26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

# Interpretable deep learning for energy forecasting and outlier detection

**Joshua Chopin** [a]

[a] *Industrial AI Research Centre, University of South Australia, Mawson Lakes, 5095, Australia*
*Email: josh.chopin@unisa.edu.au*

**Abstract:** The increasing complexity of modern energy markets, driven by fluctuating demand, renewable integration, and weather variability, presents significant challenges for accurate forecasting and anomaly detection. In recent years machine learning techniques have presented as a popular tool for overcoming some of the challenges involved with these tasks. However, while they are able to process large amounts of data and learn complicated relationships between variables, the outputs of machine learning models are often opaque at best. In real world decision-making scenarios, where trustworthiness of results is imperative, modern machine learning methods, like deep learning, are sometimes overlooked for more traditional, interpretable approaches. This work uses a publicly available dataset of Spanish energy consumption, generation, pricing, and meteorological conditions to investigate the use of deep learning methods for both predictive modelling and outlier detection, with a particular focus on explaining these predictions.

We develop and evaluate a range of deep learning architectures for short-term energy prediction and outlier detection. Post-hoc explanation methods such as SHAP (Lundberg & Lee 2017) are used to explain the predictions and classifications for each model with a focus on model trustworthiness. When the same model architecture and parameter selections lead to different predictions due to different random initialisations, which do we trust? How much can we trust them? And what can we say about the sensitivity of these methods? We will systematically evaluate how model explanations differ across different random initialisations, different parameter choices and different architectures altogether.

By contrasting multiple explainability techniques against multiple model architectures and characteristics, we highlight differences in prediction quality, stability, and interpretive usefulness. Our contributions are threefold: an empirical assessment of deep learning models for Spanish energy forecasting using a comprehensive real-world dataset; an interpretable framework for outlier detection in high-dimensional energy time series; and a comparative analysis of explanation methods that underscores the importance of transparency for decision-making in critical infrastructure applications. Results show that interpretability should not be treated as an after-thought but as a central requirement for deploying trustworthy machine learning in energy systems.

## REFERENCES

[1] Lundberg S & Lee S (2017) A unified approach to interpreting model predictions. NeurIPS 2017.

***Keywords:*** *Deep Learning, Forecasting, Outlier Detection, SHAP, Renewable, Electricity*