Forecasting requirements

1. PV Generation Forecasting

1.1 Inputs

The following inputs are recommended. Most studies show that the most important factor is global solar irradiance, but other weather factors, such as cloudiness, temperature, wind speed and humidity are important out of the weather variables

Solar position is also an important factor that we can use,

Category	Features	Notes
NWP Forecasts (primary)	Forecast global irradiance (GHI), DNI, DHI; 2 m temperature; cloud fraction; wind speed; relative humidity	KNMI HARMONIE-AROME model; 2 km resolution; updated hourly
Solar Position	Solar zenith angle, azimuth angle	Computed via pvlib; adding these reduced RMSE by 13.1% in [1]
Clear-Sky Index	Actual GHI / clear-sky GHI	Normalises seasonal/diurnal trends; stabilises model variance
Historical PV Lags	t−96 (same time yesterday), t−672 (same time last week)	Valid lags at forecast issuance time
Calendar Features	Hour of day, day of year, month, is_weekend	Cyclical encoding avoids discontinuities at midnight/year-end
Panel System Config	Tilt angle (35°), azimuth (180° south-facing), system size (~100 kWp), panel temperature coefficient	Fixed inputs; used to calibrate pvlib simulations

1.2 Models

Tree-based ensemble methods — Random Forest (RF), XGBoost, and LightGBM — are computationally efficient baselines. A 2022 benchmark comparing 24 machine learning models across 16 PV plants at 15-minute resolution confirmed that these methods are competitive following proper hyperparameter tuning [1]. However, they forecast each timestep independently, which ignores temporal dependencies across the 96-step output sequence.

For more model-comparisons, should read [1.1] and [1.2]

For the primary model, a sequence-to-sequence architecture is recommended. Hybrid CNN-LSTM-Transformer models have demonstrated state-of-the-art accuracy for day-ahead multi-step PV forecasting: Kim et al. [2] showed that a combined LSTM-Transformer model outperformed a standalone LSTM by 48.3% in mean absolute error. A 2024 systematic review in Heliyon confirmed that Transformer-based architectures achieve the best accuracy on PV forecasting tasks [3].

The recommended model stack is therefore:

- LightGBM or Random Forests (my preferred choice) – fast, interpretable baseline; also used in the production ensemble
- LSTM or GRU – captures intra-day temporal dependencies
- Temporal Fusion Transformer (TFT) – best multi-horizon architecture with native probabilistic output
- Ensemble (simple average of LightGBM + TFT) – consistently reduces error versus any single model

1.3 Recommended metrics

Metric	Role	Notes
nRMSE (% of capacity)	Primary	Normalised by installed capacity; enables comparison across systems
MAE (kW)	Primary	Robust to outliers; most intuitive for operations

Metric	Role	Notes
Forecast Skill Score	Essential	SS = 1 − RMSE_model / RMSE_persistence. Values > 0 indicate improvement over smart persistence
MBE	Secondary	Detects systematic over- or under-prediction bias
CRPS	Probabilistic	Continuous Ranked Probability Score; evaluates the full predictive distribution

2. Hospital electricity demand forecasting

2.1 Inputs

Category	Features	Availability
Historical Load	Hourly/15 min lags: t-1, t-24, t-168 (same time last week), rolling 24h and 168h statistics (should discuss this further)	God help us (maybe we can go and measure ourselves IDK)
Weather	Outside temperature !!!!, relative humidity, wind speed, global solar radiation etc. (not sure how many of these are how important, gotta research it more)	KNMI
Calendar	Hour of day, day of week, weekend, dutch public holidays, school holidays period etc.	This should be easy
Operational info	Shift schedules, elective surgery schedule binary (on/off) (I have no idea what this is, recommended by AI)	Got to beg for this i guess

2.2 Models

For 24 hour ahead multi-step forecasting, realistically Temporal Fusion Transformers (TFT) is king. I mean it's almost provably best for time-series forecasting. TFT processes static metadata, known future inputs, and past-observed variables through separate attention pathways, and natively supports quantile regression for probabilistic output [6] (important if we do RL cause of uncertainty). Multiple building energy studies confirm it outperforms standalone LSTM architectures at short-term multi-step load forecasting. A directly relevant 2025 study used LSTM forecasting combined with PPO reinforcement learning for a 1,500-bed hospital microgrid and reported strong performance [7].

Ensamble methods such as Boosted Trees and Random forests are not bad as baselines, but there's no native sequence modeling. Normal linear time-series forecasting methods do generally perform worse, but something like SARIMAX wouldn't be bad to use as it's computationally way less expensive than TFT's and LSTMN's

Model	Strengths	Weaknesses	Vedicts (by AI, didn't want to give them myself)
Temporal Fusion Transformer(TFT)	Handles mixed input types; native quantile output; interpretable attention	Higher training complexity (too much compute needed)	Primary model — recommended
LightGBM / XGBoost	Fast, interpretable, competitive on tabular data	No native temporal sequence modelling	Ensemble component and baseline
SARIMAX	Statistically rigorous; handles seasonal patterns	Cannot capture non-linear relationships	Baseline benchmark
CNN-BiLSTM	Captures local and bidirectional temporal patterns	Needs more tuning than TFT	Optional comparison model

Metrics

Evaluation must use rolling walk-forward cross-validation — not random train-test splits — to preserve temporal ordering and prevent data leakage. Test data should span at least four full seasons. Always report forecast skill relative to a naive same-weekday-same-hour persistence baseline. (AI recommendations)

Metric	Role	Notes
MAPE (%)	Primary	Target < 5%; minimum acceptable < 10% for 24h-ahead hourly forecasts
CV-RMSE (%)	Primary	Target < 20% (hourly); ASHRAE Guideline 14 threshold for model calibration
R^2	Secondary	Target > 0.95; provides intuitive variance-explained measure
NMBE (%)	Bias check	Normalised Mean Bias Error; target ±5%; flags systematic over/under-prediction
Peak MAE (kW)	Operational	Accuracy specifically during morning and evening demand peaks

3. Day-ahead Electricity Price Forecasting

3.1 Inputs

Feature selection for Dutch electricity price forecasting is well-established in the literature. The strongest predictor is the price itself — autoregressive lags dominate — while renewable generation forecasts provide the key exogenous signal

Feature Category	Specific Variables	Source
Autoregressive price lags	All 24 hours from days d−1, d−2, d−3, d−7 (96 lags total); constitutes the core of the LEAR model	ENTSO-E Transparency Platform

Feature Category	Specific Variables	Source
Renewable generation forecast	Wind generation forecast (onshore + offshore separately); solar PV forecast. In 2024, wind contributed ~27% and solar ~18% of Dutch generation. Correlation with price: r = −0.84	ENTSO-E Transparency Platform
Load forecast	Day-ahead total electricity demand forecast for the Netherlands bidding zone	ENTSO-E Transparency Platform
Fuel price	TTF natural gas spot price (Dutch Title Transfer Facility): the marginal cost signal for gas-fired generation (~34% of Dutch generation)	ICE / Trading Economics / Investing.com
Carbon price	EU ETS CO_2 allowance price; affects marginal cost of fossil generators	ICE Futures Europe
Cross-border flows	Net scheduled import/export with Germany (NL–DE), Belgium (NL–BE), UK, Norway. Netherlands has >9 GW interconnector capacity	ENTSO-E Transparency Platform
Calendar features	Hour of day, day of week, is_weekend, Dutch public holiday flags	Obvious

3.2 Models

The electricity price forecasting (EPF) literature has converged on a clear hierarchy of approaches, anchored by the benchmark work of Lago, Weron, and collaborators [8][9]. The LEAR model (LASSO-Estimated AutoRegressive) has become the gold-standard baseline: it uses 96 autoregressive price lags (all 24 hours from days d−1, d−2, d−3, and d−7) plus exogenous variable regressors, with LASSO regularisation for automatic feature selection. Many complex deep learning models fail to outperform LEAR despite orders of magnitude greater computational cost, making it an essential benchmark.

Deep neural networks with 24 (or 96) joint outputs and NBEATSx — NBEATS extended with exogenous variables by Olivares et al. [10] — improve upon LEAR by approximately 5–20%. The Temporal Fusion Transformer shows strong results for probabilistic price forecasting [11]. Crucially, the most consistent finding across the EPF literature is that ensemble averaging of 2–4 models reliably outperforms any individual model [8].

Model	Strength	Weaknesses	AI verdict
LEAR	Fast, interpretable, strong baseline; well-studied for EU markets	Linear; cannot capture non-linear price spike dynamics	Essential baseline — always implement first
NBEATSx	~20% improvement over NBEATS; strong exogenous handling	Higher implementation complexity	Strong alternative to DNN
DNN (24-output)	Captures non-linearities; modest improvement over LEAR	Requires hyperparameter tuning; less interpretable	Recommended primary DL model
Temporal Fusion Transformer (TFT)	Native quantile forecasting; handles mixed inputs	Heaviest architecture	Recommended for probabilistic forecasts
Ensemble (LEAR + DNN + TFT)	Consistently best overall accuracy	Requires all three models to be maintained	Recommended production deployment

AI remark: *Models must be recalibrated daily using a rolling 730-day training window to capture evolving market dynamics. All test sets must span at least one full year: the 2022–2023 energy crisis fundamentally altered Dutch price distributions, and models trained exclusively on pre-crisis data fail to generalise.*

3.3 Metrics

Metric	Role	Notes
MAE (€/MWh)	Primary	The standard metric in EPF literature; robust to outliers and sign changes. Typical range: €2–8/MWh in normal market conditions
RMSE (€/MWh)	Secondary	Penalises large errors (spikes) more heavily than MAE; useful for tail-risk assessment
rMAE	Relative benchmark	Relative MAE versus the naive same-day-last-week persistence benchmark; enables cross-market comparison
Diebold-Mariano test	Statistical significance	Use the multivariate DM test (Ziel & Weron, 2018) for a single significance statistic over the full daily output vector

Challenges

Several structural features of the Dutch market make electricity price forecasting harder than demand or PV forecasting:

- Negative prices are now routine: the rapid growth of offshore wind (Hollandse Kust projects) and rooftop solar creates regular oversupply conditions, especially on weekend afternoons. Standard models struggle with the bimodal price distribution this produces.
- Price spikes during scarcity events: cold, calm periods (Dunkelflaute) with low wind and solar output can drive prices to the market cap of €3,000/MWh. Models must handle heavy-tailed distributions.
- Market regime changes: the 2022–2023 energy crisis caused a structural break in price distributions. All models must include sufficient post-crisis training data.

- 15-minute MTU transition (October 2025): the shift to 15-minute market time units quadruples the forecast output dimension from 24 to 96 values per day, requiring models that scale efficiently to higher-dimensional outputs.

References:

[1] Mayer, M. J. & Gróf, G. (2022). *Extensive comparison of physical models for photovoltaic power forecasting.* Applied Energy, 324, 119526. DOI: 10.1016/j.apenergy.2022.119526

[1.1] Husein M, Gago EJ, Hasan B, Pegalajar MC. Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies. Heliyon. 2024 Jun 27;10(13):e33419. doi: 10.1016/j.heliyon.2024.e33419. PMID: 39050417; PMCID: PMC11268202.

[1.2] R. Ahmed, V. Sreeram, Y. Mishra, M.D. Arif. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, Renewable and Sustainable Energy Reviews, Volume 124,2020,109792,ISSN 1364-0321, https://doi.org/10.1016/j.rser.2020.109792 Forgot to add these two before :(

[2] Kim, T. et al. (2024). *Day-ahead multi-step PV generation forecasting using LSTM-Transformer hybrid.* Renewable and Sustainable Energy Reviews, 199, 114479. DOI: 10.1016/j.rser.2024.114479

[3] Al-Rakhami, M. et al. (2024). *Deep learning strategies for photovoltaic power forecasting: a systematic review.* Heliyon, 10(9), e30940. DOI: 10.1016/j.heliyon.2024.e30940

[4] González González, A. et al. (2018). *Evaluation of energy consumption in German hospitals: benchmarking in the public sector.* Energies, 11(9), 2279. DOI: 10.3390/en11092279

[5] Parker, S. A. et al. (2019). *Energy consumption analysis and characterisation o healthcare facilities in the United States.* Energies, 12(19), 3775. DOI: 10.3390/en12193775

[6] Lim, B. et al. (2021). *Temporal Fusion Transformers for interpretable multi-horizon time series forecasting.* International Journal of Forecasting, 37(4), 1748–1764. DOI: 10.1016/j.ijforecast.2021.03.012

[7] Ahmad, M. et al. (2025). *AI-driven energy management for hospital microgrids integrating LSTM forecasting and reinforcement learning.* Scientific Reports, 15(1).

DOI: 10.1038/s41598-025-28907-5

[8] Lago, J. et al. (2021). *Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark.* Applied Energy, 293, 116983. DOI: 10.1016/j.apenergy.2021.116983

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

[10] Olivares, K. G. et al. (2023). *Neural basis expansion analysis with exogenous variables: forecasting electricity prices with NBEATSx.* International Journal of Forecasting, 39(2), 884–900. DOI: 10.1016/j.ijforecast.2022.01.002

[11] Jiang, W. et al. (2024). *Probabilistic electricity price forecasting using LASSO-TFT hybrid models.* Journal of Forecasting, 43(5). DOI: 10.1002/for.3097