

ORIGINAL RESEARCH

Open access

A Foundation-Model Architecture for Multi-Horizon Smart-Grid Forecasting Using Smart-Meter Sequences, Weather Context, and Grid-Topology Tokens

Amira Khalil^{1*}, Sara Fathy¹, Mahmoud Adel², Hany Nabil¹, Rania Mostafa²

Abstract

Smart-grid forecasting requires models that can reason across heterogeneous temporal, environmental, and structural information. Existing forecasting architectures often treat load sequences, weather context, and grid topology as separate modeling inputs rather than as jointly learnable forecasting evidence. This article proposes a foundation-model architecture for multi-horizon smart-grid forecasting. The architecture is designed to integrate smart-meter sequences, weather context, and grid-topology tokens within a unified transformer framework. The proposed architecture could encode smart-meter data through patch-based or convolutional sequence tokenization. Weather information would be represented through learned meteorological embeddings, while feeder topology could be incorporated through graph-derived or learnable topology tokens. The design specifies how heterogeneous tokens may be fused through cross-modal attention and routed toward horizon-specific forecasting heads. It also outlines an evaluation strategy based on pre-training across diverse feeders and fine-tuning on held-out forecasting tasks. The architecture offers a conceptual blueprint for generalizable smart-grid forecasting systems. Further validation would be required to determine whether unified tokenization improves transferability across feeders, substations, and regional grid configurations.

Keywords Foundation model, Multi-horizon forecasting, Smart-meter sequences, Weather context, Grid-topology tokens, Transformer architecture

*Correspondence:

Amira Khalil

amira.khalil@gmail.com

¹ Department of Smart Energy Systems, Faculty of Engineering, Alexandria University, Alexandria, Egypt

² Department of Power Grid Optimization, Faculty of Engineering, Cairo University, Cairo, Egypt

Introduction

Smart-grid forecasting increasingly requires models that can represent short-term fluctuations, daily load cycles, local weather sensitivity, and feeder-level operating differences within a single forecasting framework. Deep learning surveys have shown that modern time-series forecasting has moved beyond fixed statistical models toward representation-learning architectures that can capture nonlinear temporal patterns across many domains [1]. In electricity systems, this shift is especially important because smart-meter data provide dense temporal traces but do not by themselves fully explain demand variation caused by weather, topology, or operational constraints [2]. A foundation-model architecture for smart-grid forecasting should therefore be conceptualized as a representation system rather than as a single task-specific predictor.

Most existing load-forecasting models are still organized around specific datasets, fixed horizons, or single-modality input pipelines. Transformer-based load-forecasting studies demonstrate that attention mechanisms can capture long-range dependencies in electrical demand sequences [3], while short-term load forecasting with transformer models suggests that sequence attention can be adapted to power-system time series [4]. However, these architectures are often developed as forecasting models for particular settings rather than as reusable pre-trained backbones for many feeders, substations, and horizons. The proposed article addresses this gap by treating load sequences, weather context, and topology information as distinct but jointly tokenized forecasting modalities.

The emergence of pre-trained time-series models indicates that forecasting systems could increasingly benefit from large-scale representation learning before task-specific fine-tuning. A survey on time-series pre-trained models frames pre-training as a way to learn transferable temporal representations across tasks [5], and foundation-model thinking in transformer-load forecasting has begun to connect pre-training with overload detection and grid forecasting use cases [6]. For smart grids, this direction would enable a model to learn generic load-weather-topology interactions before being adapted to feeder-specific or horizon-specific forecasting tasks. Such a design requires careful architectural choices because continuous load values, meteorological variables, and graph-structured connectivity cannot be tokenized in the same way.

This article contributes a conceptual foundation-model architecture for multi-horizon smart-grid forecasting using unified tokenization of smart-meter sequences, weather context, and grid-topology tokens. The architecture is informed by deep time-series forecasting literature that emphasizes representation learning [7], experimental comparisons of deep forecasting architectures [8], and long-sequence forecasting surveys that identify scalability and horizon length as central design issues [9]. Its contribution is not an experimental benchmark but an architecture-driven blueprint that specifies how smart-meter patches, learned weather embeddings, and topology-aware tokens could be integrated into a cross-modal transformer. The model would enable pre-training on diverse grid data and fine-tuning for very-short-term, short-term, and day-ahead forecasting tasks.

Multi-Horizon Forecasting Problem

Time horizons and granularity

Multi-horizon smart-grid forecasting can be conceptualized as the simultaneous estimation of electricity demand across several temporal resolutions, such as 15-minute, hourly, intra-day, and day-ahead horizons. Prior work on short-term residential load forecasting using recurrent neural networks shows that load dynamics vary across temporal scales and customer aggregation levels [10]. Deep residual networks for short-term load forecasting further indicate that extracting multi-scale temporal features is important when demand patterns contain both local fluctuations and broader daily cycles [11]. A foundation-model design should therefore avoid treating horizon length as a post-processing detail and should instead incorporate horizon awareness into the output structure.

Probabilistic and point forecasts

Smart-grid forecasting systems often require both point forecasts for operational scheduling and probabilistic forecasts for uncertainty-aware decision support. Probabilistic forecasting with temporal fusion transformer models has shown how deep architectures can generate uncertainty-aware load forecasts for medium-term hourly demand [12]. Interpretable probabilistic forecasting of residential net load also suggests that uncertainty information can be useful when load is affected by distributed energy resources and household-level variability [13]. In the proposed architecture, multi-horizon output heads could be designed to support both deterministic estimates and probabilistic intervals, although the article does not present empirical results.

Operational use cases

The operational value of multi-horizon forecasting lies in its ability to support feeder monitoring, transformer loading assessment, demand response planning, substation scheduling, and resilience-oriented grid management. Deep learning approaches to household load forecasting have shown that residential demand can be modeled from high-resolution consumption patterns, but their operational usefulness depends on how forecasts are connected to planning decisions [14]. Multi-source load forecasting using gated recurrent units demonstrates that combining demand data with contextual variables can improve the representation of operational demand drivers [15]. The foundation-model architecture proposed here would allow forecasting tasks to share a common backbone while using horizon-specific heads for different grid-management use cases.

Data Modalities

Smart-meter load sequences

Smart-meter load sequences form the primary temporal evidence stream for the proposed foundation model because they provide repeated observations of consumption behavior at household, feeder, or substation levels. Sequence-to-sequence load prediction research shows that encoder-decoder models can represent temporal dependencies in electrical demand [16], while deep learning frameworks for electricity demand forecasting demonstrate how historical load traces can be transformed into predictive features [17]. In this architecture, each smart-meter sequence could be converted into load tokens through patching, down-sampling, or convolutional encoding before being combined with weather and topology tokens. The key design requirement is that load tokens preserve temporal granularity while remaining compact enough for transformer pre-training.

Numerical weather predictions

Numerical weather predictions would provide exogenous context for demand variation, especially where temperature, humidity, solar irradiance, wind speed, and calendar-weather interactions influence electricity use. Feature selection and genetic optimization in LSTM-based load forecasting highlight the importance of selecting relevant exogenous variables rather than simply appending all available covariates [18]. Hybrid LSTM and convolutional neural network models also show that temporal weather-load interactions may be represented through learned nonlinear transformations [19]. In the proposed foundation model, meteorological values could be embedded into the same token space as load patches so that cross-modal attention can learn weather-sensitive demand representations.

Grid topology information

Grid topology information represents the structural context in which smart-meter and feeder load patterns occur, including connectivity, feeder hierarchy, transformer grouping, node distance, and local aggregation relationships. Short-term load forecasting studies based on graph convolutional recurrent networks suggest that spatial or network relationships can add predictive information beyond independent time-series modeling [20]. A topology-aware forecasting foundation model should therefore represent feeder structure through graph-derived embeddings, learnable topology tokens, or relative structural encodings that can interact with load and weather tokens.

Figure 1 presents the proposed multimodal tokenization pathway through which smart-meter sequences, weather context, and grid-topology information are converted into a shared transformer-ready representation.

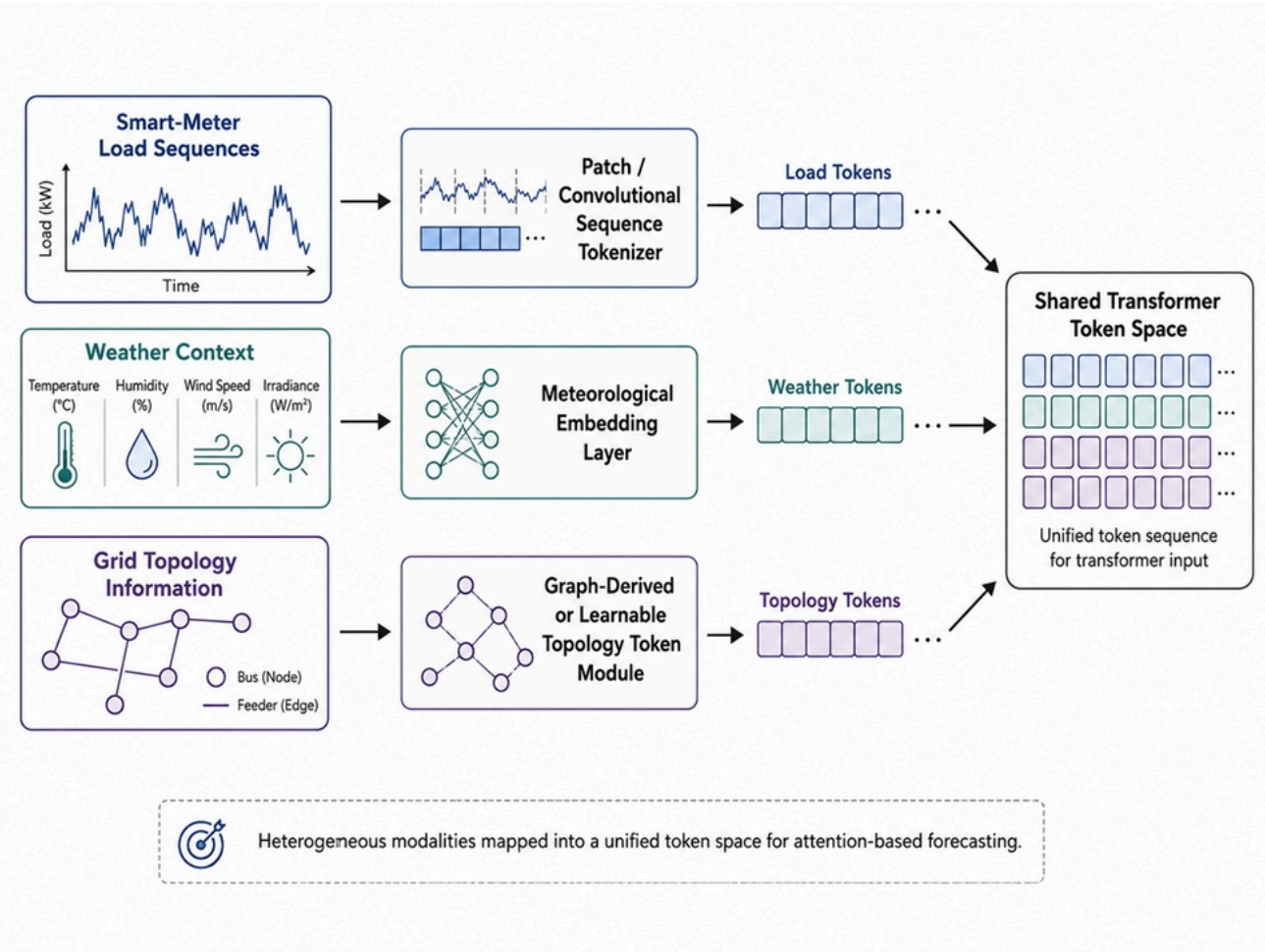


Figure 1. Multimodal Tokenization Framework for Smart-Grid Foundation-Model Forecasting

Table 1 summarizes the three data modalities and clarifies their expected role in the proposed architecture.

Table 1. Summary of Data Modalities and Their Characteristics

Data modality	Typical source	Core characteristics	Proposed token form	Role in the foundation model	Main architectural challenge
Smart-meter load sequences	Residential, commercial, feeder, or substation metering systems	High-frequency temporal demand observations with seasonality, volatility, missingness, and customer heterogeneity	Load patches, convolutional latent tokens, or down-sampled temporal tokens	Provides the main temporal demand signal for multi-horizon forecasting	Preserving short-term variation while keeping token length computationally manageable

Weather context	Numerical weather prediction feeds, local sensors, meteorological services, or reanalysis products	Temperature, humidity, wind, irradiance, precipitation, and forecast-time uncertainty	Learned continuous weather embeddings aligned to forecast horizons	Explains exogenous demand variation and weather-sensitive load shifts	Aligning spatially coarse weather data with feeder-level load sequences
Grid topology information	Distribution-system models, feeder maps, transformer records, meter-to-node mappings, or connectivity graphs	Graph-structured relationships among meters, transformers, feeders, and substations	Graph neural network embeddings, learnable topology tokens, or relative topological position encodings	Provides structural context for aggregation, congestion exposure, and feeder-specific transferability	Encoding changing or incomplete topology without requiring a fixed graph for all regions

Smart-Meter Sequence Encoding

Sequence patching and tokenization

Smart-meter sequence encoding could begin by dividing continuous load histories into fixed-length temporal patches that act as compact forecasting tokens. Patch-based transformer logic is consistent with long-sequence forecasting research, where sequence compression and attention scalability are central architectural concerns [9]. Transformer-based electrical load forecasting has already shown that load trajectories can be represented through attention-based sequence modeling [3]. In the proposed foundation model, patch length would need to balance local resolution against pre-training efficiency because very small patches preserve fine variation but increase transformer token length.

Positional and temporal embeddings

Load tokens require positional and temporal embeddings so that the transformer can distinguish chronological order, time-of-day effects, weekday-weekend differences, seasonality, and forecast issue time. Short-term load forecasting with time-augmented transformer models shows that explicitly enriching transformer inputs with temporal information can improve the model's ability to represent demand cycles [21]. Demand-side load forecasting with deep neural networks also indicates that calendar and time structure are important for short-term electricity prediction [22]. In this architecture, absolute temporal embeddings could encode clock time, while relative embeddings could represent lag structure and horizon distance.

Handling missing and irregular data

Smart-meter sequences are often affected by missing intervals, meter outages, delayed readings, irregular sampling, and data-quality inconsistencies that complicate direct tokenization. Deep learning frameworks for electricity demand forecasting emphasize the need for preprocessing pipelines that make raw load data usable for predictive models [17]. Transfer learning with seasonal and trend adjustment also suggests that load histories may require normalization or decomposition before representations can transfer across buildings or regions [23]. The proposed architecture should therefore include missingness masks, learned imputation indicators, and data-quality tokens so that the transformer can distinguish true low consumption from absent or unreliable readings.

Output representation

The smart-meter encoder should not produce a single static vector but rather a sequence of latent demand tokens that can support multiple forecasting horizons. Sequence-to-sequence load forecasting provides a useful precedent because it frames forecasting as a transformation from historical sequences to future trajectories [16]. Probabilistic temporal fusion transformer work further shows that output representations can be structured to support uncertainty-aware multi-step prediction [12]. **Table 2** compares alternative smart-meter tokenization strategies and identifies trade-offs that would shape the proposed foundation-model design.

Table 2. Comparison of Smart-Meter Sequence Tokenization Strategies

Tokenization strategy	How it would operate	Advantages	Limitations	Best suited forecasting role
Fixed-length patching	Divides the historical load sequence into non-overlapping or overlapping temporal windows and projects each window into a token	Reduces sequence length, preserves local temporal structure, and aligns well with transformer pre-training	Patch boundaries may obscure abrupt events, outages, or rapid ramps	General-purpose pre-training and multi-horizon forecasting across many feeders
Down-sampling	Aggregates raw readings into coarser intervals before token creation	Improves computational efficiency and reduces noise	May lose high-frequency demand spikes and short operational disturbances	Day-ahead and longer short-term forecasting where fine temporal volatility is less central
Convolutional encoding	Uses one-dimensional convolutional filters to transform raw load sequences into latent temporal tokens	Captures local motifs, ramps, periodicity, and short-term demand signatures	Encoder design may be less interpretable and sensitive to kernel size choices	Very-short-term and short-term forecasting where local sequence shape is important
Mask-aware tokenization	Adds missingness indicators, quality flags, or learned mask embeddings to load tokens	Allows the model to separate missing readings from real consumption behavior	Requires reliable data-quality labels or consistent missingness representation	Pre-training across heterogeneous smart-meter datasets with variable completeness
Hybrid patch-convolution encoding	Applies convolution within each patch before projecting the patch into transformer token space	Combines local feature extraction with scalable token length	More complex to tune and validate conceptually	Foundation-model pre-training where both local motifs and long-range attention are needed

Weather Context Encoding

Weather variable selection

Weather context encoding should begin with a principled selection of meteorological variables that plausibly influence electricity demand over the target horizons. LSTM-based load forecasting with feature selection and genetic optimization demonstrates that weather and contextual features should be selected according to their predictive relevance rather than included indiscriminately [18]. Multi-source load forecasting using gated recurrent units similarly indicates that exogenous variables can enrich load representations when they are integrated into the temporal modeling pipeline [15]. In the

proposed foundation model, selected weather variables could include temperature, humidity, solar irradiance, wind speed, precipitation, and forecast uncertainty descriptors.

Spatial interpolation and alignment

Weather forecasts are rarely measured at the same spatial resolution as smart-meter or feeder data, so the architecture should include a conceptual alignment layer between meteorological grids and electrical network locations. Deep learning models for solar power forecasting show that weather-driven energy prediction depends on representing meteorological context in ways that are relevant to the local asset being forecast [24]. Cross-building energy forecasting with transfer learning further suggests that regional and building-specific differences can affect how weather-load relationships generalize across sites [23]. A smart-grid foundation model would therefore require spatial interpolation, feeder assignment, or learned location embeddings to connect weather tokens with the relevant load and topology tokens.

Embedding into token space

Weather variables should be embedded into the same latent token space as smart-meter and topology information so that the transformer can learn cross-modal interactions. Deep learning surveys on time-series forecasting emphasize that representation learning enables models to encode heterogeneous temporal signals in shared latent spaces [1]. Time-series pre-trained model research also suggests that reusable embeddings can support transfer across related forecasting tasks [5]. In the proposed architecture, numerical meteorological variables could be passed through continuous feature encoders, temporal projection layers, or variable-specific embedding modules before entering cross-modal attention.

Temporal alignment with load sequences

Temporal alignment is necessary because weather forecasts, load histories, and target forecasting horizons may have different issue times, sampling frequencies, and lead-time structures. Long-sequence forecasting surveys identify horizon length and temporal dependency modeling as central challenges for transformer-based forecasting systems [9]. Temporal fusion transformer approaches to probabilistic load forecasting also show that future-known covariates can be organized differently from historical observed variables [12]. In this architecture, weather tokens should be indexed by forecast lead time and aligned with load-history tokens through horizon-aware embeddings that allow the model to learn how meteorological context affects each forecast horizon.

Grid-Topology Tokens

Representing feeder topology

Feeder topology can be represented as a graph in which nodes correspond to meters, transformers, buses, feeder segments, or substations, and edges represent electrical or operational connectivity. Graph-based short-term load forecasting shows that neighborhood structure can be useful when demand at one location is related to demand patterns at electrically or spatially connected locations [20]. Temporal graph convolution approaches further suggest that topology-aware forecasting can combine network structure with time-varying demand signals [25]. In the proposed architecture, topology should not be treated as static metadata but as a tokenizable structural modality that conditions how load and weather information are interpreted.

Graph neural network embeddings

Graph neural network embeddings could be pre-computed from feeder connectivity and then inserted into the transformer as topology tokens associated with meters, transformers, or feeder regions. Multi-resolution graph neural network forecasting demonstrates that electricity load can be represented across multiple spatial or structural resolutions [26]. Spatial-temporal dynamic graph and multi-scale transformer work also indicates that graph structure and transformer

attention can be conceptually combined for short-term load forecasting [27]. A foundation model could use graph embeddings to provide structural priors while allowing the transformer to learn long-range temporal and cross-modal dependencies.

Learnable topology tokens

Learnable topology tokens offer an alternative when detailed feeder graphs are incomplete, inconsistent, or unavailable across regions. In this design, the model could assign each feeder, substation, transformer group, or network zone a learnable embedding that captures recurring topology-related forecasting behavior during pre-training. Time-series pre-trained model literature supports the broader idea that transferable latent representations can be learned before downstream adaptation [5]. For smart grids, learnable topology tokens would enable approximate structural conditioning while still requiring validation against explicit graph-derived embeddings.

Integration into the transformer

Topology tokens should be integrated into the transformer so that structural information can influence both sequence interpretation and horizon-specific prediction. Transformer-based electrical load forecasting shows how attention can capture dependencies across temporal positions [3], while graph-based load forecasting highlights the value of explicitly modeling spatial or network relationships [20]. The proposed architecture could concatenate load patches, weather embeddings, and topology tokens, or it could use separate modality streams followed by cross-modal attention. Table 3 summarizes topology-token generation methods and their possible integration with the transformer backbone.

Table 3. Grid-Topology Token Generation Methods

Topology-token method	Conceptual mechanism	Transformer integration pathway	Strengths	Limitations
Graph neural network embeddings	A graph neural network processes feeder connectivity and generates node, feeder, or substation embeddings	Inserted as structural tokens linked to corresponding load-token groups	Preserves explicit feeder relationships and supports topology-aware forecasting	Requires reliable graph data and may be difficult to standardize across utilities
Multi-resolution graph embeddings	Topology is encoded at meter, transformer, feeder, and substation levels	Added as hierarchical tokens that condition attention at different aggregation scales	Captures structural relationships across several grid levels	Increases architectural complexity and token-management burden
Learnable topology tokens	The model learns feeder or zone embeddings during pre-training without requiring full graph specification	Concatenated with load and weather tokens as structural identifiers	Useful when topology is incomplete or partially unavailable	Less physically grounded than graph-derived representations
Relative topological encoding	Electrical distance, feeder depth, or transformer grouping is encoded as relative position information	Used as attention bias or relative positional embedding	Allows topology to influence attention without adding many tokens	Depends on well-defined structural distance measures
Hybrid graph-token	Combines pre-computed graph embeddings with	Uses graph-derived tokens for known	Balances physical structure with	Requires careful validation to avoid

approach	learnable topology identifiers	structures and learnable tokens for missing or generalized structures	transferability	misleading structural assumptions
----------	--------------------------------	---	-----------------	-----------------------------------

Model Architecture

Overall transformer backbone

The proposed foundation model could be designed around a transformer backbone that receives heterogeneous smart-grid tokens and learns shared representations for multi-horizon forecasting. Transformer-based load forecasting demonstrates the suitability of attention mechanisms for electrical demand sequences [4]. Broader deep forecasting literature also shows that transformer-style architectures can model long temporal dependencies when tokenization and positional encoding are carefully designed [2]. In this architecture, the transformer would serve not as a generic sequence model but as a cross-modal representation engine for load, weather, and topology tokens.

Multi-modal input composition

The model input could be composed of smart-meter load patches, weather-context embeddings, topology tokens, time embeddings, missingness indicators, and horizon-query tokens. Interpretable transformer-based probabilistic net-load forecasting shows that transformer models can be adapted to residential electricity prediction while supporting uncertainty-aware outputs [13]. Power substation load forecasting using temporal fusion neural networks also indicates that interpretable attention-based structures can be useful at higher grid aggregation levels [28]. The proposed foundation model would extend this logic by creating a unified token sequence that can represent both local smart-meter behavior and broader feeder or substation context.

Cross-modal attention mechanisms

Cross-modal attention should allow load tokens to attend to weather context, topology tokens to condition spatial relevance, and horizon tokens to extract task-specific forecasting evidence. Deep learning surveys emphasize that modern forecasting systems benefit from representation learning that can combine heterogeneous signals rather than relying on hand-crafted covariate interactions [1]. Multi-source load forecasting using gated recurrent units provides a precedent for integrating different information streams, although the proposed architecture would use attention-based fusion rather than recurrent concatenation [15]. Cross-modal attention would therefore enable the model to learn when a demand shift is primarily temporal, meteorological, structural, or jointly determined.

Multi-horizon output heads

The output layer could include horizon-specific heads for very-short-term, short-term, intra-day, and day-ahead forecasts, with optional point and probabilistic outputs. Probabilistic mid-term hourly load forecasting using improved temporal fusion transformers shows that deep architectures can be organized to produce uncertainty-aware outputs [12]. Short-term load forecasting based on CEEMDAN and transformer modeling also illustrates how transformer-based representations can be connected to load prediction targets [29].

Figure 2 illustrates the proposed end-to-end foundation-model architecture, showing how multimodal tokens are fused through cross-modal attention and routed to horizon-specific forecasting heads.

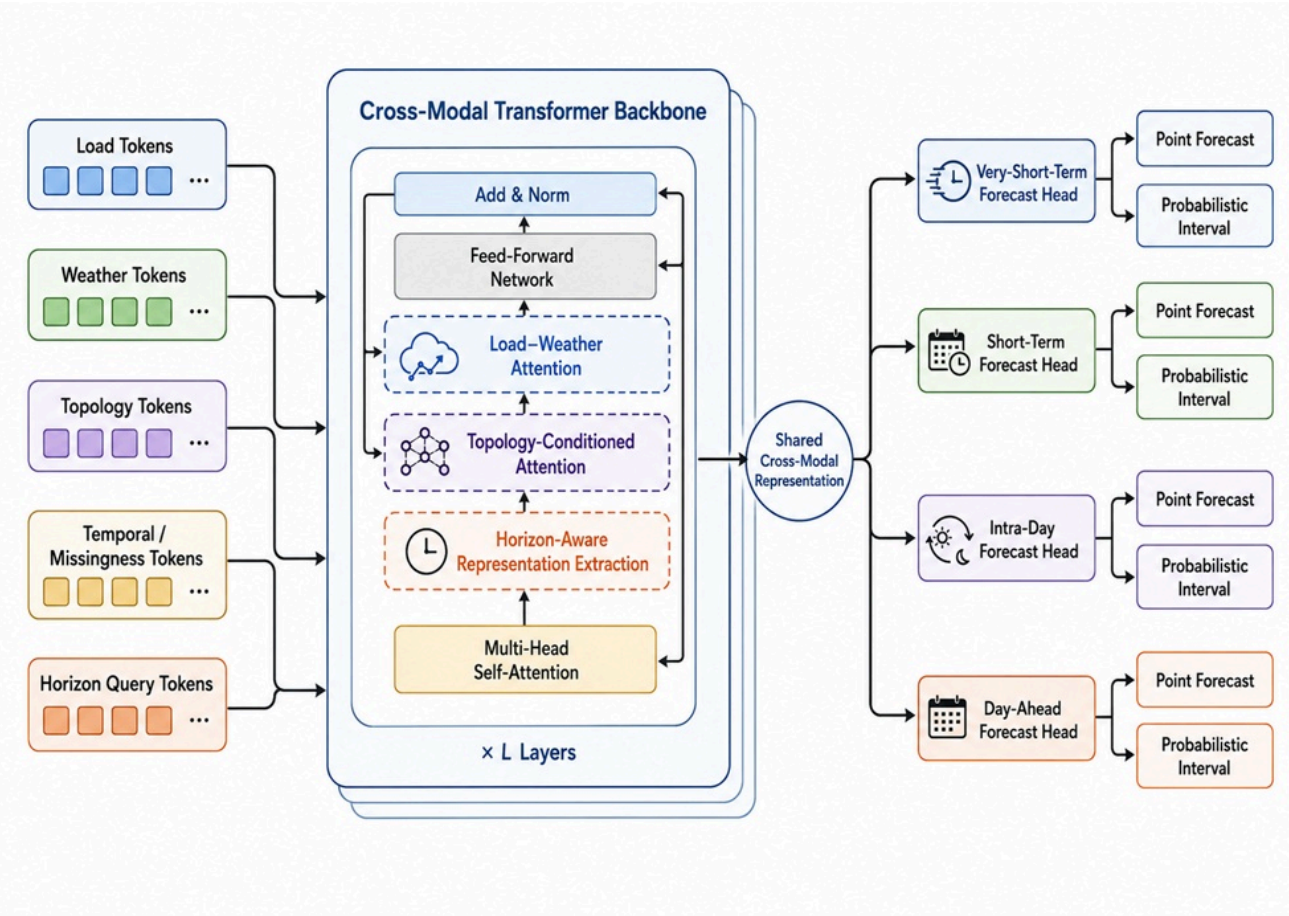


Figure 2. End-to-End Foundation-Model Architecture for Multi-Horizon Smart-Grid Forecasting

Table 4 summarizes the proposed architecture, including input composition, attention mechanisms, and output-head design.

Table 4. Proposed Foundation-Model Architecture Summary

Architectural component	Proposed design	Input or output handled	Function within the foundation model	Validation need
Smart-meter tokenizer	Patch-based, convolutional, or hybrid patch-convolution encoder	Historical smart-meter and feeder load sequences	Converts raw load histories into compact temporal tokens	Compare resolution preservation and computational scalability
Weather encoder	Continuous meteorological embedding module with lead-time encoding	Temperature, humidity, irradiance, wind, precipitation, and weather forecast lead times	Represents exogenous demand drivers in the same latent space as load tokens	Test sensitivity to weather-variable selection and spatial alignment

Topology-token module	Graph-derived embeddings, learnable topology tokens, or relative topological encodings	Feeder connectivity, transformer grouping, electrical distance, or zone identity	Conditions forecasting on structural grid context	Assess whether topology tokens improve cross-feeder transferability
Transformer backbone	Shared multi-layer attention architecture	Combined load, weather, topology, temporal, missingness, and horizon tokens	Learns reusable cross-modal representations during pre-training	Benchmark against single-modality and task-specific transformer variants
Cross-modal attention	Attention blocks or attention biases linking modality-specific token groups	Interactions among demand history, weather context, and topology structure	Enables modality-specific evidence to inform horizon-specific forecasts	Inspect attention behavior and compare fusion strategies
Multi-horizon output heads	Separate or shared heads for different forecast horizons	Very-short-term, short-term, intra-day, and day-ahead targets	Produces horizon-specific point and optional probabilistic forecasts	Evaluate transfer performance across unseen feeders and regions

Evaluation Strategy

Pre-training setup and datasets

The evaluation strategy should begin with pre-training on diverse smart-grid datasets that span multiple feeders, substations, customer mixes, weather zones, and grid configurations. Foundation-model forecasting for transformer load forecasting and overload detection suggests that pre-training can be conceptualized as a way to learn reusable grid representations before downstream adaptation [6]. Deep learning surveys on time-series forecasting also emphasize that generalizable models require exposure to diverse temporal patterns rather than narrow task-specific histories [7]. The proposed pre-training task could involve masked load-token reconstruction, next-window prediction, weather-load alignment, and topology-conditioned sequence modeling.

Fine-tuning tasks and metrics

Fine-tuning should be designed around held-out feeders, unseen substations, new geographic regions, and multiple forecasting horizons. Transfer learning with seasonal and trend adjustment demonstrates that cross-site forecasting requires adaptation mechanisms when target domains differ from source domains [23]. Short-term residential load forecasting with recurrent neural networks provides a useful baseline for feeder- or household-level temporal adaptation [10]. The proposed model should be evaluated conceptually through point-error metrics, probabilistic calibration metrics, robustness to missing data, and transfer-performance comparisons, without presenting experimental results in this architecture article.

Baseline models

Baseline comparisons should include single-modal load-only models, weather-augmented models, topology-aware graph models, single-horizon transformers, recurrent sequence models, and task-specific deep networks. Deep residual networks for short-term load forecasting [11], pooling recurrent neural networks for household forecasting [14], and optimal LSTM models with feature selection [18] provide representative single-task or task-specific deep-learning baselines. Graph neural network models [25] and multi-resolution graph forecasting architectures [26] would serve as

topology-aware baselines. Transformer-based load forecasting models [3] and temporal fusion forecasting models [28] would provide attention-based baselines for evaluating whether the proposed unified tokenization adds conceptual value.

Expected outcomes

Because this article is conceptual, expected outcomes should be framed as evaluation hypotheses rather than empirical claims. The proposed model would be expected to show advantages if unified tokenization enables stronger cross-feeder adaptation, better horizon consistency, improved robustness to missing data, and more coherent use of weather and topology information. Long-sequence forecasting research suggests that scalable tokenization and attention design are central to such performance questions [9]. **Table 5** outlines an evaluation plan that could be used in future empirical work to validate the architecture.

Table 5. Evaluation Plan for the Foundation Model

Evaluation dimension	Proposed design	Comparator group	Success criterion	Reason for inclusion
Pre-training diversity	Use multiple feeders, substations, meter groups, weather regions, and topology configurations	Pre-training on a single feeder or narrow region	Improved transfer to unseen feeders and regions	Tests whether the model learns general smart-grid representations
Smart-meter tokenization	Compare patching, convolutional encoding, down-sampling, and hybrid tokenization	Raw sequence models or fixed-window recurrent models	Better balance between forecast accuracy, robustness, and computational cost	Assesses whether load-token design affects multi-horizon usability
Weather integration	Compare weather-token fusion with no-weather and late-fusion weather baselines	Load-only and weather-appended baselines	Improved performance under weather-sensitive demand conditions	Tests whether weather embeddings add useful contextual information
Topology integration	Compare graph-derived tokens, learnable topology tokens, and no-topology variants	Non-topological transformer and graph-only forecasting models	Improved cross-feeder adaptation and structural generalization	Evaluates whether topology tokens improve transferability
Multi-horizon forecasting	Use shared backbone with horizon-specific heads	Separate models for each horizon	More consistent forecasts across very-short-term, short-term, and day-ahead horizons	Tests whether shared representations support multiple forecast horizons
Robustness and missingness	Evaluate masked readings, irregular intervals, and incomplete topology metadata	Models without missingness or data-quality tokens	Lower degradation under incomplete data conditions	Assesses deployment relevance in real smart-meter environments
Interpretability and auditability	Inspect attention patterns, token salience, and modality contribution	Black-box single-output models	More transparent evidence use across load, weather, and topology inputs	Supports responsible model assessment before operational deployment

Transferability

Cross-feeder generalization

Cross-feeder generalization is central to the proposed foundation-model architecture because utilities need forecasting models that can adapt to new feeders without full retraining from scratch. Transfer-learning research for cross-building energy forecasting shows that forecasting knowledge can be transferred when seasonal and trend differences are addressed explicitly [23]. Time-series pre-trained model surveys similarly suggest that pre-training can create reusable temporal representations that later support downstream adaptation [5]. In this architecture, feeder-level adaptation could be enabled through topology tokens, local normalization layers, and horizon-specific fine-tuning heads.

Adaptation to new grid topologies

Adaptation to new grid topologies requires the model to handle different feeder structures, transformer groupings, customer densities, and connectivity patterns. Graph convolutional recurrent load forecasting shows that explicit network structure can improve modeling of spatially related load behavior [20]. Multi-resolution graph neural network forecasting further indicates that electricity demand may need to be represented at several structural scales rather than at a single aggregation level [26]. The proposed foundation model could adapt to new topologies by recalculating graph embeddings, updating learnable topology tokens, or using relative topological encodings that are less dependent on a fixed feeder graph.

Limitations and future enhancements

The proposed architecture remains a conceptual design and would require empirical validation before any claims could be made about forecasting superiority, computational feasibility, or operational reliability. Experimental reviews of deep learning architectures for time-series forecasting caution that model performance can vary substantially by dataset, horizon, and preprocessing choice [8]. Deep learning surveys also emphasize that forecasting models must be assessed against meaningful baselines rather than evaluated only through architectural novelty [2]. Future enhancements could include physics-informed constraints, privacy-preserving pre-training, uncertainty calibration, topology-change detection, and operational interpretability modules.

Conclusion

This article proposed a foundation-model architecture for multi-horizon smart-grid forecasting using smart-meter sequences, weather context, and grid-topology tokens. The central innovation is the conceptual unification of temporal demand traces, meteorological context, and feeder structure into a shared transformer-based token space.

The architecture would allow smart-meter sequences to be represented through patches or convolutional latent tokens, weather information through learned continuous embeddings, and grid topology through graph-derived or learnable structural tokens. Cross-modal attention would enable these modalities to interact before forecasts are generated through horizon-specific output heads.

The proposed model could support pre-training on diverse grid data and fine-tuning for very-short-term, short-term, intra-day, and day-ahead forecasting tasks. This design would be especially relevant for utilities seeking models that can transfer across feeders, substations, regions, and changing grid configurations.

Further work should validate the architecture through rigorous benchmarking, baseline comparison, uncertainty assessment, and deployment-oriented evaluation. The framework provides a blueprint for future empirical studies but does not itself claim experimental performance or operational readiness.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Oct 2025

Revised: 21 Nov 2025

Accepted: 30 Dec 2025

Published online: 20 March 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Lim B, Zohren S. Time-series forecasting with deep learning: a survey. *Philos Trans R Soc A Math Phys Eng Sci*. 2021;379(2194).
- Benidis K, Rangapuram SS, Flunkert V, Wang Y, Maddix D, Turkmen C, et al. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Comput Surv*. 2022;55(6):1-36.
- L'Heureux A, Grolinger K, Capretz MA. Transformer-based model for electrical load forecasting. *Energies*. 2022;15(14):4993.
- Zhao Z, Xia C, Chi L, Chang X, Li W, Yang T, et al. Short-term load forecasting based on the transformer model. *Information*. 2021;12(12):516.
- Ma Q, Liu Z, Zheng Z, Huang Z, Zhu S, Yu Z, et al. A survey on time-series pre-trained models. *IEEE Trans Knowl Data Eng*. 2024;36(12):7536-55.
- Hou Y, Ma C, Li X, Sun Y, Yu H, Fang Z. Time series foundation model for improved transformer load forecasting and overload detection. *Energies*. 2025;18(3):660.
- Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep learning for time series forecasting: A survey. *Big Data*. 2021;9(1):3-21.

Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. *Int J Neural Syst*. 2021;31(3):2130001.

Chen Z, Ma M, Li T, Wang H, Li C. Long sequence time-series forecasting with deep learning: A survey. *Inf Fusion*. 2023;97:101819.

Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans Smart Grid*. 2017;10(1):841-51.

Chen K, Chen K, Wang Q, He Z, Hu J, He J. Short-term load forecasting with deep residual networks. *IEEE Trans Smart Grid*. 2018;10(4):3943-52.

Li D, Tan Y, Zhang Y, Miao S, He S. Probabilistic forecasting method for mid-term hourly load time series based on an improved temporal fusion transformer model. *Int J Electr Power Energy Syst*. 2023;146:108743.

Xu C, Chen G. Interpretable transformer-based model for probabilistic short-term forecasting of residential net load. *Int J Electr Power Energy Syst*. 2024;155:109515.

Shi H, Xu M, Li R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Trans Smart Grid*. 2017;9(5):5271-80.

Wang Y, Liu M, Bao Z, Zhang S. Short-term load forecasting with multi-source data using gated recurrent unit neural networks. *Energies*. 2018;11(5):1138.

Gong G, An X, Mahato NK, Sun S, Chen S, Wen Y. Research on short-term load prediction based on Seq2seq model. *Energies*. 2019;12(16):3199.

Bedi J, Toshniwal D. Deep learning framework to forecast electricity demand. *Appl Energy*. 2019;238:1312-26.

Bouktif S, Fiaz A, Ouni A, Serhani MA. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*. 2018;11(7):1636.

Tian C, Ma J, Zhang C, Zhan P. A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies*. 2018;11(12):3493.

Arastehfar S, Matinkia M, Jabbarpour MR. Short-term residential load forecasting using graph convolutional recurrent neural networks. *Eng Appl Artif Intell*. 2022;116:105358.

Zhang G, Wei C, Jing C, Wang Y. Short-term electrical load forecasting based on time augmented transformer. *Int J Comput Intell Syst*. 2022;15(1):67.

He W. Load forecasting via deep neural networks. *Procedia Comput Sci*. 2017;122:308-14.

Ribeiro M, Grolinger K, ElYamany HF, Higashino WA, Capretz MA. Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy Build*. 2018;165:352-63.

Rajasundrapandiyanleebanon T, Kumaresan K, Murugan S, Subathra MS, Sivakumar M. Solar energy forecasting using machine learning and deep learning techniques. *Arch Comput Methods Eng*. 2023;30(5):3059-79.

Sun C, Ning Y, Shen D, Nie T. Graph neural network-based short term load forecasting with temporal convolution. *Data Sci Eng*. 2024;9(2):113-32.

Mo S, Yao Z, Yang C, Wang H, Fan S, Liu R, et al. MRGNN: Multi-resolution graph neural network for electricity load forecasting. *Energy Rep*. 2025;13:4624-34.

Zhu L, Gao J, Zhu C, Deng F. Short-term power load forecasting based on spatial-temporal dynamic graph and multi-scale Transformer. *J Comput Des Eng*. 2025;12(2):92-111.

Ferreira AB, Leite JB, Salvadeo DH. Power substation load forecasting using interpretable transformer-based temporal fusion neural networks. *Electr Power Syst Res*. 2025;238:111169.

Ran P, Dong K, Liu X, Wang J. Short-term load forecasting based on CEEMDAN and Transformer. *Electr Power Syst Res*. 2023;214:108885.