Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Imbalance Price Forecasting in the Italian Energy Market Using Deep Neural Networks and Temporal Fusion Transformers

Davide Milillo

davide.milillo@uniroma3.it

Roma Tre University

Valentino Caputi

Roma Tre University

Rafiq Asghar

Roma Tre University

Husam S. Samkari

University of Tabuk

Michele Quercio

Roma Tre University

Francesco Riganti Fulginei

Roma Tre University

Research Article

Keywords: Deep Learning, Electricity Market, Imbalance Price Forecasting, Italian Energy Market, Temporal Fusion Transformer

Posted Date: September 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10499115/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Imbalance Price Forecasting in the Italian Energy Market Using Deep Neural Networks and Temporal Fusion Transformers

Davide Milillo[1*], Valentino Caputi[1], Rafiq Asghar[1], Husam S. Samkari[2,3], Michele Quercio[1], Francesco Riganti Fulginei[1]

[1*]Department of Industrial, Electronic and Mechanical Engineering (DIIEM), Roma Tre University, Via Vito Volterra 62, Rome, 00146, Italy.
[2]Department of Electrical Engineering, University of Tabuk, Tabuk, 47713, Saudi Arabia.
[3]Artificial Intelligence and Sensing Technologies Research Center, University of Tabuk, Tabuk, 47713, Saudi Arabia.

*Corresponding author(s). E-mail(s): davide.milillo@uniroma3.it;

Abstract

The Italian electricity imbalance market exhibits high volatility and non-linear dynamics, rendering deterministic forecasting inadequate. This paper proposes a probabilistic forecasting framework based on the Temporal Fusion Transformer (TFT) tailored to this highly stochastic context. The model integrates heterogeneous data—including historical imbalance prices, load and generation metrics, and future-known Day-Ahead zonal prices—enforcing a strictly causal, 12-hour multi-horizon prediction at the native 15-minute resolution. Methodologically, the TFT is formalized via quantile loss and Variable Selection Networks to meet operational trading constraints. Experimental results show that the model converges to a robust mean-value predictor with a sign accuracy of 62–65%, performing comparably to recurrent and tree-based baselines, thus confirming a deterministic accuracy ceiling. However, we demonstrate that the model's primary value lies in its probabilistic and interpretable nature. Quantile forecasts provide calibrated risk bands to identify low-risk biddable hours, while attention diagnostics reveal the dominant market drivers across temporal scales. The proposed framework operates as an explainable decision-support system for risk management rather than a deterministic price oracle.

Keywords: Deep Learning, Electricity Market, Imbalance Price Forecasting, Italian Energy Market, Temporal Fusion Transformer.

1 Introduction

Electricity is a highly stochastic commodity whose trading and settlement must continuously reconcile grid security with market efficiency. In Italy, this balance is achieved through a zonal market design, progressively integrated with European frameworks, in which institutional roles are strictly partitioned: the *Gestore dei Servizi Energetici* (GSE) manages renewable incentives and priority dispatch [1]; *Terna*, as Transmission System Operator, plans the grid, performs real-time balancing, and defines bidding zones [2–4]; the *Gestore dei Mercati Energetici* (GME) acts as Nominated Electricity Market Operator, clearing auctions and computing equilibrium prices [1, 2, 5]; and ENTSO-E coordinates pan-European data exchange and system integration [3, 6]. Prices are cleared through a zonal marginal mechanism in the Day-Ahead (MGP) and Intraday (MI) markets, while the Ancillary Services Market (MSD) uses pay-as-bid pricing [1, 3, 7].

Since the September 2021 integration into the Single Intraday Coupling under Regulation (EU) 2015/1222 [2, 8, 9], the Italian intraday segment has combined three sequential auctions (MI-A1, MI-A2, MI-A3) with continuous cross-border trading (MI-XBID), the latter extended to the French, Austrian, Slovenian, and Greek borders in June 2024 via the Euphemia algorithm [1, 10]. The resulting auction timeline, summarized in Figure 1, sets the operational deadlines within which traders must finalize their positions before delivery.

At delivery, Terna settles the deviation between scheduled and actual flows by activating balancing resources and imposing imbalance charges on Points of Delivery, following a three-tier pricing regime that distinguishes enabled (dual-priced) from non-enabled (single-priced) units, with dedicated tolerance bands for non-programmable renewables [8, 11]. Imbalances are classified as **Long** (excess injection, from over-production or under-consumption) or **Short** (deficit, from under-production or over-consumption) relative to the binding schedule [1, 4], with penalties or rewards assigned according to whether an individual position aggravates or offsets the aggregate zonal imbalance [11]. The settlement cost can be expressed as

$$\text{Cost} = (\text{Actual Energy} - \text{Scheduled Energy}) \times \alpha \times \text{Imbalance Price}, \tag{1}$$

where α — hereafter *Sign* — encodes the alignment between a trader's imbalance and the system-level imbalance direction [11]. This formulation identifies three complementary levers for cost minimization: (i) *load-energy forecasting*, which reduces the deviation term itself and remains an active, mature research area [12–15]; (ii) *sign prediction*, which allows traders to align or counter-position exposures; and (iii) *imbalance price forecasting*, the focus of this work, which quantifies exposure and supports hedging and bidding decisions.

Forecasting the Italian imbalance price is particularly challenging because redispatch costs and localized congestion can decouple intraday prices from real-time operating costs [16], while renewable penetration, cross-border flows, and an evolving regulatory framework add further non-stationarity. Notably, ARERA Resolution 304/2024 replaced the national single price (PUN) with full zonal pricing across seven zones, alongside a progressive shift to 15-minute market time units [17, 18], both of which materially affect forecast accuracy. Robust day-ahead prediction of these dynamics is therefore essential for traders seeking to optimize bidding strategies across the MGP and continuous MI-XBID phases.

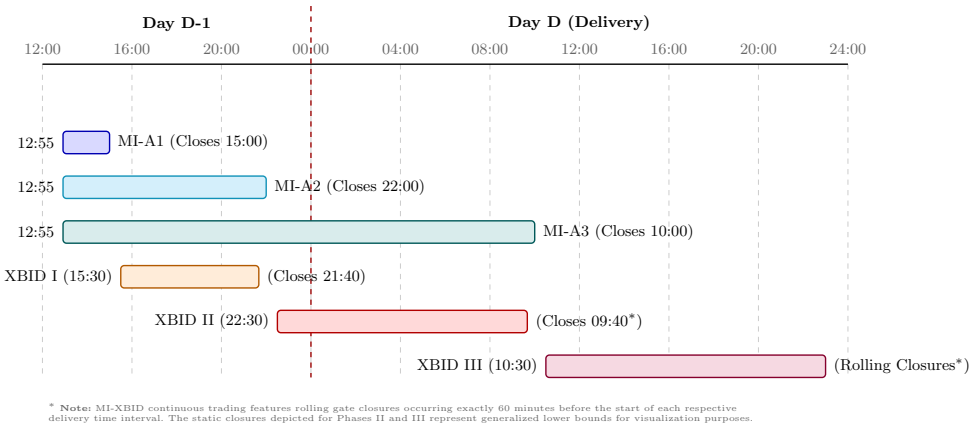


Fig. 1 Timeline of Italian Intraday Market (MI) and XBID Auction Sessions relative to Day D delivery. The chart illustrates the hybrid architecture combining sequential auctions with continuous cross-border trading windows.

A price forecast is, however, only an intermediate input: translating it into economic value requires embedding it within a broader decision layer, such as a metaheuristic optimization strategy [19], a reinforcement-learning trading agent [20, 21], generation-side forecasting [22], or a Smart Energy Management System coordinating flexible resources in real time [23, 24]. Existing imbalance-price studies, in contrast, largely stop at deterministic point forecasts, offering limited support for risk-aware operational decision-making.

This paper addresses that gap with an interpretable, uncertainty-aware forecasting framework aimed at supporting economic decision-making in the Italian imbalance market. Its contributions are three-fold:

- **A decision-support forecasting framework:** to the authors' knowledge, the first application of the Temporal Fusion Transformer (TFT) to the Italian electricity imbalance market, designed to feed directly into trading and risk-management decisions rather than to serve as a standalone predictive exercise.
- **Probabilistic rather than deterministic forecasting:** we show empirically that, in a congestion-driven zonal market, deterministic point predictions face

a hard accuracy ceiling, and reframe the operational objective around quantile-based uncertainty modeling, giving traders dynamic confidence intervals for risk mitigation.
- **Economically interpretable outputs:** Variable Selection Networks and temporal attention maps expose the market drivers behind each forecast, turning the model into a transparent decision-support tool rather than a black-box price oracle.

2 Literature Review

Accurate forecasting of the imbalance price is currently recognized as one of the most complex challenges in electricity market modeling. While load forecasting is a mature field for most Points of Delivery (PODs), the projection of settlement prices is plagued by extreme volatility. Consequently, energy traders and institutional players often bypass point price prediction, optimizing their Profit & Loss (P&L) strategies by focusing solely on anticipating the market sign (Long/Short) through statistical or deep learning models [25–28]. Existing literature demonstrates that forecasting the imbalance price in Italy is strictly bounded by a high degree of stochasticity, which depends only marginally on historical trends or exogenous fundamental features [29]. Analyzing zonal imbalance signs, traditional statistical approaches such as logit and Markov models achieve an accuracy fluctuating between 65% and 70% [25]. Even the application of state-of-the-art deep learning frameworks for day-ahead sign forecasting yields an Area Under the Curve (AUC) of approximately 0.7, translating to an estimated accuracy of 60%-65% [26]. Although these results grant evident economic benefits for market players via reduced imbalance costs, they remain structurally below the 80% target typically expected in more stationary financial markets. Similar predictive ceilings are observed internationally: in the Spanish market, Random Forest algorithms hit a comparable 68% sign accuracy [30]. When shifting from directional classification to point price forecasting, the intrinsic unpredictability becomes even more pronounced. In affine continental markets like Germany, advanced probabilistic machine learning models (including Lasso and GAMLSS) struggle to drop below a Mean Absolute Error (MAE) of 60 €/MWh, ultimately performing comparably to simple naïve baselines [31]. Only in highly liquid and geographically distinct systems like the UK, deep learning models can reach an MAE of approximately 7–8 £/MWh (around 9 €/MWh) [32]. In contrast, the predictive limit in the Italian market consistently manifests with MAE values remaining above 50–60 €/MWh [26, 29], despite the ingestion of advanced fundamental data, including day-ahead prices (PUN), zonal prices, load, renewable generation, and gas costs. This pronounced performance gap between Italy and other European hubs stems from structural, topological, and regulatory intrinsic characteristics rather than algorithmic limitations. Recent comparative studies highlight that, under identical LSTM network architectures, the forecasting error in Italy is approximately 2.3 times higher than in Germany [33]. This discrepancy is heavily driven by the transmission topology. Unlike central European grids, the Italian network is highly susceptible to internal congestion and structural network splits particularly regarding the electrical islands of Sardinia and Sicily which act as a primary driver for non-linear, extreme price volatility [34]. The magnitude of these

structural spikes is confirmed by European regulatory reports, which place the 90th percentile of the Italian imbalance price around 180 €/MWh, starkly contrasting with Germany's 85 €/MWh [35]. Furthermore, capacity projections indicate that Italian interconnection limits will persist well into 2030, maintaining a low-liquidity environment that inherently complicates algorithmic learning [36]. Finally, Italy's unique regulatory framework introduces further aleatoric uncertainty. The official methodology for calculating the imbalance premium includes a macro-zonal margin and specific coefficients published exclusively ex-post by the TSO [37]. This introduces a purely stochastic variable into the settlement equation, rendering a significant portion of the price variance strictly unobservable during the day-ahead and continuous trading phases, thereby confirming the necessity of a probabilistic, rather than deterministic, forecasting paradigm. Finally, Italy's unique regulatory framework introduces further aleatoric uncertainty. The official methodology for calculating the imbalance premium includes a macro-zonal margin and specific coefficients published exclusively ex-post by the TSO [37]. This introduces a purely stochastic variable into the settlement equation, rendering a significant portion of the price variance strictly unobservable during the day-ahead and continuous trading phases, thereby confirming the necessity of a probabilistic, rather than deterministic, forecasting paradigm.

While current literature thus establishes a clear performance ceiling for recurrent and ensemble models applied to electricity price forecasting [26, 29], these architectures suffer from a fundamental operational limitation that extends beyond error metrics: they operate as opaque deterministic oracles. Standard LSTM and GRU networks, despite their well-documented effectiveness at capturing sequential dependencies in univariate or low-dimensional settings [13, 15], are architecturally constrained to point outputs and cannot natively incorporate future known covariates—such as Day-Ahead zonal prices or load forecasts—without bespoke encoder-decoder modifications [38]. This structural rigidity is particularly penalizing in the Italian imbalance market, where the joint exploitation of heterogeneous data streams (past observations, deterministically known future inputs, and time-invariant static covariates) is essential to approaching the predictive ceiling identified in the literature [26, 29].

Alternative deep learning architectures address subsets of these limitations, but none simultaneously satisfies all requirements of the target task. N-BEATS [39] achieves strong univariate benchmark performance through its doubly residual stacking structure, but is inherently designed for univariate settings and provides no mechanism for ingesting exogenous future inputs or generating probabilistic outputs. DeepAR [40] natively supports probabilistic forecasting via autoregressive Gaussian sampling; however, it does not natively support future-known covariates and offers no feature-level interpretability over the driving market variables. Other study focus on probabilistic foundation models, like PriceFM [41], that uses complex graph neural network with mixture of expert to have a foundation model pre-trained on price-forecasting. Vanilla Transformer architectures [42], while theoretically suited to long-range dependency modelling via self-attention, have been shown empirically to underperform recurrent baselines in noisy, non-stationary financial time series [31, 33], likely due to their lack of a locality inductive bias. Table 1 summarises the architectural limitations of each competing approach relative to the requirements of this study.

Table 1 Architectural comparison of candidate deep learning models against the requirements of Italian imbalance price forecasting.

Model	Probabilistic	Future Covariates	Multi-input	Interpretable
LSTM / GRU [15]	✗	✗	Partial	✗
N-BEATS [39]	✗	✗	✗	Partial
DeepAR [40]	✓	✗	Partial	✗
Transformer [42]	✗	Partial	✓	✗
TFT [38]	✓	✓	✓	✓

The Temporal Fusion Transformer (TFT) [38] is uniquely positioned to address all four requirements simultaneously. First, its Variable Selection Networks (VSN) enable dynamic, input-specific gating via Gated Residual Networks (GRN), effectively suppressing non-informative stochastic noise from the large set of generation and load variables ingested in this study—a property absent in both standard LSTM and vanilla Transformer architectures [38]. Second, the TFT architecture explicitly partitions the input space into past observed, future known, and static covariate streams, allowing the decoder to exploit deterministically available signals such as Day-Ahead (MGP) zonal prices without introducing data leakage [1, 38]. Third, and critically for the risk management application studied here, the TFT is natively trained via quantile loss to produce calibrated uncertainty intervals rather than point estimates, transforming the forecast output into an actionable confidence band for energy traders [38, 40]. Finally, unlike black-box sequence models, the TFT's attention weights are directly inspectable through temporal attention maps and Variable Selection Networks, enabling operators to trace which market drivers and historical time steps contributed to each forecast—a capability that is indispensable in a regulatory environment subject to sudden structural shifts and non-stationary regime changes [17, 26, 38]. For these reasons, the TFT represents not merely an incremental architectural upgrade over existing approaches, but the only existing framework that simultaneously addresses the probabilistic output, multi-input heterogeneity, and interpretability requirements posed by the Italian imbalance market.

3 Methodology

3.1 Mathematical Formulation of the Temporal Fusion Transformer (TFT)

To navigate the extreme stochasticity and non-stationary regime shifts inherent to the Italian imbalance market, this study employs the Temporal Fusion Transformer (TFT), a state-of-the-art deep learning architecture specifically designed for multi-horizon time series forecasting [38]. Unlike traditional deterministic models, the TFT excels at aligning heterogeneous data streams—past observed inputs, future known inputs, and static time-invariant covariates—while utilizing specialized attention mechanisms to capture complex temporal dynamics [38].

Because point forecasting is structurally inadequate for grid topologies prone to severe congestion and localized price spikes, the true value of the TFT lies in its

probabilistic forecasting formulation. The network is trained to output an uncertainty interval rather than a singular scalar prediction, optimizing jointly across the $q \in \{0.1, 0.5, 0.9\}$ quantiles [38]. This is achieved by minimizing the Quantile Loss (or Pinball Loss) function over the prediction horizon:

$$L_q(y, \hat{y}) = \max(q(y - \hat{y}), (q - 1)(y - \hat{y})) \tag{2}$$

where y is the actual imbalance price and $\hat{y}$ is the quantile prediction. By providing a strict confidence interval, the model operates as an actionable decision-support system for risk management rather than a brittle deterministic oracle.

A fundamental contribution of the TFT architecture is its granular interpretability, distinguishing it from traditional "black-box" neural networks [38]. This is driven by the Variable Selection Network (VSN), which leverages a Gated Residual Network (GRN) to dynamically apply non-linear processing only where needed. For a given input variable c and primary input x, the GRN is defined as:

$$GRN(x, c) = LayerNorm(x + GLU(W_1 x + W_2 c + b_1)) \tag{3}$$

where GLU represents a Gated Linear Unit, which suppresses non-informative stochastic noise. The VSN computes continuous variable weights through a Softmax distribution over the GRN outputs, generating the specific attention maps that allow operators to trace exact market drivers [38]. Furthermore, to capture long-term cyclical dependencies, the TFT employs a temporal Multi-Head Attention mechanism within its decoder block:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{4}$$

where Q, K, and V represent the query, key, and value matrices, respectively [42].

3.2 Operational Architecture and Data Engineering

To validate the algorithmic efficacy in a production-ready environment, a robust, latency-tolerant MLOps pipeline was engineered. The complete data integration architecture, mapping the distinct feature spaces into the TFT encoder and decoder, is illustrated in Figure 2.

As depicted in Figure 2, the inputs are strictly partitioned to prevent data leakage. Static inputs encompass time-invariant or cyclical calendar features (day of week, month, hour). Past historical inputs, ingested by the TFT encoder, include the target variable (Imbalance Price North/South), aggregate grid load, and generation metrics segmented across 20 distinct energy sources. Crucially, the decoder leverages future known inputs—such as projected load, forecasted generation, and Day-Ahead (MGP) Zonal prices—which are deterministically established prior to the continuous intraday market gate closures [43].

Initially, the data collection layer relied on web scraping protocols interfacing with the GME portal [44]; however, this approach proved highly fragile under production

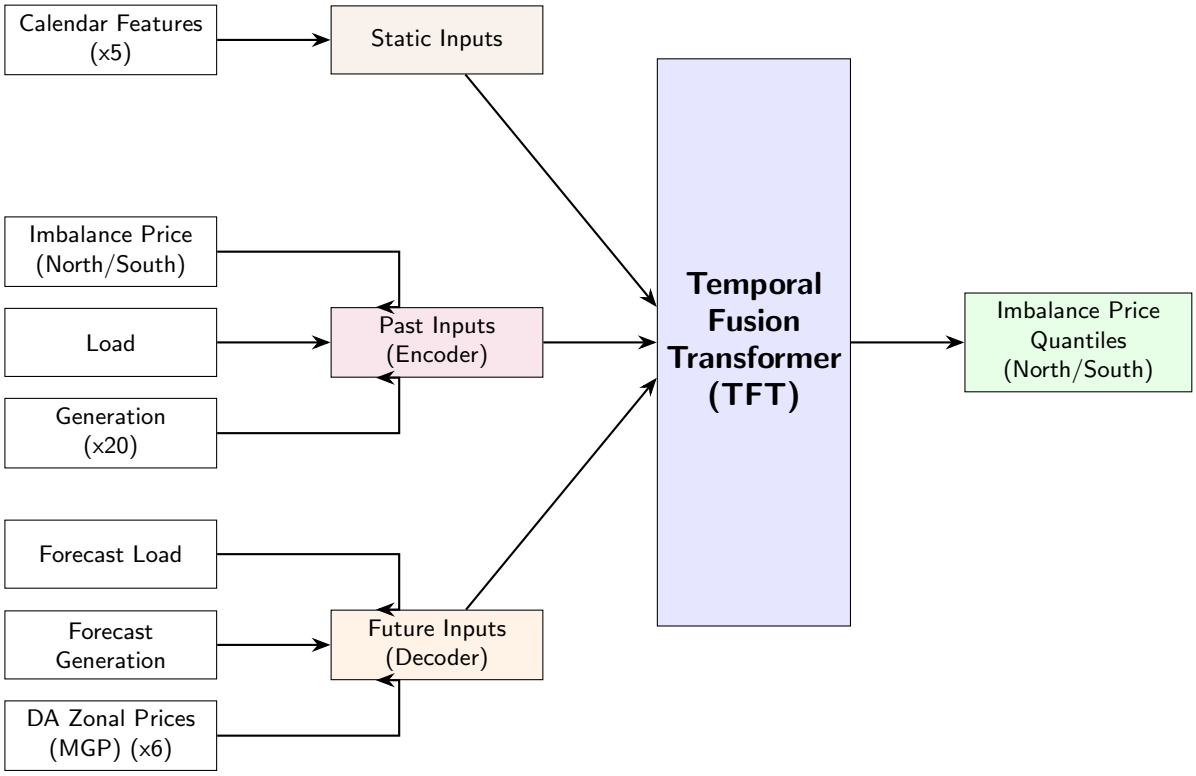


Fig. 2 Data integration architecture mapping static, historical (encoder), and future known (decoder) variables into the Temporal Fusion Transformer.

conditions due to anti-bot countermeasures. Consequently, the system architecture was migrated to ingest real-time streams directly via the Terna and ENTSO-E application programming interfaces (APIs) [43, 45].

A central operational challenge in real-time grid forecasting is the asynchronous and non-deterministic nature of API data availability, with Terna real-time metrics frequently subject to micro-delays (e.g., D-1 or a few hours delay) [43]. To counter this, a custom algorithm orchestrator was deployed on a Google Cloud Server. This orchestrator intelligently manages missing data by attempting to download the latest available data, executing rigorous validation, and stepping back hourly if data is null. Given these intrinsic API latencies and the operational requirements of energy traders, the optimal prediction horizon was calibrated to 12 hours, equivalent to 48 discrete 15-minute market time intervals.

The fully automated cloud-based workflow executes the following sequential tasks:

- **Data Collection:** Continuous downloading of real-time metrics from ENTSO-E and Terna APIs [43, 45].
- **Validation:** Algorithmic checking for null values and physical grid plausibility.

- **Database:** Structuring and storing raw metrics for historical traceability and scheduled online-training.
- **Preprocessing:** Application of standardization, normalization, and strict temporal partitioning to generate sequence blocks (known past, known future, static).
- **Training/Prediction:** Forward-pass execution of the TFT to generate the 48-period multi-horizon forecasts.
- **Evaluation:** Continuous real-time calculation of Mean Absolute Error (MAE), Mean Squared Error (MSE), and directional Market Sign accuracy (Long/Short).
- **Reporting:** Dispatching of operational summaries via a secure Telegram bot to facilitate remote monitoring.

3.3 Design Justification and Hyperparameter Selection

The prediction horizon was set to 48 discrete time steps, corresponding to 12 hours at the 15-minute market resolution mandated by the Italian electricity settlement framework [1, 4]. The 15-minute granularity is a hard operational constraint imposed by the MSD and imbalance settlement mechanisms [11]. The 12-hour horizon was selected following an explicit comparative evaluation across four candidate horizons (3, 6, 12, and 24 hours): accuracy degraded marginally beyond 12 hours, while shorter horizons yielded no statistically significant improvement. From an operational standpoint, a 3-hour horizon leaves insufficient reaction time for position adjustment across MI-XBID continuous trading phases [1], whereas a 24-hour horizon introduced additional non-stationary uncertainty without a corresponding gain in predictive accuracy [8].

The TFT was trained to jointly optimize three quantile levels $q \in \{0.1, 0.5, 0.9\}$. The median $q = 0.5$ serves as the primary point forecast. Empirical analysis of the residuals confirmed that the upper and lower prediction bounds were approximately equidistant from the realized price, indicating no systematic bias that would justify an asymmetric quantile configuration. The outer quantiles define an 80% confidence interval, providing traders with a calibrated risk band for hedging decisions [31]. Given the extreme spike behaviour of the Italian imbalance market [26], a wider confidence band is operationally more prudent, as underestimating tail risk carries a higher economic cost than overestimating it [11].

The final configuration (hidden dimension: 4, attention heads: 2, LSTM layers: 2, dropout: 0.2, learning rate: 2×10^{-4}, batch size: 64, encoder length: 168 periods) was arrived at through an extensive manual search conducted over approximately two months, spanning hidden dimensions from 4 to 128 and attention heads from 2 to 16. A fundamental observation emerging from this search is what we term the *overfitting–generalization dichotomy*: regardless of architectural scale, the TFT converged to one of only two qualitatively distinct regimes. In the first, induced by insufficient regularization, the network achieved high training-set fidelity by mapping directional changes with near-perfect accuracy, yet suffered catastrophic out-of-sample decay. In the second, enforced through dropout and early stopping, the network abandoned peak tracking and converged to a stable mean-value predictor, achieving consistent sign accuracy of approximately 62–65%. Critically, increasing model capacity did not yield any intermediate regime; larger models merely overfit more rapidly. Early stopping with patience of 20 epochs was applied monitoring validation loss; in practice,

the TFT consistently ceased improving between epochs 3 and 8, indicating rapid saturation of the learnable signal in the training data [26, 29]. This behavioral pattern strongly suggests that the performance ceiling is driven by the intrinsic aleatoric uncertainty of the Italian imbalance market rather than by model capacity or architectural choice [37], consistent with the theoretical arguments presented in Section 2.

4 Experimental Results

Before detailing the specific market outcomes, this section provides a synthetic overview of the extensive experimental campaign conducted to evaluate the Temporal Fusion Transformer's capabilities. The architecture was rigorously tested across multiple Italian market segments. The performance on sequential auction prices—specifically the Day-Ahead (PUN) and continuous Intraday (XBID) markets—yielded highly promising results with remarkably low error margins. Conversely, forecasting the final imbalance price demonstrated only marginal improvements over baseline models, such as LSTM, XGBoost, or Naive approaches.

This pronounced discrepancy empirically suggests that the extreme, congestion-driven volatility of the imbalance market acts as an intrinsic structural limit, rendering deterministic point accuracy practically unattainable. Nevertheless, despite this hard performance ceiling, the true operational value of the TFT is realized through its probabilistic outputs and granular interpretability. By shifting the focus from absolute precision to explainable uncertainty intervals and attention-driven market insights, the framework overcomes the limitations of traditional "black-box" models, providing energy traders with a uniquely powerful decision-support tool.

4.1 Day-Ahead and Intraday Markets (PUN, Zonal, XBID)

The model was first applied to auction markets (Day-Ahead and Intraday). While the primary interest lies in imbalance prices, these results demonstrate the model's feasibility.

- **PUN:** Results were excellent, with errors below 1 €/MWh. The model correctly predicted peaks, though sometimes exhibiting overshooting in intensity.
- **Zonal/XBID:** Performance was good (MAE $<$ 10 €/MWh), though not at the PUN level. Accuracy degraded drastically when extending the horizon beyond a few hours to 12 hours or 1 day.

Comparisons with existing TFT applications [46] show that our methodology leverages the architecture more fully by incorporating static and future known inputs, whereas previous academic exercises often relied solely on past PUN/GAS data and incorrect train-validation splits (predicting past vs. future).

4.2 Imbalance Price Forecasting

For the imbalance price, the results show significant room for improvement, demonstrating that market volatility is intrinsic to the imbalance price and depends only minimally on historical trends or exogenous variables. The results are summarized as follows:

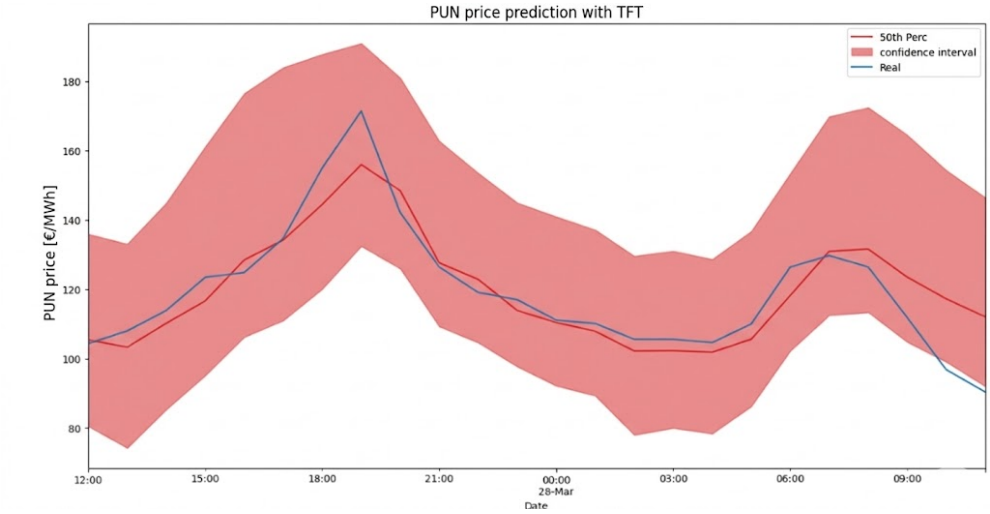


Fig. 3 PUN prediction using TFT. Data taken from GME with web scraping, in a real time test. The Deep Neural Network can successfully predict the market behaviour, the price peak and also the stress period such as the valley of 30 March at 12:00. The overall MAE was of 4.52 €/MWh.

- **Performance:** The model reached an error of MAE $\approx$ 62 €/MWh and a sign prediction accuracy of $\approx$ 65%. While these values are scientifically aligned with existing literature, they remain insufficient for high-stakes commercial trading where margins are extremely tight [47].
- **Variable Sensitivity:** Input variables such as generation, load, and zonal prices were found to have a very limited weight. Accuracy remained nearly constant regardless of the feature set, suggesting that the model struggles to extract actionable information from the data beyond a general baseline.
- **Horizon Invariance:** Shifting the forecasting horizon from 12h to 3h yielded no tangible improvements in MAE or sign accuracy. This indicates that the model's predictions are essentially mean-value estimates that are already saturated with intrinsic uncertainty.
- **Model Behavior:** The model predicts in a robust way the averages of the energy price, also during heavy changes in the electricity market, but struggled to extract useful information beyond averages.
- **Baseline Comparison:** The complex TFT performed similarly to LSTM and XGBoost and only marginally better than a Naïve model that predicts the price as the price of yesterday (Price[t] = Price[t-96], with 96 periods equal to 24 hours), Table 2, with the main advantage being prediction stability rather than error reduction.

A critical aspect of this analysis is the "Peak-Tracking vs. Generalization" trade-off. The model initially achieved high fidelity on the training set, mapping directional changes with 100 % accuracy . However, this was the result of heavy training without validation, leading to severe overfitting. When applied to the test set, the attempt to

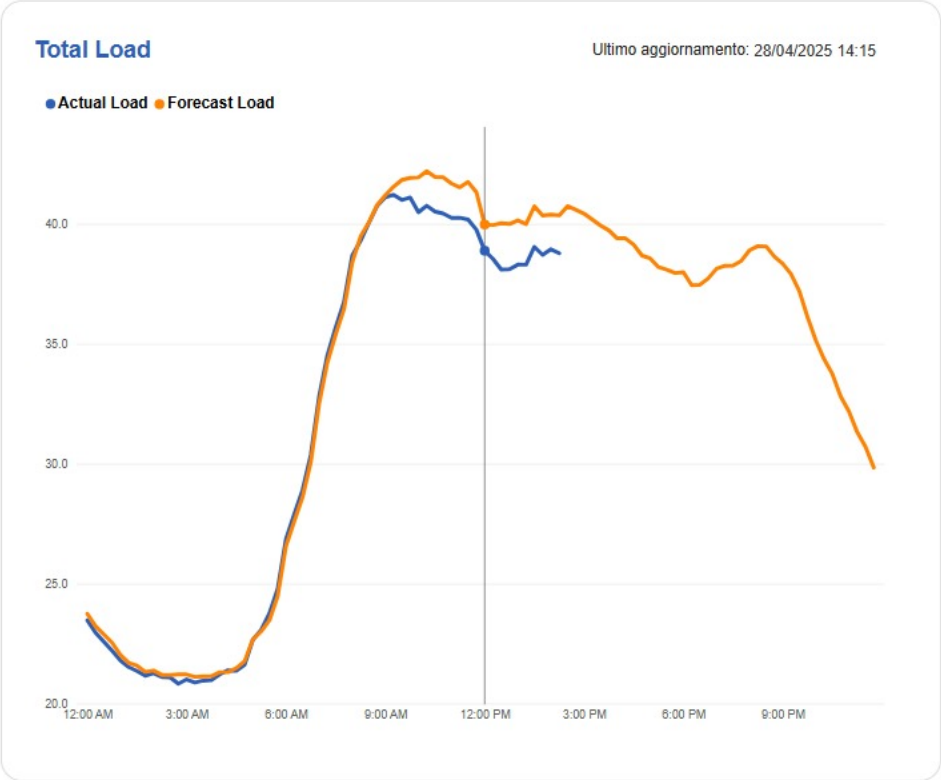


Fig. 4 PUN forecasting known only past PUN, load and generation as input variables. Increasing the input variables increased the model performance, since the quantile q0.5 follow more precisely the real value and the uncertainty interval is narrower then test made with less variables.

track volatile peaks resulted in a sharp performance decline (Figures 6 and 8), with sign accuracy dropping significantly (Figure 7. This confirms that peak behavior on day x does not necessarily correlate with day $x + 1$ due to the stochastic nature of market drivers.

4.3 Interpretability Analysis

The primary advantage of the TFT architecture lies in its intrinsic interpretability through attention mechanisms. Our training process revealed that early experiments were prone to "overthinking" the input space. In the August 26 attention maps, the network assigned spurious weights to variables such as geothermal and offshore wind generation, which resulted in catastrophic sign accuracy failures of 29.2 % (Figure 9).By implementing validation sets and early stopping (test of August 27), we forced the model to generalize. The optimized variable selection map (Figure 11) shows a more selective focus on fundamental drivers: the NORD imbalance price and the "Year" variable. The latter likely acts as a structural baseline to capture non-stationary

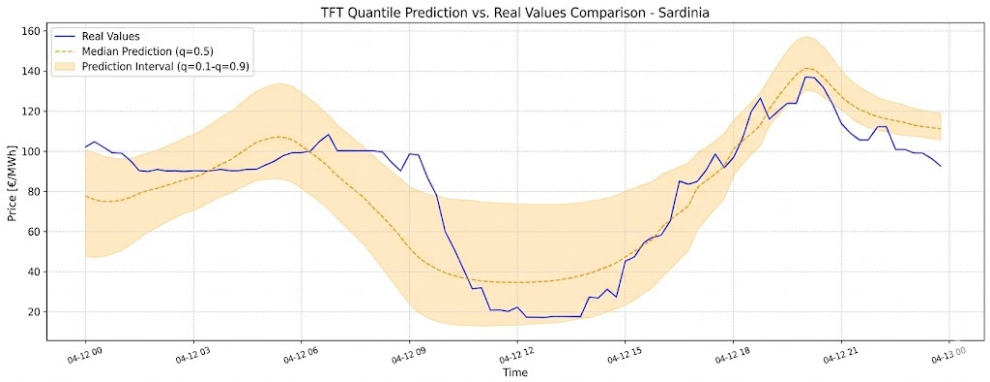


Fig. 5 XBID forecasting in Market Zone of Sardinia. In blue the real vlaue of the XBID price, in yellow the TFT predictions. On the y-axis there is the price in €/MWh; on the x-axis there is the date (12th of April). This is the first version of the orchestrator that takes data with web scraping. The MAE for this simulation, not yet optimized, was of $MAE \approx 17$ €/MWh .

Table 2 Performance Comparison of Forecasting Models across Italian Electricity Markets. Metrics are reported in Mean Absolute Error (MAE) [€/MWh].

Model	PUN	Zonal Prices	XBID	Imbalance Price
Naïve	17	20	25	67.50
XGBoost	5	11	12	63.80
LSTM	4	9	10	63.10
TFT	4	8	9	62.15

shifts in market regulations or energy mix between 2024 and 2025.While the model eventually learned to predict robust average prices (Figure 12) and estimated uncertainty intervals during high-volatility periods , it effectively "gave up" on deterministic peak forecasting. This aligns with expert feedback: peaks are driven by unpredictable, daily-changing factors. Consequently, the TFT's real value for a trader lies in providing a stable directional baseline ($\approx$ 62.5 % accuracy and quantifying the risk of potential spikes through uncertainty estimation rather than pinpointing them exactly. The attention weights, that can be seen from the attention maps in Figures 10, 11, also were extremely useful for a machine learning developer to understand what the model is learning from, and from an energy trader point of view to give insight in the energy market that can help proposing new strategy. Despite the error rate reached by the model, the TFT can give some useful insight to energy trader.

5 Conclusion

In conclusion, this research presents a robust, production-ready pipeline for real-time imbalance price forecasting within the highly volatile Italian electricity market. By

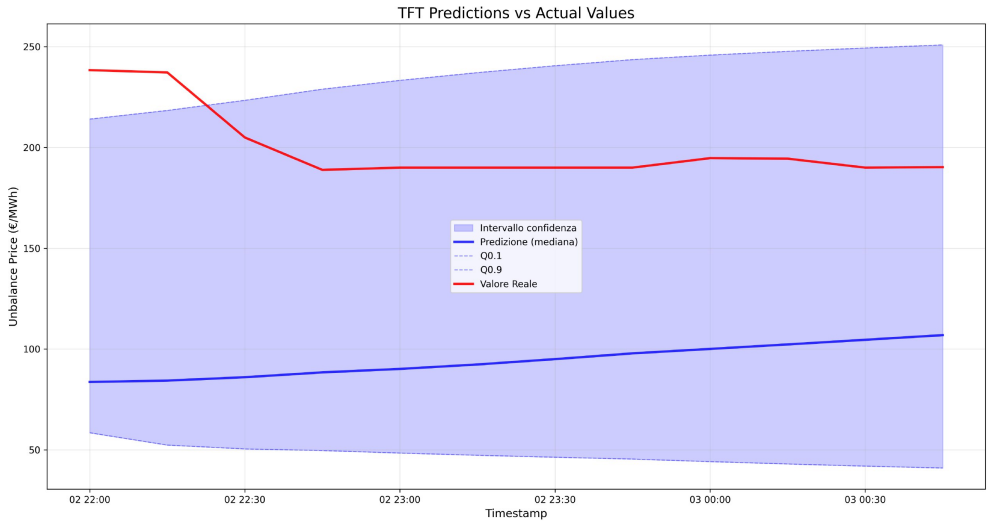


Fig. 6 Performance decay on the test set (March 21, 2025). The divergence between forecast and actual peaks demonstrates the intrinsic difficulty of deterministic tracking in volatile markets where peaks on day x do not correlate with those on day $x + 1$.

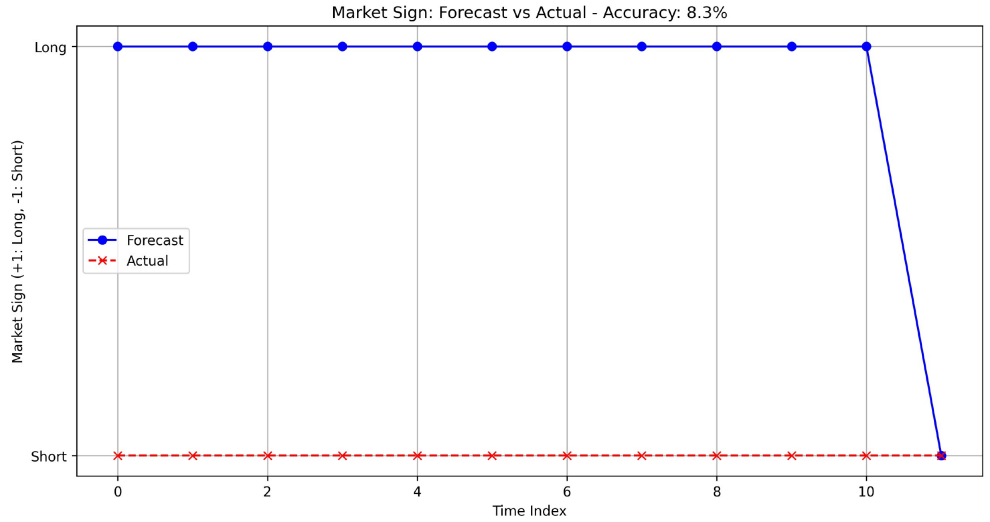


Fig. 7 Test set sign accuracy (60.4%), of the prediction of 6 The sharp decline from training levels highlights the model's limited ability to extract useful information from highly volatile data without specialized regularization.

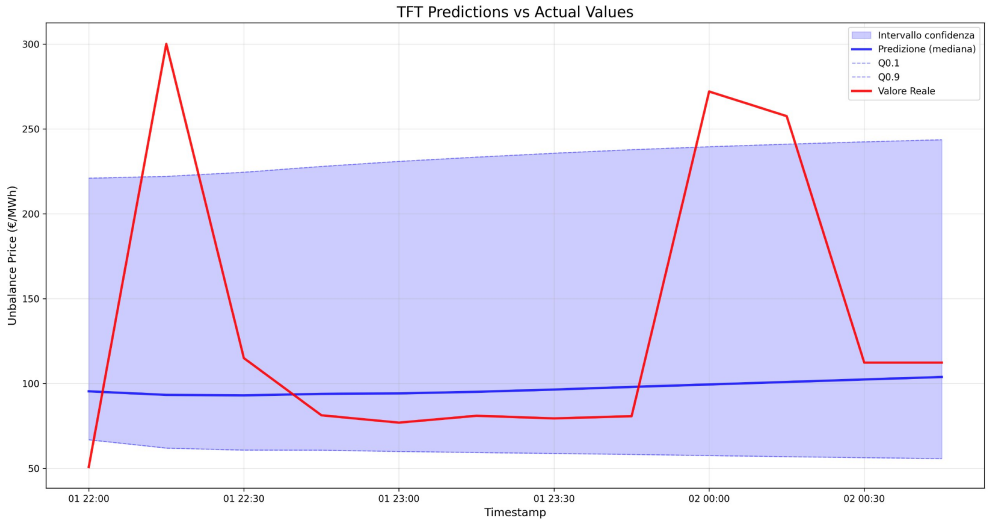


Fig. 8 Further validation on the March 22nd test set, confirming the structural challenge of imbalance forecasting. The model's failure to capture extreme price spikes remains consistent across different test periods. The Market sign accuracy for this prediction is of 54.2 %.

integrating live data streams from Terna and ENTSO-E via an orchestrated architecture, the system successfully enables continuous online training and inference. However, the experimental evaluation confirms a fundamental characteristic of the Italian grid: the intrinsic stochasticity and non-stationary dynamics of the imbalance market impose a hard limit on deterministic predictability. As demonstrated by the results, the Temporal Fusion Transformer (TFT) reached a performance ceiling with a Mean Absolute Error (MAE) of approximately 62 €/MWh. This metric, while aligned with recent state-of-the-art literature [26, 29], underscores that attempting to pinpoint extreme, congestion-driven price spikes often leads to severe network overfitting. The model's initial tendency to map training set volatility resulted in rapid out-of-sample performance decay. Through rigorous validation and early stopping, the network instead optimized for a stable mean-value prediction, achieving a directional sign accuracy of roughly 62.5%. Consequently, the primary contribution of this work lies not in overcoming the deterministic error barrier, but in shifting the forecasting paradigm. The TFT architecture provides critical value through its probabilistic quantile outputs and granular interpretability. By analyzing Variable Selection Networks and Temporal Attention maps, traders can decipher the underlying drivers of market behavior and recognize periods of extreme unobservable variance. This allows operators to identify low-risk "biddable hours" and optimize hedging strategies based on confidence intervals rather than relying on brittle point estimates. Ultimately, the framework operates as an advanced decision-support system for sophisticated risk management and market sign prediction rather than a simple price oracle. Regarding future developments to improve the model's utility, efforts should first focus on introducing new external variables to provide a richer informational context. Once this broader dataset is

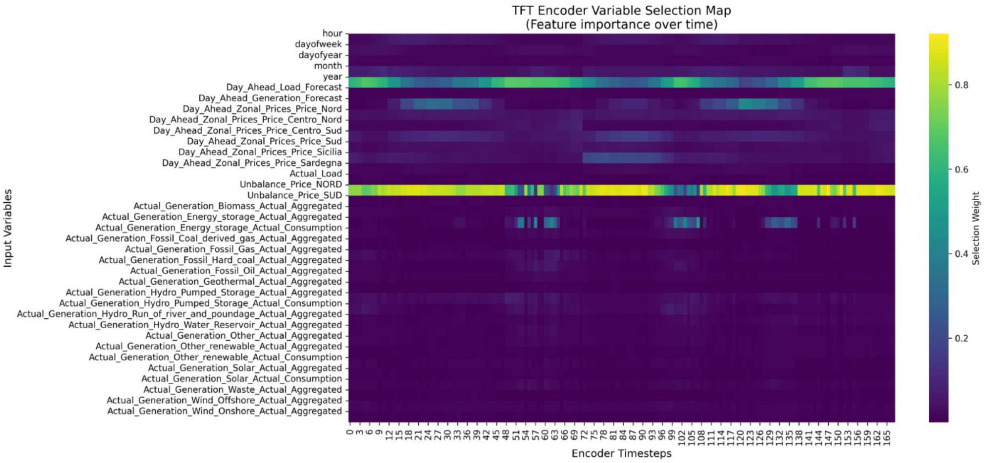


Fig. 9 Forecast under heavy overfitting (August 26). While the model approximates historical peaks in the training set, it fails to translate these insights into viable future predictions. In sign accuracy it performed with a very low accuracy, 29.2% due to overfitting. This result highlights that a model tuned too tightly to training-set volatility can perform significantly worse than a random baseline in real-world test sets.

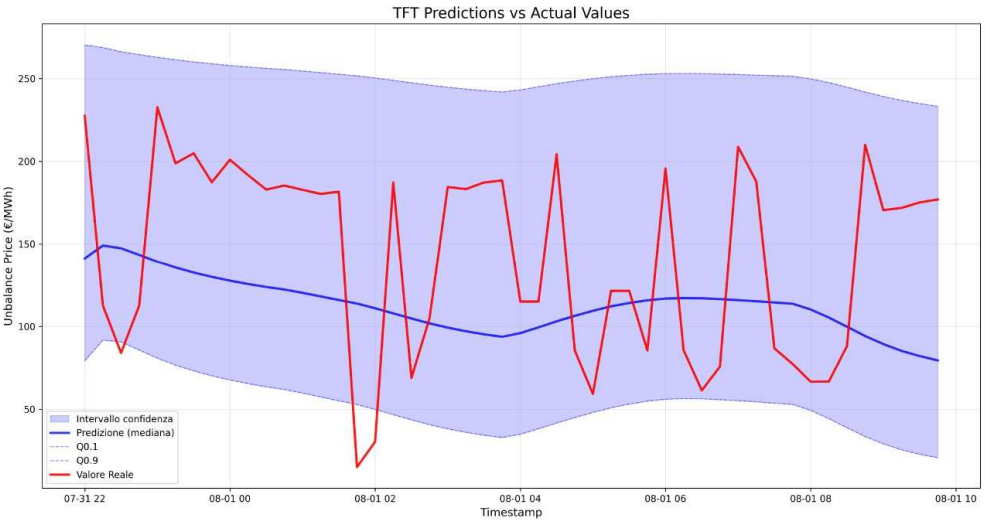


Fig. 10 Optimized decoder attention (August 27). The network identifies a more selective set of future drivers, reflecting the benefits of early stopping and validation-set regularization.

established, advanced feature engineering can be applied to extract specific temporal dynamics and better capture underlying market states. Furthermore, to achieve a comprehensive performance validation, the preliminary comparative metrics currently presented in Table 2 must evolve into a definitive benchmark through standardized

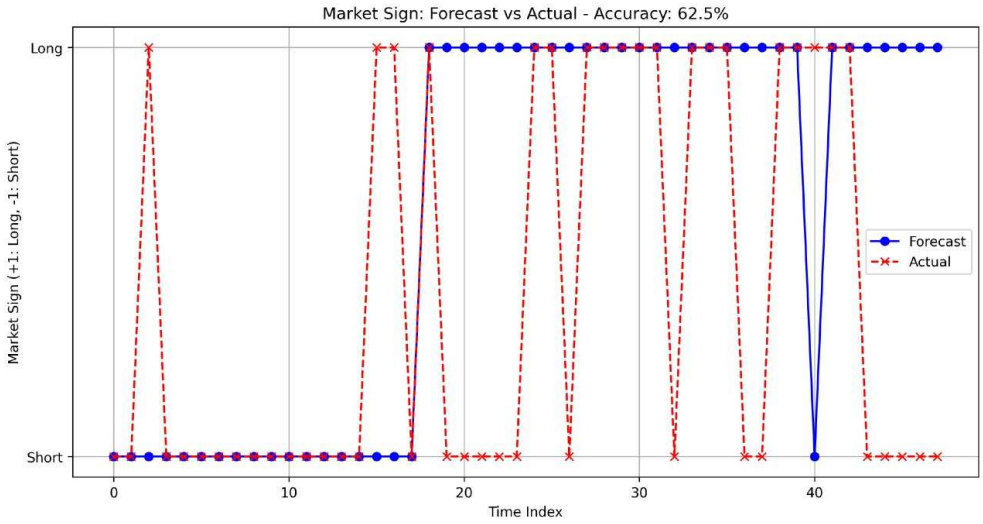


Fig. 11 Variable selection map (August 27) highlighting the dominance of the NORD imbalance price and the "Year" variable. The latter serves as a structural baseline for non-stationary market shifts that occur between different operational years.

simulations that isolate environmental and operational variability. A multi-seasonal study, spanning an entire year, is therefore essential to strictly validate the TFT's reliability against long-term data drifts and varying grid conditions, ultimately providing a statistically significant baseline for future real-time applications.

Declarations.

- **Funding:** The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- **Competing interests:** The authors have no relevant financial or non-financial interests to disclose.
- **Ethics approval:** Not applicable.
- **Consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Data availability:** The datasets analyzed during the current study are publicly available from the official Terna and ENTSO-E transparency platforms.
- **Materials availability:** Not applicable.
- **Code availability:** The custom code supporting the findings of this study is available from the corresponding author upon reasonable request.
- **Author contribution:** All authors contributed to the study conception, methodology, software implementation, data analysis, and manuscript drafting. All authors read and approved the final manuscript.

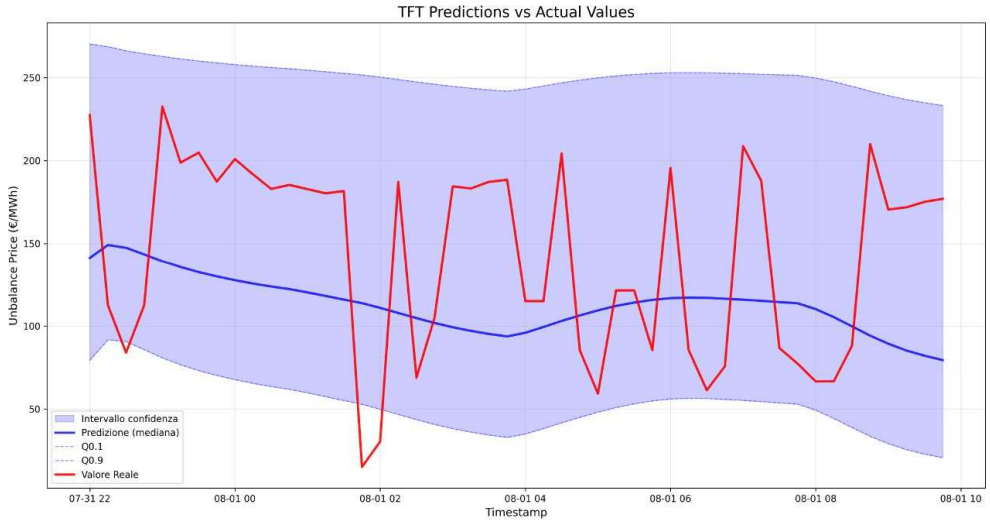


Fig. 12 Robust mean-value forecasting (August 27). The model successfully avoids the erratic behavior of Fig. 9, focusing on a sustainable average price level. Stable sign accuracy (62.5%) achieved through regularized training (August 27). This level of performance is consistent with the current state-of-the-art for imbalance markets characterized by intrinsic volatility.

References

[1] Gestore dei Mercati Energetici (GME): Vademecum to the Italian Power Exchange. Available: https://www.mercatoelettrico.org (2025)

[2] Alberizzi, A., Zani, A., Di Barba, P.: Impact of renewables resources into the italian intraday electricity market. International Journal of Smart Grid and Clean Energy (2023) https://doi.org/10.12720/sgce.12.4.81-90

[3] Gianfreda, A., Parisio, L., Pelagatti, M.: The RES-induced switching effect across fossil fuels: An analysis of day-ahead and balancing prices. Energy Journal (2019) https://doi.org/10.5547/01956574.40.1.agia

[4] Terna S.p.A.: Codice di Rete - Regole Tecniche del Dispacciamento. [Online]. Available: https://www.terna.it (2024)

[5] Zedda, S., Masala, G.B.: Price spikes in the italian electricity market: An economic analysis. Modern Economy **10**(9), 2145–2156 (2019) https://doi.org/10.4236/me.2019.109135

[6] Imani, M.H., Bompard, E., Colella, P., Huang, T.: Impact of wind and solar generation on the italian zonal electricity price. Energies (2021) https://doi.org/10.3390/en14185858

[7] Carlini, E., Caprabianca, M., Falvo, M., *et al.*: Proposal of a new procurement

strategy of frequency control reserves in power systems: The italian case in the european framework. Energies (2021) https://doi.org/10.3390/en14196105

[8] Ventura, C., Tina, G.M., Rizzo, S.: Imbalance charge reduction in the italian intra-day market using short-term forecasting of photovoltaic generation. Energies (2025) https://doi.org/10.3390/en18154161

[9] European Commission: Commission regulation (EU) 2015/1222 of 24 july 2015 establishing a guideline on capacity allocation and congestion management. Official Journal of the European Union (2015). [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R1222

[10] Agency for the Cooperation of Energy Regulators (ACER): Decision No 01/2019 on the intraday cross-zonal gate opening and gate closure times. [Online]. Available: https://www.acer.europa.eu (2019)

[11] Autorità di Regolazione per Energia Reti e Ambiente (ARERA): Resolution 345/2023/R/eel - Testo Integrato della Disciplina del Dispacciamento Elettrico (TIDE). [Online]. Available: https://www.arera.it (2023)

[12] Girolimetto, D., Di Fonzo, T.: Energy load forecasting using Terna public data: a free lunch multi-task combination approach. Preprint at https://arxiv.org/abs/2502.11873 (2025)

[13] Di Persio, L., Cecchin, A., Cordoni, F.: Novel approaches to the energy load unbalance forecasting in the italian electricity market. Journal of Mathematics in Industry **7**, 1–15 (2017) https://doi.org/10.1186/s13362-017-0035-y

[14] Tarsitano, A., Amerise, I.L.: Short-term load forecasting using a two-stage sarimax model. Energy **133**, 108–114 (2017) https://doi.org/10.1016/j.energy.2017.05.126

[15] Vasenin, D., Pasetti, M., Rinaldi, S., Golovinski, P., Savvin, N.: Long-term electrical energy forecasting of the residential sector using the LSTM model: The italian use case. In: 2023 International Conference on Future Energy Solutions (FES), pp. 1–6 (2023). https://doi.org/10.1109/fes57669.2023.10182653

[16] Graf, C., Quaglia, F., Wolak, F.: Simplified electricity market models with significant intermittent renewable capacity: Evidence from italy. Social Science Research Network (2020) https://doi.org/10.3386/w27262

[17] Autorità di Regolazione per Energia Reti e Ambiente (ARERA): Resolution 304/2024/R/eel - Amendments to the Integrated Text of the Electricity Dispatching Rules (TIDE). [Online]. Available: https://www.arera.it (2024)

[18] Ministry of Environment and Energy Security: Decree of 18 April 2024 implementing the provisions of art. 13 of Legislative Decree 210/2021. [Online]. Available:

https://www.mase.gov.it (2024)

[19] Asghar, R., Quercio, M., Milillo, D., Ullah, M., Shah, S.Z., Riganti Fulginei, F.: ADHHO: an adaptive harris hawk optimization algorithm for solving global optimization problems. Engineering Computations **43**, 1–37 (2026) https://doi.org/10.1108/EC-12-2025-1486

[20] Perquoti, F., Milillo, D., Sabino, L., Quercio, M., Riganti Fulginei, F., Lazaroiu, G.C., Crescimbini, F.: Reinforcement learning-based energy storage management for microgrid power exchanges. Eng **7**(3) (2026) https://doi.org/10.3390/eng7030126

[21] Zhao, H., et al.: TimeSeriesScientist: A General-Purpose AI Agent for Time Series Analysis. Preprint at https://arxiv.org/abs/2510.01538 (2025)

[22] Asghar, R., Quercio, M., Sabino, L., Milillo, D., Mahrouch, A., Riganti Fulginei, F.: Solar irradiance estimation using a hybrid deep learning model with meta-heuristic algorithms. In: ISAS 2025 - 9th International Symposium on Innovative Approaches in Smart Technologies, Proceedings (2025). https://doi.org/10.1109/ISAS66241.2025.11101966

[23] Quercio, M., Milillo, D., Micallef, A., Riganti Fulginei, F., Licari, J.: Smart energy management for industrial buildings: A solar and EV charging case study in malta. In: 2025 International Conference on Clean Electrical Power (ICCEP 2025), pp. 595–600 (2025). https://doi.org/10.1109/ICCEP65222.2025.11143640

[24] Becchi, L., Bindi, M., Grasso, F., Intravaia, M., Lozito, G.M., Luchetta, A.: Distributed EMS coordination via price-signal control for renewable energy communities. Energies **18**(22) (2025) https://doi.org/10.3390/en18226072

[25] Lisi, F., Edoli, E.: Analyzing and forecasting zonal imbalance signs in the italian electricity market. The Energy Journal **39**(5) (2018) https://doi.org/10.5547/01956574.39.5.flis

[26] Carnevale, D., Cavaiola, M., Mazzino, A.: A novel AI-assisted forecasting strategy reveals the energy imbalance sign for the day-ahead electricity market. Energy Reports **11**, 4115–4126 (2024) https://doi.org/10.1016/j.egyr.2024.03.058

[27] Andreotti, D., Spiller, M., Scrocca, A., *et al.*: Modeling and analysis of BESS operations in electricity markets: Prediction and strategies for day-ahead and continuous intra-day markets. Sustainability (2024) https://doi.org/10.3390/su16187940

[28] Bigerna, S., Bollino, C.: A system of hourly demand in the italian electricity market. The Energy Journal (2015) https://doi.org/10.5547/01956574.36.4.sbig

[29] Golia, S., Grossi, L., Pelagatti, M.: Machine learning models and intra-daily market information for the prediction of italian electricity prices. Forecasting **5**(1), 81–101 (2023) https://doi.org/10.3390/forecast5010003

[30] Balázs, I., Fodor, A., Magyar, A.: Short-term system imbalance forecast using linear and non-linear methods. Energy Systems (2024) https://doi.org/10.1007/s12667-024-00667-7

[31] Narajewski, M.: Probabilistic forecasting of german electricity imbalance prices. Energies **15**(14), 4976 (2022) https://doi.org/10.3390/en15144976

[32] Deng, S., Inekwe, J., Smirnov, V., Wait, A., Wang, C.: Seasonality in deep learning forecasts of electricity imbalance prices. Energy Economics **137**, 107770 (2024) https://doi.org/10.1016/j.eneco.2024.107770

[33] Ziel, F., Steinert, R.: Electricity price forecasting using sale and purchase curves: The x-model. Energy Economics **59**, 435–454 (2016) https://doi.org/10.1016/j.eneco.2016.08.008

[34] Sapio, A., Spagnolo, N.: The effect of a new power cable on energy prices volatility spillovers. Energy Policy **144**, 111488 (2020) https://doi.org/10.1016/j.enpol.2020.111488

[35] Agency for the Cooperation of Energy Regulators (ACER): Activation of Balancing Energy 2023. Cap. 4. [Online]. Available: https://www.acer.europa.eu (2023)

[36] European Network of Transmission System Operators for Electricity (ENTSO-E): Ten-Year Network Development Plan (TYNDP) 2024 - National Trends. [Online]. Available: https://tyndp.entsoe.eu (2023)

[37] Terna S.p.A.: Metodologia di calcolo PREMIO DI SBILANCIAMENTO. [Online]. Available: https://www.terna.it (2022)

[38] Lim, B., Arık, S.Ö., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting **37**(4), 1748–1764 (2021) https://doi.org/10.1016/j.ijforecast.2021.03.012

[39] Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: Proceedings of the 8th International Conference on Learning Representations (ICLR) (2020). [Online]. Available: https://arxiv.org/abs/1905.10437

[40] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting **36**(3), 1181–1191 (2020) https://doi.org/10.1016/j.ijforecast.2019.07.001

[41] Yu, R., Gu, C., Stiasny, J., Wen, Q., Dilov, W.S., Qi, L., Cremer, J.L.: PriceFM: Foundation Model for Probabilistic Electricity Price Forecasting. Preprint at https://arxiv.org/abs/2508.04875 (2025)

[42] Vaswani, A., *et al.*: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017). https://papers.nips.cc/paper/7181-attention-is-all-you-need

[43] Terna S.p.A.: Codice di Rete per la trasmissione, il dispacciamento, lo sviluppo e la sicurezza della rete. [Online]. Available: https://www.terna.it

[44] Gestore dei Mercati Energetici (GME): Vademecum della Borsa Elettrica Italiana. [Online]. Available: https://www.mercatoelettrico.org

[45] Entso-E: Transparency Platform. https://transparency.entsoe.eu/

[46] D'Elia, P.: Italian Electric Market Data. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/paolodelia/italian-electric-market-data (2022)

[47] Neuhoff, K., Batlle, C., Brunekreeft, G., et al.: Flexible Short-Term Power Trading: Gathering Experience in EU Countries (2015). https://doi.org/10.2139/ssrn.2639088