INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2020.14.103 Volume 14, 2020

Statistics and forecasting analysis on hospital garbage data

Benyou Wang
West Anhui University
Lu'an, Anhui,
P.R. China

Sihai Zhang
University of Science & Technology of China
Hefei, Anhui,
P.R. China

Li Gu
School of Information Science and Technology, University of Science & Technology of China
Hefei, Anhui,
P.R. China

Received: October 10, 2020. Revised: November 12, 2020. Accepted: November 16, 2020. Published: November 19, 2020.

Abstract—For a long time, China's medical problem is very serious. There are very few researches on medical waste data, which can not provide enough evidence for managers. Therefore, combining some methods of data analysis and data mining to analyze the medical waste data.In this study, based on the collection of the hospital garbage data of over five years from some area in China, the hospital garbage data are analyzed with consideration of the location, the hospital level, hospital beds and number of doctors and staff members, by using some data analysis and data mining methods. The time series analysis of garbage data proves that the medical wastes so produced are on the rise and the sharing of the burden of medical missions is unbalanced with regard to the hospital location and levels. By establishing an auto regressive integrated moving average（ARIMA）(0,1,1) model, the prediction and analysis for the every-day production of the hospital wastes in the area are made.The research results of the K-Means clustering analysis and the PARETO contribution analysis provide some empirical evidences for the future planning and development of the hospitals in this area.

Keywords—ARIMA Model, Correlation Analysis, Hospital Garbage Data, K-Means, PARETO Chart.

I. Introduction

MEDICAL garbage may help epidemic spreading and bring serious pollution to the environment and human health, so medical garbage must be treated as special garbage [1]. Attached great importance, at the call of the Chinese government to establish a medical and health care system in rural and urban areas by 2020 for people to achieve universal access to basic medical and health service, China's health care, medical treatment, medical services and medical management and so on have made remarkable achievements and progress, and gradually formed a relatively perfect system of health laws and regulations.

China's medical resources highly are concentrated in cities and the supply of rural medical services is relatively inadequate. There are 16,098 professional doctors and nurses in the region, including 954 in medical services, such as blood stations, family planning services, maternal and child health care and so on, and only 1248 in township hospitals. The statistics shows that there are11,450 hospital beds in the cities, but only 1,226 beds in township hospitals. Thus, the urban and rural medical resources are unbalanced, which leads to the problem of the people's "difficulty to see a doctor, and high cost of getting medical treatment" in urban and rural areas.

A. Related Works

Through statistics on the garbage data, it shows the distribution of three kinds of garbage(infectious, damaging and pathological) [2], the daily output in the region, and the correlation analysis of three kinds of garbage. Using ARIMA(0,1,1) model to predict the garbage output. By analyzing the garbage data, we know the source of the garbage, and infer the medical pressure of the relevant medical units. The K-means clustering analysis divides the hospital into 4 categories [3], and thus the garbage contribution of different types of hospitals are determined. Combining hospital staff and the number of hospital beds [4], it is concluded that the disequilibrium of medical resources leads to the problem of "difficult to see a doctor, and high cost of getting medical treatment".

Medical data is very large. This research dedicates to the analysis of the garbage produced from the hospital every day, combining with the numbers of doctors and beds. From the perspective of data the research shows to the public some undetermined point of view. Even more research focus on how to predict the future medical garbage [4-5]. That's the basic motivation of the study.

As a result, some status quo of medical care service is found from the data, thus providing a different perspectives for medical people in their management [6]. The contributions of this study fall into three folds:the study of the number of three types of garbage, the proportion of each type of garbage and the three types of correlation in the area, the study of the unbalanced distribution of the urban and rural medical institutions in terms of manpower and resources and the unbalanced sharing of the burden of medical missions, and a prediction of the yield of medical garbage in the area related to the hospital level, location and health insurance policy.

II. Basic Statistics

In this paper, the information is collected such as garbage classification, garbage weight, medical team, medical sickbed number, hospital location etc.

A. Categories and Weight of Medical Garbage

The medical garbage is divided into three categories: infectious garbage, pathological garbage and damaging garbage. Infectious garbage refers to medical garbage which carries pathogenic organisms and causes infectious diseases and spreads dangerous diseases. Pathological garbage refers to the body garbage and the corpse of medical experimental animals produced in the course of the treatment. Damaging garbage is an abandoned medical device that can stab or cut the body.Data collection began in October 2011 and ended in August 2016. A total of 612,729 records were recorded. In Fig.1, there are three kinds of garbage maps in five years. Infectious garbage accounts for 79.96% of total garbage, damaging garbage accounts for 19.51%, and pathology garbage accounts only for 0.53%.

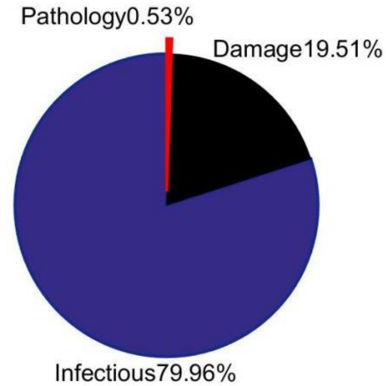

Fig. 1 Three Classes of Garbage Distribution

From Table 1, the statistics show that there is a steady rise in the risk of infectious and damaging garbage, and the pathological garbage fluctuates greatly. It reflected the garbage disposal was in time, no garbage heap.

Table 1 Daily average garbage amount of hospitals in jurisdictions(Kg)

Annual	Infectious	Damage	Pathology	Total
2012	1842.0233	497.45689	22.831742	2362.312
2013	2188.4149	523.52552	13.660501	2725.601
2014	2639.1363	521.59078	15.677556	3176.405
2015	2929.8448	776.19133	17.427444	3723.464
2016	3346.2819	865.07423	18.368539	4229.725

There should be a certain correlation between the three kinds of garbage that are produced during medical treatment . So we use the garbage data to test the correlation between them [8]. Unsurprisingly, there is a strong correlation between infectious garbage and damaging garbage, and there is also a certain correlation between garbage pathological and infectious garbage,also between pathological and damaging garbage. Table 2 shows the correlation analysis results.

Table 2 Correlation analysis results

Attribute	Infectious	Damage	Pathology
Infectious	1	0.92216	0.47711
Damage	0.92216	1	0.50087
Pathology	0.47711	0.50087	1

B. Information for Hospitals

There are two 3A hospitals, twelve 2A hospitals, six Grade 2 hospitals, twenty-five Grade 1or private hospitals, and thirteen medical services. Table 3 shows the typical hospital information, among them the Place Identity document(ID) is the administrative demarcation area, using a digit in the bar code, the region is set up in nine districts.

Table 3 Basic information of the hospital

Hos pital ID	Hospital level	Hospital Type	Staff Nu mbe	Sick bed Num	Hos pital Nu	Place

INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING
DOI: 10.46300/9106.2020.14.103 Volume 14, 2020

			r	ber	mbe r	
1	3A Hospitals	Comprehensive Hospital	3778	3311	2	Downtown
2	2A Hospitals	General Hospital	8512	6848	12	Downtown and County
3	Grade 2 Hospitals	general hospital	1584	1126	6	County
4	Grade 1 Hospitals	Township Hospital	1248	1226	24	Township
5	Private Hospitals	Private Hospital	22	70	1	Downtown
6	medical service institutions	service institutions	954	95	13	Downtown and County

C. Statistic Analysis of Urban and Rural Medical Services

To prove the wastes production in consistent proportion with the number of hospital staff members and beds in cities and countries, the hospitals are grouped into two categories: urban area and country, and a comparative study is made in terms with the above three types.

Figure 2 shows that the wastes production covers a clearly high proportion viewed with its staff members and beds in the first three hospitals while most of the rest are in consistent proportion, and only 4 of the 23 experimented hospitals hold a low proportion, and they share a smaller burden of medical missions and more resources.

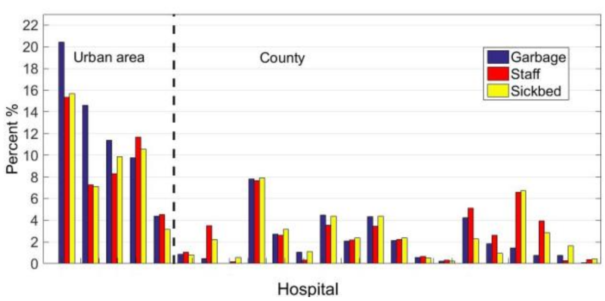

Fig. 2 Relationship of the urban and country garbage production,staff members, beds

Figure 3 shows that only 5 of the 22 experimented hospitals produce a larger quantity of wastes in proportion to their staff members and beds, and the rest are found an opposite ratio, which suggest that due to the smaller sharing of the burden of medical missions, the medical resources in the country hospitals are idled, even wasted, and then the medical level is hard to improve [11].

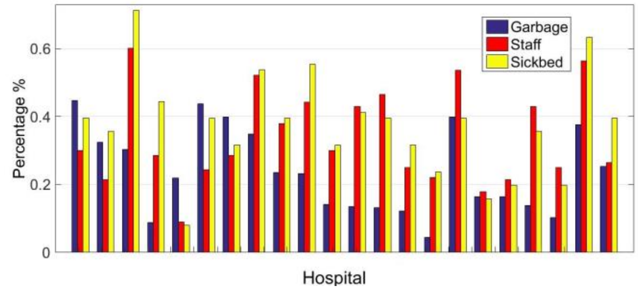

Fig. 3 Relationship of the rural hospital garbage production, staff

D. Medical Garbage Management in the Area

There is growing awareness of the need to impose stricter controls on the handling and disposal of garbage produced by hospitals. The city medical garbage in 80 hospitals in the region has been designated to companies for processing, using special packing box, and bar code pasted on the package. Before being disposed, the garbage must be registered. Using barcode to collect spam information, the computer synchronizes the garbage weight and avoids the error of manual operation.

III. DATA PROCESSING

The collected data shows that the garbage production varies with the time, but the stability of the change and the non-white noise sequence may be tested by ARIMA(p, d, q) model.

It can be seen that the change of the three kinds of garbage leveled off.

A. ARIMA Model

ARIMA(p, d, q) model, called the Auto-Regressive Moving Average model. It's a famous time series prediction by Box and Jenkins in the early 1970s.

ARIMA(p, d, q) model of the basic idea is: it predicts objects to form a sequence of data over time as a random sequence, with a certain approximate mathematical model to describe the sequence. Mathematical expression:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q} \quad (1)$$

Where Y_t is a realization of the time-series, β and α are parameters of the model and ε is an Independent and Identically Distributed (IID) error term with mean of zero and constant variance. The equation can be considered as having two main parts. The first part:

$$Y_t' = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \quad (2)$$

It represents an auto regressive (AR) element of order p. It represents the time-series being influenced by a linear function of previous values. In other words to make a forecast one needs to know the p previous values. The second part:

$$Y_t{''} = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} + \cdots + \alpha_q\varepsilon_{t-q} \quad (3)$$

It is a moving average (MA) process of order q. The number of difference times is d.

Value of a random variable Y_t is not only associated with the sequence value of the first p, q before and random disturbance, is called time series to obey (p, q) order Auto-Regressive Moving Average hybrid model.The main steps include data preprocessing, model identification, parameter estimation, and prediction analysis.Once the model can be identified we can predict the future value from the past value and present value.

1. Data Preprocessing

(1) Check the stability of the sequence

Figure 4 is a sequence diagram of the original sequence, which shows that the sequence has a distinct monotone increasing tendency, which can be judged as a non-stationary sequence.

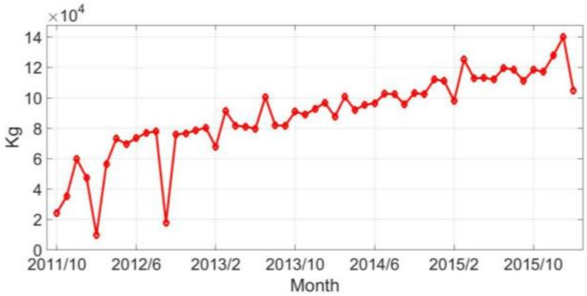

Fig. 4 Original Garbage Data Difference Sequence Diagram

Figure 5 shows a self-correlation diagram of the original sequence, showing that the correlation coefficients are always greater than 0, indicating a strong long-term correlation between the sequences.The stability of the original data is measured,the results shows:

p: 0.6538, stat: 0.018753, cValue: -1.9466

The p value is significantly greater than 0.05, so the sequence is judged to be a non-stationary sequence. The white noise test for raw data shows:

p: 0, stat: 111.1658, cValue: 12.5916

The p value is 0, so the original time sequence is a non-white noise sequence [9].

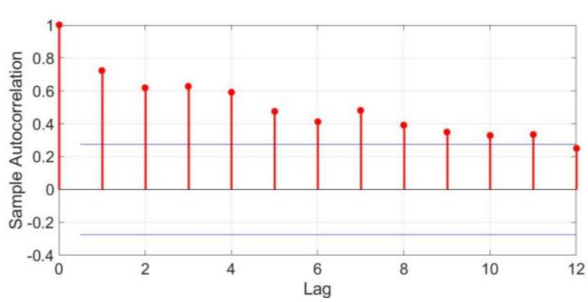

Fig. 5 Original Garbage Data Sample Autocorrelation Function

(2) Check the stability of the first order difference sequence

Difference operation has strong uncertainty information extraction ability, after many non-stationary sequence difference shows the nature of the stationary series, and then makes the non-stationary sequence difference stationary series [10].The ARIMA model can be used to fit the difference stationary sequence [11].

A first order difference of the original sequence is performed, and the sequence of the difference is determined again.The result of the stability test shows:

p: 0.001, stat: -10.6403, cValue: -1.9467

The sequence of difference is stable after the first difference.

The result of a first order difference in white noise shows:

p: 0.014777, stat: 15.8159, cValue: 12.5916

The sequence of difference is the sequence of smooth non-white noise after the first difference.

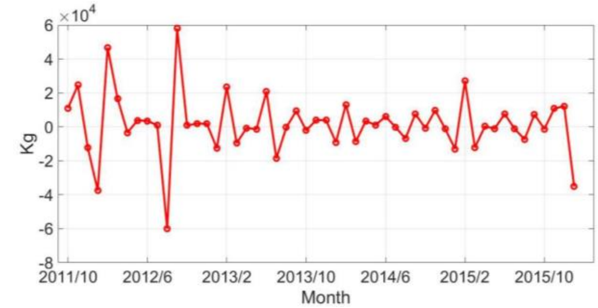

Fig. 6 Total Garbage Difference Sequence Diagram

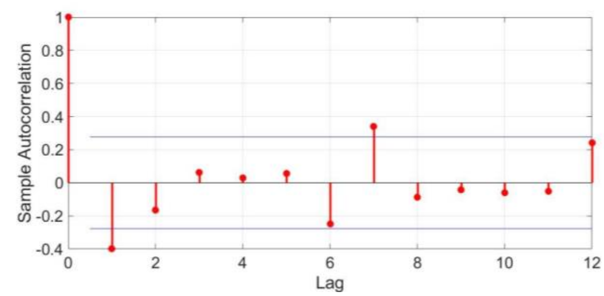

Fig. 7 Self-correlation of Sequence After First Order Difference

It can be seen from the Fig. 6 and Fig. 7 that after the first order difference sequence of the sequence diagram near the average more smoothly, auto-correlation figure has a strong short-term correlation, unit root test p value is less than 0.05, so after the first order difference sequence is stationary series.

2. Matching ARIMA Model

The matching ARIMA model is to determine the value of p and q.In this paper, to calculate p and q of the ARIMA (p, q) by using the relative optimal model identification method, p and q is less than or equal to 5 all combinations of Bayesian Information Criterion(BIC) information, access to the information model of minimum order.

Compute the BIC matrix as follows:

1.0e+03 *

1.1438	1.1232	1.1314	1.1337	1.1303	1.1335
1.1380	1.1412	1.1273	Inf	1.1664	1.1326
1.1321	1.1238	1.1286	1.1695	1.1254	1.1365
1.1307	1.1277	1.1400	1.1373	1.1288	1.1401
1.1343	1.1316	1.1439	1.1469	Inf	1.1316

DOI: 10.46300/9106.2020.14.103 Volume 14, 2020

1.1379 1.1373 1.1457 1.1482 1.1341 Inf
p: 0, q: 1, minimum BIC value: 1123.2, p, q is done.

3. Test Prediction

MA (1) was used to simulate fitting sequence, after the order difference of the original data to establish ARIMA (0, 1, 1) model, the actual amount of garbage for five years and predicted values are compared, check the error size, actual value and predicted value and error value as shown in table 4. It is shown that the overall error is within 6 percent, and the individual data errors are large, which is consistent with the MA (1) simulation of the first order difference, which is larger than the random perturbation. Because medium-sized hospitals and above timely deal with the garbage every day, and some small town hospitals process garbage within 2 days, even for a period of time no garbage, so random disturbance is in objective existence.

Table 4 Predictive data error rate

Actual data: Kg	135214.8	122910	132279.1	129860.8
Predicted data:Kg	127810	129620	135040	136850
Error value	5.48%	5.46%	2.1%	5.38%

B. Clustering Analysis for Hospitals

Hospitals in the area are tested by cluster analysis through the actual data, for the clustering characteristics, hospital ID, hospital level, the number of hospital staff, hospital beds, medical garbage.

According to the data analysis, among the hospitals most trusted by the masses in the region. Among them, the first category is the third class a hospital, there are two hospitals in total, which produce the largest amount of garbage and are also the hospitals with the greatest medical pressure. Secondly, there are three class 2A hospitals in the region, whose reimbursement ratio is high, resulting in a large amount of garbage and medical pressure; the third category is the second class a county people's Hospital, in order to guide ordinary patients to see a doctor and relieve the pressure of the first two types of hospitals, increasing the proportion of medical insurance reimbursement, the amount of garbage and medical pressure are basically balanced; the last category is Township and private hospitals, medical waste The yield is very small.The classification research results show the differences in garbage production, share of burden of medical missions and hospital scale.

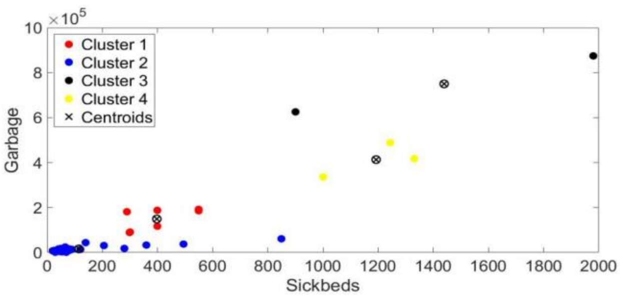

Fig. 8 K-means Clustering Analysis

C. Contribution Analysis for Hospitals

The contribution analysis is called PARETO analysis, the PARETO rule conform to the 20/80 law. As seen in Fig. 9, the total quantity of medical garbage in the top 9 hospitals accounts for 81.40 percent of the 45 hospitals in the region, and the 9 hospitals are urban and county hospitals. In the 9 hospitals, the number of doctors and staff numbers is 65.27% of the total, and the number of inpatient sickbeds is 65.35% of the total.

The region has a population of 7.18 million, with a permanent population of 5.8 million. **Medical institutions are the main body responsible for classified management of domestic waste in medical institutions. The proportion of garbage volume, the number of workers and the number of beds relate to the size of medical pressure. The proportion of garbage volume is higher than that of the number of workers and the number of beds. The medical pressure faced by enterprises is relatively large and resources are relatively scarce. On the contrary, medical pressure is small and resources are abundant.**As a result, the medical pressure in the 10 hospitals is very high and the doctors' burden is heavy. China's medical resources are highly concentrated in cities and the supply for rural medical services is relatively inadequate. Only by optimizing the allocation of human resources and material resources, and realizing the balanced allocation of medical resources in urban and rural areas, can we effectively solve the problems with "difficulties in going to a doctor, and the high expenses for getting medical treatment".

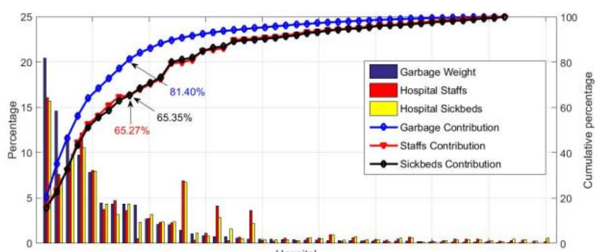

Fig. 9 Pareto Diagram and Contribution Analysis. (A) The bar chart represents the waste weight, the number of staffs and the number of beds in a hospital. (B) Left Y axis:The percentage of the total number about garbage, staff and sickbeds. (C) Right Y axis:The cumulative percentage of the total number about garbage, staff and sickbeds. (D) X axis: 45 Hospital in the Prefecture.

IV. CONCLUSIONS

Based on the data processing of medical waste collected by 68 medical institutions within 5 years, combined with the number of hospital staff and beds, through clustering and contribution analysis, this paper predicts the medical pressure and garbage volume in urban and rural areas, which plays an important role in China's medical waste. However, due to the limitation of research area and medical garbage data acquisition method, we can only analyze the imbalance of medical pressure according to the existing resources, and put forward corresponding suggestions for managers.

Based on the acquisition of the data of medical garbage in the past five years,combined with the number of doctors and staff members and hospital beds, after an analysis on the hospital contributions to the garbage,the author makes an estimation on the garbage production, which may be of practical significance. According to the analysis of"Statistic Analysis of Urban and Rural Medical Services" and"Medical Garbage Management in the Area", the trust from the people is different towards different levels of hospitals, leading to their different consumption preferences, which so results in the local central hospitals' heavy medical burden. Of course, there are limitations with the study, the hospital's proportion is changing, and the number of hospital staff and beds will change also[12]. Further researches should be made with more extensive medical data analysis.

In China, with the deepening of health care reform, it will get good grades through health insurance reform, the development of the dsignated hospitals, which can alleviate the problems with "difficulties in going to a doctor and the high cost of getting medical treatment".

ACKNOWLEDGMENTS

Benyou Wang received support from Natural Science Foundation of China under Grant (61375121), the Teaching Research Projects of Anhui China under Grant (2015zy051).

References

[1] Shujun Z, Yingying L, Dong L I, et al. Statistics and Analysis on Production of Hospital Preparation in China-Japan Friendship Hospital During 2014-2016[J]. Evaluation and Analysis of Drug-Use in Hospitals of China, 2017.
[2] Asma M, Syeda B, Muhammad C. Environmental Quantification of the Existing Waste Management System in a Developing World Municipality Using EaseTech: The Case of Bahawalpur, Pakistan[J]. Sustainability, 2018, 10(7):2424.
[3] Niyongabo E, Jang Y C, Kang D, et al. Generation, management practices and rapid risk assessment of solid medical wastes: a case study in Burundi[J]. Journal of Material Cycles and Waste Management, 2019, 21(4):950-961.
[4] Lee M Y, Soriano R P, Fallar R, et al. Assessment of Medication Management Competency Among Medical Students Using Standardized Patients[J]. Journal of the American Geriatrics Society, 2020, 68(3).
[5] Mhady A I A, Awad M A, Al-Aghah M R, et al. Assessment of Medical Waste Management in El Shifa and Al Aqsa Hospitals-Gaza Strip[J]. Health, 2019, 11(8):1028-1042.
[6] Alam I, Alam G, Ayub S, et al. Assessment of Bio-medical Waste Management in Different Hospitals in Aligarh City[M]// Advances in Waste Management. 2019.
[7] B M H D A, A H D A, C R N A, et al. Medical waste generation and management in medical clinics in South of Iran[J]. MethodsX, 2019, 6:727-733.
[8] Yuanyuan Z, Chunhe Z. Problems and countermeasures of medical waste management in the context of "Internet + Nursing Service"[J]. Chinese Nursing Management, 2019.
[9] Aung T S, Luan S, Xu Q. Application of multi-criteria-decision approach for the analysis of medical waste management systems in Myanmar[J]. Journal of Cleaner Production, 2019, 222(JUN.10):733-745.
[10] Bin L. Economic Analysis of Shanghai Medical Waste Management System[J]. Environmental ence and Management, 2018.
[11] Li-Wei Z, Cheng-Yi F, Ti L, et al. Effect of quality control circle on medical waste reduction and standardized management[J]. Chinese Journal of Nosocomiology, 2019.
[12] Stachowitsch M. Medical Wastes[J]. 2019.

Benyouo Wang, born in 1970, has received the master's degree. He is an associate professor in West Anhui University. He is interested in data mining and intelligent control.

Sihai Zhang, born in 1972, received the M.S. and Ph.D. degrees from the Department of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China, in 2002 and 2006, respectively. He is currently an Associate Professor of Electronic Engineering with the Department of Electronic Engineering and Information Science, USTC. His research interests include wireless networks, opportunistic networks, and intelligent algorithms.

Li Gu, born in 1974, has received the master's degree. He is a lecturer in University of Science and Technology of China. He is interested in data mining.