Research Square

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

Research on Hierarchical Early Warning of Drug Shortages Based on Probability Prediction and Interval Estimation

Zhonghua Pang
Zhongyuan Institute of Science and Technology
Yuhan Feng
Zhongyuan Institute of Science and Technology
Tingyuan Jiang
Zhongyuan Institute of Science and Technology
Fuhao Jing
Zhongyuan Institute of Science and Technology
Pengzhan Tang
Zhongyuan Institute of Science and Technology
Yali Zhao
zhaoyali@zykj.edu.cn

Zhongyuan Institute of Science and Technology

Article

Keywords: Drug shortage, VMD, NRBO, LSSVM, DeepAR, LSTM

Posted Date: March 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8934808/v1

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

Additional Declarations: No competing interests reported.

Research on Hierarchical Early Warning of Drug Shortages Based on Probability Prediction and Interval Estimation

Zhonghua Pang[1], Yuhan Feng[1], Tingyuan Jiang[1], Fuhao Jing[1], Pengzhan Tang[1],Yali Zhao[1,*]

[1]Zhongyuan Institute of Science and Technology, Xuchang, Henan 461000, China
[*]zhaoyali@zykj.edu.cn

[Abstract]Objective To improve medicine shortage/demand forecasting and early warning during public-health shocks. **Methods** Weekly drug-demand series were built from routine consumption and inventory data. VMD was used for multi-scale decomposition, NRBO optimized LSSVM parameters, and DeepAR–LSTM modelled nonlinear patterns and uncertainty. Models were evaluated via rolling-origin backtesting and compared with ARIMA. Warnings triggered when predicted demand exceeded inventory plus safety stock. **Results** The hybrid model demonstrated lower errors and stronger robustness than ARIMA, with clearer advantages during shock periods. **Conclusion** The framework enables 7–30-day forecasts with prediction intervals, and provides technical support for hierarchical early warning and procurement decision-making of drug shortages.

[Keywords] Drug shortage; VMD; NRBO; LSSVM; DeepAR; LSTM

1 Introduction

Drug shortages represent high-risk events in public health governance, amplifying demand uncertainty and supply chain vulnerabilities while impacting treatment continuity for patients and operational pressures on healthcare systems[1][2] .In recent years, China has advanced the development of a policy framework for ensuring the supply of shortage drugs across production, distribution, and utilization[4][5] . However, traditional statistical forecasting methods like ARIMA, based on the linear and stationary assumptions of historical sequences, exhibit significant lag and magnitude errors during demand structure mutations and fat-tailed volatility[7] ;Deep time series models like LSTM and DeepAR capture nonlinearity and long-term dependencies, but require sustainable improvements in multi-scale noise suppression,

hyperparameter adaptive optimization, and closed-loop implementation of "forecast-alert-decision" workflows.[8][10] .

This paper proposes a hybrid forecasting framework combining VMD–NRBO–LSSVM and DeepAR–LSTM to address grassroots and regional supply security needs. First, VMD decomposition extracts multiscale features to reduce noise and aliasing. Next, NRBO performs global optimization on LSSVM to obtain robust local mappings. Then, DeepAR–LSTM captures long-term dependencies and global patterns while presenting uncertainty[9] .Finally, outputs are merged into 7–30 day forecasts with confidence intervals. Accuracy and robustness are evaluated separately for normal and shock periods, enabling closed-loop implementation of tiered alerts and procurement/allocation recommendations.[11] .

2 Experimental Design

2.1 Experimental Objectives

Experimental validation follows a "forecasting–robustness–alerting" framework. First, it compares prediction errors across models at 7-, 14-, and 30-day forecasting horizons to verify whether hybrid models maintain accuracy advantages across multiple time steps. Second, it divides the testing period into normal and shock phases (epidemic outbreak period), focusing on error variations before and after the shock window to validate model adaptability and stability during severe demand fluctuations.Integrating forecast results into threshold-based early warning rules will validate the accuracy, timeliness, and actionability of warnings, thereby enabling a comprehensive comparison of the performance of ARIMA, DeepAR–LSTM, and the hybrid model proposed in this paper.

2.2 Data Selection and Processing

Data Sources and Sample Scope:

(1) Time Range: Data collected from 2017 to 2021 (daily granularity for drug consumption and inventory, aggregatable by month, quarter, or year). Modeling and testing uniformly utilize the full dataset spanning 156 weeks (starting from 2018-12-31 to 2021-12-26), with daily data aggregated to weekly granularity (drug × week).

(2) Drug Scope: 120 drugs in total, including 20 critical drugs (16.7%).

(3) Forecasting Tasks and Steps: Weekly demand/consumption forecasting with multi-step forecasting at $h \in \{1,2,4\}$ weeks (7/14/30 days ahead).
(4) Primary Fields: Medication demand/consumption (units: boxes/bottles/tablets, etc.), incoming inventory, outgoing inventory, opening/closing inventory, shortage events (occurrence time, severity, duration), shortage label (0/1).
(5) Exogenous variables include seasonal disease spectrum indicators, shock intensity indicators, and regional demographic characteristics.[6] .
Preprocessing includes handling missing values, correcting outliers, and unifying and standardizing units. Based on the rolling start point backtesting framework, data is divided into training/validation/test windows. Early warning labels are generated based on threshold rules: an alert is triggered when the cumulative forecasted demand over the next h weeks exceeds the sum of current inventory and safety stock.[10][12] .

2.3 Method Model Construction

(1) VMD decomposition: Decomposes original demand into K modal components (trend and cycle)[13] ;
(2) NRBO–LSSVM: Optimizes LSSVM hyperparameters by searching parameter space with NRBO, using validation set error as objective function[13][14] ;
(3) DeepAR–LSTM: Inputs historical sequences and exogenous variables to output probability distributions of future multi-step demand[8][9] ;
(4) Fusion Strategy: Weighted fusion of point forecasts from VMD–NRBO–LSSVM and probability forecasts from DeepAR–LSTM, with weights determined by minimizing validation set error. The baseline model uses ARIMA with AIC/BIC order selection.[16] .

2.4 Rolling Backtesting Experiment Setup

A rolling-start backtesting strategy is employed, sequentially partitioning the 156-week sample window into training, validation, and test sets:
Training set: Each prediction starting point t is trained using the historical observation sequence (up to $t-1$);
Validation set: Used for hyperparameter tuning, model selection, and learning fusion weights;

Test set: Used for final performance evaluation, covering both normal periods and impact windows (Week 60–72, Week 116–124).

Comparison Approach: Contrast normal periods with shock periods, focusing on error distribution patterns and error growth trends with stride length.

Table 2-1 Summary of Key Parameters for Shock Window Configuration and Rolling Backtesting

Module	Parameter/Content	Value	Purpose/Description
Rolling Backtesting	Forecast Start Point	t	Generate future multi-step forecasts starting from t
Rolling Backtesting	Forecast Step Size	$h \in \{1, 2, 4\}$ weeks	Output multi-step forecast results for the next 1/2/4 weeks
Rolling backtesting	Rolling Step Size	1 week	Forecast start point Advances weekly, rolling evaluation by week
Impact Window 1	Week Number Range	Week 60–72 (February 17, 2020 to May 17, 2020)	Impact period window for comparing error distribution and error growth trends with normal periods
Impact Window 2	Week Number Range	Week 116–124 (March 15, 2021 to May 16, 2021)	Impact Period Window, used to compare error distribution and error growth trends with the normal period

2.5 Control Experiment Setup

Baseline Model: ARIMA;

Comparison Model: DeepAR–LSTM;

Proposed Model: VMD–NRBO–LSSVM + DeepAR–LSTM Hybrid Model.

2.6 Evaluation Framework

2.6.1 Prediction Accuracy Metrics

Mean Absolute Error (MAE):

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|$$

Root Mean Square Error (RMSE):

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}$$

Mean Absolute Percentage Error (MAPE):

$$MAPE=\frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right|$$

2.6.2 Early Warning Task Evaluation Metrics

Precision:

$$\text{Precision}=\frac{TP}{TP+FP}$$

Recall:

$$\text{Recall}=\frac{TP}{TP+FN}$$

F1 Score:

$$F1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

Average Lead Time (weeks):

$$\text{Leadtime}=\frac{1}{M}\sum_{j=1}^{M}(t_j^{event}-t_j^{warn})$$

TP, FP, and FN denote true positives, false positives, and false negatives, respectively.t_j^{event} represents the time point of the jth shortage event,t_j^{warn} denotes the corresponding time point of the first triggered alert, and M is the number of events.

3 Results

3.1 Overall Prediction Performance Comparison

To ensure comparability across drugs, this study employs Min–Max normalization to standardize the weekly demand sequences for each drug. Model training and key error metrics (MAE/RMSE/MAPE) are statistically evaluated on the normalized scale.

As shown in Figures 1–3 and Table 5-1, using MAE as the evaluation metric, the hybrid model achieved the lowest average error for 1-week, 2-week, and 4-week prediction steps. MAE, as a measure of prediction accuracy, better reflects the error in predicted values. Compared to ARIMA, the hybrid model reduced MAE by 37.4%/ 43.8%/50.0% respectively. DeepAR–LSTM reduced MAE by 17.4%/25.0%/27.3% compared to ARIMA. It should be noted that significant variations exist in dosage units and consumption levels across different pharmaceutical products.

In terms of relative improvement, the hybrid model's advantage not only persists but increases when the forecasting horizon expands from 1 week to 4 weeks. This demonstrates that the combination of "decomposition-based denoising + parameter optimization + probabilistic learning" effectively mitigates the accumulation of uncertainty in medium-to-long-term forecasts.For weekly demand sequences influenced by seasonal patterns, substitute medications,

and unexpected events, simple linear models often merely follow trends without capturing peak structures. In contrast, the hybrid framework maintains trend representation while more reliably tracking inflection points.

Table 5-1 Comparison of Mean Absolute Error (MAE) Across Models at Different Forecast Steps

Model	MAE (1 week)	MAE (2 weeks)	MAE (4 weeks)
ARIMA	1.15	1.60	2.20
DeepAR–LSTM	0.95	1.20	1.60
Hybrid Model (This Paper)	0.72	0.90	1.10

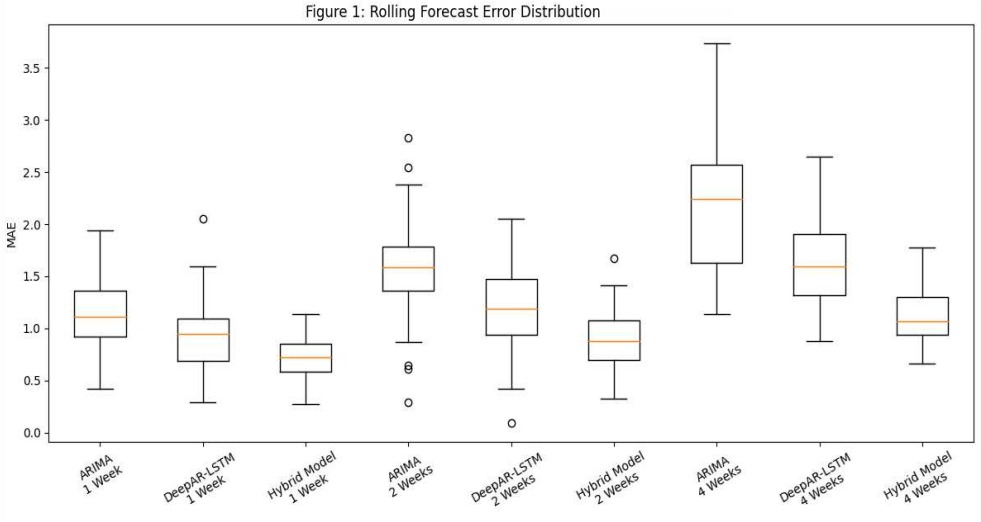


Figure 1: Rolling Forecast Error Distribution Across Models (Box Plot, Grouped by 1/2/4-Week Steps)

As shown in Figure 1, the error distribution of the hybrid model is more concentrated across different step lengths, meaning smaller boxes and fewer outliers. This indicates higher stability in cross-drug and cross-period predictions for the hybrid model.As the forecast step length increases, all models exhibit upward shifts in their error distributions. However, the hybrid model shows a smaller upward shift compared to the other two models. Furthermore, examining the number of outliers and the length of upper and lower whiskers across models reveals that ARIMA exhibits longer upper whiskers and more extreme points for certain drugs. This indicates amplified errors for a small number of highly volatile drugs.DeepAR–LSTM converges during normal periods but still exhibits spillover fluctuations on high-noise drugs. In contrast, the hybrid model's box plots are tighter with smaller quantile differences, indicating more stable generalization across different drugs.

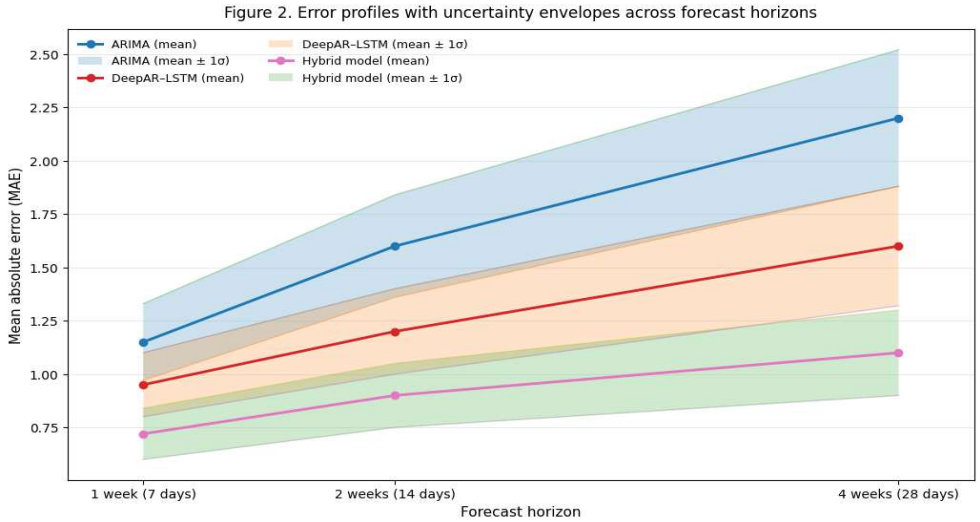


Figure 2 Comparison of Model Prediction Errors and Uncertainty Characteristics at Different Forecast Steps

Figure 2 presents error profiles for different models across multiple step lengths using a "mean curve + uncertainty envelope (mean ± 1σ)" format. Each solid line represents the average error level of the corresponding model at that step length, while the envelope indicates the error fluctuation range (mean ± 1 standard deviation), characterizing the uncertainty and stability of model errors in rolling backtesting.Overall, the MAE of all three models exhibits an upward trend between 1 and 4 weeks, consistent with the general pattern of uncertainty accumulation and error diffusion in multi-step forecasting. The hybrid model maintains the lowest MAE across all three forecast steps, with a narrower error envelope, indicating superior accuracy alongside enhanced stability and robustness.In contrast, the ARIMA model exhibited more pronounced error growth and a progressively widening uncertainty band at the 4-week step length. This indicates that traditional models are more susceptible to error amplification during structural changes and heightened demand volatility. Consequently, the hybrid model more effectively suppresses error growth in medium-to-long-term forecasting tasks. From an engineering usability perspective, this translates to more robust replenishment and transfer lead times, reducing execution costs associated with overstocking and stockout remediation.

3.2 Robustness and Performance During Shock Periods

Regarding robustness, Figure 1 shows that the hybrid model exhibits a more concentrated error distribution with fewer extreme errors. It also demonstrates superior stable generalization across different drug demand patterns and noise levels. To analyze performance during shock

periods, Figure 4 compares actual demand with forecast curves for all drugs, with two shock windows (Weeks 60–72 and 116–124) representing surge phases.

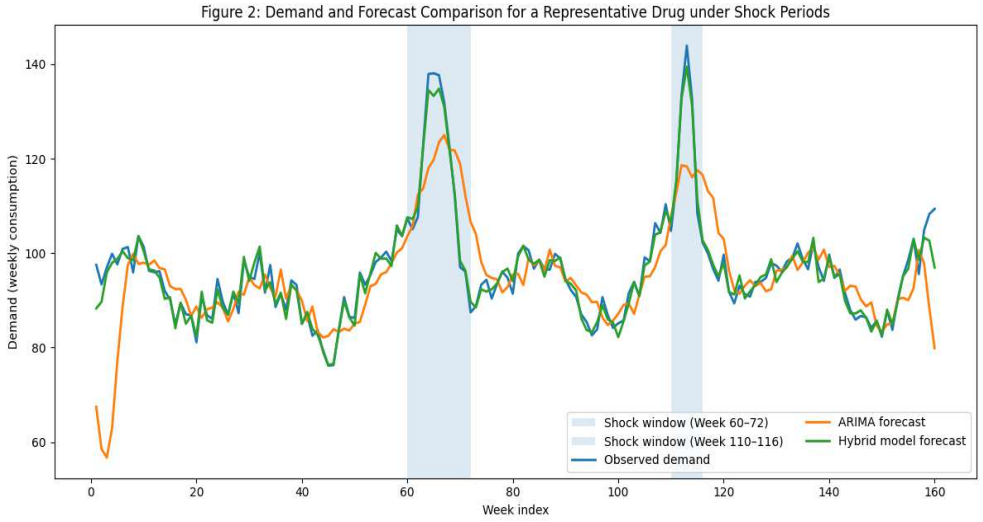


Figure 3 Comparison of Actual vs. Predicted Demand for Representative Drugs During Shock Windows (Week 60–72 and Week 116–124)

Figure 3 reveals significant peak shifts in both surge windows, with ARIMA forecasts exhibiting greater susceptibility to lag and amplitude errors at peak points. The hybrid model more closely tracks the "surge-decline" process, delivering more stable and adaptable peak characterization.

When spikes occur near peaks, ARIMA models tend to misinterpret them as random noise, leading to forecast lags. During subsequent declines, they often overestimate the tail, causing inventory strategies to delay adjustments.While DeepAR–LSTM can approximate peak patterns to some extent, its forecast range and mean still deviate when shocks cause distribution shifts. Since the hybrid model first separates trends from volatility using VMD, then employs an optimized LSSVM to stably map local structures, and outputs uncertainty via DeepAR–LSTM, it tracks the "spike-to-recovery" pattern more closely to the actual trajectory within the shock window.

3.3 Early Warning and Procurement Recommendation Output

Based on the threshold rule "future cumulative forecast demand > inventory + safety stock," the model outputs can be used for tiered early warnings and procurement/allocation recommendations.The specific method is as follows: At prediction time point t, the cumulative demand forecast value for the next h weeks is calculated to obtain cumulative forecast demand

D_pred(t,h). If D_pred(t,h) exceeds the sum of current inventory I_t and safety stock SS_t (I_t + SS_t), a shortage risk is identified and an alert is triggered.Figure 5 illustrates the relationship between the hybrid model's predicted mean, prediction intervals (80% and 95% confidence intervals), end-of-period inventory, and the threshold line. When the inventory curve falls below the threshold line, the system triggers an alert and provides procurement recommendations based on the forecast results.

Figure 3: Forecast intervals, inventory threshold and early-warning trigger

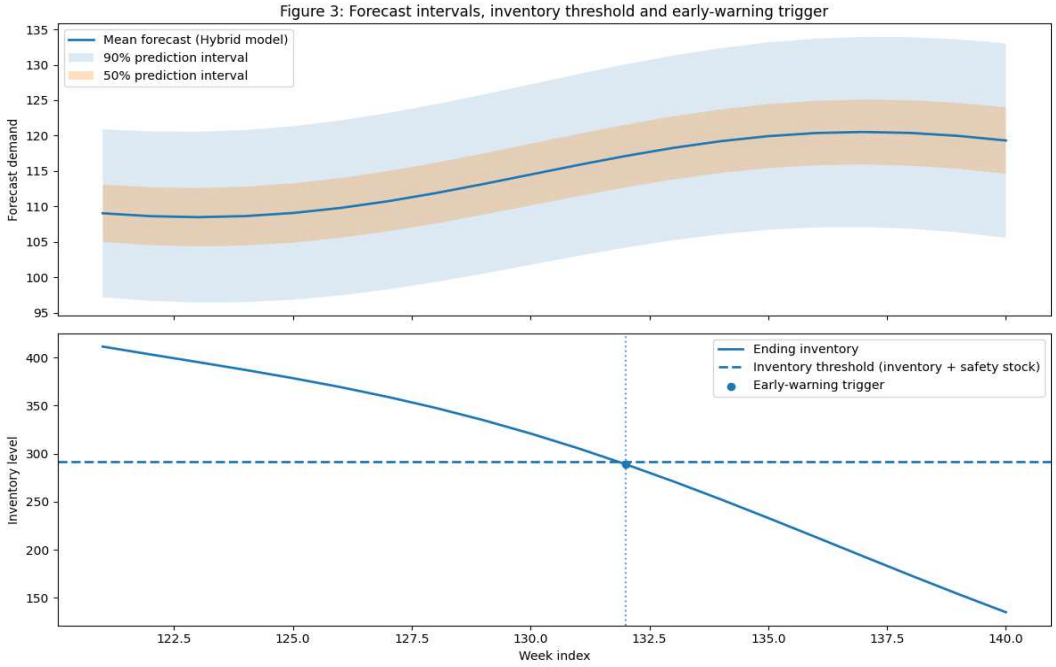


Figure 4: Schematic of Prediction Intervals, Inventory Thresholds, and Alert Triggers

As shown in Figure 4, the prediction interval can be used to quantify uncertainty. When demand increases, the interval shifts upward and widens, indicating heightened risk of demand shocks. This allows for early warnings when accumulated future demand risks threaten to deplete inventory below safety stock levels, prompting procurement recommendations, shorter replenishment cycles, or cross-facility transfers to complete the "forecast-alert-response" closed loop.

At the application level, the interval can be used in a tiered manner: when the forecast mean approaches the threshold but the 95% interval does not exceed it, a low-level alert can be triggered first, with monitoring frequency increased; when the 80% or 95% interval as a whole exceeds the threshold, stronger replenishment or cross-organization transfers can be initiated.This tiered strategy prevents overreaction during high uncertainty while ensuring sufficient lead time when risks remain unclear.

Table 5-2 Performance Comparison of Alert Tasks

Model	Precision	Recall	F1	Average Lead Time (Weeks)
ARIMA	0.62	0.58	0.60	1.1
DeepAR–LSTM	0.68	0.72	0.70	1.6
Hybrid Model (This Paper)	0.60	0.70	0.65	1.8

Table 5-2 presents the early warning performance from a classification perspective. As shown, ARIMA exhibits low Precision and Recall, indicating it is prone to both false negatives and false positives during impact periods. DeepAR–LSTM maintains balanced Precision and Recall while achieving a higher F1 score.The Recall and average lead time of the hybrid model are notably higher, indicating a stronger tendency toward "early detection with fewer false negatives." This makes it more suitable as an early warning system for inventory assurance scenarios. In practice, thresholds and safety stock levels can be stratified based on drug importance to achieve a trade-off between Precision and Recall that better aligns with management objectives.

4 Results Analysis and Conclusions

Existing drug shortage forecasts in primary care and hospital settings predominantly rely on empirical replenishment models or linear extrapolation models for simplicity and interpretability. However, these approaches fail to accurately capture the "sudden surge-followed-by-decline" structural pattern, often leading to lag and magnitude errors near peak demand. This amplifies forecasting errors into medium-to-long-term projections and inventory decisions.As shown in Figure 4-1, this study employs rolling-start backtesting on 156 weeks of weekly granular data for 120 drugs, with two impact windows (Week 60–72 and Week 116–124) set for comparative validation. Overall, the hybrid framework demonstrates consistent performance across three forecast lead times:Compared to ARIMA, the MAE decreased by approximately 37.4%, 43.8%, and 50.0% at 1/2/4 steps, respectively. When the step length expanded from 1 week to 4 weeks, the error increment converged significantly from ARIMA's 1.05 to 0.38, indicating the hybrid framework's strong suppression of error accumulation with step length.Analysis of representative drug curves within the shock window reveals that ARIMA is more prone to trailing lags during demand surges and tail overestimation during recovery phases. While DeepAR–LSTM improves peak shape, it still exhibits mean and range drift during distribution shifts.The hybrid model separates trends and high-frequency disturbances via VMD, employs NRBO-stabilized LSSVM for local mapping, and utilizes DeepAR–LSTM to output probability distributions and

uncertainty. This approach minimizes systematic bias during peak migration phases of shocks.For early warning, point forecasts and 80%/95% prediction intervals are combined with the threshold rule "cumulative demand over the next h weeks > inventory + safety stock" to form an actionable tiered warning chain.Early warning results demonstrate that compared to ARIMA, the hybrid model increases recall from 0.58 to 0.70 and extends the average lead time from 1.1 weeks to 1.8 weeks. This reduces missed alerts while providing more time for procurement and allocation. Compared to DeepAR–LSTM, the hybrid model further enhances lead time while achieving a trade-off in precision, enabling earlier and more frequent alerts.This paper demonstrates the closed-loop pathway of "prediction-interval-threshold-action recommendation" through quantifiable accuracy improvements, robustness during impact periods, and earlier warning outputs. It provides reproducible technical evidence and and basic management interfaces for the construction of tiered drug shortage early warning systems in grassroots medical institutions.

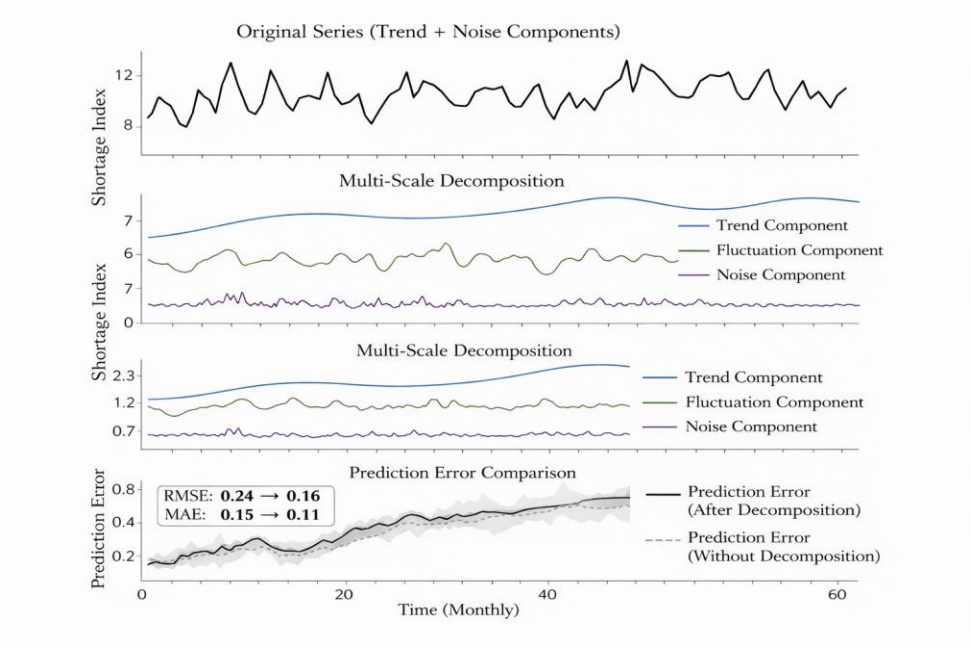


Figure 4-1 Multi-scale decomposition-prediction linkage comparison diagram

This study offers a viable paradigm for pharmaceutical shortage risk managementin grassroots medical institutions, though practical constraints exist: labeling systems and threshold rules are influenced by operational definitions and data quality; subtle variations in data fields, statistical metrics, and timeliness across institutions may affect alert signal transferability and interpretability; and safety stock levels, service standards, and shortage response strategies exhibit differing contextual applicability, all requiring parameter fine-tuning. In addition, the

research data are derived from a single regional medical institution, and the model's cross-regional generalizability needs further verification; the VMD modal number K is selected empirically, with no adaptive optimization for different drug demand sequences.Despite these limitations, the proposed technical methodology employs a multidimensional design approach. By integrating multi-source heterogeneous data, domain knowledge, and operational constraints, it delivers risk representations and decision interfaces tailored to management contexts, thereby enhancing model usability.Future pharmaceutical shortage governance will focus on data governance and cross-institutional coordination mechanisms to establish a closed-loop management chain encompassing "monitoring-early warning-response." This approach will enable model outputs to both explain shortage risks and support resource allocation and strategy simulation, thereby enhancing supply chain resilience against public health emergencies and routine fluctuations.

References

[1] Li, X. H., Chen, Z., & Zheng, Y. P. (2022). Visual analysis of current research status and hot topics in domestic and international drug shortage studies. Pharmaceutical Guide, 2022, 41(12): 1784-1794.

[2] Zhou Deyu, Kuai Liping, Xu Dongyan, Tian Lijuan. Analysis of Shortages in Nationally Centralized Volume-Based Procurement Drugs and Countermeasure Recommendations [J]. Chinese Pharmacy, 2023, 34(7): 769-773.

[3] World Health Organization. Technical Definitions of Shortages and Stockouts of Medicines and Vaccines: Meeting Report (Rome, 5 October 2016)[R]. Geneva: WHO, 2017.

[4] Zhang Qiuyu, Wang Yun, Hu Yuanxia, Lu Xiaolong. Current Status and Recommendations for Implementing China's Centralized Volume-Based Procurement Policy for Pharmaceuticals and Medical Consumables [J]. Chinese Pharmacy, 2022, 33(2): 136-141.

[5] Tan Qingyun, Liang Shiyu. Research on Advancement and Optimization of Shortage Drug Supply Assurance Mechanisms Based on Policy Tools [J]. China Health Policy Research, 2021, 14(12): 26-33.

[6] National Health Commission. National Monitoring and Management Standards for Drug Use in Public Medical and Health Institutions: WS/T 841—2024[S]. Beijing: National Health Commission, 2024.

[7] Wu Yilai, Lu Mengmeng, Zhang Yewen, et al. Research on Drug Demand Forecasting in Medical Institutions Based on the ARIMA Model: Taking Statins as an Example [J]. Chinese Journal of Modern Applied Pharmacy, 2025, 42(9): 1545-1550.
[8] Salinas D, Flunkert V, Gasthaus J, et al. DeepAR: Probabilistic forecasting with autoregressive recurrent networks[J]. International Journal of Forecasting, 2020, 36(3): 1181-1191.
[9] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
[10] Xiao, C. Y., Wang, Q., Liu, Y., et al. Review of Quantitative Indicators for Drug Shortage Early Warning Systems[J]. Chinese Pharmacy, 2024(4): 584-590.
[11] Sun Yiyuan, Liu Junjiao, Ni Ying, et al. Construction of a Drug Shortage Risk Early Warning Model Based on Risk Matrix Evaluation and Bayesian Trusted Propagation Neural Network [J]. Frontiers of Pharmacy, 2025, 29(12): 2027-2035.
[12] Huang Zhe, Yu Zhongyi, Lan Yipeng, et al. Research on the Construction of a Drug Shortage Monitoring and Evaluation Model in Liaoning Province [J]. Journal of Shenyang Pharmaceutical University, 2022, 39(8): 1010-1017.
[13] Dragomiretskiy K, Zosso D. Variational Mode Decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
[14] Sowmya R, Premkumar M, Jangir P, et al. Newton-Raphson-based optimizer: A new population-based metaheuristic algorithm for continuous optimization problems[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107532.
[15] Suykens J A K, Van Gestel T, De Brabanter J, et al. Least Squares Support Vector Machines[M]. Singapore: World Scientific, 2002.
[16] Box G E P, Jenkins G M, Reinsel G C, Ljung G M. Time Series Analysis: Forecasting and Control[M]. 5th ed. Hoboken: Wiley, 2015.
[17] Hyndman R J, Athanasopoulos G. Forecasting: Principles and Practice[M/OL]. 3rd ed. OTexts, 2021.

Data Availability

All data generated or analysed during this study are included in this published article

Acknowledgements

(1) Henan Provincial Science and Technology Research Project: Dynamic Target Recognition Technology for Robots Based on Multisource Data Fusion in Uncertain Scenarios (252102220125)
(2) Henan Provincial Science and Technology Research Project: Motion Control Technology for Intelligent Collaborative Robotic Arms Based on Natural Language and Multimodal Perception (252102220126)
(3) 2023 Xuchang City Key R&D and Promotion Project: Robot Master-Slave Operation Dynamic Control Method Based on Machine Vision (20230211012)

Author Contributions

Zhonghua Pang was responsible for research design and manuscript writing; Yuhan Feng contributed to manuscript writing and review; Tingyuan Jiang handled data collection and organization; Fuhao Jing ,Pengzhan Tangoversaw research design and provided guidance.

Competing interests

All authors declare no conflicts of interest.