



Hybrid LSTM-DGWO Framework for Forecasting Insulin Aspart Demand: A Case Study of NovoRapid in the Pharmaceutical Supply Chain

Mohammad Ali Ariyamanesh¹, Ebrahim Niknaghsh^{1,*}, Mehdi Alirezanejad², Amir Gholamabri³

¹ Department of Industrial Management, F.i.c., Islamic Azad University, Firoozkooh, Iran

² Department of Computer Engineering, F.i.c., Islamic Azad University, Firoozkooh, Iran

³ Department of Mathematics, F.i.c., Islamic Azad University, Firoozkooh, Iran

*Corresponding Author: Department of Industrial Management, F.i.c., Islamic Azad University, Firoozkooh, Iran. Email: eniknaghsh@iau.ac.ir

Received: 9 February, 2026; Revised: 29 July, 2026; Accepted: 1 August, 2026

Abstract

Background: Accurate demand forecasting is critical for operational resilience and efficient resource allocation in pharmaceutical supply chains, particularly for life-critical medications such as Insulin Aspart (NovoRapid). However, the demand for these products is often characterized by nonlinearity and structural complexity, which frequently limits the predictive accuracy of conventional statistical methods.

Objectives: This study aims to improve demand forecasting accuracy by developing and evaluating a hybrid long short-term memory (LSTM)-Dynamic Grey Wolf Optimizer (DGWO) framework. The objective is to leverage deep learning for temporal sequence modeling while employing metaheuristic optimization to systematically refine model hyperparameters.

Methods: The proposed framework integrates an LSTM network to capture temporal dependencies and nonlinear demand patterns with a DGWO for hyperparameter tuning. The hybrid model's forecasting performance was empirically evaluated against benchmark approaches, specifically an autoregressive integrated moving average (ARIMA) (1, 1, 1) model and a naive baseline, using root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2) as accuracy indicators.

Results: The empirical results show that the proposed hybrid framework outperforms the benchmark models. The final model achieved an RMSE of 131,400, an MAE of 81,112, a MAPE of 10.59%, and an R^2 of 0.6823. These findings indicate that the hybrid structure, which combines deep sequence learning with intelligent hyperparameter optimization, provides a more robust representation of demand dynamics than do traditional forecasting methods.

Conclusions: Integrating deep learning with metaheuristic optimization provides a robust methodological foundation for pharmaceutical demand forecasting. The LSTM-DGWO framework serves as a practical decision-support tool for supply chain stakeholders, enabling improved inventory planning, optimized safety stock determination, and more accurate supply coordination. Although the model shows substantial promise, its generalizability is constrained by the 32-month dataset and the single-product focus. Future research should scale this framework across broader product portfolios and incorporate exogenous market variables to further strengthen predictive robustness.

Keywords: Demand Forecasting, Pharmaceutical Supply Chain, Insulin Aspart, NovoRapid, LSTM-DGWO, Inventory Management

1. Background

Accurate demand forecasting is a cornerstone of pharmaceutical supply chain management, directly influencing production planning, inventory control, and distribution efficiency (1). In this sector, forecasting errors can have serious consequences: shortages may threaten patient health, whereas excess inventory results in financial waste and operational inefficiency. In particular,

insulin analogues such as NovoRapid are critical medicines for which reliable demand forecasting is essential; however, temporal variability, uncertainty, and nonlinear demand behavior make accurate prediction challenging (2, 3). These factors generate complex, nonlinear, and nonstationary demand patterns, rendering reliable forecasting difficult. Despite the importance of pharmaceutical demand forecasting, accurate prediction at the individual product level remains challenging,

Copyright © 2026, Ariyamanesh et al. This open-access article is available under the Creative Commons Attribution 4.0 (CC BY 4.0) International License (<https://creativecommons.org/licenses/by/4.0/>), which allows for unrestricted use, distribution, and reproduction in any medium, provided that the original work is properly cited.

How to Cite: Ariyamanesh MA, Niknaghsh E, Alirezanejad M, Gholamabri A. Hybrid LSTM-DGWO Framework for Forecasting Insulin Aspart Demand: A Case Study of NovoRapid in the Pharmaceutical Supply Chain. Middle East J Rehabil Health Stud. 2026;13(4):e169627. doi: <https://doi.org/10.5812/mejrh-169627>

particularly in real-world environments in which historical datasets are often limited in size (15). Small datasets increase the risk of overfitting and may reduce the effectiveness and generalizability of data-driven forecasting models. Consequently, there is a growing need for forecasting frameworks that can capture nonlinear temporal dependencies while maintaining robust predictive performance under limited-data conditions.

Deep learning models, particularly long short-term memory (LSTM) networks, have demonstrated promising capabilities for modeling complex sequential patterns. However, the forecasting performance of LSTM models is highly dependent on appropriate hyperparameter selection. Inadequate hyperparameter configurations may lead to poor convergence, unstable learning behavior, and reduced prediction accuracy. To address this challenge, metaheuristic optimization algorithms have increasingly been used to automate hyperparameter tuning and improve model performance.

1.1. Literature Review

Demand forecasting has long been recognized as a critical component of supply chain management and operational planning. Traditional statistical forecasting approaches, particularly the Autoregressive Integrated Moving Average (ARIMA) model, have been widely applied because of their simplicity, interpretability, and established theoretical foundation (4). Previous studies have shown that LSTM-based models are well suited for forecasting complex sequential data because of their ability to capture nonlinear and long-term temporal dependencies (5, 6). Insulin analogues such as NovoRapid are widely used in glycemic control and are therefore important products in pharmaceutical supply planning (7, 8).

Despite their advantages, LSTM networks are highly sensitive to hyperparameter selection. Parameters such as the learning rate, number of hidden neurons, batch size, and look-back window length significantly influence forecasting performance. Traditional hyperparameter tuning approaches, including grid search and random search, are often computationally expensive and may be inefficient in complex search spaces (9, 10). As a result, metaheuristic optimization algorithms have increasingly been integrated with deep learning frameworks to improve hyperparameter optimization. Among these methods, the Grey Wolf Optimizer (GWO) has attracted considerable attention because of its effective balance between global

exploration and local exploitation, inspired by the leadership hierarchy and hunting behavior of grey wolves (11, 12). Several studies have reported that improved or developed variants of GWO can enhance neural network training and improve predictive accuracy across different forecasting applications (13). Recent research further suggests that hybrid deep learning-metaheuristic frameworks can significantly improve forecasting accuracy and model robustness (14).

Nevertheless, the application of such hybrid approaches in pharmaceutical demand forecasting remains relatively limited, particularly at the individual product level (15). Furthermore, many previous studies have relied on large datasets, whereas practical pharmaceutical forecasting problems often involve limited historical observations, increasing the risk of overfitting and reducing model reliability. Although traditional statistical models such as ARIMA continue to provide useful forecasting benchmarks, the primary objective of the present study is not to compare multiple forecasting paradigms. Instead, this research focuses on evaluating whether Dynamic Grey Wolf Optimizer (DGWO)-based hyperparameter optimization can enhance the predictive performance of an LSTM forecasting model under limited-data pharmaceutical demand conditions. Therefore, the study compares a standard LSTM model with its DGWO-optimized counterpart to isolate and assess the contribution of the proposed optimization mechanism (Table 1).

2. Objectives

In this study, we propose a hybrid LSTM-DGWO framework for forecasting monthly NovoRapid demand. The framework integrates a DGWO with an LSTM forecasting model to automatically optimize key hyperparameters. In addition, a walk-forward validation strategy is used to preserve temporal consistency and minimize the risk of data leakage during model evaluation. By combining deep learning with metaheuristic optimization, this study aims to provide a robust and practical framework for forecasting pharmaceutical demand in limited-data environments.

3. Methods

3.1. Research Framework

This study proposes a hybrid LSTM-DGWO framework for the monthly forecasting of NovoRapid insulin demand.

Table 1. Summary of the Literature Review

Study	Model/Method	Optimization	Application Domain	Strength	Limitation	Comparison With Proposed LSTM-DGWO
Makridakis et al., 2018 (3)	Statistical vs ML models	–	Forecasting benchmarks	Comprehensive comparison of forecasting techniques	Limited exploration of deep learning optimization	The proposed model integrates deep learning with metaheuristic optimization.
Hochreiter and Schmidhuber, 1997 (5)	LSTM	–	Sequential prediction	Captures long-term temporal dependencies	Sensitive to hyperparameter configuration	DGWO automatically optimizes LSTM parameters to enhance forecasting performance.
Gers et al., 2000 (6)	Improved LSTM	–	Sequence prediction	Improved memory control using forget gates	Computational complexity and parameter sensitivity	Optimization in LSTM-DGWO mitigates parameter sensitivity.
Mirjalili et al., 2014 (9)	Grey Wolf Optimizer	GWO	Optimization problems	Effective swarm-based optimizer	Not applied to demand forecasting models	This study integrates a developed GWO with LSTM for forecasting.
Senapati and Dash, 2018 (11)	Deep neural network	PSO	Energy load forecasting	Metaheuristic tuning improves prediction accuracy	Domain limited to energy forecasting	LSTM-DGWO applies optimization-based deep learning in pharmaceutical demand forecasting.
Wang et al., 2021 (13)	Optimized LSTM	Metaheuristic	Time-series forecasting	Demonstrated improved forecasting accuracy	Limited healthcare/pharmaceutical applications	The proposed model applies optimized LSTM to pharmaceutical demand forecasting.
Zhang and Liu, 2020 (14)	Deep learning	–	Pharmaceutical demand forecasting	Applied deep learning to drug demand prediction	No optimization mechanism for model parameters	LSTM-DGWO introduces optimization to enhance prediction accuracy.
Dogan and Tuncer, 2022 (15)	Deep learning models	–	Healthcare demand forecasting	Effective deep learning forecasting	Lack of hybrid optimization strategies	The proposed hybrid model combines deep learning with swarm optimization.
This study	Hybrid LSTM-DGWO	Dynamic Grey Wolf Optimizer (DGWO)	Pharmaceutical demand forecasting and inventory decision support (NovoRapid case)	Integrates deep temporal learning with metaheuristic optimization for improved forecasting accuracy and operational relevance	Small dataset and limited generalizability to other products or settings	Proposed as an integrated forecasting-optimization framework tailored to pharmaceutical demand patterns.

The LSTM component captures nonlinear temporal dependencies, whereas the DGWO algorithm optimizes its hyperparameters. The workflow includes data preprocessing, feature construction, normalization, sliding-window generation, LSTM forecasting, DGWO-based hyperparameter tuning, walk-forward validation, and performance evaluation. Monthly demand data spanning March 2023 to October 2025, comprising 32 observations, were collected from pharmacy sales reports and official pharmaceutical information systems and aggregated after cleaning. Lagged demand variables and temporal indicators were constructed as predictors. Following an exploratory analysis of the relationships among variables, the overall architecture of the proposed hybrid forecasting model is introduced. The framework integrates data preprocessing, feature engineering, LSTM-based demand prediction, and DGWO-based hyperparameter optimization within a unified forecasting pipeline. The complete forecasting process of the proposed framework is illustrated in Figure 1.

3.2. Dataset Description and Feature Construction

In this study, D_t denotes the observed demand for NovoRapid insulin at time t . Monthly data were collected from pharmacy sales records and pharmaceutical information systems and, after preprocessing, were organized into a chronological time series. To model temporal dependencies, lagged demand variables and seasonal indicators were used as predictors. In addition, feature relevance was assessed using mutual information and Spearman's rank correlation to account for potential nonlinear relationships among the variables.

3.3. Proposed Hybrid LSTM-DGWO Forecasting Model

3.3.1. Problem Formulation and Forecasting Setting

The forecasting problem is formulated as a univariate time-series prediction task. Given a sequence of historical

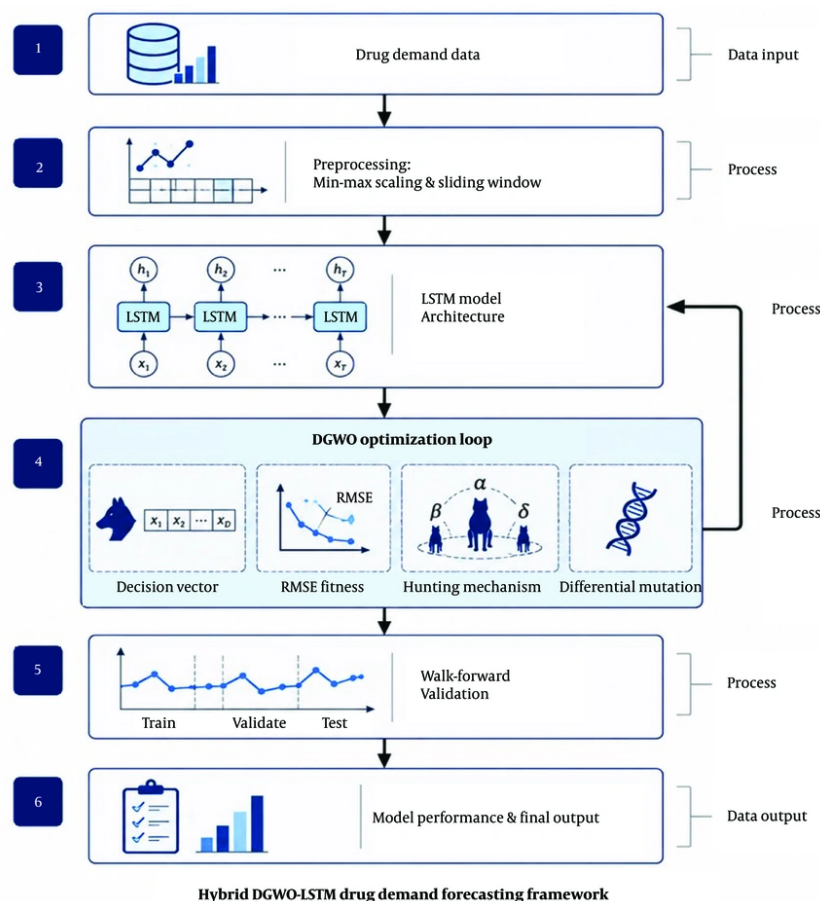


Figure 1. Hybrid LSTM-DGWO Drug Demand Forecasting Framework

demand observations $\{D_1, D_2, \dots, D_t\}$, the objective is to estimate demand at the next time period:

$$\widehat{D}_{t+1}$$

using the information contained in the previous -->FORMULA<-- observations. A sliding-window representation is used to construct input sequences for the forecasting model. To maintain the strictly univariate nature of the proposed framework, only historical lagged observations of NovoRapid demand are used as input features for training and prediction. Other environmental indicators discussed in later sections of this study are treated solely as exploratory variables for analyzing

demand patterns and are not included in the model's feature space.

3.3.2. Data Normalization, Input Construction, and LSTM Cell

3.3.2.1. Data Normalization

All numerical variables were normalized using Min-Max scaling to improve the numerical stability of neural network training. To prevent potential information leakage, the scaling parameters $D_{\min}$ and $D_{\max}$ were estimated exclusively from the training data at each forecasting step and then applied to the validation and test observations. The normalized demand at time -->FORMULA<-- was calculated as:

$$\widetilde{D}_t = \frac{D_t - D_{\min}}{D_{\max} - D_{\min}}$$

where $D_{\min}$ and $D_{\max}$ represent the minimum and maximum demand values from the training data.

3.3.2.2. Sliding-Window Representation

For a look-back window length L , the input vector is defined as:

$$X_t = [\widetilde{D}_{t-L}, \widetilde{D}_{t-L+1}, \dots, \widetilde{D}_{t-1}]^T$$

where $X_t \in \mathbb{R}^L$.

3.3.2.3. LSTM Cell Formulation

The LSTM architecture uses gating mechanisms to regulate the flow of information through the network.

Forget gate:

$$f_t = \sigma(W_f X_t + U_f h_{t-1} + b_f)$$

Input gate:

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i)$$

Candidate state:

$$\widetilde{C}_t = \tanh(W_c X_t + U_c h_{t-1} + b_c)$$

Cell state update:

$$\widetilde{C}_t = \tanh(W_c X_t + U_c h_{t-1} + b_c)$$

Output gate:

$$o_t = \sigma(W_o X_t + U_o h_{t-1} + b_o)$$

Hidden state:

$$h_t = o_t \odot \tanh(C_t)$$

The predicted normalized demand is computed as:

$$\widehat{D}_t = W_y h_t + b_y$$

The forecast on the original scale is obtained through inverse normalization:

$$\widehat{D}_t = \widehat{D}_t (D_{\max} - D_{\min}) + D_{\min}$$

3.3.3. Decision Vector and Search Space

The performance of the LSTM model depends on the selection of several hyperparameters. In this study, these hyperparameters are optimized using DGWO. The decision vector is defined as:

$$x = [L, n_h, lr, e]$$

where:

L = look-back window length

n_h = number of neurons in the hidden layer

lr = learning rate

e = number of training epochs

The search space is defined as:

$$L^{\min} \leq L \leq L^{\max}$$

$$n_h^{\min} \leq n_h \leq n_h^{\max}$$

$$lr^{\min} \leq lr \leq lr^{\max}$$

$$e^{\min} \leq e \leq e^{\max}$$

Because the look-back window, hidden-neuron count, and epoch count are integer variables, whereas the learning rate is continuous, the optimization problem corresponds to a mixed discrete-continuous search space. To mitigate the risk of overfitting—which is common in models with excessive capacity—structural regularization was integrated directly into the search-space design. Specifically, the upper bounds for the look-back window and the number of hidden units were strictly constrained.

This constraint prevented the network from becoming overparameterized and memorizing noise within the limited historical dataset of 32 monthly observations. By restricting these architectural parameters, the DGWO algorithm was forced to search within a low-capacity hypothesis space, thereby promoting better generalization performance.

3.3.4. Time-Series Validity Constraint and Objective Function

To ensure valid temporal forecasting, the following constraint must hold:

$$N > L + 1$$

The optimization objective is to minimize the forecasting error measured by root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N-L} \sum_{t=L+1}^N (D_t - \widehat{D}_t)^2}$$

The RMSE value obtained from walk-forward predictions is used as the fitness function in the DGWO optimization process.

3.3.5. DGWO-Based Hyperparameter Optimization

The DGWO is an enhanced version of the Grey Wolf Optimizer that improves population diversity and reduces the risk of premature convergence. Candidate solutions (wolves) are organized in a hierarchical structure:

α - best solution

β - second-best solution

δ - third-best solution

ω - remaining solutions

The control parameter decreases over iterations:

$$a = 2 \left(1 - \frac{t}{T_{\max}} \right)$$

Coefficient vectors:

$$A = 2ar_1 - a$$

$$C = 2r_2$$

Position update rule:

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3}$$

where X_1 , X_2 , and X_3 denote candidate positions guided by the α , β , and δ wolves. Each wolf represents a candidate LSTM hyperparameter configuration. Model performance is evaluated using RMSE from walk-forward forecasting, and the population is iteratively updated until the maximum number of iterations is reached, yielding the optimal hyperparameter set. During the evaluation of each candidate configuration, training-stage regularization was employed to prevent spurious learning. Models were trained for a maximum of 250 epochs using the Adam optimizer, incorporating weight decay to penalize excessive weight magnitudes. To ensure that the model did not overfit the training partition, an early stopping mechanism was implemented, terminating training if the validation loss failed to improve over 12 consecutive iterations (patience = 12). Furthermore, a dropout rate of 0.2 was applied to the hidden layers to prevent feature co-adaptation and further enhance robustness.

3.3.6. Walk-Forward Validation Strategy

Walk-forward validation was used to avoid information leakage and preserve temporal order. The model was first trained on an initial window and then used for one-step-ahead forecasting. After each step, the new actual value was added to the training set, and the process continued sequentially. This expanding-window approach maintains chronological consistency and reflects real-world iterative forecasting.

3.3.7. Performance Evaluation Metrics

Forecasting performance was evaluated using multiple accuracy measures.

Root mean square error:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (D_t - \widehat{D}_t)^2}$$

Mean absolute error:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |D_t - \widehat{D}_t|$$

Mean absolute percentage error:

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^N \left| \frac{D_t - \widehat{D}_t}{D_t} \right|$$

Coefficient of determination:

$$R^2 = 1 - \frac{\sum (D_t - \widehat{D}_t)^2}{\sum (D_t - \bar{D})^2}$$

Directional accuracy:

$$\text{DA} = \frac{1}{N} \sum \mathbb{I}[(D_t - D_{t-1})(\widehat{D}_t - \widehat{D}_{t-1}) > 0]$$

3.4. Experimental Settings and Reproducibility

All experiments were implemented in Python 3.11 using standard scientific computing and machine learning libraries. The LSTM model was trained using the Adam optimizer with adaptive learning rate updates. To ensure reproducibility, consistent preprocessing and normalization procedures were applied across all experiments, fixed random seeds were used for model initialization, and a uniform walk-forward validation protocol was employed. In addition, identical stopping criteria were used for all optimization runs. The DGWO algorithm was executed with a predefined population size and number of iterations, and the corresponding hyperparameter search ranges are reported below.

Hyperparameter search ranges:

L: 2 - 12

Hidden neurons: 8 - 128

Learning rate: 0.0001 - 0.01

Epochs: 50 - 300

3.5. Baseline Models and Fair Comparison Protocol

To evaluate the proposed LSTM-DGWO framework, its forecasting performance was benchmarked against three models: a baseline LSTM without hyperparameter optimization, the naive method, and ARIMA. The baseline LSTM isolates the specific contribution of the DGWO optimizer, whereas the naive and ARIMA models represent persistence-based and linear stochastic benchmarks, respectively.

The naive model provides a parsimonious lower bound, assuming that the next forecast equals the most recent

observation:

$$\hat{y}_{t+1} = y_t$$

As a stronger statistical benchmark, the ARIMA (p, d, q) process was formulated to capture linear temporal dependencies:

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t$$

Following the Box-Jenkins procedure and ADF stationarity testing, ARIMA (1, 1, 1) was selected for the monthly demand series. To ensure a rigorous comparison, a strict protocol was implemented: all models were evaluated under identical experimental conditions, including the dataset, sliding-window structure, and walk-forward validation strategy. No future information was permitted to enter the estimation or optimization processes, ensuring that observed performance gains reflect genuine model capability rather than inconsistencies in design. To supplement the quantitative error metrics (RMSE, MAE, and MAPE), a visual trajectory analysis was integrated into the evaluation protocol. This involved plotting the forecasted values of the hybrid LSTM-DGWO model against the baseline ARIMA and naive models. The primary objective of this visual analysis was to assess the structural robustness of the models—specifically, their ability to capture nonlinear turning points and volatility clusters within the test set. Unlike aggregate metrics, which may mask local forecasting failures, this graphical comparison reveals the degree of temporal lag (persistence) in simpler models and highlights the hybrid model's capacity to synchronize with rapid demand fluctuations in the pharmaceutical market.

3.6. Integrated Operational Supply Chain Framework for Forecast-Driven Inventory and Replenishment Decisions

As shown in Figure 2, the operational framework and forecasting methodology were designed as an integrated structure to bridge the gap between predictive data analysis and logistical decision-making. This structure consists of two complementary layers: the forecasting methodology layer, which employs computational intelligence through the LSTM-DGWO model to transform complex environmental variables into demand forecasting patterns, and the supply chain operational framework layer, which transmits these patterns to the physical

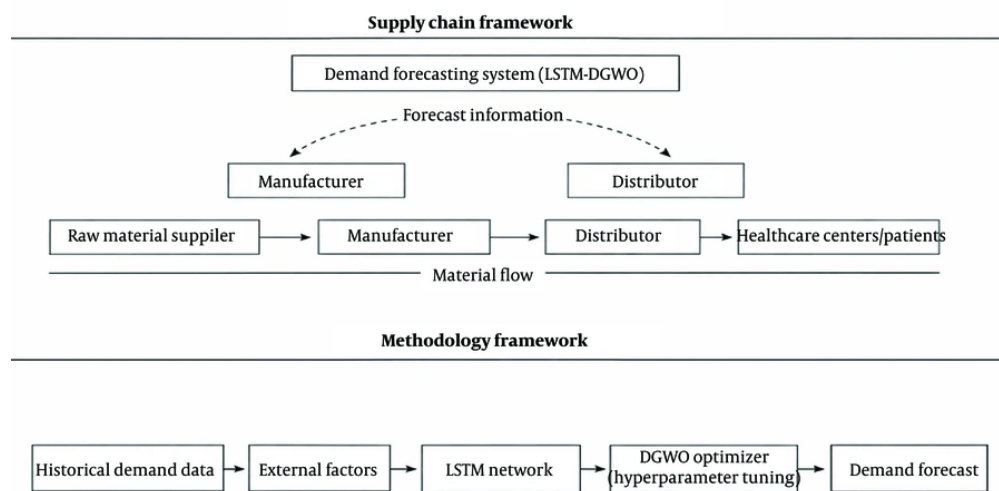


Figure 2. Integrated operational supply chain framework linking the proposed hybrid forecasting model with downstream inventory and replenishment decision-making layers.

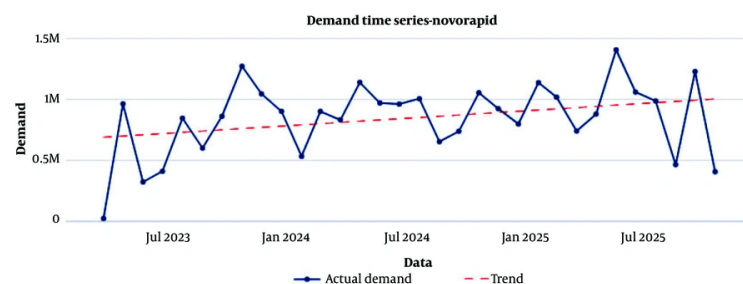


Figure 3. Monthly demand time series of the NovoRapid product from March 2023 to October 2025.

components of the supply chain as informational signals. Coordination between the computational and operational layers constitutes a key strength of the proposed integrated system, contributing to reduced uncertainty in inventory management and improved material flow optimization.

4. Results

4.1. Descriptive Analysis of the Demand Time Series

The empirical analysis begins with an examination of the monthly demand time series for the pharmaceutical

product under study. As shown in Figure 3, demand evolved over a 32-month period from March 2023 to October 2025, exhibiting noticeable fluctuations over time. The mean monthly demand was 847,138 units, with a standard deviation of 298,876 units, resulting in a coefficient of variation of 35.28%. This level of dispersion indicates a relatively volatile demand pattern and suggests that the series may require a flexible forecasting approach capable of capturing nonlinear and time-varying dynamics. From an operational perspective, the magnitude of demand variation is substantial, ranging from approximately 0.5 million to over 1.0 million units. Such

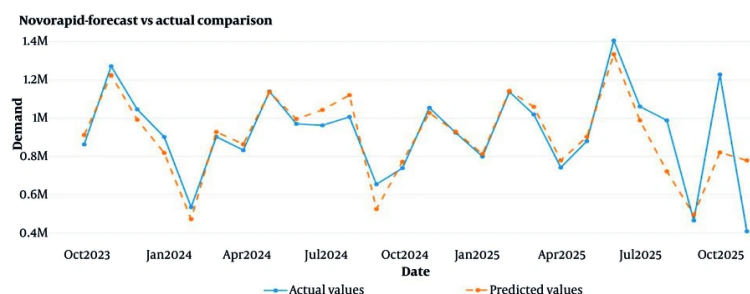


Figure 4. Actual and predicted NovoRapid demand over time.

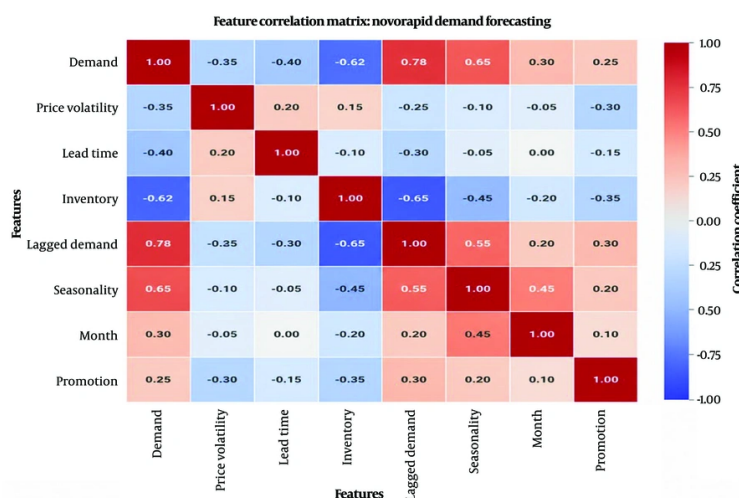


Figure 5. Correlation matrix of the exploratory variables.

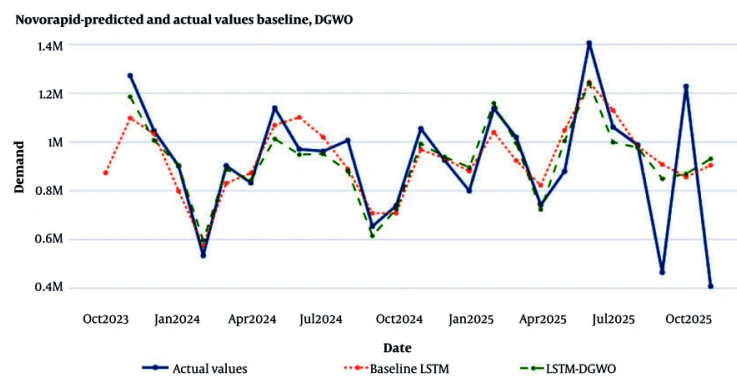
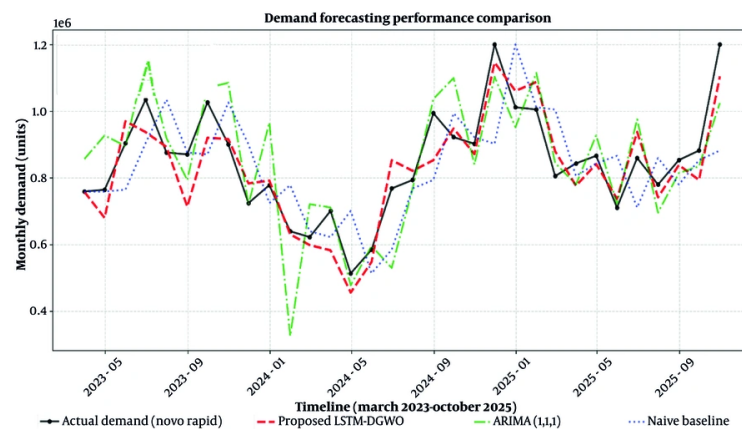
variability is particularly critical in pharmaceutical supply chains, where forecasting errors can directly influence inventory availability, procurement timing, replenishment planning, and the risks of both stockouts and overstocking. In addition, given the relatively short historical horizon of the dataset, model training required careful control to preserve generalizability. For this reason, the regularization mechanisms described in Section 3 were incorporated to mitigate overfitting. Overall, the observed instability in the demand series provides an empirical basis for adopting a dynamic and nonlinear forecasting framework.

4.2. Temporal Comparison of Actual and Predicted Demand and Inter-Variable Relationships

Figure 4 presents a temporal comparison between the actual and predicted NovoRapid demand values over the study period. The results show that the proposed forecasting model closely tracks the overall movement of demand over time, including several short-term fluctuations. This pattern indicates that the model captures the underlying demand dynamics with a reasonable level of accuracy and is capable of following variations in observed demand across different periods. In addition, Figure 5 presents the correlation matrix of the exploratory

Table 2. Performance Comparison Between the Baseline LSTM Model and the Optimized LSTM-DGWO Model

Evaluation Metric	LSTM Baseline	LSTM-DGWO	Performance Interpretation
RMSE	223,957	131,400	Reduction of second-order error
MAE	174,407	81,112	Reduction in mean absolute error
MAPE (%)	23.26	10.59	Improvement in scalable relative error
R ²	0.0772	0.6823	Improved explanation of variance

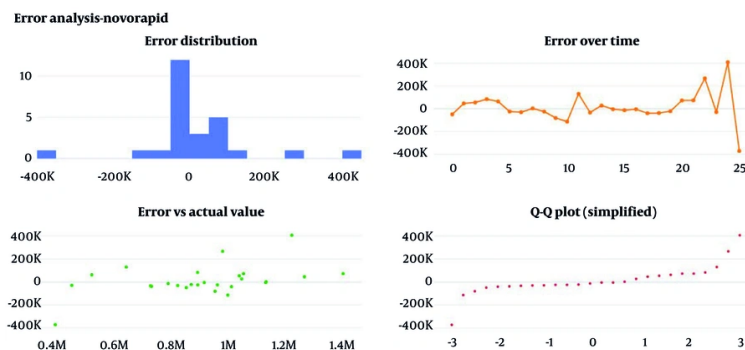
**Figure 6.** Comparison of actual demand with the forecasts of the baseline LSTM model and the optimized LSTM-DGWO model.**Figure 7.** Comparison of actual monthly demand for NovoRapid versus forecasted values using the proposed LSTM-DGWO model, ARIMA (1, 1, 1), and naive baseline (March 2023 - October 2025).

variables. The results reveal associations among several contextual factors, including price, logistics, capacity, and sustainability-related indicators. However, these variables are considered only for exploratory analysis and are not

used as input features in the proposed LSTM-DGWO forecasting model. Therefore, the forecasting framework remains strictly univariate, relying exclusively on historical lagged demand observations. Overall, the

Table 3. Comparative Performance of the Proposed LSTM-DGWO Model, ARIMA (1, 1, 1), and the Naive Baseline for NovoRapid Demand Forecasting

Model	RMSE	MAE	MAPE (%)	R ²
LSTM-DGWO (Proposed)	131,400	81,112	10.59	0.6823
ARIMA (1, 1, 1)	158,240	102,450	13.82	0.5410
Naive baseline	192,100	124,300	17.15	0.3250

**Figure 8.** Residual analysis, including error distribution, error over time, error versus actual values, and Q-Q plot.

temporal agreement between actual and predicted demand, together with the exploratory correlation structure, supports the applicability of an adaptive, data-driven forecasting approach for pharmaceutical demand analysis.

4.3. Forecasting Performance Evaluation

The forecasting performance of the proposed framework was assessed by comparing actual demand values with the predictions generated by the model. As illustrated in Figure 6, the LSTM-DGWO model generally followed the overall trajectory of the demand series and captured the major fluctuations and turning points with reasonable accuracy. For a quantitative evaluation, Table 2 compares the baseline LSTM model with the optimized LSTM-DGWO model across the selected performance metrics. The results show that the optimized model outperformed the baseline model on all major evaluation criteria, yielding an overall average improvement of 49.76%. This finding indicates that integrating the DGWO algorithm into the hyperparameter optimization process contributed substantially to forecasting performance. This improvement is particularly noteworthy given the limited size of the historical dataset. As described in Section 3, the

training procedure and search space were designed under regularization and model-capacity control constraints to reduce overfitting and support generalizability.

4.4. Final Forecasting Performance and Residual Analysis

Actual monthly NovoRapid demand and forecasted values are compared in Figure 7.

The comparative results in Table 3 show that the proposed LSTM-DGWO model outperformed both ARIMA (1, 1, 1) and the naive baseline across all evaluation metrics. It achieved the lowest RMSE, MAE, and MAPE, along with the highest R² value, indicating superior predictive accuracy and stronger explanatory power. Residual analysis in Figure 8 indicates that the prediction errors are largely centered around zero, suggesting that the model captures the general pattern of the series without severe systematic bias. Nevertheless, slight asymmetry in the residual distribution, heavier tails, and increasing error dispersion at higher demand levels indicate reduced predictive stability during extreme or highly volatile demand periods.

4.5. Hyperparameter Optimization and Comparative Analysis

The primary objective of employing the DGWO was to overcome the limitations of baseline LSTM models by identifying an optimal set of hyperparameters. Because the performance of deep learning models is highly sensitive to configurations such as the look-back window and learning rate, DGWO was used to navigate the mixed-integer search space efficiently. The final optimized hyperparameters, which yielded the minimum validation error, are summarized in Table 4.

Table 4. Final Optimized Hyperparameters for the LSTM-DGWO Model

Parameter	Value
Look-back window	7 (mo)
Number of LSTM neurons	48
Learning rate	0.005
Epochs	250
Loss function	MSE
Optimizer algorithm	Adam
Hyperparameter tuning algorithm	DGWO
Iterations to convergence	12

To validate the superiority of the proposed LSTM-DGWO framework, its performance was compared with several well-known metaheuristic optimization algorithms, including the standard GWO, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Differential Evolution (DE), and Artificial Bee Colony (ABC). To ensure a fair comparison, all algorithms were evaluated using Normalized RMSE (N-RMSE) values, as shown in Table 5.

Table 5. Performance Comparison of Different Optimization Algorithms Using RMSE on Normalized Data

Algorithm	Rank	RMSE
DGWO	1	0.28
GWO	2	0.29
PSO	3	0.30
GA	4	0.33
DE	5	0.35
ABC	6	0.38

The results indicate that DGWO outperformed the other metaheuristics by achieving the lowest N-RMSE value of 0.28. This superior performance can be attributed to the improved exploration-exploitation balance in DGWO, which prevents the model from becoming trapped in local optima, a common issue in standard GWO and GA. Furthermore, the convergence analysis illustrated in Figure 9 confirms the efficiency of the DGWO algorithm. While other algorithms required more iterations to stabilize, DGWO converged in only 12 iterations, showing a

steeper and smoother reduction in the objective function value. This rapid convergence demonstrates that DGWO is not only more accurate but also computationally efficient for tuning complex deep learning architectures in pharmaceutical supply chain forecasting.

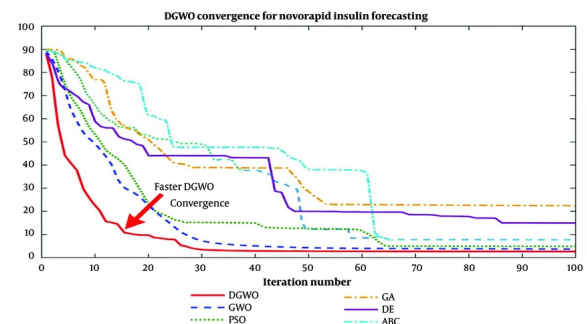


Figure 9. Comparison of the optimization algorithms in terms of RMSE and convergence behavior.

The results show that LSTM-DGWO achieved the best performance across all evaluation metrics, with the lowest RMSE, MAE, and MAPE, as well as the highest R^2 . Specifically, the hybrid model obtained an RMSE of 131,400, an MAE of 81,112, a MAPE of 10.59%, and an R^2 of 0.6823, indicating more accurate forecasting performance than the benchmark models. In contrast, ARIMA (1, 1, 1) and naive produced larger forecast errors and lower explanatory power. Overall, these findings demonstrate that the proposed hybrid LSTM-DGWO model provided the most accurate demand forecasts among the evaluated approaches.

4.6. Comparative Analysis

As illustrated in Figure 10, the baseline models struggle to capture the substantial volatility observed in the latter part of the test set, particularly the sharp decline and subsequent rebound between August and October 2025. The naive model exhibits a typical one-step lag (persistence) effect, effectively shifting the previous month's value forward, which results in a failure to anticipate trend reversals. Meanwhile, the model produces a smoothed trend that fails to track the high-amplitude, nonlinear peaks and troughs of the actual demand series. This visual evidence further justifies the transition to the proposed hybrid deep learning framework (LSTM-DGWO),

as traditional statistical and persistence models lack the necessary complexity to model such rapid demand fluctuations in the pharmaceutical supply chain.

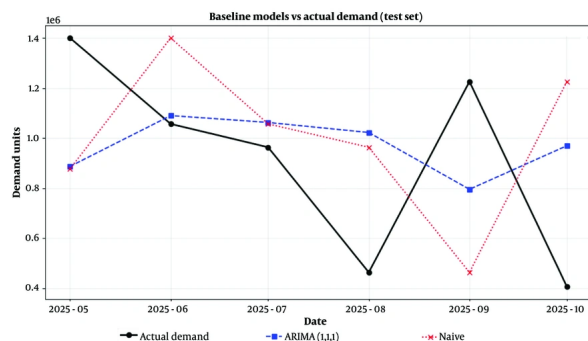


Figure 10. Performance comparison of baseline models (ARIMA and Naive) against actual NovoRapid demand for the test period (May 2025 - October 2025).

5. Discussion

The results indicate that the proposed LSTM-DGWO framework outperformed ARIMA (1, 1, 1) and the naive baseline in forecasting NovoRapid demand. This improvement reflects the complementary roles of the two components: LSTM captures temporal dependencies and nonlinear patterns, whereas DGWO optimizes the network's hyperparameters and should therefore be regarded as a tuning mechanism rather than an independent forecasting model. These findings are consistent with the optimization principles described by Mirjalili et al. (9), the value of intelligent search strategies noted by Senapati and Dash (11), and the efficiency of careful hyperparameter selection emphasized. They also reinforce the limitations of classical statistical benchmarks for nonlinear series, as discussed by Makridakis et al. (3).

Another important contribution of this study is that the model was developed using a constrained architecture. The use of a limited look-back window, a restricted neuron count, dropout, and L2 regularization helped control model complexity and reduce overfitting, which is particularly important for the 32-month dataset used here. In this context, DGWO helped identify a more robust and efficient parameter setting rather than simply increasing model complexity. Practically, these results suggest that the proposed framework can improve pharmaceutical demand forecasting, support inventory planning and

replenishment decisions, reduce uncertainty and risk, and strengthen supply-chain decision-making, in line with previous studies on data-driven performance improvement and forecasting under disruption (16, 17) and the findings of Sadeghi and Mohammadzadeh (18).

5.1. Limitations

This study has several limitations. First, it was based on only 32 monthly observations; therefore, although dropout, early stopping, L2 regularization, walk-forward validation, and constrained model complexity were used to reduce overfitting, some residual overfitting cannot be excluded. Second, the analysis focused on a single product, NovoRapid, which limits generalizability to other medicines or supply-chain settings. Third, the benchmark comparison was limited to ARIMA (1, 1, 1) and naive; thus, future work should include a wider range of modern forecasting models.

5.2. Conclusions

This research addressed the challenge of demand forecasting in a volatile pharmaceutical market by developing a hybrid LSTM-DGWO framework. Empirical results demonstrated that the proposed model outperformed conventional benchmark approaches, including the ARIMA (1, 1, 1) model and the naive baseline. The final hybrid model achieved a MAPE of 10.59%, an RMSE of 131,400, an MAE of 81,112, and an R^2 of 0.6823, indicating a superior capability in capturing the complex, nonlinear dynamics of the demand series within the studied dataset.

The performance gains are attributed to the synergistic integration of deep sequence learning and metaheuristic optimization. In this framework, the LSTM network was tasked with learning the temporal dependencies and nonlinear patterns inherent in the demand data, whereas the DGWO algorithm systematically identified optimal hyperparameters. Consequently, the superior forecasting performance is not attributable to the optimizer alone, but rather to the combined effect of deep learning architecture and rigorous hyperparameter tuning, which facilitated a more stable modeling process under constrained data conditions.

Practically, the proposed framework serves as a robust decision-support tool for pharmaceutical supply chain management. By enhancing forecasting precision, the model facilitates more efficient inventory planning, optimized distribution strategies, and reduced risk of

shortages, particularly in operational environments in which demand volatility challenges conventional statistical methods. Despite these contributions, the findings must be interpreted within the study's limitations. The analysis was based on a 32-month longitudinal dataset focused on a single pharmaceutical product, which inherently limits the generalizability of the results. Future research should aim to scale this framework across broader product portfolios, larger datasets, and diverse market scenarios. Additionally, investigating the integration of exogenous variables—such as market trends or epidemiological data—could further enhance the predictive power and robustness of the proposed hybrid model.

Acknowledgements

The authors would like to express their sincere gratitude to the Iranian Food and Drug Administration (IFDA) for providing the NovoRapid insulin datasets and for offering valuable technical insights into the dynamics of the pharmaceutical market.

Footnotes

AI Use Disclosure: The authors declare that no generative AI tools were used in the creation of this article.

Authors' Contribution: Study concept and design: M. A. A. and E. N.; Acquisition of data: M. A. A.; Analysis and interpretation of data: M. A. A., E. N., and M. A.; Drafting of the manuscript: M. A. A.; Critical revision of the manuscript for important intellectual content: E. N., M. A., and A. G.; Statistical analysis: M. A. A., under the supervision of E. N.; Administrative, technical, and material support: A. G. and E. N.; Study supervision: E. N.

Conflict of Interests Statement: The authors declare that there are no scientific, financial, or personal conflicts of interest associated with this research. All stages of study design, analysis, manuscript preparation, and reporting of results were conducted independently and without the influence of any organizational or commercial interests.

Data Availability: The data used for modeling and analysis in this study are available from the corresponding author upon reasonable request. Due to the nature of the study, the data are not publicly available.

Funding/Support: This research was conducted through the authors' personal resources and without receiving financial support from any specific institution or company. All costs related to data processing, model implementation, and preparation of the final documentation were fully covered by the researchers.

Informed Consent: Informed consent was not required for this study as it did not involve human participants or the use of identifiable personal data.

References

- Shah N. Pharmaceutical supply chains: Key issues and strategies for optimization. *Comput Chem Eng*. 2004;**28**(6 - 7):929-941. <https://doi.org/10.1016/j.compchemeng.2003.09.022>.
- Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;**50**:159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
- Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*. 2018;**13**(3). e0194889. [PubMed ID: 29584784]. [PubMed Central ID: PMC5870978]. <https://doi.org/10.1371/journal.pone.0194889>.
- Kelle P, Woosley J, Schneider H. Pharmaceutical supply chain specifics and inventory solutions. *Oper Res Health Care*. 2012;**1**(2 - 3):54-63. <https://doi.org/10.1016/j.orhc.2012.07.001>.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;**9**(8):1735-1780. [PubMed ID: 9377276]. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. *Neural Comput*. 2000;**12**(10):2451-2471. [PubMed ID: 11032042]. <https://doi.org/10.1162/089976600300015015>.
- Hirsch IB. Insulin analogues. *N Engl J Med*. 2005;**352**(2):174-183. [PubMed ID: 15647580]. <https://doi.org/10.1056/NEJMra040832>.
- Home PD. Pharmacokinetics and pharmacodynamics of insulin analogues. *Diabetes Obes Metab*. 2012;**14**(9):780-788. [PubMed ID: 22321739]. <https://doi.org/10.1111/j.1463-1326.2012.01580.x>.
- Mirjalili S, Mirjalili SM, Lewis A. Grey Wolf Optimizer. *Adv Eng Softw*. 2014;**69**:46-61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- Saremi S, Mirjalili S, Lewis A. How important is a transfer function in discrete heuristic algorithms. *Neural Comput Appl*. 2015;**26**:625-640. <https://doi.org/10.1007/s00521-014-1743-5>.
- Senapati MR, Dash PK. Hybrid deep learning and particle swarm optimization for forecasting. *Appl Soft Comput*. 2018;**68**:408-418.
- Amjady N, Keynia F, Zareipour H. Short-term load forecasting using hybrid artificial intelligence models. *IEEE Trans Smart Grid*. 2010;**1**(3):1-11. <https://doi.org/10.1109/TSG.2010.2078842>.
- Wang J, Li Y, Guo Z. Metaheuristic optimization of LSTM for time series forecasting. *Expert Syst Appl*. 2021;**173**:114686.
- Zhang Y, Liu Q. Pharmaceutical demand forecasting using deep learning approaches. *Comput Ind Eng*. 2020;**149**:106781.
- Dogan K, Tuncer T. Deep learning-based demand forecasting in healthcare systems. *Comput Methods Programs Biomed*. 2022;**215**:106606.

16. Baryannis A, Validi S, Dani S, Antoniou G. Supply chain risk management and artificial intelligence: state of the art and future research directions. *Int J Logist Res Appl.* 2019;**22**(3):235-252. <https://doi.org/10.1080/13675567.2018.1534060>.
17. Ivanov D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the example of COVID-19. *Transp Res E Logist Transp Rev.* 2020;**136**:101922.
18. Sadeghi S, Mohammadzadeh M. Evaluating the Impact of Strategic Alignment on Performance Components of Iranian Pharmaceutical Companies Using Machine Learning Techniques. *Iran J Pharm Res.* 2025;**24**(1). e165722. <https://doi.org/10.5812/ijpr-165722>.