engineering proceedings

MDPI

Proceeding Paper

Short-Horizon Grid Frequency Forecasting Using Hybrid Machine Learning Models with a Frequency-Aware Demand–Response Action Layer †

Elohor Ekurume *, Triantafyllos Kanakis and Michael Opoku-Agyeman

Department of Computer Science, University of Northampton, Northampton NN1 5PH, UK; triantafyllos.kanakis@northampton.ac.uk (T.K.); michael.opoku-agyeman@northampton.ac.uk (M.O.-A.)

* Correspondence: elohorekurume@gmail.com

† Presented at the 8th International Conference on Nuclear and Renewable Energy Resources (NURER 2026), Almaty, Kazakhstan, 10–12 September 2026.

Abstract

Maintaining grid frequency stability is increasingly challenging in low-inertia power systems due to the growing penetration of renewable energy sources. Conventional control mechanisms such as primary frequency response and under-frequency load shedding (UFLS) operate reactively, limiting their effectiveness under rapid system dynamics. This study proposes a unified forecasting-to-control framework that integrates short-horizon grid frequency prediction with a frequency-aware demand–response (DR) action layer for proactive stabilisation. The proposed approach combines XGBoost, CNN–LSTM, and Transformer models within a stacking architecture, where an XGBoost meta-learner generates the final forecast. The model is trained on high-resolution UK grid frequency data resampled to 10 s intervals and enriched with domain-informed features, including frequency deviation, rate of change of frequency (RoCoF), rolling statistics, and cyclical time encoding. A 5 min lookback window is used to predict frequency 30 s ahead. Results show that XGBoost achieves the best standalone performance (R^2 = 0.9264), while the stacking ensemble provides comparable accuracy and improved robustness. Transformer predictions contribute the highest importance within the ensemble, highlighting their complementary value. The forecast is mapped to operational DR actions, enabling approximately 30 s of lead time. The results demonstrate the potential of integrating forecasting with control for intelligent grid management.

Keywords: grid frequency forecasting; demand response; low-inertia power systems; stacking ensemble; XGBoost; Transformer; CNN–LSTM; short-horizon prediction; smart grid control; uncertainty-aware forecasting

Check for updates

Academic Editors: Sümer Şahin, Hacı Mehmet Şahin and Bekir Dursun

Published: 26 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Power system frequency reflects the instantaneous balance between electricity generation and demand and is a critical indicator of grid stability [1]. In modern power systems, the increasing penetration of inverter-based renewable energy sources has significantly reduced system inertia, leading to faster and more volatile frequency dynamics [2–4]. Under such low-inertia conditions, disturbances propagate rapidly, reducing the time available for corrective actions and increasing the risk of instability.

Conventional control mechanisms, including primary frequency response and under-frequency load shedding (UFLS), operate in a reactive manner, triggering corrective actions

only after frequency deviations occur. While effective in preventing system collapse, these approaches provide limited response time and may result in excessive or unnecessary load disconnections [5]. As grid dynamics become increasingly stochastic and nonlinear, there is a growing need for predictive approaches that enable proactive intervention.

Recent advances in smart grid technologies and demand-side flexibility have introduced demand response (DR) as a viable mechanism for frequency control. By dynamically adjusting controllable loads such as electric vehicles, heating systems, and household appliances, DR can provide fast and scalable support for grid stability [6,7]. However, most existing DR frameworks remain reactive, relying on real-time measurements rather than anticipating future system conditions.

Machine learning techniques have emerged as promising tools for short-horizon grid frequency forecasting. Tree-based models such as XGBoost are effective for structured feature learning [8], while deep learning architectures such as LSTM and CNN-LSTM can capture temporal dependencies in sequential data [9,10]. More recently, Transformer-based models have demonstrated strong capability in modelling complex temporal relationships through attention mechanisms [11]. Despite these advances, no single model consistently outperforms others across all operating conditions, motivating the use of ensemble approaches.

Despite this progress, a clear gap remains in the literature. Existing frequency-forecasting studies predominantly optimise standalone predictive accuracy without establishing an explicit, operational link between forecast outputs and control actions [12–14]. Conversely, demand–response frameworks are typically reactive, responding only after a frequency deviation is measured rather than anticipating it [6,7,15,16]. Moreover, although hybrid ensembles combining tree-based and deep learning models have improved robustness in general energy-forecasting tasks [17], their application to sub-minute, short-horizon grid-frequency prediction that is directly coupled to an actionable demand–response decision layer has not been systematically demonstrated. This absence of a validated, unified forecasting-to-control pipeline for short-horizon frequency management in low-inertia grids is the specific research gap addressed in this study.

This study proposes a unified forecasting-to-control framework that integrates short-horizon grid frequency prediction with a frequency-aware demand–response action layer. A stacking ensemble is developed by combining XGBoost, CNN-LSTM, and Transformer models, with an XGBoost meta-learner producing the final forecast. Unlike existing work that focuses primarily on predictive accuracy, the proposed approach explicitly links forecasting outputs to operational decision-making, enabling proactive intervention approximately 30 s ahead of system disturbances.

The main contributions of this paper are as follows:

- Development of a hybrid stacking framework combining tree-based and deep learning models for short-horizon grid frequency forecasting;
- Demonstration of complementary learning, where models with lower standalone accuracy contribute to improved ensemble robustness;
- Integration of forecasting with a frequency-aware demand–response action layer for proactive control;
- Empirical analysis of ensemble behaviour using meta-learner feature importance;
- Validation of operational feasibility through accurate classification of grid states with actionable lead time.

The remainder of this paper is organised as follows. Section 2 reviews related work, Section 3 presents the proposed methodology, Section 4 reports experimental results and Section 5 concludes the paper.

2. Related Work

Short-horizon grid frequency forecasting is essential for maintaining stability in modern low-inertia power systems. The literature reviewed below is organised methodologically, progressing from classical statistical models, through machine learning and deep learning approaches, to Transformer-based and ensemble methods before turning to demand–response and forecast-to-control integration, which motivates the present work.

2.1. Statistical and Model-Based Forecasting Approaches

Traditional approaches model frequency dynamics using statistical techniques such as autoregressive integrated moving average (ARIMA) and Kalman filtering [18]. While effective under stationary conditions, these methods struggle to capture nonlinear behaviour and rapid transient dynamics associated with disturbances, motivating a shift toward data-driven approaches. Related machine-learning work on short-term kinetic-energy forecasting further demonstrates the value of data-driven prediction in low-inertia systems [19].

2.2. Machine Learning and Deep Learning Approaches

Machine learning models such as XGBoost have demonstrated strong performance on structured feature representations [8]. Similarly, recurrent neural networks, particularly LSTM-based architectures, have shown effectiveness in modelling temporal dependencies in sequential data [9], with CNN-LSTM hybrids further improving the extraction of local temporal patterns [10]. More recent studies extend these approaches to very short-term forecasting and frequency nadir prediction, highlighting the growing role of machine learning in real-time grid stability analysis [12,13,20].

2.3. Transformer-Based and Ensemble Forecasting Methods

Transformer architectures have recently been adapted for time-series forecasting due to their ability to capture long-range dependencies using self-attention mechanisms [11]. Models such as Informer and Autoformer introduce efficient attention and decomposition techniques for temporal modelling [21,22]. While these approaches have shown strong performance in applications such as load forecasting, their advantages in short-horizon, low-latency frequency prediction remain less consistent [14]. In parallel, ensemble learning methods, particularly stacked generalisation, have been widely adopted to improve predictive performance by combining heterogeneous models [23]. Hybrid approaches integrating tree-based and deep learning models have demonstrated improved robustness in energy forecasting tasks, although gains are often limited when base models share similar input representations [17].

2.4. Demand–Response and Forecast-to-Control Integration

On the control side, DR has emerged as a key mechanism for enhancing grid flexibility by enabling dynamic adjustment of electricity consumption [6,7]. Conventional UFLS schemes remain essential for system protection but operate reactively based on predefined thresholds [5]. Recent research has explored adaptive and intelligent control strategies, including machine learning-based approaches for demand response [15,16]. However, most existing studies treat forecasting and control as separate problems. This work addresses this gap by integrating short-horizon forecasting with a decision-making layer, enabling proactive demand–response actions based on predicted system conditions.

https://doi.org/10.3390/engproc2026160029

3. Proposed Framework

Figure 1 (Claude and ChatGPT assisted with drawing the system architecture diagram in Figure 1) illustrates the overall architecture of the proposed hybrid AI framework. The pipeline comprises four stages: (i) acquisition of grid frequency and related system data; (ii) data preprocessing and feature engineering; (iii) a hybrid stacked ensemble forecasting model combining XGBoost, CNN-LSTM, and Transformer base learners with an XGBoost meta-learner; and (iv) a decision-support module that converts the resulting forecast into demand–response recommendations. Each stage is described in the following subsections.

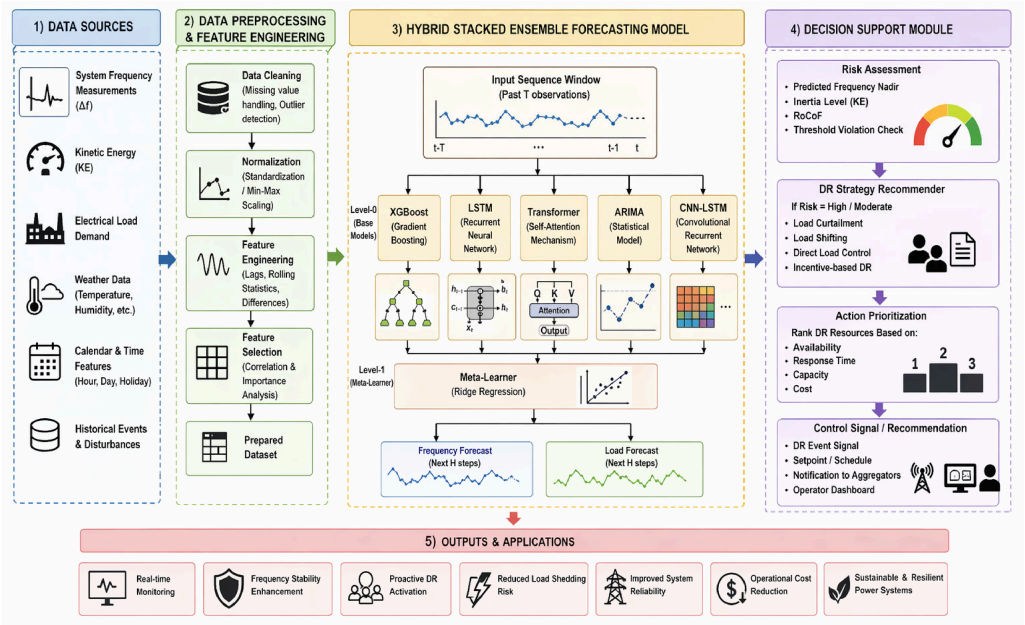


Figure 1. Overall architecture of the proposed hybrid AI framework for short-term frequency and load forecasting with DR recommendations. The colours are different to show the different stages.

3.1. Data and Problem Formulation

This study uses high-resolution grid frequency data from the UK National Grid ESO, originally sampled at 1 s resolution. To balance computational efficiency with temporal fidelity, the data are resampled to 10 s intervals using averaging, resulting in 267,840 observations for December 2019.

The forecasting task is formulated as a supervised learning problem. Given a sequence of past observations over a lookback window L, the objective is to predict the future frequency value at horizon H:

$$\hat{f}_{t+H} = \mathcal{F}(\mathbf{X}_t), \tag{1}$$

where $\mathbf{X}_t \in \mathbb{R}^{L \times d}$ represents the input feature sequence. In this work, L = 30 (5 min) and H = 3 (30 s ahead).

3.2. Feature Engineering

To capture system dynamics, the raw frequency signal is enriched with domain-informed features. Frequency deviation is computed relative to the nominal value

$$\Delta f_t = f_t - 50 \tag{2}$$

and the rate of change of frequency (RoCoF) is approximated using first-order differencing:

$$\mathrm{RoCoF}_t = (f_t - f_{t-1})/dt. \tag{3}$$

Short-term temporal behaviour is captured using rolling statistics computed over window sizes $w \in \{3,6,12\}$:

$$\mu_t^{(w)} = \frac{1}{w}\sum_{i=0}^{w-1} f_{t-i}, \qquad \sigma_t^{(w)} = \sqrt{\frac{1}{w}\sum_{i=0}^{w-1}\left(f_{t-i} - \mu_t^{(w)}\right)^2}. \tag{4}$$

Cyclical time encoding is applied to represent periodic patterns:

$$x_{\sin} = \sin\left(\frac{2\pi h}{T}\right), \qquad x_{\cos} = \cos\left(\frac{2\pi h}{T}\right). \tag{5}$$

All features are normalised using Min–Max scaling.

3.3. Hybrid Forecasting Model

The proposed framework integrates three heterogeneous base learners.

XGBoost: a tree-based ensemble model that captures nonlinear relationships in structured data:

$$\hat{y} = \sum_{k=1}^{K} f_k(\mathbf{x}). \tag{6}$$

CNN-LSTM: a hybrid deep learning model that extracts local temporal patterns via convolution and models sequential dependencies using LSTM layers.

Transformer: a self-attention-based model that captures global temporal dependencies:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{7}$$

3.4. Model Architectures and Training Configuration

To make the architectures and training configuration of each model fully reproducible, Table 1 summarises the layer structure, hyperparameters, and optimisation settings used for the three base learners and the meta-learner, as implemented in the accompanying training pipeline.

Table 1. Architecture, hyperparameters, and optimisation configuration of the base learners and meta-learner. Data are split 80/20 (train/test) in chronological order, with random_state = 42 used throughout for reproducibility.

Component	Architecture/Layers	Key Hyperparameters	Optimiser & Training
XGBoost (base)	Gradient-boosted decision trees on engineered tabular features	n_estimators = 150; max_depth = 5; learning_rate = 0.05; subsample = 0.8; colsample_bytree = 0.8	Objective: squared error; random_state = 42
CNN–LSTM (base)	Conv1D(32,k = 3)–BN–MaxPool(2) → Conv1D(64,k = 3)–BN–MaxPool(2) → LSTM(32, return_sequences)–Dropout(0.2) → LSTM(16)–Dropout(0.2) → Dense(16, ReLU) → Dense(1)	Lookback L = 30 (5 min); batch size = 128; epochs = 12	Adam optimiser; MSE loss; EarlyStopping (patience = 4) and ReduceLROnPlateau (factor = 0.5, patience = 2) on validation loss

Table 1. *Cont.*

Component	Architecture/Layers	Key Hyperparameters	Optimiser & Training
Transformer (base)	Dense(32) input projection + sinusoidal positional encoding → 2× encoder blocks (multi-head self-attention: heads = 2, key_dim = 16; feed-forward dim = 32; dropout = 0.1; residual + layer norm) → GlobalAveragePooling1D → Dense(32, ReLU) → Dropout(0.2) → Dense(1)	Lookback L = 30 (5 min); batch size = 128; epochs = 12	Adam optimiser; MSE loss; same EarlyStopping/ReduceLROnPlateau callbacks as CNN–LSTM
XGBoost (meta-learner)	Gradient-boosted trees over the 3 base-model predictions $\left[\hat{y}^{(1)}, \hat{y}^{(2)}, \hat{y}^{(3)}\right]$	n_estimators = 120; max_depth = 3; learning_rate = 0.05; subsample = 0.9; colsample_bytree = 0.9	Objective: squared error; random_state = 42

3.5. Stacking Ensemble

Predictions from the base models are combined using stacked generalisation. Let $\hat{y}^{(1)}, \hat{y}^{(2)}, \hat{y}^{(3)}$ denote outputs from XGBoost, CNN–LSTM, and Transformer respectively. These are used as inputs to a meta-learner:

$$\hat{y} = g\left(\left[\hat{y}^{(1)}, \hat{y}^{(2)}, \hat{y}^{(3)}\right]\right), \tag{8}$$

where $g(\cdot)$ is implemented using XGBoost, as detailed in Table 1.

3.6. Forecast-to-Control Action Layer

To enable operational decision-making, predicted frequency values are mapped to demand–response actions:

$$a\left(\hat{f}\right) = \begin{cases} \text{Critical}, & \hat{f} < 49.8 \\ \text{Warning}, & 49.8 \leq \hat{f} < 49.9 \\ \text{Normal}, & \hat{f} \geq 49.9 \end{cases} \tag{9}$$

These thresholds are not arbitrary: they are aligned with the operational frequency limits defined by the Great Britain system operator. National Grid ESO targets a normal operating range of 49.8–50.2 Hz, deploying frequency response and reserve services once frequency deviates beyond this band, while the wider statutory limit is 49.5–50.5 Hz [24,25]. As frequency approaches or breaches the 49.8 Hz lower operational bound, primary frequency response and, in more severe cases, low-frequency demand disconnection may be triggered [5]. The Critical and Warning thresholds adopted here (49.8 Hz and 49.9 Hz) therefore reflect established GB grid operational practice rather than arbitrary cut-offs, consistent with the operating margins reported for low-inertia GB scenarios [24]. This mapping enables proactive intervention approximately 30 s ahead of disturbances.

3.7. Evaluation Metrics

Model performance is evaluated using three standard regression metrics, each assigned its own equation number for clarity:

$$\text{MAE} = \frac{1}{N}\sum|y - \hat{y}|. \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum(y-\hat{y})^2}. \tag{11}$$

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}. \tag{12}$$

4. Results and Analysis

4.1. Forecasting Performance

The predictive performance of all models was evaluated using MAE, RMSE, and R^2 on the held-out test set. Table 2 summarises the results.

Table 2. Performance comparison of forecasting models on the held-out test set.

Model	MAE (Hz)	RMSE (Hz)	R^2
XGBoost	0.01295	0.01794	0.9264
Stacking Hybrid	0.01309	0.01819	0.9244
Transformer	0.01560	0.02126	0.8967
CNN–LSTM	0.01719	0.02294	0.8796

XGBoost achieves the strongest standalone performance across all metrics. The stacking hybrid performs comparably, with only marginal degradation in accuracy. The Transformer and CNN–LSTM exhibit lower performance, particularly during rapid fluctuations. Figure 2 compares the actual and predicted frequency trajectories for the first 500 test samples: as shown in Figure 2, XGBoost achieves the closest fit to the observed signal, while the stacking hybrid produces a smoother trajectory with comparable accuracy. The Transformer captures major trends but exhibits reduced precision during rapid fluctuations, whereas the CNN–LSTM shows noticeable lag and underestimation. Overall, the stacking hybrid closely follows the XGBoost trajectory while reducing local volatility, producing a more stable signal suitable for downstream control decisions.

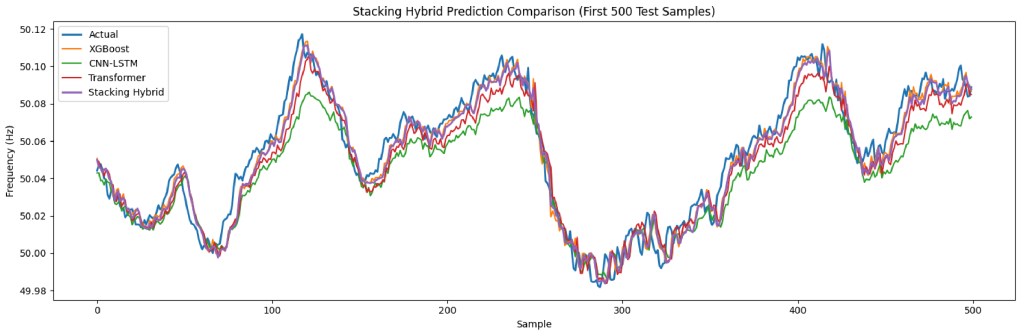


Figure 2. Comparison of actual and predicted grid frequency for the first 500 test samples. XGBoost achieves the closest fit to the observed signal, while the stacking hybrid produces a smoother trajectory with comparable accuracy.

4.2. Ensemble Behaviour

Although the stacking model does not outperform XGBoost, it provides enhanced robustness through complementary learning. The performance gap is minimal (MAE difference $< 1.5 \times 10^{-4}$ Hz), indicating that ensemble fusion preserves predictive strength while reducing reliance on a single model.

Figure 3 shows the feature importance of the XGBoost meta-learner based on the three base-model predictions. As illustrated in Figure 3, Transformer predictions contribute the highest importance (48.6%), followed by XGBoost (41.2%) and CNN–LSTM (10.2%). This indicates that Transformer outputs contain complementary temporal information not

fully captured by tree-based models, despite lower standalone accuracy—that is, the most influential model within the ensemble is not necessarily the best standalone predictor.

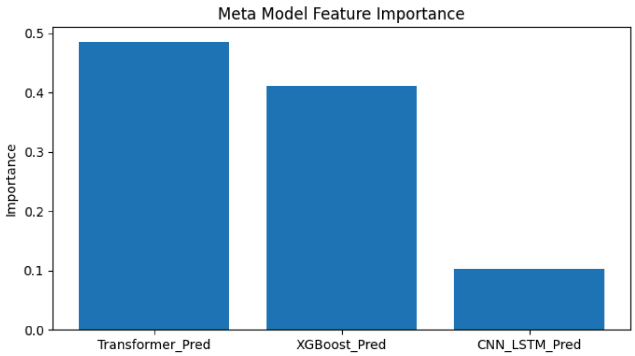


Figure 3. Feature importance of the XGBoost meta-learner based on base-model predictions. Transformer predictions contribute the highest importance, followed by XGBoost, while CNN–LSTM contributes least.

4.3. Operational Relevance

From a control perspective, prediction accuracy must be evaluated relative to decision thresholds rather than absolute error alone. The average prediction error is significantly smaller than the separation between action thresholds (0.1 Hz, per Section 3.6), ensuring reliable classification into operational states.

The forecasting-to-control pipeline enables approximately 30 s of lead time, which is sufficient for demand–response actions such as load shifting and peak shaving. This demonstrates that even modest improvements in short-horizon prediction can translate into meaningful operational benefits.

4.4. Uncertainty and Decision Reliability

Residual-based uncertainty estimation shows that true frequency values largely fall within the predicted confidence intervals, indicating reasonable calibration. Figure 4 presents the resulting probabilistic forecast with 95% confidence intervals: as shown in Figure 4, the predicted values closely follow the actual signal, and the confidence bands capture the majority of observations. This is particularly important near threshold boundaries, where uncertainty-aware forecasting can prevent premature or unnecessary control actions.

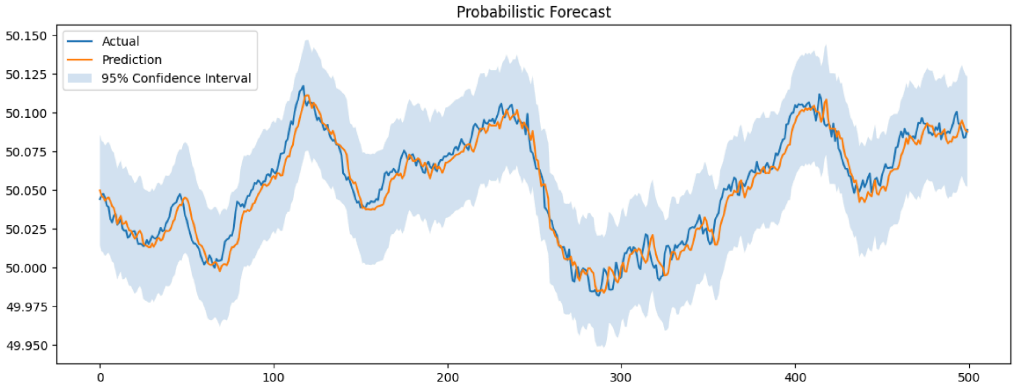


Figure 4. Probabilistic forecast of grid frequency with 95% confidence intervals. The predicted values closely follow the actual signal, while the confidence bands capture the majority of observations, indicating well-calibrated uncertainty estimation.

4.5. Key Insights

The confidence intervals in Figure 4 remain consistent across different operating conditions and successfully capture most true values, which is particularly valuable near decision thresholds, where uncertainty-aware predictions enhance the reliability of demand–response actions. Taken together with the results in Sections 4.1–4.4, the findings highlight several important points:

- Tree-based models remain highly effective for short-horizon forecasting with engineered features;
- Stacking improves robustness rather than raw accuracy;
- Transformer models provide valuable complementary information within ensembles;
- Predictive accuracy is sufficient for reliable demand–response decision support;
- Integrating forecasting with control provides practical value beyond standalone prediction.

5. Conclusions

This study presented a unified forecasting-to-control framework for short-horizon grid frequency management in low-inertia power systems. By integrating heterogeneous machine learning models like XGBoost, CNN-LSTM, and a Transformer within a stacking architecture (Table 1), the proposed approach achieves high predictive accuracy while leveraging complementary learning across models.

Experimental results (Table 2) demonstrate that XGBoost provides the strongest standalone performance, while the stacking ensemble maintains comparable accuracy and enhances robustness. Importantly, meta-learner analysis (Figure 3) reveals that Transformer predictions contribute the highest importance within the ensemble, highlighting the value of complementary temporal representations even when individual model performance is lower.

A key contribution of this work is the integration of forecasting with a frequency-aware demand–response action layer, whose thresholds are grounded in GB grid operational practice [5,24,25]. By mapping predicted frequency values into operational states (Normal, Warning, Critical), the framework enables proactive control with approximately 30 s of lead time. This capability is particularly significant in low-inertia systems, where rapid frequency deviations require fast and reliable intervention.

The inclusion of probabilistic forecasting (Figure 4) further strengthens the framework by providing uncertainty-aware decision support, improving reliability near operational thresholds. Collectively, these results demonstrate that short-horizon predictive models can move beyond passive forecasting to support real-time control strategies in modern power systems.

Future work will focus on incorporating additional system variables such as load, renewable generation, and weather data, as well as developing adaptive control policies using reinforcement learning. Extending the framework across longer time horizons and multiple grid environments will further validate its scalability and operational applicability.

Overall, this work contributes toward the development of intelligent, predictive, and control-oriented solutions for next-generation smart grids.

Author Contributions: Conceptualization, E.E.; methodology, E.E.; software, E.E.; validation, E.E.; formal analysis, E.E.; investigation, E.E.; data curation, E.E.; writing—original draft preparation, E.E.; writing—review and editing, E.E., T.K. and M.O.-A.; visualization, E.E.; supervision, T.K. and M.O.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The grid frequency data used in this study are publicly available from the UK National Grid ESO. The processed dataset, feature-engineering, training, and evaluation pipeline are available from the corresponding author on reasonable request.

Acknowledgments: During manuscript preparation, Claude (Opus 5.5), ChatGPT (6.1 Sol Light) and Grammarly were used for grammar correction, proofreading and improving wording. Claude and ChatGPT also assisted with drawing the system architecture diagram in Figure 1. The authors take full responsibility for the accuracy and integrity of the manuscript, including its text and figures.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DR—Demand Response; UFLS—Under-Frequency Load Shedding; RoCoF—Rate of Change of Frequency; LSTM—Long Short-Term Memory; CNN—Convolutional Neural Network; MAE—Mean Absolute Error; RMSE—Root Mean Squared Error; R^2—Coefficient of Determination; GB—Great Britain; ESO—Electricity System Operator.

References

1. Kundur, P. *Power System Stability and Control*; McGraw-Hill: New York, NY, USA, 1994.
2. Milano, F.; Dörfler, F.; Hug, G.; Hill, D.J.; Verbič, G. Foundations and challenges of low-inertia systems. In Proceedings of the Power Systems Computation Conference (PSCC), Dublin, Ireland, 11–15 June 2018; pp. 1–25. [CrossRef]
3. Homan, S.; Mac Dowell, N.; Brown, S. Grid frequency volatility in future low inertia scenarios: Challenges and mitigation options. *Appl. Energy* **2021**, *290*, 116723. [CrossRef]
4. Cheng, Y.; Azizipanah-Abarghooee, R.; Azizi, S.; Ding, L.; Terzija, V. Smart frequency control in low inertia energy systems based on frequency response techniques: A review. *Appl. Energy* **2020**, *279*, 115798. [CrossRef]
5. Alhelou, H.H.; Golshan, M.E.H.; Njenda, T.C.; Hatziargyriou, N.D. An overview of UFLS in conventional, modern, and future smart power systems: Challenges and opportunities. *Electr. Power Syst. Res.* **2020**, *179*, 106054. [CrossRef]
6. Palensky, P.; Dietrich, D. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Trans. Ind. Inform.* **2011**, *7*, 381–388. [CrossRef]
7. Siano, P. Demand response and smart grids—A survey. *Renew. Sustain. Energy Rev.* **2014**, *30*, 461–478. [CrossRef]
8. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
9. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
10. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*; NeurIPS: San Diego, CA, USA, 2017; pp. 5998–6008.
12. Yurdakul, O.; Eser, F.; Sivrikaya, F.; Albayrak, S. Very short-term power system frequency forecasting. *IEEE Access* **2020**, *8*, 141234–141245. [CrossRef]
13. Jiménez-Navarro, M.J.; Riquelme, J.M.; Carranza-García, M.; González-Longatt, F.M. A real-time machine learning-based methodology for short-term frequency nadir prediction in low-inertia power systems. *Neurocomputing* **2025**, *626*, 129583. [CrossRef]
14. Chan, J.W.; Yeo, C.K. A Transformer-based approach to electricity load forecasting. *Electr. J.* **2024**, *37*, 107370. [CrossRef]
15. Antonopoulos, I.; Robu, V.; Couraud, B.; Kirli, D.; Norbu, S.; Kiprakis, A.; Flynn, D.; Elizondo-Gonzalez, S.; Wattam, S. Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109899. [CrossRef]
16. Tsybina, E.; Winstead, C.; Ollis, B.; Olama, M.; Kuruganti, T. Demand response for frequency regulation: Research continuity and knowledge gaps. *Renew. Sustain. Energy Rev.* **2025**, *207*, 114958. [CrossRef]
17. Eren, Y.; Küçükdemiral, İ. A comprehensive review on deep learning approaches for short-term load forecasting. *Renew. Sustain. Energy Rev.* **2024**, *189*, 114031. [CrossRef]
18. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.

19. Riquelme-Domínguez, J.M.; Carranza-García, M.; Jiménez-Navarro, M.J.; González-Longatt, F.M. A machine learning-based methodology for short-term kinetic energy forecasting. *Int. J. Electr. Power Energy Syst.* **2024**, *156*, 109730. [CrossRef]
20. Xie, J.; Sun, W. A Transfer and Deep Learning-Based Method for Online Frequency Stability Assessment and Control. *IEEE Access* **2021**, *9*, 75712–75721. [CrossRef]
21. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [CrossRef]
22. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*; NeurIPS: San Diego, CA, USA, 2021; Volume 34, pp. 22419–22430.
23. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]
24. Hong, Q.; Khan, M.A.U.; Henderson, C.; Egea-Álvarez, A.; Tzelepis, D.; Booth, C. Addressing frequency control challenges in future low-inertia power systems: A Great Britain perspective. *Engineering* **2021**, *7*, 1057–1063. [CrossRef]
25. National Energy System Operator (NESO). Frequency Response Obligations—Statutory, Code and Operational Standards; Technical Report. 2021. Available online: https://www.neso.energy/document/10411/download (accessed on 24 July 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.