Energy Reports 11 (2024) 4115–4126

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

A novel AI-assisted forecasting strategy reveals the energy imbalance sign for the day-ahead electricity market

Daniele Carnevale [a,d], Mattia Cavaiola [b], Andrea Mazzino [a,c,*]

[a] DICCA, Department of Civil, Chemical and Environmental Engineering, Via Montallegro 1, Genova, 16145, Italy
[b] CNR - National Research Council of Italy, Institute of Marine Sciences, Via S.Teresa S/N, 19032, Pozzuolo di Lerici, La Spezia, Italy
[c] INFN, Istituto Nazionale di Fisica Nucleare, Genova section, Via Dodecaneso 33, Genova, 16146, Italy
[d] ARPAL, Regional Agency for Environmental Protection Liguria, Via Bombrini 8, Genova, 16149, Italy

ARTICLE INFO

Keywords:
Energy markets
Energy management
Decision-making
Artificial intelligence
Numerical Weather Prediction models

ABSTRACT

An advanced artificial intelligence (AI)-assisted methodology for predicting the sign of energy imbalances within the day-ahead energy market is introduced in this study, with a focus on the integration of renewable energy sources. By leveraging deep learning techniques and Numerical Weather Prediction (NWP) models, a nuanced understanding of energy market dynamics over a comprehensive five-year period is provided by the research. The findings reveal the substantial predictive advantage of the AI model over traditional forecasting methods, with fold-averaged Area Under the Curve (AUC) values of about 0.7 achieved for the two distinct macro-zones N and S. Economically, the model indicates potential for significant market participant gains, with mean efficiencies reaching 16% and 11% for macro-zones N and S, respectively. The implications extend beyond the Italian market, suggesting transformative potentials for European energy markets at large. This work not only fills a critical gap in the literature but also sets a new benchmark for predictive accuracy and economic viability in energy market forecasting.

1. Introduction

A transformative shift is being undergone by the global energy landscape with the increasing integration of renewable energy sources (Erdiwansyah et al., 2021) and the drive towards sustainable and efficient energy production and consumption (Hirth and Ziegenhagen, 2015). In this context, a pivotal role is played by the day-ahead energy market, serving as a critical mechanism for balancing electricity supply and demand (Gan et al., 2020). Accurate forecasting of the energy imbalance, which represents the discrepancy between scheduled energy generation and actual generation, has become paramount for grid operators, policymakers alike, and energy traders (Dudek et al., 2023). The ability to anticipate whether the market will experience a surplus (positive imbalance) or deficit (negative imbalance) of energy in the day-ahead period is crucial for making informed decisions related to scheduling generation (Salkuti, 2019), adjusting demand (You et al., 2022), and managing grid stability (Liu et al., 2022). The accurate anticipation of energy imbalance sign is considered of primary importance for traders seeking to optimize their trading strategies and maximize profitability in this highly competitive environment. Indeed, the main characteristic of the electricity markets is that one has to propose bids in advance, and is then charged for any imbalance. Owing to the problem of predictability, the market value of renewable energy is reduced by the cost of regulation (Nielsen et al., 1999). Knowing the market imbalance sign in advance can also significantly aid network managers by enabling them to preemptively adjust supply and demand strategies, ensuring optimal resource allocation and maintaining network stability against potential fluctuations in energy demand.

The inherent volatility and uncertainty associated with renewable energy sources, such as solar and wind, present significant challenges to energy market participants in predicting the sign of the energy imbalance accurately. Valuable insights into the energy market's behavior have been provided by conventional forecasting methods, such as time series analysis and statistical modeling. In way of example, in Di Persio et al. (2017) different statistical methods, including the exponential smoothing model, the ARMA-ARIMA model, and the ARIMA-GARCH model, were exploited to guess the right sign of the energy load unbalance in the Italian electricity market. The strategies were fully data-driven with air temperature being the only weather feature entering the model. However, linear techniques struggle to capture the complex interplay between diverse variables, non-linear relationships, and dynamic patterns exhibited in the modern energy landscape. Consequently, there is a growing interest in exploring the potential of advanced AI-based techniques to enhance the accuracy and robustness

* Corresponding author at: DICCA, Department of Civil, Chemical and Environmental Engineering, Via Montallegro 1, Genova, 16145, Italy.
E-mail address: andrea.mazzino@unige.it (A. Mazzino).

https://doi.org/10.1016/j.egyr.2024.03.058
Received 9 January 2024; Received in revised form 19 March 2024; Accepted 30 March 2024
Available online 8 April 2024

Nomenclature

Acronym

AI	Artificial Intelligence
NWP	Numerical Weather Prediction
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic (curve)
DA	Day-Ahead (Market)
SMP	System Marginal Price
PUN	Prezzo Unico Nazionale (Unique National Price)
P_{DA}	Day-Ahead price for a specific time interval
BM	Balancing Market
TSO	Transmission System Operator
MO	Market Operator
ANNs	Artificial Neural Networks
ECMWF	European Centre for Medium-Range Weather Forecasts
HRES	High Resolution Model from ECMWF
PTU	Program Time Unit
UTC	Coordinated Universal Time

Symbol

TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
FPR	False Positive Rate
TPR	True Positive Rate
P	Precision
R	Recall
Acc	Accuracy
F_1	F1 Score
SS	Skill Score
$\wp$	Prevalence
ρ	Probability of assigning a positive class by a random model
Ap	Actual energy production
Pp	Programmed energy production
Ac	Actual energy consumption
Pc	Programmed energy consumption
P_{DA}	Price in the Day-Ahead market
p^+	Price in a positive imbalanced market
p^-	Price in a negative imbalanced market
n_+	Total number of positive events
n_-	Total number of negative events
$P(+)$	Precision for positive class
$P(-)$	Precision for negative class
$R(+)$	Recall for positive class
$R(-)$	Recall for negative class
$\overline{F}$	Mean value of F1-score between positive and negative class
g	Hourly revenue/loss per MWh per unit volume variation
G	Accumulated gain over a period
G_{max}	Maximum possible accumulated gain over a period
$msbil$	Unitary revenue/loss at a program time unit
$qsbil_{mat}$	Materialized macro-zonally-aggregated energy imbalance
$msbil_{mat}$	Materialized unitary revenue/loss at a program time unit
$qsbil$	Macro-zonally-aggregated energy imbalance
ΔR	Revenue associated to the volume variation
δV	Energy volume variation
Day-2	Day where $qsbil$ materialized
Day-1	Day where $qsbil$ of Day-2 is available
Day 0	Day prior to delivery in the energy market
Day 1	Day of delivery in the energy market
K	Kelvin
m s^{-1}	meter per second
J m^{-2}	Joule per meter square

of energy imbalance sign forecasting. Surprisingly, the current literature on this problem is sparse and limited, especially when compared to the extensive literature on forecasting day-ahead market prices, reviewed in Lago et al. (2021) and Nowotarski and Weron (2018). The additional benefits of the data-mining technique, when compared with classical algorithms, were showcased by the employment of classical time series models (ARIMA and exponential smoothing) together with a novel data-mining technique for system imbalance forecasting in Garcia and Kirschen (2006).

A novel approach for both deterministic and probabilistic forecasting of system imbalance has recently been released by Elia Group (2023), the Belgian Transmission System Operator (TSO). In the deterministic model, they employ a conventional linear regression framework with external predictors, while for the probabilistic model, they utilize binomial logistic regression. As a result, a minute-by-minute forecast is generated and continuously updated for the current and subsequent quarter-hour periods.

An investigation reported in Contreras (2023) proposed a regression based on the random forest algorithm to forecast the imbalance in the Spanish energy system. The study emphasized the advantages of using an ensemble technique for predicting system imbalances, even if the robustness of the results needs to be deepened because of the limited period of the considered data analysis.

An imbalance forecasting tool based on quantile random forests, thus producing probabilistic outputs, was developed for the case of the Norwegian TSO, Statnett, in Salem et al. (2019). Significant improvements were observed when comparing the new strategy against the usual approach of TSO.

In Plakas et al. (2023), a forecasting tool for the prediction of system imbalance in the Greek power system was presented. Three different algorithms have been compared for two different forecasting horizons. In both cases, the random forest algorithm proved to be the most accurate and computationally efficient compared to other approaches.

In Lisi and Edoli (2018) the dynamics of the historical time series of the imbalance sign in the Italian electricity market was analyzed and different predictive models were identified, estimated, and compared. The authors also showed that the information produced by a suitable model, together with a proper strategy, leads to a significant economic return. The authors also pointed out that their outcomes allow a more in-depth knowledge in a topic surely important even if not enough studied.

All analyzed studies aiming at predicting the imbalance sign in the day-ahead market do not take advantage of the huge potential of state-of-the-art Numerical Weather Prediction (NWP) models as providers of weather features possibly correlated with the imbalance sign to

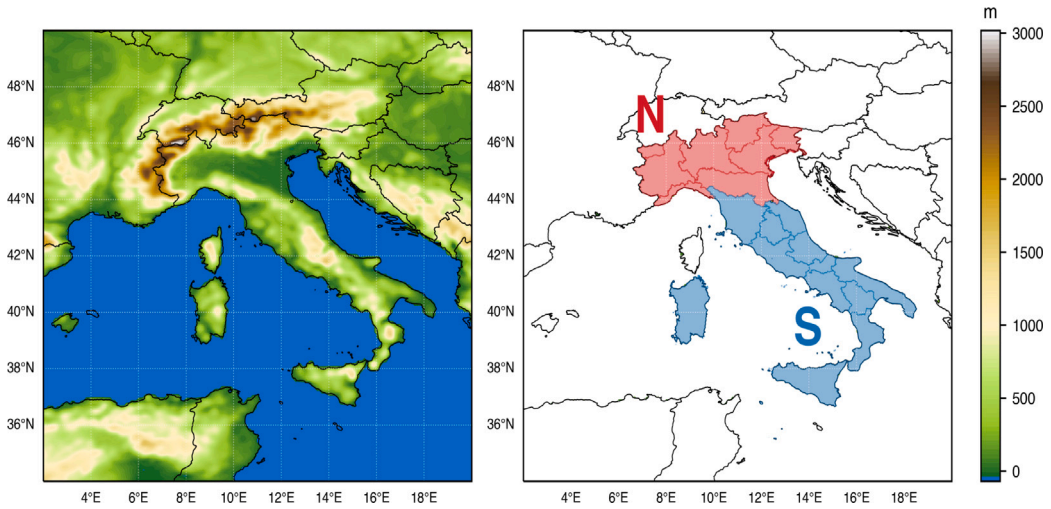

Fig. 1. Left panel: European domain for HRES output extraction. Right panel: HRES output masked to focus on Italian territory, excluding offshore areas without energy activities. Colors denote elevation. Labels N and S indicate two distinct Italian macro-zones for imbalance sign prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

be predicted. Also, none of them uses state-of-the-art deep learning methods to non-linearly map weather and market features into the imbalance sign. The objective of our paper is to contribute filling this gap by using NWP models in concert with a powerful deep-learning method to capture non-linear structures. Also, the resulting strategy will be tested in a time span of 5 years, an interval much longer than those usually considered to test the predictive skills of proposed models. Our testing phase will be carried out both to assess the accuracy of the classification-based approach and its degree of calibration, and thus its reliability, together with a skill assessment of indices having a direct economic meaning. A similar complete and rigorous assessment was not found in existing papers dealing with the prediction of energy imbalance sign.

In Section 2, our AI-assisted approach is outlined detailing the methodologies and models employed for energy imbalance forecasting. This includes an overview of the electricity market challenges, the imbalance settlement under the single-price scheme, a discussion on the global ECMWF-HRES numerical weather prediction model, and the preparation of feature/target datasets. This section also introduces the indices used for evaluating the AI model, crucial for understanding the metrics behind the model's evaluation.

Section 3 presents our model's results, comparing it with traditional forecasting benchmarks and discussing its economic implications.

Finally, in Section 4, the paper concludes with a summary of our findings, highlighting the impact and potential of our AI strategy in energy market forecasting.

2. Materials and methods

2.1. Generalities on the electricity market

Novel obstacles are encountered by the challenge to uphold equilibrium between electricity supply and demand in liberalized electricity markets. Primarily, the task of coordination is no longer rested solely within the domain of a single vertically integrated enterprise; instead, it becomes a responsibility shared by the Transmission System Operator (TSO), which is tasked with managing the transmission network to ensure its stability and reliability, and by the Market Operator (MO), who is made responsible for the organization of the electricity market. The competitive bidding process in the day-ahead (DA) market is managed by the MO, where bids and offers indicating the quantity of electricity that producers and consumers are willing to buy or sell and the price at which they are willing to do so are submitted. The electricity price, which can be set in different ways depending on specific market regulations, is determined by aggregating these bids and offers by the MO. In way of example, the System Marginal Price (SMP) serves as the reference price for all transactions that occur in many European Markets including, among others, Nordic, Iberian, French, and Baltic Markets. The price at which the last marginal unit of electricity needed to meet forecast demand is dispatched is represented by SMP. Based on the highest accepted bid among all bids submitted by electricity producers and consumers, this price is set.

In the day-ahead Italian electricity market, the energy is scheduled and traded at a price (referred to as PUN) calculated as the weighted average of all accepted bids and offers, separately for the Northern and Southern Italian macro-zone depicted in the right panel of Fig. 1, and for a specific time interval. Here, this reference price will be denoted by P_{DA}.

Any residual deviations are rectified under the obligation shouldered by the TSO, which fulfills this role by tapping into the flexible generation capacity accessible through the balancing market (BM). This latter, situated as the ultimate market in the sequence of temporal events within the wholesale power exchange, addresses discrepancies in supply and demand.

In order to diminish the residual requirements for balancing in electricity markets, it becomes essential that the organizational frameworks dictating the responsibilities for balancing offer efficient motivations for participants to meticulously plan their energy programs. This entails aligning their programs with their programmed production or consumption quantities. About this aspect, European regulations stipulate that any divergence from an energy program is resolved through compensation at a standardized rate. This rate reflects the actual instantaneous value of electricity (EU, 2017). Nevertheless, the actual value of electricity often diverges from the prices established in the day-ahead market. As a result, avenues for capitalizing on price disparities between consecutive market stages can thus emerge.

The relationship between market agents' balancing responsibility and the economic incentives provided by imbalance pricing rules is governed by two different pricing schemes: the *single-price scheme* and the *dual-price scheme*. The incentive properties of the two pricing schemes, concerning a market agent's responsibility to be balanced, are analyzed in detail in Clò and Fumagalli (2019). While the former pricing scheme rewards or penalizes market agents according to the impact of their individual program deviation on the system imbalance, the latter penalizes, at best does not reward, all individual imbalances.

More details on the single-price scheme, that is the one considered in the present paper, are provided in the next subsection.

2.2. Imbalances settlement under the single-price scheme

In the Italian market that is focused on, a certain number of consumption/production units ('units' in short) is found in each of the macro-zones alluded to above. For each day and load period, the difference between the scheduled and realized consumption/production is defined as the imbalance generated by each unit ('individual imbalance' in short).

The macro-zonally-aggregated energy imbalance (market imbalance in short) is defined as (Terna, 2023a):

$$qsbil = (Ap - Pp) - (Ac - Pc) \quad (1)$$

where, Ap/Pp are the actual/programmed energy production and Ac/Pc are the actual/programmed energy consumption. A positive sign of $qsbil$ (corresponding to a 'long' market) thus means a positive difference between the actual macro-zonal production (evaluated relatively to the programmed one) and the actual consumption (evaluated relatively to the programmed one).

Significant economic implications are held by the market imbalance sign. This is because the settlement price for the imbalance of a unit typically diverges from the one established in the DA market. Being imbalanced offers the ability to capitalize on or face disadvantages, which hinges upon the correlation between the market imbalance and the unit's individual imbalance. Since the latter is greatly influenced by the unit's choices, the key is to predict the market's positive or negative imbalance as accurately as possible. If imbalanced on the favorable side, a unit sees the magnitude of the profits arising from that imbalance depending on the price disparities between the DA and the balancing market. According to the single-price scheme (Terna, 2023b), when the market is positively imbalanced, the price, p^+, at which the energy is settled is the minimum between the price in the DA market, P_{DA}, and the average bid price in the balancing market. When the market is negatively imbalanced, the price, p^-, of the energy is the maximum between the price in the DA market and the average ask price in the balancing market. Since $p^+ < P_{DA} < p^-$, for a given sign of the market imbalance, unit profit/penalization only depends on its imbalance. For a unit positively imbalanced belonging to a positively imbalanced macro-zone, the extra injected energy (i.e. the energy amount exceeding the one programmed in the DA market) by the unit into the system through the balancing market is paid by the TSO at the price, p^+, less than P_{DA} (i.e. $msbil = p^+ - P_{DA} < 0$), the price it would have been paid in the DA market. Thus, the production unit is penalized. The opposite situation happens when the unit's imbalance is negative: the unit has in this case to pay for the energy not produced at the price p^+. This price, however, is lower than P_{DA} at which it was sold on the DA market, a mechanism leading to a unit profit.

A similar way of reasoning applies to positive/negative unit's imbalance in a negative market imbalance where now $msbil = p^- - P_{DA} > 0$. In summary, the single-price scheme leads to a negative payoff when the unit imbalance exhibits the same sign as the market imbalance: agents are penalized because they increase the size of the market imbalance and the related system costs. Conversely, the single pricing scheme rewards economic agents when their unit imbalance is opposite in sign to the market imbalance. Indeed, a unit imbalance opposing the market imbalance lowers the size of the market imbalance and the related system costs.

Even if the market imbalance is supposed to be almost unpredictable and producers should not create an imbalance on purpose (in the Italian market this is forbidden by the Italian law Terna, 2023c), the issue of market sign imbalance predictability is still poorly addressed and understood. The main aim of the present paper is to contribute to filling this gap.

Table 1
Meteorological features extracted from HRES' output in the forecast horizons 0–48 h.

Short name	Long name	Units
2d	2 meter dew point temperature	K
2t	2 meter temperature	K
10u	10 meter U wind component	m s^{-1}
10u	10 meter V wind component	m s^{-1}
tcc	Total cloud cover	Unity fraction
ssrd	Surface solar radiation downwards	J m^{-2}

2.3. The global ECMWF-HRES numerical weather prediction model

The HRES model is currently run by ECMWF with outputs collected on a latitude-longitude grid at a resolution of 0.1° × 0.1° (Malardel et al., 2016) four times per day (starting at the so-called synoptic hours 00, 06, 12, 24 UTC) up to 10 days ahead. On this grid, and for the sole 00 UTC run, forecast data relative to the entire Italian territory sketched in the left panel of Fig. 1, and for a time span of 5 years (from 2018 to 2022) have been collected every hour for the present study. When mimicking an operational data handling as in the present study, it must be considered that HRES outputs relative to the 00 UTC run are disseminated by ECMWF at 6 UTC (ECMWF, 2023). For the sake of notation, 'Day 0' will be called the starting day of the considered 00 UTC HRES' run. Day 0 also corresponds to the day prior to delivery in the energy market. Coherently, 'Day 1' will refer to the HRES 24–48 h forecast horizon and corresponds to the day of delivery in the energy market.

Among the whole set of variables provided by HRES model (relative to surface fields, fields integrated along the column, and fully three-dimensional fields), only a small subset of features have been selected to be successively handled in a way to be ingested by our AI algorithm. In Table 1, the selected features are listed. Observables that are expected to be related to the energy demand/consumption in the Italian territory are referred to. Due to the association of energy demand and consumption with ground areas, offshore regions have been excluded by masking the region in the left panel of Fig. 1 (as seen in the right panel of Fig. 1). In this masked area, the features of the HRES listed in Table 1 have been considered.

2.4. Preparation of feature/target datasets

To summarize the meteorological information on the area of the right panel of Fig. 1 and build a tabular data set of hourly meteorological variables, a histogram with a fixed range of minimum and maximum values, specific for each variable of Table 1, has been extracted, with the number of bins fixed to 20. In this way, the total number of occurrences, e.g., of a given value of the temperature field on the whole Italian territory for a given forecast lead time, is contained by each bin. Moreover, maximum, minimum, average, and standard deviation values have been also included in a way to account for spatial variability of the selected atmospheric features in the region under consideration.

Besides the meteorological features from the HRES model, other features have been considered to be ingested by our AI-based algorithm. Namely, the hourly-varying energy imbalance volume (qsbil, measured in MWh, being positive for a positively imbalanced macro-zone and negative elsewhere), the total load (tl, [GW]), and the hourly-varying renewable photovoltaic and wind energy generation volume (rg_photo and rg_wind, respectively), each one of them delivered by the electricity transmission system operator (Terna, 2023d). This information was acquired in the same time interval, from 2018 to 2022, during which HRES' outputs were also available. t must be emphasized that, when daily operational actions on the day-ahead energy market are mimicked (see Fig. 2 for a sketch), awareness must be had of the fact that the values of qsbil are made available at 5 CET of Day-1 while tl is made available in quasi-real-time.

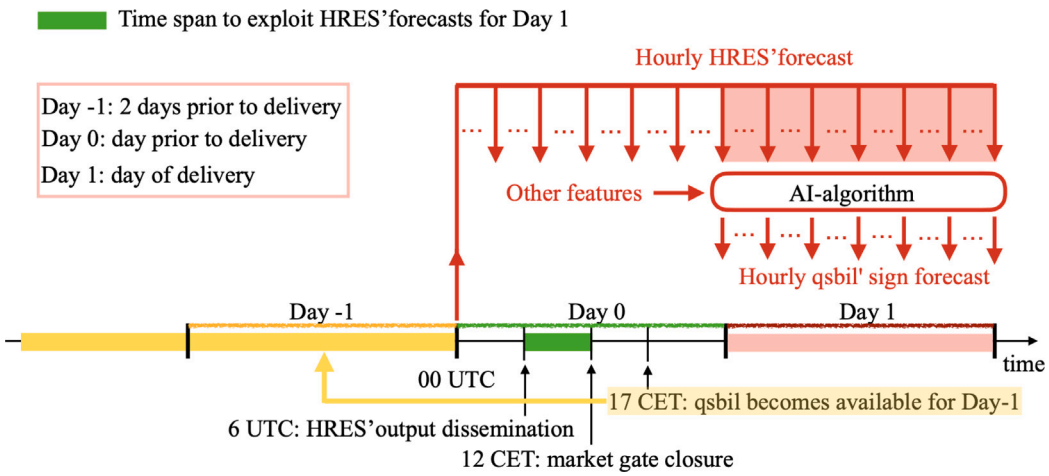

Fig. 2. Sketch highlighting the relevant information of the day-ahead market. In particular, the green region represents the time span in which operators can take advantage of our AI algorithm, to make predictions at Day 1. Also highlighted is the fact that information on the materialized qsbil sign arrives at Day 0 after the market gate closure and refers to Day-1. Only the information of the qsbil sign available at Day-1 (for Day-2) can be thus used during operations in the green region. This information enters our AI algorithm (as 'Other features') simply replicating in time for Day 1 the observed time behavior of qsbil at Day-2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This means that, to be used as features in the useful window of Day 0 to operate for Day 1, qsbil must be extrapolated from Day-2 to Day 1 (see Fig. 2), with their temporal behavior (on an hourly basis) occurring on Day-2 being preserved. Because of its quasi-real-time availability, tl, rg_photo and rg_wind can be extrapolated to Day+1 partly from Day 0 and partly from Day-1. A further extrapolation to Day 1 is based on the materialized value of the qsbil sign relative to the year preceding the target one. In more detail, if the target is the energy market in 2020, the most frequent sign in the year 2019 is computed and its value is used in our algorithm as a feature for the year 2020 ('Other features' in Fig. 2). This extrapolation is thus a sort of climatology-based feature.

The information related to qsbil also serves for testing the resulting predictive skills of the algorithm following the pipeline described in Section 2.5.

2.5. Tuning and testing the AI algorithm: the nested k-fold cross-validation

Nested cross-validation is an advanced technique used for model evaluation and selection (Stone, 1974; Cawley and Talbot, 2010), especially in machine learning tasks where data is limited, and model performance needs to be accurately estimated. It is an extension of the standard cross-validation method and provides more reliable estimates of a model's generalization performance. In standard cross-validation, the dataset is divided into multiple, k, folds, and the model is trained and evaluated multiple times. The process involves splitting the data into a training set and a test set, and this split is repeated several times. However, standard cross-validation has some limitations, particularly when it comes to hyperparameter tuning and model selection. It can lead to optimistic performance estimates, which might result in the selection of suboptimal hyper-parameters or models.

These limitations are addressed by nested cross-validation by adding an additional layer of cross-validation. This involves an inner loop of cross-validation dedicated to selecting the best hyperparameters, ensuring that the model's performance is optimized without directly testing on the outer validation set. Upon completion of all folds, the performance metrics from each iteration are aggregated by calculating their mean, to offer an overall performance estimate. This aggregated result gives a robust indication of the model's generalization capability across different subsets of the data. The variation in performance metrics across folds can also be analyzed to assess the model's stability. Finally, with the optimal hyperparameters determined and the model's generalization capability validated, a final model can be trained on the entire dataset for operative exploitation, such as predicting energy imbalance signs in the electricity market on daily basis. This comprehensive evaluation method, while computationally demanding, is invaluable for its ability to provide a reliable estimate of a model's performance, significantly reducing the risk of overfitting and ensuring the model's applicability to real-world data. In the context of energy markets, where predictions directly inform trading and operational decisions, the rigorous assessment provided by k-fold cross-validation is crucial for developing dependable forecasting tools.

The value of k in the present study is k = 10 and each considered fold is related to its start/end date as reported in Table A.3.

A sketch of the nested loops is shown in Fig. 3 for k = 3 for the sake of example.

2.6. The deep learning framework to predict the energy imbalance sign

Artificial Neural Networks (ANNs) are state-of-the-art machine learning algorithms, that have the ability to extract meaning patterns from huge and complex amounts of data and produce non-linear prediction models. ANNs were shown to outperform any machine learning algorithms, approaching human-level performance on various tasks, such as pattern and handwriting recognition. The architecture considered in this work is a kind of Deep Feed-forward Neural Network. One of the major efforts of dealing with ANNs is to find the best combination of hyperparameters for a specific problem, through *tuning*. In other words, tuning allows us to find the best architecture able to generalize well to new, unseen data. Tuning is typically an iterative process where a performance metric (in our case accuracy) is monitored. Tuning can rapidly become unmanageable due to the almost infinite number of possible hyperparameter combinations. So compromise choices were made. Table 2 reports the ranges of values of hyperparameters used during the tuning processes. In any case, the output layer is provided by a single neuron that exploits a sigmoid activation function in order to get a value between 0 and 1 which represents the probability associated with a positive or negative imbalance prediction. As mentioned before, in this work a nested cross-validation method was used to correctly deal with this task. Networks and tuning are implemented using the Python libraries Keras (Chollet et al., 2015), TensorFlow (Abadi et al., 2016) and Scikit-Learn (Pedregosa et al., 2011). For each macro-zones a nested cross-validation with tuning has been performed, leading to different best models.

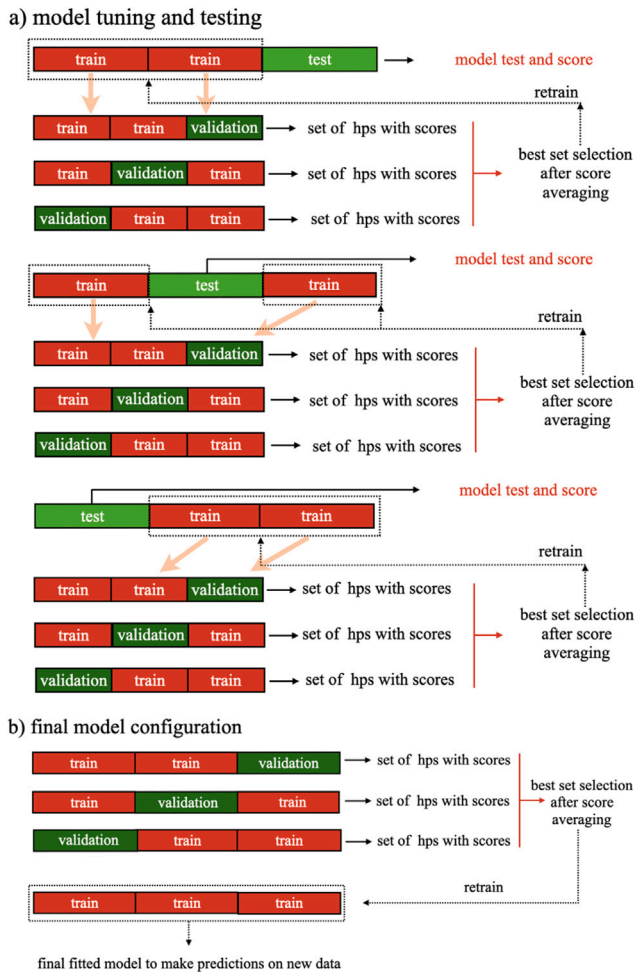

Fig. 3. Panel (a): The dataset is divided into k folds (k = 3 illustrated). Each outer loop iteration uses one fold for testing and the rest for training/validation, where models are tuned with various hyperparameters. After tuning, models are retested on the test set. Panel (b): Post tuning/testing, hyperparameters are finalized for model predictions on new data.

Table 2
Hyperparameters ranges tuned via nested cross-validation. Note that the feature scalers are preprocessing classes of the Scikit-Learn library, while kernel regularizers and kernel initializers are options available in the Tensorflow library.

Hyperparameters	Possible choices
Batch size	32, 64, 128
Epochs	25, 50, 75, 100
Learning rate	10^{-3}, 10^{-4}
Number of hidden layers	1, 2, 3
Units per hidden layers	16, 32, 64, 128
Kernel regularized	None, L2(10^{-2}), L2(10^{-4})
Drop-out rate	0.0, 0.1, 0.25
Kernel initializer	Glorot uniform, Random normal
Loss function	Binary Cross-Entropy, Brier Score
Feature scaler	StandardScaler, MinMaxScaler

2.7. Indices to assess model skills and its reliability

In this section, how the skills of our AI-assisted algorithm have been assessed in predicting the energy imbalance sign in the 0–48 h forecast horizon through the use of suitable statistical indices, which are detailed below, is described. Two types of indices are used. The first type is a classic of binary classifier evaluation. The second type is specific to the problem at hand and it follows from our definition of a skillful prediction as the one allowing agents participating in the market to gain from operations carried out in the day-ahead market once the imbalance sign is correctly predicted.

Starting from the first type, separately for each class, denoted in short class '+1' (positively imbalanced market) and class '−1' (negatively imbalanced market), the model skill evaluation is built in terms of the 'precision' [Eq. (2)], 'recall' [Eq. (3)], 'accuracy' [Eq. (4)], and 'F_1 Score' [Eq. (5)] statistical indices (Sofaer et al., 2019), denoted by P, R, Acc, and F_1, respectively. Namely,

$$P = \frac{TP}{TP + FP} = \frac{TP}{n_+} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{n_+ + n_-} \tag{4}$$

$$F_1 = \frac{2TP}{2TP + FP + FN}\,. \tag{5}$$

In the above equations, TP refers to the true positives relative to one of the two possible classes (i.e. an outcome where the model correctly predicts the event occurrence), FP refers to the false positives (i.e. an outcome where the model incorrectly predicts the event occurrence), FN refers to the false negatives (i.e. an outcome where the model incorrectly predicts the negative class corresponding to the absence of the event), and TN refers to the true negative (i.e. an outcome where the model correctly predicts the absence of an event), n_+ (n_-) refers to the total number of positive (negative) events. To determine TP, TN, FP, and FN a discrimination threshold acting on the probability outputted by the AI algorithm must be set in order to define event '+1' and event '−1'. The above indices thus depend on the class under consideration. For instance, $P(+)$ will denote the precision for the class '+1' while $P(-)$ will denote the precision for the class '−1'. A similar notation will be used for the other indices when necessary. By varying the threshold from 0 to 1 one obtains the so-called ROC curve (Fernández et al., 2018). This curve is built by reporting along the abscissa the False Positive Rate (FPR) defined as

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

corresponding to the probability that a '−1' observed event is classified as positive and along the ordinate the True Positive Rate (TPR) defined as

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

corresponding to the probability that a '+1' event is classified as positive. The goal is to have a classifier with a ROC curve well above the diagonal. This property can be quantified via the so-called 'area-under-the-curve' index (AUC), corresponding to the integral of the ROC curve. Of course, an optimal prediction would have $P = R = Acc = F_1 = AUC = 1$.

A useful benchmark random model can be constructed in terms of the so-called 'Prevalence', e.g. of the '+1' class, here denoted by $\wp$. This latter is defined as the ratio of the number of observed event occurrences n_+ to the total number of events $n_+ + n_-$. Namely, for the positive imbalance class, $\wp = n_+/(n_+ + n_-)$. Note that the prevalence associated with the negative class is $1 - \wp$.

Assuming that our classifier consists of randomly assigning the positive class with probability ρ and the negative class with probability $1-\rho$, the confusion matrix of a random classifier will have the following expected proportions

$$TP = \rho\wp \tag{8}$$

$$FP = \rho(1 - \wp) \tag{9}$$

$$FN = (1 - \rho)\wp \tag{10}$$

$$TN = (1 - \rho)(1 - \wp) \tag{11}$$

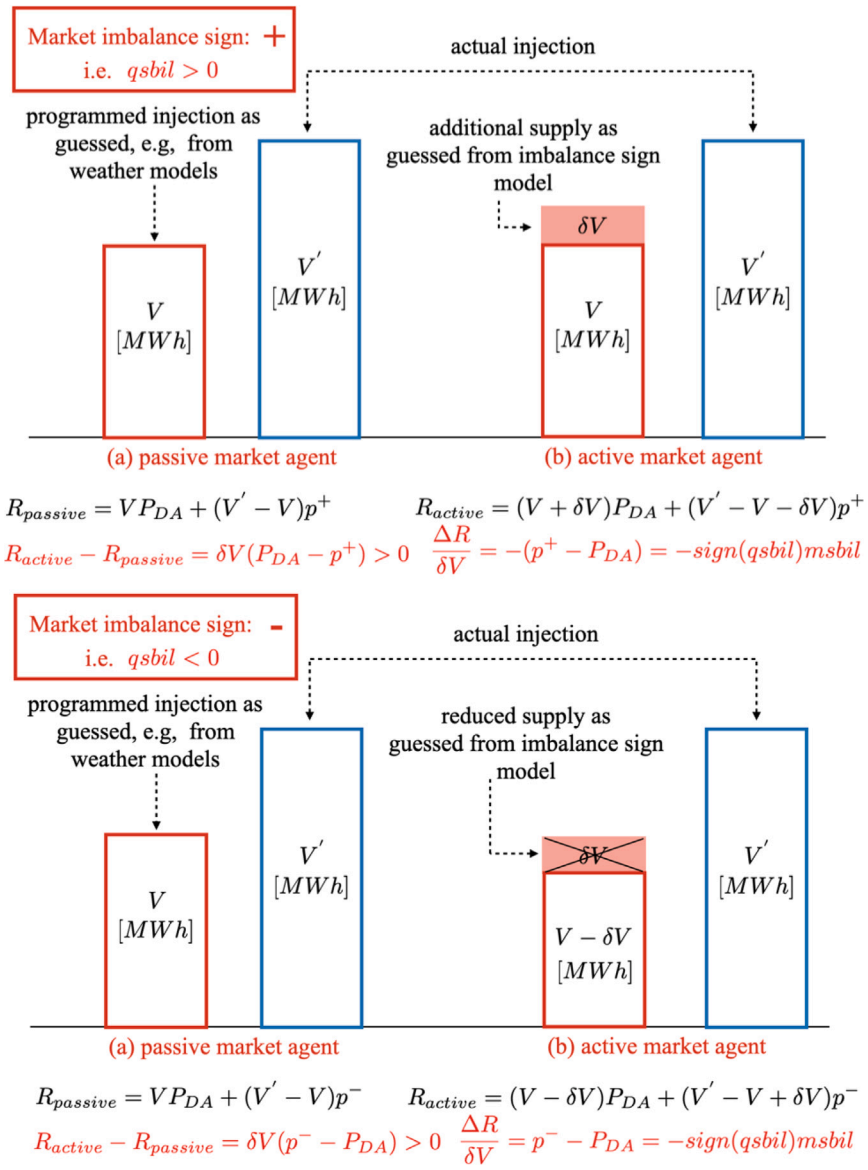

Fig. 4. Sketch of a simple strategy leading to rewards in the single-price Italian market scheme once the zonal sign is known two days in advance. The imbalance prices, p^+ and p^-, depend on the sign of the aggregated (macro-zonal) imbalance and are such that $p^+ < P_{DA} < p^-$, P_{DA} being the average bid price in the balancing market. Here $\Delta R = R_{active} - R_{passive}$ is the revenue associated to the volume variation $\pm\delta V$.

From these expressions, by direct substitutions, one obtains

$$TPR = FPR = \rho \tag{12}$$

This result holds independently of $\wp$. A random classifier thus always gives a ROC curve coinciding with the diagonal and the corresponding value of the AUC index is 0.5. It is also easy to verify that for the random model, one has:

$$P(+) = \wp \qquad P(-) = (1 - \wp) \qquad P(+) + P(-) = 1 \tag{13}$$

and

$$R(+) = \rho \qquad R(-) = (1 - \rho) \qquad R(+) + R(-) = 1 \tag{14}$$

Obtaining $P(+) + P(-) > 1$ and $R(+) + R(-) > 1$ with more complex prediction strategies is thus a direct measure of success compared to the simpler random-model based prediction.

Assuming now $\rho = 1/2$ (i.e. the mean prevalence of the two classes) it is easy to verify that the random model predicts:

$$Acc = \frac{1}{2} \qquad \overline{F} = \frac{F_1(+) + F_1(-)}{2} = \frac{\wp}{1 + 2\wp} + \frac{1 - \wp}{1 + 2(1 - \wp)} \tag{15}$$

All these random-model based predictions will be used as a benchmark in Section 3.

Let the discussion now pass to how the reliability of a forecast will be assessed. The concept of reliability is related to the one of calibration, meaning that a reliable forecast denotes the goodness of calibration (Silva Filho et al., 2023). Calibration pertains to the agreement between a forecaster's predictions and the actual observed relative frequency of a given phenomenon. Focusing on the problem at hand, a forecast is perfectly calibrated if it predicts a set of cases with, say, x% probability of being a positive imbalance, and the frequency of events having a positive imbalance contained in that set is equal to x%.

Let the focus now be on the second type of indices used to assess prediction skills. The starting point is the expression for the unitary (hourly) revenue/loss per MWh associated with the volume variation $\pm\delta V$ sketched in Fig. 4.

$$g \equiv \frac{\Delta R}{\delta V} = -sign(qsbil)\ msbil \qquad \text{'sign' being the sign function} \tag{16}$$

This quantity will be a revenue as long as the materialized sign of the market will be of the same sign of the variation of the program. For

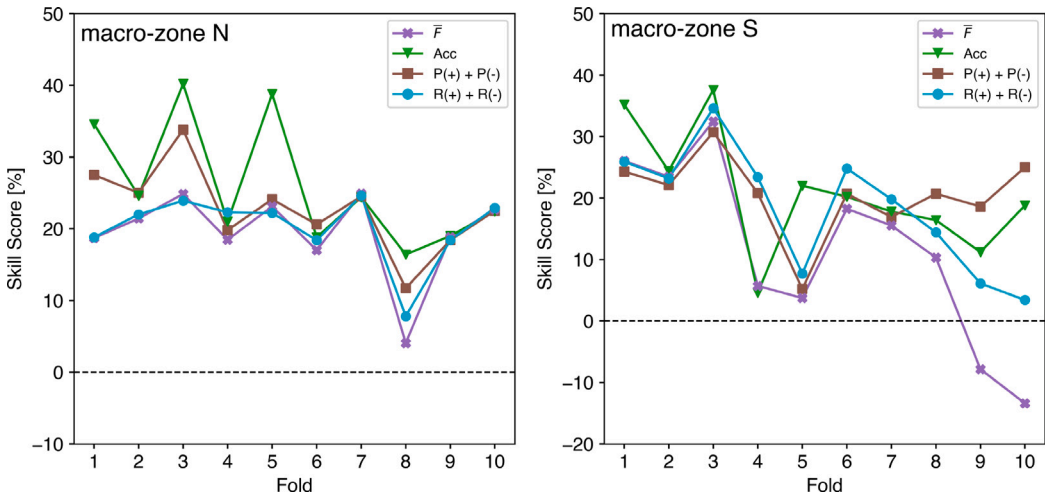

Fig. 5. Skill scores of $\overline{F}$, Acc, $P(+)+P(-)$, $R(+)+R(-)$ as a function of the fold. These indices have been defined in Section 2.7. The reference model, defined in Section 2.7, is the prevalence-based random model with $\rho = 1/2$. Averaging the fold-dependent skill scores over the 10 folds the following mean skill scores (SS) are obtained: $SS_{\overline{F}} = 0.19$, $SS_{Acc} = 0.26$, $SS_{P(+)+P(-)} = 0.23$, $SS_{R(+)+R(-)} = 0.20$ (macro-zone N); $SS_{\overline{F}} = 0.11$, $SS_{Acc} = 0.21$, $SS_{P(+)+P(-)} = 0.21$, $SS_{R(+)+R(-)} = 0.18$ (macro-zone S).

comparison, the maximum hourly revenue (known only a posteriori) can also be defined as

$$g_{max} = -sign(qsbil_{mat})\, msbil_{mat} \tag{17}$$

$qsbil_{mat}$ and $msbil_{mat}$ being the materialized $qsbil$ and $msbil$, respectively, and refer to the same time instant of Eq. (16).

Fig. 4 provides a sketch of the mechanism leading to a revenue. In this figure, a passive market agent has been defined as one who does not implement any strategy; on the other hand, an active agent has been defined as one who, having an accurate prediction of the macro-zone sign, relies on it and acts to find herself/himself in the BM market in a situation of imbalance opposite to that of the macro-zone.

Failing to predict the correct energy imbalance sign turns the revenue in loss as detailed in Section 2.2. A further score index can be thus constructed proceeding as follows. For each test fold of the 10-fold nested cross-validation (each fold corresponding to 6 consecutive months) the predicted $sign(qsbil)$ is multiplied by the collected $msbil$ to determine g from Eq. (16) and accumulating it on the six months of the test fold under consideration (with hourly frequency) to obtain the 6-month accumulated gain, G. By denoting g_i the revenue/loss at the ith (hourly step) program time unit (PTU), $qsbil_i$ the macro-zonally-aggregated energy imbalance (positive when the sum of the actual aggregated energy production exceeds the programmed value) at the ith PTU, and $msbil_i$ the unitary revenue/loss at ith PTU, Hm the number of hours in 6 months, one has:

$$G = \sum_{i=1}^{Hm} g_i = -\sum_{i=1}^{Hm} sign(qsbil_i)\, msbil_i \tag{18}$$

and, similarly, for the maximum revenue:

$$G_{max} = \sum_{i=1}^{Hm} g_{max\,i} \tag{19}$$

The resulting G can be compared with the maximum possible 6-month accumulated gain, G_{max}. The ratio G/G_{max} is a direct measure of the prediction skill, with $G/G_{max} = 1$ corresponding to a perfect prediction.

3. Results and discussion

Below, the results obtained using the AI-enhanced strategy for predicting the sign of the zone are reported. The probability thresholds for determining the various indices to evaluate the classifier's skills were determined by optimizing the product of the F_1 Scores for the '+1' and '−1' classes during the tuning phase within the 10-fold cross-validation pipeline. In Fig. 5, the skill score of $\overline{F}$, Acc, $P(+)+P(-)$, $R(+)+R(-)$ has been reported. These quantities have been defined in Section 2.7. The left panel refers to the macro-zone N, while the right panel refers to the macro-zone S. Both macro-zones have been described in Fig. 1. The skill score is defined as (Wilks, 2011; Casciaro et al., 2022)

$$SS = \frac{a - a_{ref}}{a_{opt} - a_{ref}} \tag{20}$$

where a is an error index to assess the forecast quality, a_{ref} is the error-index associated with a reference forecast, and a_{opt} refers to the index value corresponding to optimality. In the present case, a represents all indices described above, obtained from our AI-based algorithm, and a_{ref} refers to the same indices but obtained from the random model described in Section 2.7. Except for one case concerning the macro-zone S, related only to the $\overline{F}$ index and for only folds 9 and 10, all curves in Fig. 5 are always well above zero, indicating a clear advantage of our AI-enhanced strategy compared to using a random model.

The good performance of our strategy is also confirmed by the analysis of the ROC curve shown in Fig. 6 (macro-zone N in the right panel, macro-zone S in the left panel) and the corresponding AUC index values, shown as insets. Despite fluctuations from one fold to another (orange curves), the ROC curves are always well above the diagonal, and the AUC values are always appreciably greater than 0.5, the value corresponding to the prediction of the random model.

To assess calibration, for each test set of the 10-fold cross-validation, the predicted probabilities for events of class '+1' from our AI-based model have been gathered, and the corresponding actual outcomes (whether the event happened or not) have been collected. The range of predicted probabilities has been successively partitioned into 10 subsets (bins along the x-axis), each representing a disjoint interval of probabilities between 0 and 1. The length of each subset is determined to have the same number of events in each interval. Each predicted probability has been assigned to a bin based on its value. For each bin, the observed frequency of each event has been calculated. This has been done by dividing the number of times the event actually happened by the total number of forecasts in that bin. The resulting reliability diagram has been obtained by plotting the bins of predicted probabilities on the x-axis and the corresponding observed frequencies on the y-axis. The resulting plot is reported in Fig. 7. The upper panel refers to the macro-zone N, while the lower panel to the macro-zone S. The bisector corresponds to perfectly calibrated forecasts, while the shaded region represents the skill area (Weisheimer and Palmer, 2014). Points above the diagonal indicate under-forecasting (the event happens more

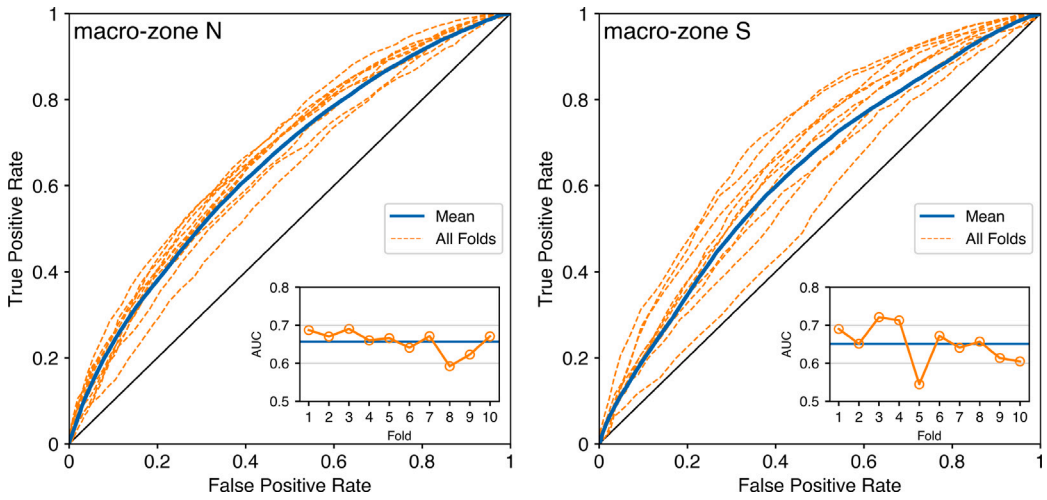

Fig. 6. The ROC curves for the two considered macro-zones are presented. Orange lines represent each single fold, while the blue line indicates the average ROC curve over the 10 folds. The two insets display the AUC as a function of the fold. The resulting fold-averaged AUCs are 0.66 for macro-zone N and 0.65 for macro-zone S. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

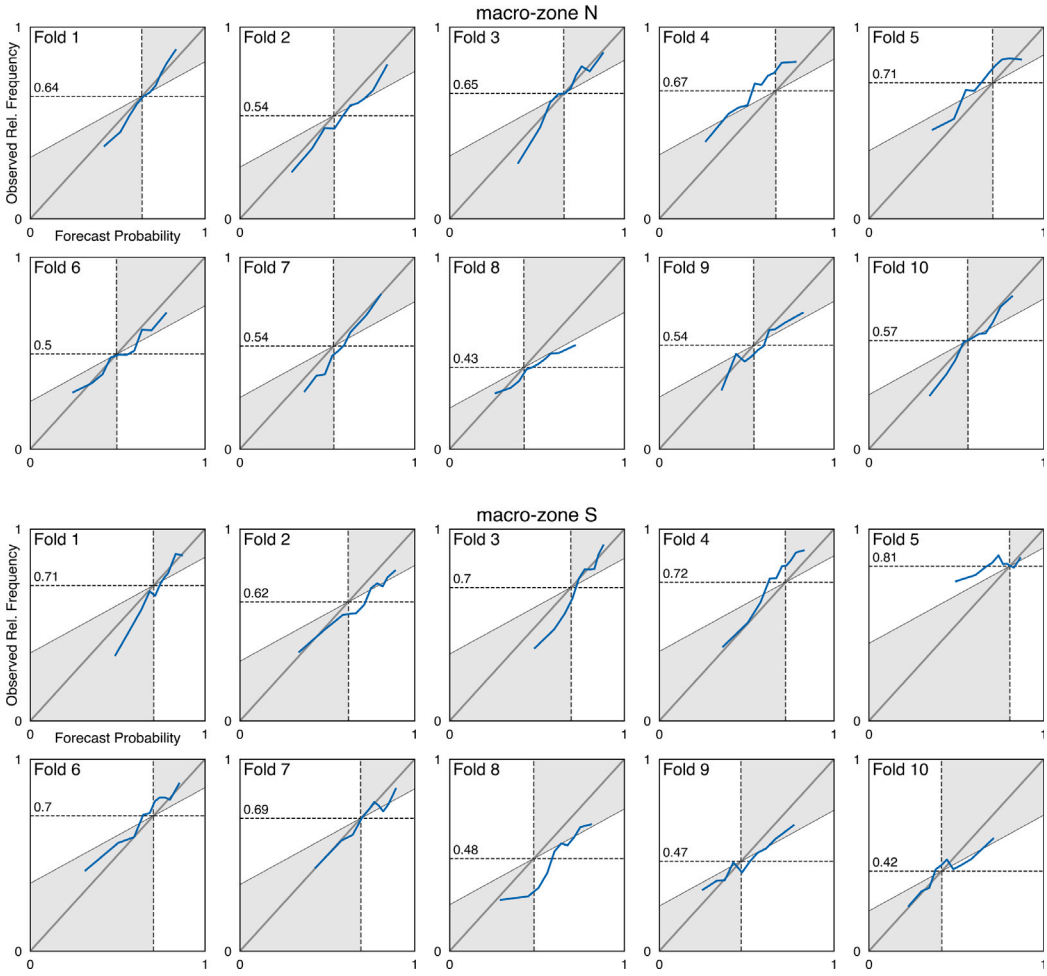

Fig. 7. Reliability diagrams are presented for the two considered macro-zones across all folds. Shaded regions indicate the skill regions, and the diagonal represents a perfectly calibrated prediction. The horizontal and vertical dashed lines represent the prevalence, $\wp(+)$, of each fold.

frequently than predicted), while points below the diagonal indicate over-forecasting (the event happens less frequently than predicted).

As can be easily seen, our AI algorithm produces forecasts within the skill area for the majority of the folds, a hallmark of reliability, even though the prediction in some folds (folds 4 and 5 for macro-zone N and S, and fold 8 solely for the macro-zone S) appears less reliable than others. Having had the quality of our AI strategy as a classifier assessed, it can now be analyzed whether the strategy can yield a gain for a market agent who knows the prediction of the macro-zone sign. Fig. 8 shows for the macro-zone N (left panel) and the macro-zone S (right panel) the behavior of the unitary (per unit volume variation) hourly revenue/loss per MWh defined in Section 2.7,

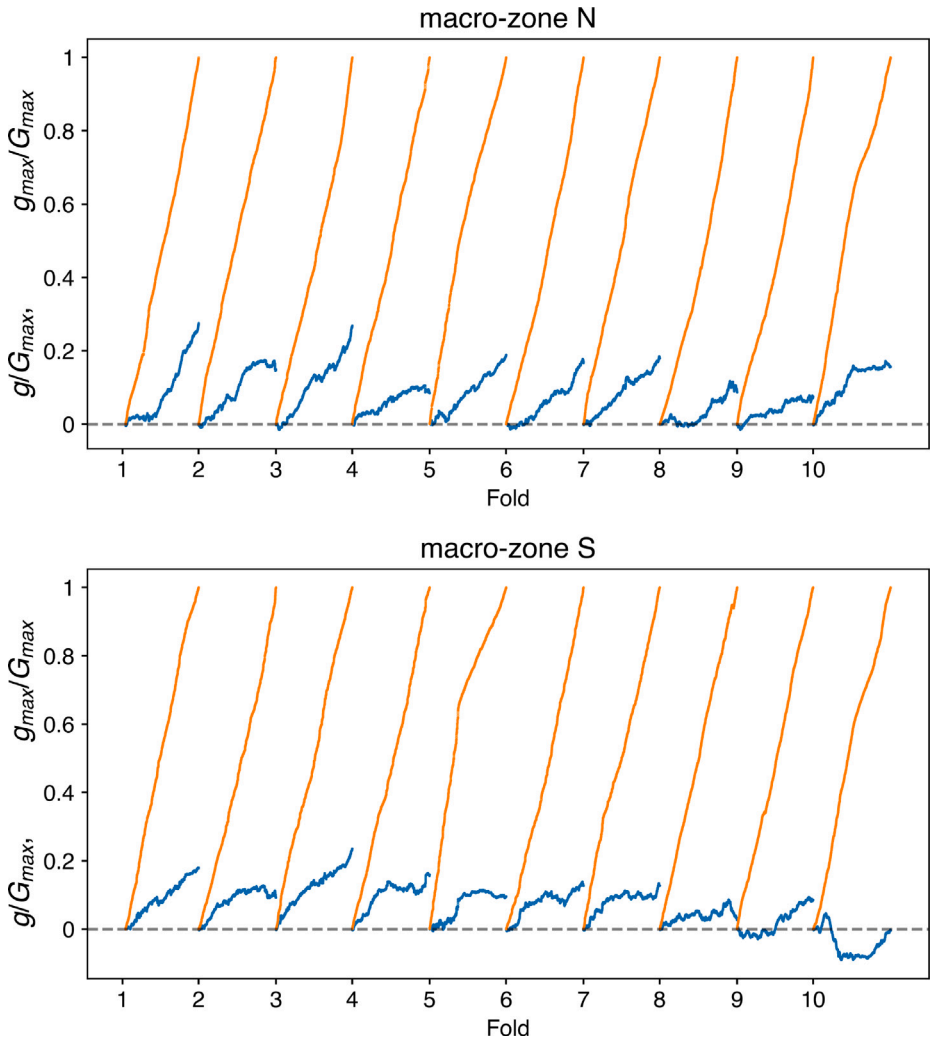

Fig. 8. The hourly revenue/loss per MWh and per unit volume variation, g/G_{max}, defined in Section 2.7, is shown as a function of the program time unit (PTU) varying on an hourly basis, spanning in each fold a time interval of 6 months. The efficiency index G/G_{max} of each fold is the integral over time of each blue curve. For comparison, g_{max}/G_{max} is also shown (orange lines). Averaging over all folds, the resulting mean efficiencies are 0.16 for macro-zone N and 0.11 for macro-zone S. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

g/G_{max}, as a function of the program time unit (PTU) varying on hourly basis, spanning in each fold a time interval of 6 months. The integral over time of each blue curve gives G/G_{max} for each fold. The orange curves represent the maximum possible gain, g/G_{max} (normalized to 1 when integrated over 6 months), which would materialize if the agent knew the macro-zone sign for certain. The integral of the blue curves represents a sort of efficiency of our strategy evaluated over a time span of 6 months, indicating what fraction of the maximum gain it is capable of producing. Except for fold 10 of the macro-zone S, which has a gain integrated over the semester equal to zero, all the remaining folds show a significantly positive net gain. The average over the 10 folds of G/G_{max} in the macro-zone N is about 0.16, while in the macro-zone S, it is about 0.11. So, these are not negligible gains of just a few percent. The fact that the performance obtained in the macro-zone N is superior to that of the macro-zone S can be attributed to the greater penetration of renewable energy sources and, therefore less predictable sources, present in the latter macro-zone.

4. Conclusions and future work

An AI-assisted approach for enhancing the prediction of energy imbalance signs in the day-ahead energy market has been effectively established by the presented research. By combining deep learning with numerical weather prediction models, it has significantly outperformed traditional forecasting methodologies, substantiating the model's superior predictive accuracy and economic utility. Specifically, the study achieved noteworthy fold-averaged AUC values and demonstrated the model's capacity to facilitate substantial economic advantages for market participants, with mean efficiencies of 16% and 11% for macro-zones N and S, respectively.

Several avenues for future research are opened by this study beyond its immediate findings. Firstly, exploring the integration of more granular weather and market data could further refine prediction accuracies. Secondly, extending the model's application to other geographic regions and energy markets could validate its versatility and adaptability. Thirdly, investigating the impact of emerging renewable energy technologies and changing consumption patterns on market dynamics could provide deeper insights into future energy market behaviors.

Moreover, the development of real-time predictive models that can dynamically adjust to new data presents an exciting frontier for operational efficiency and market responsiveness. Lastly, examining the regulatory and economic implications of widespread adoption of AI in energy markets could offer valuable perspectives on future policy development and market strategies.

In essence, this study not only contributes to the ongoing evolution of energy market forecasting but also highlights the transformative

potential of AI in addressing the challenges and opportunities presented by the integration of renewable energy sources. Future research in these areas will be crucial in realizing the full potential of AI in enhancing the sustainability and efficiency of global energy systems.

Despite the demonstrated effectiveness of the AI-assisted forecasting strategy, it is important to acknowledge certain limitations inherent to the present approach. Firstly, the model's performance, while superior to traditional methods, is inherently dependent on the quality and granularity of the input data, including weather predictions and market dynamics. Inaccuracies in the NWP models or unexpected market behaviors can adversely affect the predictive accuracy. Secondly, the approach assumes a relatively stable energy market structure and may not fully account for rapid changes in energy policies, market regulations, or significant technological advancements in renewable energy sources. Additionally, the model's generalizability across different geographic regions and energy markets remains to be thoroughly tested, as varying climatic conditions and market mechanisms could impact its applicability. Recognizing these limitations is crucial for guiding future research efforts aimed at enhancing the robustness, adaptability, and operational feasibility of AI-assisted forecasting strategies in the energy market.

CRediT authorship contribution statement

Daniele Carnevale: Methodology, Investigation, Formal analysis, Data curation, Software, Validation, Writing – review & editing. **Mattia Cavaiola:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Data curation, Investigation, Methodology, Project administration. **Andrea Mazzino:** Conceptualization, Methodology, Project administration, Supervision, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Useful discussions with Francesco Ferrari, Massimo Rivarolo, and Alessandro Sorce are warmly acknowledged.

Ethical approval statement

This research did not involve human participants, human data, or human tissue, and therefore did not require ethical approval.

Appendix. Definition of the 10 folds

See Table A.3.

Table A.3
The 10 folds considered are related to their start and end dates. Each fold spans six consecutive months.

Fold number	Initial date	Final date
1	2018-01-01	2018-06-30
2	2018-07-01	2018-12-31
3	2019-01-01	2019-06-30
4	2019-07-01	2019-12-31
5	2020-01-01	2020-06-30
6	2020-07-01	2020-12-31
7	2021-01-01	2021-06-30
8	2021-07-01	2021-12-31
9	2022-01-01	2022-06-30
10	2022-07-01	2022-12-31

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. OSDI '16, USENIX Association, USA, pp. 265–283.

Casciaro, G., Cavaiola, M., Mazzino, A., 2022. Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring. Atmos. Environ. 287, 119259. http://dx.doi.org/10.1016/j.atmosenv.2022.119259.

Cawley, G.C., Talbot, N.L.C., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11 (70), 2079–2107.

Chollet, F., et al., 2015. Keras, https://keras.io.

Clò, S., Fumagalli, E., 2019. The effect of price regulation on energy imbalances: A difference in differences design. Energy Econ. 81, 754–764.

Contreras, C., 2023. System imbalance forecasting and short-term bidding strategy to minimize imbalance costs of transacting in the spanish electricity market. https://repositorio.comillas.edu/xmlui/handle/11531/16621.

Di Persio, L., Cecchin, A., Cordoni, F., 2017. Novel approaches to the energy load unbalance forecasting in the Italian electricity market. J. Math. Ind. 7 (1), 5, URL https://doi.org/10.1186/s13362-017-0035-y.

Dudek, G., Piotrowski, P., Baczyński, D., 2023. Intelligent forecasting and optimization in electrical power systems: Advances in models and applications. Energies 16 (7).

2023. ECMWF Dissemination Schedule. URL https://confluence.ecmwf.int/display/DAC/Dissemination+schedule.

Elia Group, 2023. Documentation of the methodology used for the system imbalance forecast publications. URL https://www.elia.be/en/grid-data/balancing/system-imbalance-forecasts.

Erdiwansyah, Mahidin, Husin, H., Nasaruddin, Zaki, M., Muhibbuddin, 2021. A critical review of the integration of renewable energy sources with various technologies. Prot. Control Mod. Power Syst. 6 (1), 3.

2017. EU commission regulation (EU) 2017/2195 of 23 november 2017 establishing a guideline on electricity balancing. URL eur-lex.europa.eu.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from imbalanced data sets. In: Cambridge International Law Journal. URL https://api.semanticscholar.org/CorpusID:53046396.

Gan, L., Jiang, P., Lev, B., Zhou, X., 2020. Balancing of supply and demand of renewable energy power system: A review and bibliometric analysis. Sustain. Futures 2, 100013.

Garcia, M., Kirschen, D., 2006. Forecasting system imbalance volumes in competitive electricity markets. IEEE Trans. Power Syst. 21 (1), 240–248.

Hirth, L., Ziegenhagen, I., 2015. Balancing power and variable renewables: Three links. Renew. Sustain. Energy Rev. 50, 1035–1051.

Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Appl. Energy 293, 116983.

Lisi, F., Edoli, E., 2018. Analyzing and forecasting zonal imbalance signs in the Italian electricity market. Energy J. 39, 1–19.

Liu, T., Song, Y., Zhu, L., Hill, D.J., 2022. Stability and control of power grids. Annu. Rev. Control Robot. Auton. Syst. 5 (1), 689–716.

Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzynski, G., Hamrud, M., Smolarkiewicz, P., 2016. A new grid for the IFS. ECMWF Newsl. 146 (23–28), 321.

Nielsen, L.H., Morthorst, P.E., Skytte, K., et al., 1999. Wind Power and a Liberalised North European Electricity Exchange. Denmark.

Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: A review of probabilistic forecasting. Renew. Sustain. Energy Rev. 81, 1548–1568.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Plakas, K., Andriopoulos, N., Birbas, A., Moraitis, I., Papalexopoulos, A., 2023. A forecasting model for the prediction of system imbalance in the greek power system. Eng. Proc. 39 (1), URL https://www.mdpi.com/2673-4591/39/1/18.

Salem, T.S., Kathuria, K., Ramampiaro, H., Langseth, H., 2019. Forecasting intra-hour imbalances in electric power systems. Proc. AAAI Conf. Artif. Intell. 33 (01), 9595–9600.

Salkuti, S.R., 2019. Day-ahead thermal and renewable power generation scheduling considering uncertainty. Renew. Energy 131, 956–965.

Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M., Flach, P., 2023. Classifier calibration: a survey on how to assess and improve predicted class probabilities. Mach. Learn. 1–50.

Sofaer, H.R., Hoeting, J.A., Jarnevich, C.S., 2019. The area under the precision-recall curve as a performance metric for rare binary events. Methods Ecol. Evol. 10 (4), 565–577.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (2), 111–133. http://dx.doi.org/10.1111/j.2517-6161.1974.tb00994.x.

Terna, 2023a. Chapter 7, page 14 of the Italian Grid Code. URL https://www.terna.it/en-gb/sistemaelettrico/codi-cedirete.aspx.

Terna, 2023b. Chapter 7, page 18 of the Italian Grid Code. URL https://www.terna.it/en-gb/sistemaelettrico/codi-cedirete.aspx.

Terna, 2023c. Chapter 7, par. 3 of the Italian Grid Code. URL https://www.terna.it/en-gb/sistemaelettrico/codi-cedirete.aspx.

Terna, 2023d. Download Center. URL https://www.terna.it/it/sistema-elettrico/transparency-report/download-center.

Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. J. R. Soc. Interface 11 (96), 20131162.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, vol. 100, Academic Press.

You, M., Wang, Q., Sun, H., Castro, I., Jiang, J., 2022. Digital twins based day-ahead integrated energy system scheduling under load and renewable energy uncertainties. Appl. Energy 305, 117899.