, S., 2021. Explainable activity recognition for smart home systems. *arXiv preprint arXiv:2105.09787*.
- [8] Došilović, F.K., Brčić, M., Hlupić, N., 2018. Explainable artificial intelligence: A survey. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics.
- [9] MIPRO, IEEE, pp. 0210–0215. Efat, M.I.A., Hajek, P., Abedin, M.Z., Azad, R.U., Jaber, M.A., Aditya, S., Hassan, M.K., 2022. Deep-learning model using a hybrid adaptive trend-estimated series for modelling and forecasting sales. *Ann. Oper. Res.* 1–32. Ehsan, U., Wintersberger, P., Liao, Q.V., Mara, M., Streit, M., Wachter, S., Riener, A., Riedl, M.O., 2021.
- [10] Adadi, A., Berrada, M., 2020. Explainable AI for healthcare: from black box to interpretable models. In: *Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco*. Springer, pp. 327–337.
- [11] Ahmad, J., Tahir, A., Larijani, H., Ahmed, F., Aziz Shah, S., Hall, A.J., Buchanan, W.J., 2020. Energy demand forecasting of buildings using random neural networks. *J. Intell. Fuzzy Systems* 38 (4), 4753–4765.
- [12] Ali, U., Shamsi, M.H., Hoare, C., Mangina, E., O'Donnell, J., 2021. Review of urban building energy modelling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy Build.* 246, 111073.
- [13] Arrotta, L., 2021. Multi-inhabitant and explainable activity recognition in smart homes. In: 2021 22nd IEEE International Conference on Mobile Data Management. MDM, IEEE, pp. 264–266.
- [14] Kabir, M.H., Hasan, K.F., Hasan, M.K., Ansari, K., 2021. Explainable artificial intelligence for smart city application: A secure and trusted platform. *arXiv preprint arXiv:2111.00601*.
- [15] Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., Rebholz Schuhmann, D., Decker, S., 2023. Explainable AI for Bioinformatics: Methods, Tools, and Applications. *Brief. Bioinform.* 24 (5), bbad236.
- [16] Karim, M.R., Dey, S.K., Islam, T., Sarker, S., Menon, M.H., Hossain, K., Hossain, M.A., Decker, S., 2021. Deep Hate Speech Explainer: Explainable hate speech detection in the under-resourced Bengali language. In: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics. DSAA, IEEE, pp. 1–10.
- [17] Shajalal, M., Boden, A., Stevens, G., 2022a. Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses. *Electron. Mark.* 32 (4), 2107–2122.
- [18] Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In: *International Conference on Machine Learning*. PMLR, pp. 3145–3153.

- [19] Vanting, N.B., Ma, Z., Jørgensen, B.N., 2021. A scoping review of deep neural networks for electric load forecasting. *Energy Inform.* 4 (2), 1–13.
- [20] Mucha, H., Robert, S., Breitschwerdt, R., Fellmann, M., 2021. Interfaces for explanations in human-AI interaction: Proposing a design evaluation approach. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. pp. 1–6.
- [21] Saadallah, A., Jakobs, M., Morik, K., 2021. Explainable online deep neural network selection using adaptive saliency maps for time series forecasting. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 404–420.
- [22] Saluja, R., Malhi, A., Knapič, S., Främling, K., Cavdar, C., 2021. Towards a rigorous evaluation of explainability for multivariate time series. *arXiv preprint arXiv: 2104.04075*.
- [23] Sharma, H., Kumar, P. and Sharma, K., 2025. Intelligent Time Series Analysis for Intrusion Detection in the Internet of Things: A Generative-Adversarial-Network-Enhanced Convolutional-Neural-Network–Long-Short-Term-Memory Framework Using Signal Features. *Intelligent Computing*, 4, p.0127.
- [24] Schwartz, T., Denef, S., Stevens, G., Ramirez, L., Wulf, V., 2013a. Cultivating energy literacy: results from a longitudinal living lab study of a home energy management system. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1193–1202.
- [25] Kim, D., Song, Y., Kim, S., Lee, S., Wu, Y., Shin, J., Lee, D., 2023a. How should the results of artificial intelligence be explained to users?-Research on consumer preferences in user-centred explainable artificial intelligence. *Technol. Forecast. Soc—change* 188, 122343.
- [26] Kim, J., Lee, G., Lee, S., Lee, C., 2022. Towards expert–machine collaborations for technology valuation: An interpretable machine learning approach. *Technol. Forecast. Soc—change* 183, 121940.
- [27] Zhang, W., Liu, F., Wen, Y., Nee, B., 2021. Toward explainable and interpretable building energy modelling: an explainable artificial intelligence approach. In: *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. pp. 255–258.
- [28] Zdravković, M., Ćirić, I., Ignjatović, M., 2022. Explainable heat demand forecasting for the novel control strategies of district heating systems. *Annu. Rev. Control*.
- [29] Khodabandehloo, E., Riboni, D., Alimohammadi, A., 2021. HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline. *Future Gener. Comput. Syst.* 116, 168–189.