mathematics

MDPI

Article

The Multivariate Fusion Distribution Characteristics in Physician Demand Prediction

Jiazhen Zhang *, Wei Chen and Xiulai Wang

School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 202311240009@nuist.edu.cn (W.C.); wangxiulai@126.com (X.W.)
* Correspondence: 202491000016@nuist.edu.cn; Tel.: +86-17302590255

Abstract: Aiming at the optimization of the big data infrastructure in China's healthcare system, this study proposes a lightweight time series physician demand prediction model, which is especially suitable for the field of telemedicine. The model incorporates multi-head attention mechanisms and generates statistical information, which significantly improves the ability to process nonlinear data, adapt to different data sources, improve the computational efficiency, and process high-dimensional features. By combining variational autoencoders and LSTM units, the model can effectively capture complex nonlinear relationships and long-term dependencies, and the multi-head attention mechanism overcomes the limitations of traditional algorithms. This lightweight architecture design not only improves the computational efficiency but also enhances the stability in high-dimensional data processing and reduces feature redundancy by combining the normalization process with statistics. The experimental results show that the model has wide applicability and excellent performance in a telemedicine consulting service system.

Keywords: deep learning; multi-attention mechanism; demand forecast; enhancement of data; long short-term memory network

MSC: 37M10



Academic Editor: MinJae Lee

Received: 4 December 2024
Revised: 7 January 2025
Accepted: 10 January 2025
Published: 11 January 2025

Citation: Zhang, J.; Chen, W.; Wang, X. The Multivariate Fusion Distribution Characteristics in Physician Demand Prediction. *Mathematics* **2025**, *13*, 233. https://doi.org/10.3390/math13020233

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Digital information services offer significant advantages and growth potential in domestic medical systems. These services are essential for hospital consultations, providing professional advice and staff lists via mobile devices and data platforms [1]. High-speed wireless technologies and tailored processing models support these advancements. A centralized database evaluates patient data to plan equipment allocation, ensuring the continuous supply of the necessary resources and uninterrupted care [2].

As the medical system evolves, integrating intelligent allocation architectures is crucial to optimize resource use. This enhances the efficiency, especially in areas with uneven service distribution. Remote consultations and real-time adjustments improve the healthcare effectiveness, ensuring better resource availability for patients. Consistent technology implementation strengthens healthcare optimization efforts, aligning resource allocation with patient needs [2,3]. The support of national policies and the steady development of efficient digital processing technologies have established a basic framework for an interconnected medical service system. This system serves as a critical module in the overall healthcare framework, enabling the efficient utilization of data. The differences and similarities in healthcare environments across regions influence the efficacy and output ratios of healthcare programs. As such, the design of healthcare programs must ensure

the adequate allocation of specific, accessible resources. Satisfaction with healthcare services has been shown to correlate directly with the rational and predictable allocation of resources [4]. Zhai, Qiao et al. [5] proposed a short-time manpower demand prediction model for telemedicine workforce allocation, highlighting the significant impact of the vacation effect. Lv, Su, and Yang [6] demonstrated that the allocation of primary healthcare resources is significantly related to medical visit rates and that resource scheduling aligned with temporal cycles is essential for efficient visit processes. However, regional constraints do not guarantee that the program can be migrated effectively to other healthcare units within the same stratum, as the factors influencing data collection and information storage vary significantly across China, weakening the generalization performance of the model.

In evolving healthcare systems, intelligent predictive programs are being developed to meet diverse assessment needs. An integrated intelligent architecture is essential for efficient resource allocation across regions. A connected medical service system, supported by national policies and advanced digital technologies, is a critical component of the healthcare framework [1,2]. This system enables efficient data use and resource allocation through remote consultations and real-time adjustments. The proposed predictive scheme aims to enhance resource utilization and improve healthcare effectiveness. This research focuses on two main directions.

(1) The formulation and optimization of a logical architecture and theoretical algorithms based on the algorithm itself. The autoregressive model is one of the research branches of traditional fitting linear equations, in which both the independent and dependent variables in a time series originate from the same series. The former comprises historical data, whereas the latter consists of current observed values. The ARIMA autoregressive equation theory examines the inferential direction and mapping relationship between these two variables. Romanuke [7] proposed incorporating two sets of fitting effect indicators to enhance the model. These indicators establish a threshold for the fitting of the fluctuation limits, resulting in smoother fitting outcomes; however, the horizontal interactions among the elements require additional investigation. Regarding deep learning, Ouyang H and Huang K et al. [8] extended the application logic of the long short-term memory (LSTM) network. The network architecture and test factor selection function well together in the framework when it comes to the factor learning problem, and the outcomes follow the law of the economic effect turnover. BiLSTM is a bidirectional computational model based on LSTM networks. The literature discusses whether the incorporation of two forms of information, the original order and reverse order, enhances the performance stability and generalization of information splicing in bidirectional models. Regarding hybrid modeling, Chan S et al. used a combination of deep convolution and fine-tuned machine learning techniques [9] for a multifactor forecasting initiative regarding power usage, minimizing the distinctions in their convergence methods. Zou et al. [10] classified unlabeled prediction data from a clustering perspective, segmented the prediction intervals, identified the prominent temporal center of gravity of the sequence features, aggregated similar time periods inside the center of gravity region, and classified the predicted credit scores. Zhang et al. [11] used the mRMR feature extraction algorithm to diminish the redundancy of features, minimize the redundancy percentage, and preserve high-efficiency application features.

(2) Scenario-specific prediction optimization according to the principle of factor correlation. Zhang and Ma [12] developed a metro traffic prediction model by considering the time and space information found in historical data. Their findings show that the spatial and temporal distribution influences the subject's flow trend and determines the direction of traffic movement. Meanwhile, Zhang, Liu, and Li [13] and Wang,

Jia, and Guo [14] measured the numerical relationships between various factors and integrated them as much as possible. Yu, Ge, and Jin [15] introduced climate factors and quantitatively derived multivariate correlation matrices to identify potential correlation information. SMGformer [16], proposed by Wang, W. and Gu, M. et al., integrates STL decomposition and multi-head self-attention to construct inputs from period, trend, and residual features. It connects the informer encoding layer with BiGRU, enhancing the model's ability to capture sequence dependencies. This leads to improved accuracy in multi-step runoff predictions, highlighting the effectiveness of multi-head self-attention in handling multivariate complex sequences. Indrajeet Kumar et al. [17] proposed a deep learning model that dynamically modulates the degree of attention allocated to features at different temporal points via an attention mechanism. This study employs attention weights as the primary representation of time series data, leveraging the temporal relationships between data points to enhance the predictive performance. Facilitated by the attention mechanism, the representation of specific features greatly supports the prediction task.

Following the aforementioned current model research cases and applications, a few additional significant issues can be summarized. In data integration, statistical indicators have been a common method of processing. The mathematical distribution characteristics can be used to detect the integration of a variety of indicators. However, the quantitative outline does not provide a unified multivariate prediction. The performance of the model can be improved by single-channel statistical indicators. Deep learning, as a time series prediction tool, is more inclined to gradually train data by reasoning about sample trends. Certain models employ the Gaussian mixture distribution principle [18] for data processing. However, the absence of a probabilistic factor fusion feature generation technique, coupled with frequent empirical parameter adjustments for varying data volumes, may diminish the generalizability. Owing to the constrained, asynchronous, and unique characteristics of healthcare systems globally, along with the evident high capacity and robust compatibility of big data servers [13], generalized prediction algorithms must confront the integration and augmentation of consensus attribute sets to achieve time series forecasting while ensuring data equity.

With the onset of the rapidly advancing artificial intelligence era, deep learning is increasingly focused on training data, as they progressively reflect the trends in sample changes. As previously discussed, the core mechanism of the LSTM algorithm is to preserve the short-term distribution law of the subsample [19]. This was integrated with prior fitting patterns during iterative training to merge the current distribution domain while effectively eliminating incompatible tail values. In 2017, the attention mechanism algorithm [20] emerged as a focal point in deep learning research, demonstrating the significant, interactive fusion of multi-scale subspace potentials. In particular, the directional aggregation facilitated by the multi-head self-attention mechanism [21] plays a key role across various domains, including text, sequences, and images.

The above shows that statistical theory and deep learning regarding the distribution law of the learning purpose are consistent. If the training of the consensus collection of information can generate statistical indicators and subsequently be utilized in the fusion module of the same type of model, it mitigates the non-homogeneous characteristics arising from different inference logic based on the same data source, thereby enhancing the accuracy of reasoning in relation to actual demand quantities.

In response to the above problems, the following work is carried out.

(1) The proposed lightweight time series prediction model, which integrates a multi-head attention mechanism with generative statistical information, demonstrates significant advantages in nonlinear modeling, data adaptability, computational efficiency,

and high-dimensional feature processing. This model provides an enhanced and reliable solution for time series prediction tasks. Specifically, it employs a variational autoencoder to generate internal statistical measures from sample data, which are subsequently concatenated with the original dataset. This integrated approach, combined with LSTM units, effectively captures complex nonlinear relationships and long-term dependencies. The dynamic feature importance adjustment capability of the multi-head attention mechanism addresses several limitations inherent in traditional algorithms. By surpassing the constraints imposed by autoregressive models, which rely on linear assumptions and stationarity, the model enhances its capacity to manage complex nonlinear data through the application of attention mechanisms.

(2) The variational autoencoder reduces the model's dependence on large-scale, high-quality datasets, thereby improving its robustness in the presence of noise and intricate data relationships. The lightweight architectural design of the model enhances the computational efficiency while mitigating the complexities and optimization challenges associated with mixed modeling. Additionally, the integration of normalization processes with statistical information allows the model to reduce feature redundancy and improve its stability in high-dimensional data processing, thereby mitigating the susceptibilities of traditional clustering and mRMR feature selection methods to noise and parameter sensitivity.

(3) Three input influencing variables of the model were determined from the big data service planning level: weather conditions, search traffic values of the day, and shift breaks. This allows us to demonstrate that the data variables are applicable to a wide range of multi-level medical consultation service systems. The experimental results demonstrate that the generative prediction model's performance clearly surpasses that of the multi-class model when comparing the three sets of influencing factors, historical demand values, and mixed statistical factors of the combined sample multivariate sequences. Simulation experiments were conducted to evaluate the efficacy of the dynamic resource allocation model in comparison with the static resource allocation model for medical communication resources, specifically focusing on bandwidth utilization. Two distinct health-related datasets were employed to validate the model's adaptability to diverse medical information.

2. Related Research

This paper describes the generative principles and realizable ideas for the exploration of indicator features based on the internal sequences of sample units, introduces the application of the multi-head self-attention module in multivariate time series prediction, and explains the multi-faceted influences that make the three medical treatments cited in this paper of strong generalizable value.

2.1. Selection and Interpretation of Multi-Perspective Pervasive Factors in Healthcare

The factors selected in this study correspond to the three macroscopic elements of medical manpower prognosis: climate, attention, and humanistic law. These three categories pertain to non-identical perspective global interference factors [22], which mutually impact and constrain one another, lacking the usual linear connection. Specifically, alongside extreme climatic phenomena, the gradual change in the weather does not significantly influence the other two factors, whereas the generation of search traffic is not solely constrained by the immediate group environment, being only directly related to historical conditions and the current policy environment.

One of the advantages of employing converged technologies to analyze ubiquitous information in short-term healthcare big data is that the integration of fusion features is

essential in optimizing LSTM algorithms when processing heterogeneous and complex data sources. By selecting the most pertinent features, this approach minimizes noise and reduces the dimensionality, thereby enhancing the training efficiency, prediction accuracy, and model robustness. Consequently, LSTM models can concentrate on critical information, effectively capture long-term dependencies, and sustain high performance even in noisy or diverse data environments [9].

2.2. Regularization Techniques for Internal Sequence Patterns in Physician Demand Data

From a data analysis perspective, basic data normalization is an effective method in addressing multi-attribute scale differences in samples, a process closely aligned with the batch processing layer in neural networks. For example, the initial batch normalization [23] layer was designed to unify the scale of batch data, normalizing each feature to the range (0, 1) and decoupling feature channels. However, this decoupling limits the ability to capture temporal correlation patterns in medical demand data, which often exhibit sequential dependencies. To address this, layer normalization [24] was introduced, allowing normalization at the individual sample level and demonstrating significant success in sequence-based learning tasks [25]. This approach facilitates the preservation of temporal dependencies, which is critical in improving the precision of healthcare demand forecasting models. Moreover, the embedding of input features as measurable numerical units enables the extraction of internal metrics that can enhance the resource allocation accuracy. In this context, advanced models such as Bert, a large-scale non-generative coding model, have shown compatibility with lightweight self-encoders through guided model distillation [26], offering a pathway for the generation of latent vector representations of medical demand patterns. Similarly, variational autoencoders (VAEs) [27], as self-coding models, provide a mechanism for the capture and expression of statistical information through output sampling distribution parameters, which can be leveraged to refine the predictions in healthcare systems. These methods underscore the potential of integrating advanced normalization and coding architectures to improve the granularity and accuracy of medical resource forecasting.

2.3. Short-Term Forecasting of Multivariate Series for Physician Demand Using Multi-Attention Mechanisms

The attention mechanism's theoretical framework, akin to comparative reasoning, plays a crucial role in capturing sequence vectors by using a corpus as a reference for semantic recognition or imitation learning. This approach is particularly valuable for healthcare demand forecasting, where accurate and reliable data are essential to minimize unnecessary constraints and biases. By reducing the overreliance on reference data, the self-attention mechanism enhances weight bias learning through iterative updates of weights based on previous output vectors, allowing the model to focus on patterns within the training set. This adaptability is key to improving the precision of predictions in dynamic and resource-intensive healthcare environments.

The multi-head attention mechanism [21], by analyzing time series variables from multiple perspectives, further refines the forecasting process. It aggregates local patterns into global trends, identifying critical segments where weight extremes overlap. This process minimizes the discrepancy between the fitted subsegments and actual values, enabling the more effective modeling of the temporal dependencies inherent in healthcare demand data. By optimizing the alignment of the predicted and observed values, the multi-head attention mechanism significantly enhances the accuracy and reliability of medical resource allocation models, supporting informed decision-making in healthcare systems.

3. A Medical Demand Forecasting Model with a Multi-Head Attention Mechanism Embedded with the Distribution Characteristics of Enhanced Mixed Factors

3.1. Problem Definition

Short-term physician demand forecasting and evaluation: This study uses time series and historical demand data, transformed for supervised learning. A stacked LSTM network with an attention mechanism is used for short-term demand prediction. Statistical metrics from normalized sequences generated by a latent variable model serve as supplementary features. The input includes multi-step historical demand data and associated factors from multivariate analysis, concatenated into a connection sequence. This sequence, along with its statistical indicators, forms the network's input vector. The theoretical framework is shown in Figure 1.

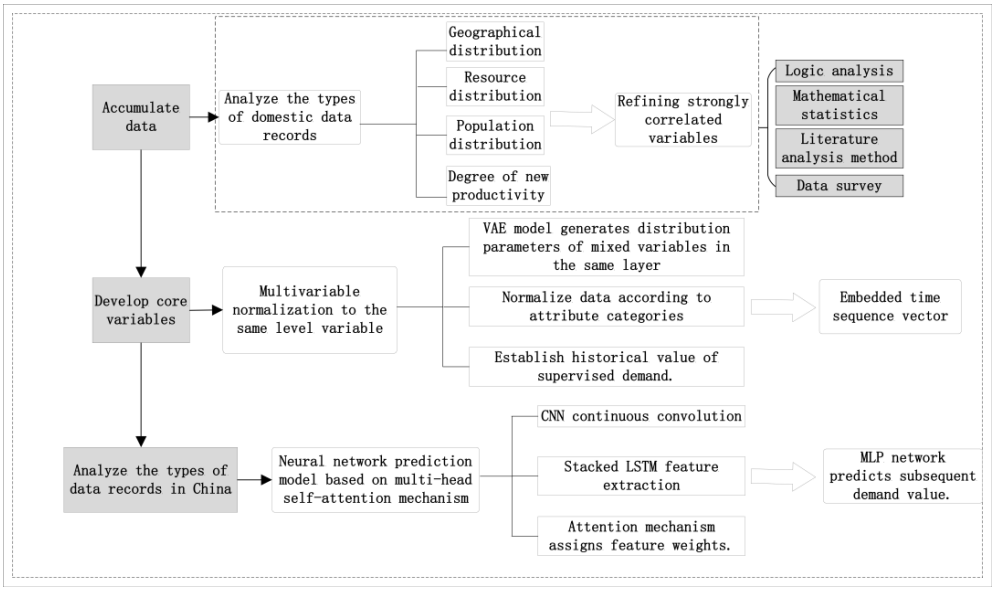


Figure 1. Theoretical logic structure.

3.2. Data Processing

The total physician demand on the same day is designated as the target variable. The sequence features encompass the weather conditions in the vicinity of the hospital, search volume trends, the holiday status, and the prior day's physician demand. These features encapsulate critical factors that influence actual medical demand patterns, thereby enhancing the accuracy of forecasting and resource allocation [12–15]. The weather data are a text sequence, which is processed into a numerical vector by single heat coding technology. All numerical sequences are normalized according to distinct properties, as indicated in the following formula:

$$X = \frac{X\textbf{original} - X\textbf{original.min}}{X\textbf{original.max} - X\textbf{original.min}} \quad (1)$$

where X**original** represents the original elements' integrated sequence points, X denotes the normalized elements' integrated sequence points, X**original.min** is the minimum value in the sequence points, and X**original.max** is the maximum value in the sequence points.

3.3. Model Design Methods

We propose a lightweight time series prediction model with a multi-head attention mechanism to enhance medical demand forecasting [28]. As illustrated in Figure 2, the framework consists of a generation module, stacked multi-LSTM layers, and an attention mechanism. The model leverages variational autoencoders to extract statistical measures and integrate them with multivariate sequences, thereby effectively capturing nonlinear relationships and long-term dependencies in medical needs. Additionally, the dynamic adjustment capabilities of the multi-head attention mechanisms improve the adaptability to noisy or incomplete data. The lightweight design ensures computational efficiency, reduces the dependence on large datasets, and enhances the stability, making it a robust and practical solution for real-time healthcare resource forecasting. The functions of these three modules are elaborated upon in the following sections.

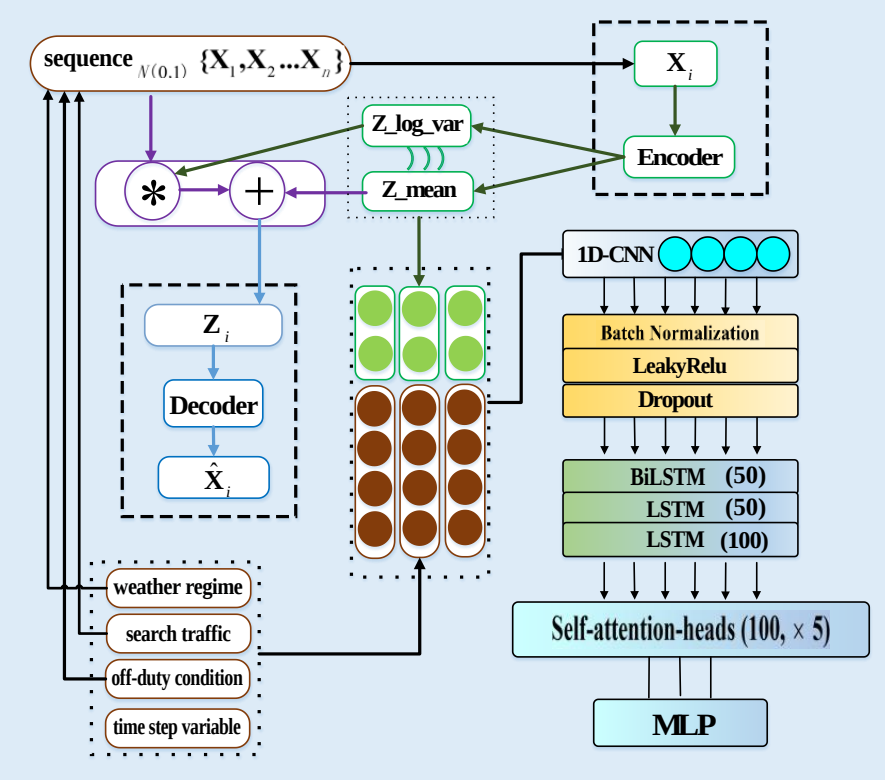


Figure 2. Lightweight medical demand forecasting model embedded with enhanced mixed factor distribution characteristics.

3.3.1. The Design of the Generation Encoding Method

In this paper, the VAE is used to integrate multiple medical information sources (historical demand values and influencing factors) and extract statistical indicators. The two substructures of the VAE are symmetrically coupled, while the encoder of the VAE [27] fits the sample x with the average value of the sequence z_mean and the logarithmic variance z_log_var one by one, and the two types of statistics are used as the sampling parameters for the decoder to calculate the sample distribution corresponding to y. The VAE loss function is shown in Equation (2).

$$\mathcal{L} = -\mathbb{E}_{z\sim q}\left[\log p(x|z)\right] + D\left[q(z|x) \parallel p(z)\right] \tag{2}$$

The former term of the loss function is the reconstruction error between the latent variable $\mathcal{L}$ of the sample data and the actual sample x, and the latter term is the KL divergence.

The similarity between the posterior probability distribution $q(z|x)$ and the sampling generation distribution $p(z|x)$ is calculated to mitigate the issue of fluctuation discrepancies in the spatial distribution of the latent variable that arise from model overfitting.

As shown in Figure 3, the encoded influencing factors are incorporated into a normalized sequence of sample units. By minimizing the loss function during the training process, the encoder output vectors, which represent distribution parameters, are extracted as statistical features of the underlying variables and integrated with labeled data characteristics. The VAE's strength resides in its capability to model the underlying data distribution, generate comprehensive statistical indicators, and capture implicit information.

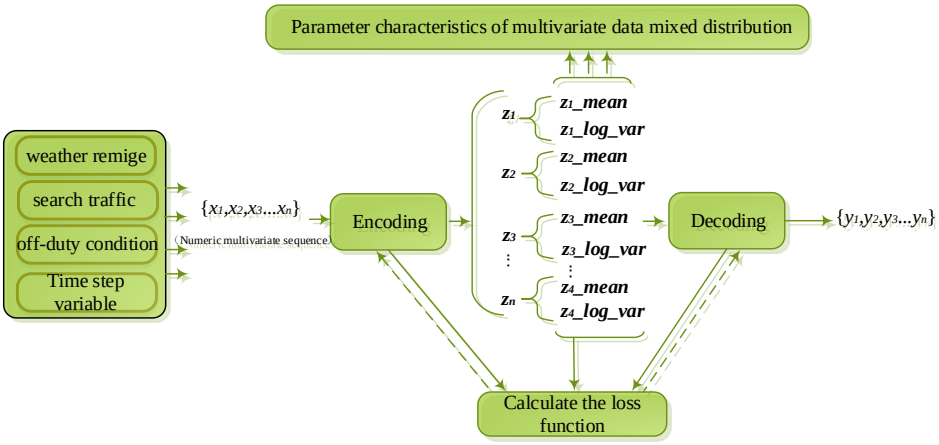


Figure 3. Generation model for statistical parameters of mixed distribution based on VAE.

3.3.2. Convolution Stacked LSTM Network

A convolutional stacked LSTM network is designed to capture the fusion of time series and statistical indicators of discrete factors that affect the change in the target. This framework enables accurate forecasting and resource allocation for medical demands, as shown in Figure 4.

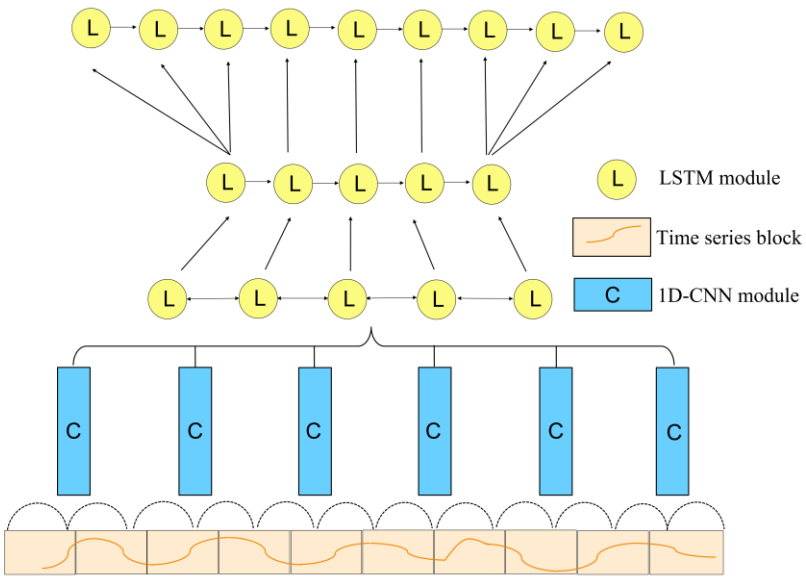


Figure 4. Convolutional stacked LSTM network architecture.

The convolution filtering technique [29] effectively manages dispersed multivariate sequences, making it ideal for the analysis of complex healthcare demand patterns. It excels

in capturing correlations within overlapping regions, enabling the detection of subtle variable interdependencies. These capabilities are crucial in improving the accuracy of medical resource allocation models by understanding temporal, environmental, and operational relationships. This relationship is expressed mathematically in the following formula:

$$f_{\mathbf{cnn}}(x_t) = \sum_{\mathbf{k}=1}^{\mathbf{K}} W_k x_{t-k+1} \tag{3}$$

W_k is a filter coefficient, and its product with sequence point x_{t-k+1} is a sub-convolution element, while the cumulative product of continuous sub-convolutions is the convolution sum of one unit x_t.

The stacking module computes convolutional features using a sliding window in the 1D-CNN layer, preserving the continuity through overlapping windows. These features are then passed to a bidirectional LSTM layer to capture periodic fluctuations in the physician demand. To prevent convergence issues, the model expands the feature space with historical data and uses two unidirectional LSTM layers to increase the neuron count. Finally, the output vector from the three-layer network is embedded as a feature potential vector in the prediction module.

3.3.3. Multi-Head Attention Mechanism Prediction Module

The stacked network faces challenges with strong feature fusion and high complexity in demand forecasting. To address this, internal interaction information is ranked before the output layer prediction. The multi-head attention mechanism includes Q (query), K (key), and V (value), derived from input vector X with different weights. The product of Q and K (Equations (4)–(6)) shows the similarity between the query and key–value vectors.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

$$head_{\,i} = \text{Attention}\left(XW_i^Q, XW_i^K, XW_i^V\right) \tag{5}$$

$$\text{MultiHead}\ (Q, K, V) = \text{Concat}(\ head_{\,1},\ head_{\,2}, \ldots,\ head_{\,h})W^O \tag{6}$$

Multi-head attention generates a weight vector β_i, as shown in Equation (7):

$$\beta_i = \frac{\exp(e_i)}{\sum_{k=1}^{m} \exp(e_k)} \tag{7}$$

e_i represents an individual component of the multi-head attention combination matrix, corresponding to the weight coefficient associated with the LSTM output for demand prediction information. The components x_i and β_i of the LSTM feature information X as the input vector of the attention mechanism are weighted and summed to obtain a multi-angle feature extraction set, as shown in Equations (8) and (9):

$$a_i = \sum_{i=1}^{m} x_i \beta_i \tag{8}$$

$$A = [a_1, a_2, a_3 \ldots a_n] \tag{9}$$

The output vector A of the attention module is taken as the input vector of the MLP. The MLP obtains the prediction result through a fully connected layer and completes the model task.

4. Experiment and Result Analysis

4.1. Dataset

Based on the data of a tertiary hospital in Nanjing, Jiangsu Province, this study records the off-duty circumstances, the information flow from Baidu search hospital consultations on that day, and the weather conditions around the hospital (types: sunny, cloudy, rainy, and snowy) during the experiment. The dataset encompasses the period from 1 December 2022 to 1 March 2024, comprising 457 data points. The binary random forest algorithm and extreme value screening method were used to clean the anomalous data, resulting in the utilization of 436 data points. Fifty-nine data points at the conclusion of the dataset were designated as the test set, without altering the order of the data.

4.2. Experimental Setup

The CPU model in all of the experiments involved in this study was an Intel i7-12700H, the programming language was Python 3.6, and the deep learning environment was the TensorFlow 2.0 advanced programming platform. Several algorithms were implemented based on the scikit-learn algorithm library. The training batch size, number of round epochs, unit settings of the convolution module and LSTM module, and number of heads of the attention mechanism of the model were taken as super parameters. See Table 1 for the parameters; the output unit and number of heads of LSTM were set to prepare a set for the comparative analysis of the core parameter relationship.

Table 1. Experimental parameter settings.

Hyperparametric Variable	Parameter Setting
Batch_size	32
Epochs	300
CNN_filters	8
CNN_kernel_size	5
CNN_strides	1
BiLSTM_units	50
First_LSTM_units	50
Second_LSTM_units	[90,100,110,120]
Multi_Head_Attention_heads	[2,3,4,5]

4.3. Predictive Performance Evaluation Indicators

The root mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), and root mean square percentage error (*RMSPE*) are commonly used to evaluate the prediction performance of time series, and the formulas are as follows:

$$RMSE = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(y_i - \hat{y}_i)^2} \tag{10}$$

$$MAE = \frac{1}{I}\sum_{i=1}^{I}|\hat{y}_i - y_i| \tag{11}$$

$$MAPE = \frac{1}{I}\sum_{i=1}^{I}\frac{100 \times |\hat{y}_i - y_i|}{y_i} \tag{12}$$

$$RMSPE = \sqrt{\frac{1}{I}\sum_{i=1}^{I}\left|\frac{y_i - \hat{y}_i}{y_i}\right|^2} \times 100 \tag{13}$$

Among them, I represents the sample size, $\hat{y}_i$ denotes the predicted value of the model at time i, and y_i represents the actual observed value at time i. The four categories of indicators reflect the forecast accuracy, i.e., the overall error magnitude, mean absolute error, relative error proportion, and root percentage error.

4.4. Multi-Factor Significance Verification

A larger *F-value* indicates a larger contribution to the model. We calculated and compared the *F-values* and *p-values* of each factor. In this study, we used statistical analysis techniques to determine the validity of the selected factors. The *F-value* is a measure of a factor's ability to explain changes in the model, while the *p-value* assesses its statistical significance. A *p-value* less than or equal to 0.05 is considered statistically significant [30]. See Table 2 for the analysis results.

Table 2. Significance analysis of influencing factors.

Factor	*F-Value*	*p-Value*
Weather regime	3.958	0.00161
Search traffic	10.73	0.00114
Off-duty conditions	4.981	0.0261

Among them, the *F-value* of the search flow is the highest, which is 10.73, and the individual significance of the weather regime and off-duty conditions is weak, indicating that they should be used together. In terms of the *p-values*, the *p-value* of the weather regime and search traffic is less than 0.001, and the *p-value* of off-duty conditions is less than 0.05, which proves that the three factors can be accepted as significant independent variables.

4.5. Result Analysis

4.5.1. Influence of Stacked LSTM Output Dimension on the Number of Attention Mechanism Heads

The output vector dimension of stacked LSTM is a key parameter in measuring the attribute density of time series, because the sparseness of potential variables will greatly hinder the efficiency of information transmission between the head mechanism when the attention mechanism decomposes the sequence vector program in parallel.

As can be seen from Table 3, the optimal matching parameters for the number of neurons are [90,100,120] according to the ascending order of the number of concerned heads [2,3,4,5]. By observing the prediction results of the test samples corresponding to each parameter unit in Figures 5–8, it is found that the smaller the *RMSE* error indicators, the better the overall fitting effect. Among them, the *RMSE* value of the parameter unit (2, 90) is 6.279, and the *RMSE* value of the parameter unit (5, 100) is 6.576, which are close. Therefore, other variables must be set to determine the parameter configuration.

Table 3. *RMSE* comparison of parameter combination schemes and attention level.

Attention Level Header Number	Number of Neurons in LSTM Output Layer			
	90	**100**	**110**	**120**
2	6.279	6.703	7.737	8.781
3	8.027	\	\	8.338
4	\	8.113	\	7.299
5	7.994	6.576	7.700	9.576

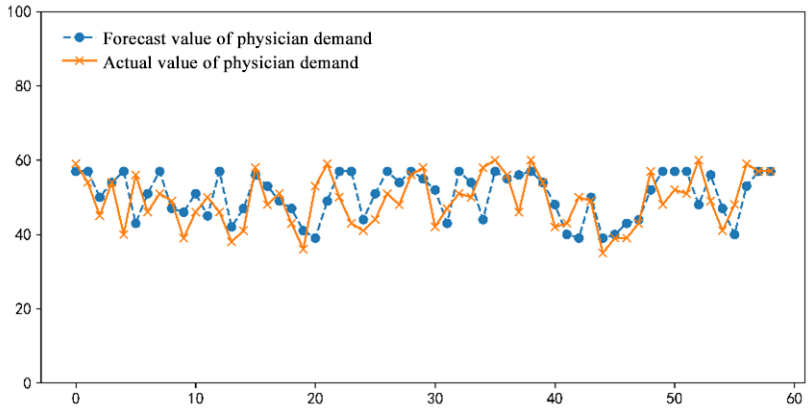


Figure 5. Prediction results of CNN–Multi-Head Attention–Multi-LSTM (heads = 2, dims = 90).

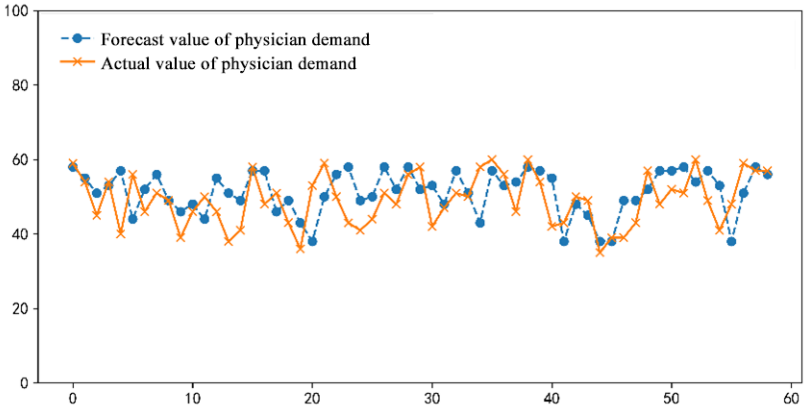


Figure 6. Prediction results of CNN–Multi-Head Attention–Multi-LSTM (heads = 3, dims = 90).

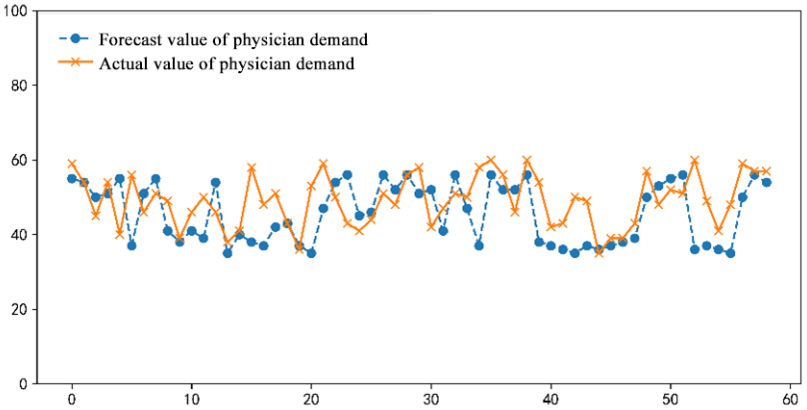


Figure 7. Prediction results of CNN–Multi-Head Attention–Multi-LSTM (heads = 4, dims = 120).

Deep learning converges the structural weights of a model based on a specified number of training rounds, and the stability of the convergence process is directly proportional to the stability of the model. Figure 9 shows the change in the model loss error under the condition of setting two sets of parameter units with the same number of training rounds. Here, we refer to the model set by unit (2, 90) as Model-1 and the model set by unit (5, 100) as Model-2. The position marked with the black circle in the figure shows that, within the range of the number of rounds (0:50) and (250:300), the loss transformation amplitude of Model-1 is large, and the peak value is high when it falls in the early stage; thus, it is unstable when it changes, and the loss difference fluctuates before the end of training. Model-2 exhibits a small fluctuation and stability in the convergence stage, and the model

weight plays a stabilizing role. Based on this parameter, the performance reliability of the composite frame is stronger; therefore, the number of headers is 5, and the output dimension of the LSTM is 100.

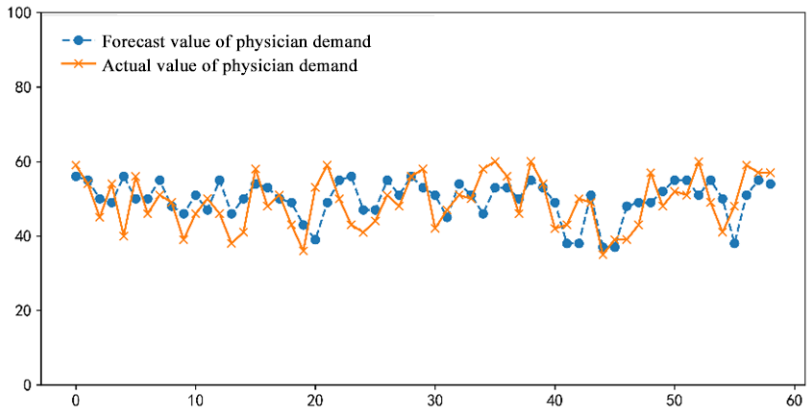
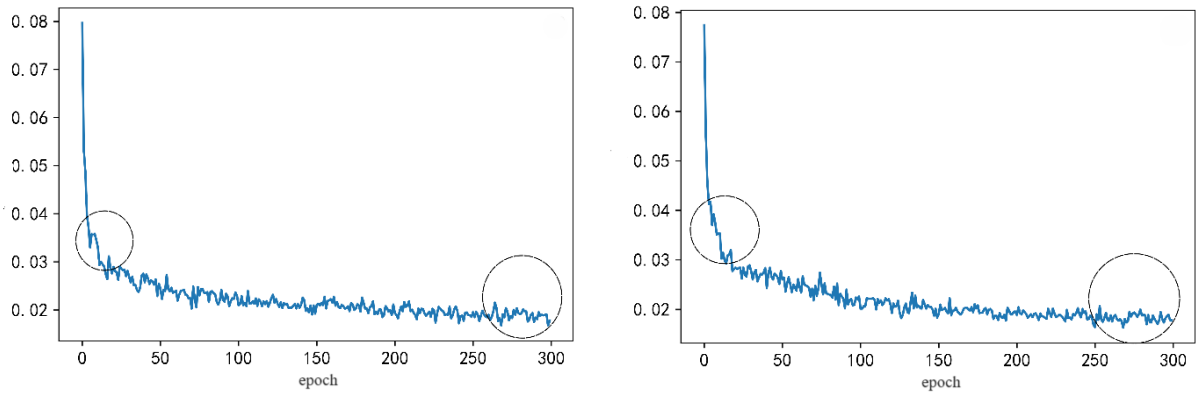


Figure 8. Prediction results of CNN–Multi-Head Attention–Multi-LSTM (heads = 5, dims = 100).



(**a**) Training set error variation chart (heads = 2, dims = 90)

(**b**) Training set error variation chart (heads = 5, dims = 100)

Figure 9. Comparison of training errors of undetermined parameter combinations.

4.5.2. Analysis of Validity of Multivariate Fusion Distribution Characteristics

To further validate the feature advantages of VAE-generated indicators and multi-sequence fusion, we conducted performance comparison experiments using five distinct CNN–Multi-Head Attention–Multi-LSTM models with varying feature input vectors. These experiments were built upon the parameter verification detailed in Section 4.5.1. The performance metrics used were the *RMSE* and *MAPE*. The feature types included the following: historical relative numbers of one-step time variables, influencing factors (excluding demand), multivariate data combinations, one-step time variables along with VAE statistics, and multivariate data combined with VAE statistics. The results are summarized in Table 4 and Figure 10.

Among them, the *RMSE* and *MAPE* of the models with influencing factors as input values are the highest, which shows that it is difficult for the model to obtain the fluctuation law of the measurement target sequence from the relevant information without the historical data of the target quantity itself. Compared with the historical data of a single demand, the performance indicators of the multivariate data show a downward trend, which shows that the influencing factors can help the model to make better use of historical data to predict the period, but the effect is not great. Multivariate variables with VAE generation indicators play a highly positive role in model prediction. Compared with the multivariate

data, the *RMSE* decreased by 4.46, and the *MAPE* decreased by 8.709. Without the support of relevant influencing factors, the *MAPE* of the model is still not ideal, indicating that the enhancement effect of the indicators is directional.

Table 4. Performance comparison of CNN–Multi-Head Attention–Multi-LSTM with fusion of different features.

Feature Type	*RMSE*	*MAPE*
Time step variable	11.503	19.660
Influencing factors	14.737	22.653
Multivariate data	11.036	19.097
Statistical indicators of VAE + time step variable	8.769	17.766
Statistical indicators of VAE + multivariate data	6.576	10.388

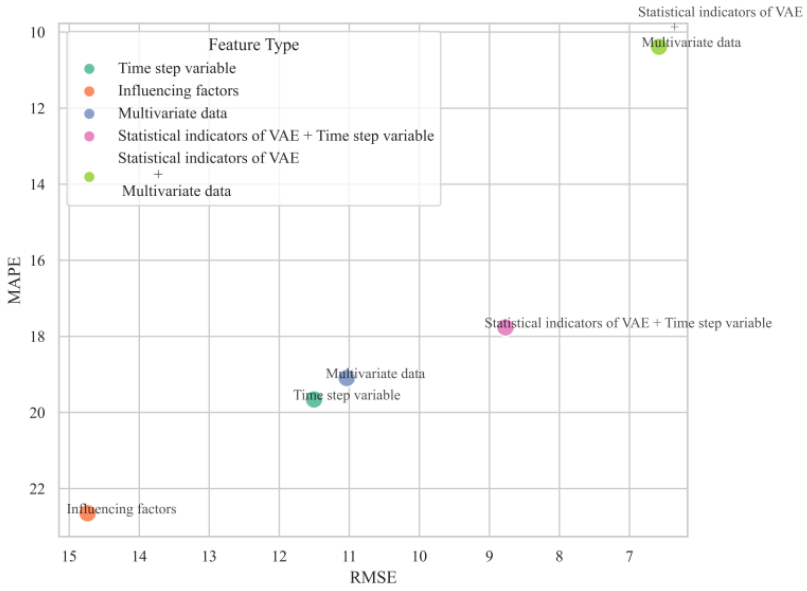


Figure 10. Performance variations in CNN–Multi-Head Attention–Multi-LSTM due to feature fusion.

It can be concluded that the statistical characteristics of the stitching sequence inside the sample enhance the reasoning of the original information, and the characteristics of the training sample unit mark the deviation of the multivariate data in the sample individuals in the form of numerical distribution characteristics, thereby reducing the difficulty in distinguishing samples using the model. Identifying similar samples in different periods will lead to obvious differences in the weight vectors, making the training more stable and inclined to search for the distribution clues of the sequence. It can be considered that the correlation strength between the model weight distribution and the predicted target is maximized, and the prediction effect of a single time point is more accurate.

4.5.3. Comparative Analysis of Multi-Model Errors

After verifying the effectiveness of feature fusion and the model framework suitable for design, eleven groups of contrast models were established, which included random forest (RF), support vector regression (SVR), LSTM, CNN–BiLSTM, CNN–Attention–BiLSTM, CNN–Multi-Head Attention–BiLSTM, Transformer, CNN–Transformer, CNN–Multi-LSTM and CNN–Attention–Multi-LSTM and CNN–Multi-Head Attention–Multi-LSTM. According to the fusion data proposed in this paper, the total demand of doctors was predicted;

for the comparison with our model, the prediction error indicators were the *RMSE*, *MAE*, *MAPE*, and *RMSPE*. The error comparison is presented in Table 5 and Figure 11.

Table 5. Error comparison of prediction results of each model.

Model	Testing Set			
	RMSE	***MAE***	***MAPE***	***RMSPE***
Random Forest	11.749	9.520	18.476	24.231
SVR	8.305	6.525	13.380	16.950
LSTM	8.080	6.373	12.980	16.240
CNN–BiLSTM	7.706	6.271	12.800	15.719
CNN–Attention–BiLSTM	7.115	5.644	11.710	14.706
CNN–Multi-Head Attention–BiLSTM	9.275	7.102	14.063	17.854
Transformer	10.057	7.949	16.680	21.368
CNN–Transformer	7.360	5.729	11.178	13.925
CNN–Multi-LSTM	7.418	5.983	12.227	15.022
CNN–Attention–Multi-LSTM	7.456	6.000	11.899	14.400
CNN–Multi-Head Attention–Multi-LSTM	6.576	5.169	10.388	13.00

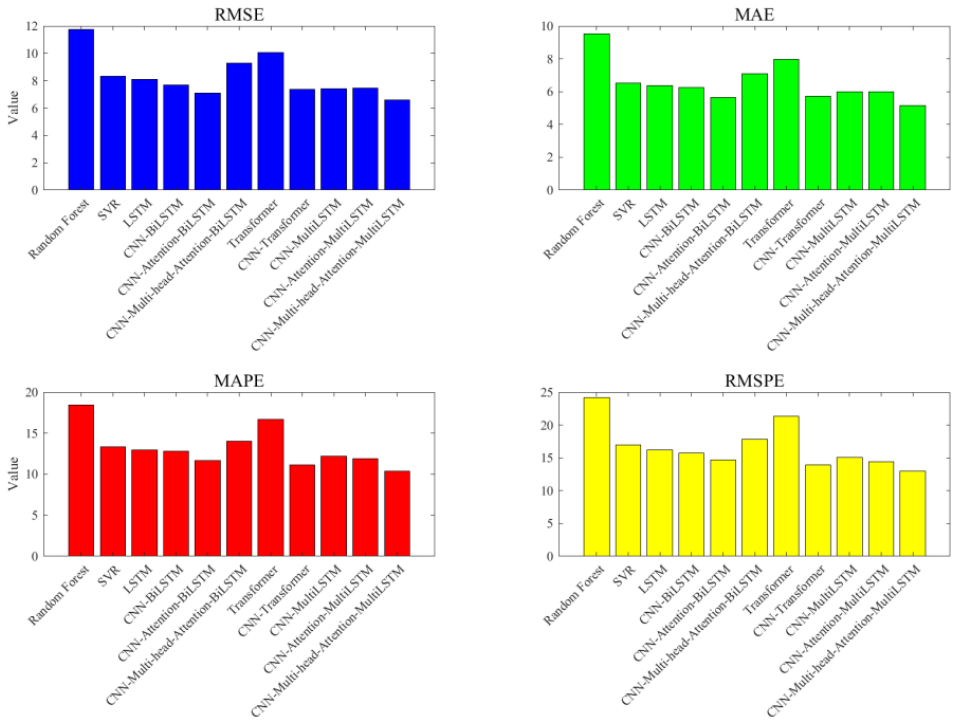


Figure 11. Comparative analysis of performance across multiple models.

One of the indicators of the random forest results is the least satisfactory, as the selection algorithm's mechanism depends on the random allocation of sequences of subdata [31]. Multiple forests use a voting mechanism to generate concurrent predictions and select final outcomes. This approach reduces overfitting but struggles in capturing temporal information due to the parallel module arrangement. Support vector regression (SVR), based on support vector machines, creates a hyperplane to maximize the margins between data points [32]. The kernel function expands the data dimensions nonlinearly, projecting the features into a higher-dimensional space as linear combinations. In this study, the

dataset showed excellent prediction performance, with an *RMSE* of 8.305. Compared to the traditional random forest (RF), four key indicators were improved significantly.

Regarding the deep learning model, the LSTM single-layer algorithm directly received fused features, and the prediction results were close to those of the SVR algorithm. Its feature processing ability is limited to the time-flow direction information of sequence vectors, and it fails to completely fuse multidimensional cells. CNN–BiLSTM joins the CNN processing layer to generate a large number of correlation matrices via continuous convolution, and the BiLSTM layer recognizes the periodic information between elements in both directions. In order to deepen the sequence analysis function of the module, the CNN–Multi-LSTM stack deepens the LSTM layer and explores the deeper internal laws. From the *MAPE* and *RMSPE* indicators, the unit prediction levels were still limited to 12% and 15%, respectively. CNN–BiLSTM is connected to the CNN processing layer to generate a large number of correlation matrices by continuous convolution, and the BiLSTM layer recognizes the periodic information between elements in both directions. If it is directly combined with the attention mechanism, the overload features are not transitionally processed by other processing layers, and the performance is unstable. In terms of the performance results, short sequence tasks generally have lower complexity. Within this experimental context, the Transformer and its composite models are less optimal choices due to their excessive parameter requirements and high computational costs. CNN–Attention–Multi-LSTM introduces the single-headed attention mechanism layer, and LSTM uses the attention mechanism to infer the areas with high weight changes and guide the model to identify the potential variables associated with sample differences, so that the overall and individual prediction effects are strong. The multi-head attention mechanism of this model first divides the attention to guide the direction, explores the global multi-faceted trends of the time series characteristics, and excavates the potential factors to guide the model more hierarchically. In this comparative experiment, all indicators of the model proposed in this paper had the smallest errors, and the test performance was the best.

4.5.4. Supplementary Experiments for Model Enhancement and Adaptation

(1) Simulation experiment on medical communication

In the simulation experiment, the dynamic resource allocation process of the remote consultation network integrated physician demand prediction and network resource optimization, thus demonstrating the cross-domain advantages of telemedicine. Key performance indicators, including the bandwidth performance, were compared between a static allocation strategy with a fixed daily bandwidth of 50 Mbps and a dynamically adjusted strategy based on predictive consulting. The dynamic allocation strategy adjusts the bandwidth based on the predicted physician demand and calculates the required bandwidth using Formula (14), which is derived by multiplying the predicted number of patients by the demand per unit width. During periods of high demand, when the demand exceeds 80% of the total bandwidth, dynamic strategies are used to optimize the resource availability by applying high demand multipliers while controlling the costs. Conversely, static allocation strategies provide a fixed 50 Mbps bandwidth per day, irrespective of the patient needs, which can result in overutilization during off-peak hours or underutilization during periods of peak demand. The performance results and associated costs are illustrated in Figure 12.

$$Daily\ bandwidth = Number\ of\ patients \times Patient\ bandwidth \quad (14)$$

In telemedicine networks, dynamic resource allocation is the key to improving the network performance and cost efficiency. This approach optimizes resource allocation

during peak hours and manages the costs by adjusting the bandwidth to the patient needs. In contrast, static allocation strategies that provide a fixed bandwidth can lead to wasted medical communication resources when the demand changes. The simulation experiments validate existing research [33,34] and show that dynamic resource allocation has significant advantages in adapting to patient needs and fluctuations in network conditions, thereby improving the quality and efficiency of telemedicine services.

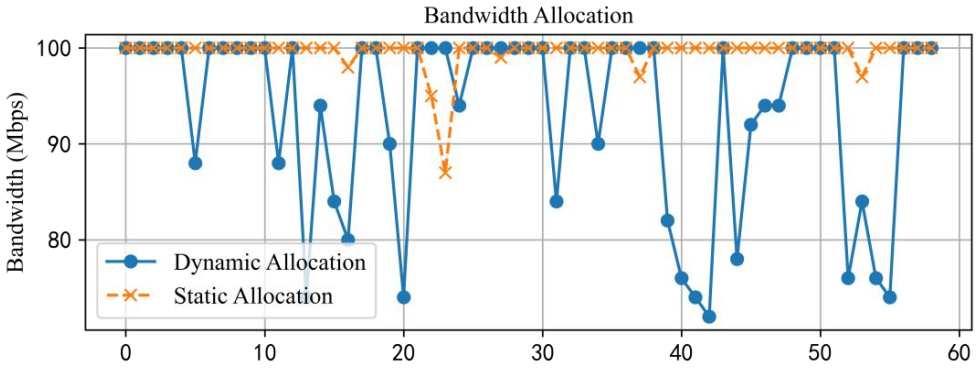
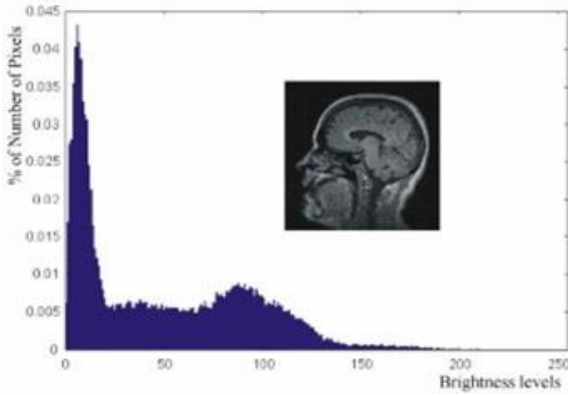


Figure 12. Bandwidth allocation comparison.

(2) Experimental analysis of model generalization capabilities

Two types of medical-related time series datasets from the UCR dataset were used to test the universality of the model: one set comprised histograms of the pixel intensity from medical images, with classes representing different human body regions; the other set corresponded to ECG recordings from the left and right chest. In this dataset, each type of noninvasive fetal ECG is represented as a two-dimensional time series. The contents of the dataset and the associated parameters are presented in Figure 13 and Table 6.



(**a**) MedicalImages dataset image



(**b**) NonInvasiveFetalECGThorax1 dataset image

Figure 13. Dataset visualization.

In this experiment, the *MAE* and *RMSPE* were used to assess the prediction accuracy of the model on the two different medical time series datasets. The *MAE* provides an absolute measure of the discrepancy between the predicted and actual values, while the *RMSPE* provides a relative measure. The primary objective of this experiment was to evaluate the model's versatility and accuracy in processing different types of medical time series data, as well as to investigate its generalizability and adaptability across different medical institutions. As shown in Figure 14, on the MedicalImages dataset, the *MAE* was 1.83 and the *RMSPE* was 2.57, indicating that the model could accurately distinguish different

regions of the human body. On the NonInvasiveFetalECGThorax1 dataset, the *MAE* and *RMSPE* were 7.74 and 9.91, respectively, which, although relatively high, still indicate considerable accuracy when dealing with complex time series data, with the *RMSPE* not exceeding 10%.

Table 6. Dataset information.

Dataset	MedicalImages	NonInvasive FetalECGThorax1
Train Size	381	1800
Test Size	760	1965
Length	99	750
Classses	10	42
Dimensions	1	1
Type	IMAGE	ECG

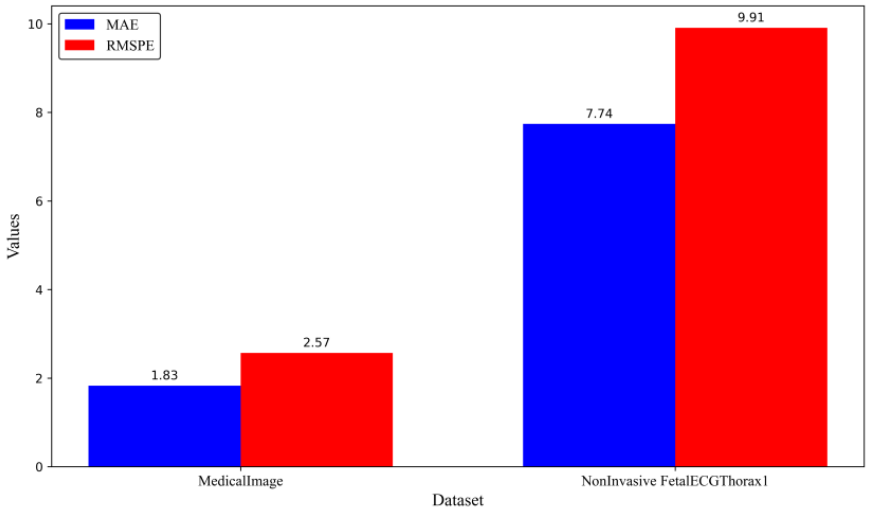


Figure 14. Presentation of model generalization performance metrics.

The preliminary validation of its potential application value across diverse medical settings has been conducted. Previous studies have confirmed the model's efficacy in predicting medical staffing requirements, and the current study further substantiates its versatility for multiple applications within the medical field. The experimental results confirm that the model demonstrates robust accuracy and adaptability in processing different types of medical time series data, particularly in medical image analysis. Additionally, the generalizability and adaptability of the model across different medical institutions have been preliminarily verified.

4.6. Discussion

In China, the rise of telemedicine has brought new opportunities for the prediction and management of medical needs. Through the integration of medical resources across the region by technical means, telemedicine allows patients to enjoy cross-regional medical services, which greatly improves the accessibility of medical services. At the same time, this model has also produced a large number of diversified medical data, including traditional medical records, remote diagnosis and treatment records, and patient behavior data. Incorporating these data into models helps to better understand the diversity and complexity of healthcare needs. In addition, the hierarchical diagnosis and treatment system being implemented in China has also reshaped the distribution and pattern of the

medical demand to a certain extent. Models that incorporate the effects of these policy changes will be able to more accurately predict and respond to the evolution of healthcare demands. Therefore, combined with the practice and policy background of telemedicine in China, the application value of this model will be further improved and expanded.

5. Conclusions

This study examines recent advancements in sequence prediction algorithms integrated into medical big data systems. It highlights the key concepts and applications of high-performance sequence processing in healthcare. Research shows that normalization-based methods and multi-attention mechanisms reduce regional data variability, ensuring reliable predictions in complex medical environments [35]. The proposed model leverages a CNN–LSTM module architecture, drawing inspiration from text normalization techniques, to effectively handle the variability in non-isomorphic attributes in telemedicine data. It constructs one-dimensional sequences by concatenating normalized attributes, which are then processed by a generative encoder. This encoder integrates these elements based on their distinct characteristics and inserts the enhanced sequence into the original dataset, preserving the statistical independence of the samples. The model employs a CNN–Multi-LSTM module combined with a multi-attention mechanism to monitor fluctuations in the latent variable series. Additionally, it applies advanced logical reasoning to identify potential sources of ambiguity. For instance, a review of historical physician demand data from a tertiary hospital yielded the following conclusions.

(1) In comparison to analogous models, the complexity of CNN–Multi-Head Attention–Multi-LSTM results in a consistent performance enhancement: the LSTM network continuously processes the convolutional embedding features, addressing the issue of excessive discreteness in the input vector caused by the direct concatenation of the original sequence data; the stacked architecture effectively manages nonlinear complex feature functions; and the attention module ensures that the final model performance reaches its optimal value with appropriate parameter allocation.

(2) The three factors selected in this study passed the statistical test of significance, which is in accordance with the theory of variable correlation. The comparison experiment of features shows that the accumulation effect of the elements is weak, whereas the core indicators that the VAE purifies enhance the discernability of the accumulated variables, showing that similar vectors with specific values have good identifiability in the network.

(3) The prediction results provide a realistic scheme for the short-term demand allocation prediction of medical databases. The benefit of this data universality technique is that it allows a big data platform-supported medical information system to create a prediction framework model, making it easy for medical units to connect with one another based on shared characteristics. In the follow-up, the ecological environment of the information platform will be improved, and more detailed and effective variables will be recorded in an organized manner. The fusion logic of new variables does not conflict with the previously established embedding vector, and its applicability and updating coexist.

Despite the promising findings, it is important to acknowledge certain inherent limitations introduced by the dataset utilized in this study. The dataset primarily consists of records from a single hospital and encompasses a limited range of data dimensions, such as weather conditions and search traffic. This restricts the dataset's ability to fully represent the complex dynamics of the medical demand. The regional focus and limited dimensionality of the data constrain the generalizability of the model's findings, thereby limiting their applicability to other healthcare settings or institutions. To enhance the

model's adaptability and generalizability, future research should aim to broaden both the scope and depth of the data. This can be achieved by incorporating data from diverse geographical areas and various levels of healthcare facilities, as well as enriching the dataset with crucial factors such as patient demographics, hospital operations, and policy changes. Additionally, integrating multi-model data sources, such as electronic medical records and medical imaging, will enhance the model's ability to capture a more comprehensive picture of the medical demand. These advancements will not only increase the reliability and generalizability of the model in more complex medical contexts but also provide more accurate and holistic support for healthcare decision-making.

The popularization of large-scale models and high-tech modules that can be embedded in many fields, owing to their technical extensibility, indicates that tens of millions of parametric models benchmark the increasing large-scale resource demand information at the order of magnitude level. Beck et al. [36] proposed an xLSTM module similar to the architecture used in this study and suggested that the model be migrated to a high-order framework to explore a medical information decision-making platform for the multi-channel collaborative computing of big data.

Author Contributions: Conceptualization, J.Z. and W.C.; methodology, J.Z.; formal analysis, J.Z.; resources, X.W.; J.Z., W.C. and X.W. developed the idea for the study. J.Z.did the experiments. And all authors analyzed the results and were involved in writing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The data supporting the reported results were obtained from a tertiary hospital in Nanjing and are subject to privacy and ethical restrictions, thus they cannot be publicly shared.

Conflicts of Interest: The authors have no conflict of interest to declare.

References

1. Gu, H.; Feng, Z.; Wu, D.; Yang, N. Research Status and Trends of Telemedicine in China—Quantitative Analysis Based on CiteSpace. *J. Inf. Resour. Manag.* **2020**, *10*, 119–129. [CrossRef]
2. Wang, T. Research on Cost Forecast and Control Strategy of Telemedicine Service. Master's Thesis, Zhengzhou University, Zhengzhou, China, 2020.
3. Kamel, F.O.; Magadmi, R.; Qutub, S.; Badawi, M.; Badawi, M.; Madani, T.A.; Alhothali, A.; Abozinadah, E.A.; Bakhshwin, D.M.; Jamal, M.H.; et al. Machine Learning-Based Prediction of COVID-19 Prognosis Using Clinical and Hematologic Data. *Cureus* **2023**, *15*, e50212. [CrossRef] [PubMed]
4. Liu, C.; He, T.; Wang, Q. Responses to the Supportive Policies for the Implementation of Contracted Family Doctor Services and Job Satisfaction in Primary Healthcare Workers. *Chin. Gen. Pract.* **2019**, *22*, 1160–1164. [CrossRef]
5. Zhai, Y.; Qiao, Z.; Qiao, Y. Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism. *Comput. Sci.* **2024**, *51*, 230800119–230800127. [CrossRef]
6. Lv, C.; Su, Y.; Yang, J.; Yu, F.; Wang, N. Study on Correlation Factors Analysis and Prediction of Visit Rate of Grassroots Medical and Health Institutions Among Residents in Ningxia. *Med. Soc.* **2024**, *37*, 71–77. [CrossRef]
7. Romanuke, V. Arima model optimal selection for time series forecasting. *Marit. Tech. J.* **2022**, *224*, 28–40. [CrossRef]
8. Chan, S.; Oktavianti, I.; Puspita, V. A deep learning cnn and ai-tuned svm for electricity consumption forecasting: Multivariate time series data. In Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 17–19 October 2019; IEEE: Piscataway, NJ, USA, 2019.
9. Lim, B.; Zohren, S. Time-series forecasting with deep learning: A survey. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200209. [CrossRef]
10. Zou, Z.; Li, C.; Deng, J. Research on Fuel Consumption Prediction Modeling Based on Max-Relevance and Min-Redundancy Feature Optimization Algorithm. *Comput. Simul.* **2023**, *40*, 186–191.
11. Zhang, H.; Ma, W. Subway Passenger Flow Forecasting Model Based on Temporal and Spatial Characteristics. *Comput. Sci.* **2019**, *46*, 292–299.

12. Zhang, X.; Liu, G.; Wang, Z.; Luo, F.; Wu, J. Microscopic diffusion prediction based on multifeature fusion and deep learning. *J. Tsinghua Univ. (Sci. Technol.)* **2024**, *64*, 688–699. [CrossRef]
13. Zhang, H.; Liu, M.; Li, L.; Cheng, Y.; Zhou, Y. Study on Data Governance of the Healthcare Big Data Platform Based on Common Data Model. *J. Med. Inf.* **2022**, *43*, 13. [CrossRef]
14. Wang, B.; Jia, P.; Guo, F.; Sun, L.; Lin, K. Deep prediction model of coal spontaneous combustion temperature based on multi-feature fusion. *China Min. Mag.* **2024**, *33*, 84–90. [CrossRef]
15. Yu, Y.; Ge, L.; Jin, Z.; Wang, Y.; Ding, L. Short-term load prediction method of distribution networks considering weather featuresand multivariate correlations. *Power Syst. Prot. Control* **2024**, *52*, 131–141. [CrossRef]
16. Wang, W.-C.; Gu, M.; Hong, Y.-H.; Hu, X.-X.; Zang, H.-F.; Chen, X.-N.; Jin, Y.-G. SMGformer: Integrating STL and multi-head self-attention in deep learning model for multi-step runoff forecasting. *Sci. Rep.* **2024**, *14*, 23550. [CrossRef] [PubMed]
17. Kumar, I.; Tripathi, B.K.; Singh, A. Attention-based LSTM network-assisted time series forecasting models for petroleum production. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106440. [CrossRef]
18. Wu, Y.; Ni, J.; Cheng, W.; Zong, B.; Song, D.; Chen, Z.; Liu, Y.; Zhang, X.; Chen, H.; Davidson, S.B. Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 651–659. [CrossRef]
19. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]
20. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30. [CrossRef]
22. Chen, W.; Li, J. Forecasting Teleconsultation Demand Using an Ensemble CNN Attention-Based BILSTM Model with Additional Variables. *Healthcare* **2021**, *9*, 992. [CrossRef] [PubMed]
23. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the PMLR, International Conference on Machine Learning, Montreal, QC, Canada, 11–12 December 2015; pp. 448–456. [CrossRef]
24. Xu, J.; Sun, X.; Zhang, Z.; Zhao, G.; Lin, J. Understanding and improving layer normalization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32. [CrossRef]
25. Frick, T.; Antognini, D.; Giurgiu, I.; Grewe, B.F.; Malossi, C.; Zhu, R.J.; Rigotti, M. MC Layer Normalization for calibrated uncertainty in Deep Learning. *Trans. Mach. Learn. Res.* **2024**. Available online: https://openreview.net/forum?id=bG3ICt3E0C (accessed on 10 October 2024).
26. Bai, Y.; Yi, J.; Tao, J.; Tian, Z.; Wen, Z.; Zhang, S. Fast end-to-end speech recognition via non-autoregressive models and cross-modal knowledge transferring from BERT. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 1897–1911. [CrossRef]
27. Greco, G.; Guzzo, A.; Nardiello, G. FD-VAE: A feature driven VAE architecture for flexible synthetic data generation. In *Database and Expert Systems Applications*; DEXA 2020. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12391. [CrossRef]
28. Fan, S.; Yu, H.; Cai, X.; Geng, Y.; Li, G.; Xu, W.; Wang, X.; Yang, Y. Multi-Attention Deep Neural Network Fusing Character and Word Embedding for Clinical and Biomedical Concept Extraction. *Inf. Sci.* **2022**, *608*, 778–793. [CrossRef]
29. Che, C.; Wang, H.; Ni, X.; Lin, R.; Xiong, M. Residual Life Prediction of Aeroengine Based on 1D-CNN and Bi-LSTM. *J. Mech. Eng.* **2021**, *57*, 304–312. [CrossRef]
30. Wang, J.; Chen, L.; Yu, S.; Jiang, C.; Wu, Y. Multi-factor perceived short-term tourist number prediction model based on gated recurrent unit. *J. ZheJiang Univ. (Eng. Sci.)* **2019**, *53*, 2357–2364. [CrossRef]
31. Li, Y.; Jia, Y.; Li, L.; Hao, J.; Zhang, X. Short term power load forecasting based on a stochastic forest algorithm. *Power Syst. Prot. Control* **2020**, *48*, 117–124. [CrossRef]
32. Valente, J.M.; Maldonado, S. SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression. *Expert Syst. Appl.* **2020**, *160*, 113729. [CrossRef]
33. Ezeamii, V.C.; Okobi, O.E.; Wambai-Sani, H.; Perera, G.S.; Zaynieva, S.; Okonkwo, C.C.; Ohaiba, M.M.; William-Enemali, P.C.; Obodo, O.R.; Obiefuna, N.G. Revolutionizing Healthcare: How Telemedicine Is Improving Patient Outcomes and Expanding Access to Care. *Cureus* **2024**, *16*, e63881. [CrossRef] [PubMed] [PubMed Central]
34. Ahmed, S.T.; Patil, K.K.; Shanraj, R.K.; Dhanaraj, R.K.; Khan, S.B.; Alzahrani, S.; Rani, S. 6GTelMED: Resources Recommendation Framework on 6G-Enabled Distributed Telemedicine Using Edge-AI. *IEEE Trans. Consum. Electron.* **2024**, *70*, 5524–5532. [CrossRef]

35. Gavgani, V.Z.; Pourrasmi, A. Data Governance Navigation for Advanced Operations in Healthcare Excellence. *Taṣvīr-i Salāmat* **2024**, *15*, 249–254. [CrossRef]
36. Beck, M.; Pöppel, K.; Spanring, M.; Auer, A.; Prudnikova, O.; Kopp, M.; Klambauer, G.; Brandstetter, J.; Hochreiter, S. xLSTM: Extended Long Short-Term Memory. *arXiv* **2024**, arXiv:2405.04517. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.