Engineering Informatics and Intelligent Systems 1 (2026)
journal homepage: http://eiis.informaticsfoundry.org/

Research Article

An Explainable Day-Ahead Electricity Price Forecasting Framework with Temporal Decoupling and Cross-Market Validation

Xuan Wu[a], Yifei Liu[b,*]

[a]*School of Health Economics and Management, Jiangxi University of Chinese Medicine, China;*
[b]*Department of Economics, University of Oxford, United Kingdom;*

* Correspondence: sant6806@ox.ac.uk

ABSTRACT Accurate day-ahead electricity price forecasting is important for electricity market operation, risk assessment, and renewable energy integration. Under increasing variable renewable energy penetration, strong intraday cycles and co-movement between renewable generation and demand-related activity can obscure marginal relationships between explanatory variables and prices. This study proposes an explainable day-ahead electricity price forecasting framework with temporal decoupling and cross-market validation. Using hourly records from Denmark West, Great Britain, and Northern Italy, we construct a 12-dimensional feature vector covering demand-side, supply-side, and market-side information. Hourly de-meaning is introduced for PV-related variables before Pearson and hourly-correlation screening. A Transformer–BiLSTM hybrid model captures long-range dependence and local sequential dynamics, while SHAP-based attribution examines whether learned feature contributions remain aligned with expected market mechanisms. The framework obtains lower MAE and RMSE than the compared benchmark models in all three markets. Ablation results indicate that hourly de-meaning, feature screening, the Transformer module, and the BiLSTM module each contribute to forecasting performance.

Keywords: Electricity price forecasting; Sustainable energy systems; Renewable energy integration; Temporal decoupling

1. Introduction

Electricity price forecasting (EPF) is an important foundation for power market trading, risk hedging, unit dispatch, and renewable energy integration [1, 2]. With the increasing share of wind and solar power, whose output is highly variable and whose marginal cost is low, the price formation mechanism in power markets is undergoing structural changes. Market prices are no longer mainly determined by demand alone; instead, they are increasingly driven by changes in net load, meaning the combined effect of demand and variable renewable output such as wind and solar. As a result, electricity prices exhibit strong nonlinearity, price spikes, and complex time-series coupling. At the same time, driven by global warming and environmental governance, the penetration of variable renewable energy (VRE), represented by wind and solar, is rising rapidly on the supply side [3, 4]. The global energy system is currently experiencing an unprecedented transition, which not only reshapes the physical structure of power systems but also changes the economic logic of market operation [5–7]. However, because wind and solar output are inherently unstable, supply side uncertainty increases significantly, and prices become harder to predict. Rapid changes in renewable output can amplify supply and demand imbalance, making short-term price spikes or sudden drops more likely. In addition, when renewable output is high while demand is low, markets are more likely to experience low prices or even negative prices, which increases trading risk and return volatility. Therefore, achieving high-accuracy day-ahead EPF under high VRE penetration has become a key issue in both EPF research and practice.

At present, extensive research has been conducted on EPF. Contreras et al. proposed an ARIMA-based day-ahead forecasting framework that captures linear dependence and intraday periodic patterns, achieving workable performance under traditional market conditions [8]. Garcia et al. further introduced GARCH to model volatility clustering, allowing the model to describe the dynamics of both the mean and the variance, thereby improving stability during high-volatility periods [9]. Liu et al. developed ARMA-GARCH models and their extensions for short-term price forecasting, and showed that jointly modeling the price level and volatility intensity can reduce forecasting errors and better reflect nonlinearity and heteroskedasticity in electricity prices [10]. To better handle strong non-stationarity and spike shocks, Tan et al. proposed a hybrid strategy that combines wavelet decomposition with ARIMA and GARCH. This strategy first decomposes the price series into multiple scales, then models each component separately and reconstructs the final forecast, which improves the adaptability of a single linear model under spikes and sudden changes [11]. In addition, Gupta et al. compared seasonal ARIMA (SARIMA) with methods such as multilayer perceptron (MLP) in an empirical study of trading-oriented markets; however, they noted that purely linear models can underfit when volatility is complex and market structure changes, so nonlinear learning is often needed for robustness [12]. However, these traditional econometric models, while effective at capturing linear trends and seasonal fluctuations, mainly rely on the price series itself and are limited in their ability to integrate nonlinear effects from exogenous variables such as demand and renewable output. As a result, under high VRE penetration and stronger variability, they face bottlenecks in accuracy and generalization [13]. Data-driven machine learning and deep learning methods offer a practical way forward by learning nonlinear patterns in price series more effectively. For example, Saini et al. used linear regression and support vector machines (SVM) for price forecasting to address nonlinear mappings between prices and their drivers, and confirmed that SVM performs better in handling nonlinear relationships [14]. Tschora et al. proposed a machine learning XGBoost approach to capture local patterns and window dependence in power time-series learning, and demonstrated the effectiveness of such algorithms for electricity time-series modeling [15]. In deep learning, Lago et al. used an MLP to combine exogenous inputs, such as temperature, demand, and gas prices, for hourly forecasting, demonstrating that feed-forward networks can learn nonlinear relationships driven by exogenous factors [16]. Later, recurrent neural networks such as LSTM and GRU were applied to model temporal dependence in electricity prices and improve the fit to sequence dynamics [17, 18]. Furthermore, Vaswani et al. proposed the Transformer architecture, which introduces attention mechanisms to better capture long-range dependencies in long sequences [19]. In EPF tasks, Wang et al. incorporated dynamic Bayesian networks, achieving improved forecasting performance, indicating that probabilistic network structures can enhance forecast accuracy [20]. Lago et al. provided a systematic review of day-ahead EPF methods and best practices, and released open benchmarks to support model comparison and reproducible research [21].

However, although deep learning and machine learning can achieve accurate EPF, most deep neural networks are essentially 'black-box" systems, making it difficult to clearly explain the economic rationale between input variables and predicted prices [22]. In addition, in statistics, Simpson's Paradox shows that correlations computed on the full sample can mask or even distort the true patterns within specific time periods [23]. This issue is especially common in EPF. For example, in PV-dominated markets, if correlation analysis is conducted on full-day data, because PV peaks often overlap with daytime demand peaks, one observes a weak or even positive correlation between PV generation and prices. However, this apparent result conflicts with the merit-order effect, where zero-marginal-cost renewable generation tends to reduce market clearing prices [24]. Moreover, markets differ greatly in renewable energy structure and market rules. For instance, Jónsson et al. reported that wind power forecasts have a strong impact on market behavior and prices, with higher price sensitivity in markets with high wind penetration [25]. Hogan emphasized the role of scarcity pricing and reserve mechanisms in market clearing [26]. Clò et al. analyzed the impact of solar and wind power on wholesale prices in the Italian market and found a clear merit-order effect [27]. Overall, focusing only on fitting accuracy in a single market is not sufficient for accurate cross-regional EPF, and a mechanism-consistent modeling pipeline with explainability analysis is of interest.

To address these issues, we propose an explainable and modular framework for day-ahead EPF. First, based on related literature and public information, we summarize renewable shares, price volatility features, and key market rules for three representative market types: wind-dominated markets, island-type markets with scarcity pricing, and PV-dominated markets. This summary guides the construction of a candidate set of driving factors. Next, on the data side, we use feature engineering to transform basic variables such as demand, wind generation, and PV generation into derived features, including net load, renewable penetration, lag terms, and ramp terms. These features are used as multivariate time-series inputs and key features are selected through correlation tests. For the model, we build a Transformer with multi-head attention and then connect it in series with a bidirectional long short-term memory network (BiLSTM) to capture both long-range dependence and local dynamics, improving the ability to track price spikes and structured intraday changes [28]. To reduce the statistical masking caused by strong intraday cycles, we further introduce an hourly de-meaning decoupling strategy, which helps keep the marginal price-lowering direction of PV generation consistent with the expected mechanism. Finally, we use SHAP-based feature attribution to quantify the importance and contribution direction of each input, identify key driving factors and their roles, and provide explainability analysis for the forecasts [29]. To validate robustness and general applicability, we conduct empirical benchmarking and ablation studies on day-ahead price data from three European markets with clearly different structures, and systematically evaluate forecasting performance, explainability, and robustness across heterogeneous market settings.

2. Methodology

The overall workflow of the proposed framework is illustrated in Figure 1.

This study proposes an explainable day-ahead EPF framework with temporal decoupling and cross-market validation, aiming to improve forecasting accuracy while maintaining explainability. First, based on existing literature and publicly available market rules, we analyze the energy mix, market clearing mechanisms, and typical price patterns of different markets. We then use correlation tests on the data side and analyze the contribution of model inputs to establish market mechanism-consistent evidence on the key drivers of prices in three market types. To address intraday cycle confounding and statistical masking that often occur in markets with high PV penetration, we introduce an hourly de-meaning strategy. By applying centering for each hour to PV generation and related variables, this step reduces the shared intraday trend within fixed hours and allows deviations within the same hour to better reflect the marginal effect direction of PV generation on prices. Based on the selected feature set, we develop a Transformer-BiLSTM hybrid model that combines multi-head-attention with bidirectional sequence modeling. We conduct day-ahead EPF experiments on the Great Britain, Denmark West, and Italy North markets. The forecasting task is formulated as a sequence-to-sequence problem. The model takes the past 168 hours

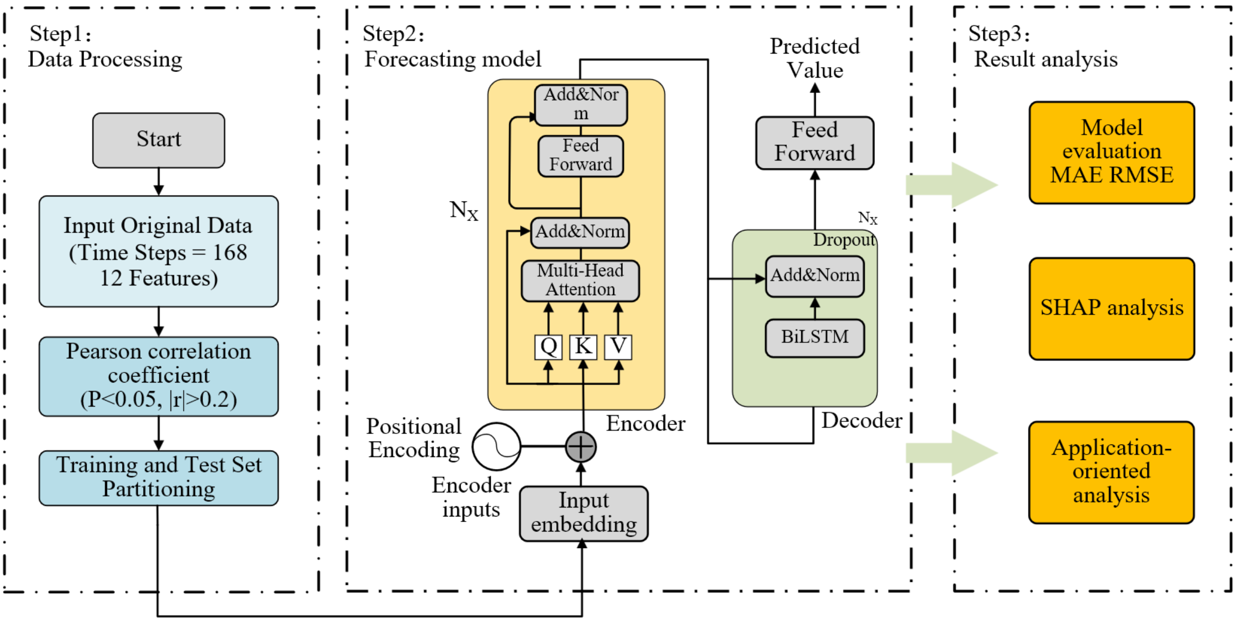


Fig. 1. Overall workflow of the proposed framework.

of multivariate observations (12 features) as input and directly produces a 24-hour ahead price trajectory for the following day. In addition, we apply SHAP attribution to quantify the contribution direction and importance of key driving factors and to examine their consistency with expected mechanisms, thereby improving explainability and trust for sustainable power system operation.

2.1. Cross-Market Electricity Price Data

To examine the proposed framework across different market settings, we use cross-market data from three European day-ahead market zones with clear differences in generation mix, physical connection, and pricing characteristics, thereby providing data support for validating the proposed framework under heterogeneous market conditions. The data are obtained from the ENTSO-E Transparency Platform and cover hourly records from 1 January 2015 to 31 December 2020 [5]. The forecasting target is the day-ahead electricity market price. Based on the market-mechanism-oriented feature design adopted in this study, we construct a 12-feature input set that includes demand-side, supply-side, and market-side information (Table 1). All explanatory variables are derived from historical observations available prior to the forecasting day, using lagged values and related transformations within predefined look-back windows.

To verify the general applicability of the model under different market mechanisms, we select three European day-ahead market zones with clear differences in physical connection, generation mix, and pricing mechanisms. Their geographic locations are shown in Figure 2.

Fig. 2. Experimental area map.

DK_1 (Denmark West) is a wind dominated market. Wind penetration is high, and wind variability has a strong impact on supply side marginal cost and market settlement prices [30–32].

GB_GBN (Great Britain) is an island market with scarcity pricing. When wind output is low and gas marginal cost is high, scarcity pricing is more likely, and price spikes occur more frequently [33–35].

IT_NORD (Italy North) is a PV dominated market with a duck curve pattern. Under high PV penetration, prices often show a clear intraday structure with a midday valley and an evening rebound. The concentrated rise of PV output around midday reduces net load and pushes prices down, while the fast drop of PV output in the evening, together with still high demand, drives prices up, leading to strong intraday valley peak separation [36–38]. The statistical characteristics of the three market datasets are summarized in Table 2.

Table 1
Definition of input features (hourly, ENTSO-E).

Label	Formula	Feature description
Load_Total	–	Hourly actual electricity load of the bidding zone, representing realized demand level.
Load_Lag_24	Load_{t-24}	Load at the same hour one day earlier, capturing diurnal demand periodicity.
Load_Ramp	$\text{Load}_t - \text{Load}_{t-1}$	Hour-to-hour load change, reflecting short-term demand ramping requirement.
Solar_Gen	–	Hourly photovoltaic generation, representing solar supply shock with near-zero marginal cost.
Wind_Onshore_Gen	–	Hourly onshore wind generation, representing wind supply contribution.
Wind_Offshore_Gen	–	Hourly offshore wind generation.
Wind_Total	$\text{Wind}_{\text{Onshore}} + \text{Wind}_{\text{Offshore}}$	Total wind generation (onshore + offshore), representing overall wind supply level.
Renewable_Total	Solar_Gen + Wind_Total	Total variable renewable generation (solar and wind), representing overall VRE supply level.
Net_Load	Load_t − Renewable_Total	Residual (net) load that must be served by conventional generation and imports.
RES_Penetration	$\dfrac{\text{Renewable\_Total}}{\text{Load}_t + \epsilon}$	Renewable penetration ratio, measuring the share of renewable supply relative to demand; ϵ avoids division by zero and is set to 1 in the implementation.
Price_Lag_24	–	Electricity price at the same hour one day earlier, capturing market memory and diurnal periodicity.
Price_Lag_168	–	Electricity price at the same hour one week earlier, capturing weekly seasonality and market memory.

Table 2
Statistical characteristics of the three market datasets (2015–2020, hourly).

Market	Mean price	Share of negative prices	Maximum price	Share of Net_Load < 0	Notes
DK_1	31.33	1.07%	200.04	20.53%	Typical wind dominated market with frequent supply surplus.
GB_GBN	43.41	0.16%	999.00	0.05%	Demand dominated with clear spikes (17 hours above 300).
IT_NORD	49.99	0%	206.12	0%	Solar accounts for 98.7% of renewables.

2.2. Data Preprocessing and Feature Selection

Market-mechanism-guided statistical diagnosis is used to screen candidate driving factors by checking mechanism consistency and statistical robustness under the merit order effect and intraday cycle confounding. We conduct the diagnosis under different market structures. Denmark West is characterized by high wind penetration and frequent negative prices, Great Britain is characterized by demand driven scarcity spikes, and Italy North is characterized by high PV penetration and clear intraday valley peak patterns. Based on these differences, we test the correlation strength of variables such as wind and PV generation, net load, renewable penetration, and ramp terms under both full-sample and hour-by-hour conditions. For PV related variables, we apply an hourly de-meaning step to reduce bias caused by intraday co movement and to recover the marginal price lowering signal. In this way, we obtain a key feature set that is both physically explainable and statistically reliable.

Let the day-ahead price forecasting target be Y_i and the input feature be X_i. The Pearson correlation coefficient is:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2}}. \tag{1}$$

Significance screening rule (two-sided test). We apply a two-sided significance test for Pearson correlation and set the screening threshold as $p<0.05$ and $|r|>0.2$. However, when a variable strongly depends on the hour of the day, the global correlation can be masked by confounding factors. Therefore, we also compute hourly correlation. The hourly correlation is:

$$r_h = \mathrm{Corr}(X_i, Y_i \mid h(t) = h), \quad h = 0, 1, \ldots, 23. \tag{2}$$

where h(t) denotes the hour index of time t. Hour by hour correlation reveals local relationships under the same time condition and reduces confounding caused by intraday cycle patterns.

Among the candidate inputs, PV-related variables show a pronounced intraday profile and overlap with periods of relatively high demand-related activity. Under full-sample aggregation, this co-movement can weaken or even distort the marginal relationship between PV-related variables and prices, making the correlation pattern less consistent with the merit-order effect. This issue is particularly relevant in IT_NORD and can be understood as a Simpson's-Paradox-type masking effect. For this reason, hourly de-meaning is applied only to PV-related variables, including Solar_Gen and renewable-share-related indicators. For a feature value X_t, we compute its deviation at a specific hour h as follows:

$$X'_t = X_t - \mathrm{E}[X \mid \mathrm{hour}(t) = h(t)]. \quad (3)$$

Here, $\mathrm{E}[X \mid \mathrm{hour}(t) = h(t)]$ is the long-term conditional mean for hour h. To avoid data leakage, this conditional mean is computed over the training set.

2.3. Transformer–BiLSTM Forecasting Model

We propose a Transformer-BiLSTM hybrid model for EPF. First, the historical multivariate time series is fed into a Transformer encoder for feature extraction. Using the multi-head self-attention mechanism, the encoder models global dependencies across time steps, so it can represent multiple seasonal patterns in electricity prices and interactions across input features. Next, we connect the latent representation produced by the Transformer encoder with the sequence modeling output of a bidirectional long short-term memory network (BiLSTM). This step further captures local dynamics and short-term fluctuation patterns and complements the global information learned by the attention module. Finally, a fully connected layer performs feature mapping and dimension transformation, and outputs the predicted day-ahead electricity price series for the next 24 hours. In this way, the proposed Transformer-BiLSTM hybrid model improves the ability to model long range dependence caused by multiple seasonal patterns and the non-stationarity caused by real-time market fluctuations.

2.3.1. Transformer Encoder for Feature Extraction

The model input is a multivariate time-series matrix with a window length of L, denoted as $X \in \mathbb{R}^{L\times d}$, where d is the number of features. Because the Transformer is based on set style operations and does not have an inherent sense of time order, positional encoding is required. We use the standard sine cosine positional encoding to inject time position information into the input embeddings, so the model can recognize the absolute time positions in electricity price data.

The core component of the Transformer is multi head self-attention (MHSA). For an input sequence, MHSA computes the interaction among Query, Key, and Value and assigns different attention weights to different historical time steps in a data driven manner. The attention formula is:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V. \quad (4)$$

In EPF tasks, the main advantage of self-attention is that it can directly connect the current forecast point (t+k) with distant historical points such as t and t–168. This means that when the model predicts the electricity price at 9:00 a.m. next Monday, it can directly refer to the market state at 9:00 a.m. last Monday, without being limited by the 167 intermediate time steps. This property is critical for capturing weekly seasonality. Figure 3 shows the structure of the Transformer model.

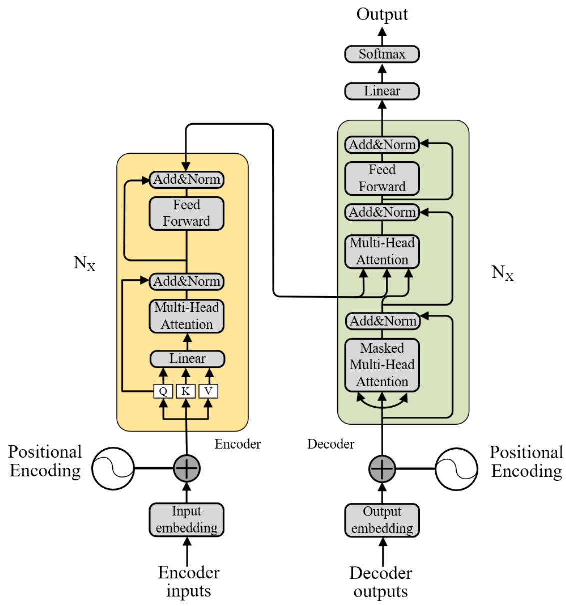


Fig. 3. The structure of the Transformer model.

2.3.2. BiLSTM Layer for Temporal Dependency Modeling

The long short-term memory (LSTM) network is a special type of recurrent neural network (RNN) that is widely used for time series data because it can better capture dependencies in sequences. By introducing an input gate, a forget gate, and an output gate, LSTM controls the flow of information and effectively reduces the gradient vanishing and gradient explosion problems that are common in standard RNNs. A bidirectional LSTM (BiLSTM) consists of two LSTM components. The forward LSTM processes the input sequence in the forward direction, while the backward LSTM processes the sequence in the reverse direction. This design strengthens the ability to capture context information from both directions.

As shown in Figure 4, BiLSTM can model bidirectional dependencies by using information from both past and future time steps within the input window. The two directional outputs are concatenated at each time step to form the final representation.

The BiLSTM formulas are:

$$\overrightarrow{h}_t = \mathrm{LSTM}(x_t, \overrightarrow{h}_{t-1}), \quad (5)$$

$$\overleftarrow{h}_t = \mathrm{LSTM}(x_t, \overleftarrow{h}_{t+1}), \quad (6)$$

$$y_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]. \quad (7)$$

where x_t is the input feature vector at time t, $\overleftarrow{h}_t$ is the hidden-state output of the backward LSTM at time t, and $[\cdot,\cdot]$ denotes concatenation along the feature dimension, which fuses sequence information from both directions.

2.3.3. The architecture of Transformer-BiLSTM Model

In machine translation, the standard Transformer uses a masked multi-head attention mechanism in the decoder so that the model attends only to previous tokens during sequence gener-

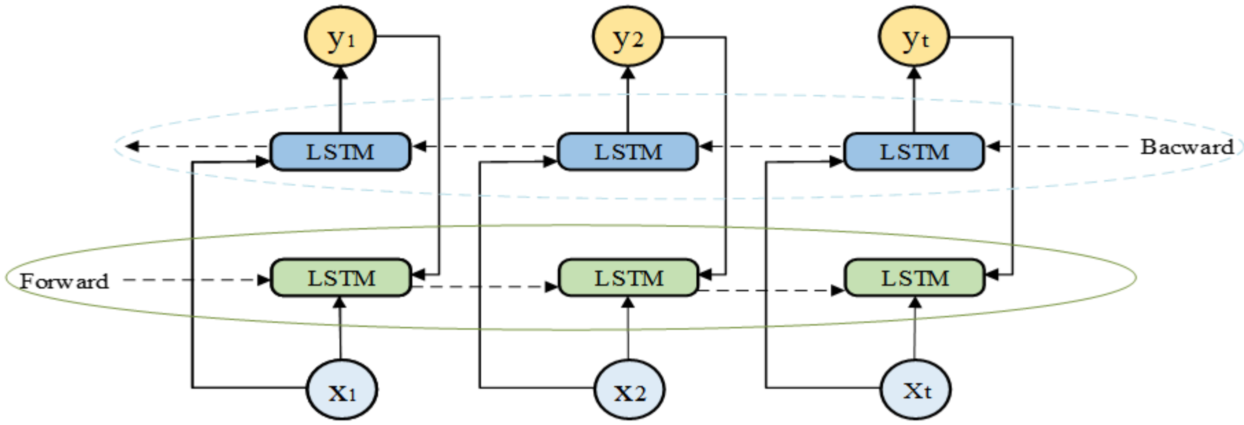


Fig. 4. The structure of BiLSTM.

ation. This helps keep the generated text consistent. However, in time series forecasting, the input is a known historical window and the goal is to predict future values. The dependency structure is different from translation, so a masked decoder is not required.

To address this, we use BiLSTM as a sequence modeling head/decoder, and we apply residual connections to process the encoded sequence. This modification preserves the information extracted by the encoder and helps handle long range dependence in time series. To reduce overfitting during training, we add a Dropout layer within the BiLSTM module, which randomly drops part of the neuron outputs. Finally, a fully connected feedforward network is used to generate the final day-ahead electricity price forecasts. The overall architecture that combines the Transformer encoder and the BiLSTM module is shown in Figure 5.

2.3.4. Experimental setup

The proposed Transformer-BiLSTM based EPF method is implemented and validated in a server environment. The hardware uses an NVIDIA GeForce RTX 4090 GPU, and GPU acceleration is applied for both model training and inference. The software environment is based on Python 3.11.

In the model design, the Transformer encoder is used to capture long range dependence in electricity prices and interactions across input features, while the BiLSTM module is used to capture short term fluctuations and local time series patterns. During training, the batch size is set to 256, and Dropout is set to 0.1. We use Adam as the optimizer with a learning rate of 1×10^{-3}. Mean squared error (MSE) is used as the loss function to support stable convergence and improve fitting accuracy. We use data from 2015 to 2019 as the training set and data from 2020 as the test set.

To evaluate forecasting performance, we use mean absolute error (MAE) and root mean squared error (RMSE) as the evaluation metrics. Their formulas are given in Equations (8) and (9).

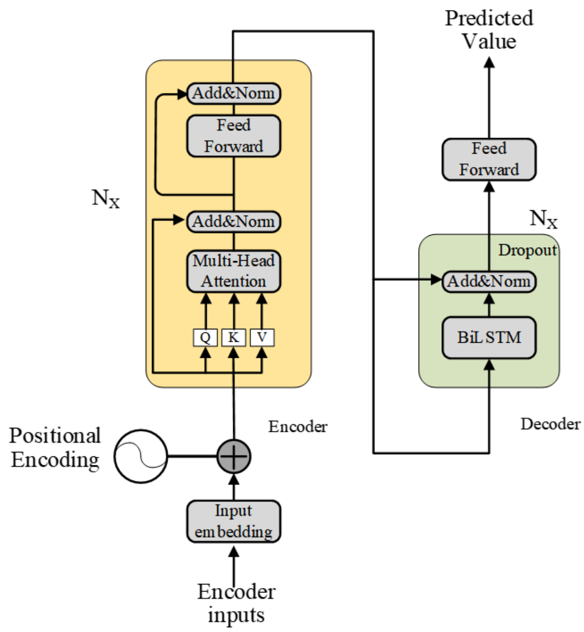


Fig. 5. The structure of the Transformer–BiLSTM model.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \tag{9}$$

where y_i is the actual price for sample i, $\hat{y}_i$ is the predicted price, and N is the number of test samples.

3. Results and analysis

3.1. Cross-Market Correlation Analysis

Before building the forecasting model, we conduct statistical tests to check the consistency between market mechanisms and data patterns. Because EPF (spot market price) and variables such as demand and renewable generation show strong intraday cycles, computing Pearson correlation on the full sample can mix time-of-day structure with marginal effects. We therefore use raw global correlations to characterize the training data and evaluate the preprocessing step separately through the ablation study.

First, we compute the Pearson correlation coefficients between each input variable and EPF for the three markets using the training set (2015–2019), as shown in Figure 6. In general, load-related variables (Load_Total, Load_Lag_24, Load_Ramp, and Net_Load) are positively associated with EPF, indicating that demand-side level, short-term load variation, and residual system tightness remain major upward drivers of prices. In addition, the market-side lag features Price_Lag_24 and Price_Lag_168 capture daily and weekly price memory.

In DK_1, which is wind dominated, wind related variables show a more stable negative correlation with EPF, consistent with the expected supply side impact of wind generation. In GB_GBN, which is demand dominated, prices are more strongly explained by demand and lagged price terms, and spike events make the correlation structure more driven by demand and market memory. In IT_NORD, which is PV dominated, the raw correlation between Solar_Gen and EPF is negative but smaller in magnitude than the price-lag relationships. This motivates the hourly de-meaning preprocessing step used in the forecasting pipeline.

Specifically, because offshore wind capacity is very low in Italy North, the ENTSO-E Transparency Platform does not provide valid Wind_Offshore_Gen records for this zone during the study period. The variable is retained only to maintain the common 12-variable input schema and is not interpreted as a market driver for IT_NORD; its contribution is negligible in the corresponding SHAP plot.

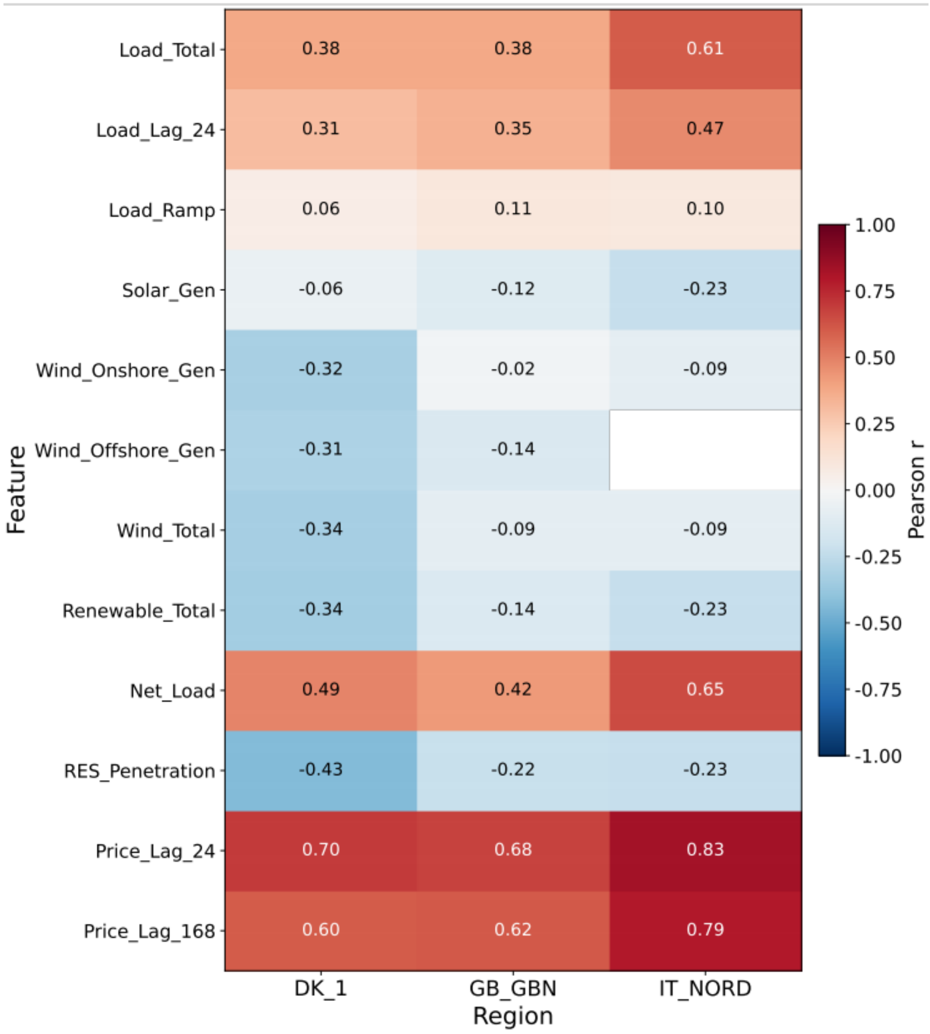


Fig. 6. Pearson correlation coefficients on the training set.

To examine the effect of intraday aggregation, Figure 7 compares the global raw relationship, a noon-hour slice, and the relationship after hourly de-meaning of Solar_Gen. The corresponding correlations shown in the three panels are −0.049, −0.339, and −0.124, respectively. The noon-hour result is more negative than the global raw correlation, illustrating why a single pooled correlation can obscure within-hour variation.

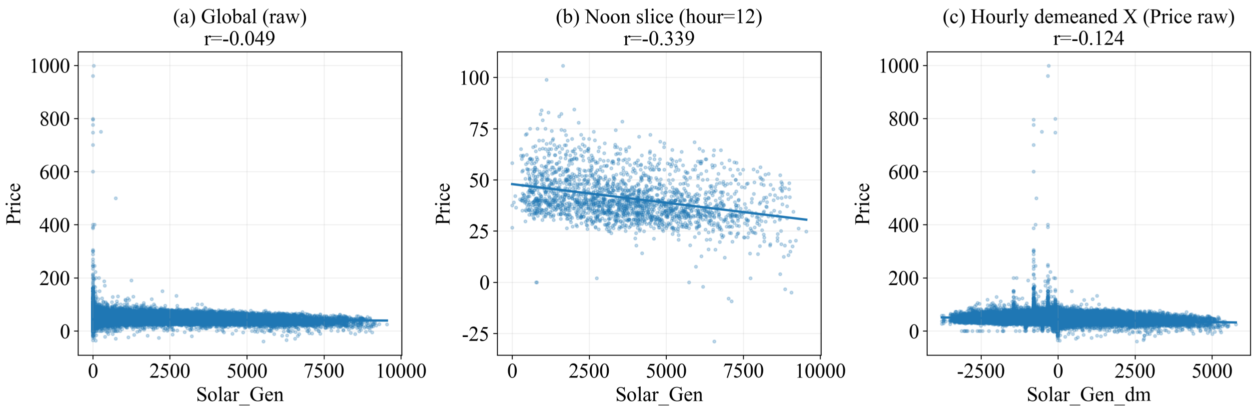


Fig. 7. Solar generation and price relationships before and after hourly de-meaning in IT_NORD.

We also report the Pearson correlation matrix obtained with hourly de-meaned predictors and the raw price series in Figure 8. This diagnostic view documents the correlation structure used to examine the temporal-decoupling preprocessing step.

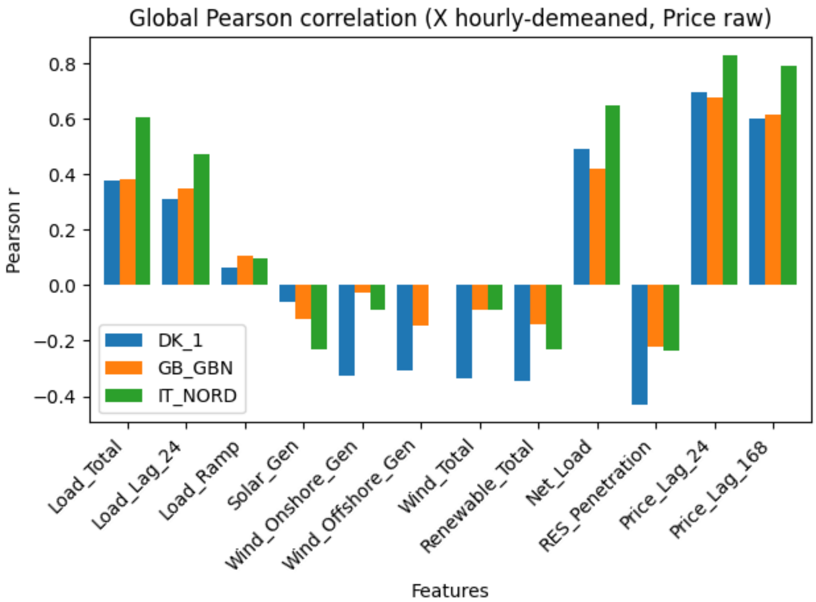


Fig. 8. Pearson correlation coefficients with hourly de-meaned predictors and raw prices.

3.2. Day-Ahead Electricity Price Forecasting Results for Cross-Regional European Markets

We compare the proposed Transformer-BiLSTM model with seasonal naive baselines (Naive-24 and Naive-168), classical machine learning models (SVM, XGBoost, and LightGBM), and commonly used deep learning sequence models (LSTM, GRU, and CNN-LSTM). These baselines are widely used in EPF studies to ensure a fair and comprehensive comparison. As shown in Figure 9, Transformer-BiLSTM achieves the lowest MAE and RMSE in all three markets, indicating consistently strong forecasting performance across different market structures. In DK_1, Transformer-BiLSTM reaches MAE = 7.182 and RMSE = 10.223, outperforming the strongest competing benchmark SVM (MAE = 7.615, RMSE = 10.892). In GB_GBN, Transformer-BiLSTM achieves MAE = 6.982 and RMSE = 7.945, showing clear improvements over all benchmark models, including CNN-LSTM and LightGBM. In IT_NORD, Transformer-BiLSTM achieves MAE = 4.696 and RMSE = 6.245, again providing the best overall forecasting accuracy among all compared models. These results suggest that combining the global dependency modeling of the Transformer with the local sequence modeling of BiLSTM helps capture both long-range temporal dependencies and local sequential dynamics in EPF data, leading to improved predictive accuracy across heterogeneous markets.

For error metrics, the comparison shows a consistent advantage of Transformer-BiLSTM over benchmark models in terms of both MAE and RMSE. In DK_1, although SVM and LightGBM remain competitive, the proposed model still yields the lowest forecasting errors. In GB_GBN, Transformer-BiLSTM achieves the best RMSE (7.945) and also the best MAE (6.982), indicating a clear advantage in a more volatile market environment. In IT_NORD, the proposed model also maintains the lowest MAE and RMSE, further confirming its robustness

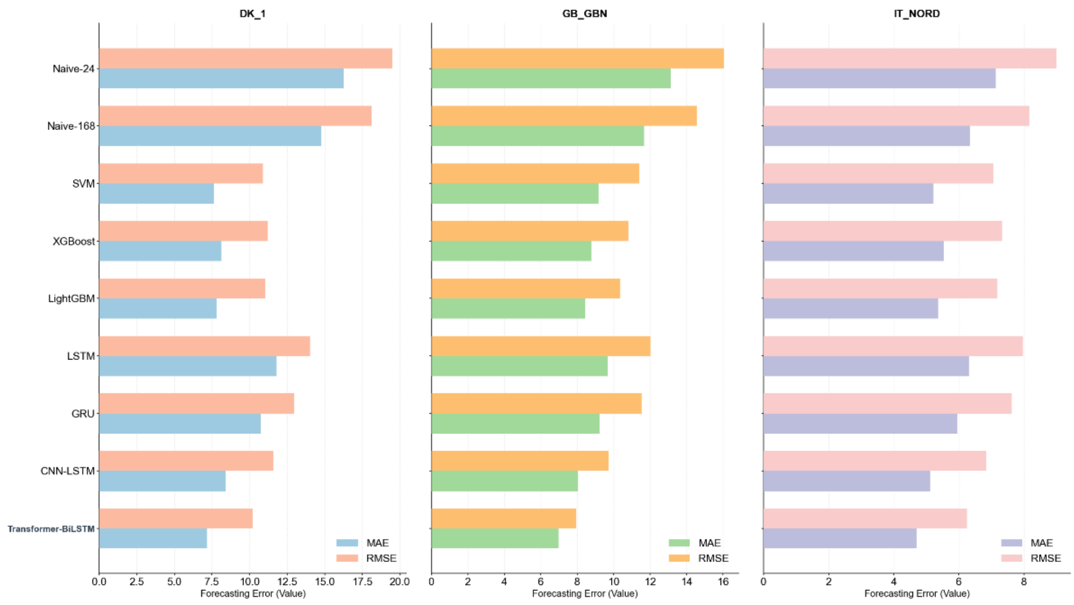


Fig. 9. Benchmark forecasting performance of all models in the three electricity markets.

in a market with strong renewable-related variability. Overall, the comparison across MAE and RMSE confirms that Transformer-BiLSTM provides consistently superior forecasting performance across the three markets, offering a reliable basis for cross-market day-ahead electricity price forecasting and subsequent interpretation analysis.

3.3. Ablation Study

This section uses a set of ablation experiments to systematically evaluate how each key module affects day-ahead electricity price forecasting performance. The results are reported in Table 3.

Table 3
Ablation results of the proposed model in the three electricity markets.

Model setting	Market	MAE	RMSE
Full model (Transformer–BiLSTM, hourly de-meaning, feature screening)	DK_1	7.182	10.223
	GB_GBN	6.982	7.945
	IT_NORD	4.696	6.245
Experiment 1, without hourly de-meaning	DK_1	7.654	10.876
	GB_GBN	7.432	8.567
	IT_NORD	5.123	6.789
Experiment 2, without BiLSTM	DK_1	8.234	11.567
	GB_GBN	7.987	9.234
	IT_NORD	5.765	7.345
Experiment 3, without Transformer	DK_1	8.345	11.789
	GB_GBN	8.123	9.456
	IT_NORD	5.876	7.567
Experiment 4, without feature screening	DK_1	7.890	11.123
	GB_GBN	7.654	8.89
	IT_NORD	5.345	7.012

Experiment 1 removes the hourly de-meaning decoupling strategy. Performance decreases in all three markets. In DK_1, MAE increases to 7.654 and RMSE increases to 10.876. In GB_GBN, MAE increases to 7.432 and RMSE increases to 8.567. In IT_NORD, the degradation is also clear, with MAE = 5.123 and RMSE = 6.789. These results show that hourly de-meaning can reduce statistical masking caused by strong intraday cycles, helping the model learn clearer marginal driving signals and improving forecasting accuracy.

Experiment 2 removes the BiLSTM module. After removing BiLSTM, performance degrades clearly. In DK_1, MAE is 8.234 and RMSE is 11.567. In GB_GBN, MAE is 7.987 and RMSE is 9.234. In IT_NORD, MAE is 5.765 and RMSE is 7.345. This indicates that BiLSTM is important for learning local time-series patterns and short-term dynamics. Without it, the model cannot capture complex time dependence well, leading to higher errors across all three markets.

Experiment 3 removes the Transformer module. Performance further worsens in all markets. In DK_1, MAE is 8.345 and RMSE is 11.789. In GB_GBN, MAE is 8.123 and RMSE is 9.456. In IT_NORD, MAE is 5.876 and RMSE is 7.567. This shows that the Transformer module is crucial for modeling global dependencies and interactions across features. Without it, the model captures long-range dependence and structured fluctuations less effectively, which reduces overall forecasting performance.

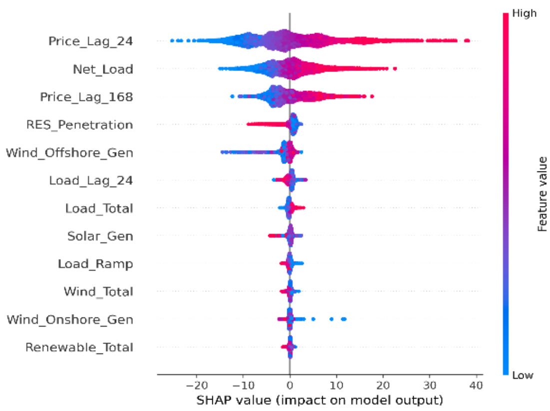
Experiment 4 removes Pearson-based feature screening. This setting causes a moderate performance drop. In DK_1, MAE is 7.890 and RMSE is 11.123. In GB_GBN, MAE is 7.654 and RMSE is 8.890. In IT_NORD, MAE is 5.345 and RMSE is 7.012. This suggests that Pearson-based feature screening reduces redundancy and noise, improves learning efficiency, and supports more stable generalization, especially when market differences are large.

Overall, the Transformer, BiLSTM, hourly de-meaning, and Pearson-based feature screening jointly improve EPF performance. The Transformer strengthens long-range dependence modeling and cross-variable interaction learning, BiLSTM improves local dynamics and short-term pattern capture, hourly de-meaning reduces intraday co-movement masking, and feature screening reduces feature redundancy and noise. Their combined effect enables more accurate and more robust day-ahead EPF across markets with clearly different structures.

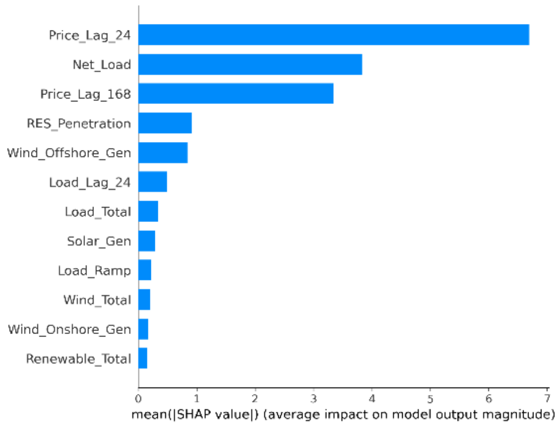
3.4. SHAP-Based Market Interpretability Analysis

To improve explainability, we use the Shapley additive explanation method to decompose feature contributions. For each sample, the marginal effect of each feature among the 12 input features is quantified as a SHAP value. We then use the mean absolute SHAP value to rank global feature importance. From the SHAP beeswarm plots and global-importance bar plots across the three markets (Figures 10–12), two common patterns can be observed. First, Price_Lag_24 is one of the most important features in all three markets, indicating strong daily cycle persistence and market memory in day-ahead prices. Second, Net_Load and RES_Penetration show stable contributions in most cases, reflecting the basic role of supply demand tightness and renewable supply shocks in price formation. Although the detailed feature rankings vary across markets, the overall contribution patterns remain broadly consistent with the corresponding market structures.

In DK_1, the importance ranking shows that Price_Lag_24 and Net_Load contribute the most, followed by renewable-related and wind related variables such as Wind_Offshore_Gen and Wind_Onshore_Gen or Wind_Total. The SHAP directions are consistent with expected mechanisms. Higher Net_Load is associated with positive SHAP values and pushes the model output upward. In contrast, higher RES_Penetration and higher Wind_Offshore_Gen tend to produce negative SHAP values, reflecting the price lowering effect of increased zero marginal cost generation. This interpretation matches the wind dominated structure of DK_1, where supply side variability has a stronger impact on prices.



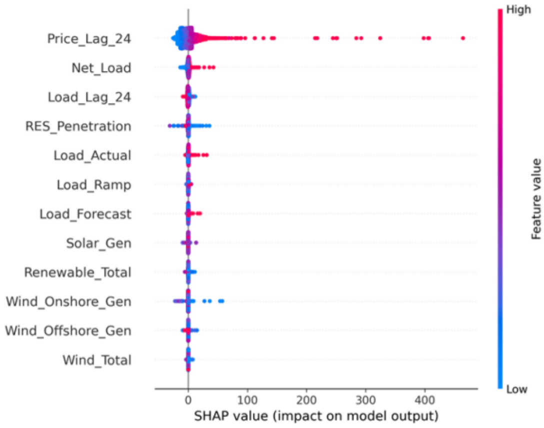
(a) SHAP beeswarm DK_1



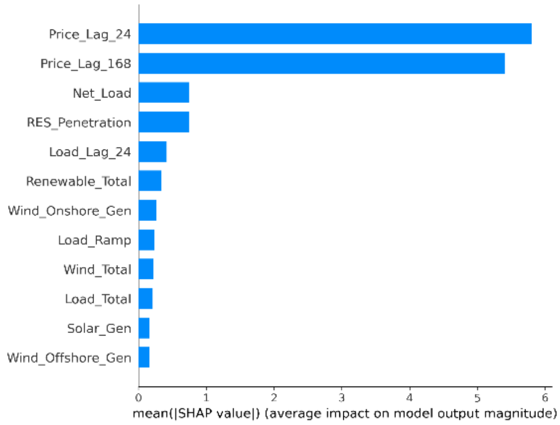
(b) SHAP contribution plot DK_1

Fig. 10. SHAP beeswarm and global mean absolute feature-importance plots for DK_1.

In GB_GBN, Price_Lag_24 and Price_Lag_168 jointly dominate the global importance ranking, indicating that daily and weekly price persistence are the main drivers of the model output. Net_Load and RES_Penetration form a second tier of importance, while Load_Lag_24, Load_Total, and Load_Ramp make smaller contributions. This suggests that under tight supply demand conditions, demand-side changes and market memory can amplify price responses and make tightness signals more visible in the explanation.



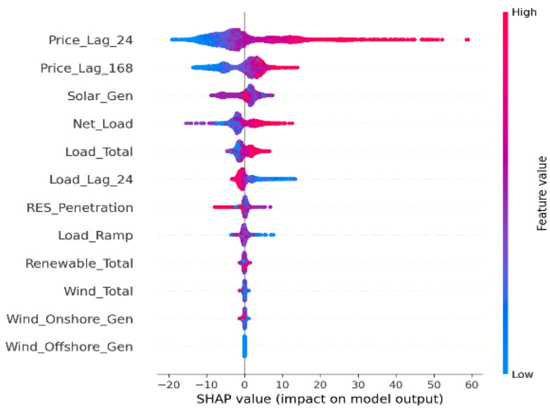
(a) SHAP beeswarm GB_GBN



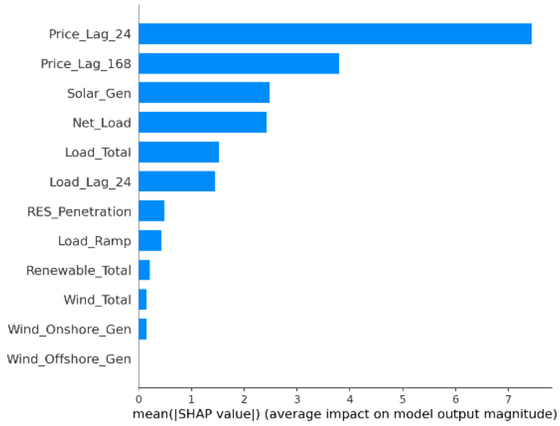
(b) SHAP contribution plot GB_GBN

Fig. 11. SHAP beeswarm and global mean absolute feature-importance plots for GB_GBN.

In IT_NORD, Solar_Gen is among the leading non-price features, and the SHAP direction shows that higher Solar_Gen is more often associated with negative SHAP values. This indicates that the model can capture a PV-related price-lowering pattern. The result is aligned with the intraday valley and peak structure and the duck curve pattern in Italy North under high PV penetration [36–38]. At the same time, wind related variables such as Wind_Onshore_Gen, Wind_Offshore_Gen, and Wind_Total have weaker overall contributions in this market, which is consistent with its generation mix.



(a) SHAP beeswarm IT_NORD



(b) SHAP contribution plot IT_NORD

Fig. 12. SHAP beeswarm and global mean absolute feature-importance plots for IT_NORD.

Overall, the SHAP results identify Price_Lag_24 as a major contributor across the three markets, while Net_Load represents clearing pressure. The beeswarm panels show contribution directions for DK_1, GB_GBN, and IT_NORD, and the paired bar plots report global importance magnitudes. Solar_Gen, Wind_Onshore_Gen, Wind_Offshore_Gen, and RES_Penetration provide model-based evidence broadly consistent with the merit-order effect [24, 27]. These results support the rationality of the proposed feature set and provide explainable evidence for cross-market comparison and downstream analysis.

4. Discussion

This study provides empirical evidence on three aspects of day-ahead electricity price forecasting across heterogeneous electricity markets [27, 36–38]. First, the proposed framework shows cross-market predictive robustness. Across DK_1, GB_GBN, and IT_NORD, the full Transformer–BiLSTM model with hourly de-meaning and feature screening achieves the lowest MAE and RMSE among the compared benchmark settings, and the ablation results further show that removing any of the four key components leads to performance degradation. These results indicate that the proposed framework is not only effective in a single market context, but also remains empirically robust across markets with clearly different renewable structures and pricing behaviors.

Second, the SHAP analysis provides post hoc explanatory evidence that is broadly consistent with expected market mechanisms. Across the three markets, Price_Lag_24 remains a major contributor, reflecting strong daily persistence in day-ahead prices, while Net_Load and RES-related variables show contribution directions that are consistent with market tightness and merit-order effects. In DK_1, renewable and wind-related features contribute more strongly and are generally associated with price-lowering effects [26, 33–35]; in GB_GBN, lagged price and demand-side tightness signals are more prominent under scarcity-pricing conditions; and in IT_NORD, Solar_Gen shows clearer negative SHAP contributions, consistent with the PV-driven midday valley pattern. Taken together, these results do not establish causal mechanisms, but they do indicate that the learned attribution patterns are broadly aligned with the known economic logic of the three market types [39].

Third, the results support the practical value of hourly de-meaning as a temporal decoupling strategy. The empirical motivation of this step is that strong intraday co-movement can mask the marginal relationship between PV-related variables and prices when correlations are examined on the full sample. The ablation study shows that removing hourly de-meaning leads to a clear loss in forecasting accuracy in all three markets, while the post hoc attribution results suggest that the de-meaned setting helps recover feature-contribution directions that are more consistent with the expected merit-order effect in PV-related contexts [37, 38]. Therefore, the contribution of hourly de-meaning is not merely statistical simplification; rather, it helps reduce intraday masking and improves the interpretability of feature-response patterns under structured intraday cycles [40, 41].

Beyond these empirically supported findings, the present study also suggests several directions for future application-oriented use, although these downstream values were not directly validated here. In PV-dominated market settings such as IT_NORD, the ability to better identify structured intraday low-price windows may support future storage-oriented scheduling assessments, especially where charging and discharging decisions depend on the timing and depth of predictable midday price valleys. In wind-dominated settings such as DK_1, earlier recognition of abnormal low-price or negative-price periods may support spike-risk and downside-risk monitoring for market participants exposed to volatile renewable supply conditions [42]. In scarcity-pricing settings such as GB_GBN, improved sensitivity to high-price episodes may provide useful reference signals for future decision-oriented analysis for renewable-integrated electricity markets, such as bidding assessment, reserve-aware monitoring, or hedging trigger design [43]. Lastly, although the proposed framework was validated in three representative European markets, further evaluation across additional market settings, rolling forecasting windows, and formal statistical significance tests is still needed to more fully establish its broader robustness and generalizability.

5. Conclusions

Day-ahead electricity price forecasting in cross-regional markets remains challenging because electricity price series are jointly affected by strong intraday patterns, heterogeneous market structures, and changing renewable generation conditions. These factors increase the complexity of price formation and lead to substantial differences in market behavior across regions. As a result, it is difficult for a single forecasting framework to effectively capture both common temporal patterns and region-specific characteristics. This makes it challenging to achieve both forecasting accuracy and interpretability in cross-regional electricity price forecasting.

To address this issue, this study examines three day-ahead electricity markets, DK_1, GB_GBN, and IT_NORD, and develops an explainable forecasting framework for day-ahead electricity prices. Specifically, this study proposes a Transformer–BiLSTM forecasting framework combined with hourly de-meaning, feature screening, and SHAP-based analysis. The empirical results show that the proposed framework improves forecasting performance across the selected markets and performs better than the benchmark settings in terms of MAE and RMSE. The ablation analysis further indicates that hourly de-meaning, the Transformer module, the BiLSTM module, and feature screening all contribute to model performance. In addition, the SHAP results provide a model-based view of feature contributions and help relate the forecasting results to market characteristics in different regions.

This study still has some limitations. The analysis is based on three representative markets, and the generalizability of the framework to other market environments remains to be further examined. In addition, the SHAP results reflect feature contribution patterns learned by the model and should not be interpreted as causal relationships. Future work may extend the analysis to more market regions, incorporate a wider range of policy and structural variables, and further examine the application of the framework in trading, storage scheduling, and risk management under high VRE penetration.

List of abbreviations

Abbreviation	Definition
BiLSTM	Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
CNN-LSTM	Convolutional Neural Network–Long Short-Term Memory
Transformer	Transformer Network
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
R^2	Coefficient of Determination
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
EPF	Electricity Price Forecasting

Funding

This research received no external funding.

Author Contributions

Xuan Wu: Conceptualization, Data curation, Methodology, Validation, Software, Writing–original draft. **Yifei Liu**: Supervision, Formal analysis, Project administration, Writing–review and editing.

Data Availability Statement

Data will be made available on request.

Conflict of Interest

The authors declare no competing interests.

References

[1] Maciejowska, K.; Uniejewski, B.; Weron, R. Forecasting electricity prices. arXiv 2022, arXiv:2204.11735. https://doi.org/10.48550/arXiv.2204.11735.

[2] Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 2014, 30(4), 1030-1081. https://doi.org/10.1016/j.ijforecast.2014.08.008.

[3] Arantegui, R.L.; Jäger-Waldau, A. Photovoltaics and wind status in the European Union after the Paris Agreement. Renew. Sustain. Energy Rev. 2018, 81, 2460–2471. https://doi.org/10.1016/j.rser.2017.06.052.

[4] Wang, Y.; Wang, R.; Tanaka, K.; Ciais, P.; Penuelas, J.; Balkanski, Y.; Sardans, J.; Hauglustaine, D.; Cao, J.; Chen, J.; et al. Global Spatiotemporal Optimization of Photovoltaic and Wind Power to Achieve the Paris Agreement Targets. Nat. Commun 2025, 16, 2127. https://doi.org/10.1038/s41467-025-57292-w.

[5] ENTSO-E. System Flexibility Needs for the Energy Transition; ENTSO-E: Brussels, Belgium, 2024. Available online: https://www.entsoe.eu/system-flexibility/ (accessed on 15 November 2025).

[6] Cavus, M. Advancing Power Systems with Renewable Energy and Intelligent Technologies: A Comprehensive Review on Grid Transformation and Integration. Electronics 2025, 14, 1159. https://doi.org/10.3390/electronics14061159.

[7] Conejo, A.J.; Morales, J.M.; Baringo, L. Real-time demand response model. IEEE Transactions on Smart Grid, 2010, 1(3), 236-242. https://doi.org/10.1109/TSG.2010.2076840.

[8] Contreras, J.; Espínola, R.; Nogales, F.J.; Conejo, A.J. ARIMA models to predict next-day electricity prices. IEEE Trans. Power Syst. 2003, 18, 1014–1020. https://doi.org/10.1109/TPWRS.2002.804943.

[9] Garcia, R.C.; Contreras, J.; van Akkeren, M.; Garcia, J.B.C. A GARCH forecasting model to predict day-ahead electricity prices. IEEE Trans. Power Syst. 2005, 20, 867–874. https://doi.org/10.1109/TPWRS.2005.846044.

[10] Liu, H.; Shi, J. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. Energy Econ. 2013, 37, 152–166. https://doi.org/10.1016/j.eneco.2013.02.006.

[11] Tan, Z.; Zhang, J.; Wang, J.; Xu, J. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. Appl. Energy 2010, 87, 3606–3610. https://doi.org/10.1016/j.apenergy.2010.05.012.

[12] Gupta, S.; Chakrabarty, D.; Kumar, R. Predicting Indian electricity exchange-traded market prices: SARIMA and MLP approach. OPEC Energy Rev. 2023, 47, 271–286. https://doi.org/10.1111/opec.12287.

[13] Chai, S.; Li, Q.; Abedin, M.Z.; Lucey, B.M. Forecasting electricity prices from the state-of-the-art modeling technology and the price determinant perspectives. Res. Int. Bus. Finance 2024, 67, 102132. https://doi.org/10.1016/j.ribaf.2023.102132.

[14] Saini, D.; Saxena, A.; Bansal, R.C. Electricity price forecasting by linear regression and SVM. In Proceedings of the 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE 2016), Jaipur, India, 23–25 December 2016; pp. 1–7. https://doi.org/10.1109/ICRAIE.2016.7939509.

[15] Tschora, L.; Pierre, E.; Plantevit, M.; Robardet, C. Electricity price forecasting on the day-ahead market using machine learning. Applied Energy, 2022, 313, 118752. https://doi.org/10.1016/j.apenergy.2022.118752.

[16] Lago, J.; De Ridder, F.; De Schutter, B. Forecasting Spot Electricity Prices: Deep Learning Approaches and Empirical Comparison of Traditional Algorithms. Applied Energy 2018, 221, 386–405, https://doi.org/10.1016/j.apenergy.2018.02.069.

[17] Zhou, S.; Zhou, L.; Mao, M.; Tai, H.-M.; Wan, Y. An optimized heterogeneous structure LSTM network for electricity price forecasting. IEEE Access 2019, 7, 108161–108173. https://doi.org/10.1109/ACCESS.2019.2932999.

[18] Yang, H.; Schell, K.R. HFNet: Forecasting real-time electricity price via novel GRU architectures. In Proceedings of the 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2020), Liège, Belgium, 18–21 August 2020; pp. 1–6. https://doi.org/10.1109/PMAPS47429.2020.9183697.

[19] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008. https://doi.org/10.48550/arXiv.1706.03762.

[20] Wang, H. Prediction of Electricity Price Intervals Using Dynamic Bayesian Networks. Energy Inform 2025, 8, 127, http://doi.org/10.1186/s42162-025-00578-6.

[21] Lago, J.; Marcjasz, G.; De Schutter, B.; Weron, R. Forecasting Day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-source benchmark. Appl. Energy 2021, 293, 116983. https://doi.org/10.1016/j.apenergy.2021.116983.

[22] Cu, Y.; Wang, K.; Zhang, L.; Liu, Z.; Liu, Y.; Mo, L. A Time Series Decomposition-Based Interpretable Electricity Price Forecasting Method. Energies 2025, 18, 664. https://doi.org/10.3390/en18030664.

[23] Blyth, C.R. On Simpson's paradox and the sure-thing principle. J. Am. Stat. Assoc. 1972, 67, 364–366. https://doi.org/10.1080/01621459.1972.10482387.

[24] Sensfuß, F.; Ragwitz, M.; Genoese, M. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. Energy Policy 2008, 36, 3086–3094. https://doi.org/10.1016/j.enpol.2008.03.035.

[25] Jónsson, T.; Pinson, P.; Madsen, H. On the market impact of wind energy forecasts. Energy Econ. 2010, 32, 313–320. https://doi.org/10.1016/j.eneco.2009.10.018.

[26] Hogan, W.W. Electricity scarcity pricing through operating reserves. Econ. Energy Environ. Policy 2013, 2, 65–86. https://doi.org/10.5547/2160-5890.2.2.4.

[27] Clò, S.; Cataldi, A.; Zoppoli, P. The merit-order effect in the Italian power market: The impact of solar and wind generation on national wholesale electricity prices. Energy Policy 2015, 77, 79–88. https://doi.org/10.1016/j.enpol.2014.11.038.

[28] Khan, A.A.A.; Ullah, M.H.; Tabassum, R.; Kabir, M.F. Enhanced Transformer–BiLSTM deep learning framework for day-ahead energy price forecasting. IEEE Trans. Ind. Appl. 2025, Early Access. https://doi.org/10.1109/TIA.2025.3599812.

[29] Marcjasz, G.; Narajewski, M.; Weron, R.; Ziel, F. Distributional neural networks for electricity price forecasting. Energy Economics, 2023, 125, 106843. https://doi.org/10.1016/j.eneco.2023.106843.

[30] Jørgensen, B.H.; Wolter, C.; Moreno, C. IEA Wind TCP Annual Report 2024: Denmark; International Energy Agency Wind Technology Collaboration Programme: Paris, France, 2025. Available online: https://iea-wind.org/about-iea-wind-tcp/annual-reports/ (accessed on 8 February 2026).

[31] Energinet. Annual Report 2018: Energy across borders; Energinet: Fredericia, Denmark, 2018. Available online: https://en.energinet.dk/media/fjnfkw54/annual-report-2018.pdf (accessed on 8 February 2026).

[32] Rintamäki, T.; Siddiqui, A.S.; Salo, A. Does renewable energy generation decrease the volatility of electricity prices– An analysis of Denmark and Germany. Energy Economics, 2017, 62, 270-282. https://doi.org/10.1016/j.eneco.2016.12.019

[33] Newbery, D. Missing money and missing markets: Reliability, capacity auctions and interconnectors. Energy Policy 2016, 94, 401–410. https://doi.org/10.1016/j.enpol.2015.10.028.

[34] Ganepola, C.N.; Shubita, M.; Lee, L. The electric shock: Causes and consequences of electricity prices in the United Kingdom. Energy Econ. 2023, 126, 107030. https://doi.org/10.1016/j.eneco.2023.107030.

[35] Elexon. System Prices Analysis Report: January 2025; Elexon: London, UK, 2025. Available online: https://assets.elexon.co.uk/wp-content/uploads/sites/11/2025/02/21160740/System-Prices-Analysis-Report_-January-2025-Elexon-BSC.pdf (accessed on 8 February 2026).

[36] Durán-Castillo, G.; Weis, T.; Leach, A.; Fleck, B.A. Toward Sustainable Electricity Markets: Merit-Order Dynamics on Photovoltaic Energy Price Duck Curve and Emissions Displacement. Sustainability 2025, 17, 4618. https://doi.org/10.3390/su17094618.

[37] Pierro, M.; Moser, D.; Perez, R.; Cornaro, C. The value of PV power forecast and the paradox of the single pricing scheme: The Italian case study. Energies 2020, 13, 3945. https://doi.org/10.3390/en13153945.

[38] Hosseini Imani, M.; Bompard, E.; Colella, P.; Huang, T. Impact of wind and solar generation on the Italian zonal electricity price. Energies 2021, 14, 5858. https://doi.org/10.3390/en14185858.

[39] Kath, C.; Ziel, F. Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. Energy Econ. 2018, 76, 411–423. https://doi.org/10.1016/j.eneco.2018.10.005.

[40] Sioshansi, R.; Denholm, P.; Jenkin, T.; Weiss, J. Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects. Energy Econ. 2009, 31, 269–277. https://doi.org/10.1016/j.eneco.2008.10.005.

[41] Schill, W.-P. Residual load, renewable surplus generation and storage requirements in Germany. Energy Policy 2014, 73, 65–79. https://doi.org/10.1016/j.enpol.2014.05.032.

[42] Prokhorov, A.; Dreisbach, N. The impact of renewable energy generation on electricity price volatility in European markets. Renewable and Sustainable Energy Reviews 2022, 153, 111819. https://doi.org/10.1016/j.rser.2021.111819.

[43] Green, R.; Vasilakos, N. Market behaviour with large amounts of intermittent generation. Energy Policy 2010, 38, 3211–3220. https://doi.org/10.1016/j.enpol.2009.07.038.