Hindawi
Mobile Information Systems
Volume 2022, Article ID 8966266, 9 pages
https://doi.org/10.1155/2022/8966266

Research Article

Explainable and Personalized Medical Cost Prediction Based on Multitask Learning over Mobile Devices

Lin Sun,[1,2] Tingqi Wang,[1] Bei Hui,[3,4] Yun Li,[5] and Ling Tian[1,6]

[1]*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2]*West China Hospital of Sichuan University, Chengdu 610000, China*
[3]*School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610031, China*
[4]*Kashi Institute of Electronic and Information Industry, Kashi 844000, China*
[5]*Department of Thoracic Surgery, The Seventh Affiliated Hospital, Sun Yat-Sen University, Shenzhen 528406, China*
[6]*Shenzhen Institute of Information Technology, Shenzhen 518172, China*

Correspondence should be addressed to Bei Hui; bhui@uestc.edu.cn

Received 15 August 2022; Accepted 6 September 2022; Published 9 October 2022

Academic Editor: Yan Huang

Currently, the forecasting of healthcare costs is of significant importance for the finance management of both government and individual citizens. However, the existence of dramatic individual diversity in health status, as well as the extensive complexity of the factors influencing the cost, has made the prediction a challenging task. Thanks to the unprecedented adoption of mobile devices, regular individuals may contribute diverse dimensions of data for the medical cost prediction. Hospitals and healthcare service providers are all setting up their own mobile services and collect user data for analysis. Previous methods usually employed traditional machine learning or simple neural network methods, which are difficult to be applied to the nonlinear medical cost and diverse dimensions of data. Therefore, this paper proposes a multitask learning-based framework for interpretable medical cost interval prediction to address these issues. The framework proposed in this paper first predicts subcost intervals by applying the multidimensional data collected from mobile ends and following the multitask learning paradigm. The total cost interval is then predicted based on this prediction. Simultaneously, the framework derives a decision tree from the parameters of the multitask learning network and calculates the importance of each feature in predicting the cost intervals. This paper demonstrates the method's effectiveness using real-world data experiments.

1. Introduction

The management of healthcare cost is one of the largest challenges in the field of health insurance and healthcare, which can easily lead to a shortage or waste of healthcare resources when poorly managed [1–4]. Owing to the extensive development of mobile devices, patients and regular citizens can freely contribute their own data for the prediction of the medical cost. Typical organizations like healthcare service providers and hospitals are setting up their own applications towards this trend. Patients and subscribers can use those mobile apps to contribute multiple types of data like demographic attributes, manually inputted daily healthcare records, and even sensing data from smart watches [5]. Therefore, it is of great significance to study the adoption of these data for medical cost prediction, which can bring personalized and understandable services for patients.

Currently, DRG (Diagnosis Related Group)-based payment methods are being widely used to predict costs through characteristic groupings [6, 7], which has strongly motivated the research on reliable medical cost prediction. Various methods are proposed to accurately predict cost ranges and identify key factors for grouping, allowing for efficient resource deployment and timely identification of potential risks. These methods are assumed to bring significant implications for reducing pressure on healthcare resources and improving resource utilization [8, 9] while concealing no significant personal information of patients

[10–12]. However, because of the various treatment options chosen by individual patients, the amount and composition of medical cost are highly personalized and divergent [13]. Moreover, due to the different conditions of different patients and the influence of factors such as healing time and degree of recovery, it is difficult to fit medical costs with simple linear models. Therefore, the prediction of medical cost requests both the application of various dimensions of data available from mobile ends and feedbacks to users with a deep understanding on the impact of individual and personalized characteristics on healthcare costs [14, 15].

Considering these challenges, traditional methods rely heavily on machine learning models like linear regression [16] and regression trees [17, 18], as well as simple neural network models. However, the overall representation is inadequate owing to the sophisticated correlations among factors. In recent research, deep learning methods [19] outperformed traditional computational methods in various prediction tasks due to their ability to adapt the composition of individual feature factors for better representation [20]. Given the complexity of the components and data dimensions in medical cost prediction [21], deep learning methods can make more accurate and reasonable predictions of overall costs by depicting the correlation between the various costs in addition to predicting individual costs.

Based on the above, this paper proposes a multi-task learning-based interpretable medical cost interval prediction framework. The model takes multiple sources of information about the patient into account, including (1) the patient's natural characteristics and (2) the stage of the patient's condition. (3) The patient's lesion attributes, and outputs the prediction results for each type of cost interval.

The framework is made up of two parts: (1) A multitask-learning framework for interval prediction over data collected from mobile ends. The cost intervals are predicted by the prediction framework in two steps. To begin, a logistic regression approach is combined as a preprocessing of the input neural network data, which is then fed into the neural network to calculate predictions for the various subcost intervals. The total costs are then predicted based on the prediction of the subcost intervals. Among these, the logistic regression method is used to improve the network's convergence and training speeds. (2) An explainable and personalized decision tree based on the analysis of factor importance in a multilearning task. The Gini coefficient is reconstructed using the multitask learning framework weights obtained from training to build a decision tree, and the importance of each feature is calculated using the decision tree.

The proposed framework owns two advantages for medical cost prediction. On the one hand, the framework predicts total costs by coupling subcosts, allowing all subcost prediction intervals to be in obtained while also capturing the links between subcosts and global payments; on the other hand, the framework can analyze the importance of different factors in the prediction of cost intervals based on the prediction process. Corresponding observations can serve as a foundation for physician triage.

To the best of our knowledge, this is the first time that a multiclassification approach to cost interval forecasting has been used. The remainder of this paper is organized as follows: Section 2 presents work related to cost interval prediction. Section 3 presents the cost interval prediction model for multitask learning. Section 4 presents the experimental results. Section 5 analyses the factor impact. Section 6 presents the conclusions.

2. Related Work

The study of cost prediction tasks is becoming more widespread, and one of the widely used methods for health care cost prediction is the regression-based model [22, 23]. To avoid the requirement of general linear models for data to follow a normal distribution, Moran et al. performed prediction using generalized linear models [16]. Panay et al. used the evidence regression method, which is based on the idea that other elements in a set that are correlated for a specific element are placed in a set of similar patients, and the overall predicted expectation is calculated for optimization [24]. Tkachenko R et al. used SGTM-like neural structures for segmented linear prediction [25]. Takeshima et al. defined experimental valuables on which regression models with minimum absolute shrinkage and selection operators (lasso) were built. Explanatory valuables were selected by LASSO avoiding overfitting using the validation data [26]. Based on regression methods, various machine learning methods have been introduced [27, 28]. Taloba et al. in [17] compare the performance of linear regression type Lasso, gradient augmentation of regression decision trees, M5 regression decision trees, random forests, linear regression, and CART regression trees in this task and analyze the advantages and disadvantages of each method.

Due to properties such as end-to-end training and good fitting ability to nonlinear data, neural network methods, in addition to machine learning methods, have been introduced into the prediction of medical costs. Morid et al. compared various methods and found that ANN (Artificial Neural Network) performed the best [20]. In [29], Zeng et al. used multilayer neural networks to construct unsupervised learning models to learn patient representation from medical data. The collection of medical data from mobile devices are also extensively studied. Issues like efficiency [5, 30] and data utilities are thoroughly considered. These studies are complementary to our work.

Generally, for cost prediction, current work is primarily based on patients' natural attributes and health data, but there are fewer methods for predicting the costs of specific conditions during treatment. At the same time, current methods are based on simple statistical learning and neural networks, and they are incapable of fully exploiting the value of data contributed by patients from mobile ends.

3. Framework: A Multitask Learning Based Framework for Interpretable Medical Cost Interval Prediction

3.1. Problem Definition. For a patient set $U = \{U_1, U_2, ..., U_i, ..., U_N\}$ containing N patients, where each patient U_i has a feature set X_i and an element $x_i \in R$ for each feature

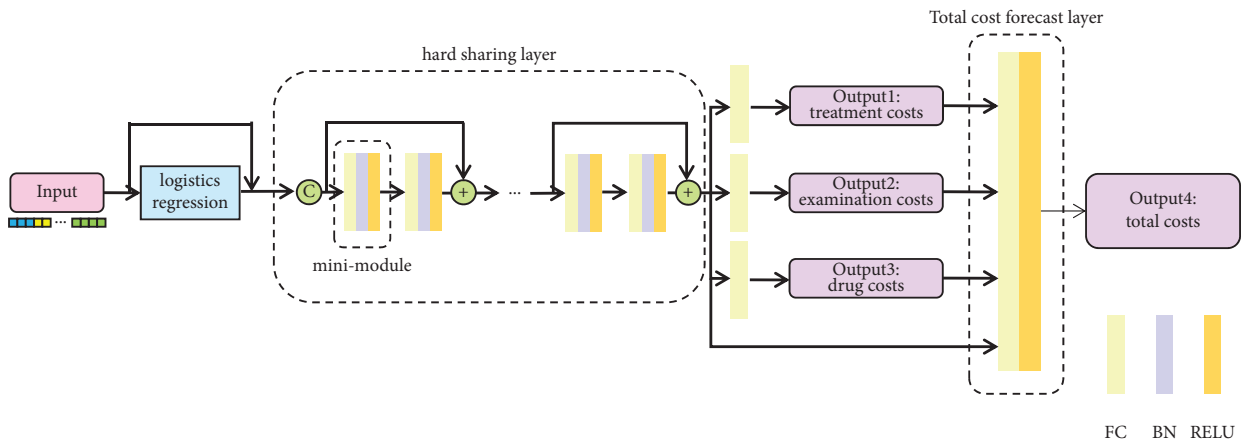

Figure 1: The cost forecasting framework.

dimension. For example, the feature set may include natural features such as "age," daily collected data like heartbeat records and related events inputted by patients. These features are collected and submitted through mobile devices. Corresponding features also involve the disease stage such as "TMN-stage," and focal features such as "type of comorbidity," which are both terminologies used for clinic diagnosis of breast cancer. We define a cost interval set $Y = \{Y_1, Y_2, \ldots, Y_i, \ldots, Y_N\}$, where $Y_i = Y_{i,\text{sub}} \cup Y_{i,\text{total}}$, $Y_{i,\text{sub}} = \{y_{i,1}, y_{i,2}, \ldots, y_{i,k}\}$, $Y_{i,\text{total}} = \{y_{i,\text{total}}\}$.

In this paper, we take $k = 3$ as an example, $y_{i,1}, y_{i,2}, y_{i,3}$ correspond to the intervals of treatment cost, examination cost, and drug cost, respectively, and $y_{i,\text{total}}$ is the total cost interval. $y_{i,1} \in \{1, 2, \ldots, k_1\}$, $y_{i,2} \in \{1, 2, \ldots, k_2\}$, $y_{i,3} \in \{1, 2, \ldots, k_3\}$, $y_{i,\text{total}} \in \{1, 2, \ldots, k_{\text{total}}\}$.

For a given set of patient features X, after inputting it into the model, the set $Y = Y_{\text{sub}} \cup Y_{\text{total}}$ of its corresponding cost intervals is output.

3.2. A Framework for Predicting Medical Cost Intervals Based on Multitask Learning. The framework proposed in this paper consists of three components: data preprocessing; a hard sharing network for subcost interval prediction; and a total task prediction network based on sub-cost intervals. The results obtained from predicting subcost intervals and the raw data outputted by hard sharing are used as inputs for the total cost prediction. An illustration of the framework is shown in Figure 1.

3.2.1. Data Preprocessing: Logistic Regression. The training of neural networks for such data suffers from slow convergence and long training times due to the weak linear nature of the association between medical data and medical cost intervals. Traditional machine learning methods like logistic regression may extract shallow nonlinear association among data, which can benefit the overall training performance of the framework. As a result, in this paper, $\hat{\theta}_{\text{log istic}}$ is calculated using logistic regression before being fitted with a neural network. When the user U_i set of features X_i is entered, the auxiliary information $h_{\hat{\theta}_{\text{log istic}}}(X_i)$ can be obtained.

$$h_{\hat{\theta}_{\text{log istic}}}(X_i) = \text{softmax}\left(X_i; \hat{\theta}_{\text{log istic}}\right) = \begin{bmatrix} p\left(y_{i,\text{total}} = 1 | X_i; \hat{\theta}_{\text{log istic}}\right) \\ p\left(y_{i,\text{total}} = 2 | X_i; \hat{\theta}_{\text{log istic}}\right) \\ \cdot \\ p\left(y_{i,\text{total}} = k_{\text{total}} | X_i; \hat{\theta}_{\text{log istic}}\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\hat{\theta}_j^T X_i}} \begin{bmatrix} e^{\hat{\theta}_1^T X_i} \\ e^{\hat{\theta}_2^T X_i} \\ \cdot \\ e^{\hat{\theta}_{k_{\text{total}}}^T X_i} \end{bmatrix}, \quad (1)$$

where $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{k_{\text{total}}} \in R$ is the model parameter and the parameter is obtained by optimising the loss function by an iterative method, the loss function is a great likelihood function, $\hat{\theta}_{\text{log istic}} = \underset{\theta_{\text{log istic}}}{\text{argmin}}\, \text{Loss}_{\text{log istic}}(h_{\theta_{\text{log istic}}}(X), Y)$.

$h_{\hat{\theta}_{\text{log istic}}}(X_i)$ concatenated with the original data, to obtain $\text{INPUT}(X_i)$:

$$\text{INPUT}(X_i) = \text{concat}\left(X_i, h_{\hat{\theta}_{\text{log istic}}}(X_i)\right). \quad (2)$$

$\text{INPUT}(X_i)$ as input to the multitask learning hard sharing layer can speed up the convergence and training of the neural network.

3.2.2. Hard Sharing Network for Subcost Interval Forecasting. The hard sharing layer consists of the mini-module and Resnet.

Each mini-model consists of a full connection layer, a BatchNormalization layer, and an activation layer (ReLu is used as an example in this paper) which, after the hard sharing layer, gives a hidden layer representation of the data $m_i^{(l)}$:

$$\begin{aligned} FC_i^{(l)} &= FC\left(m_i^{(l-1)}\right), \\ BN_i^{(l)} &= BN\left(FC_i^{(l)}\right), \\ m_i^{(l)} &= \sigma\left(BN_i^{(l)}\right), \end{aligned} \tag{3}$$

where l is the number of layers in the network of the mini-module.

In order not to degrade the performance of the network due to degradation caused by nonconstant mapping, a residual network is used in this paper. A residual connection [19] is made for every two mini-modules to obtain the hidden layer $h(X_i)$:

$$\begin{aligned} \text{Resnet}_i^{(L)} &= \text{Resnet}\left(m_i^{(2L)}, m_i^{(2L+1)}\right), \\ h(X_i) &= \text{Resnet}_i^{(L)}. \end{aligned} \tag{4}$$

Put $h(X_i)$ into different full connection layers to obtain predictions for each subcost interval:

$$\begin{aligned} \hat{y}_{i,1} &= FC_1\left(h(X_i)\right), \\ \hat{y}_{i,2} &= FC_2\left(h(X_i)\right), \\ \hat{y}_{i,3} &= FC_3\left(h(X_i)\right). \end{aligned} \tag{5}$$

Based on this, $\hat{Y}_{i,\text{sub}} = \left\{\hat{y}_{i,1}, \hat{y}_{i,2}, \hat{y}_{i,3}\right\}$ of X_i is obtained. Where the loss function Loss_{sub} for the subcost prediction network is defined as follows:

$$\text{Loss}_{\text{sub}} = \sum_{j=1}^{3} \alpha_j \text{Loss}\left(\hat{Y}_{i,j}, Y_{i,j}\right). \tag{6}$$

3.2.3. Total Cost Interval Forecasting Network Based on Subcost Intervals. A fully connected layer and an activation layer comprise the total cost interval prediction network. The predicted values of the three subcost intervals, along with the output of the hard sharing layer, are fed into the total cost prediction layer, which produces a prediction of the total cost interval as follows:

$$\begin{aligned} \hat{X}_{i,\text{total}} &= \text{concat}\left(h(X_i), \hat{y}_1, \hat{y}_2, \hat{y}_3\right), \\ \hat{y}_{i,\text{total}} &= \sigma\left(FC\left(\hat{X}_{i,\text{total}}\right)\right), \end{aligned} \tag{7}$$

From this, the predicted value $\hat{Y}_i = \left\{\hat{y}_{i,\text{total}}\right\}$ is obtained for four cost intervals of X_i

The loss function $\text{Loss}_{\text{total}}$ for the overall cost is defined as follows:

$$\text{Loss}_{\text{total}} = \text{Loss}\left(\hat{Y}, Y\right). \tag{8}$$

It distinguish the differences between Loss_{sub} and $\text{Loss}_{\text{total}}$:

$$\text{Loss} = \beta_1 \text{Loss}_{\text{sub}} + \beta_2 \text{Loss}_{\text{total}}, \tag{9}$$

where β_1 and β_2 are hyper parameters. In this paper, (8) is selected as the loss function.

4. Feature Importance Analysis

Simply predicting the cost interval may confuse the doctors even if a highly accurate performance is guaranteed. Therefore, a feature-importance-based framework in explaining the prediction of cost interval is further proposed in this part. The whole framework is based on an improved version of decision tree, where multiple factors considered in the prediction model are involved. An illustration is shown in Figure 2.

A decision tree approach is used in this paper to analyze the importance of factors obtained through the multitask neural network in section 3. In contrast to previous decision tree methods simply estimating information gain for a single task, an method tailored to couple with multitasks is designed for the information gain estimation.

Based on the weight parameters of the whole prediction network obtained from training, the weight parameters corresponding to each sub-cost interval is first calculated as a percentage of the total cost prediction layer, which is used as the weight for the Gini coefficient calculation of the decision tree nodes. The original Gini coefficient calculation formula:

$$\text{gini} = \sum_{k=1}^{k} p_k\left(1 - p_k\right). \tag{10}$$

For a given matrix of patient features $(x_{1,0}, x_{1,1}, x_{2,0}, x_{2,1}, x_{2,2})$, the input total cost prediction layer is subject to the following calculation: $(x_{1,0}, x_{1,1}, x_{2,0}, x_{2,1}, x_{2,2}) \times (w_{1,0}, w_{1,1}, w_{2,0}, w_{2,1}, w_{2,2})^T$, where the feature elements with the same first numerical ordinal number of the subscript, e.g., $x_{1,0}, x_{1,1}$ refer to common features. Then, for the same feature element X_i, having a matrix of weights W_i, the weight of each feature element in the calculation of the Gini coefficient is calculated as follows:

$$\alpha_k = \frac{\left\|W_k\right\|_{l_2}}{\sum_{i=1}^{N}\left\|W_i\right\|_{l_2}}, \tag{11}$$

Then, the weighted Gini coefficient calculation formula:

$$\begin{aligned} \text{gini} = \alpha_1 \sum_{K_1=1}^{K_1} p_{k_1}\left(1 - p_{k_1}\right) + \alpha_2 \sum_{k_2=1}^{K_2} p_{k_2}\left(1 - p_{k_2}\right) + \ldots \\ + \alpha_n \sum_{k_n=1}^{K_n} p_{k_n}\left(1 - p_{k_n}\right). \end{aligned} \tag{12}$$

We build a decision tree using the CART classification tree method [20]. The main idea of the method is to

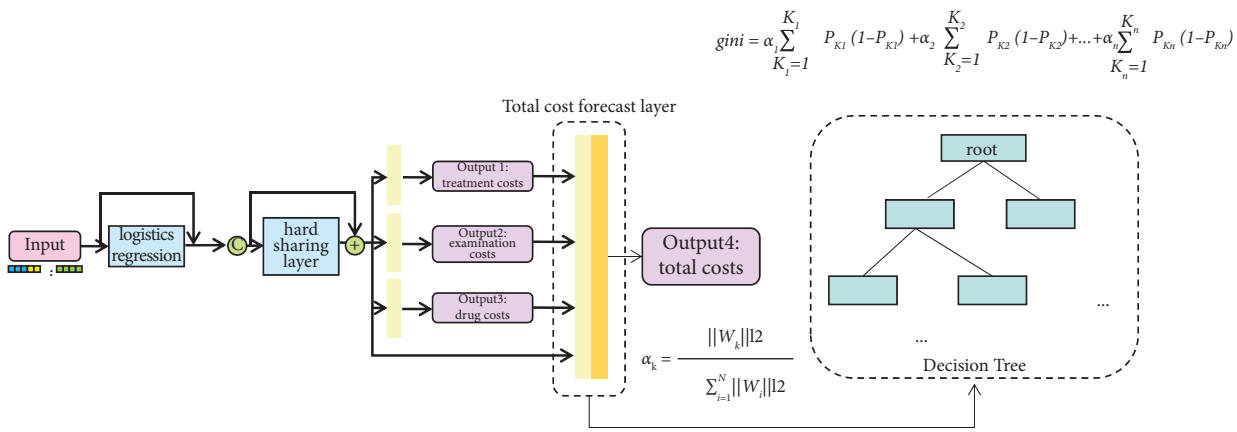

Figure 2: Decision tree analysis method based on neural network weights.

iteratively split the patient set where each subset share identical value on some features. Specifically, when a feature F takes the value f in a sample U with N users, the sample U is divided into two parts U_1 and U_2, where U_1 is the set of samples with $F \neq f$ and U_2 is the set of samples with $F \neq f$. The method calculates the Gini coefficient of each feature at each value, choose the case with the smallest Gini coefficient, and use it to generate this node, with U_1 and U_2 as patient sets in two child nodes. When a node's number of samples U falls below a predefined threshold, or when the number of features is zero, the current node's decision making process is terminated.

The method described above is used to create a decision tree. The essence is to create a binary tree by selecting the features that will give the greatest Gini gain as nodes at each layer. The importance of each feature in nodes is calculated using the following formula based on the generated decision tree:

$$\frac{N_t}{N} \times \left(\text{gini} - \frac{N_{t_R}}{N_t} \times \text{gini}_R - \frac{N_{t_L}}{N_t} \times \text{gini}_L \right), \tag{13}$$

where N is the total number of samples, N_t is the number of samples at this node, N_r is the number of samples at the right child node, and N_{t_L} is the number of samples at the left child node.

5. Experiments

5.1. Dataset. The experiments in this paper are based on a real breast cancer medical cost dataset. We give links to the data demos at the end of this article. Patient features include age, T stage, M stage, N stage, histological classification, complication and comorbidity, and HER2 attributes. The representation for the feature is shown in Table 1.

In this paper, treatment costs, examination costs, and drug costs are selected as the three subcosts to be predicted and the total costs are predicted by using these three subcosts as auxiliary information. For the different costs, the paper divides the cost intervals as shown in Table 2.

The experiments in this paper use one-hot coding for the representation of the data.

Table 1: Patient feature categories classification.

Features	Feature categories	Values
Age	25 -	0
	25–40	1
	40–50	2
	50–60	3
	60+	4
T-stage	T0	0
	Tis	1
	Tx	2
	T1	3
	T2	4
	T3	5
	T4	6
Histologic classification	Invasive carcinoma	0
	Invasive ductal carcinoma	1
	Papillary carcinoma	2
	Infiltrating lobular carcinoma	3
	Medullary carcinoma	4
N-stage	N0	0
	N1	1
	N2	2
	N3	3
	Nx	4
M-stage	M0	0
	M1	1
	Mx	2
Complication and comorbidity	serious	0
	General	1
	—	2
HER2	0	0
	1+	1
	2+	2
	3+	3

5.2. Parameter Settings. The logistic regression model in this paper employs Newton's method as the optimization method for the loss function, and the regularization method employs the l_2 norm with a regularization strength of 0.5; the neural network employs ASGD as the optimization method, with a starting learning rate of 0.1 and decreasing to 50% of the original every 100 epochs; and the linear layer has a dimension of 128.

Table 2: Cost interval.

Treatment costs	1000–	1000–2000	2000+	—
Examination costs	5000–	5000–10000	10000+	—
Drug costs	500–	500–1500	1500–2000	2000+
Total costs	10000–	10000–20000	20000–25000	25000+

Table 3: Accuracy comparison with classical methods.

Method	Accuracy			
	Treatment costs	Examination costs	Drug costs	Total costs
SVM	0.54	0.48	0.53	0.43
DecisionTreeClassifier	0.61	0.54	0.54	0.45
Naive_Bayes	0.59	0.58	0.61	**0.61**
Logistic regression	**0.65**	**0.64**	**0.62**	0.51
k-means	0.48	0.43	0.37	0.42
Multi-task	**0.7**	**0.73**	**0.72**	**0.71**

The decision tree for analyzing the importance of the influencing factors in this paper uses a CART decision tree with a maximum number of layers of 7. The Gini coefficient is used to calculate the information gain, but unlike the traditional Gini coefficient, the Gini coefficient is improved in this paper, and the specific method is described in Section 4.

5.3. Experimental Results

5.3.1. Cost Interval Forecast Results. First, to verify the effectiveness of the methods, SVM, decision tree, plain Bayesian, logistic regression, and k-means methods were tested on the same dataset in this paper. The results are shown in Table 3. Compared with traditional machine learning methods, the multi-task learning method has significantly improved the prediction accuracy for the four types of cost intervals, which raises the accuracy by 5% on treatment cost, 9% on examination cost, 10% on drug cost, and 10% on the total cost.

It can also be seen that the prediction accuracy of our method is also significantly improved compared to that of the logistic regression-only method. The multitask learning model can effectively reduce the reliance on the linear nature of the data using the logistic regression-only method. According to Figure 3, the model convergence speed is improved after the inclusion of the logistic regression approach.

To verify the effectiveness of the framework in this paper, we test the prediction results when the network layer in the framework is replaced by traditional machine learning methods. Moreover, the accuracy of the prediction of subcosts is tested under different total cost prediction results. According to the results in Table 4, when the total cost prediction is correct, the method fails in all cases for the subcost intervals only 21% of the time, which is much lower than traditional machine learning methods. Correspondingly, according to the results in Table 4, when the total cost prediction is incorrect, the sub-cost interval prediction fails

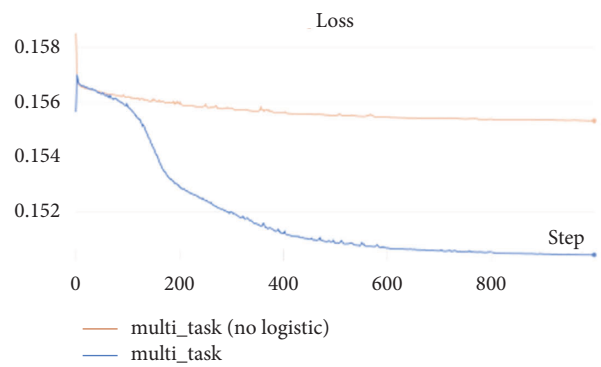

Figure 3: Comparison of loss before and after using the logistic regression method.

in all cases by 24% compared to the correct case, which is the largest improvement compared to the other cases and remains lower than the traditional method. Thus, it can be demonstrated that the neural network method used in this paper, which can better capture the non-linear relationship between subcosts and total costs, outperforms traditional machine learning methods.

Finally, to verify the robustness of the framework, the operation of the network is tested at different learning rates in this paper, and the results are shown in Figure 4. The convergence rate is fast at higher learning rates, but the accuracy as well as the loss gradually converge to the same level at the end. This proves that the network is stable.

5.3.2. Experimental Results on the Feature Importance. Decision trees based on the trained network are shown in Figure 5. Compared with the decision trees built by the traditional method, the prediction accuracy of the our decision tree for the total cost is 0.71, which is much greater than the 0.45 of the traditional decision tree method. The decision tree generated by this method has more Gini nodes with 0 and a clearer judgement process.

Table 4: Percentage of cost forecast (1) Table of other cost projections when total cost projections are correct. (2) Table of other cost projections when total cost projections are incorrect.

Methods	Accuracy	
	Number of correct projections for other costs ≥ 1	Number of correct projections for other costs = 0
SVM	0.54	0.45
DecisionTreeClassifier	0.62	0.38
Naive_Bayes	0.51	0.49
Logistic	0.64	0.36
k-means	0.53	0.47
Multi-task	0.79	0.21
	Accuracy	
Methods	Number of correct projections for other costs ≥ 1	Number of correct projections for other costs = 0
SVM	0.47	0.53
DecisionTreeClassifier	0.51	0.49
Naive_Bayes	0.48	0.52
Logistic	0.41	0.59
k-means	0.49	0.51
Multi-task	0.55	0.45

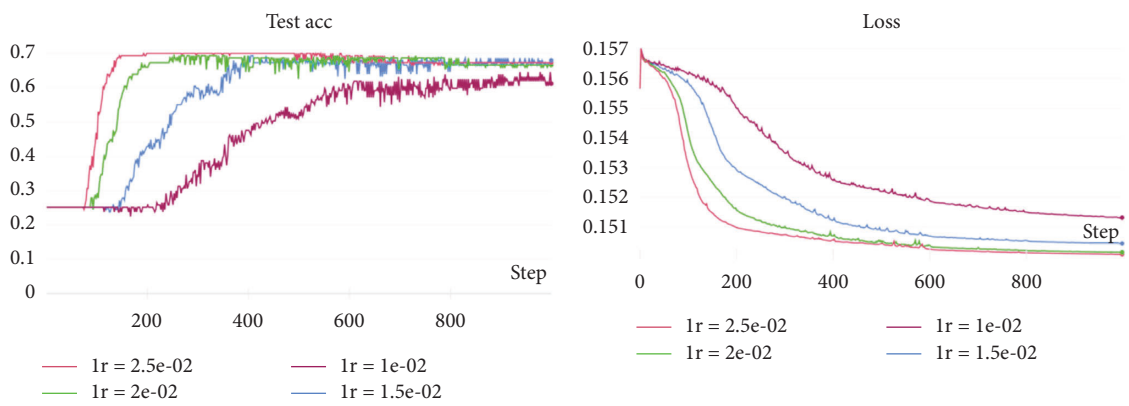

Figure 4: Accuracy and loss of training at different learning rates. (a) Accuracy. (b) Loss.

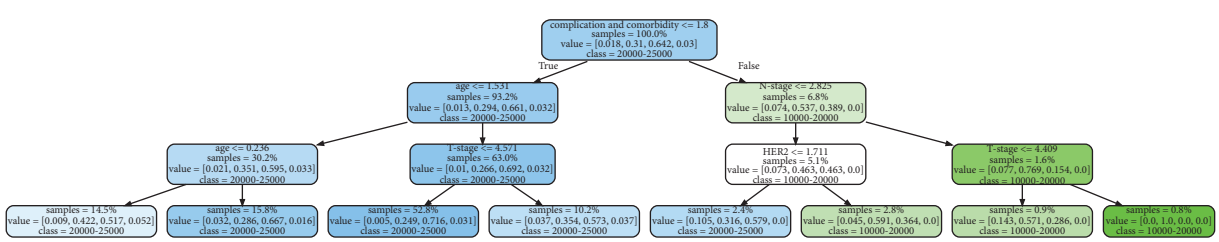

Figure 5: Illustration of the decision tree constructed by our framework.

Table 5: Importance of each feature.

Feature	Importance (descending order)
N-stage	0.1934
T-stage	0.1747
Age	0.1344
Complication and comorbidity	0.1276
HER2	0.1033
Histologic classification	0.0997
M-stage	0.0701

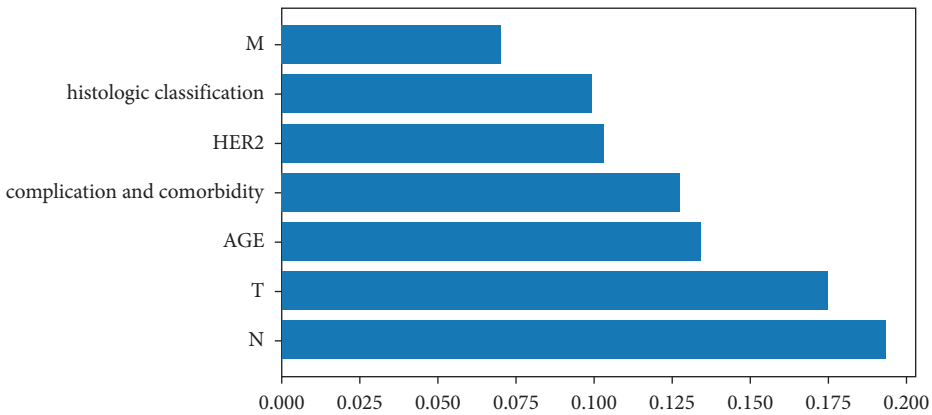

Figure 6: Importance of features.

Based on the generated decision tree, the importance of the features is calculated and the results are shown in Table 5 and Figure 6. Among them, N-stage and T-stage are significantly more important than the last five features, and M-stage is significantly less important than the first six features.

6. Conclusion

This paper presents an interpretable and personalized medical cost interval prediction framework based on multitask learning over data on mobile ends. It can predict total cost intervals based on the subcost intervals of the medical process, and the importance of each feature for cost interval prediction can be obtained using a decision tree approach based on the trained neural network's weight parameters. To begin, this paper uses a multitask learning approach to obtain the subcost intervals in the medical process and mine their correlation to exploit the value of the data; second, the subcosts pass through the full connection layer to predict the total cost intervals; finally, in order to determine the importance of patient characteristics in predicting cost intervals, the decision tree's Gini coefficient calculation method is reconstructed by using full connection layer weights of subcosts to predict total costs. Furthermore, to improve the speed of model training and convergence, the data is preprocessed using logistic regression methods, and ResNet structure is used to keep the network identity Mapping.

Data Availability

The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Ministry of Science and Technology of Sichuan Province Program (No. 2021YFG0018, 2022YFG0038).

References

[1] S. M. Bartsch, M. C. Ferguson, J. A. McKinnell, W. O'Shea, and B. Y. SiegmundLee, "The potential health care costs and resource use associated with COVID-19 in the United States," *Health Affairs*, vol. 39, no. 6, pp. 927–935, 2020.

[2] R. Tipirneni, M. C. Politi, J. T. Kullgren, E. C. Kieffer, S. D. Goold, and A. M. Scherer, "Association between health insurance literacy and avoidance of health care services owing to cost," *JAMA Network Open*, vol. 1, no. 7, p. e184796, 2018.

[3] J. Pang, Y. Huang, Z. Xie, and Z. LiCai, "Collaborative city digital twin for the COVID-19 pandemic: a federated learning solution," *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 759–771, 2021.

[4] A. Agarwal, S. Sharma, V. Kaur, and M. Kaur, "Effect of E-learning on public health and environment during COVID-19 lockdown," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 104–115, June 2021.

[5] S. Cheng, Z. Cai, J. Li, and H. Gao, "Extracting kernel dataset from big sensory data in wireless sensor networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 4, pp. 813–827, 2017.

[6] N. Mihailovic, S. Kocic, and M. Jakovljevic, "Review of diagnosis-related group-based financing of hospital care," *Health Services Research and Managerial Epidemiology*, vol. 3, 2016.

[7] G. Robinson, M. Goldstein, and G. M. Levine, "Impact of nutritional status on DRG length of stay," *Journal of Parenteral and Enteral Nutrition*, vol. 11, no. 1, pp. 49–51, 1987.

[8] H. Fahlevi, I. Irsyadillah, M. Indriani, and S. O. Rina, "DRG-based payment system and management accounting changes in an Indonesian public hospital: exploring potential roles of big data analytics[J]," *Journal of Accounting and Organizational Change*, vol. 18, p. 4, 2021.

[9] A. A. H. de Hond, A. M. Leeuwenberg, L. Hooft et al., "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review[J]," *Npj Digital Medicine*, vol. 5, no. 1, pp. 1–13, 2022.

[10] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Network*, vol. 32, no. 4, pp. 8–14, 2018.

[11] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.

[12] L. Yang, X. Chen, Y. Luo, X. Wang, and W. Wang, "IDEA: a utility-enhanced approach to incomplete data stream anonymization," *Tsinghua Science and Technology*, vol. 27, no. 1, pp. 127–140, 2022.

[13] R. Manne and S. C. Kantheti, "Application of artificial intelligence in healthcare: chances and challenges," *Current Journal of Applied Science and Technology*, vol. 40, no. 6, pp. 78–89, 2021.

[14] V. Shankaran, S. Chennupati, H. Sanchez, L. F. Sun, and B. AlyHealeySeal, "Clinical characteristics, treatment patterns, and healthcare costs and utilization for hepatocellular carcinoma (HCC) patients treated at a large referral center in Washington state 2007-2018," *Journal of Hepatocellular Carcinoma*, vol. 8, pp. 1597–1606, 2021.

[15] B. E. Saelens, R. T. Meenan, E. M. Keast, Y. Frank, D. Kuntz, and S. P. Fortmann, "Transit use and health care costs: a cross-sectional analysis," *Journal of Transport & Health*, vol. 24, Article ID 101294, 2022.

[16] J. L. Moran, P. J. Solomon, A. R. Peisach, and J. Martin, "New models for old questions: generalized linear models for cost prediction," *Journal of Evaluation in Clinical Practice*, vol. 13, no. 3, pp. 381–389, 2007.

[17] A. I. Taloba, A. El-Aziz, M. Rasha, H. M. Alshanbari, and A. A. El-Bagoury, "Estimation and prediction of hospitalization and medical care costs using regression in machine learning[J]," *Journal of Healthcare Engineering*, vol. 2022, Article ID 7969220, 2022.

[18] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 577–590, 2018.

[19] I. K. Nti, J. A. Quarcoo, J. Fosu, and G. K. Fosu, "A mini-review of machine learning in big data analytics: applications, challenges, and prospects," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 81–97, 2022.

[20] M. A. Morid, K. Kawamoto, T. Ault, J. Dorius, and S. Abdelrahman, "Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation," in *Proceedings of the AMIA Annual Symposium Proceedings. American Medical Informatics Association*, pp. 1312–1321, Beijing China, June 2017.

[21] Z. Zheng and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.

[22] J. Pang, Y. Huang, Z. Xie, and Z. HanCai, "Realizing the heterogeneity: a self-organized federated learning framework for IoT," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3088–3098, 2021.

[23] J. Tie, X. Pan, and Y. Pan, "Metabolite-disease association prediction algorithm combining DeepWalk and random forest," *Tsinghua Science and Technology*, vol. 27, no. 1, pp. 58–67, Feb. 2022.

[24] B. Panay, N. Baloian, J. A. Pino, S. Peñafiel, H. Sanson, and N. Bersano, "Predicting health care costs using evidence regression[J]," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 31, no. 1, p. 74, 2019.

[25] R. Tkachenko, I. Izonin, N. Kryvinska, V. Chopyak, N. Lotoshynska, and D. Danylyuk, "Piecewise-linear approach for medical insurance costs prediction using SGTM neural-like structure," *IDDM*, vol. 21, pp. 170–179, 2018.

[26] T. Takeshima, S. Keino, R. Aoki, and K. MatsuiIwasaki, "Development of medical cost prediction model based on statistical machine learning using health insurance claims data," *Value in Health*, vol. 21, p. S97, 2018.

[27] D. Bertsimas, M. V. Bjarnadóttir, M. A. Kane, P. Kryder, and W. G. Vempala, "Algorithmic prediction of health-care costs," *Operations Research*, vol. 56, no. 6, pp. 1382–1392, 2008.

[28] S. Sushmita, S. Newman, J. Marquardt, P. Ram, V. Prasad, and A. Teredesai, "Population cost prediction on public healthcare datasets[C]," in *Proceedings of the 5th International Conference on Digital Health*, pp. 87–94, Florence, Italy, May 2015.

[29] X. Zeng, S. Moosavinasab, E. J. D. Lin, L. Simon, B. Razvan, and L. Chang, "Distributed representation of patients and its use for medical cost prediction," 2019, https://arxiv.org/abs/1909.07157.

[30] Z. He, Z. Cai, S. Cheng, and X. Wang, "Approximate aggregation for tracking quantiles and range countings in wireless sensor networks," *Theoretical Computer Science*, vol. 607, pp. 381–390, 2015.