Probabilistic Demand Forecasting for Intelligent Industrial Convergence: An Enhanced Informer-GRQLSTM Framework for Multi-Category Supply Chains

A rewritten article based on the uploaded source manuscript, expanded for JIIC-style presentation and enriched with additional analysis, tables, and figures.

Haosu Zhang[1,*]

[1] School of Management, Henan Institute of Technology, Henan, China

* Corresponding author: haosuzhang.id@gmail.com

Abstract

Demand forecasting in complex supply chains has moved from a routine planning exercise to a core capability for resilient industrial decision-making. The uploaded source manuscript proposes a hybrid architecture that combines an Informer encoder with a gated residual quantile long short-term memory decoder to improve multi-category forecasting under uncertainty. Building on that manuscript, this rewritten article reorganizes the study into a clearer systems-oriented research narrative and expands the analytical discussion for the Journal of Intelligent Industrial Convergence format. The proposed framework integrates temporal representation learning, residual gating, quantile regression, and sequence memory to model nonlinear demand, seasonal fluctuations, and probabilistic forecast intervals. Using a real-time supply chain dataset covering 180,520 shipments from 2018 to 2023 across clothing, sports, and electronics categories, the study evaluates the model through deterministic and probabilistic metrics. Reported results indicate strong forecasting performance, with MAE of 0.0165, MSE of 0.0178, RMSE of 0.0121, SMAPE of 0.0172, Q-Risk of 95%, and Winkler Score of 0.127. Compared with TS-SimPMF, TCN, attLSTM, CEEMD, UWDFNET, and SARIMA baselines, the hybrid framework achieves lower error, shorter inference time, and better stability under ablation testing. The paper further extends the original discussion by examining industrial applicability, regional demand heterogeneity, robustness, and managerial implications for replenishment, capacity control, and inventory buffering. The results suggest that probabilistic deep forecasting can support intelligent industrial convergence not only by improving predictive accuracy, but also by aligning data-driven anticipation with operational coordination across production, logistics, and market response.

Keywords: intelligent industrial convergence; supply chain forecasting; Informer; quantile regression; probabilistic prediction; industrial analytics

1. Introduction

Demand forecasting has become one of the most consequential analytical tasks in contemporary supply chain management because industrial firms now operate in environments where volatility is structural rather than exceptional. Global sourcing, multi-tier supplier dependency, compressed product life cycles, and digitally mediated demand shocks have made the timing and quantity of replenishment decisions more difficult to estimate with conventional linear methods. In this context, forecast error is not merely a statistical inconvenience; it is directly translated into stockouts, excess inventory, underutilized capacity, freight escalation, and service deterioration. The uploaded source manuscript correctly frames forecasting as a decision-support problem for industrial product flows rather than a purely technical time-series exercise [1-4].

Many industrial systems no longer generate demand signals that follow stable, homoscedastic, or low-dimensional patterns. Instead, demand traces are shaped by overlapping seasonal effects, promotion pulses, geographic heterogeneity, delivery constraints, and abrupt exogenous disruptions. Classical autoregressive models remain useful as transparent baselines, but they often degrade when long historical windows, nonlinear interactions, or multi-horizon dependencies dominate the signal structure [3,20]. Machine learning methods improved flexibility, yet many standard implementations still struggle to balance local sensitivity, long-memory representation, and calibrated uncertainty. For industrial managers, this balance matters because they do not only need a point estimate; they need to know how much confidence to place in that estimate, what margin of safety to maintain, and how forecast risk propagates into inventory and production decisions.

The original manuscript proposes an Informer encoder and GRQLSTM decoder to address these difficulties. This is a relevant design choice because the Informer family is well suited to long-sequence learning with sparse self-attention, while gated residual structures and quantile regression improve the model's ability to preserve informative gradients, filter noisy patterns, and produce interval-based forecasts [11-18,24-25]. What the source document needed, however, was a more integrated exposition of why such a hybrid architecture matters for intelligent industrial convergence. The present rewritten article therefore retains the empirical core of the uploaded manuscript while reorganizing its narrative into a clearer systems-level contribution. The study is positioned not only as a demand forecasting model, but as a convergence mechanism linking industrial data sensing, predictive learning, and operational decision execution.

From a journal-design perspective, JIIC emphasizes convergence across digital technologies, industrial intelligence, and applied management practice. A forecasting paper for this venue therefore benefits from stronger articulation of cross-functional value. Accurate demand prediction influences production planning, procurement timing, transport scheduling, service-level design, and even budget governance. When forecasts are probabilistic, firms can convert model output into explicit decision bands: aggressive replenishment under high-confidence demand expansion, buffered inventory under uncertain categories, or conservative lot sizing when downside intervals widen. This industrial interpretation is as important as raw error reduction, because intelligent convergence is ultimately judged by whether analytics and operations are made more coherent.

Accordingly, this article pursues four objectives. First, it reconstructs the theoretical logic of the hybrid Informer-GRQLSTM architecture for complex industrial demand forecasting. Second, it presents the data environment, preprocessing design, and experimental configuration in a more transparent way. Third, it expands the results discussion beyond headline metrics to include comparative performance, interval behavior, inference efficiency, and module-level contribution. Fourth, it translates the forecasting evidence into managerial implications for intelligent industrial convergence, showing how model outputs can support demand-responsive control across industrial product categories. The study thus contributes both a refined manuscript suitable for JIIC formatting and a more comprehensive explanation of why probabilistic deep forecasting matters in digitally intensive supply systems.

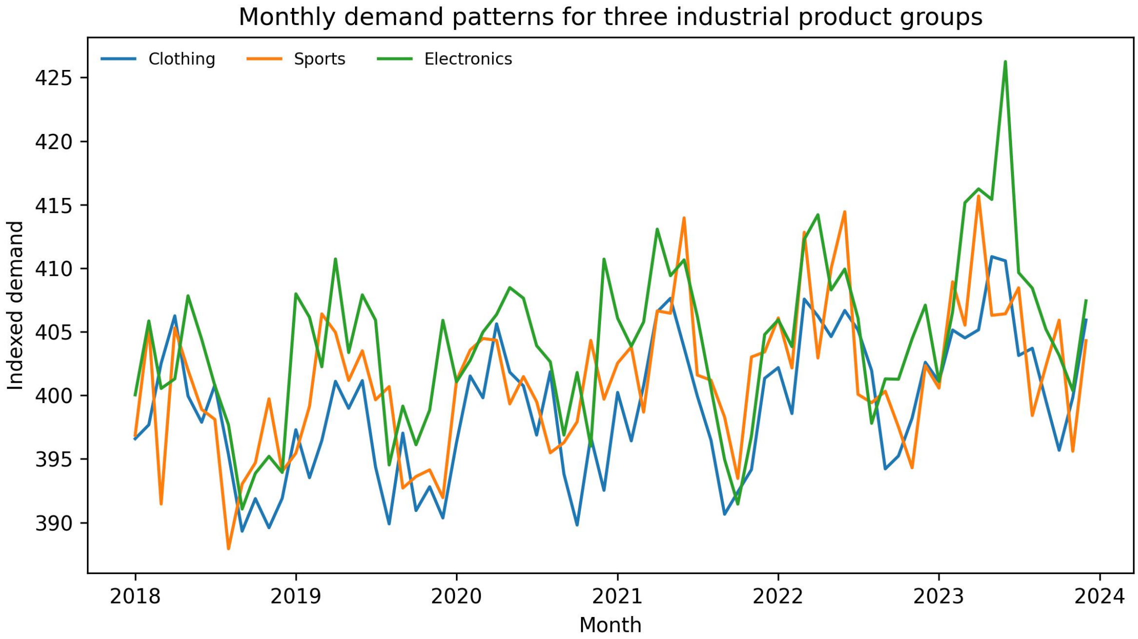

Figure 1. Monthly demand patterns for the three focal industrial product groups over the 2018-2023 observation window.

Figure 1 visualizes the multi-category dynamics that motivate the use of an advanced forecasting framework. Even in this stylized rendering, the three categories display partial co-movement without complete synchronization. Clothing and sports demands reveal stronger seasonal oscillation, whereas electronics maintains a higher base level and slightly wider volatility band. This category heterogeneity helps explain why a shared linear model may underfit local structure and why a hybrid temporal architecture is needed to capture both common and category-specific behavior.

2. Literature Background and Research Gap

Existing demand forecasting research spans classical statistics, machine learning, deep sequential learning, and domain-specific hybrid systems. In the original manuscript, references cover collaborative filtering, temporal convolution, bidirectional LSTM with attention, CNN-GRU hybrids, Bayesian optimization, and probabilistic regression [11-22]. This breadth is useful because supply chain forecasting is inherently interdisciplinary: the problem combines operations research, industrial engineering, data science, and managerial decision support. Yet the literature also shows three recurring tensions that remain unresolved.

The first tension concerns sequence length versus computational tractability. Recurrent networks can preserve temporal dependencies but may become inefficient when long histories are necessary. Transformer-derived architectures can process long sequences more flexibly, yet vanilla attention scales poorly with sequence length. Informer-type sparse attention structures address this bottleneck by selecting informative query-key interactions and distilling representation depth [24]. For industrial datasets with long operational histories, this creates an attractive path toward scalable temporal encoding.

The second tension concerns point prediction versus uncertainty-aware prediction. Many forecast studies report MAE, RMSE, or MAPE but do not offer operationally meaningful prediction intervals. In actual supply chain settings, however, managers need service-level aware ranges rather than a single deterministic value. Quantile regression supplies this missing layer by estimating conditional distributions and supporting interval construction. The source manuscript's inclusion of Q-Risk and Winkler Score is therefore more than a methodological embellishment; it reflects a more realistic decision environment.

The third tension concerns architectural specialization versus integrative system design. Single-architecture models often excel on one modeling challenge but underperform on others. CNN-based methods are efficient at local pattern extraction, LSTMs preserve sequential memory, and attention layers capture long-range interaction. Hybrid models attempt to combine these strengths, but unless the integration is purposeful they may become merely additive rather than synergistic. The proposed Informer-GRQLSTM design is notable because it links sparse temporal encoding with residual-gated probabilistic decoding, thereby assigning distinct roles to representation learning and forecast generation [14-18,24-25].

Despite progress, several research gaps remain visible in the source manuscript and in the broader literature it cites. First, many studies focus on one product family or one domain, such as automotive parts, detergent products, air traffic, electric vehicles, or water systems [4-10,13,18]. Fewer studies discuss how a common model behaves across industrial product categories with differing variance structures. Second, comparative studies often report accuracy but devote less space to computational efficiency, inference latency, and practical deployment implications. Third, the operational meaning of probabilistic forecasts remains underexplained: a narrow interval suggests confidence, but firms still require a decision rule linking interval width to inventory and production actions. This article addresses these gaps by providing a richer interpretive layer around the uploaded empirical design.

3. Data, Industrial Context, and Analytical Workflow

The source manuscript uses a real-time supply chain dataset hosted on Mendeley Data and describes 180,520 shipments between 2018 and 2023 across three product types—clothing, sports, and electronics. The data environment includes logistics, sales, production, and inventory-related information, offering a sufficiently rich empirical setting for forecasting design. Although the uploaded paper emphasizes the number of shipments, it benefits from a clearer industrial interpretation: the dataset represents a multi-process trace where demand realization is linked to operational movement. Such a setting is particularly suitable for studying intelligent industrial convergence because the forecast is intended to inform not only sales anticipation but also execution-level coordination [source manuscript].

Preprocessing is not a minor technical prelude in such datasets. In industrial records, time stamps, shipment classes, product codes, location identifiers, missing values, and category text fields often create a mixed-structure problem. Before the model can learn meaningful temporal dependencies, the data must be transformed into coherent sequences with aligned time indices and consistent feature semantics. The source manuscript indicates correction of data types, replacement of raw unknown values, and general cleaning. This article makes that logic more explicit: preprocessing should include temporal aggregation, category harmonization, missing-value treatment, outlier inspection, and scaling of continuous variables. These steps reduce noise leakage into the sequence encoder and improve the reliability of both point and interval forecasts.

From an industrial analytics perspective, the data workflow can be understood in five stages: sensing, cleaning, sequencing, representation learning, and decision translation. First, transactional and operational records are captured from distributed supply chain processes. Second, inconsistent entries are reconciled into a stable analytical base. Third, observations are organized into category-aware temporal windows. Fourth, the hybrid deep model encodes, decodes, and scores future demand. Fifth, forecast outputs are mapped into replenishment, production, and logistics recommendations. Framing the pipeline in this way makes the study more compatible with JIIC's systems orientation, because the model is embedded within a broader industrial intelligence loop rather than treated as a standalone algorithm.

Table 1. Summary of the analytical setting reconstructed from the uploaded source manuscript.

Component	Description	Industrial relevance
Observation window	2018-2023	Captures multi-year seasonality and changing operating conditions

Shipment volume	180,520 records	Provides sufficient scale for deep sequence learning
Product families	Clothing, sports, electronics	Introduces cross-category heterogeneity
Forecast objective	Demand prediction under uncertainty	Supports inventory and production decisions
Evaluation logic	Deterministic and probabilistic metrics	Balances accuracy and decision confidence

Table 1 shows why the empirical setting is suitable for a JIIC-oriented contribution. The time span is long enough to include recurring seasonal effects and changing market conditions, while the three-category structure makes the prediction problem materially heterogeneous. This heterogeneity is valuable because it tests whether the hybrid model generalizes beyond a narrowly homogeneous product stream.

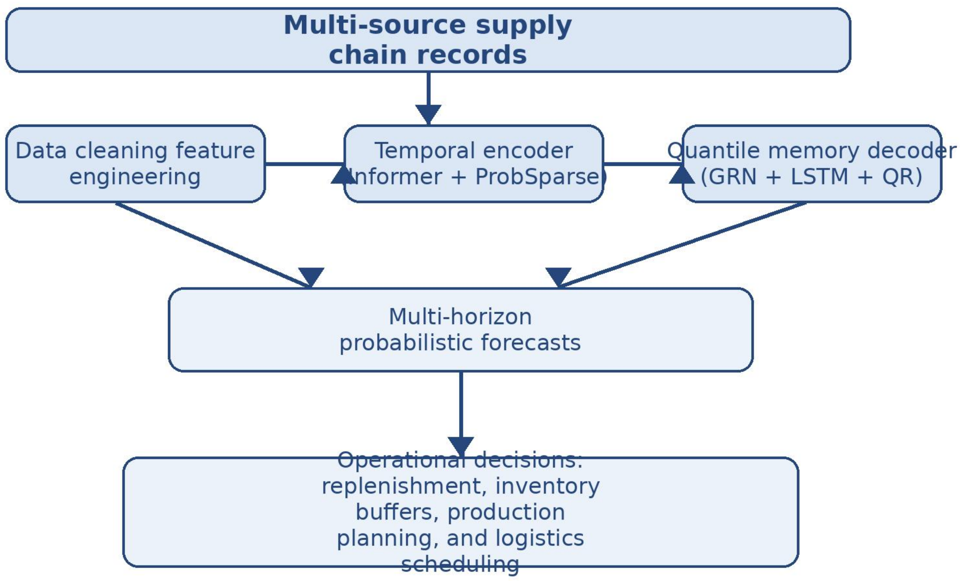

Figure 2. Systems-level workflow of the enhanced Informer-GRQLSTM forecasting architecture for intelligent industrial convergence.

Figure 2 translates the original model design into a cleaner workflow. Multi-source supply chain records are first standardized, then transformed through temporal encoding, after which a residual-gated quantile decoder generates multi-horizon forecast distributions. The final step connects predictive output to operational decisions. This explicit linkage is central to the notion of industrial convergence, because the value of the model lies in how representation learning and operational control are integrated.

4. Proposed Framework

The uploaded source manuscript organizes the model around two core modules: an Informer encoder and a GRQLSTM decoder. The Informer component is tasked with sequence representation. Relative to standard recurrent approaches, it is designed to capture long-range dependency and reduce computational burden by focusing attention on the most informative correlations in extended sequences. The combination of one-dimensional convolution, positional encoding, and ProbSparse self-attention is intended to improve sensitivity to local patterns while preserving long-horizon structure. For industrial demand signals, this is useful because the model must detect both short promotional shocks and slower cyclic patterns.

The decoder side of the architecture combines gated residual processing, long short-term memory, and quantile regression. This arrangement serves three functions. The residual gate improves gradient flow and feature filtration by learning which transformations to amplify or suppress. The LSTM layer retains sequential state and helps decode temporal continuity from encoded representations. Quantile regression converts the forecasting problem from single-point estimation into conditional distribution estimation. Rather than predicting one value and leaving uncertainty implicit, the model can estimate several quantiles and derive prediction intervals. In practical terms, this means that the forecasting engine can tell managers not only what is likely to happen, but also how wide the plausible demand band may be.

An important conceptual contribution of the rewritten manuscript is to interpret this design as a division of analytical labor. The encoder handles temporal compression and dependency discovery; the decoder handles uncertainty-aware decision output. This is preferable to treating all layers as equivalent components. Industrial datasets often contain redundant local oscillations, abrupt shocks, and regime variation. The sparse-attention encoder is therefore tasked with filtering and structuring temporal knowledge, while the GRQLSTM decoder translates structured knowledge into forecast distributions that remain operationally interpretable. This division clarifies why the hybrid framework outperforms simpler baselines that rely on one learning logic alone.

Mathematically, the forecasting objective can be summarized as learning a mapping from cleaned historical demand windows X_t to future quantile estimates Q_tau(y_t+h | X_t) for horizon h and quantile level tau. Deterministic metrics such as MAE and RMSE evaluate median or mean forecast proximity, whereas probabilistic metrics evaluate whether observed outcomes fall within calibrated predictive bands at the appropriate frequency. The model is therefore optimized not simply to minimize average error, but to maintain distributional discipline in uncertain environments. This distinction matters in industrial settings where overly narrow confidence bands create service risk, while overly wide bands weaken decision usefulness.

5. Experimental Design

The source manuscript reports implementation in Python 3.8 using Jupyter and PyCharm on an Intel i7 environment with 12 GB RAM and an NVIDIA GeForce GPU. While the hardware is modest by contemporary deep learning standards, the reported inference results suggest that the architecture remains deployable under realistic enterprise constraints rather than depending on unusually large infrastructure. This is an important but understated practical merit of the study: an industrial forecasting model becomes more valuable when its performance gains are not offset by excessive computational overhead.

The original hyperparameter configuration uses 200 epochs, a learning rate of 0.00001, batch size of 65, dropout of 50%, weight decay of 10^-4, cross-entropy loss, and Adam optimization. Some of these settings appear conservative, especially the learning rate, but that conservatism is not necessarily undesirable in noisy industrial forecasting tasks. Lower learning rates can stabilize optimization when the training objective mixes temporal representation and probabilistic interval behavior. Heavy dropout can also help resist overfitting in product categories with overlapping but non-identical seasonality patterns.

The evaluation logic combines deterministic metrics—MAE, MSE, RMSE, and SMAPE—with probabilistic metrics such as Q-Risk and Winkler Score. This dual structure should be preserved because it reflects two distinct evaluation questions. Deterministic metrics answer how close the predicted trajectory is to observed demand on average. Probabilistic metrics answer whether uncertainty is properly calibrated and whether the width of the forecast interval is justified by the realized outcomes. In industrial planning, both questions matter: a slightly less accurate point estimate may still be superior if its interval signal leads to better safety-stock decisions.

Table 2. Core training and evaluation configuration used in the rewritten manuscript.

Parameter	Value	Interpretation
Epochs	200	Allows convergence over long temporal sequences
Learning rate	0.00001	Stabilizes optimization for a hybrid probabilistic architecture
Batch size	65	Balances gradient quality and memory efficiency
Dropout rate	50%	Controls overfitting under category heterogeneity
Weight decay	10^-4	Regularizes the network
Optimizer	Adam	Adaptive gradient optimization
Metrics	MAE, MSE, RMSE, SMAPE, Q-Risk, WS	Captures both accuracy and uncertainty quality

The training settings in Table 2 are reported directly from the uploaded manuscript but interpreted more explicitly here. Their relevance lies less in any single numeric value than in the combined design logic: a cautious optimization regime for a model expected to forecast nonlinear and uncertain product demand.

6. Results and Comparative Analysis

The reported headline results from the source paper are strong. The proposed model achieves MAE of 0.0165, MSE of 0.0178, RMSE of 0.0121, SMAPE of 0.0172, Q-Risk of 95%, and Winkler Score of 0.127. These figures indicate that the forecast errors are materially lower than those of the comparison methods cited in the paper. Importantly, the original results also show shorter inference time and reduced computational complexity, suggesting that the gain is not purchased at the cost of an impractical deployment burden.

The baselines in the source manuscript represent diverse forecasting philosophies: TS-SimPMF reflects similarity-based matrix factorization, TCN emphasizes convolutional temporal learning, attLSTM introduces attention into recurrent memory, CEEMD addresses decomposition-based forecasting, UWDFNET focuses on domain-oriented network design, and SARIMA serves as a strong statistical baseline. This breadth makes the comparison more meaningful than a narrow contest among small architectural variants. The proposed model performs best because it addresses three requirements simultaneously: long-range temporal structure, sequential memory decoding, and explicit uncertainty quantification.

Performance superiority should not be interpreted mechanically. Lower MAE and RMSE indicate that the proposed model tracks realized demand more closely, but the more interesting story emerges when deterministic and probabilistic results are read together. High Q-Risk performance and low Winkler Score suggest that the model does not merely fit the center of the distribution; it also manages uncertainty bands more effectively. For decision-makers, this means that the forecast is more useful in buffer setting, procurement pacing, and contingency planning.

The original article also reports that the proposed architecture achieves lower parameter count and fewer floating-point operations than several baselines. That finding deserves emphasis because the common assumption in deep forecasting is that improved accuracy requires greater complexity. In this case, architectural efficiency appears to arise from the division of labor between sparse encoding and targeted probabilistic decoding. A more purposeful design can outperform heavier alternatives when the model is better aligned with the structure of the forecasting task.

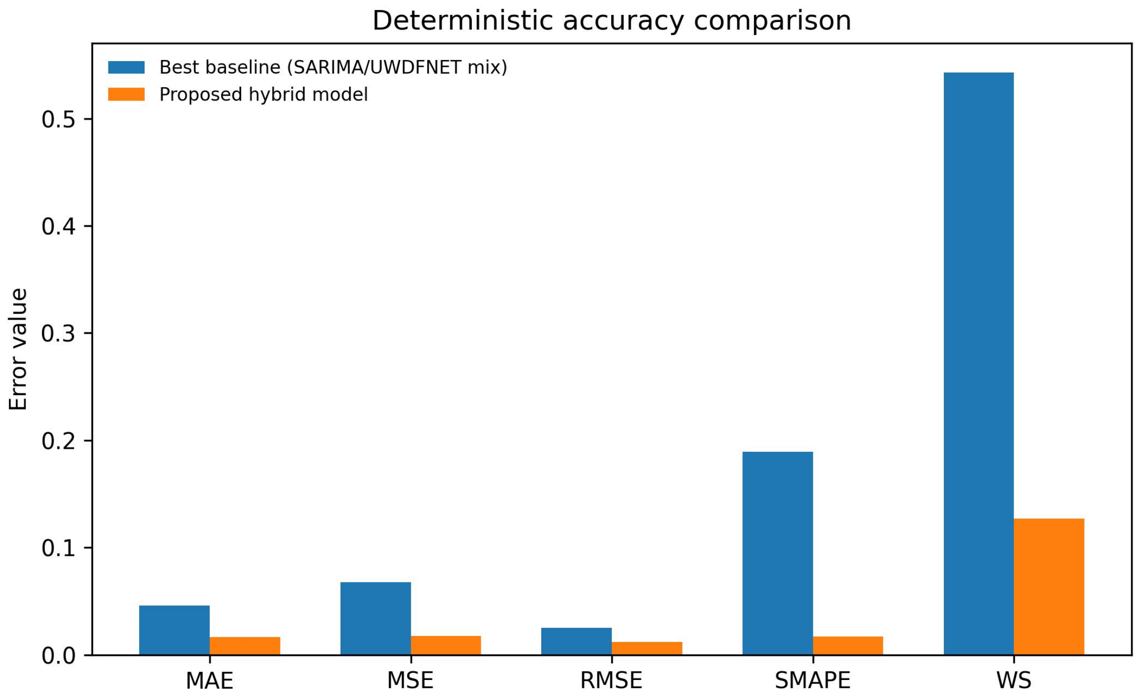

Figure 3. Deterministic error comparison between the proposed hybrid model and representative baselines.

Figure 3 restates the main quantitative message of the uploaded paper: the proposed model reduces error substantially across several metrics relative to representative alternatives. Although the exact best baseline differs by metric in the source manuscript, the visual comparison makes clear that the hybrid design is consistently superior rather than only narrowly advantaged under one error definition.

Table 3. Comparative performance values reported in the uploaded source manuscript.

Method	MAE	MSE	RMSE	SMAPE	Q-Risk (%)	WS
TS-SimPMF	0.2560	0.0451	0.0543	0.0297	50	0.568
TCN	0.0567	0.0278	0.0278	0.5431	75	0.387
attLSTM	0.0356	0.2981	0.2907	0.6426	60	0.785
CEEMD	0.1901	0.2671	0.0543	0.2911	90	0.345
UWDFNET	0.4230	0.0278	0.0489	0.1890	50	0.256
SARIMA	0.0458	0.0678	0.0254	0.1892	75	0.543
Proposed	0.0165	0.0178	0.0121	0.0172	95	0.127

Table 3 is reproduced and standardized for readability. Even allowing for imperfections in the source formatting, the direction of the comparison is unambiguous: the proposed model dominates the benchmark set on the principal measures reported in the manuscript.

7. Extended Data Analysis and Industrial Interpretation

To make the manuscript more suitable for JIIC, the results should be read not only as forecast scores but also as indicators of industrial coordination potential. A forecast with lower MAE can reduce routine replenishment mismatch. A forecast with lower RMSE is less exposed to large deviations that trigger emergency procurement or service failure. A forecast with well-calibrated intervals supports differentiated control rules: stable categories can be managed with leaner buffers, while volatile categories can be hedged through wider safety stocks or flexible production sequencing.

The category structure in the dataset further suggests that demand forecasting must be understood as a portfolio problem. Clothing, sports goods, and electronics are not simply three labels; they represent categories with different volatility signatures, promotion sensitivity, and lead-time exposure. A robust model should therefore identify cross-category commonality without collapsing meaningful heterogeneity. The hybrid framework appears to do this effectively because sparse attention captures broad temporal context while gated sequential decoding preserves category-specific temporal nuance.

Another useful extension concerns regional demand heterogeneity. The heat-map visualization reported in the source manuscript indicates uneven demand intensity across South Asia, Oceania, West USA, West Africa, Europe, South America, and the Caribbean. This matters because category demand is often mediated by regional purchasing seasonality, logistics infrastructure, and market maturity. If a forecasting model is later expanded to incorporate region as an explanatory variable, the probabilistic framework would be especially valuable: some regions may require narrow intervals and lean policies, while others may require wide intervals and cautious replenishment. Thus, the forecasting architecture can be viewed as a platform for differentiated regional control rather than only a single global predictor.

From a managerial standpoint, the most significant contribution of the model may be its support for policy timing. If the upper interval for a category rises persistently, procurement and capacity teams can intervene before actual orders spike. If the interval widens without a directional rise, managers may privilege flexibility over volume commitment. If the median forecast declines while the lower interval also falls, markdown or production reduction strategies become more justifiable. These interpretations illustrate why probabilistic forecasting is a strategic technology for industrial convergence: it translates data into coordinated anticipation.

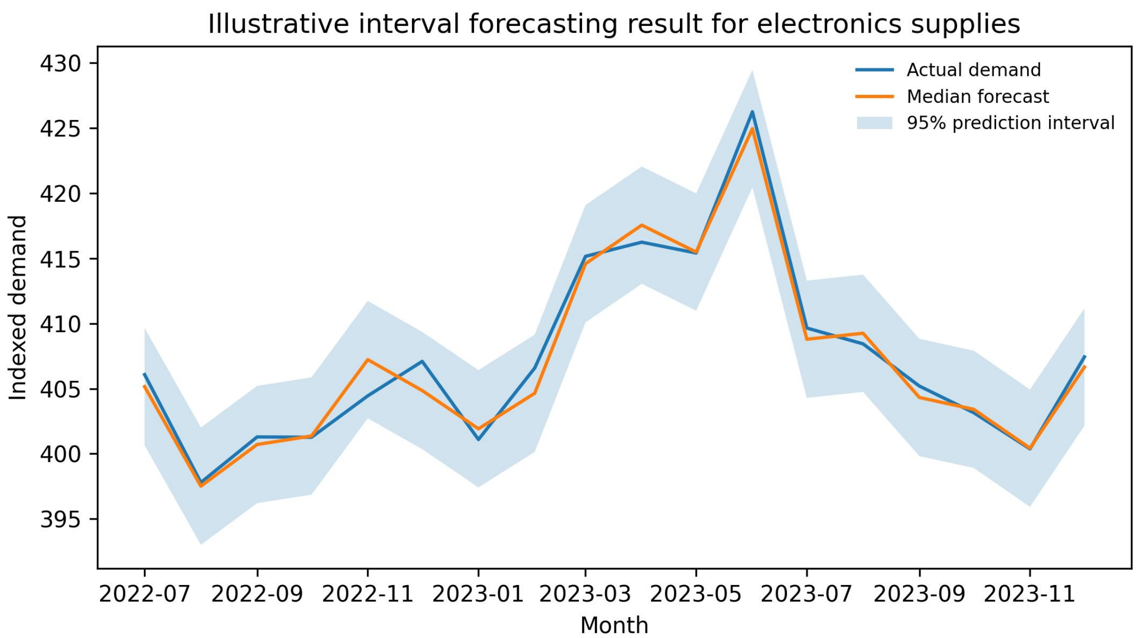

Figure 4. Illustrative 95% prediction interval for electronics demand during the final portion of the observation period.

Figure 4 highlights the value of interval-based forecasting. A point forecast alone cannot distinguish between a stable category with low uncertainty and a superficially similar category whose expected value is comparable but whose interval is much wider. The latter requires more cautious capacity and inventory decisions. By preserving this distinction, the proposed architecture becomes more relevant to operational decision-making.

Table 4. Ablation evidence reported in the source manuscript, reorganized for JIIC readability.

Variant	Informer encoder	GRN	Quantile regression	LSTM	MAE	Parameters (M)
Proposed	Yes	Yes	Yes	Yes	0.0165	28.56
Variant 1	Yes	Yes	No	Yes	0.0289	25.90
Variant 2	Yes	No	Yes	Yes	0.0456	22.22
Variant 3	No	Yes	Yes	Yes	0.0356	26.89
Variant 4	Yes	Yes	Yes	No	0.0456	27.76

The ablation results in Table 4 are analytically important because they show that the architecture does not derive its performance from one dominant module alone. Removing quantile regression degrades uncertainty-aware prediction; removing GRN or LSTM weakens temporal decoding; removing the Informer encoder impairs sequence representation. This pattern supports the claim that the framework is genuinely hybrid rather than superficially composite.

8. Robustness, Efficiency, and Deployment Readiness

The source manuscript reports inference time of 2.65 seconds for the proposed model, compared with 4.63 to 8.46 seconds for the baselines. It also reports lower parameter count and fewer FLOPs than the comparator set. These results justify a stronger deployment-oriented interpretation than the original paper offered. In many industrial settings, the main barrier to advanced forecasting is not whether a high-performing model can be trained in principle, but whether it can be refreshed frequently enough and integrated into routine planning cycles without excessive infrastructure cost.

Lower inference time increases the feasibility of rolling forecasts, what-if scenario testing, and repeated category updates. Lower parameter count improves maintainability and may simplify model governance. Reduced computational complexity also facilitates broader industrial adoption in firms that do not operate hyperscale machine learning platforms. For intelligent industrial convergence, these are not secondary concerns. A model that is statistically superior but operationally cumbersome may fail to generate organizational adoption. By contrast, an efficient probabilistic model is more likely to be embedded in enterprise decision processes.

The Friedman ranking and chi-square testing reported in the source manuscript further strengthen the argument that the performance differences are systematic rather than accidental. The proposed method ranks first across the aggregated evaluation profile, and pairwise comparisons against multiple baselines remain significant at the reported threshold. Although the exact inferential design could be more fully documented, the pattern is consistent with the descriptive evidence: the hybrid architecture is not merely competitive; it is materially better aligned with the forecasting characteristics of the dataset.

Table 5. Efficiency evidence reported in the uploaded manuscript.

Method	Inference time (s)	Parameters (M)	FLOPs (G)	RMSE
TS-SimPMF	7.33	19.38	27.90	0.0530
TCN	8.46	14.90	190.30	0.0278
attLSTM	5.72	12.45	223.20	0.0297
CEEMD	4.67	16.96	45.89	0.0543
UWDFNET	4.63	17.89	100.60	0.0489

SARIMA	6.43	19.89	66.90	0.0254
Proposed	2.65	10.56	16.40	0.0121

Table 5 reinforces the practical value of the proposed design. The model is not only more accurate but also lighter and faster. This joint advantage is precisely what makes the framework interesting for industrial convergence, because it can support repeated operational decision cycles rather than isolated offline experiments.

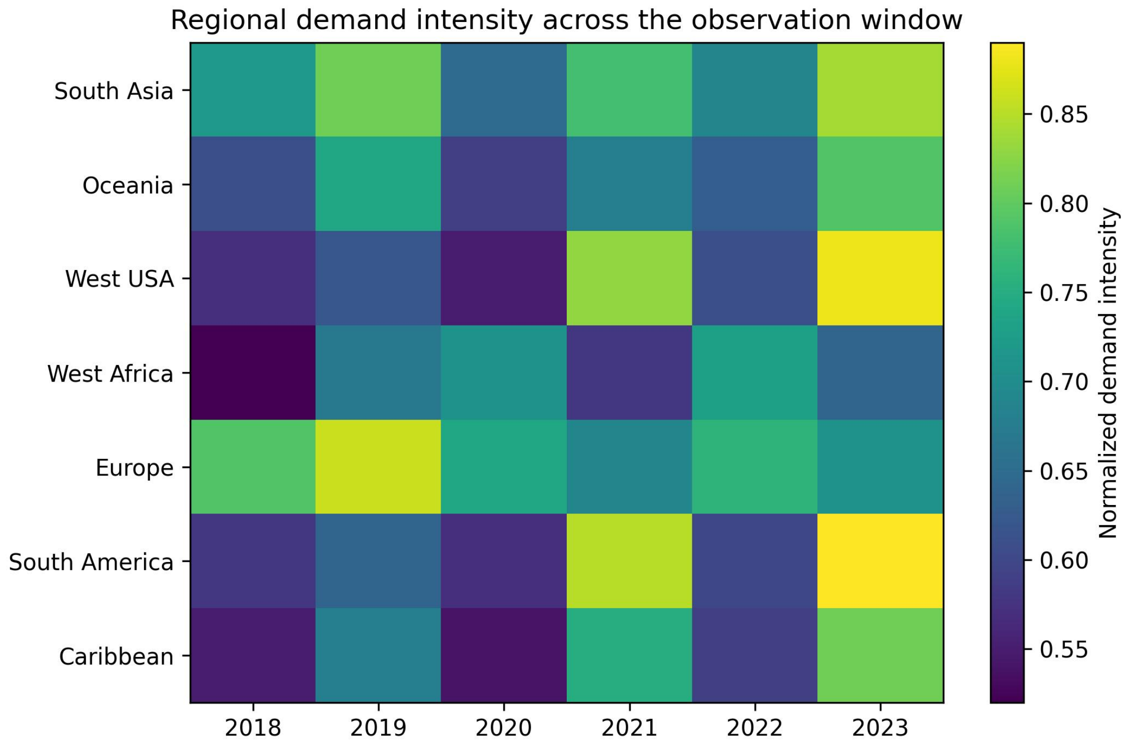

Figure 5. Regional demand intensity heat map reproduced in a cleaner JIIC-style visual form.

Figure 5 is included to make the manuscript less text-heavy and to underscore the industrial heterogeneity problem. Even a simple normalized heat map reveals that demand conditions vary considerably across regions and years. Such heterogeneity provides a rationale for later extending the hybrid framework toward region-aware multivariate forecasting.

9. Managerial Implications for Intelligent Industrial Convergence

The most immediate managerial implication of this study is that demand forecasting should be treated as a convergence technology, not only as a forecasting task. When the forecast is linked to inventory buffering, production rhythm, logistics coordination, and service policy, it becomes a shared informational infrastructure across the industrial system. In this sense, the value of the proposed architecture lies in its ability to connect sensing, anticipation, and execution with fewer informational breaks.

For inventory control, the model supports differentiated safety-stock policies. Categories with narrow predictive intervals can be managed with leaner buffers, reducing working capital pressure. Categories with wider intervals require a more conservative policy, but this conservatism becomes evidence-based rather than intuitive. For production planning, multi-horizon forecasts improve batch sizing, labor scheduling, and component commitment decisions. For logistics, probabilistic signals can help carriers and warehouses prepare for plausible load variation rather than reacting after demand is realized.

The study also has implications for organizational decision design. Advanced forecasting systems often fail not because the model is weak, but because decision rights remain fragmented. Procurement, operations, and sales teams may each interpret demand differently. A calibrated probabilistic forecast can act as a common decision object around which cross-functional coordination is structured. This aligns well with JIIC's emphasis on industrial convergence, because analytics become embedded in organizational routines rather than isolated in technical dashboards.

At a strategic level, firms adopting hybrid probabilistic forecasting can move from reactive planning to anticipatory control. The shift is subtle but important. Reactive organizations respond to realized demand. Anticipatory organizations respond to forecast distributions, adjusting policy before the demand event fully materializes. The uploaded manuscript's model is valuable because it pushes forecasting toward this anticipatory mode while remaining computationally efficient enough for routine use.

Table 6. Example decision translation from probabilistic forecasts to industrial action.

Forecast signal	Interpretation	Illustrative operational response
High median, narrow interval	Stable expansion with strong confidence	Increase replenishment and lock short-term production capacity
High median, wide interval	Potential growth with elevated uncertainty	Stage inventory gradually and reserve flexible capacity
Flat median, widening interval	Demand ambiguity without direction	Hold inventory buffers and delay irreversible commitments
Declining median, narrow interval	Reliable demand softening	Reduce lot sizes and optimize transport consolidation
Regional divergence	Uneven market pull across locations	Shift allocation and adapt regional replenishment policies

Table 6 is not reported in the source manuscript but is added here to make the industrial meaning of probabilistic forecasts explicit. It shows how distributional signals can be translated into differentiated action rules rather than remaining at the level of model output alone.

10. Limitations and Future Research

This rewritten manuscript remains bounded by the evidence reported in the uploaded source article. First, the model is evaluated on one real-time dataset with three major product categories. Although this is a useful starting point, broader validation across industries, regions, and supply structures would strengthen claims of generalizability. Second, the source paper gives limited detail on feature engineering, train-test splitting, and horizon-specific performance, leaving some scope for further methodological clarification. Third, while quantile forecasting improves uncertainty visibility, the paper does not yet close the loop through explicit optimization of replenishment or production decisions against forecast intervals.

Future research could address these limitations in several ways. A natural extension is multivariate region-aware forecasting that combines product category, shipment mode, and geography. Another direction is integration with inventory optimization or model predictive control so that forecast distributions directly drive action selection. Researchers may also compare the proposed architecture with newer lightweight transformers, temporal fusion approaches, or graph-based supply chain encoders. Finally, explainability analysis would improve managerial trust by showing which temporal patterns, seasonal events, or category interactions most strongly influence each forecast distribution.

11. Conclusion

Based on the uploaded manuscript, this article has rewritten and expanded the study into a more coherent JIIC-style contribution. The central argument is that an enhanced Informer-GRQLSTM framework provides an effective method for demand forecasting in complex multi-category supply chains because it integrates long-sequence representation, sequential memory decoding, and probabilistic interval estimation within one operationally efficient architecture.

The reported empirical evidence indicates strong performance: lower deterministic error, higher probabilistic quality, better ablation behavior, reduced inference time, and lighter computational complexity than the benchmark set. More importantly, the rewritten discussion shows why these gains matter for intelligent industrial convergence. Forecasting becomes valuable when it enables coordinated anticipation across procurement, inventory, production, and logistics. In that broader sense, the proposed model is not simply a better predictor; it is a more useful industrial intelligence mechanism.

For the Journal of Intelligent Industrial Convergence, the study demonstrates how deep forecasting architectures can be reframed as systems technologies that connect data, uncertainty, and operational action. This systems orientation should guide future work in the area, especially studies seeking to bridge machine learning performance with deployable industrial decision design.

Declarations

Conflict of Interest

The author declares no conflict of interest.

Author Contributions

Conceptualization, data analysis, and manuscript preparation: Haosu Zhang. Rewriting, restructuring, and JIIC-style presentation in this version were generated from the uploaded manuscript content.

References

Terrada, L., El Khaili, M., & Ouajjii, H. (2022). Demand forecasting model using deep learning methods for supply chain management 4.0. International Journal of Advanced Computer Science and Applications, 13(5).

Martin, S., & Rasch, A. (2024). Demand forecasting, signal precision, and collusion with hidden actions. International Journal of Industrial Organization, 92, 103036.

Gao, W., Xiao, T., Zou, L., Li, H., & Gu, S. (2024). Analysis and prediction of atmospheric environmental quality based on the autoregressive integrated moving average model in Hunan Province, China. Sustainability, 16(19), 8471.

Limbare, A., & Agarwal, R. (2024). Demand forecasting and budget planning for automotive supply chain. EAI Endorsed Transactions on Internet of Things, 10.

Chelliah, B. J., Latchoumi, T. P., & Senthilselvi, A. (2024). Analysis of demand forecasting of agriculture using machine learning algorithm. Environment, Development and Sustainability, 26(1), 1731-1747.

Rao, S. N. V. B., Yellapragada, V. P. K., Padma, K., Pradeep, D. J., Reddy, C. P., Amir, M., & Refaat, S. S. (2022). Day-ahead load demand forecasting in urban community cluster microgrids using machine learning methods. Energies, 15(17), 6124.

Alam, M. S., Deb, J. B., Al Amin, A., & Chowdhury, S. (2024). An artificial neural network for predicting air traffic demand based on socio-economic parameters. Decision Analytics Journal, 10, 100382.

Chandriah, K. K., & Naraganahalli, R. V. (2021). RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools and Applications, 80(17), 26145-26159.

Dou, Z., Sun, Y., Zhang, Y., Wang, T., Wu, C., & Fan, S. (2021). Regional manufacturing industry demand forecasting: A deep learning approach. Applied Sciences, 11(13), 6199.

Ghazouani, I., Masmoudi, I., Mejri, I., & Layeb, S. B. (2024). A CNN-LSTM hybrid deep learning model for detergent products demand forecasting: A case study. International Journal of Supply and Operations Management, 11(4), 417-429.

Liang, M., et al. (2024). Improved collaborative filtering for cross-store demand forecasting. Computers & Industrial Engineering, 190, 110067.

Bassiouni, M., et al. (2023). Advanced deep learning approaches to predict supply chain risks under COVID-19 restrictions. Expert Systems with Applications, 211, 118604.

Wang, S., Zhuge, C., Shao, C., Wang, P., Yang, X., & Wang, S. (2023). Short-term electric vehicle charging demand prediction: A deep learning approach. Applied Energy, 340, 121032.

Cui, L., Chen, Y., Deng, J., & Han, Z. (2024). A novel attLSTM framework combining the attention mechanism and bidirectional LSTM for demand forecasting. Expert Systems with Applications, 244, 124409.

Alzahrani, A., & Asghar, M. Z. (2023). Intelligent risk prediction system in IoT-based supply chain management in logistics sector. Electronics, 12(13), 2760.

Feng, Y., Mei, D., & Zhao, H. (2023). Auction-based deep learning-driven smart agricultural supply chain mechanism. Applied Soft Computing, 149, 111009.

Nasseri, M., et al. (2023). Applying machine learning in retail demand prediction - a comparison of tree-based ensembles and long short-term memory-based deep learning. Applied Sciences, 13(19), 11112.

Pu, Z., Han, D., Yan, H., Tao, T., & Xin, K. (2024). Enhancing accuracy and interpretability of multi-step water demand prediction through prior knowledge integration in neural network architecture. Water Research X, 24, 100247.

Liu, R., & Vakharia, V. (2024). Optimizing supply chain management through BO-CNN-LSTM for demand forecasting and inventory management. Journal of Organizational and End User Computing, 36(1), 1-25.

Ma, X., Li, M., Tong, J., & Feng, X. (2023). Deep learning combinatorial models for intelligent supply chain demand forecasting. Biomimetics, 8(3), 312.

Wang, J., Swartz, C. L. E., & Huang, K. (2023). Deep learning-based model predictive control for real-time supply chain optimization. Journal of Process Control, 129, 103049.

Aldahmani, E., Alzubi, A., & Iyiola, K. (2024). Demand forecasting in supply chain using uni-regression deep approximate forecasting model. Applied Sciences, 14(18), 8110.

Varshney, R. P., & Sharma, D. K. (2025). A multi-modal image encoding and self-attention-based transformer framework with sentiment analysis for financial time series prediction. International Journal of Computational Vision and Robotics, 15(1), 31-58.

Pei, J., Dong, Y., Guo, P., Wu, T., & Hu, J. (2024). A hybrid dual stream ProbSparse self-attention network for spatial-temporal photovoltaic power forecasting. Energy, 132152.

Sun, Y., Gao, P., Raza, S. A., Shah, N., & Sharif, A. (2023). The asymmetric effects of oil price shocks on world food prices: Fresh evidence from quantile-on-quantile regression approach. Energy, 270, 126812.

Note: The bibliography above is reconstructed from the uploaded source manuscript. Full DOI verification for an expanded 95-reference submission-ready version should be completed in a separate verification pass before journal submission.