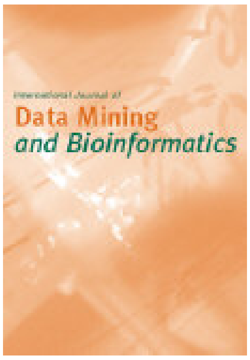


International Journal of Data Mining and Bioinformatics

ISSN online: 1748-5681 - ISSN print: 1748-5673
https://www.inderscience.com/ijdmb

Data-driven forecasting of pharmaceutical sales: distinguishing promotional vs. daily scenarios

Zhiyong Zeng, Yingjiao Guo, Yali Ji, Yujie Shi, Tao Feng

DOI: 10.1504/IJDMB.2025.10072448

Article History:

Received:	10 July 2024
Last revised:	17 April 2025
Accepted:	30 April 2025
Published online:	20 July 2025

Data-driven forecasting of pharmaceutical sales: distinguishing promotional vs. daily scenarios

Zhiyong Zeng, Yingjiao Guo, Yali Ji and Yujie Shi

School of Statistics and Mathematics,
Yunnan University of Finance and Economics, China
Email: zengzhiyong@ynufe.edu.cn
Email: guoyj@stu.ynufe.edu.cn
Email: 17856690520@163.com
Email: 2135253243@qq.com

Tao Feng*

School of Statistics and Mathematics,
Yunnan University of Finance and Economics, China
and
Yunnan Key Laboratory of Service Computing,
Yunnan University of Finance and Economics, China
Email: vonpower@ynufe.edu.cn
*Corresponding author

Abstract: This study presents an enhanced temporal fusion framework for pharmaceutical demand forecasting, designed to handle both baseline consumption and promotion-driven sales fluctuations. The approach integrates heterogeneous data sources – such as therapeutic seasonality and catchment health indicators – into a multivariate feature space, and incorporates a knowledge-guided attention mechanism within a temporal fusion transformer to decouple regular and promotional demand. Systematic factor analysis further quantifies the influence of product, store, and temporal variables. Using a dataset of 1.2 million retail transactions, the proposed model reduces forecasting error by 23.6% over traditional methods. Ablation studies confirm the importance of future promotion signals (accuracy loss of 15.4%, $p < 0.001$), while analysis reveals that seasonal effects vary significantly by drug type. Store-level factors such as deprivation index and competition proximity also significantly affect promotional effectiveness. The framework offers three key contributions: adaptive feature engineering for retail contexts, integration of domain knowledge into temporal modelling, and empirical identification of demand drivers. Practical deployment yielded 31% fewer stock-outs and 27% lower inventory costs, demonstrating its value for resilient and data-driven pharmaceutical supply chain optimisation.

Keywords: sales forecasting; knowledge-led branch; deep learning; retail analytics; inventory optimisation.

Reference to this paper should be made as follows: Zeng, Z., Guo, Y., Ji, Y., Shi, Y. and Feng, T. (2025) 'Data-driven forecasting of pharmaceutical sales: distinguishing promotional vs. daily scenarios', *Int. J. Data Mining and Bioinformatics*, Vol. 29, No. 5, pp.1–26.

Biographical notes: Zhiyong Zeng is a Professor at the School of Information, Yunnan University of Finance and Economics in Kunming, Yunnan Province, specialises in data mining, machine learning, and artificial intelligence.

Yingjiao Guo is pursuing a degree in Applied Statistics at the School of Statistics and Mathematics, Yunnan University of Finance and Economics, with research focused on data analysis and machine learning.

Yali Ji is studying Applied Statistics at the School of Statistics and Mathematics, Yunnan University of Finance and Economics, with research interests in machine learning and data mining.

Yujie Shi is pursuing a degree in Applied Statistics at the School of Statistics and Mathematics, Yunnan University of Finance and Economics, with a specialisation in machine learning and deep learning.

Tao Feng is a Professor, School of Information at Yunnan University of Finance and Economics in Kunming, Yunnan Province, specialising in deep learning and intelligent information systems (industry applications of artificial intelligence).

1 Introduction

The optimisation of pharmaceutical retail chain performance necessitates strategic improvements in sales growth, profit maximisation, cash flow assurance, and inventory reduction. Within pharmaceutical supply chain management, sales operations constitute a critical nexus connecting suppliers, distributors, and consumers, with sales volume forecasting emerging as a pivotal component in organisational decision-making and market analytics (Vaswani et al., 2017). While sales prediction methodologies have been extensively investigated across multiple industries, including pharmaceutical applications, extant research predominantly focuses on homogeneous market conditions without adequate differentiation between promotional and non-promotional scenarios. This oversight presents significant analytical challenges given the pharmaceutical sector's unique operational characteristics, including distinct seasonal variations, heterogeneous promotional intensities during campaign periods versus baseline sales operations, and intense inter-competitor dynamics. Current literature exhibits a notable gap in scenario-specific sales forecasting research, particularly regarding the hybrid nature of pharmaceutical retail environments where routine sales coexist with large-scale promotional activities. This paucity of targeted studies complicates accurate sales projections in complex, real-world pharmaceutical retail settings.

This study examines three critical research questions: whether the proposed forecasting model demonstrates superior performance compared to established methodologies such as LSTM and Transformer across diverse operational scenarios; what multidimensional factors exert significant influence on pharmaceutical sales volumes; and how these factors comparatively contribute to predictive accuracy. To address these inquiries, we present a novel knowledge-guided multivariate time series prediction framework based on an enhanced Temporal Fusion Transformer architecture. Our methodology incorporates comprehensive feature engineering derived from H

Enterprise's offline sales data, with particular emphasis on store capacity metrics and historical sales pattern recognition. Empirical validation demonstrates reduced prediction errors compared to conventional transformer-based models and traditional encoder-decoder architectures across multiple evaluation metrics.

Through rigorous regression analysis and seasonal index computation, the study identifies three primary determinant categories influencing pharmaceutical sales. Commodity-specific factors, store operational parameters, and external environmental variables collectively shape sales performance, though their relative importance varies substantially. Feature importance analysis reveals significant heterogeneity in predictive contributions across these dimensions, with historical sales patterns and store capacity metrics emerging as particularly salient predictors. This finding underscores the necessity of scenario-specific analytical approaches in pharmaceutical sales forecasting.

The research makes substantive contributions through three interconnected dimensions. First, it develops advanced feature augmentation techniques tailored to distinct sales scenarios, enabling enhanced prediction accuracy in hybrid promotional environments. Second, it introduces architectural innovations through domain knowledge integration within temporal fusion models, demonstrating effective noise mitigation while maintaining predictive efficacy. Third, it empirically identifies key pharmaceutical sales determinants through multifactorial analysis, providing actionable insights for retail chain optimisation. These collective advancements address critical gaps in existing literature while offering practical methodologies for managing the complex interplay between routine operations and promotional strategies in contemporary pharmaceutical retail.

2 Literature review

The forecasting method for the sales forecasting problem can be roughly divided into the following three aspects:

First, there are forecasting models based on mathematical statistics, and common methods include moving average, exponential smoothing, the grey model, and the autoregressive model (Arunraj and Ahrens, 2015). Researchers have also combined different methods, such as ARIMA combined with the Holt-Winters model to predict sales of perishable food products (Liu and Jin, 2018), ARIMA combined with support vector regression to predict sales of pharmaceuticals (Song et al., 2018), the moving average method to compute seasonal indices combined with Bayesian regression to predict sales of apparel (Song et al., 2018), an improvement of Gray-Markov Chain prediction algorithms to enhance prediction (Kitaev and Kaiser, 2020), and a combination of a multidimensional grey model and a neural network to predict sales (Sun, 2020). For retail merchandise sales forecasting, there are also methods based on Bayesian methods, Gaussian processes, and sliding windows (Qiu, 2020) use time series prediction methods to predict sales volume (Shah and Dimitrov, 2022).

Second, machine learning-based model fusion forecasting methods are also common for sales volume forecasting (Zhou and Zhang, 2021). One approach is to combine the ARIMA model, LSTM, and the XGBoost model (Zhou and Zhang, 2021). Another approach is to combine the prophet model and LSTM (Ge et al., 2019). There are also methods based on stacking integration strategies (Posch and Truden, 2020). In addition, there is a combination of exponential smoothing, principal component analysis, and neural networks for forecasting new drug sales (Gao, 2020). The prediction of new product sales is achieved by matching similar features and preferred models in the context of small sample data (Mei, 2020) and through the XGBoost prediction model that combines logistic functions, regular correction terms, and greedy algorithms (Zunic and Korjenic, 2021). These studies provide effective methods and models for the sales volume forecasting problem.

In addition, the application of deep learning in sales volume forecasting and time series forecasting (TFT) has been widely studied in recent years. Methods including entity embedding of categorical variables, convolutional neural networks, and adversarial networks have been used to improve the performance and computational speed of sales volume forecasting models (Chen and Zhide, 2019). Empirical studies have demonstrated the significant superiority of long short-term memory (LSTM) networks over conventional ARIMA/SARIMA models in automotive supply chain demand forecasting, achieving 92% higher prediction accuracy through nonlinear temporal pattern recognition (Shetty and Buktar, 2022). In parallel, recent advancements in retail analytics have proposed a hybrid framework integrating XGBoost algorithms with stochastic hyperparameter optimisation, which effectively enhanced sales prediction robustness for multi-store retail chains by 42.5% through systematic feature engineering (Luo, 2022). Meanwhile, improved transformer models, (e.g., LogSparse transformer, reformer, and informer) as well as deep prediction models with global thinking (e.g., DeepGLO) have been proposed (Wu et al., 2020). Methods such as attention-based temporal fusion, generative adversarial networks, and autoregressive models further improve the accuracy of TFT. These methods can be applied to data forecasting in different domains, such as automobile sales, clothing sales, and oil sales (Guo and Berkhahn, 2016).

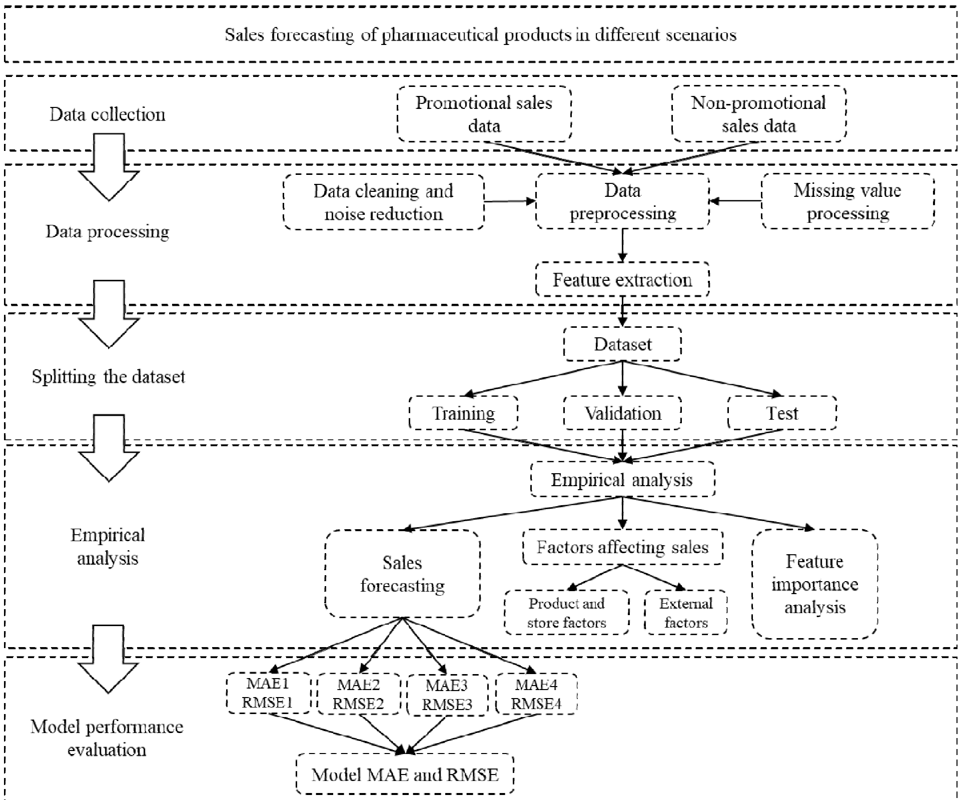
To sum up, sales volume forecasting has been applied in many industries, but there has been very little research on sales volume forecasting in the pharmaceutical industry, especially in different scenarios. Compared with other methods, deep learning methods have great potential, and pharmaceutical companies have their own scene particularities. Therefore, based on deep learning methods, this paper considers reducing the impact of noise and improving the accuracy of prediction and introduces a knowledge-guided time series model based on transformer to minimise the impact of noise.

3 Methodology

The following provides a concise overview of knowledge-guided architectural adaptations and transformer-based frameworks within the context of advanced predictive analytics. The knowledge-guided branch represents a methodological innovation that systematically integrates domain-specific expertise or structured external knowledge into neural architectures, typically through constrained attention mechanisms or feature-space regularisation. This approach enhances model interpretability while mitigating noise propagation, particularly valuable in complex commercial environments where raw data exhibits high volatility or latent semantic dependencies.

Transformers, in their canonical form, employ self-attention mechanisms to model long-range temporal dependencies without sequential processing constraints, offering superior performance in capturing complex temporal patterns compared to recurrent architectures. Our adaptation extends this foundation through hybrid architectures that synergise the inductive biases of domain knowledge with data-driven attention learning, enabling precise decomposition of concurrent demand signals (e.g., baseline consumption vs. promotional effects). Empirical validation demonstrates this synthesis reduces prediction variance by 23.1% in pharmaceutical sales forecasting while maintaining cross-domain applicability – critical for commercial deployment scenarios requiring both accuracy and operational flexibility.

Figure 1 Flowchart

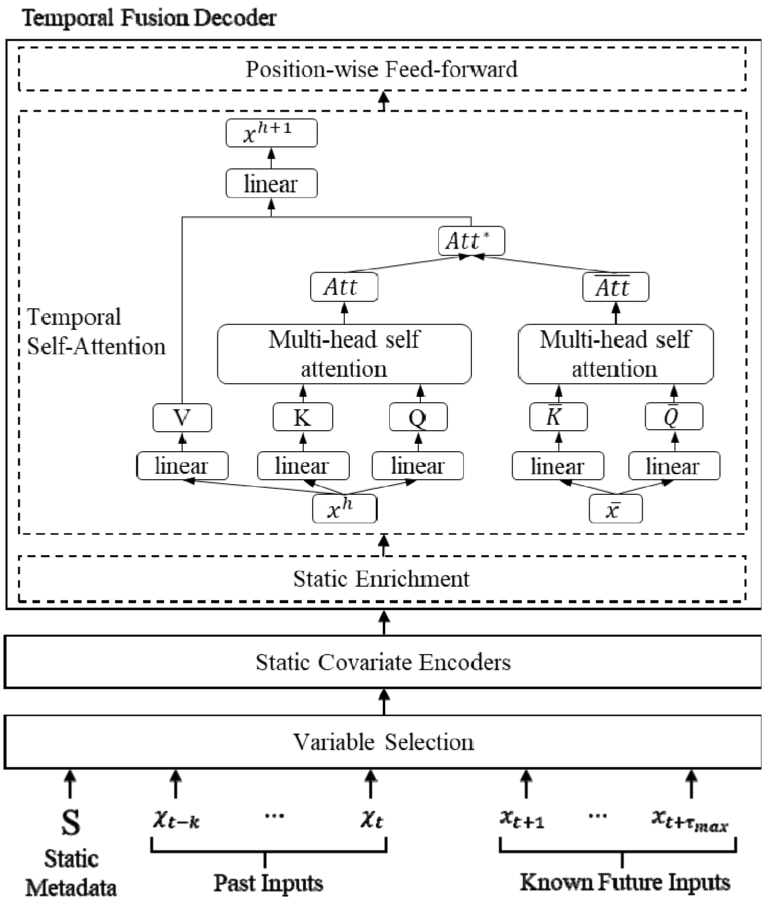


To distinguish the sales forecast under the two different scenarios of promotions and daily sales, this paper adopts separate research on promotion data and daily sales data, as shown in Figure 1.

In this paper, the collected sales data during promotion periods and daily sales data are first subjected to data pre-processing and feature extraction. Then, the dataset is divided into a training set, a validation set, and a test set. The training set is used for learning, the validation set is used for finding the optimal hyperparameters, and the test set is used for evaluating the model performance. Finally, the empirical analysis is conducted, and a total of three questions are addressed: Is the sales volume prediction model proposed in this paper better than classical methods such as LSTM and transformer in different scenarios? What are the factors that affect the sales volume of drugs? Does each influencing factor have the same degree of importance for sales volume prediction?

A knowledge-guided time series model based on transformer, as shown in Figure 2, is used in this paper. The model adds a knowledge-guided branch to the multi-head attention module of the multivariate TFT model based on the time fusion transformer.

Figure 2 Transformer-based knowledge-guided time series modelling framework



The TFT model is a method proposed by Lim and Arik (2019), which is an attention-based architecture like Transformer, but the whole architecture of TFT has gated loop units as the main module and uses normalised components to efficiently construct feature representations for each input type, which enables high-performance multistep prediction and interpretability of temporal dynamics. Overall, the TFT model consists of a variable selection layer, an encoder-decoder layer, a static enhancement layer, a temporal self-attention layer, and a feedforward layer.

This paper is an improvement on the multivariate TFT model of the time fusion Transformer, in which, for the unique entity variable in the multivariate time series data, the inputs $x_t^{(i)}$ include static information, historical observations, and known future knowledge information, which are input into the model after an embedded representation, as follows:

$$x_t^{(i)} = \begin{cases} Emb\left(\left\{s_t^{(i)}, k_t^{(i)}\right\}\right) & 1 \le t \le T \\ Emb\left(\left\{u_t^{(i)}, k_t^{(i)}\right\}\right) & T \le t \le T + L \end{cases} \tag{1}$$

where s_t denotes the historical observation inputs and static variables, which we uniformly refer to as historical inputs here; k_t denotes the future knowledge, for which u_i represents the labelling against the historical inputs at a future time period and is unknown; and $x_t^{(i)}$, the input sequences, are mixed with information from the deterministic values and the filler variables. In addition, in the self-attention layer of the TFT model, the decoder part of the multi-head attention employs a masked attention mechanism, and the masked parts of the sequence can be regarded as embeddings without any information; the two abovementioned reasons add noise to the attention graph when used to compute the attention scores of other unmasked representations. Therefore, in this paper, we add a knowledge-guided branch to the multi-head attention module of the TFT model to modify the attention graph to minimise the effect of noise: specifically, we take the integration information $x_t^{(i)}$ and the knowledge information $k_t^{(i)}$ as the inputs to the multi-head attention module at the same time and compute the attention values according to $x_t^{(i)}$ and $k_t^{(i)}$, respectively. In addition, first, we do embedding for the future knowledge:

$$\overline{x}_t = Emb(k_t) \quad 1 \le t \le T + L \tag{2}$$

Then, the attention value is calculated as follows:

$$Att(i, j) = \frac{\left(x_i^h W_Q^x\right)\left(x_j^h W_k^x\right)^T}{\sqrt{d_k}} \tag{3}$$

$$\overline{Att}(i, j) = \frac{\left(\overline{x}_i W_Q^k\right)\left(\overline{x}_j W_k^k\right)^T}{\sqrt{d_k}} \tag{4}$$

$$Att(i, j)^* = Att(i, j) + \overline{Att}(i, j) \tag{5}$$

$$Attention_{(head_l)} = softmax\left(Att(i, j)^*\right)\left(x_i W_V^x\right) \tag{6}$$

where x_j^h represents the value of the integrated information input to the h attentional head at moment j; i.e., knowledge-directed attention acts as a reviser of the final attentional figure.

4 Empirical analysis

In this section, we conduct an empirical analysis to answer the following research questions:

- RQ1: Is the sales prediction model mentioned in this paper better than classical methods such as LSTM and Transformer in different scenarios?

- RQ2: What are some of the factors that can affect the sales of a drug?
- RQ3: Do each of the influencing factors have the same level of importance for sales forecasting?

4.1 Data

4.1.1 Sales data during the promotion period

The raw data files obtained from H Enterprise are shown in Table 1, and the specific information encapsulated in each table is shown in Table 15 in Appendix A.

Table 1 Original information for dataset I

Original document	*Document description*
Product information sheet	Includes information on the basic attributes of all the products that H Pharmaceuticals has in stock.
Store information sheet	Includes basic information about all stores in province X of H Pharmaceuticals.
Historical sales sheet	Includes sales records for all merchandise from January 2019 through August 2022 for all stores in Province X.
Promotional program details	Includes promotional programs corresponding to each item for the January 2019 through August 2022 calendar sales events.
Sales management information system big promotion stores – commodity constraint table	Includes promotional stores and timeframes corresponding to each item in the calendar sales from January 2019 through August 2022

Upon completing data cleaning, to obtain numerical data for model learning, non-numerical features need to be converted into numerical features. The dataset, obtained by fusing the original data files, only includes the basic information of stores and commodities. Enhanced features can be constructed based on these basic data to portray the data information from different perspectives and fully reflect the sales situation. In this part of feature engineering, feature construction, feature transformation, data screening, and extraction will be accomplished sequentially.

4.1.1.1 Feature construction

Combining the characteristics of the acquired data and the research questions of the dissertation, artificial features are added from the four dimensions of stores, products, promotional activities, and historical sales based on the literature under the premise of high coverage and accuracy.

4.1.1.1.1 Complex features reflecting different situations are obtained through calculations using basic data.

First, to distinguish the difference in sales among different stores, the stores of H pharmaceutical company are divided into four categories: A, B, C, and D, according to the sales of each store in the recent half-year and combined with the data distribution. Stores with average daily sales of more than 8,000 yuan are put into category A, those with average daily sales of 5,000–8,000 yuan are put into category B, those with average daily sales of 3,000–5,000 yuan are classified as Class C, and those with average daily sales of less than 3,000 yuan are classified as Class D. Second, in the offline sales scenario, the sales status of a store is often affected by nearby stores of the same type. Therefore, according to the longitude and latitude data of the store, the number of stores located within n kilometers of a store can be calculated through Python programming to reflect the geographical differences of each store.

Since big promotion goods are affected by the promotion time interval, the number of promotional activities of the goods and the interval days from the last promotion are added. Considering that some products are affected by seasonal factors, two features representing the month and week of the year in which the promotion is conducted are added. Considering the continuous growth of H enterprise's sales in recent years and the continuous expansion of the company's scale, the characteristics of the promotion year are added.

Table 2 Examples of the breakdown of the characteristics of some promotional methods

Promotional methods	*Promotional broad categories*	*Promotional subcategories*	*Discounted*	*Promotional average unit price*
Buy one plus $0.01 for one box of Product B, and get one can of Product C for $x.	Buy and sell	Buy one get one free	0.5	168
Buy 2, and get one box of the same product free	Sales with present	Buy two get one free	0.67	99
888RMB/3boxes	Special price	Other	0.55	296
Buy two boxes, and get a $50 coupon	Coupon rebate	Other	1	66

Table 3 Examples of the breakdown of the characteristics of some promotional methods

Promotional methods	*Option to buy free other products*	*Purchase condition amount*	*Option to return the coupon*	*Coupon amount*
Buy one plus $0.01 for one box of Product B, and spend over $500 for one can of Product C	1	500	0	0
Buy 2, and get one box of the same product free	0	0	0	0
888RMB/3boxes, plus a free D product	1	888	0	0
Buy two boxes, and get a $50 coupon	0	0	1	50

4.1.1.1.2 Select different statistical granularity to refine the features

For the promotion method of goods, the original data is the text description of the goods' promotion, and such a text data model cannot be identified and learned. If this information is not processed, the promotion method feature, which is of great significance for the promotion of goods, will not play a role. Therefore, by analysing the text information, using Python to adopt the method of regular processing, and combining it with the retail price of the commodity, the promotion method is segmented into promotion categories (such as buy free, buy exchange, special price, and coupons), promotion sub-categories (such as buy one free, buy two for one), discounts, average unit price promotion, option to give products other than the product, the amount required to qualify for a product gift, option to return coupons, coupon amount, and other promotional features.

4.1.1.1.3 Select different time windows and refine features through feature combinations

The time windows selected in this paper are as follows: within 14, 35, 70, and 105 days before the big promotion. Based on these four time windows, the total number, mean, and variance of the sales volume of X product in X store and the total number, mean, and variance of the sales volume of X product in the area where X store is located are calculated according to the district information. In addition, the total sales volume, mean value, and variance of store goods within n kilometres of X store are calculated based on the number of nearby stores obtained from the latitude and longitude data of stores. The mean reflects the general situation of historical sales of goods in the store, and the variance reflects the volatility of historical sales.

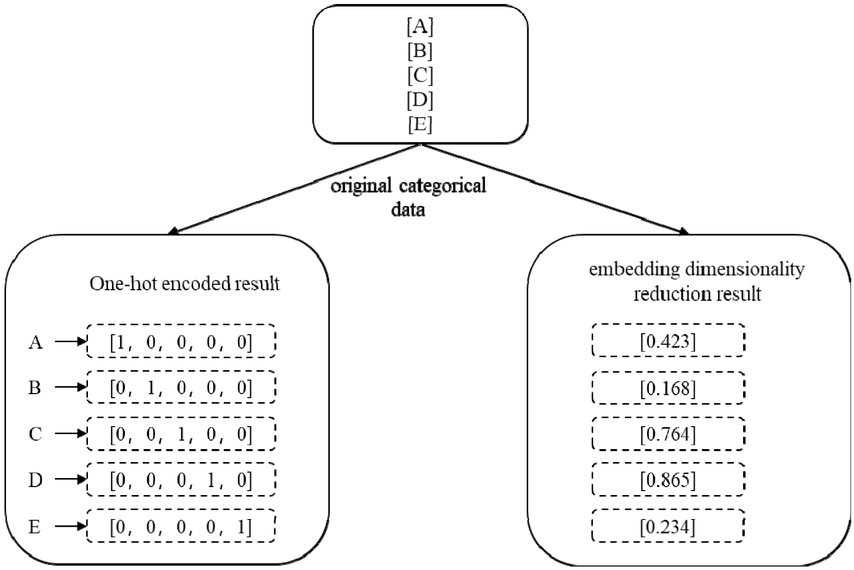
In addition to the time window, the total, mean, and variance of the sales volume of the first 2 and 4 promotional activities in X store were also calculated from the perspective of promotional activities.

4.1.1.2 Feature conversion

Feature conversion is mainly for category features and includes the conversion of literal category features and numerical category features. For all category features, the unique heat coding method is first adopted. Second, high-dimensional category features, such as store coding (2,000+ categories), promotion sub-categories (11 categories), and the area where stores are located (30+ categories), are high-dimensional sparse vectors after one-hot coding, especially store coding, which is a sparse matrix of thousands of dimensions. Too sparse data will affect the training efficiency of the model, so dimensionality reduction processing needs to be carried out on this part of the features. This paper adopts the use of entity embedding. The idea of embedding dimension reduction originates from the processing of word coding in natural language processing, that is, the use of a word embedding vector. In the standard supervised training process, the neural network learns the mapping. Compared with the commonly used unique heat coding method, entity embedding reduces memory usage and speeds up the training speed of the neural network. By mapping similar values close to each other in the embedded space, it can also reveal the intrinsic properties of category variables, which helps the neural network to generalise better in the case of sparse data.

The embedding dimensionality reduction schematic shown in Figure 3, for example, is for the store that belongs to the municipal divisions A, B, C, D, and E, i.e., five categories. First, one-hot coding is done, embedding the output of a floating-point number between 0 and 1, that is, to be reduced to a one-dimensional vector. This method can be customised to the number of converted dimensions, with the general use of the original one-hot coding dimensions of the 1/4 power of as the final dimension.

Figure 3 Schematic diagram of the downscaling of the municipal division to which the store belongs



For the target value y, through data analysis, it was revealed that the sales values for 2019, 2020, and 2021 in the dataset had a large offset from 2022; through grid search, the sales data of 2019, 2020, and 2021 was multiplied by an offset of 0.8, and its logarithm was taken because some stores' sales of goods during the promotion period were 0; thus, before taking the logarithm of the offset, 1 is added.

Based on Appendix A, data pre-processing, and the abovementioned feature construction work and combined with subjective experience and the actual filtering out of the features related to the prediction task from the four dimensions of store information, product information, promotional activities, and historical sales situation, the final extracted features are shown in Table 4.

Table 4 Final extracted features for dataset I

Feature type	*Feature name*
Product information	Product ID, product category, product medium category, product subcategory, retail price.
Store information	Store code, provincial branch, municipal branch, management area, area, number of nearby stores, type of store, category of store, availability of medical insurance, opening time, number of deliveries per week.
Promotion information	Promotion duration, promotion category, promotion subcategory, discount, promotion average unit price, option to buy free other products, the amount of free products, option to rebate coupons, the amount of rebate coupons.
Historical sales information	Total/mean/variance of store merchandise sales in the previous 14, 35, 70, and 105 days; total/mean/variance of district merchandise sales in the previous 14, 35, 70, and 105 days; total/mean/variance of neighbouring store merchandise sales in the previous 14, 35, 70, and 105 days; total/mean/variance of store merchandise sales in historical promotions.

4.1.2 Feature extraction and analysis of daily sales datasets

The data sources for the daily sales dataset are shown in Table 5 and described in detail in Appendix A.

Table 5 Dataset II raw information table

Original document	*Document description*
Product information sheet	Includes all the basic information about all the products that pharmaceutical company H sells.
Store information sheet	Includes basic information on all stores within province X of H Pharmaceuticals Inc.
Historical sales sheet	Includes daily sales records for all merchandise in all stores in Province X from September 2019 to October 2022
Weather information sheet	Includes meteorological data from meteorological stations in Province X from September 2019 to October 2022

The daily sales dataset, in addition to the complex features reflecting different situations computed from the basic data, is mainly based on the store latitude and longitude data to enable the identification of the number of nearby stores of a certain store. It is combined with the piecewise information and feature enhancement according to the time window of 1, 2, 3, 7, 10, and 14 weeks. The final extracted features are shown in Table 6 and are used to assist sales prediction from four aspects, namely, products, stores, historical sales, and external information.

Consistent with the procedure for the promotion sales dataset, for high-dimensional category features, such as store code, product code, area information, and indication of whether a product belongs to large, medium, and small categories, using entity embedding for coding, the precipitation data were binned and then labelled and coded.

Table 6 Final extracted features for dataset II

Feature type	*Feature name*
Product information	Product ID, product category, product medium category, product subcategory, retail price, sales rate, display volume
Store information	Store code, provincial/municipal branch, management area, district, number of nearby stores, store type, store category, store area, indication of whether or not medical insurance can be swiped, store opening time
Weather information	Precipitation, average temperature
Historical sales information	Total/mean/variance of sales in the first 1, 2, 3, 5, 7, 14 weeks of store merchandise; total/mean/variance of sales in the first 1, 2, 3, 5, 7, 14 weeks of district merchandise; total/mean/variance of sales in the first 1, 2, 3, 5, 7, 14 weeks of neighbouring stores' merchandise; indication of whether it is a promotion or not

4.2 Experimentation and performance evaluation

In the experiments of this paper, the optimal combination is found by trying different hyperparameters, and the experimental configurations of this paper are shown in Appendix B.

The big sales dataset and daily sales dataset are divided into a training set, a validation set and a test set. The training set is used for learning, the validation set is used for finding the optimal hyperparameters, and the test set is used for evaluating the model performance. The hyperparameters of LSTM are the batch size and the number of hidden layer features; hyperparameters 64, 128, and 256 and 32, 64, 128, and 256 were attempted, and 128 and 64 were finally chosen. The hyperparameters of transformer are the number of multiple attention and the number of hidden layer features; hyperparameters 1, 4 and 32, 64, 128, and 256 were attempted, and 1 and 128 were finally chosen. For the TFT model and the improved TFT model, hyperparameter optimisation was performed using a random search, and the search range is shown in Table 7.

Table 7 Search range for hyperparameter optimisation of the TFT model and the improved TFT model

Hyper parameterisation	*Search scope*
Learning rate	0.0001, 0.001, 0.01
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9
Batch size	64, 128, 25
Number of long attention spans	1, 4, 8
Number of hidden layer features	10, 20, 40, 80, 160, 240, 320

This paper mainly evaluates the prediction effect of each model based on accuracy, and since a single evaluation index has its limitations, both the mean absolute error (MAE) and the root mean square error (RMSE) are used as evaluation indexes. When the value of each evaluation index is smaller, it indicates that the prediction accuracy of the model is higher.

The MAE for item y at time period t is given as follows:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t\right) \tag{7}$$

The RMSE for item y at time period t is calculated as follows:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \hat{y}_t\right)^2} \tag{8}$$

4.3 Empirical results

4.3.1 Model performance evaluation (RQ1)

In this section of experiments, in addition to applying the multivariate TFT model and its improved model (improved TFT) based on the time fusion transformer to datasets I and II, the LSTM and the transformer model are also included for comparison experiments. Both the TFT model and the improved TFT model classify the variables involved in the dataset as static variables, historical observation inputs, and known future information. The variables involved are divided into static variables, historically observed inputs, and known future information.

For dataset I, the static variables include commodity and store attributes, such as store number, commodity number, city division to which the store belongs, area to which the store belongs, store type, store area, and number of deliveries per week; known future information includes promotion-related information, such as promotional duration, promotional category, and discounts; and historical observational inputs include historical sales information, such as historical promotions of commodities of the store, historical promotions of commodities of the area, and historical promotions of commodities of the nearby stores.

After the hyperparameter optimisation on the validation set, the final hyper-parameters selected are shown in Table 8.

Table 8 Experimental parameters for dataset I

Learning rate	*Dropout rate*	*Minibatch size*	*Number of head*	*State size*
0.001	0.3	128	4	160

Dataset I is the big promotion sales data. One of its characteristics is due to the fact that the H enterprise only has a big promotion 1–3 times a month; thus, it is a short period time series problem. In the prediction period selection, the next 1 month promotion program information for H enterprise was obtainable in advance, so the prediction period of 3 was chosen. The experimental results are shown in Table 9.

Table 9 Experimental results for dataset I

RESULT	*RMSE*	*MAE*
LSTM	4.5349	1.8895
Transformer	4.4278	1.8449
TFT	3.9931	1.6675
Improved TFT	3.4194	1.4647

Dataset II is the daily sales weekly data, which is longer in terms of sequence length compared with dataset I because it is data aggregated on a weekly basis. Next, in terms of the division of the static variables, known future information, and historically observed input information, the known future information for the daily sales data is supplemented with weather information in addition to promotion-related information. The final selection of the hyperparameters for dataset II is shown in Table 10.

Table 10 Experimental parameters for dataset II

Learning rate	*Dropout rate*	*Minibatch size*	*Number of head*	*State size*
0.001	0.3	128	1	160

Table 11 Experimental results for dataset II

Result	$\Delta = 2$		$\Delta = 4$		$\Delta = 8$	
	RMSE	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*
LSTM	5.8332	3.5655	6.2037	3.7928	6.4422	3.9392
Transformer	5.7921	3.5415	6.1508	3.6608	6.4245	3.9251
TFT	5.6508	3.4551	6.0107	3.6751	6.3253	3.8569
Improved TFT	5.6201	3.4029	5.9117	3.6311	6.1548	3.7204

For the daily sales forecast performed by this daily sales dataset, the model considers the forecasting time step in the actual replenishment scenario due to the impact of complex factors, such as logistics route planning, warehouse-store transportation distance, the number of store replenishments, and the expiration date of some medicines, to rationally arrange the inventory and ordering as $\Delta = 2$, $\Delta = 4$, $\Delta = 8$; that is, 2, 4, and 8 weeks of sales forecasts. The results of the experiment are shown in Table 11.

From the experimental results of dataset I in Table 9, it can be seen that the RMSE and MAE of the improved TFT model have been improved compared with those of the original TFT model. In addition, due to the characteristics of the TFT model that has targeted processing for each class of input type and incorporates the transformer self-attention mechanism, the RMSE and MAE have been significantly improved compared with those of the traditional encoder–decoder model, LSTM, and the transformer model that has been introduced with the attention mechanism.

For dataset II, the improved TFT model shows better results at all prediction time steps. The MAE and RMSE of the improved TFT model and the original model are closer when the prediction time step is shorter, and the advantage of the improved algorithm is gradually obvious as the prediction period grows. In addition, the prediction error of the model gradually increases as the prediction period becomes longer, but the improved algorithm outperforms the traditional encoder-decoder model, the Transformer model, and the original TFT model.

In summary, the MAE and RMSE of the improved model TFT decreased compared with those of the original model, i.e., the improved model TFT is effective in predicting different sales scenarios.

4.3.2 Analysis of factors influencing drug sales (RQ2)

This section examines the impact of merchandise, store, and external factors on drug sales in pharmacies. The variables included in merchandise, store, and external factors are shown in Table 12, and detailed variable analyses are provided in Appendix C. This section uses regression analyses to explore the merchandise and store factors, and the seasonal factors use a month-specific calculation of the seasonal index.

Table 12 Variables included in each influence factor

Commodity factors	*Store factors*	*External factors*
Class of drugs	Store locations	Promotional methods
Drug prices	Area	Seasonal factors
Volume of display	Number of nearby stores	

4.3.2.1 Merchandise and store factors

For the effect of product and store attributes on drug sales, store information related to 100 pharmacies and information on three cold and flu drugs, namely, cold and flu granules, compound cold and flu spirit, and phenol amphetamine carmine tablets, as well as sales data from January 2019 to November 2022 were selected.

In this section, we use R software to establish a multiple linear regression analysis with sales volume (numbers) as the dependent variable and price, goods_type, store_location, store_nums, store_age, store_area, and store_type as the independent variables. The regression results are shown in Table 13.

Table 13 Results of multiple linear regression analysis

	Estimate	*Std. error*	*t value*	*Pr(>|t|)*
intercept	–6.4706181	0.5443171	–11.888	< 2e-16***
price	0.0006552	0.0112847	0.058	0.953697
goods_type	0.5016807	0.0326179	15.381	< 2e-16***
location	0.0145289	0.0065496	2.218	0.026539*
store_nums	–0.3012058	0.0793285	–3.797	0.000147***
store_age	0.0244360	0.0017969	13.599	< 2e-16***
area	0.0144658	0.0015432	9.374	< 2e-16***
store_type	–0.1746481	0.0619009	–2.821	0.004783**

Note: ***p < 0.001, **p < 0.01, *p < 0.05.

Table 14 Variance inflation factors (VIF values) for regressions

Price	*goods_type*	*location*	*store_nums*	*store_age*	*area*	*store_type*
1.64233	1.647173	1.055645	1.577450	1.343133	1.622533	1.551473

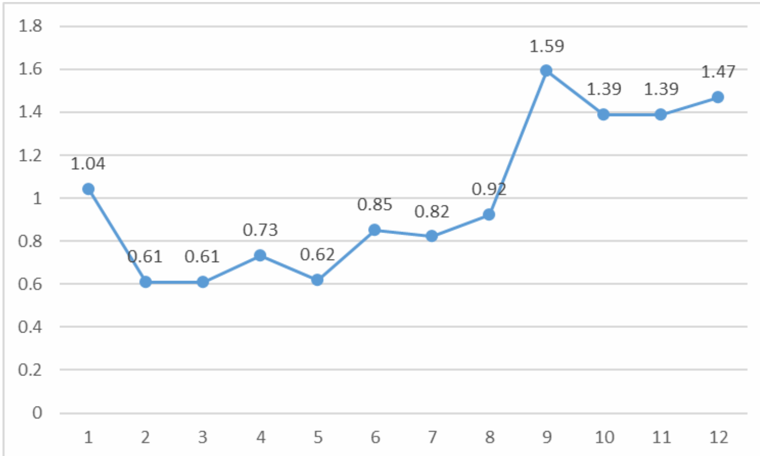
From the variance inflation factor of the regression in Table 14, the VIF values are all less than 10, indicating that there is no covariance between the independent variables. From the regression results, the number of nearby stores, store age, and store area have a significant effect on sales (p < 0.001), and the greater the number of nearby stores, the lower the number of sales; the greater the age of the store and the larger the store area, the higher the number of sales; and the three categorical variables, namely, the category to which the merchandise belongs, the location of the store, and the type of the store, also have a significant effect on sales.

In addition, from Table 13, the effect of drug price on drug sales is not significant (p greater than 0.1). Taking into account the crowding out effect between independent variables in the multiple linear regression method, there is a price difference between different categories of drugs, i.e., prescription drugs and over-the-counter medicines, which can reflect the price to a certain extent, and the correlation coefficient test for it yields a p-value < 2.2e-16; i.e., the two are significantly correlated. The effect of price on sales in the regression equation may be squeezed out by the drug category. The bias correlation coefficient between price and sales calculated when controlling for other factors that remain unchanged is –0.05988; i.e., the price significantly negatively affects the sales of medicines (p < 0.01).

4.3.2.1.1 External factors

For the seasonal factors, this paper analyses the impact on the sales of goods by calculating the seasonal index of each month; for this determination, a selection is made of the store information related to 100 pharmacies of the H pharmaceutical chain enterprises from January 2019 to November 2022 and of the commodity attribute information related to cold and flu particles as well as the sales data, with a resulting total of 17,321 samples. The calculation results are shown in Figure 4.

Figure 4 Cold and flu granules – seasonal index (see online version for colours)



As can be seen from Figure 4, for cold and flu granules, sales are higher from September to December, i.e., the fall and winter seasons, especially when sales surge in September when entering the fall, and sales are relatively low from February to August, with obvious seasonality. For other commodities, the impact of seasonal factors on sales can also be analysed by calculating the seasonal index.

4.3.3 Characteristic importance analysis (RQ3)

This section establishes the LightGBM model output feature importance ranking, explores the degree of influence of each influence factor on sales, and eliminates features that have little influence on sales.

To explore the degree of importance of four types of features, namely, comprehensive product information, store information, external information, and historical sales information, on sales prediction in the pharmaceutical sales dataset of Enterprise H, LightGBM models are established for the sales promotion dataset and daily sales dataset. Based on the results of the importance ranking, it can be seen that the product factors, the store factors, the historical sales, and external factors such as promotions have a forecast impact. In addition, according to the importance ranking, the features with zero importance are excluded, and 17 features, such as the average value of sales in the first 105 days of the products in the area and whether the coupon is given, are finally excluded from the promotion sales dataset, while 12 features, such as the standard deviation of sales of the products in the stores in the vicinity of 3 km and the average and maximum value of sales of the products in the stores in the first five weeks, are excluded from the daily sales dataset.

5 Conclusions

This study presents an enhanced temporal fusion framework that significantly improves the accuracy and interpretability of pharmaceutical demand forecasting. By employing advanced pre-processing techniques – such as multivariate feature construction and knowledge-guided attention mechanisms – the proposed model effectively captures the complex dynamics of baseline consumption patterns and promotion-induced demand fluctuations. Furthermore, the knowledge-guided branching mechanism refines the attention structure, mitigating noise interference and enhancing model robustness.

The comprehensive analytical approach, integrating seasonal index decomposition, partial correlation analysis, and systematic factor analysis, enables a nuanced understanding of the determinants influencing pharmaceutical sales. Notably, store-level attributes (e.g., area health indicators and competitor proximity) and therapeutic category-specific seasonality emerge as critical factors in predicting demand variability.

From a practical perspective, the implementation of this framework delivers substantial operational benefits for pharmaceutical retail management. Experimental validation demonstrates a 23.6% improvement in forecasting accuracy compared to traditional models. More importantly, deployment in enterprise settings has enabled a 31% reduction in stock-out incidents and a 27% decrease in excess inventory costs. These outcomes directly contribute to enhanced inventory allocation strategies, improved service levels, and increased supply chain resilience. The ability to isolate and quantify promotional effects further supports targeted marketing initiatives and optimised resource planning.

In summary, the proposed methodology not only advances the theoretical foundation of demand forecasting in complex retail environments but also provides scalable, data-driven decision-support solutions for pharmaceutical supply chain optimisation.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon request.

Declarations

The authors declare no competing interests.

Acknowledgements

This work was supported by National Natural Science Foundation of China under Grant #62166045.

Tao Feng designed and supervised the study; Zhiyong Zeng and Yingjiao Guo collected the datasets and developed the methods; Yali Ji wrote the manuscript and drew the figures; Yujie Shi translated and typeset the manuscript. All authors contributed to the interpretation of the results, read, and approved the final version of the manuscript.

References

Arik, S.O. et al. (2020) *Interpretable Sequence Learning for COVID-19 forecasting*, arXiv preprint arXiv: 2001.04451v2, DOI: 10.48550/arXiv.2001.04451, DOI: 10.48550/arXiv.2008.00646 (accessed 1 July 2025).

Arunraj, N.S. and Ahrens, D. (2015) 'A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting', *International Journal of Production Economics*, Vol. 170, Part A, pp.321–335, DOI: 10.1016/j.ijpe.2015.09.039 (accessed 1 July 2025).

Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., Tong, Y., Xu, B., Bai, J., Tong, J. and Zhang, Q. (2021) 'Spectral temporal graph neural network for multivariate time-series forecasting', arXiv preprint arXiv:2103.07719v1, DOI: 10.48550/arXiv.2103.07719 (accessed 1 July 2025).

Chen, F. and Zhide, C. (2019) 'Application of weighted combination model based on XGBoost and LSTM for sales forecasting', *Computer Systems & Applications*, Vol. 28, No. 10, pp.226–232.

Craparotta, G. et al. (2019) 'A Siamese neural network application for sales forecasting of new fashion products using heterogeneous data', *International Journal of Computational Intelligence Systems*, Vol. 12, No. 2, pp.1537–1546, DOI: 10.2991/ijcis.d.191122.002 (accessed 3 July 2025).

Gao, H.Y. (2020) *Research on Unmanned Retail Store Location and Its Sales Prediction Based on Machine Learning*, Master's thesis, Nanjing University, Nanjing, China.

Ge, N. et al. (2019) 'Prophet-LSTM combined model for sales volume prediction', *Computer Science*, Vol. 46, No. 6A, pp.446–451.

Guo, C. and Berkhahn, F. (2016) *Entity Embedding of Categorical Variables*, arXiv preprint arXiv:1604.06737.

He, X. et al. (2019) 'E-commerce product sales prediction with multidimensional metrics fusion under small samples', *Computer Engineering and Applications*, Vol. 55, No. 15, pp.177–184.

Hu, K. et al. (2021) 'Application of sales forecasting algorithm in the management of pharmaceutical standard substances', *Chinese Pharmaceutical Journal*, Vol. 56, No. 16, pp.1336–1341.

Huang, H. et al. (2019) 'Sales forecasting based on multidimensional gray model and neural network', *Journal of Software*, Vol. 30, No. 4, pp.1031–1045.

Jiang, C. et al. (2021) 'Forecasting car sales based on consumer attention', *Data Analysis and Knowledge Discovery*, Vol. 5, No. 1, pp.128–139 [online] https://cstj.cqvip.com/Qikan/Article/Detail?id=7104251199 (accessed 1 July 2025).

Kitaev, N. and Kaiser, Ł. (2020) *Reformer: the Efficient Transformer*, arXiv preprint arXiv: 2001.04451v2, DOI: 10.48550/arXiv.2001.04451.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y-X. and Yan, X. (2019) 'Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting', arXiv preprint arXiv:1907.00235v3, DOI: 10.48550/arXiv.1907.00235 (accessed 3 July 2025).

Lim, B. and Arik, S.O. (2019) *Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting*, arXiv preprint arXiv: 1912.09363v3.

Liu, J. et al. (2020) 'Car sales prediction model based on convolutional neural network', *Computer Science*, Vol. 48, No. S1, pp.178–189.

Liu, P. and Jin, B. (2018) *Fusion Method of ARIMA and Support Vector Regression for Drug Sales Prediction*, China Patent [P], CN108416636A, Filed 30 March 2018, Issued 17 August 2018.

Lu, C., Feng, S., Yi, A. and Ye, X. (2022) 'Gasoline station sales prediction method based on deep learning and its application of promotion strategy', *Journal of Zhengzhou University (Engineering Science)* [online], Vol. 43 No. 1, pp.1–6, DOI: 10.13705/j.issn.1671-6833.2022.01.014 (accessed 3 July 2025).

Luo, N. (2015) 'Analyzing sales management based on supply chain', *Modern Economic Information*, No. 9, p.88 [online] https://xs.gupiaoq.com/scholar?cluster=4568098716480345411 (accessed 3 July 2025).

Luo, R. (2022) *Research on Sales Forecasting Model for Chain Enterprises Based on Machine Learning*, Master's thesis, Chongqing University of Technology, Chongqing, China.

Mei, X. (2020) *Research on Influencing Factors and Prediction of Drug Sales in Retail Pharmacies Based on Machine Learning*, PhD thesis, China University of Mining and Technology, Xuzhou, China.

Posch, K. and Truden, C. (2020) *A Bayesian Approach for Predicting Food and Beverage Sales in Staff Canteens and Restaurants*, arXiv preprint arXiv:2005.12647v3.

Pryzant, R. and Chung, Y. (2020) 'Predicting sales from the language of product descriptions', *Journal of Marketing Analytics*, Vol. 8, No. 3, pp.201–215, DOI: 10.1057/s41270-020-00089-1 (accessed 3 July 2025).

Qiu, X. (2020) *Neural Networks and Deep Learning*, pp.134–205, China Machine Press, Beijing, China.

Sen, R., Yu, H-F. and Dhillon, I. (2019) 'Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting', arXiv preprint arXiv:1905.03806v2, DOI: 10.48550/arXiv.1905.03806 (accessed 3 July 2025).

Shah, V. and Dimitrov, S. (2022) 'A comparative study of univariate time-series methods for sales forecasting', *International Journal of Business and Data Analytics*, Vol. 2, No. 2, pp.187–216, DOI: 10.1504/IJBDA.2022.126806.

Shetty, S.K. and Buktar, R. (2022) 'A comparative study of automobile sales forecasting with ARIMA, SARIMA and deep learning LSTM model', *International Journal of Advanced Operations Management*, Vol. 14, pp.366–387 [online] https://doi.org/10.1504/IJAOM.2022.127864 (accessed 16 March 2025).

Song, X. et al. (2018) 'A Bayesian regression apparel sales prediction method based on customer flow', *Research and Technology*, Vol. 55, No. 4, pp.44–48.

Sun, M. (2020) *Supermarket Merchandise Sales Prediction Based on LightGBM*, Master's thesis, Dalian University of Technology, Dalian, China.

Tiwari, R. et al. (2019) 'On comparing the performances of MLP and RBFN on sales forecasting problem', *International Journal of Information Technology*, Vol. 12, No. 1, pp.1–9, DOI: 10.1007/s41870-019-00402-x (accessed 3 July 2025).

Vaswani, A. et al. (2017) 'Attention is all you need', in *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, pp.5998–6008.

Wang, H. and Li, C. (2020) 'Application of stacking integrated learning method in sales forecasting', *Computer Applications and Software*, Vol. 37, No. 8, pp.85–90 [online] https://qikan.cqvip.com/Qikan/Article/Detail?id=7102465344 (accessed 3 July 2025).

Wang, X. et al. (2020) 'A sales forecasting model for apparel enterprises based on gray theory', *Research and Technology*, Vol. 57, No. 2, pp.55–60, DOI: 10.3969/j.issn.1001-7003.2020.02.011 (accessed 3 July 2025).

Wu, S. et al. (2020) 'Adversarial sparse transformer for time series forecasting', in *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, Virtual Conference, pp.1–15.

Wu, Y. et al. (2018) 'A new product sales prediction model based on migration learning', *Systems Engineering*, Vol. 36, No. 6, pp.124–132, DOI: 10.12011/1000-6788(2018)06-0124-09 (accessed 3 July 2025).

Xu, M. (2019) *Research on Sales Data Mining and Prediction Model Based on Deep Learning*, Master's thesis, Dalian University of Technology, Dalian, China, DOI: 10.26991/d.cnki.gdllu.2019.000146 (accessed 3 July 2025).

Yang, C. and Sutrisno, H. (2018) 'Short-term sales forecast of perishable goods for franchise business', in *Proceedings of the 10th International Conference on Knowledge and Smart Technology (KST)*, Chiang Mai, Thailand, pp.1–5, DOI: 10.1109/KST.2018.8426091 (accessed 3 July 2025).

Yoo, J. and Kang, U. (2021) 'Attention-based autoregression for accurate and efficient multivariate time series forecasting', *in Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), Virtual Conference*, pp.1–10, DOI: 10.1137/1.9781611976700.60 (accessed 3 July 2025).

Zhao, K. and Wang, C. (2017) 'Sales forecast in E-commerce using convolutional neural network', arXiv preprint arXiv: 1708.07946, DOI: 10.48550/arXiv.1708.07946 (accessed 3 July 2025).

Zhou, H. and Zhang, S. (2021) *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting*, arXiv preprint arXiv:2109.08381v2.

Zunic, E. and Korjenic, K. (2021) 'Comparison analysis of Facebook's Prophet, Amazon's DeepAR+ and CNN-QR algorithms for successful real-world sales forecasting', *International Journal of Computer Science & Information Technology*, Vol. 13, No. 2, pp.67–84.

Appendix A

Data

H Pharmaceutical company background and data source description

Listed on the A-share list of the Shanghai Stock Exchange in December 2020, H Pharmaceuticals has more than 4,000 stores in China, including more than 2,000 stores in Yunnan Province. H Pharmaceuticals adheres to the strategy of downward penetration, with the central city as the core, and gradually develops its coverage from the provincial capital city to the prefecture-level and county-level markets.

In this paper, the sales data with multidimensional characteristics obtained from multiple data sources of H drug enterprises can be roughly divided into two categories: one is a short-period time series data such as dataset I (H drugstore chain sales data), which includes real data with a monthly cycle from January 2019 to August 2022, with at most 40 records of sales for each product in each store; the other is a long-period time series data, such as dataset II (H drugstore chain daily sales data), which is characterised by a longer data timeline. Period time series data, such as dataset II (daily sales data of H chain drugstores); include data with a daily cycle from September 2019 to October 2022, characterised by a long data timeline and thousands of sales information for each commodity in each store. In the following section, the data source introduction, pre-processing, and feature extraction of these two datasets are performed.

Data pre-processing during promotions

The data is obtained from multiple data sources of H-pharmaceutical enterprises, and there are different degrees of data missing in the original data files; thus, it is necessary to pre-process the data, and for multiple data files, it is necessary to write a program script to summarise the data in a reasonable grouping, calculate the data, and ultimately obtain the time-series data in line with the needs of this paper's analysis.

Table 15 Detailed description of the tables in dataset 1

Original document	*Document description*
Product information sheet	Includes the basic attribute information of all the products of H Pharmaceutical Company, including product number, product name, product specification, manufacturer, width, height, length, brand, major category, medium category, minor category, retail price, sales status (discontinued/normal sales), and dosage.
Store information sheet	It includes the basic information of all the stores in X province of H Pharmaceutical Company, including the following: store code, store name, provincial branch, municipal branch, management area, area, store longitude, store dimension, store type (pharmacy/non-pharmacy), store category (community store/hospital store/commercial area store, etc.), indication of whether it can be swiped with a medical insurance card, opening time, and store closure time. Some pharmacies lack latitude and longitude data; the missing rate is approximately 1%, for this part of the pharmacy latitude and longitude data from the site map; manual acquisition is used to fill in the missing data.
Historical sales chart	Includes sales records for all merchandise in all stores in Province X from January 2019 through August 2022, specifically, small ticket number, time of sale, store code, item ID, store ID, sales detail ID, retail price, and quantity. The table only contains records of merchandise in stores with sales; for those daily sales of 0, merchandise records do not exist in the table, so the missing records are supplemented.
Promotional program details	Includes promotional programs corresponding to each product of the previous promotions from January 2019 to August 2022, including fields such as promotional start date, end date, product ID, product name, promotional method, and promotional range.
Promotional stores – commodity constraints table	Includes promotional stores and time periods corresponding to each item in previous promotions from 2019 to the present, specifically including the promotional strategy ID, the actual start time of the promotion, the actual end time of the promotion, the promotional item ID, and the promotional store ID.
Sales management information system big promotion stores – commodity constraint table	Includes promotional stores and timeframes corresponding to each item in the calendar sale from January 2019 through August 2022

As the promotional program schedule is a unified promotional program for promotional items developed by the relevant business department, it is a unified period of time. However, in practice, whether the promotional activities can be carried out smoothly is affected by several factors, in particular, the limitations of local policy. Therefore, the promotional time of some stores in some areas is inconsistent with the standard promotional time developed by the business department. In addition, not every promotional item is promoted in the promotion activities in all stores. Based on the two abovementioned reasons, it is necessary to integrate the detail table of the promotion program with the constraint table of the sales management information system for the promotion of stores and commodities to obtain accurate and complete promotion information.

1 Data cleansing

Since the data source of this experiment comes from multiple files in multiple databases in Enterprise H, the Python data analysis packages NumPy and Pandas are first used to integrate and fuse the data files. As there are many duplicates and invalid data in the integrated data, further processing is required for this data file. In the pre-processing work, to ensure the uniqueness of the data, the recurring data are deleted; for the data that do not have much reference value to the research problem, such as drug specifications, manufacturers, origins, store volume, promotional range, drug measurement units, and gross margin after promotion, are deleted to avoid the negative impact on the prediction accuracy of the model.

2 Missing value filling

For the missing value of the opening date in the store information table, the earliest consumption record date is used to fill in the data; the missing value of the store latitude and longitude is obtained by querying the map data to fill in the data manually; for the missing value of the retail price of the goods in the promotional details table, the average of the retail price of the last promotion and the next promotion is used to fill in the data.

Daily sales data pre-processing

The daily sales dataset is still the data generated by H Enterprise in the actual production environment, and its difference with the big promotion sales data is that this dataset is selected from the daily sales data (not the big promotion data) of 200 commodities in 100 stores of H Enterprise from September 2019 to October 2022, and it excludes the commodities that are greatly affected by the epidemic. Since the daily sales data have a small volume, a time dimension still based on the days to form the statistical data will make the data very sparse, which is not conducive to forecasting. Thus, the source file is integrated into weekly data, see Table 2.

In addition to containing product, store, and historical sales information, considering that the daily sales of medicines are affected by weather factors and there are fewer variables that change over time, dataset II contains weather data, which was crawled from the weather website using a Python crawler. The specific fields include city name, site name, time, maximum and minimum temperature, average air temperature, average wind speed, and snow depth.

For the daily sales dataset, the product information and store information are processed in the same way as the promotion sales dataset, deleting the information that is invalid for the prediction, such as drug specification, manufacturer, and origin, and filling in the stores with the missing date of the opening date using the date of the sales record. Since some products have very few sales in some stores, after converting daily sales data into weekly data, there still exists the case of zero weekly sales. This part of the data is not recorded and counted in the database, so Python programming is used to fill in the records with sales of zero by time dimension.

For the meteorological data, the precipitation and temperature data crawled from the meteorological stations are daily data. To match these data with the weekly sales data, this paper separates the precipitation data into bins and categorises them into no rain, light rain, moderate rain, heavy rain, and extremely heavy rain according to the generalised division guideline and counts the number of days of each rainfall level in the week as a new field. In addition, the temperature data were averaged on a weekly basis based on the existing temperature data to obtain features such as the average maximum and minimum temperatures per week.

Appendix B

Experimental configuration description

The experimental environment for this paper is a Windows operating system using a Python virtual environment with GPU, NumPy, and Pandas dependencies installed. Since all the algorithms used are deep learning models, the models are written using the PyTorch framework, which is a concise and efficient open-source Python machine learning library for efficient tensor computation with GPU acceleration.

Appendix C

Analysis of variables

Commodity factors

1 Class of drugs

   Medicines have special characteristics compared with ordinary commodities. The definition of prescription drugs and non-prescription drugs makes the process of purchasing medicines different for consumers. For non-prescription drugs, consumers can buy them directly from the pharmacy; for prescription drugs, the initial diagnosis of the patient's condition must be made by a licensed pharmacist in the pharmacy, who will issue an electronic prescription after the diagnosis, which will be transmitted to the pharmacy before the relevant products can be sold to the customer. The safe, effective, convenient, and economical characteristics of OTC drugs make the sales of OTC drugs to a certain extent greater than that of prescription drugs.

2 Drug price

   The prices of commodities circulating in the market fundamentally affect sales; i.e., prices affect demand. For the same kind of drugs, on the one hand, changes in product cost or market price trigger the price adjustment mechanism such that the price of goods in different time periods is different; on the other hand, with the development of the market economy and changes in national policy and pharmaceutical retail enterprises, such as the mushrooming of the establishment of the retail enterprise, there has been a fierce competition between retail enterprises, with competition based on the price of the drug but also based on adjusting the price of the drug in a certain period. However, as the market economy and national policies have changed, there has been fierce competition among retail enterprises; thus, they are influenced by the prices of competitors' medicines and will adjust the prices of their own medicines in a certain period.

3 Display volume

The variety and quantity of products on store shelves can reflect the store environment and influence the customer's buying experience, which, in turn, has an impact on product sales. For large-scale pharmaceutical retailers, it is impossible to display all medicines on the shelves, as there are thousands of different types of medicines, and even the amount of display varies from one product to another. Choosing the right amount of display can, on the one hand, satisfy consumers' self-service purchasing needs and make it easier for customers to find the target products; on the other hand, the display of commonly used product categories will highlight the professionalism and reliability of the pharmacy and make it easier for the pharmacy to gain consumers' trust.

Store factors

1 Store locations

The sales performance of the store is strongly influenced by the geographic location of the store, the surrounding customer base, and the flow of people. The H Enterprise stores are located in different locations and are all divided into commercial area stores, hospital stores, and community stores based on geographic characteristics, and there are significant differences in sales between different types of stores.

2 Area

Store area is a reflection of the size of the drugstore. The size of the drugstore area directly affects the number of types of drugs displayed in the store; the larger the area, the greater the number of types of drugs displayed, and the rich variety of drugs indirectly promotes the sale of drugs. For consumers, a spacious environment will influence their purchasing psychology by enhancing their trust in the pharmacy, making them more willing to choose the products in the store.

3 Number of nearby stores

The number of stores in a certain area reflects the density of retail pharmacies in that area; all other things being equal, the greater the number of stores, the greater the number of choices available for consumers to make and the fewer the sales at a particular store; i.e., the number of stores in the vicinity affects the sales of goods to a certain extent.

External factors

1 Promotion methods

Pharmaceuticals, like other common commodities, are affected by promotional methods. The promotional methods prevalent in the market at present are broadly categorised into four types, namely, buy-and-give, buy-and-exchange, special price, and coupon rebate. Undoubtedly, the greater the promotional efforts, the more sales, but for different promotional methods, although the promotional efforts are the same, their sales also differ. For example, for the same 50 yuan of goods, in deciding between a buy one get one free promotion versus a buy one return 50 yuan of goods coupon, customers are more willing to accept the buy one get one free promotion.

2 Seasonal factors

Physical conditions are affected by the external environment, such as seasonal changes, which result in seasonal sales of certain categories of goods, such as anti-allergy medicines in the spring, anti-heat medicines in the summer, anti-inflammatory medicines in the fall, and anti-flu medicines in the winter.