IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 2, 2019, accepted September 15, 2019, date of publication September 24, 2019,
date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2943547

A Privacy Security Risk Analysis Method for Medical Big Data in Urban Computing

RONG JIANG[1,2,3], MINGYUE SHI[1,2,3], AND WEI ZHOU[2,3,4]

[1]School of Information, Yunnan University of Finance and Economics, Kunming 650221, China
[2]Key Laboratory of Service Computing and Security Management of Yunnan Provincial Universities, Kunming 650221, China
[3]Kunming Key Laboratory of Information Economy & Information Management, Kunming 650221, China
[4]School of Finance, Yunnan University of Finance and Economics, Kunming 650221, China

Corresponding author: Rong Jiang (jiang_rong@aliyun.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71972165, Grant 61763048, Grant 61263022, and Grant 61303234, in part by the National Social Science Foundation of China under Grant 12XTQ012, in part by the Innovation and Promotion of Education Foundation Project of Science and Technology Development Center, Ministry of Education, under Grant 2018A01042, in part by the Science and Technology Foundation of Yunnan Province under Grant 2017FB095 and Grant 201901S070110, in part by the 18th Yunnan Young and Middle-Aged Academic and Technical Leaders Reserve Personnel Training Program under Grant 2015HB038.

ABSTRACT As an emerging computing mode, urban computing is mainly used to integrate, analyze and reuse urban resources by using perceptual computing, data mining and intelligent extraction to eliminate the phenomenon of data islands and provide wisdom for people to make decisions. But in the era of big data, the security and privacy leakage of users has become a major obstacle in urban computing. Taking medical big data as an example, this paper analyzed the risk of security and privacy leakage in the collection, transmission, storage, use and sharing of medical big data, and established a medical big data security and privacy leakage risk indicator system with 4 primary indicators and 35 secondary indicators. In addition, the weight of each indicator was calculated by GI method and entropy weight method. Then the fuzzy comprehensive evaluation model was established to verify the risk of medical big data security and privacy disclosure in urban computing. The results show that the risk of medical big data security and privacy leakage in the Grade II Level A hospitals is higher than that in the Grade III Level A hospitals, and in the life cycle of medical big data, the two stages of data storage, data use and sharing may cause more prominent problems of data security and privacy disclosure, while the data collection and data transmission are slightly less. Finally, the comparison of performance further proved the scientificity and effectiveness of this method.

INDEX TERMS Urban computing, medical big data, security and privacy, risk analysis, indicator system.

I. INTRODUCTION

Driven by the wave of information technology, urban computing with urban backgrounds has emerged as an emerging field [1], with categories including transportation, environment, economics, social, medical services, and urban planning [2]. It mainly provides data acquisition and analysis of various types of data in urban through intelligent extraction, data mining, and sensing technology to provide predictions and references for urban traffic conditions, disease spread, and house price trends, etc. [3]. From the perspective of the Internet, the core problem of urban computing is the use of intelligent sensors to collect and transmit data in the urban, while managing and analyzing the data [4]. From the perspective of the Internet, the core problem of urban computing is the use of intelligent sensors to collect and transmit data in the urban, while managing and analyzing the data [5]. However, in the category of urban computing research, the field of medical services is quite special, because the data it involves is basically human-based [6]. As shown in Figure 1, with the development of science and technology, the medical information data flow in urban computing contains four aspects: medical resources (electronic medical records, clinical testing, doctor-patient behavior, etc.), subject-related data resources (such as life sciences, demography, etc.), industry related data resources (such as medical insurance government), Internet data resources (such as social media) [7], involves data on physiology, psychology, disease

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

prevention and medical management generated during the whole life cycle of people's birth, aging, illness and death, basic necessities, transportation, industry, agriculture and business, etc. [8].

These data have important implications for medical research, commercial development, and more. According to the report, if hospitals can use the technology of intelligent extraction and data mining in urban computing to fully realizethe value of these data, it will bring hundreds of billions of profits to the medical industry every year. For example, by means of big data analysis tools in urban computing to help doctors to develop more scientific and effective diagnostic programs. In addition, insurance companies can use these data to update the forecast model in time [9]. It is worth noting that the development of medical digitalization in urban computing can indeed bring some value to people's medical treatment and medical research, but we must also consider the security problems brought by technology while enjoying the convenience. At present, medical data leakage incidents are not uncommon, and there are many problems in the sharing and use of medical data etc. in urban computing. At the same time, the informationization of the medical industry is in a critical period of development, and various new forms of Internet and medical are booming. The widespread application of new technologies such as cloud computing, mobile Internet and Internet of things has brought new challenges and threats to medical big data. Therefore, it is urgent to solve the information security issues throughout the life cycle of medical big data production, collection, storage, sharing, exchange, and use in urban computing [8].

Medical data including three kinds of physical states: pictures and files, video and data flow, language and text, but no matter which scene application is hidden, it is closely related to human's privacy information, must ensure the user's personal privacy and information security [8]. If these private data are leaked, it will cause great harm to the patient's reputation and life, and even bring serious moral and ethical problems to the hospital [9]. In addition, the information leakage in the medical big data environment is not only the data itself, but more serious is that hackers steal patients' social security accounts and personal finance by mining the hidden information behind the data, endangering the patient's personal and property security [6]. Therefore, it is imperative to establish and improve the Internet and medical service security working mechanism in urban computing, and improve data security risk prevention measures and privacy protection [10].

In this study, we firstly analyzes the security and privacy leakage risk indicators of medical big data collection, transmission, storage, use and sharing through brainstorming, Delphi, questionnaire, interview and field research. The combination of methods can reduce the influence of subjective factors to some extent. Then, the comprehensive weight of secondary indicators is calculated through the combination of GI method and entropy weight method. The combination of these two methods not only avoids the influence of subjective factors, but also avoids the loss of data hidden in objective information, making the evaluation results more accurate. Next, the fuzzy comprehensive evaluation model is used to evaluate the risk level of medical big data security and privacy disclosure. Finally, the risk reduction strategies corresponding to the four stages of data collection, transmission, storage, use and sharing in the life cycle of medical big data are presented.

The rest of this paper is organized as follows. Section 2 describes the progress and deficiencies in research on information security and privacy protection. Section 3 introduces the establishment process of medical big data security and privacy leakage risk evaluation index system and the calculation method of each index weight. Section 4 introduces the risk quantification method of medical big data security and privacy disclosure. Section 5 conducts experimental analysis with examples, and gives corresponding risk reduction strategies. Section 6 summarizes this article.

II. RELATED WORK

At present, few scholars specifically study the medical information security and privacy leakage in the urban computing environment, but some scholars have analyzed the information security issues in the construction of smart cities. For example, the literature [11]–[14] mainly studied the problem from the aspects of management mechanism, technology and infrastructure, and clarified the new information security management mode and measures. Wang [15] analyzed the personal information security issues in the context of smart cities from the legal level, and focused on the contents of legal documents. Ferraz and Ferraz [16] focused on the issue of citizen privacy violations in a smart city environment. Literature [17] analyzed the relationship between information security risk factors in smart cities, and established a set of information security risk assessment indicators and risk assessment methods. The above studies are only a macroscopic discussion of information security issues in smart cities and do not address specific data security and privacy issues. If the background of urban computing is abandoned, domestic and foreign scholars have already carried out relevant research on data security and privacy protection issues, including the following key technologies: access control technology, secure retrieval, and secure computing. Among them, access control technology is a research hotspot, and this method can be summarized into two categories: one is the technical route based on cryptography, and the other is the access control based on role and risk. The latter is more flexible and suitable for the complex environment of medical big data [18]. A report published by foreign scholar JASON in 2004 [19] is the first to introduce the concept of risk into the field of access control. The report gave some guiding principles and recommendations that should be met based on risk access control and defined the risk quantitative concept. Literature [20], [21] introduced the concept of risk into fuzzy theory, increasing the flexibility of the authorization strategy.

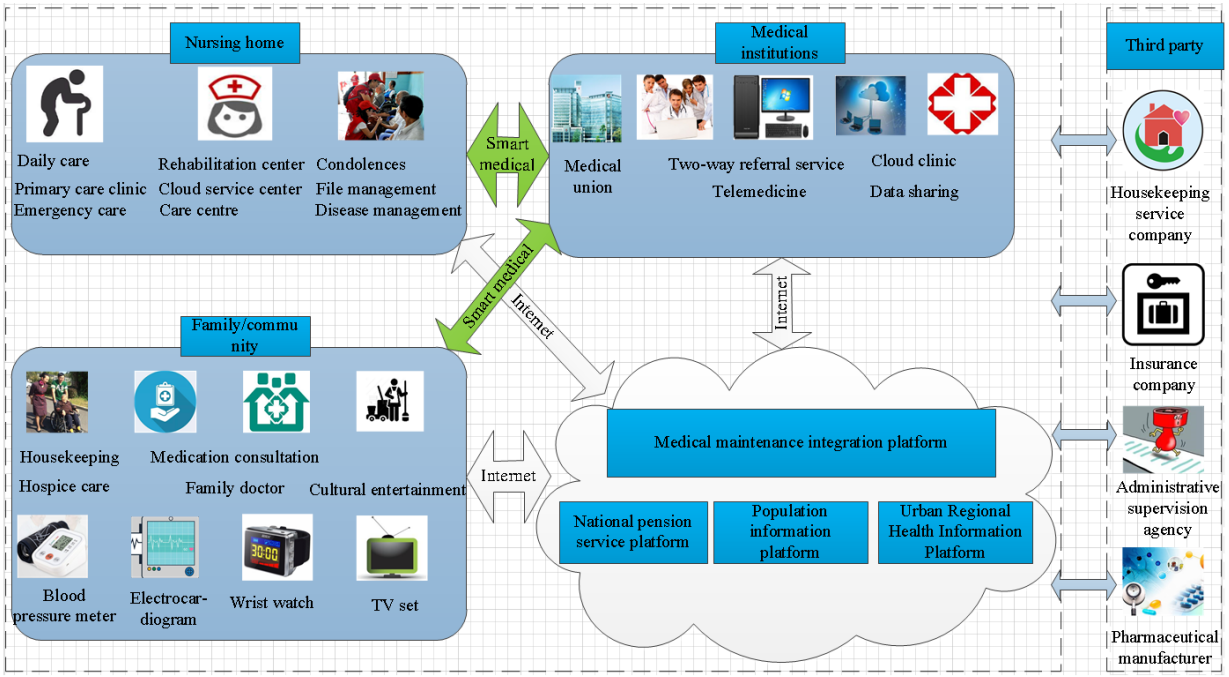

FIGURE 1. Pattern of medical information data flow and relationship between various fields in urban computing.

There are also some scholars who have proposed a risk-based access control model for medical systems to solve the problem of patient privacy leakage caused by excessive authorization or illegal access in medical systems [22], [23]. Literature [24]–[30] analyzed the information security issues in the cloud environment from different levels, and established a cloud computing security evaluation index system. Literature [31]–[34] studied the privacy leakage of users, and established a user privacy risk assessment index system from different levels. G. Zhu *et al.* [35] constructed a privacy risk analysis indicator system from 3 aspects: social network platform, user behavior and external threats. Finally, the fuzzy risk analysis of social network privacy risk is carried out by AHP and entropy method. Li [36] researched the typical risks and information protection problems faced by personal information protection in the era of big data from the aspects of theory and justice. Meanwhile, corresponding countermeasures and suggestions were given.

In summary, it is not difficult to find that the current research on information security and privacy protection most focuses on the individuals and network systems in the cloud environment. In contrast, there are few studies on data security and privacy protection issues specific to the medical industry. Although some scholars have established risk indicators for data security and privacy leakage in some aspects of the medical big data life cycle, they do not give a specific risk assessment method for the medical industry. At the same time, the establishment of the risk indicator system and the calculation process of the weights of each index are not rigorous, and the results are greatly affected by subjective factors. Therefore, this paper is a necessary supplement to the research on security and privacy protection in the medical big data environment.

III. DESIGN OF RISK EVALUATION INDICATOR SYSTEM

A. ESTABLISH RISK ANALYSIS INDICATOR SYSTEM

In the production, collection, transmission, storage, use and destruction of urban medical big data, there are different security and privacy leakage threats in each link. After consulting the medical industry and information security related professionals and analyzing a large number of domestic and foreign references, it is found that the medical big data industry security issues are most prominent in the four stages of data collection, data transmission, data storage, data usage and sharing, and the relationship between them is shown in Figure 2.

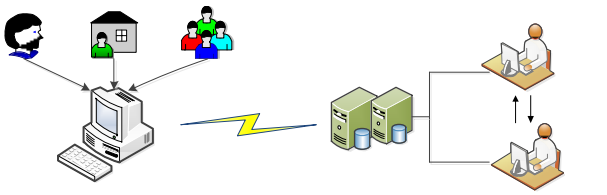

FIGURE 2. The relationship between data collection, transmission, storage, use and sharing in the medical big data life cycle.

If the indicators of security and privacy leakage risk in the medical big data collection, transmission, storage, use and

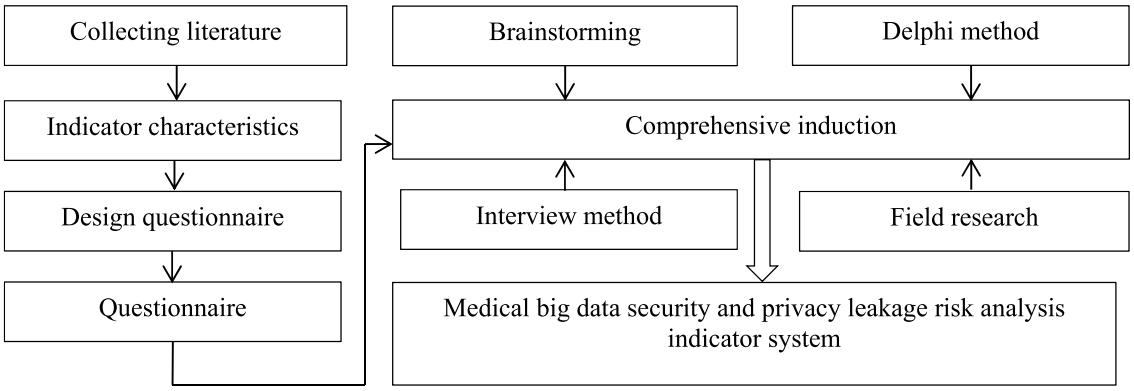

FIGURE 3. Steps for establishing risk indicators of medical big data security and privacy leakage.

sharing can be correctly combed, controlled and managed, the probability of security and privacy leakage problems can be reduced to a large extent. Due to the rapid development of Internet and medical, the risk of security and privacy leakage of medical big data is partly intersected with the risk of network security and privacy leakage. In addition, the research achievements of the academic circle in network security have been quite rich, and the relevant indicator system is relatively mature. Therefore, the initial indicators of this paper will be extracted from the literature [27], [30], [32], [35], but in order to guarantee risk factors is more perfect, objective and accurate, this article will further by consulting relevant experts and relevant personnel of medical institutions (such as doctors, nurses, management personnel, technical personnel, etc.). In addition, by combining field research, Brain Storming, Delphi, Interview, questionnaire and other research methods, we modified and improved the preliminarily established risk index system, deleted some indexes of low importance, and dealt with repeated indexes in an interactive way. Finally, the risk factors of security and privacy disclosure of medical big data are identified and condensed from the aspects of data collection, transmission, storage, use and sharing. Combined with the indicator system construction process shown in Figure 3, a risk assessment indicator system including 4 first-level indicators and 35 second-level indicators is established. The specific indicators and contents are as follows. Medical big data security and privacy leakage caused by the data collection phase (A_1), which includes 9 secondary indicators: Lack of data collection specification (B_1), Patients lack the right to know when collecting data through smart devices (B_2), Data ownership is not clear (B_3), Medical staff operation error (B_4), Lack of professional ethics staff (B_5), Third party malicious behavior (B_6), Wearable device positioning function (B_7), Lack of supporting policies and supervision mechanism (B_8), Lack of special laws and regulations (B_9). Medical big data security and privacy leakage caused by the data transmission phase (A_2), which includes 8 secondary indicators: Third party malicious behavior (B_6), Lack of unified data transfer protocol standard (B_{10}), Encryption and key management weak (B_{11}), Service engine vulnerability (B_{12}), Hardware security (B_{13}), Software security (B_{14}), Virus intrusion (B_{15}), Hacker attacks (B_{16}). Medical big data security and privacy breaches caused by the data storage phase (A_3), which includes 17 secondary indicators: Medical staff operation error (B_4), Encryption and key management weak (B_{11}), Hardware security (B_{13}), Software security (B_{14}), Virus intrusion(B_{15}), Hacker attacks (B_{16}), Internal personnel stealing information (B_{17}), Physical environment (B_{18}), Virtual vulnerability (B_{19}), Firewall vulnerability (B_{20}), Access control mechanism is not perfect (B_{21}), Identity authentication technology is not complete (B_{22}), Safety audit (B_{23}), Data monitoring (B_{24}), Digital certificate reliability (B_{25}), IDS reliability (B_{26}), Data backup and recovery (B_{27}). Medical big data security and privacy breaches caused by the data usage and sharing phase (A_4), which includes 19 secondary indicators: Data ownership is not clear (B_3), Lack of supporting policies and supervision mechanism (B_8), Lack of special laws and regulations (B_9), Encryption and key management weak (B_{11}), Internal personnel stealing information (B_{17}), Firewall vulnerability (B_{20}), Access control mechanism is not perfect (B_{21}), Safety audit (B_{23}), Data monitoring (B_{24}), Digital certificate reliability (B_{25}), Data backup and recovery (B_{27}), Data acquirer's dishonest behavior (B_{28}), Electronic certification service is not perfect (B_{29}), Electronic medical record sharing standards are not perfect (B_{30}), Hospital information platform interaction standard is not standardized (B_{31}), Telemedicine equipment and unified communication interaction standards are not standardized (B_{32}), Data usage management system lacks (B_{33}), Data sharing standard is not perfect (B_{34}), Application management organization system is not sound enough (B_{35}).

B. CALCULATE THE WEIGHT OF RISK ANALYSIS INDICATOR

Relevant risk indicators have been sorted out in section A, but our ultimate goal is to get an accurate risk value. Therefore, the factors need to be quantified. Through the analysis of the literature, the calculation methods of indicator weights are mainly divided into three categories: one is the subjective weighting method represented by Delphi method, expert

TABLE 1. Comparative analysis of the advantages and disadvantages of typical weighting methods.

Weight calculation method	Advantage	Disadvantage
Delphi method	The weight of opinions is the same, the implementation is convenient and extensive	The process is complex and takes a long time
Analytic hierarchy process	Systematic analysis method requires less quantitative data information	It is not suitable for problems with many factors and large scale, and the consistency requirement is difficult to meet
Fuzzy analytic hierarchy process	Handling fuzzy evaluation objects through precise digital means, making more scientific, reasonable and close to practical quantitative evaluation	The calculation is complex and the determination of indicator weight vector is subjective
GI method	No consistency test, in the case of a large number of indicators, the amount of calculation is relatively small	It has high subjectivity and low accuracy
Extension goodness method	It can clearly reflect the relationship between things and quantity, express the change process of objective things, and combine qualitative and quantitative analysis together	Stronger subjectivity for the determination of indicator weight vector
Multi-objective programming	High precision, can solve multi - objective problem	The calculation is extremely complicated and the calculation amount is large
Entropy weight method	Compared with subjective assignment method, it has higher accuracy and stronger objectivity, and can be used in combination with other methods	Large amount of calculation, not suitable for many factors with large scale
Principal component analysis	It can eliminate the influence between evaluation indicators and reduce the workload of indicator selection	The interpretation of the principal component has ambiguity

investigation method, analytic hierarchy process, fuzzy analytic hierarchy process, GI method and extension goodness method. The other type is the objective weighting method represented by entropy weight method, multi-objective programming method and principal component analysis method. Subjective analysis is based on the decision maker's empowerment, and the results will be affected by subjective factors. However, the objective analysis method does not need the decision-maker to provide relevant information, and the calculation is carried out strictly according to the relevant model theory, but the result is relatively harsh, and some key indicators with small data differences may be ignored. Therefore, in order to make up for the shortcomings of subjective weighting method and objective weighting method, the third comprehensive screening weighting method appeared, which can reduce the influence of subjective factors without losing the data hidden in objective information.

Based on the previous research methods, this paper compares some typical weighting methods in Table 1 [37], and combines the characteristics of medical big data security and privacy risk indicators to propose a risk indicator weighting method combining GI method and entropy weight method.

1) SUBJECTIVE WEIGHT CALCULATION BASEDON GI METHOD

The GI method was first proposed by Guo Yajun as a new algorithm for a kind of decision problem [38]. It overcomes the difficulty in constructing a two-two judgment matrix by AHP and ANP methods and the difficulty in passing the consistency test due to too many indicators. The central idea is also to compare the importance of the indicators in each indicator layer, but not to compare all the indicators in pairs. The specific calculation steps are as follows:

a: SORTING THE IMPORTANCE OF INDICATORS

Let the peer indicator set to $\{x_1, x_2, \cdots x_n\}$ and determine the order relationship according to the following rules:

1) Invite the experts in the field to select one of the most important indicators from the indicator set $\{x_1, x_2, \cdots x_n\}$ as x_1'.
2) The expert continues to select the most important indicator items from the remaining $n-1$ indicator sets as x_2'.
3) After selecting m times, the expert continues to select the most important indicator item from the remaining $n-(m-1)$ as x_m'.
4) Finally, after the selection of $n-1$, the only remaining one is marked as x_n'.

b: DETERMINE THE RELATIVE IMPORTANCE OF EACH INDICATOR

Let the ratio of the degree of importance between x_{n-1}' and $\mathrm{x_n'}$ be denoted as r_n, then $r_n = \frac{w_{n-1}}{w_n}$, where w_n represents the weight of the n-th indicator, and experts are also requested to score according to the relative importance of each indicator in Table 2

c: DETERMINE THE WEIGHT OF EACH INDICATOR

Taking the primary indicator as an example, if an expert k gives a subjective evaluation of each indicator, the weight of the j-th indicator under the k-th expert decision is w_j^k expressed as:

$$w_j^k = \left(1 + \sum_{n=2}^{j} \prod_{i=n}^{j} r_i\right)^{-1} \quad and\ w_{n-1} = r_n w_n; \quad n = j, j-1, \cdots 3, 2 \qquad (1)$$

TABLE 2. Quantification table of relative importance scores between indicators.

r_n	Description
1.0	The indicator x'_{n-1} is as important as the indicator x'_n
1.1	The ratio of the indicator x'_{n-1} to the indicator x'_n is between the same important and slightly important
1.2	The indicator x'_{n-1} is slightly important than the indicator
1.3	The ratio of the indicator x'_{n-1} to the indicator x'_n is between slightly important and obviously important
1.4	The indicator x'_{n-1} is obviously important than the indicator x'_n
1.5	The ratio of the indicator x'_{n-1} to the indicator x'_n is between obviously important and strongly important
1.6	The indicator x'_{n-1} is strongly important than the indicator x'_n
1.7	The ratio of the indicator x'_{n-1} to the indicator x'_n is between strongly important and extremely important
1.8	The indicator x'_{n-1} is extremely important than the indicator x'_n

d: CLUSTER DECISION RESULT

Assuming that a total of t experts participate in the decision-making, the weight indicator of the k-th expert in decision-making is L_k, then the cluster decision result w_j^g of the j-th indicator is expressed as:

$$w_j^g = \sum_{k=1}^{t} L_k w_j^k \tag{2}$$

2) OBJECTIVE WEIGHT CALCULATION BASED ON ENTROPY WEIGHT METHOD

The concept of entropy first appeared in thermodynamics. In 1948, Shannon proposed the concept of information entropy and solved the problem of quantitative measurement of information. In information theory, information entropy represents the uncertainty measure of the system in the disordered state. The basic idea of entropy weight method is to determine the objective weight of indicators according to their variability. The larger the entropy is, the smaller the variability of the indicator is, the less information it will provide [39]. On the contrary, the more information it will provide, the greater the role it will play in the comprehensive evaluation and the greater its weight will be.

Since the weights of the indicators calculated by GI method will be affected by experts' personal subjective factors, while the entropy weight method is an objective weight determined according to the variability of the indicators, and the subjective weights calculated by the GI method can be corrected to make the evaluation result more scientific and reasonable. Therefore, this section focuses on the specific steps of calculating indicator weight by entropy weight method.

a: CONSTRUCT JUDGMENT MATRIX X ACCORDING TO EXPERT SCORE TABLE

After the experts evaluate the indicators, the judgment matrix $X_{m\times n}$ is obtained:

$$X = (x_{ij})_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{3}$$

b: DATA STANDARDIZATION AND NORMALIZATION

Standardize the judgment matrix, so:

$$x'_{ij} = \frac{x_{ij} - min(x_i)}{max(x_i) - min(x_i)} \tag{4}$$

After normalization:

$$x''_{ij} = \frac{x'_{ij}}{\sum_{j=1}^{n} x'_{ij}} \tag{5}$$

The normalized matrix:

$$(X''_{ij})_{m\times n} = \begin{bmatrix} x''_{11} & x''_{12} & \cdots & x''_{1n} \\ x''_{21} & x''_{22} & \cdots & x''_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x''_{m1} & x''_{m2} & \cdots & x''_{mn} \end{bmatrix}$$

c: CALCULATE THE ENTROPY OF EACH INDICATOR

Assuming that the information entropy of the i-th indicator is e_i, then:

$$e_i = -\frac{1}{lnn} \sum_{j=1}^{n} x''_{ij} ln x''_{ij} \tag{6}$$

Among them, $\frac{1}{lnn} > 0, x''_{ij} \in [0, 1], e_i \in [0, 1]$.

d: CALCULATE THE WEIGHT OF EACH INDICATOR

Given that the information entropy of the i-th indicator is e_i, the difference coefficient H_i of the ith indicator can be expressed as:

$$H_i = 1 - e_i \tag{7}$$

Then the weight of the indicator w_i^s is expressed as:

$$w_i^s = \frac{H_i}{\sum_{i=1}^{m} H_i} = \frac{1 - e_i}{m - \sum_{i=1}^{m} e_i} \tag{8}$$

3) COMPREHENSIVE WEIGHT CALCULATION

In some way, the subjective weight determined by the GI method is combined with the objective weight determined by the entropy method to obtain the weight of subjective and objective integration. Looking at the relevant literature, we can find that the Lagrange multiplier method or the weighted linear combination method is mainly used to calculate the comprehensive weight.

The Lagrange multiplier method is calculated as follows:

$$w_j = \frac{w_j^s w_j^g}{\sum_{j=1}^{n} w_j^s w_j^g} \tag{9}$$

where w_j represents the combined weight of the subjective weight w_j^g obtained by the GI method and the objective weight w_j^s obtained by the entropy weight method, and n represents the number of indicators.

The calculation method of the weighted linear combination is as follows:

$$w_j = \varphi w_j^g + (1 - \varphi) w_j^s \tag{10}$$

where φ represents the subjective coefficient, $\varphi \in [0, 1]$, used to adjust the degree of influence to subjective factors on the results.

By comparing the above two formulas, it can be found that in Lagrange multiplier method, the proportion of subjective and objective factors is 1 to 1, while the weighted linear combination method adjusts the proportion of subjective and objective factors according to the actual situation. The purpose of this paper is to minimize the impact of external factors to make the indicator weight more objective and reasonable. Therefore, in order to ensure the accuracy of the results, this paper chooses the weighted linear combination method to adjust the subjective and objective proportion according to the actual situation. Invited 5 university experts who are very familiar with the medical field and 3 medical information security experts from Kunming First People's Hospital to score the subjective coefficient φ, and finally reached a consensus opinion, $\varphi = 0.45$. Therefore, the final composite indicator weight w_j is expressed as:

$$w_j = 0.45 \times w_j^g + 0.55 \times w_j^s \tag{11}$$

IV. RISK QUANTIFICATION

In risk-based access control research, risk quantification is the focus of the whole research process [40]. Section 2 has determined risk indicators that affect medical big data security and privacy breaches, and calculated the weights of related indicators. Next, we need to quantify the risk according to the relevant indicators. Risk represents the possibility of the threat, which is an uncertain factor, but our ultimate goal is to comprehensively evaluate the risk level of medical big data security and privacy leakage. Therefore, according to the two characteristics of risk, one is not easy to quantify and it is an uncertain factor; the other is that it needs to be comprehensively evaluated according to relevant indicators [41]. Finally, we decided to use the fuzzy comprehensive evaluation method to quantify the risk of medical big data security and privacy leakage.

A. BASIC IDEAS AND PRINCIPLES OF FUZZY COMPREHENSIVE EVALUATION METHOD

The fuzzy comprehensive evaluation method is a method based on fuzzy mathematics to quantify some factors whose boundaries are not clear and difficult to quantify. The main idea is to comprehensively evaluate the membership level of the evaluation object from a number of factors.

The basic principle of the fuzzy comprehensive evaluation method can be roughly summarized into three steps [42]:

Step 1: Determine the indicator evaluation set of the evaluated object

Step 2: Determine the weights and membership vectors of each indicator to obtain a fuzzy evaluation matrix

Step 3: Fuzzy evaluation matrix and indicator weight vector for fuzzy operation to obtain fuzzy comprehensive evaluation results

B. MODEL AND STEPS OF FUZZY COMPREHENSIVE EVALUATION METHOD

When quantifying the security and privacy leakage risk of medical big data through fuzzy comprehensive evaluation method, it is mainly carried out from the following six aspects.

1) DETERMINING THE INDICATOR DOMAIN OF THE EVALUATION OBJECT

$$U = \{u_1, u_2, \cdots, u_m\}$$

in which m represents the number of evaluation indicators, and in this paper refers to the number of indicators that affect the security of big data and privacy issues.

2) DETERMINE THE WEIGHT VECTOR OF EACH INDICATOR

In Section 2.2, the weights of the indicators have been calculated by the GI method and the entropy weight method. Here, $a_i (i = 1, 2, \cdots, m)$ is used to indicate the weight of the i-th indicator. Finally, the weights of the indicators are recorded as A.

3) DETERMINING THE EVALUATION LEVEL DOMAIN

The evaluation level is a set of various evaluation results that the experts may make on the evaluation object, which is represented by V:

$$V = \{v_1, v_2, \cdots, v_n\}$$

where v_i represents the i-th evaluation result and n represents the number of division intervals for evaluation results. In this paper, it refers to the number of intervals for risk grades.

4) SINGLE FACTOR FUZZY EVALUATION

The single factor evaluation here refers to determining the membership degree of the evaluation object belonging to the evaluation set V from one indicator factor alone. After the fuzzy evaluation of each factor u_i, the fuzzy membership degree of the evaluation object to each evaluation level is obtained from a single factor, so the fuzzy relation matrix R is determined.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

in which $r_{ij}(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ represents the membership degree of the evaluated object to v_j grade from factor u_i. Different rows reflect the membership degree of evaluated objects to each evaluation set from different single factors. In addition, through the above analysis, it can be found that fuzzy relation matrix R is actually a fuzzy relation between indicator factor set U and evaluation set V.

5) MULTI-FACTOR FUZZY EVALUATION

The fuzzy comprehensive evaluation result vector B can be obtained by performing an integrated operation on each of the indicator weight vector A and the fuzzy relation matrix R.

$$B = A \circ R = [a_1, a_2, \cdots, a_m] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} = [b_1, b_2, \cdots, b_n] \quad (12)$$

6) ANALYZE THE FUZZY COMPREHENSIVE EVALUATION RESULTS

Since the result of the fuzzy comprehensive evaluation is a vector rather than a specific value, we need to further process the result. According to the relevant literature, we know that the current analysis of fuzzy comprehensive evaluation results mainly includes the following two methods: one is the principle of maximum membership degree, and the other is the principle of weighted average. When dealing with problems, it is necessary to select appropriate methods according to actual conditions.

V. EXPERIMENTAL ANALYSIS

A. DATA SOURCES

This paper relies on the National Natural Science Foundation project ''Medical Big Data Privacy and Security Risk Measurement and Privacy Protection in the Cloud Environment'', with the cooperation unit ''Kunming First People's Hospital'' jointly developed the questionnaire on the importance of medical big data security and privacy leakage risk indicators. 11 experts related to the research direction were invited to fill in, and the questionnaire was collected and collected later. Finally, according to the feedback from 11 experts, the weights of 4 primary indicators and 35 secondary indicators are determined by GI method and entropy weight method respectively.

B. CALCULATE THE INDICATOR WEIGHT

1) GI METHOD DETERMINES THE INDICATOR WEIGHT

According to the feedback from the first expert, the importance order of the four primary indicators is: $A_3 > A_2 > A_1 > A_4$.

The importance relationship between the indicators is as follows: $r_2 = 1.1, r_3 = 1, r_4 = 1.1$.

Then the weight of the primary indicator is: $w^1_{A_4} = (1 + r_2r_3r_4 + r_3r_4 + r_4)^{-1} = 0.227, w^1_{A_1} = r_4w^1_{A_4} = 0.249, w^1_{A_2} = r_3w^1_{A_1} = 0.249, w^1_{A_3} = r_2w^1_{A_2} = 0.274$.

At the same time, the importance of the secondary indicators $B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8, B_9$ under the primary indicator A_1 in the feedback information of the expert is: $B_5 > B_6 > B_3 > B_4 > B_2 > B_7 > B_1 > B_8 > B_9$, the importance relationship between each indicator is as follows: $r_2 = w^1_{B_5}/w^1_{B_6} = 1.2, r_3 = w^1_{B_6}/w^1_{B_3} = 1.6, r_4 = w^1_{B_3}/w^1_{B_4} = 1.3, r_5 = w^1_{B_4}/w^1_{B_2} = 1.1, r_6 = w^1_{B_2}/w^1_{B_7} = 1.3, r_7 = w^1_{B_7}/w^1_{B_1} = 1.2, r_8 = w^1_{B_1}/w^1_{B_8} = 1.4, r_9 = w^1_{B_8}/w^1_{B_9} = 1.1$.

Then the single layer weight of the secondary indicator is:$w^1_{B_9} = (1 + r_2r_3r_4r_5r_6r_7r_8r_9 + r_3r_4r_5r_6r_7r_8r_9 + \cdots + r_8r_9 + r_9)^{-1} = 0.038, w^1_{B_8} = r_9w^1_{B_9} = 0.042, w^1_{B_1} = r_8w^1_{B_8} = 0.059, w^1_{B_7} = r_7w^1_{B_1} = 0.071, w^1_{B_2} = r_6w^1_{B_7} = 0.092, w^1_{B_4} = r_5w^1_{B_2} = 0.101, w^1_{B_3} = r_4w^1_{B_4} = 0.131, w^1_{B_6} = r_3w^1_{B_3} = 0.210, w^1_{B_5} = r_2w^1_{B_6} = 0.252$.

Therefore, it is easy to obtain the comprehensive weight of the secondary indicator as follows: $w^{1'}_{B_1} = 0.015, w^{1'}_{B_2} = 0.023, w^{1'}_{B_3} = 0.033, w^{1'}_{B_4} = 0.025, w^{1'}_{B_5} = 0.063, w^{1'}_{B_6} = 0.052, w^{1'}_{B_7} = 0.018, w^{1'}_{B_8} = 0.010, w^{1'}_{B_9} = 0.009$.

By analogy, the weights of the corresponding primary indicators and secondary indicators can be obtained through the above process according to the questionnaire of the other 10 experts. However, some secondary indicators are cross-existing. For example, B_6 in the secondary indicator belongs to both the primary indicator A_1 and A_2, so when calculating the single-layer weight and comprehensive weight of B_6, should calculate the single-level indicator weight $w^k_{B_{6A_1}}$ and the comprehensive weight $w^{k'}_{B_{6A_1}}$ of the B_6 under the primary indicator A_1, then calculate the single-level indicator weight $w^k_{B_{6A_2}}$ and the comprehensive weight $w^{k'}_{B_{6A_2}}$ of the B_6 under the primary indicator A_2. Finally, the results are added as the final statistics: single layer weight is $w^k_{B_{6A_1}} + w_{B_{6A_2}}{}^k$ and comprehensive weight is $w^{k'}_{B_{6A1}} + w^{k'}_{B_{6A2}}$.

After the weights of indicators at all levels are calculated according to the questionnaire of each expert, which need to be sorted out to calculate the cluster decision results. Therefore, we must first quantify the familiarity of experts in this research direction, and the weight indicator L_k of each

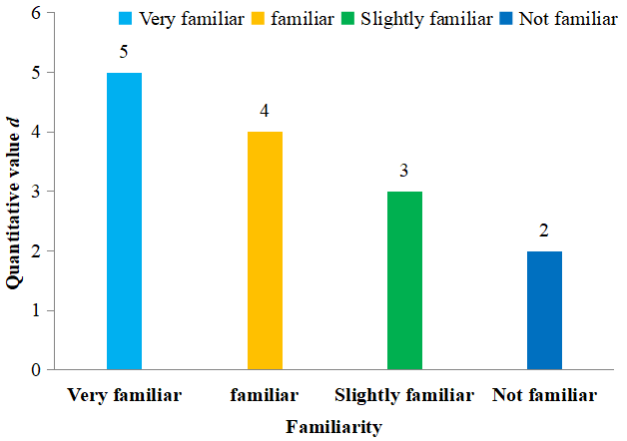

FIGURE 4. Experts familiarity with the field.

expert can be calculated from the quantitative value d of expert familiarity in Figure 4, where $L_k = d_k / \sum_{i=1}^{11} d_i$.

Finally, the weight of the secondary indicators in the medical big data security and privacy leakage risk indicator system is obtained by formula (2):

$$\begin{aligned}&\left\{w^g_{B_1}, w^g_{B_2}, w^g_{B_3}, w^g_{B_4}, w^g_{B_5}, \cdots, w^g_{B_{35}}\right\}\\ &= \{0.0018, 0.0222, 0.0246, 0.0049, 0.0417, 0.0464,\\ &\quad 0.0332, 0.0485, 0.0481, 0.0384, 0.0486, 0.0214, 0.0072,\\ &\quad 0.0080, 0.0140, 0.0277, 0.0430, 0.0023, 0.0199, 0.0193,\\ &\quad 0.0489, 0.0492, 0.0401, 0.0485, 0.0473, 0.0087, 0.0161,\\ &\quad 0.0405, 0.0332, 0.0344, 0.0376, 0.0352, 0.0017, 0.0358,\\ &\quad 0.0016\}\end{aligned}$$

2) ENTROPY WEIGHT METHOD TO DETERMINE THE INDICATOR WEIGHT

The expert score sheet shown in Table 3 was obtained by statistic the scores of the indicators in the "Medical Big Data Security and Privacy Leakage Risk Indicator Importance Questionnaire".

The judgment matrix R can be obtained from the expert scoring table:

$$R = \begin{bmatrix} 4 & 4 & 4 & 5 & 3 & 3 & \cdots & 3 \\ 4 & 7 & 6 & 5 & 4 & 5 & \cdots & 4 \\ 6 & 6 & 5 & 4 & 6 & 5 & \cdots & 5 \\ 5 & 4 & 4 & 4 & 3 & 3 & \cdots & 3 \\ 7 & 6 & 7 & 7 & 6 & 6 & \cdots & 8 \\ 8 & 7 & 5 & 7 & 6 & 5 & \cdots & 6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 4 & 4 & 4 & 3 & 3 & 3 & \cdots & 3 \end{bmatrix}$$

Then according to the formula (3) (4) (5), the data is normalized by the programming tool to obtain the result shown in Figure 5.

TABLE 3. Expert score sheet.

Indicator \ Expert	1	2	3	4	5	6	7	8	9	10	11
B_1	4	4	4	5	3	3	3	3	4	4	3
B_2	4	7	6	5	4	5	5	4	4	7	4
B_3	6	6	5	4	6	5	4	6	4	4	5
B_4	5	4	4	4	3	3	3	5	3	4	3
B_5	7	6	7	7	6	6	5	6	5	6	8
B_6	8	7	5	7	6	5	9	5	6	7	6
B_7	8	6	5	4	6	5	5	4	5	4	4
B_8	8	7	5	7	7	6	7	8	9	6	8
B_9	7	6	8	6	9	6	5	6	7	5	8
B_{10}	5	6	6	7	5	7	9	6	5	7	5
B_{11}	8	7	6	8	9	5	6	7	9	8	9
B_{12}	7	5	6	6	4	4	4	5	4	5	4
B_{13}	3	5	4	6	4	4	5	3	3	4	3
B_{14}	6	5	5	4	3	4	4	3	4	3	4
B_{15}	5	4	4	4	5	3	4	5	3	4	4
B_{16}	5	5	6	4	6	7	5	5	4	5	4
B_{17}	7	8	5	1	6	5	7	5	8	6	6
B_{18}	4	4	3	4	3	4	5	5	3	3	4
B_{19}	4	4	5	6	6	5	4	5	4	6	4
B_{20}	6	6	4	5	4	6	5	4	4	4	5
B_{21}	8	8	7	9	9	7	7	5	6	7	8
B_{22}	7	9	8	5	9	1	8	9	9	8	7
B_{23}	6	8	7	7	6	5	8	7	6	5	5
B_{24}	9	6	8	7	6	9	1	8	7	8	5
B_{25}	6	8	7	7	5	6	5	7	8	6	9
B_{26}	4	4	5	3	4	3	3	4	5	4	3
B_{27}	5	4	4	5	6	4	5	4	3	4	5
B_{28}	7	7	6	6	5	8	5	6	7	9	5
B_{29}	7	6	4	7	4	5	7	6	5	4	5
B_{30}	6	6	7	5	6	4	6	4	5	6	5
B_{31}	8	7	5	6	7	5	5	6	5	6	5
B_{32}	5	6	7	8	5	6	4	5	6	5	5
B_{33}	3	3	4	3	4	4	4	3	3	2	2
B_{34}	7	6	5	4	6	6	7	6	5	5	5
B_{35}	4	4	4	3	3	3	2	4	3	3	3

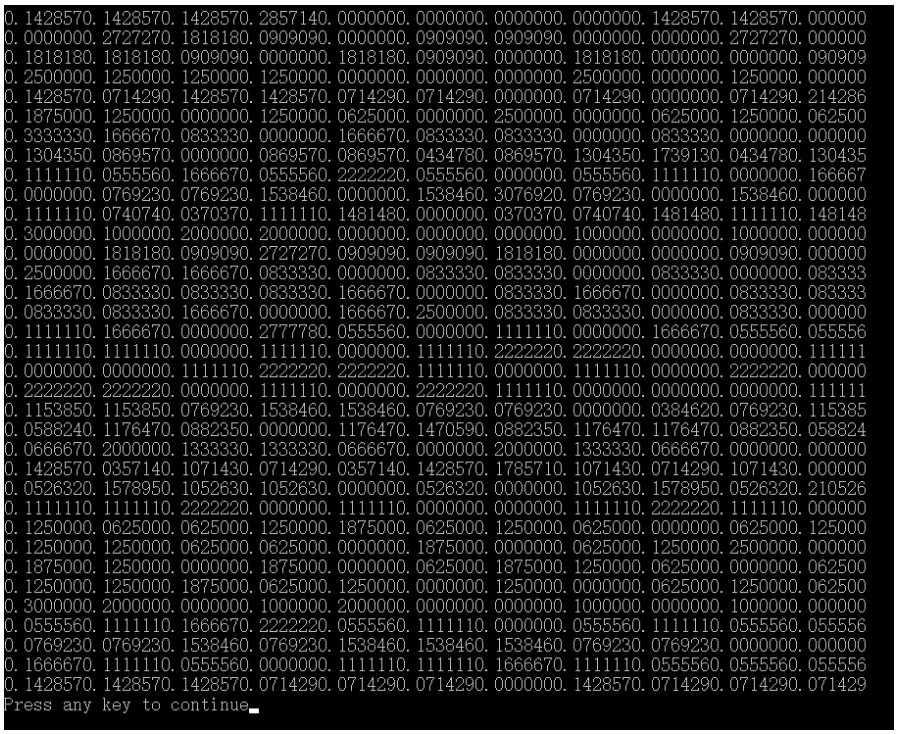

FIGURE 5. Data normalization results.

Organize the matrix X''_{ij}:

$$X''_{ij} = \begin{bmatrix} 0.142857 & 0.142857 & \cdots & 0.000000 \\ 0.000000 & 0.272727 & \cdots & 0.000000 \\ 0.181818 & 0.181818 & \cdots & 0.090909 \\ 0.250000 & 0.125000 & \cdots & 0.000000 \\ 0.142875 & 0.071429 & \cdots & 0.214286 \\ 0.187500 & 0.125000 & \cdots & 0.062500 \\ 0.333333 & 0.166667 & \cdots & 0.000000 \\ 0.130435 & 0.086957 & \cdots & 0.130435 \\ \vdots & \vdots & \ddots & \vdots \\ 0.142857 & 0.142857 & \cdots & 0.071429 \end{bmatrix}$$

The weights of the secondary indicators in the medical big data security and privacy leakage risk indicator system are obtained by formula (6) (7) (8):

$$\begin{aligned} &\left\{w^s_{B_1}, w^s_{B_2}, w^s_{B_3}, w^s_{B_4}, w^s_{B_5}, \cdots, w^s_{B_{35}}\right\} \\ &= \{0.045261, 0.050500, 0.035101, 0.046305, 0.020284, \\ &\quad 0.030484, 0.042205, 0.012149, 0.023383, 0.040344, \\ &\quad 0.013026, 0.048890, 0.038413, 0.029153, 0.018073, \\ &\quad 0.029153, 0.033062, 0.035423, 0.046148, 0.046148, \\ &\quad 0.011135, 0.009257, 0.028308, 0.014018, 0.021663, \\ &\quad 0.035423, 0.012385, 0.030484, 0.029005, 0.018418, \\ &\quad 0.048890, 0.015997, 0.018069, 0.012658, 0.010787\} \end{aligned}$$

3) DETERMINE THE COMPREHENSIVE WEIGHT

w^g_j and w^s_j have been obtained, so the weights based on the GI method and the entropy weight method can be determined from the formula (11), as shown in Figure 6:

C. RISK QUANTIFICATION

Take the Kunming city's first hospital as an example to evaluate the data security and privacy leakage risks of the hospital. Invite relevant experts and third-party evaluation agencies to conduct on-site investigations, security audits, software analysis and other research work.

Finally, experts give the hospital data security and privacy leakage risk analysis level V, and assign it: V = {high risk, slightly high risk, medium risk, slightly low risk, low risk} = {8, 6, 4, 2, 1}.

Refer to the relevant literature to count the number of voters in each risk level of the hospital [33], as shown in Table 4.

From Table 4, the indicator weight vector A and the fuzzy relation matrix R can be generated:

$$\begin{aligned} A_{1\times 35} = [&0.0257, 0.0378, 0.0304, 0.0277, \\ &0.0299, 0.0376, 0.0382, 0.0285, 0.0345, 0.0395, \\ &0.0290, 0.0365, 0.0244, 0.0196, 0.0162, 0.0285, \\ &0.0375, 0.0205, 0.0343, 0.0341, 0.0281, 0.0272, \\ &0.0336, 0.0295, 0.0332, 0.0234, 0.0141, 0.0350, \\ &0.0309, 0.0256, 0.0438, 0.0246, 0.0107, 0.0231, \\ &0.0067] \end{aligned}$$

The fuzzy comprehensive evaluation result vector B is easily obtained by the formula (12), and the normalized result is as follows:

$$\mathrm{B} = [0.2709 \quad 0.2090 \quad 0.2093 \quad 0.2115 \quad 0.1046]$$

Finally, the fuzzy comprehensive evaluation results are analyzed to calculate the comprehensive membership

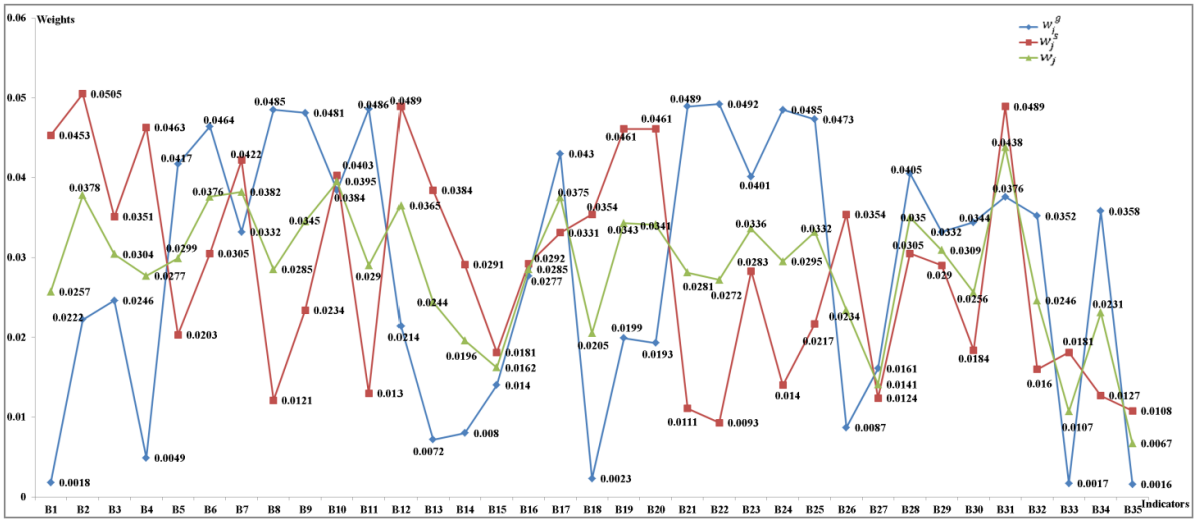

FIGURE 6. The weight results based on GI method and entropy weight method.

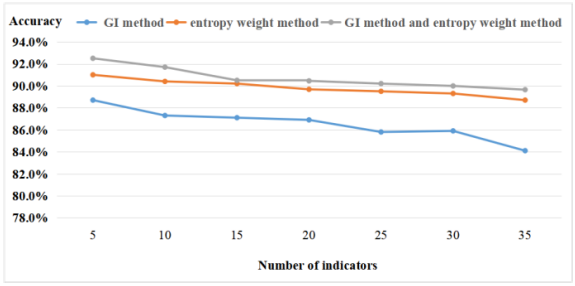

FIGURE 7. Accuracy comparison result.

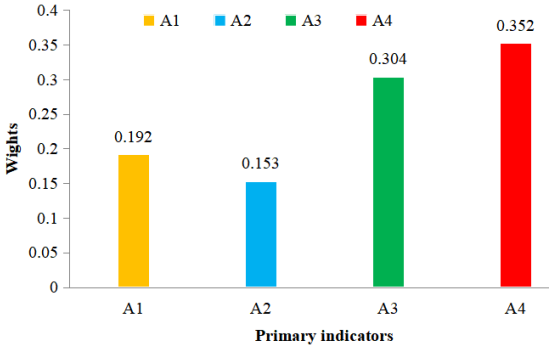

FIGURE 8. Weight of primary indicator.

degree L:

$$L = 8 * 0.2709 + 6 * 0.2090 + 4 * 0.2093 + 2 * 0.2115 + 1 * 0.104 = 4.786$$

The survey results of the Kunming First People's Hospital showed that the hospital's medical big data security and privacy leakage risk is between medium and slightly high risk, but closer to medium risk. In addition, the same method also evaluates the risks of medical big data security and privacy leakage of another Grade III Level A hospitals and two Grade II Level A hospitals, and the results are 4.9320, 5.4876 and 5.3310 respectively. Finally, ranked by risk size is: the Grade II Level A hospitals > the Grade III Level A hospitals

D. RESULT ANALYSIS

This paper mainly evaluates the risks of medical information security and privacy disclosure involved in urban computing, and selects two Grade III Level A hospitals and two Grade II Level A hospitals respectively. The results show that the Grade III Level A hospital slightly lower than the Grade II Level A hospital in terms of data security and privacy risks. When selecting the hospital, it is mainly considered whether the key risk indicators during the data collection, storage, transmission, analysis and use phases are consistent. The survey shows that the key indicators of the four hospitals are roughly the same, and there is not much difference. Therefore, this also explains the reliability of the evaluation method in this paper to some extent. In view of the above results, it is recommended that the Grade II Level A hospitals should strengthen management in medical data, conduct risk analysis on a regular basis, and adopt certain risk reduction strategies timely to ensure the security of medical data and user privacy. However, in order to further strengthen the persuasiveness, we conducted a comparative analysis of the accuracy of the method in this paper. The results are shown in Figure 7.

E. PREVENTIVE SOLUTION

The weights of the primary indicators can be obtained from Figure 6: $\{w_{A_1}, w_{A_2}, w_{A_3}, w_{A_4}\} = 0.192, 0.153, 0.304, 0.352\}$. As shown in Figure 8, in the four stages of the collection, transmission, storage, analysis and use of medical big data, the weights occupied by the data collection and data transmission stages are relatively small, while the weights of the data storage and data analysis and use phases are relatively higher, according to this situation, we specifically analyze the reasons, and combine with the comprehensive weight of

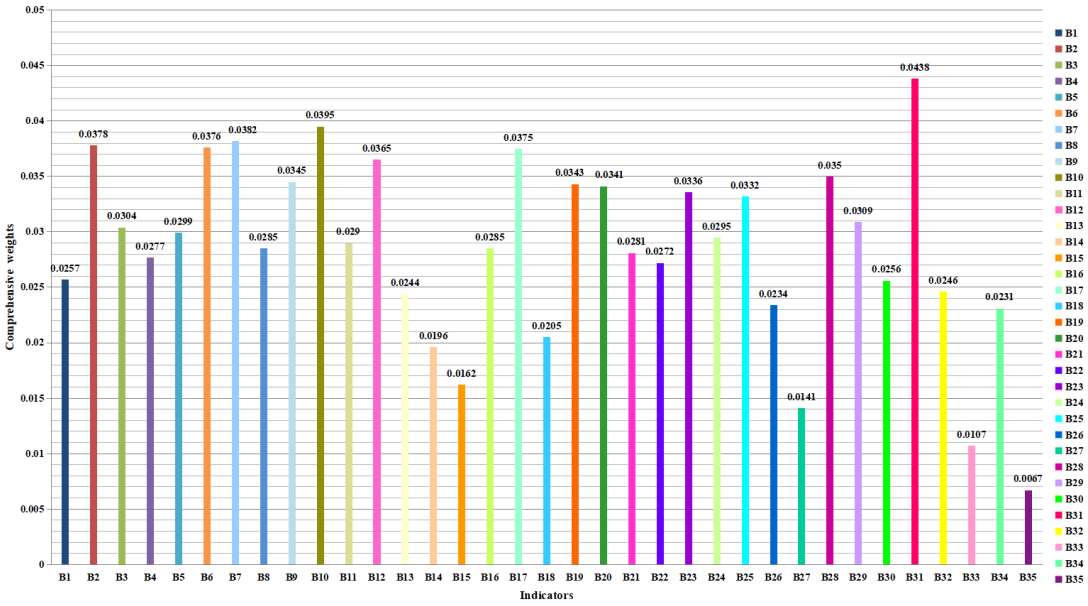

FIGURE 9. Comprehensive weight of secondary indicators.

secondary indicators in Figure 9 to give the corresponding risk reduction strategy.

1) DATA COLLECTION PHASE

The data collection phase mainly collects data generated by sensors, smart devices, etc., most of which are raw data that has not been processed, resulting in the user's privacy completely out of the user's own control, so this stage of privacy issues is even more prominent. The key factors in the corresponding secondary indicators include: wearable device positioning function (0.0382), the patient's lack of right to know (0.0378), and third party malicious behavior (0.0376).

At this stage, privacy protection technologies such as cryptography, local differential privacy, social graph privacy, and location track privacy are recommended to prevent privacy breaches caused by malicious actions of wearable devices and third parties. In addition, an informed consent approach should be established that is consistent with the cultural characteristics of our country.

2) DATA TRANSMISSION PHASE

In the data transmission phase, the collected data is transmitted to a large database through terminal devices such as smart devices and sensors. The data security problem at this stage is more prominent. The Key factors in the corresponding secondary indicators include: lack of a unified data transfer protocol standard (0.395), service engine vulnerabilities (0.0365), and hacker attacks (0.0285).

At present, there are many kinds of data transmission protocols in the market, such as Bluethtooth Medical Device Profile, IEEE 11073-104xx specification, etc. Only when a compatible data transmission protocol is established can data security be guaranteed to a certain extent. At the same time, it is recommended to use VPN technology or SSL communication protocol in the data transmission phase to prevent hackers from attacking during data transmission, and regularly scan for security vulnerabilities to prevent security problems from happening in time.

3) DATA STORAGE PHASE

A large amount of valuable data is stored together, which will not only become the target of external hackers, but also become the main target of internal personnel to steal information, and also include the unauthorized use of some data. Therefore, the security issues facing the data storage hierarchy are multifaceted, including data security, platform security, privacy security and other security requirements. The key factors in the corresponding secondary indicators include: internal personnel stealing information (0.0375), virtual vulnerability (0.0343) and firewall vulnerability (0.0341).

In response to these problems, it is recommended to establish a management system of ''multi-level authorization and consistent responsibility'', and strict implementation of medical data confidentiality regulations. Establish and improve the personal privacy information protection mechanism, severely punish the illegal stealing, trafficking of medical information. At the same time, it is necessary to update the system patch in time, fundamentally solve the vulnerability problem, strengthen the firewall configuration, and improve the defense capability. In addition, third-party software can be used to provide a certain security guarantee for data storage.

Finally, in order to further ensure the security and privacy protection of data storage, technologies such as access control, secure retrieval, and secure computing can be adopted for big data security. Privacy protection can use differential privacy, k-anonymity, etc.

TABLE 4. Indicator weights and expert survey results.

Secondary	Weights	Number of experts corresponding to each risk Level				
		H	SH	M	SL	L
B_1	0.0257	1	2	4	4	1
B_2	0.0378	4	4	2	1	0
B_3	0.0304	3	3	3	2	0
B_4	0.0277	2	1	5	3	0
B_5	0.0299	3	2	2	2	2
B_6	0.0376	4	3	3	1	0
B_7	0.0382	5	3	1	2	0
B_8	0.0285	2	3	3	2	1
B_9	0.0345	4	2	2	2	1
B_{10}	0.0395	5	3	2	1	0
B_{11}	0.0290	3	2	3	2	1
B_{12}	0.0365	5	3	1	1	1
B_{13}	0.0244	0	1	3	6	1
B_{14}	0.0196	1	0	2	4	4
B_{15}	0.0162	1	0	2	3	5
B_{16}	0.0285	2	2	3	3	1
B_{17}	0.0375	4	3	2	1	1
B_{18}	0.0205	0	1	3	4	3
B_{19}	0.0343	4	3	1	1	2
B_{20}	0.0341	5	2	1	2	1
B_{21}	0.0281	2	2	2	4	1
B_{22}	0.0272	2	1	5	2	1
B_{23}	0.0336	4	4	2	1	1
B_{24}	0.0295	3	3	1	2	2
B_{25}	0.0332	3	3	2	2	1
B_{26}	0.0234	1	0	4	6	0
B_{27}	0.0141	0	1	2	2	6
B_{28}	0.0350	5	3	1	2	0
B_{29}	0.0309	3	4	1	2	1
B_{30}	0.0256	1	1	4	3	2
B_{31}	0.0438	6	3	1	1	0
B_{32}	0.0246	0	1	4	5	1
B_{33}	0.0107	0	1	2	2	6
B_{34}	0.0231	0	1	3	5	2
B_{35}	0.0067	0	0	2	3	6

4) DATA USE AND SHARING PHASE

The data collection, transmission, and storage stages are all for data usage and analysis services. This stage mainly uses data mining and other technologies to extract information hidden inside the data. Strictly speaking, the privacy issue at this stage is more prominent, but since the data is in the hands of people, the purpose of using the data is not controlled. There are also no specific laws and regulations. So, it is not surprising that data security and privacy issues at this stage are most prominent. The corresponding secondary key indicators include: hospital information platform interaction standard is not standardized (0.0438), data acquirer's dishonest behavior (0.035), lack of special laws and regulations (0.0345), security audit (0.0336), digital certificate reliability (0.0332) and so on.

In view of these problems, medical institutions need to further supplement and improve the standardization of hospital platform interaction and electronic medical record interaction to prevent data security problems from occurring due to inconsistent standards in the interaction process. In addition, China should strengthen laws, regulations and supervision mechanisms on the application of medical big data to prevent the illegal use of data by users. Finally, medical care organizations should establish a trusted digital identity management system to ensure that access to medical data is manageable, controllable, and traceable. At the same time, each user of the access system should implement unified identity identification and management to prevent unauthorized access and behavioral repudiation.

VI. CONCLUSION

As an emerging cross-cutting area, urban computing has generated a large amount of data in urban space using sensing technology and large-scale computer infrastructure. At the same time, the arrival of the era of big data has brought challenges and opportunities for urban computing. This paper analyzed the security and privacy risk assessment methods of medical big data by taking the medical service industry in urban computing as an example. However, in a complex cross-domain environment, ensuring that users' access to data is manageable, controllable, and traceable will be the next step of research.

(Rong Jiang and Mingyue Shi contributed equally to this work.)

REFERENCES

[1] R. Y. Yu, Y. Yang, L. Y. Yang, and G. J. Han, "RAQ—A random forest approach for predicting air quality in urban sensing systems," *Sensors*, vol. 16, no. 1, p. 16, Jan. 2016.

[2] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[3] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, p. 38, 2014.

[4] A. Amairah, B. N. Al-Tamimi, M. Anbar, and K. Aloufi, "Cloud computing and Internet of Things integration systems: A review," in *Recent Trends in Data Science and Soft Computing*. Cham, Switzerland: Springer, vol. 843. Sep. 2018, pp. 406–414.

[5] Z. Huang, J. Tang, G. Shan, J. Ni, Y. Chen, and C. Wang, "An efficient passenger-hunting recommendation framework with multi-task deep learning," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2019.2901759.

[6] Y. Niu, M. M. Yan, H. Zheng, and J. J. Yang, "Research review on medical data access control in cloud computing environment," *Smart Health*, vol. 1, pp. 23–27, 2016.

[7] M. Xu, J. Shen, and H. Y. Yu, *Big Data Medical Treatment*. Beijing, China: China Machine Press, 2017, pp. 1–27.

[8] X. T. Jin, *Big Data Of Medical Care*. Beijing, China: People's Medical, 2018, pp. 37–193.
[9] Y. F. Wu, Y. Wan, and J. Fan, "Research on government behavior of medical information privacy management in the era of big data," *Manage. World*, vol. 1, no. 1, pp. 174–175, Jan. 2017.
[10] W. Wei, S. Liu, W. Li, and D. Du, "Fractal intelligent privacy protection in online social network using attribute-based encryption schemes," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 3, pp. 736–747, Sep. 2018.
[11] F. Y. Li, "Research on information security management of smart community under construction in China," M.S. thesis, Northeastern Univ., Boston, MA, USA, 2015.
[12] Z. X. Lu, "Reinforcing and weakening influence of smart city construction on information security," *Inf. Commun.*, no. 7, pp. 119–120, Jul. 2018.
[13] H. Guo and X. N. Su, "Research on environment, challenges and models of information security management in smart cities," *Library Inf. Service*, vol. 60, no. 19, pp. 49–58, Oct. 2016.
[14] D. J. Zhang, X. Y. Bi, X. Lu, and X. L. Han, "Research on smart city information security system," *Inf. Secur. Res.*, vol. 3, no. 8, pp. 710–717, Aug. 2017.
[15] K. Wang, "The legal protection of personal-information under the background of smart city," M.S. thesis, Huazhong Univ. Sci. Technol., Wuhan, China, 2013.
[16] F. S. Ferraz and C. A. G. Ferraz, "Smart city security issues: Depicting information security issues in the role of an urban environment," in *Proc. IEEE/ACM 7th Int. Conf. Utility Cloud Comput.*, Dec. 2014, pp. 842–847.
[17] S. Xiang, "Model construction and empirical study on smart city information security risk assessment," M.S. thesis, Xiangtan Univ., Xiangtan, China, 2017.
[18] D. G. Feng, *Big Data Security and Privacy Protection*. Beijing, China: Tsinghua Univ. Press, 2018.
[19] *Broader Access Models for Realizing Information DomiCorporation*, document JSR-04-132, JASON, MITRE McLean, VA, USA, Nov. 2004.
[20] Q. Ni, E. Bertino, and J. Lobo, "Risk-based access control systems built on fuzzy inferences," in *Proc. 5th ACM Symp. Inf., Comput. Commun. Secur. (ASIACCS)*, Beijing, China, 2010, pp. 250–260.
[21] P.-C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, and A. S. Reninger, "Fuzzy multi-level security: An experiment on quantified risk-adaptive access control," in *Proc. IEEE Symp. Secur. Privacy (SP)*, Oakland, CA, USA, May 2007, pp. 222–230.
[22] Q. Wang and H. Jin, "Quantified risk-adaptive access control for patient privacy protection in health information systems," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Secur. (ASIACCS)*, Hong Kong, 2011, pp. 406–410.
[23] Z. Hui, H. Li, M. Zhang, and D. G. Feng, "Risk-adaptive access control model for big data in healthcare," *J. Commun.*, vol. 36, no. 12, pp. 190–199, Jul. 2017.
[24] Q. J. Zhang, "Cloud computing privacy security risk assessment," M.S. thesis, Yunnan Univ., Kunming, China, 2015.
[25] Z. Ma, R. Jiang, M. Yang, T. Li, and Q. Zhang, "Research on the measurement and evaluation of trusted cloud service," *Soft Comput.*, vol. 22, no. 4, pp. 1247–1262, Feb. 2018.
[26] J. Zhang, "Risk assessment of privacy security of medical big data in cloud environment," M.S. thesis, Yunnan Univ. Finance Econ., Kunming, China, 2018.
[27] X. Y. Huang, "Research on cloud computing platform security evaluation index model," M.S. thesis, Donghua Univ., Shanghai, China, 2017.
[28] D.-G. Feng, M. Zhang, Y. Zhang, and Z. Xu, "Study on cloud computing security," *J. Softw.*, vol. 22, no. 1, pp. 71–83, Nov. 2011.
[29] Y. F. Cai, "Risk assessment model of information system security based on cloud computing," *China Manage. Informationization*, vol. 13, no. 12, pp. 75–77, Dec. 2010.
[30] Z. W. Jiang, W. R. Zhao, Y. Liu, and B. X. Liu, "Model for cloud computing security assessment based on classified protection," *Comput. Sci.*, vol. 40, no. 8, pp. 151–156, Dec. 2013.
[31] Y. J. Zhu, "Risk analysis and assessment of privacy disclosure in location-based services," M.S. thesis, Guizhou Univ. Finance Econ., Guiyang, China, 2016.
[32] Q. Q. Kuang, "Risk assessment based on personal privacy disclosure," M.S. thesis, Guizhou Univ. Finance Econ., Guiyang, China, 2016.
[33] B. Tian, Y. S. Zheng, P. Y. Liu, and C. H. Li, "Risk evaluation index and empirical research on mobile app users' privacy information disclosure," *Library Inf. Service*, vol. 62, no. 19, pp. 101–110, Oct. 2018.
[34] N. Y. Tong, "Personal privacy protection in the era of big data," M.S. thesis, Shanghai Normal Univ., Shanghai, China, 2015.
[35] G. Zhu, M. N. Feng, Y. Chen, and J. Y. Yang, "Research on fuzzy evaluation of social network privacy risk in big data environment," *Inf. Sci.*, vol. 34, no. 9, pp. 94–98, Sep. 2016.
[36] Y. Li, "Research on personal information protection in the era of big data," M.S. thesis, Southwest Univ. Political Sci. Law, Chongqing, China, 2016.
[37] H. Y. Li, G. H. Luan, Y. J. Song, S. L. Chu, and M. Qin, "HSE audit weight assignment method for oil companies based on method G1 and entropy weight method," *Inf. Saf. Manage.*, vol. 36, no. 5, pp. 91–95, May 2018.
[38] X. J. Wang and Y. J. Guo, "Consistency analysis of judgment matrix based on G1 method," *Chin. J. Manage. Sci.*, vol. 14, no. 3, pp. 66–69, Jun. 2006.
[39] M. L. Ran, H. Huang, and Y. Zhong, "Evaluation model construction based on entropy weight method," *Technol. Wind*, vol. 14, pp. 207–208, 2018.
[40] K. Zhu, X. He, B. Xiang, L. Zhang, and A. Pattavina, "How dangerous are your smartphones? App usage recommendation with privacy preserving," *Mobile Inf. Syst.*, vol. 4, no. 5, pp. 1–10, May 2016.
[41] G. Bansal, F. M. Zahedi, and D. Gefen, "Do context and personality matter? Trust and privacy concerns in disclosing private information online," *Inf. Manage.*, vol. 53, no. 1, pp. 1–21, Aug. 2016.
[42] H.-G. Gao, X.-Y. Li, B. Zhang, and W. Xiao, "Information security risk assessment based on information measure and fuzzy clustering," *J. Softw.*, vol. 6, no. 11, pp. 2159–2166, Nov. 2011.

RONG JIANG received the Ph.D. degree in system analysis and integration from the School of Software, Yunnan University, China. He is currently a Distinguished Professor and Doctoral Supervisor with the School of Information, Yunnan University of Finance and Economics, Kunming, China. He is an expert enjoying special government allowances in Yunnan Province, an excellent Teacher in Yunnan Province, the Director of Key Laboratory of Service Computing and Security Management of Yunnan Provincial Universities, and the Director of Kunming Key Laboratory of Information Economy and Information Management. He has published more than 40 papers and ten text books. He has received more than 50 prizes in recent years. His current research interests include cloud computing, big data, block chain, and software engineering.

MINGYUE SHI was born in Bozhou, China, in 1993. She is a graduate student majoring in computer software and theory with the Yunnan University of Finance and Economics. Her current research interests include cloud computing, big data, and information security management.

WEI ZHOU received the Ph.D. degree in management science and engineering from Southeast University, Nanjing, China, in 2012. From June 2014 to November 2014, he was an Academic Visitor with the School of Social Science, Policy and Evaluation, Claremont Graduate University, Claremont, CA, USA. He is currently a Professor with the School of Finance, Yunnan University of Finance and Economics, Kunming, China. He has published more than 30 peer-reviewed papers, many in high-quality international journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Computers and Industrial Engineering*, *Knowledge-Based Systems*, *Applied Mathematical Modeling*, *Technological and Economic Development of Economy*, the *International Journal of Intelligent Systems*, the *Journal of Intelligent and Fuzzy Systems*, and the *Journal of Applied Mathematics*. His current research interests include risk decision making, risk management, group decision making, and information fusion. He serves as a Reviewer for more than ten international journals.