Forecast of Hospitalization Costs of Child Patients Based on Machine Learning Methods and Multiple Classification

Chenguang Wang [1], Xinyi Pan [1], Lishan Ye [2], Weifen Zhuang [3], and Fei Ma [1]
[1] Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
[2] Information Department, Zhongshan Hospital of Xiamen University, Xiamen, China
[3] School of Management, Xiamen University, Xiamen, China
Email: wfzhuang@xmu.edu.cn, fei.ma@xjtlu.edu.cn

***Abstract*—In the paper, Random Forest algorithm (RF), bagging and error-correction output code model (ECOC) were employed to predict the clinic expenditure of infantile patients with data consisting of records extracted from a hospital system. Throughout the modelling, the training set utilized 80% of the records selected from the original data set in random and the rest of data were used in the test set. The RF received superior predictive accuracy than bagging and ECOC, with RMSE being 0.138, R^2 being 0.928, $|R|$ being 0.885, and Acc $\pm$ 1 being 85.5%. Additionally, both RF and bagging obtained impressive performances on different types of charges, achieving over 80% accuracy on average. Besides, among different types of information, clinic features obtained better results, with RMSE being around 0.2, R^2 being more than 0.8, $|R|$ being larger than 0.7 and Acc being nearly 65%. In comparison, the random forest and bagging performed slightly better than ECOC models in most fields. To summarize, all three types of methods could obtain good performance during prediction, with accuracy of nearly 80%, and clinic features could provide models with higher accuracy among all fields of information.**

***Index Terms*—hospitalization costs, forecasts, random forest, error-correction output code, multiple classification**

I. Introduction

Since 2004, the past decade has witnessed a large number of studies supporting the field of medical care prediction. In 2014, Muthoni *et al*. wrote the review of the forecast of the patient volume in hospitals, discussing the pros and cons of existing models and innovations [1]. For example, Joy and Jones [2] attempted to predict the bed demand by combining neural network and ARIMA model in 2005 in order to provide hospitals with an optimization plan for limited medical resources. In the same year, Mackay and Lee completed their research on bed prediction by implementing compartmental flow models [3]. In addition to predicting medical needs, studies also focused on forecasting the volume of patients. Jones *et al*., Schweigler *et al*. and Kam *et al*. achieved significant progress successively in 2008 [4], 2009 [5] and 2010 [6]. In their research, time series combined with auxiliary methods were used to predict the patient population.

In pediatric departments, statistical techniques have been widely applied to the studies of medical information and the prediction of individual status. LeDuc *et al*. attempted to forecast pediatric re-hospitalization by analyzing the features of recidivism, indicating certain relationships between features and diseases [7]. Besides, it was discussed whether or not infantile ages had impacts on the frequency of clinic admissions with comparisons between initial diagnosis and re-consultation [8]. Graauw *et al*. also wrote the review of the estimation of activity related expenditures among children and adolescents by using predictive models [9]. Our preceding study adopted ensemble methods to predict the length of hospital stay for infants with extensive clinical data [10].

The purpose of this study is to forecast the total clinical expenditure of infantile patients with modified random forest and bagging algorithm. Different from the length of stay, the clinical expenditure containing massive data could hardly be predicted directly with classification models. Therefore, we first classified the expenditure into different subgroups with the same numerical span and next labelled these subgroups in the ascending order. Similar manipulations were applied to the set of charge3 (Cost of Western Medicine), charge7 (Treatment Fee) and charge10 (Testing Fee). Furthermore, the multi-classification model and Error-Correcting-Output Codes model (ECOC) were employed to predict the expenses. Results are compared with those of random forest to test which method is superior in terms of prediction. Ultimately, the final results obtained were highly consistent with the results of the study [11].

The paper is organized as following. Section II discusses data preprocessing and predictive models in details. Section III describes the experiments, and finally Section IV discusses the performances of each model and makes conclusions.

II. Methodology

A. Data Preprocessing

There were 11,206 medical records being applied in total, containing 351 distinct feature variables. In this

Manuscript received July 3, 2018; revised November 5, 2018.

doi: 10.12720/jait.9.4.89-96

experiment, we added 43 new generated variables to the original data set, including the information of age, duration of treatment and levels of cost in epidemic diseases. Sensitive information of patients were excluded or modified slightly for the purpose of privacy.

1) General Manipulation: The data preprocessing consists of two steps: selection of existed features and manipulation of selected variables. During the selection, dozens of administrative categories and variables with insufficient information are deleted. For instance, the information of admission physicians and department directors would not be considered as relevant variables because nearly most of patients have similar label.

During the manipulation, we convert the original features into the `binary' form and generate new variables. For features with limited distinctive conditions, generated subgroups would be developed based on two indicators, '1' and '0', covering the entire information of the primary features. Specifically, the binary indicator '1' suggests that entities belong to this category while 0 indicates that they do not apply to each condition. Take gender for an example, such features will be processed as Fig. 1 shows. For features with many distinctive conditions, we have developed a method called 'dominant expansion', which focuses on the tail of entries. Entries of conditions shared among patients would be organized into one subgroup.

ID	Gender
1	Male
2	Male
3	Female
4	Female
5	Male

ID	Male	Female	Unknown
1	1	0	0
2	1	0	0
3	0	1	0
4	0	1	0
5	1	0	0

Figure 1. Manipulation of feature 'gender'.

After feature processing, we divide 351 fields into different scenarios based on the types of information: (1) Patient information includes basic personal information, such as age and gender. (2) Clinic information relates to the patients' medical conditions, such as diagnosis codes and duration of treatment. (3) The administrative field contains information on patients' administrative records in hospital. Table I demonstrates all the fields in the matrix.

TABLE I. SUMMARY OF THE DATA PREPROCESSING

Sources	Scheme	Type	Manipulation	Num
Age	P	N	B	15
Gender	P	C	B	3
Admission Times	C	N	1/2/3/4/5+	5
Charge Type	A	C	B	2
Response Type	A	C	1/2/2+	3
Admission Path	A	C	B	5
Local Flag	A	C	1/2/NULL	3
Days in Hospital	C	N	B;D	26
Charge3	N/A	N	B;D	49
Charge7	N/A	N	B;D	37
Charge10	N/A	N	B;D	34
Admission Diagnosis	C	C	B;D	45
Clinic Diagnosis	C	C	B;D	44
Discharge Diagnosis	C	C	B;D	95
Blood Type	P	C	B	7
Country	P	C	B	3
Employer District	P	C	D	2
Home Code	P	C	D	3
Nation Code	P	C	D	2
Relative Code	P	C	B	5
Occupational Code	P	C	B	7
Relation Code	P	C	D	2
Days(Diag.-Admission)	C	N	B	3
Days(Discharge-Leave)	C	N	B	3
Med.Card Condition	A	C	B	2
Change Dept (Admission, Discharge)	A	C	B	1
Change Ward (Admission, Discharge)	A	C	B	1
Comparison of A/C/D	C	C	D	3
Admission Status-1	C	N	1/2/3/4/NULL	5
Admission Status-2	C	N	1/2/3/4/NULL	5
Discharge Diag.Num	C	N	1/2/3/4/5/6/7+	7
Discharge Diag.Status	C	N	1/2/3/4/5/NULL	6
Discharge Diag.Type	C	N	1/2/3/4/5/6/7/8	8
Discharge Status	C	N	1/2/3/4/5/6/7/8/9/NULL	10
Quality Level	A	N	1/2/3/NULL	4
Special Individual	A	N	B	1
Archive Mark	A	N	B	2
Cp Status	A	N	B	5
Last Admission Times	C	N	1/2/3/4/5/6/7+	7
Medical Care Flag	A	N	B	1
Level of Expenditure	C	N	B;D	13

Level of Source Variables: Patient Information (P), Clinic Information (C), Administrative (A); Type of Source Variables: numeric (N), categorial (C); Manipulation: Computation and Expansion. Computation: the method focus on dividing numerical variables into different bins in the form of binary mode; Expansion: binary expansion (B), dominant expansion (D), not applicable (N/A); Num: Number of features.

2) Classification of Expenditure: As mentioned above, all types of clinic expense need be processed before prediction. In other words, each patient would be grouped into different subgroups according to their clinic expenditures.

Taking the total expenditure for example, the cost ranges from 65 to 62,323. Given the wide range of expenditures, we set six units to distinguish between '200', '1000', '2000', '3000', '4000' and 'high'. The unit '200' is used to measure patients whose charges are less than 10,000, the unit '1000' for [10001, 20000] and so on. Specifically, when the patient spends between 0 and 200, the patient would be marked '1' and the label '2' means

the patient spends between 201 and 400. The grouping follows such process. Finally, the group measured in "high" units contains patients who spent over 50,000, for a total of 71 unique tags.

Similarly, charge3, charge7 and charge10 follow the manipulation but change the units of length. To fit the expenditure better, charge3, charge7 and charge10 have units in terms of '100' and '1000'. There are 49 labels in charge3 and 34 labels in charge10. For charge7, patients who spend more than 10000 are labeled as 'high', and the highest label is 37.

After grouping, according to the total charge prediction, patients labelled as '5' spend between 801 and 1000, while the label '10' indicates that the individuals' costs range from 1801 to 2000.

3) Classification of Diagnostic Information: With ideas from [12]-[14], we develop new methods to manipulate diagnostic features.

The whole process works as follow: Firstly, all ICD-10 codes (the International Classification of Diseases code, version 10) are directly classified according to their medical categories. For example, the code 'J18.X' represents the subcategory of pneumonia diseases, in which J18.000 stands for bronchopneumonia and J18.004 for other Asthmatic bronchopneumonia and. The process includes codes of admission, clinic and discharge. After classifying the disease categories, we would categorize the levels of expenditure for each category. Taking 'J18.X' as an instance, the expenditure of patients with 'J18.X' diseases would be divided into three sets equally, referred as 'High', 'Middle' and 'Low' respectively. Patients are labeled to each new subgroup by previous manipulation. As illustrated in Fig. 2, these features (No.337-351) have a positive effect during modelling.

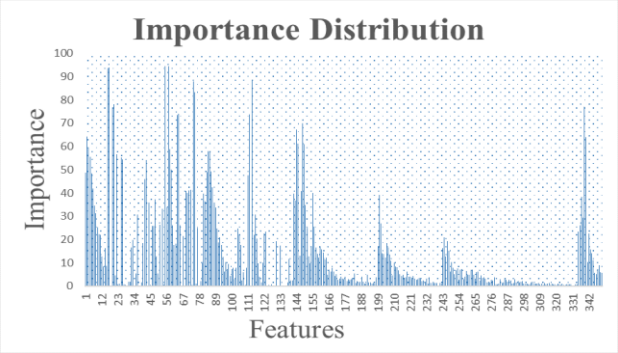

Figure 2. Distribution of importance. Importance distribution indicates the predictive power of each feature.

B. Predictive Models

1) Random Forest: The random decision forest algorithm was firstly proposed by Ho in 1995 [15] and Breiman provided a full introduction version after 6 years [16]. The revised version added more methods, such as randomized node optimization, to construct trees.

As an ensemble learning method, random forests rely on weak learner named decision trees, and these trees are developed on a bootstrap replica of tested data independently. Suppose S=$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ denotes the entire training samples. The random forests extract S*n samples out of d features randomly, and these extracted data form the training set S_i which is prepared to construct the decision tree h_i. Here, s ∈(0,1] is a size-controlling indicator.

In each bootstrap replica, a tree would be created by recursively splitting the data into two sets. Then, N_{tree} features are extracted randomly from the whole feature set at each splitting point while the best feature variable would be selected from the set of N_{tree} features to divide the data into two sets. After repeating m times, the model would generate m decision trees. In order to predict a new sample. The sample would be input into each tree to be voted for 0 or 1. Also, the result with greatest weight would be accepted as the classification/prediction of the new sample. As for the optimal number of decision trees, testing the model by setting different numbers of trees and recording their classification error together can determine the optimal tree volume. Table II reveals the details of steps of the model.

TABLE II. THE ALGORITHMS PROCESS OF RANDOM FOREST

Method	Random Forests
Input:	1.Training Set S = (x_i,y_i), i=1,2,...,n, (X,Y) ∈ R^d × R 2.Testing Sample x_t ∈ R^d For i=1,2, … , N_{tree} (1) Sampling the original training set S by boostrap and generating the training set S_i (2) Generating a tree without pruning by S_i a) Select M_{try} out of d features randomly b) Pick the optimal feature from M_{try} at each node using Gini index c) Split branches till the tree has receive its maximun End
Output:	1. Tree Set h_i, i=1,2, … , N_{tree} 2. For testing samples x_t, decision tree h_i output $h_i(x_t)$ Regression Tree: $f(x_t) = \frac{1}{N_{tree}} \sum_{i=1}^{N_{tree}} h_i(x_t)$ Classification Tree: $f(x_t)$ = majority vote $\{h_i(x_t)\}_{i=1}^{N_{tree}}$

Gini Index represents the probability that a randomly chosen sample has wrong classification in a sample set. Gini(p) = $\sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} {p_k}^2$ where p_k represents the probability that the chosen sample belongs to k class and K is the number of different classes in the sample set. An output of feature importance calculated in bagging method based on gini index is presented in Table V.

2) Bagging: The method 'Bagging', abbreviated from 'Bootstrap aggregating', was also proposed by Leo Breiman for improving the quality of classification [17]. The study tried to decrease the deviation of the model and avoid overfitting by combining classifications of developed training sets randomly.

The only difference between 'Bagging' and 'Random Forest' lies in the number of features which are used for selection at each node. During the process of bagging, all features will be selected for distinction while random forest will only pick M_{try} out of all features randomly and select the best one.

3) Error-Correcting Output Codes Model: Error-correcting code could be applied to detect and correct errors between strings. Several studies had already suggested the prediction capabilities of ECOC model [18]-[20].

In the method, each class in the row represents the levels of expenditure while the column shows 351 fields of feature. Each label (we use labels to denote classes in this case) is represented as C_i (i = 1, … , n). A unique binary string of length 351 would be assigned to each class as codeword, from which the binary functions could be learned by applying the decision tree or neural network algorithms. During the training of each class, the codewords generated by 351 binary functions could specify the outputs of each class, and each class has a unique codeword.

After the setup is completed, the model is ready for predicting through code identification. For instance, let us add a new value *M* into this model for classification/ prediction. Firstly, a 351-bit string is constructed by generating each binary function as a new identifiable codeword M = {(m_1), (m_1), ... , (m_n)}. Then, the string would be compared with C_i respecting each of codewords in the column and compute the distance/difference *D* between C_i and itself. The distance function D is defined as

$$D(M, C_i) = \sum_{j=0}^{M} \left| m_i - C_{i,j} \right| \quad (1)$$

Finally, after comparison M would be assigned to the label with the smallest D value. Fig. 3 revealed the details of process.

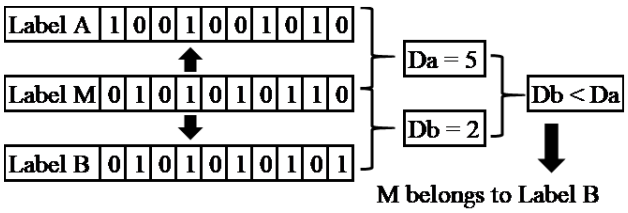

Figure 3. An example of process of ECOC.

III. EXPERIMENT

A. Performance Measurement

The report uses the indicators in Table III to evaluate the performance of each model.

In Table III, RMSE (The Root-Mean-Square Error) is the difference between the logarithm of the true expenditure and the predicted one. Then, MAPE (The Mean Absolute Percentage Error) [6], [21] and MSE (The Mean- Square Error) show the average loss of absolute percentage and the errors from the average square difference indicate the true and estimate costs respectively. Plus, we also use Spearman rank correlation coefficient. It revealed the goodness of predictability from classification models to approximate the labels of expenditures. Besides, R^2 [22] is used to display the quality of model to explain the variation, revealing the relation between the mean of actual results and the improvement of classification predictability, while $|R|$ [23], induced from R^2, would be applied to measures the decrease between the sum of absolute values of the residuals and real values. Finally, we also design a new measurement, ± 1 Acc. Different from the exact results, the measurement tolerates the generalization error to some extent, namely, the label measured by the minimum unit would be considered as correct result within ± 1 deviation.

TABLE III. PERFORMANCE METRICS

RMSE = $\sqrt{\frac{1}{N}\sum_{i=1}^{N}[\ln(p_i+1)-\ln(a_i+1)]^2}$,
$R^2 = 1 - \frac{\sum(a_i-p_i)^2}{\sum(a_i-m)^2}$,
$\|R\| = 1 - \frac{\sum\|a_i-p_i\|}{\sum\|a_i-m_{med}\|}$,
$\rho = 1 - \frac{6\sum[\ln(p_i+1)-\ln(a_i+1)]^2}{N(N^2-1)}$,
MSE = $\frac{1}{N}\sum(a_i-p_i)^2$,
MAPE = $\sum\frac{\|a_i-p_i\|}{a_i}\frac{100}{N}$,
$\pm$ Acc = $\frac{\sum\delta(a_i-p_i)}{N}$

N represents the number of entire population; p_i is the predicted value of expenditure for i^{th} patient while a_i is the actual expenditure, i ∈ [1,N]; m is the average expenditure of all records and m_{med} is the median expenditure of all records. δ(x) = 1 when $|x|=|a_i-p_i| \le 1$, while a_i, p_i are measured by minimum unit. δ (x) = 1 when $|x|=|a_i-p_i| = 0$, while a_i, p_i are measured by other units. The measurement, ± 1 Acc, assumes that the minimum expenditure unit could be accepted as a tolerable fluctuation under actual conditions.

B. Performance of Different Groups

Apart from the total expenditure, both models apply the same feature matrix to predict charge3, charge7 and charge1. Furthermore, measurements also reveal the performance of three features schemes. Table IV shows forecast results for different subsets. Throughout the process, 500 trees would be set as the optimal number of trees. As Fig. 4 has shown, after groups of tree volumes are tested by out-of-bag estimation, the estimated loss is stable around 0.45 and stops fluctuating when the number of trees reaches 500. The distribution of importance of Random Forest is shown is Fig. 2. Table V shows the best 15 features selected based on the gini index outputs calculated in Bagging method.

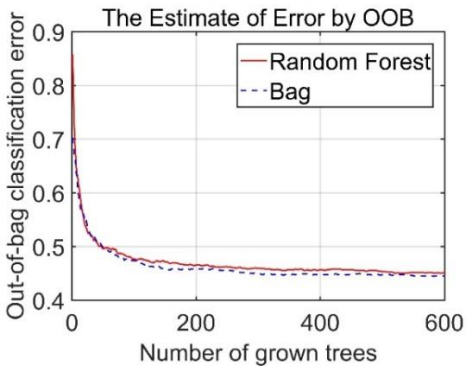

Figure 4. The out-of-bag estimation tested on two models.

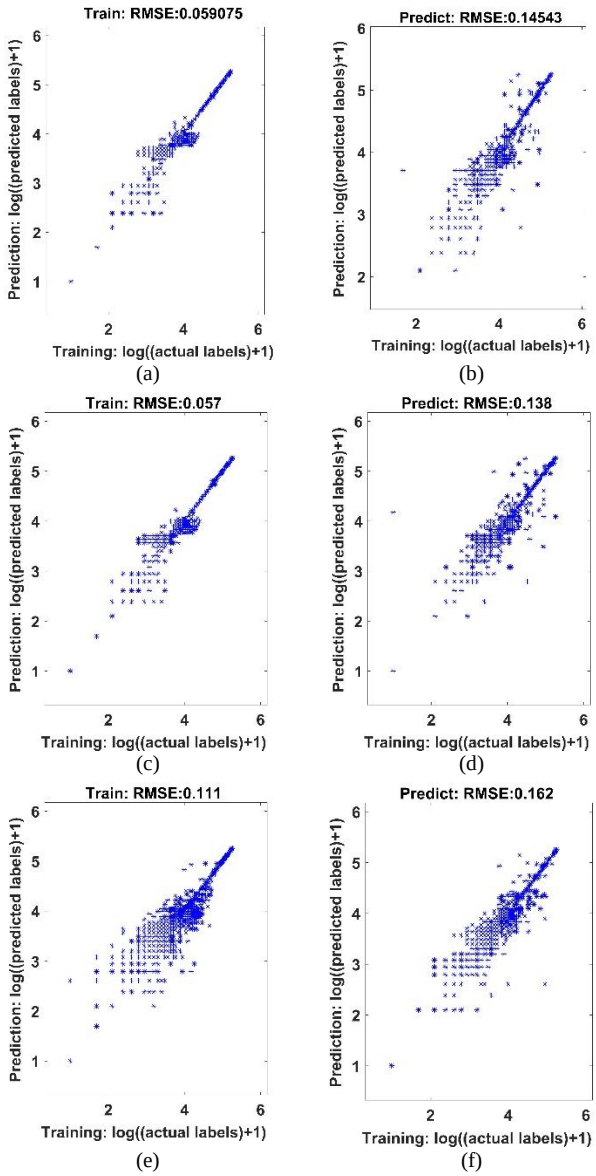

Figure 5. Scatter-plots for models of bag (a,b), Random forest (c,d) and ECOC (e,f); (a), (c), (e) base on training set while (b), (d), (e) base on testing set.

Fig. 5 is the scatter-plots for the classification results from three models. The left subplots are for training while the subplots on the right are for prediction. From these graphs, points scatter randomly when 'training is small while the results of three models start to converge as a line when 'training increases, namely, the difference between actual labels and predicted labels is narrowing as individual expenditures grow. Therefore, the performance verifies the conclusions from the study [11]. Clinic expenditures of high-cost patients are more easily predicable by utilizing medical information as compared to medical expenditures of low-costs individuals. Furthermore, the impressive outcomes also indicate that the form of data of past spending can be regarded as a convincing reference for prediction of future costs.

IV. Discussion and Conclusion

A. Result Analysis

From Table IV, in general, random forest performed better than Bagging and ECOC in most cases.

1) Random Forest: Random forest achieved the best performance in predicting charge7 (Treatment Free) with MSE being 2.843, R^2 being 0.945, $|R|$ being 0.893 and '$\pm$ 1 Acc' being 0.932. Compared with the forecast of other expenditures, the forecast of total expenditure obtained the minimum RMSE (0.138) and maximum MSE (10.894) while the performance of charge3 and charge10 was quite similar in most fields but different on MAPE (8.014 and 12.736 respectively) and MSE (10.006 and 7.021 respectively).

For schemes of prediction, clinic information had better predictive ability than others, with RMSE being 0.182, R^2 being 0.908, $|R|$ being 0.794 and '$\pm$ 1 Acc' being 0.686. However, patient and administrative information both received RMSE around 0.55, '$\pm$ 1 Acc' with nearly 0.25 and negative R^2.

2) Bagging: Bagging also achieved a pretty performance among different types of charges. Charge7 had the best result on prediction of '$\pm$ 1 Acc' and 'Exact' with 0.93 and 0.829 respectively. The result of total expenditure prediction was also impressive, with RMSE being 0.145, R^2 being 0.904, $|R|$ being 0.873 and '$\pm$ 1 Acc' being 0.847.

Still, the result of prediction by using clinic information was superior among the forecasts of all types of information, which had small RMSE (0.215) and tolerable '$\pm$ 1 Acc' (0.672). The measuring indicators R^2 still had negative results in the performance of patient and administrative information, along with the low accuracy.

3) ECOC: Although the general performance of ECOC was not as good as that of those random forest and bagging, the accuracy of '$\pm$ 1 Acc' and 'Exact' was still very high. Similarly, the superior results also stayed on the prediction of charge7, with RMSE being 0.144, MSE being 2.207, R^2 being 0.959, $|R|$ being 0.891 and '$\pm$1 Acc' being 0.912. Compared with charge7, the results on the group with total expenditure had RMSE of 0.162, R^2 of 0.922 and |R| of 0.852.

TABLE IV. Performance Metrics

Data	Dataset	RMSE	ρ	MSE	MAPE	R^2	$\|R\|$	±1 Accuracy	Exact
Random Forest									
Total Expenditure	Test	0.138	1	10.894	5.227	0.928	0.885	0.852	0.812
	Train	0.057	1	0.698	1.195	0.995	0.981	0.963	0.951
Charge 3	Test	0.266	1	10.066	8.014	0.862	0.872	0.874	0.821
	Train	0.083	1	0.381	1.278	0.995	0.983	0.974	0.961
Charge 7	Test	0.147	1	2.843	5.661	0.945	0.893	0.932	0.827

	Train	0.065	1	0.167	1.778	0.997	0.978	0.982	0.936
Charge 10	Test	0.259	1	7.021	12.736	0.827	0.814	0.846	0.803
	Train	0.111	1	0.844	c2.461	0.979	0.974	0.976	0.971
Patient Information	Test	0.512	1	162.718	50.207	-0.092	0.077	0.306	0.213
	Train	0.483	1	148.064	46.337	0.033	0.142	0.334	0.244
Clinic Information	Test	0.182	1	14.223	10.586	0.908	0.794	0.686	0.551
	Train	0.145	1	6.915	7.77	0.955	0.857	0.752	0.651
Administrative Information	Test	0.598	1	251.139	67.118	-0.626	-0.195	0.215	0.116
	Train	0.587	1	227.218	66.444	-0.484	-0.144	0.227	0.131
Bagging									
Total Expenditure	Test	0.145	1	13.934	5.627	0.904	0.873	0.847	0.788
	Train	0.059	1	0.72	1.217	0.995	0.98	0.963	0.952
Charge 3	Test	0.298	1	12.889	7.586	0.842	0.861	0.875	0.818
	Train	0.0839	1	0.418	1.29	0.995	0.983	0.974	0.961
Charge 7	Test	0.152	1	3.225	5.656	0.942	0.898	0.93	0.829
	Train	0.065	1	0.164	1.786	0.997	0.978	0.983	0.936
Charge 10	Test	0.252	1	6.193	12.132	0.843	0.823	0.851	0.809
	Train	0.106	1	0.812	2.233	0.981	0.974	0.976	0.971
Patient Information	Test	0.494	1	156.953	47.815	-0.135	0.054	0.311	0.215
	Train	0.484	1	147.972	46.546	0.034	0.141	0.334	0.245
Clinic Information	Test	0.183	1	15.551	10.981	0.911	0.792	0.672	0.552
	Train	0.144	1	6.885	7.723	0.955	0.858	0.753	0.651
Administrative Information	Test	0.592	1	234.297	66.754	-0.591	-0.209	0.212	0.112
	Train	0.586	1	229.062	66.186	-0.496	-0.149	0.227	0.131
ECOC									
Total Expenditure	Test	0.162	1	11.794	7.321	0.922	0.852	0.791	0.714
	Train	0.111	1	3.207	3.762	0.979	0.936	0.898	0.862
Charge 3	Test	0.283	1	12.043	13.143	0.843	0.813	0.812	0.736
	Train	0.161	1	2.998	6.567	0.962	0.921	0.898	0.863
Charge 7	Test	0.144	1	2.207	6.672	0.959	0.891	0.912	0.785
	Train	0.092	1	0.567	3.912	0.989	0.948	0.956	0.881
Charge 10	Test	0.305	1	9.527	19.964	0.757	0.732	0.766	0.721
	Train	0.251	1	5.568	12.169	0.862	0.851	0.877	0.854
Patient Information	Test	0.502	1	164.332	45.937	-0.089	0.063	0.291	0.193
	Train	0.503	1	166.839	49.308	-0.091	0.088	0.323	0.233
Clinic Information	Test	0.215	1	22.837	11.882	0.849	0.739	0.645	0.507
	Train	0.171	1	11.461	9.462	0.925	0.813	0.712	0.605
Administrative Information	Test	0.579	1	225.532	64.251	-0.564	-0.193	0.215	0.121
	Train	0.595	1	233.707	68.716	-0.526	-0.172	0.223	0.129

Charge3: The cost of Western Medicine; Charge7: Treatment Cost; Charge10: Testing Fee. Noticeably, ρ equaled to 1 in each result, implying that predicted and actual number of levels have a perfect tendency as models have developed.

TABLE V. 15 TOP IMPORTANT FEATURES

Features	55	57	17	16	114	74	75	20	19	64	63	112	339	148	144
Values(10^{-6})	5.06	4.93	4.92	4.89	4.72	4.65	4.36	4.07	4.04	3.93	3.87	3.87	3.85	3.67	3.54

Top 15 important features measured by using Gini index output calculated in Bagging. The detailed features are: 55: hospital location (same city); 57: home (unknown); 17: gender(female); 16: gender(male); 114: admission status (unknown); 74: admission path (emergency treatment); 75: admission path (outpatient treatment); 20: local citizen (no); 19: local citizen (yes); 64: diagnosis status (improvement); 63: diagnosis status (recovery); 112: admission status (regular); 339: J18 (low cost); 148: last admission times (1); 144: cp status (completed).

The results of prediction on patient and administrative information were also inferior, with RMSE being around 0.535, MSE being 195, MAPE being around 55. Both R^2 and |R| were negative and their '±1 Acc' were very low.

4) Feature impact: From Fig. 2 and Table V, the highly important features are usually distribute in patient information and clinic information. The best 15 features processed by the Gini index are analyzed in details. The table shows that the most important distinguishing characteristic is the location of the hospital, namely, whether the hospital and patients home are in the same city. Then, other essential features are related to the information of gender, admission states and diagnosis status. However, although the features of patients and clinic information have a significant impact on classification, the performance based on different information does not behave well. Hence, we have the same conclusion with [24], which is that the model generated by combined features provides a better performance compared to models with individual information.

5) Time Efficiency: All methods can be completed within a reasonable time. Averagely ECOC took around 75 seconds, faster than the other two models (Bagging: 92 seconds; Random Forest: 88 seconds). For this application, however, the running time within 100 seconds is acceptable.

B. Comparison with Other Studies

Although studies on clinic expenditures prediction have developed for decades, appropriate references for comparison are still difficult to find because different methods and the data used would affect results significantly.

Firstly, we would compare the results with the study [11], where R^2 is 0.22 and |R| is 0.194 from the clustering algorithm and classification tree algorithm has 0.204 in R^2 and 0.182 for |R|. As mentioned before, our modified random forest obtains 0.928 and 0.885 respectively for R^2 and |R| while bagging receives 0.904 for R^2 and 0.873 for |R|. Furthermore, ECOC also receives 0.922 and 0.852 respectively, which are superior to previous performance.

In [25], linear regression models and corresponding adjusted R^2 statistics obtains R^2 from 0.115 to 0.F181 by using different subgroups of data. Obviously, the values of R^2 are far smaller than our results. [14] utilizes Diagnostic Cost Group Hierarchical Condition Category (DCG/HCC) model to predict pharmacy costs and other medical costs. It receives the highest R^2 (0.482) by using drug-based models. The outcome of R^2 is lower than the result obtained in this study.

ACKNOWLEDGMENT

This study is supported by the Natural Science Foundation of Jiangsu Province (BK20171237) and the National Natural Science Foundation of China (71672160).

REFERENCES

[1] G. J. Muthoni, S. Kimani, and J. Wafula, "Review of predicting number of patients in the queue in the hospital using Monte Carlo simulation," *International Journal of Computer Science Issues*, 2014.

[2] M. P. Joy and S. Jones, "Predicting bed demand in a hospital using neural networks and arima models: A hybrid approach," in *Proc. European Symposium on Artificial Neural Networks*, Bruges, Belgium, 2005, pp. 127–132.

[3] M. Mackay and M. Lee, "Choice of models for the analysis and forecasting of hospital beds," *Health Care Management Science*, vol. 8, no. 3, pp. 221–230, 2005.

[4] S. S. Jones, R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch, and G. L. Snow, "A multivariate time series approach to modeling and forecasting demand in the emergency department," *Journal of Biomedical Informatics*, vol. 42, no. 1, p. 123, 2009.

[5] L. M. Schweigler, J. S. Desmond, M. L. Mccarthy, K. J. Bukowski, E. L. Ionides, and J. G. Younger, "Forecasting models of emergency department crowding," *Academic Emergency Medicine Official Journal of the Society for Academic Emergency Medicine*, vol. 16, no. 4, p. 301, 2009.

[6] H. J. Kam, O. S. Jin, and R. W. Park, "Prediction of daily patient numbers for a regional emergency medical center using time series analysis," *Healthcare Informatics Research*, vol. 16, no. 3, pp. 158–165, 2010.

[7] K. Leduc, H. Rosebrook, M. Rannie, and D. Gao, "Pediatric emergency department recidivism: Demographic characteristics and diagnostic predictors," *Journal of Emergency Nursing*, vol. 32, no. 2, pp. 131–138, 2006.

[8] G. A. Rivas, M. G. Manrique, L. L. Butragueno, G. S. Mesa, S. A. Campos, I. V. Fernandez, S. R. Moreno, and J. M. A. Mulet, "Frequent users in paediatric emergency departments. Who are they? Why do they consult?" *Anales De Pediatria*, 2016.

[9] S. M. D. Graauw, J. F. D. Groot, M. V. Brussel, M. F. Streur, and T. Takken, "Review of prediction models to estimate activity-related energy expenditure in children and adolescents," *Int. J Pediatr.*, vol. 2010, no. 2, p. 489304, 2010.

[10] C. Wang, X. Dong, L. Yu, L. Ye, W. Zhuang, and F. Ma, "Prediction of days in hospital for children using random forest," in *Proc. Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2017.

[11] D. Bertsimas, R. Pandey, R. Pandey, R. Pandey, S. Vempala, and G. Wang, "Algorithmic prediction of health-care costs," *Operations Research*, vol. 56, no. 6, pp. 1382–1392, 2008.

[12] A. S. Ash and S. Byrne-Logan, "How well do models work? Predicting health care costs," *Proceedings of the Section on Statistics in Epidemiology of the American Statistics Association*, 1998.

[13] A. S. Ash, R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, E. Mackay, and W. Yu, "Using diagnoses to describe populations and predict costs," *Health Care Financing Review*, vol. 21, no. 3, p. 7, 2000.

[14] Y. Zhao, A. S. Ash, R. P. Ellis, J. Z. Ayanian, G. C. Pope, B. Bowen, and L. Weyuker, "Predicting pharmacy costs and other medical costs using diagnoses and drug claims," *Medical Care*, vol. 43, no. 1, p. 34, 2005.

[15] T. K. Ho, "Random decision forests," in *Proc. International Conference on Document Analysis and Recognition*, 1995, p. 278.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] L. Brieman, "Bagging predictors," *Machine Learning*, vol. 24, 1996.

[18] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 263–286, 1995.

[19] J. Gareth and H. Trevor, "The error coding method and picts," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 377–387, 1998.

[20] G. Armano, C. Chira, and N. Hatami, "Error-correcting output codes for multi-label text categorization," 2013.

[21] S. Jones, A. Thomas, S. Evans, S. Welch, P. Haug, and G. Show, "Forecasting daily patient volumes in the emergency department," *Clinical Practice*, vol. 15, no. 2, pp. 159–170, 2008.

[22] R. B. Cumming, D. Knutson, B. A. Cameron, and B. Derrick, "A comparative analysis of claims-based methods of health risk assessment for commercial populations," 2002.

[23] D. Bertsimas, M. V. Bjarnadottir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala, and G. Wang, "Algorithmic prediction of health care costs and discovery of medical knowledge," *Operations Research*, vol. 56, 2007.

[24] Y. Xie, S. Neubauer, G. Schreier, S. J. Redmond, and N. H. Lovell, "Impact of hierarchies of clinical codes on predicting future days in hospital," *Conf. Proc IEEE Eng. Med. Biol. Soc.*, pp. 6852–6855, 2015.

[25] J. F. Farley, C. R. Harley, and J. W. Devine, "A comparison of comorbidity measurements to predict healthcare expenditures," *American Journal of Managed Care*, vol. 12, no. 2, p. 110, 2006.

Lishan Ye received her M.Sc. degree from the School of Computer and Information Technology, Xiamen University. She is an associate professor and a senior engineer. She has been the director of information department of Zhongshan Hospital of Xiamen University since 2011. She also serves as a member of the Health Information Standards Committee of the China Health Information and Health Medical Big Data Society, the national committee member of the Electronic Medical Record and Hospital Information Professional Committee of the Chinese Health Information Society, the national committee member of the China Hospital Association Information Committee (CHIMA), and the Medical Information of the Chinese Medical Association Branch Youth Committee, vice president of Xiamen Health Information Association.

Weifen Zhuang is an associate professor at School of Management, Xiamen University (XMU) and the director of Research Center for Operations and Technology Management in Contemporary Business and Healthcare. Prior to joining XMU in 2011, she was a Postdoctoral Research Fellow at McGill University and the Chinese University of Hong Kong. She obtained her Ph.D. in Operations Management from Nanyang Business School, Nanyang Technological University, Singapore. Dr. Zhuangs research focuses on Healthcare Operations Management, Revenue Management and Pricing.

Fei Ma received his B.Sc and M.Sc. degrees of Computational Mathematics from Xiamen University, China in 1999 and 2002. He received his Ph.D. degree in Applied Mathematics from Flinders University, Australia in 2008. He worked as a software engineering at Kingdee Co. Ltd. and also worked at Symbion as an analyst. Since 2012, Dr. Ma has been with the Xi'an Jiaotong-Liverpool University, where he is currently an Associate Professor and Deputy Head of the Mathematical Sciences Department. Dr. Ma's research interests include image processing, medical image analysis, forecasting methods, automated guided vehicle optimisation, artificial intelligence and nonnegative matrix.