Hindawi
Security and Communication Networks
Volume 2020, Article ID 5610839, 18 pages
https://doi.org/10.1155/2020/5610839

Research Article

A Privacy Risk Assessment Model for Medical Big Data Based on Adaptive Neuro-Fuzzy Theory

Mingyue Shi,[1,2,3] Rong Jiang,[1,2,3] Wei Zhou,[2,3,4] and Sen Liu[2,3,5]

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

Correspondence should be addressed to Mingyue Shi; 1420846885@qq.com

Received 19 December 2019; Revised 28 June 2020; Accepted 9 September 2020; Published 29 September 2020

Academic Editor: Savio Sciancalepore

Information leakage in the medical industry has become an urgent problem to be solved in the field of Internet security. However, due to the need for automated or semiautomated authorization management for privacy protection in the big data environment, the traditional privacy protection model cannot adapt to this complex open environment. Although some scholars have studied the risk assessment model of privacy disclosure in the medical big data environment, it is still in the initial stage of exploration. This paper analyzes the key indicators that affect medical big data security and privacy leakage, including user access behavior and trust, from the perspective of users through literature review and expert consultation. Also, based on the user's historical access information and interaction records, the user's access behavior and trust are quantified with the help of information entropy and probability, and a definition expression is given explicitly. Finally, the entire experimental process and specific operations are introduced in three aspects: the experimental environment, the experimental data, and the experimental process, and then, the predicted results of the model are compared with the actual output through the 10-fold cross verification with Matlab. The results prove that the model in this paper is feasible. In addition, the method in this paper is compared with the current more classical medical big data risk assessment model, and the results show that when the proportion of illegal users is less than 15%, the model in this paper is more superior in terms of accuracy and recall.

1. Introduction

With the development of information technology, the era of big data has come quietly, bringing opportunities and different challenges to all walks of life. Among them, the medical field is a very special field. Its particularity lies in that all its data are closely related to everyone's life, involving the whole life process such as people's food, clothing, housing, life, illness, and death, which is the core asset of big data. The outline of the "Healthy China 2030" plan issued by the CPC central committee and the State Council in October 2016 pointed out that the total size of the health service industry would reach 8 trillion by 2020 and 16 trillion by 2030. In the future holographic digital era, everyone will generate about 605 Tbit of data in their lifetime, and the country will generate 1,000 Zbit of data every year, which has extremely broad industry prospects [1]. Looking at the global development pattern, medical big data has become an emerging industry that promotes the development of economic industries, and it has provided a primitive resource base for scientific and technological innovation [2].

At present, China has realized major "Internet + medical" engineering projects such as telemedicine and cross-domain medical care, which has brought great convenience to people's lives. But technology is a double-edged sword that brings convenience to people's lives while also bringing some disadvantages. Symantec released the top ten industries with severe data leakage in the 2016 "Internet Security Threat Report," ranking the first is the medical industry. In addition, the survey found that more than 90%

of medical social security information in the United States has been sold in the recent years, and more and more medical equipment is out of control. In most hospitals in China, the HIS system has no privacy at all. If a higher-level doctor wants to obtain patient information, he only needs to log in to the terminal to obtain the entire patient's medical record information, while ordinary doctors can obtain all patient information in their own workstation. We can view the current status of the medical industry through a set of data from the United States: medical machinery without any security protection accounts for 77%, and medical equipment with a certain security strategy accounts for 27%. Of these attacks, 17% came from medical equipment, and 75% of the traffic on the hospital's LAN was not monitored and audited, and the hospital itself knew that patient privacy was leaking every day. Although the privacy leak rate of medical big data in China is slightly lower, personal privacy leaks occur from time to time, and there are currently no complete laws and regulations on personal privacy protection. Medical data have their particularity, because their data source is mainly "people." No matter what level of application, it involves human privacy and social stability [1].

Therefore, in the process of the rapid development of the big data Internet, in order to better serve the people, accelerate the development of the digital economy, and promote the integration and open sharing of medical data resources, research on the privacy protection of medical big data is imperative.

The rest of this paper is organized as follows: the second part discusses the research progress and current situation at home and abroad from the two aspects of medical big data security and privacy protection technology, risk-based access control, and summarizes the research status at home and abroad; the third part first introduces the relevant theories and principles, then formalizes the definition of risk indicators, and finally, combines fuzzy theory and a neural network to establish a risk quantification model based on adaptive neural fuzzy theory; the fourth part has carried out simulation experiments to prove that the model in this paper is feasible and efficient; and the fifth part mainly summarizes the work of this paper.

2. Related Works

2.1. Medical Big Data Security and Privacy Protection Technology. Scholars at home and abroad have carried out related research on medical big data security and privacy leakage. Through the collection of relevant literature analysis, it is found that the current academic research in this field mainly adopts technologies such as differential privacy, encryption algorithms, anonymization, and authentication. Research on access control is scarce. For example, literature [2, 3] protects sensitive information in the patient's genome sequence by the differential privacy method and homomorphic encryption method; literature [4, 5] mainly researches medical big data generated by wearable medical sensors and through agitation thresholds and the introduction of binary trees achieve user privacy protection; literature [6–10] researches personal privacy protection issues from the technical level and has established differential privacy protection models, in which the privacy protection of medical big data is given suggestions; Zhang and Zhang [11] advocated the use of data encryption technology to ensure the security of medical data from the first, middle, and last three aspects of the incident; Tian et al. [12] proposed an attribute-based encryption method that structured data access As an authorization policy, the decryptor is allowed to request access to data only when the attributes of the decryptor satisfy the structure; He et al. [13], through the anonymity of user identity and mutual authentication between the client, server, and network administrator, protect patient identity information and data confidentiality.

Xing [14] designed the disease-based secure routing protocol and emergency response scheduling mechanism based on the social layer and cloud service layer of the wireless medical network; Wei and Xu [15] proposed that, from the generation and storage of medical big data starting from the calculation of three nodes, I believe that privacy protection technology should take these three aspects as the starting point to strengthen the "medical network technology structure"; Wang et al. [16] analyzed the privacy leakage of regional medical care and suggested that the data sharing platform started to study privacy protection work; Chen [17] analyzed the possible causes of medical data leakage from a systematic perspective and suggested that information protection should be performed through means such as anonymization, access control, and hierarchical management; Mounia and Habiba [18] analyzed the opportunities and challenges faced by the medical field in the context of big data and, finally, proposed privacy protection issues and coping methods for medical information; Gao and Sang [19] first combined the characteristics of medical big data. Based on this, the entire life cycle of medical data is summarized, and finally, the problems faced at each stage are discussed.

In addition, some scholars have studied the security and privacy protection of medical big data from the perspective of ethics and law. Liu and Wang [20] analyzed the ethical problems existing in the protection of medical information privacy from the perspective of the patient's right of informed consent and the relationship between risks and benefits and gave coping strategies; we should adhere to the principles of transparency and autonomy, establish a comprehensive medical information privacy protection law, and construct a medical-oriented access control system to achieve the privacy protection of medical big data.

2.2. Risk-Based Access Control Technology. Risk refers to the harm that may occur when an event occurs. From a management perspective, risk-based access control is actually applying risk assessment as an effective decision-making tool to access control and dynamically giving subjects access. The concept of risk was proposed for the first time in literature [21], which provides the principles and suggestions that the risk-based access control model should satisfy. Kandala et al. [22] proposed an attribute-based risk access control framework. The author mainly constructed an attribute-based RAdAC model from the user's access purpose, user credibility, historical access behavior, and device attributes, but did not provide a specific risk quantification method. However, literature [23, 24] have given a method for quantifying risk based on factors such as the subject's security level, the sensitivity of the objects, and the mutually exclusive relationship between the objects. In addition, an RBAC model based on risk awareness has been proposed in literature [25], which mainly includes the following three parts: user trust, user's ability to assume roles, and the compatibility between roles and permissions. Finally, a risk assessment model combining these three factors is presented.

In literature [26, 27], according to the risk assessment principles, context, and other information, users' behaviors of viewing, modifying, and deleting medical records are evaluated from the integrity, availability, and confidentiality of medical records. Wang and Hong [28] statically calculated the doctor's access behavior risk by measuring the deviation between the resources accessed by the doctor and the objective. Literature [29] is not specifically for the medical field, but it is a risk decision access control system proposed for a dynamic environment such as the medical industry. This system not only considers the user's historical access behavior, but also considers the user's recent access behavior, and the user's trust and access risk are dynamically adjusted based on the user's access behavior. Choi et al. [30] constructed a context-based medical information risk access control framework, and this framework mainly judges whether to grant users access rights based on authority files, user access logs, and context information. Hui et al. [31] improved on literature [28] not only considering the deviation degree between medical information accessed by doctors and work objective but also considering that doctors may steal patients' privacy by forging work objectives, that is, the deviation degree between work objectives selected by doctors and patients' conditions. Literature [32, 33] mainly analyzed the risk indicator system affecting the privacy leakage of medical big data in the cloud environment from the stages of collection, transmission, storage, and use of medical big data, without designing a specific risk quantification model. Literature [34–36] established the risk assessment model of medical big data with the help of fuzzy theory, but this method has some obvious disadvantages, such as fuzzy rules and membership function are determined based on expert experience and the results are highly subjective.

2.3. Summary and Analysis of Research Status at Home and Abroad. A comprehensive analysis of the relevant research literature at home and abroad found that there are already some scholars doing research in the cross section of information and medicine and also achieved good results. However, from the perspective of the technology and method of privacy protection, it can be roughly divided into privacy protection technology based on anonymity and differential privacy; from the perspective of big data security technology, current research is mainly based on cryptography; however, from a management perspective, analysis can be summarized into the following two categories: one is the use of electronic information technology to monitor networks, platforms, and management systems; the other is the use of computer methods to analyze and mine medical data, such as machine learning.

Although there are some similar studies from the perspective of risk, this is still in the initial stage of exploration, and there is no mature theoretical model system; especially for the privacy protection of medical big data based on risk, it is extremely scarce.

The main contributions of this article are as follows:

(1) Due to the particularity of the medical field, it is difficult to determine whether a user is an "illegal user" based on the user's access behavior. Therefore, this article introduces the user's trust value as one of the risk evaluation indicators. The two jointly evaluate users' access requests to reduce the possibility of system misjudgment.

(2) This paper uses mathematical methods such as information entropy, neural network, fuzzy theory, and probability to establish an adaptive fuzzy neural network model. First, information entropy and probability are used to quantify risk indicators. Then, the knowledge expression ability of the fuzzy theory and the self-learning ability of the neural network are combined, so that the data processing process can be presented in a way that people can easily accept, and at the same time, the risk can be dynamically predicted according to scene changes.

3. Risk Assessment Model Based on Adaptive Neural Fuzzy Theory

Fuzzy theory solves the problems of unclear and uncertain boundaries in intelligent systems by imitating human perception and reasoning [37]. From the perspective of practical application, the application of fuzzy theory mainly focuses on the fuzzy system, especially on fuzzy control. For example, the fuzzy expert system in the medical field is often used for medical diagnosis and decision support [38]. However, there are some disadvantages of fuzzy theory in practical application. For example, in the fuzzy control system, the corresponding rule base should be established according to experience, and the number of rules increases exponentially with the increase of input variables. In addition, the selection of membership functions and optimization work need to be completed manually, and the workload of fuzzy systems in the big data environment becomes extremely complicated [39]. However, the biggest feature of neural networks is to automatically learn new things by imitating the thinking mode of the human brain. The introduction of neural networks into fuzzy theory can help people deal with complex tasks such as rule bases and membership function optimization in fuzzy systems. Therefore, the combination of the two methods can not only improve the expression and learning ability of fuzzy systems but also make the processing of neural networks appear in a way that people can easily accept.

Before introducing how to deal with the security and privacy issues of medical big data with an adaptive neuro-fuzzy system, the relevant theories involved in this model are firstly introduced.

3.1. Relevant Theories and Principles. The risk assessment model based on adaptive neural fuzzy theory mainly involves three key concepts of neural network, neural fuzzy theory, and adaptive neural fuzzy theory. The related content will be described in detail below.

3.1.1. Basic Principles of Neural Networks. A neural network is a network structure formed by the interconnection of many neurons. According to the different connection methods, neural networks can be divided into feed-forward neural networks, feedback neural networks, and self-organizing networks. This article mainly uses feed-forward neural networks, so the other two connection methods will not be introduced here.

Feed-forward neural networks are mainly composed of three parts: the input layer, hidden layer, and output layer. As shown in Figure 1, each circle represents a neuron node, and the output of each layer of neurons will be used as the input of the next layer of neurons [40].

A BP neural network is a typical feed-forward neural network, and the basic idea is to calculate the error value of the previous layer according to the output layer and, then, further calculate the error value of the previous layer based on this error value. At the same time, the weight coefficients of neurons in each layer are adjusted, and so, it went on until the final error value is within the acceptable range [41].

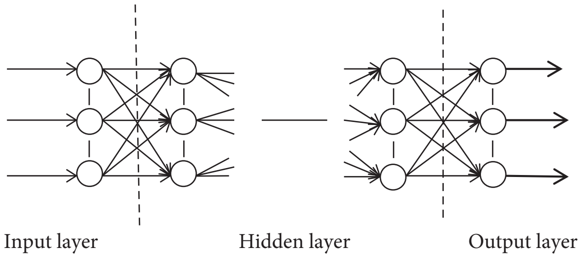


Figure 1: Schematic diagram of the feed-forward neural network structure.

3.1.2. Neural Fuzzy Theory. Although the neural network has strong self-learning ability, its modeling process and data processing process have the characteristics of black box learning, and the processing process cannot be presented in a way that people can easily accept. Therefore, combining the ability of fuzzy theory to express the learning process and the self-learning ability of neural networks is undoubtedly the best choice. In the fuzzy system, the fuzzy models can be divided into two types according to the different output results. One is a Mamdani-type fuzzy model, and the other is a Takagi–Sugeno-type fuzzy model. The former output is a fuzzy set, while the latter outputs the input result in linear combinations or constants of variables [42]. Because the specific risk value is finally calculated in this article, this section mainly introduces the combination of the Takagi–Sugeno fuzzy model and neural network, as shown in Figure 2 [43].

The first layer is used to receive the input variable x_i and pass the input variable $x = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ to the second layer. The role of the second layer is mainly to blur the input variables and calculate the membership function $u_i^j\,(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m_i)$ of each variable, where n represents the input variable and m_i represents the number of fuzzy sets corresponding to the variable x_i. The third layer is used to train the antecedents of fuzzy rules, and each node represents a rule. For a specific input $\mathbf{x}$, $a_j = \min\{u_1^{i_1}, u_2^{i_2}, \ldots, u_n^{i_n}\}$ represents the fitness of each rule, where $i_1 \in \{1, 2, \ldots, m_1\}$, $i_2 \in \{1, 2, \ldots, m_2\}$, $\cdots$, $i_n \in \{1, 2, \ldots, m_n\}$, $j = 1, 2, \ldots, m$, and m represents the total number of rules. The fourth layer performs normalization processing according to the antecedent of the rule, that is, $a_j = (a_j / \sum_{i=1}^{m} a_i)$, where $j = 1, 2, \ldots, m$. The fifth layer performs deblurring processing on the results of each rule aggregation to obtain the output result $y_i = \sum_{j=1}^{m} y_{ij} \times \overline{a}_j$, $i = 1, 2, \ldots, r$. For a more intuitive representation, y_i can be written in the form of the following vector [44]:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{r1} & y_{r2} & \cdots & y_{rm} \end{bmatrix} \begin{bmatrix} \overline{a}_1 \\ \overline{a}_2 \\ \vdots \\ \overline{a}_m \end{bmatrix}. \quad (1)$$

But, this kind of fuzzy control system completed with the help of neural networks has certain problems when dealing with practical problems, such as adjusting parameters and determining the number of hidden layers.

3.2. Quantification of Medical Big Data Risk Based on Adaptive Neuro-Fuzzy Theory. At present, hospital data are basically stored in a local area network. Generally, the outside world cannot steal the patient's private information. In addition, the patient's information will be printed out and stored in the medical record room after the patient is discharged. The workstation of the ordinary user can only query the recent patient's medical information. However, in order not to affect the normal work of the doctor, some highly qualified doctors or experts will be granted extremely high permissions. They can not only access the patient information of their workstations but also log in to the hospital's information center to view all the patients' treatment information. Therefore, their access behavior needs to be evaluated to prevent them from stealing or snooping on patient information.

For the convenience of description, this article divides users into two categories, one is called legal user and the other is called illegal user. Legal users generally only access medical records within their own scope of responsibility, while illegal users, in order to steal more patient

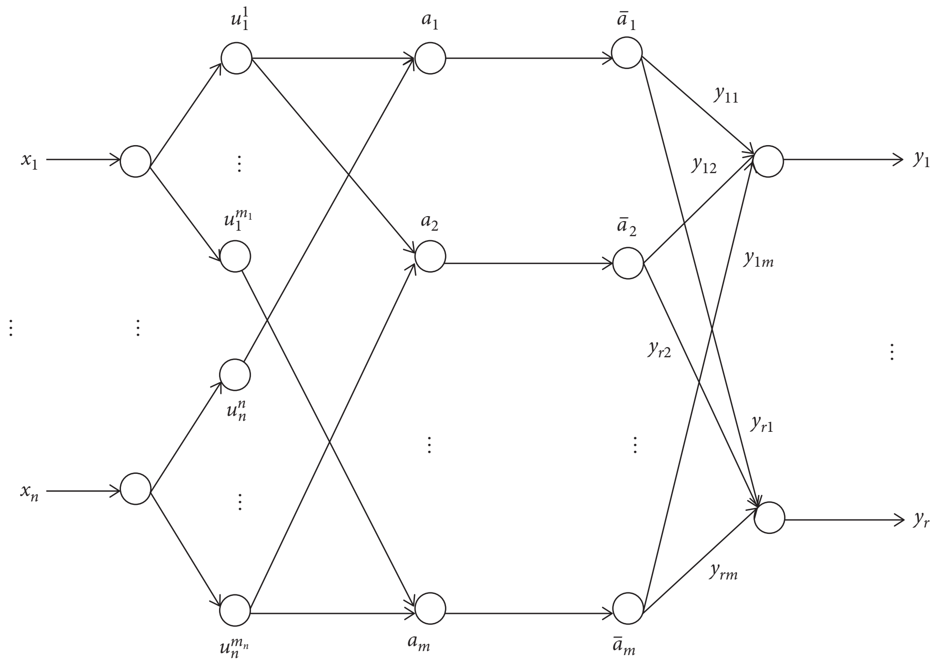


Figure 2: Schematic diagram of the neural fuzzy network structure based on the Takagi–Sugeno model.

information, will also access related medical records by falsifying the patient's condition while completing their own work or access some patients medical records unrelated to the condition [28]. Therefore, legal users can be distinguished from illegal users based on differences in user access behavior. However, we also need to consider some special situations, such as encountering a patient's condition is rare and difficult. In order to ensure the accuracy of the diagnosis, legal users may access additional information from the database that is not related to their work objective, and the more senior the expert, the more often the incurable diseases diagnosed.

In this case, it is difficult for the system to judge whether the user is legitimate or illegal just based on the access behavior, and the access request of the legitimate user may be rejected because of misjudgment. In order to solve this problem, we introduce the user's trust and the user through the user's access behavior trust together to evaluate the user's access request; when a user's access behavior is abnormal, the system will be combined to determine the user's trust, if the user's trust is very high, the system will may be allowed to access, but if the user's trust is lower, then the user will have the risk of stealing the patient's privacy, which, to some extent, can reduce the possibility of miscalculation. Therefore, this article mainly evaluates user access requests from two aspects: user access behavior and user trust.

3.2.1. Formal Definition of Risk Indicators. Before quantifying the risk of privacy leakage of medical big data, this section first formalizes the key index factors that affect the privacy of medical data and turns it into specific mathematical problems. A user's access is recorded as a six-tuple:

$$(U, S, O, M, UT, \text{Risk}), \tag{2}$$

where $U = \{u_1, u_2, \ldots, u_{I_u}\}$ represents the set of all access requesters, including doctors, nurses, technicians, administrators, or the initiators of other actions; $S = \{s_1, s_2, \ldots, s_{I_s}\}$ represents the set of all patients in the hospital, and each patient has corresponding medical records; $O = \{o_1, o_2, \ldots, o_{I_o}\}$ represents a set of task objectives, which is an activity set corresponding to a business process, and each user has his own work objective; $M = \{m_1, m_2, \ldots, m_{I_m}\}$ represents the collection of patient medical information, including basic patient information, medical conditions, and medical records; $UT = \{ut_1, ut_2, \ldots, ut_{I_u}\}$ represents the trust of all users; Risk is the result value of risk quantification; and I_u, I_s, I_o, and I_m, respectively, represent the number of users, patients, work objectives, and medical records.

Since the Adaptive Neural Fuzzy Inference System (ANFIS) cannot identify qualitative index factors, it is necessary to formally describe the user's access behavior and trust so that it turns into a quantitative mathematical problem. This section mainly quantifies the user's access behavior and trust value. The following will introduce the quantification method and process of indicators in detail:

(1) Quantification of User Access Behavior. In order to compare the differences in access behavior between users, we use information entropy to describe the user's access behavior. Suppose X is a random variable and the random distribution of $X$is $P(X)$; then, the entropy of $X$is

$$H(X) = -\sum_{x \in X} P(x) \times \log P(x). \tag{3}$$

Reference literature [31], according to the user's historical access records, respectively, defines the probability of the user choosing the work objective o_k and accessing the medical record m_l stage and, then, defines the information entropy of the user's selection of the work objective stage and access to the medical record.

Definition 1. Probability that user u_i selects work objective o_k when diagnosing patient s_j.

$$P\left(o_k \mid u_i, s_j\right) = \frac{\|f(o_k)\|}{\sum_{o_k \in O_{u_i|s_j}} \|f(o_k)\|}, \tag{4}$$

where $O_{u_i||s_j}$ represents the set of work objective that user u_i accesses when treating patients$_j$ and $\|f(o_k)\|$ represents the number of times the user selects work objective o_k.

Definition 2. Probability that user u_i selects medical record m_l under job objective o_k.

$$P\left(m_l \mid u_i, s_j, o_k\right) = \frac{\|f(m_l)\|}{\sum_{m_l \in M_{s_j|o_k}} \|f(m_l)\|}, \tag{5}$$

where $M_{s_j|o_k}$ represents the set of medical records accessed by the user when patient $s_j$and work objective are determined and $\|f(m_l)\|$ represents the number of times the user accesses medical records m_l.

Definition 3. Entropy for choosing work objectives (EFCWO) when user u_i diagnoses patient s_j.

$$H_{s_j}^{o}(u_i) = -\sum_{k=1}^{I_o} P\left(o_k \mid u_i, s_j\right) \times \log P\left(o_k \mid u_i, s_j\right). \tag{6}$$

Definition 4. Entropy of access to medical records (EATMR) when user u_i is under job objective o_k.

$$H_{o_k}^{m}(u_i) = -\sum_{l=1}^{I_m} P\left(m_l \mid u_i, s_j, o_k\right) \times \log P\left(m_l \mid u_i, s_j, o_k\right). \tag{7}$$

(2) Quantification of User Trust. Based on the existing research, this paper mainly divides it into direct trust and recommended trust according to the way of obtaining trust. The following first introduces the related concepts and definitions.

Definition 5 (trust). Trust refers to the dependency relationship between entities. In spite of believing that the other party is trustworthy and upright, trust also has certain risks because trusting the other party means bearing the losses caused and hurt by the other party's behavior.

Definition 6 (trust value). Trust is an evaluation between entities, which itself has a certain degree of uncertainty and ambiguity, and trust value is a quantification of this uncertainty and is expressed by Td.

$$\mathrm{Td}_{ij} = \begin{cases} \mathrm{DT}\left(u_i, u_j\right), \\ \mathrm{RT}\left(u_i, u_j\right). \end{cases} \tag{8}$$

Among them, $\mathrm{Td}_{ij} = \mathrm{DT}(u_i, u_j)$represents direct trust, $\mathrm{Td}_{ij} = \mathrm{RT}(u_i, u_j)$ represents recommendation trust, and u_i and u_j represent two entities.

Definition 7 (trust matrix). It refers to a matrix composed of the trust between entities in a specific context, denoted by M.

$$M = \begin{bmatrix} \mathrm{Td}_{11} & \cdots & \mathrm{Td}_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{Td}_{m1} & \cdots & \mathrm{Td}_{mn} \end{bmatrix}, \tag{9}$$

where the element Td_{ij} represents the trust degree of entity u_i to entity u_j and the diagonal elements are all 1.

According to the relevant definitions, we will evaluate the user's trust from two aspects: direct trust and recommended trust.

(a) *Direct Trust.* When evaluating the trust degree of the user u_j, if the evaluation result is the direct experience from the user u_i, the relationship between u_i and u_j is called a direct trust relationship. Assuming that, during the user's historical interaction, the number of successful interactions between user u_i and user u_j is m and the number of interaction failures is n; then, the direct trust relationship function between user u_i and user u_j is defined as

$$\mathrm{DT}\left(u_i, u_j\right) = \left(\frac{m}{m+n}\right)\delta^{(1/m+1)}, \tag{10}$$

in which $\delta\,(0 < \delta < 1)$ increases with the number of successful interactions. The premise of ensuring the validity of the relationship function is to have sufficient historical data; that is, the number of interactions between two users must be sufficient. If the number of interactions is very small, the accuracy of the results will be affected. In order to solve this problem, reference literature [25] in this paper introduces the interaction threshold π. When the number of interactions is less than the threshold π, the abovementioned formula is adjusted as follows:

$$\mathrm{DT}\left(u_i, u_j\right) = \left(0.5 + \frac{m-n}{2\pi}\right)\delta^{(1/m+1)}. \tag{11}$$

Therefore, the direct trust relationship between end user u_i and user u_j can be expressed by the following relationship function:

$$\mathrm{DT}\left(u_i, u_j\right) = \begin{cases} \left(\dfrac{m}{m+n}\right)\delta^{(1/m+1)}, & m+n > \pi, \\ \left(0.5 + \dfrac{m-n}{2\pi}\right)\delta^{(1/m+1)}, & m+n \leq \pi. \end{cases} \tag{12}$$

(b) *Recommendation Trust.* The key to recommending trust different from direct trust is that there is no direct empirical relationship between the trustee u_i and the client u_j, but the trust relationship is established indirectly through the introduction of acquaintances. When there is no direct interaction experience between u_i and u_j, or the interaction experience is very limited, in order to be able to objectively evaluate the trust of u_j, u_j can establish an indirect trust relationship with u_i through the introduction of acquaintance u_k. As shown in Figure 3, u_i and u_k are direct trust relationships, u_k and u_j are also direct trust relationships, but u_i and u_j are recommended trust relationships established through u_k.

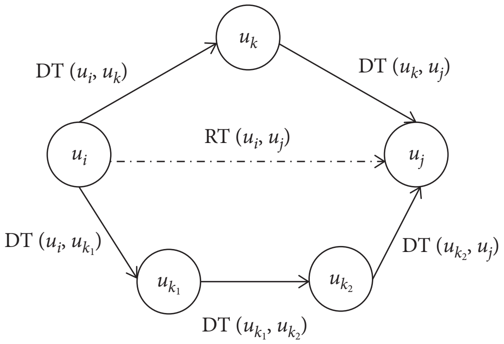


Figure 3: Recommended trust relationship diagram.

As shown in Figure 3, there are two indirect recommended paths between users u_i and u_j: $u_i \longrightarrow u_{k_1} \longrightarrow u_{k_2} \longrightarrow u_j$ and $u_i \longrightarrow u_k \longrightarrow u_j$; each path corresponds to a trust value, and the greater the path depth, the lower the trust between entities. Therefore, it is necessary to comprehensively calculate all reachable paths to obtain the final trust degree between u_i and u_j.

Assuming that the path depth is $\omega\,(\omega \geq 2)$ and the reachable path is γ, the corresponding recommendation trust degree has the following definition [45]:

$$\mathrm{RT}^{\omega}\left(u_i, u_j\right) = \frac{\sum_{l=1}^{\gamma} \mathrm{DT}\left(u_i, u_{k_1}\right) \times \mathrm{DT}\left(u_{k_1}, u_{k_2}\right) \times \cdots \times \mathrm{DT}\left(u_{k_{\omega-1}}, u_j\right)}{\gamma}. \tag{13}$$

Finally, the comprehensive recommended trust degree $\mathrm{RT}\,(u_i, u_j)$ between u_i and u_j is calculated for all possible path depths and corresponding reachable paths.

$$\mathrm{RT}\left(u_i, u_j\right) = \sum_{\omega=\min_{\mathrm{dph}}}^{\max_{\mathrm{dph}}} \alpha_\omega \times \mathrm{RT}^{\omega}\left(u_i, u_j\right),$$
$$\alpha_\omega = \frac{1 - \left(\omega / \sum_{\omega=\min_{\mathrm{dph}}}^{\max_{\mathrm{dph}}} \omega\right)}{\sum_{\omega=\min_{\mathrm{dph}}}^{\max_{\mathrm{dph}}} \left(1 - \left(\omega / \sum_{\omega=\min_{\mathrm{dph}}}^{\max_{\mathrm{dph}}} \omega\right)\right)}, \tag{14}$$

where $\min_{\mathrm{dph}}$ and $\max_{\mathrm{dph}}$, respectively, represent the minimum path depth and the maximum path depth and α_ω indicates the weight of the corresponding recommendation trust when the path depth is ω and satisfies $\alpha_\omega \in (0, 1), \sum_{\omega=\min_{\mathrm{dph}}}^{\max_{\mathrm{dph}}} \alpha_\omega = 1$.

(c) *Comprehensive Trust.* The comprehensive trust degree CT is the result of combining the direct trust degree and the recommended trust degree in a certain way. This article uses the following expression to express it:

$$\mathrm{CT}\left(u_i, u_j\right) = \alpha \mathrm{DT}\left(u_i, u_j\right) + (1 - \alpha)\mathrm{RT}\left(u_i, u_j\right). \tag{15}$$

Among them, $\alpha\,(0 < \alpha < 1)$ represents the proportion of direct trust $\mathrm{DT}\,(u_i, u_j)$ in the comprehensive trust. At present, there is no unified standard for the value of α, which is generally subjectively determined based on expert experience.

3.2.2. Risk Quantification Method Based on Adaptive Neural Fuzzy Theory. Quantifying the risk of medical big data privacy leakage is a very complicated process because the access behavior and trust of users at each stage are mutually independent and interrelated, and different indicators have different effects on the final risk, which is a nonlinear changing relationship. This article establishes a risk quantification method based on adaptive neuro-fuzzy theory, aiming at solving some problems existing in existing methods, providing a risk assessment model and method specifically for medical big data information security, achieving academic innovation, and providing reference for relevant institutions.

The adaptive neural fuzzy theory mainly uses the self-learning ability of the neural network to learn the existing data, automatically generates the rule base and membership function in the fuzzy system, and does not rely on subjective factors such as expert experience. Matlab provides an adaptive neuro-fuzzy inference system based on the Takagi–Sugeno model [46]. As shown in Figure 4, the quantification process of medical big data privacy leakage risk based on this inference system can be roughly divided into the following four steps:

Step 1: the user's access behavior data and trust data are preprocessed, and the processed data are loaded into the Matlab workshop

Step 2: fuzzy C-means clustering or subtractive clustering is used to process the input data to generate the initial fuzzy inference system (FIS)

Step 3: on the basis of the initial FIS, the adaptive neural fuzzy inference system trains the inference system according to the existing data, so as to correct and adjust the parameters of each membership function and output function and generate the final FIS

Step 4: according to the final training results, the user's access behavior, trust, and final risk membership function and rule base are recorded

4. Simulation Experiment

4.1. Experimental Environment. The experimental part mainly uses Matlab software to model and analyze the design of the network structure and the specific processing of the data in this paper. Then, the performance of the model is tested and the configuration of the specific experimental environment is shown in Table 1.

4.2. Experimental Data. It is known from the foregoing that, before training a fuzzy neural network, not only an input data set but also a corresponding output data set should be obtained. Therefore, this article not only obtains the following input data before the experimental test, entropy for choosing work objectives (EFCWO), entropy of access to medical records (EATMR), and the user's trust (UT), but also the output data, risk (Risk).

At present, we have obtained part of the user information form, the doctor's advice, and the user's access record form from a hospital. The user information table mainly includes fields such as the user's ID, department, and title; the medical order is what we usually call the electronic medical record, which mainly records the patient's medical records, medical plans, and other information; the user's access record is mainly extracted from the user's access log, which records the computer model, login time, user access information, and user's operation. In addition, for partial missing fields, we assume that there are appropriate software components that can automatically obtain information from the system, such as the number of successful and failures interactions between users and the interaction relationship and dynamically adjust these factors when the context changes. The initial value of δ is set as 0.5, the interaction threshold as $\pi = 50$, and the weight of the direct trust as $\alpha = 0.6$. Formulas (5), (6), and (14) are combined to simulate the generation of the user's EFCWO, EATMR, and UT.

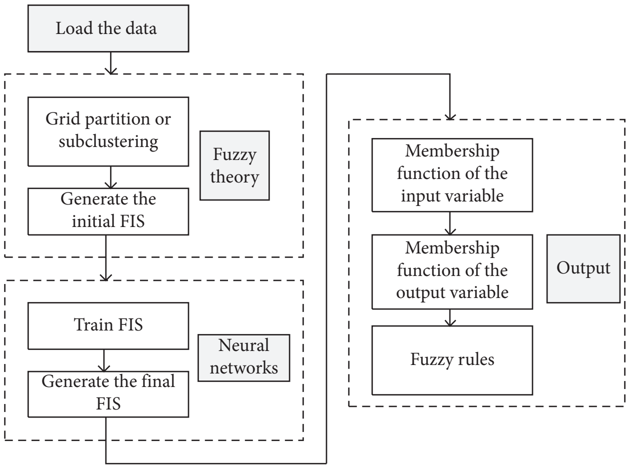

Load the data
Grid partition or subclustering
Generate the initial FIS
Fuzzy theory
Train FIS
Generate the final FIS
Neural networks
Membership function of the input variable
Membership function of the output variable
Fuzzy rules
Output

Figure 4: ANFIS data processing process based on the Takagi–Sugeno model.

Table 1: Configuration of the experimental environment.

Software/hardware	Version/model
Processor	Intel(R) core(TM) m3-7Y30
Frequency	1.00 GHz
RAM	4 GB
SSD	256 GB
Operating system	Windows10
Programming software	MATLAB2016a

The output data are calculated through the risk index; in literature [25, 28, 31, 34–36], the related risk quantitative method is introduced, and this paper is based on the existing research by cross entropy to measure individual user's access behavior deviating from all user access behavior of entropy to calculate the risk. Assume that the trust of a user is $\phi(u)$, the risk caused by choosing the work objective is risk_1, and the risk caused by accessing medical records is risk_2; then, the risk calculation formula introduced in literature [25] $(\text{Risk} = \min\{1, (W_1^* \text{Risk}_1 + W_2^* \text{Risk}_2 + W_3^* (1 - \phi(u)))\})$ can simulate and generate the corresponding output data set.

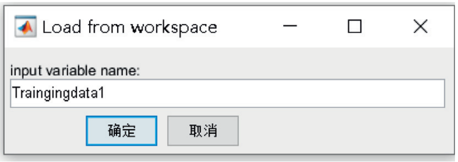
In summary, 1500 pieces of data were generated for simulation experiments in this paper. In order to avoid omission of data and reduce the chance of test results, this paper uses 10-fold cross validation to test the accuracy of the model. The data set is divided into ten equal parts. First, the first data set for testing data is used, then the remaining 9 parts for training data are used, the second data set was used for testing, and the remaining 9 parts are used for training, and ten verifications are run in turn. In this paper, the first group of data (training data1, testing data1) is used to introduce the whole experimental process in detail.

4.3. Experimental Process. As shown in Figure 4, the risk quantization method based on adaptive neural fuzzy theory is roughly divided into five steps: loading data, generating initial FIS, training FIS, generating final FIS, and outputting results. This section will follow these five steps to conduct the following operations and display the results.

4.3.1. Load Data. First, we need to load the training data into the workspace to form a multi-input single-output data matrix, where the last column defaults to output data because only single-output data formats are supported in the Takagi–Sugeno model-based fuzzy inference system. Through the graphical interface window shown in Figure 5, the training data1 is loaded into the workshop of Matlab, and the final data distribution is shown in Figure 6.

4.3.2. Generating the Initial FIS. Before training FIS, an initial FIS structure is required. In this paper, fuzzy C-means clustering is used to extract the features of the input data to generate the initial FIS. The output of the clustering method represents the membership degree of each data point to each cluster center. Through constant correction of the clustering center point, until the weighted sum of the distance from each data point to the clustering center and the membership degree is the smallest, the output result can be further used to establish the fuzzy inference system. However, the subtractive clustering method is mainly used to estimate the number of data clusters and the location of the cluster center, so the fuzzy C-means clustering algorithm is selected. Among them, the fuzzy subsets of the input variables EFCWO, EATMR, and UT are all divided into 4 categories, which are very low (VL), low (L), medium (M), and high (H). At the same time, according to the distribution characteristics of the user's risk indicators, most users' EFCWO, EATMR, and UT are distributed near the mean, and the number of users with very low or very high values of the three indicator variables accounts for only a few parts. The general trend is consistent with the characteristics of the Gaussian distribution. Therefore, the input variable of type selects Gaussian membership functions, and the type of the output variable can only be constant or linear combination of the input variables. The resulting neural network structure and membership function corresponding to each index before system training are shown in Figures 7–10.

4.3.3. Training the Initial FIS and Generating the Final FIS. Based on the initial FIS structure, the neural fuzzy inference system is trained by data loaded into the workspace. But, before training, we need to determine the training method, error accuracy, and training times. The training method mainly includes a hybrid algorithm and BP algorithm. In this paper, the hybrid algorithm is used to train the FIS, the error accuracy is set to 1e-5, and the training number is set to 20 times. As shown in Figure 11–13, after training, the range of fuzzy subsets of input variables and the membership function shape of each index have changed, but the general trend still conforms to the Gaussian distribution. In addition, the ANFIS model structure will not change after training, only some structural parameters will change.


Load from workspace

input variable name:

Traingingdata1

Figure 5: Dialog for loading data.

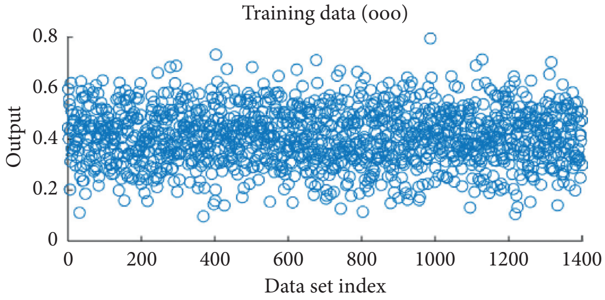


Figure 6: Distribution of training data after loading.

4.3.4. Output Results. This article combines fuzzy theory and neural networks to utilize the self-learning ability of neural networks to adaptively train membership functions and rule bases in fuzzy inference systems. At the same time, the fuzzy theory is used to present the relationship between the input and output of the neural network learning data in a way that people can easily accept. Therefore, this section mainly presents the results of the training in Section 4.3.3 in a formal way.

(1) Membership function of input variables: the membership functions after the training of input variables are given, as shown in Figures 11–13, and all conform to the Gaussian distribution. From this, the parameters of each membership function can be obtained, as shown in Table 2.

The formula of the Gaussian membership function is known as $F(x, \sigma, c) = \exp(-((x - c)^2/2\sigma^2))$; the parameters in Table 1 are brought into it, and the membership function expression of the input variables can be obtained such as $F_{\mathrm{VL}}(\mathrm{EFCWO}) = \exp(-((x - 0.1026)^2/2 \times (0.1263)^2))$.

(2) Membership function of output variables: it is known that there are 3 input variables, and each input variable corresponds to 4 fuzzy subsets. Therefore, the result of all input combinations will produce 64 output records. As shown in Figure 14, the output variable $u$corresponds to 64 membership functions. Among them, the membership function parameter corresponding to $u1$after training in the neural network is [0.03051 0.01467 0.01683 0.1191]; then, the function expression corresponding to the output function u1 is $u1 = 0.03051 * \mathrm{EFCWO} + \quad 0.01467 * \mathrm{EATMR} + 0.01683 * \mathrm{UT} + 0.1191$, and so, all the output functions can be obtained. For convenience, only the parameters corresponding to each output function are listed here, as shown in Table 3.

(3) Rule base: according to the input and output membership function of each indicator, the rule base shown in Table 4 is easy to obtain. For convenience of writing, a, b, and c are used instead of the indicators EFCWO, EATMR, and

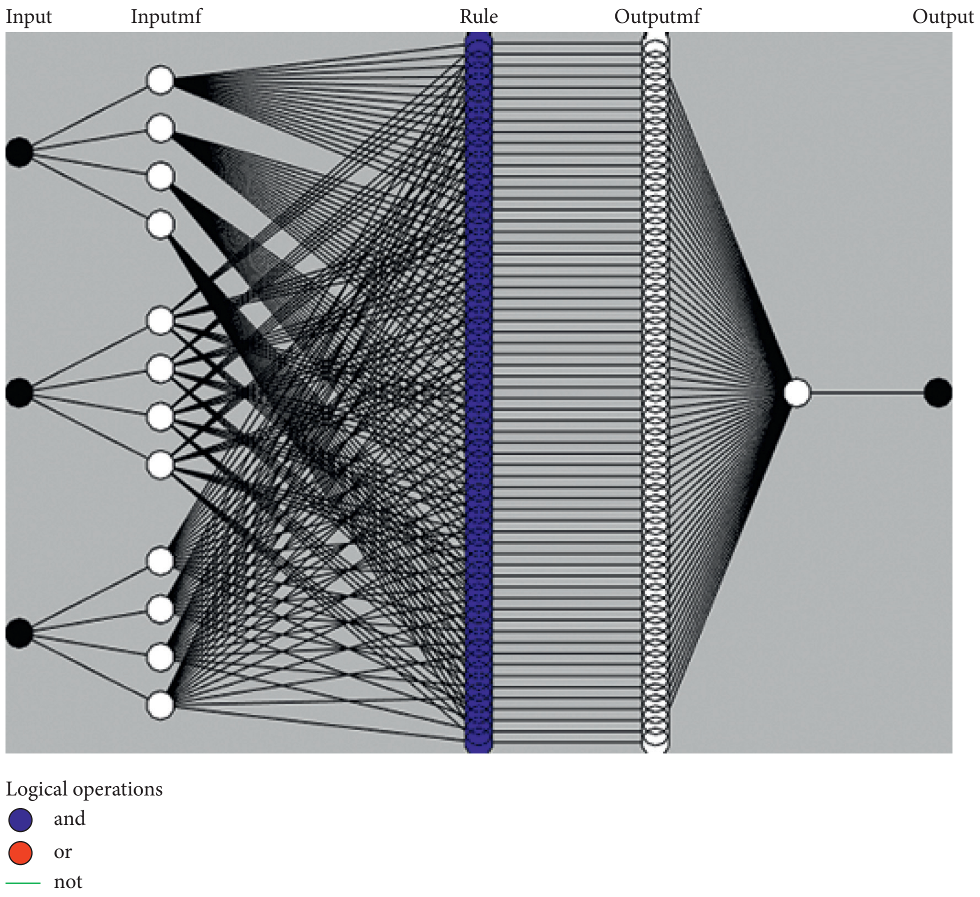


Figure 7: ANFIS model structure diagram.

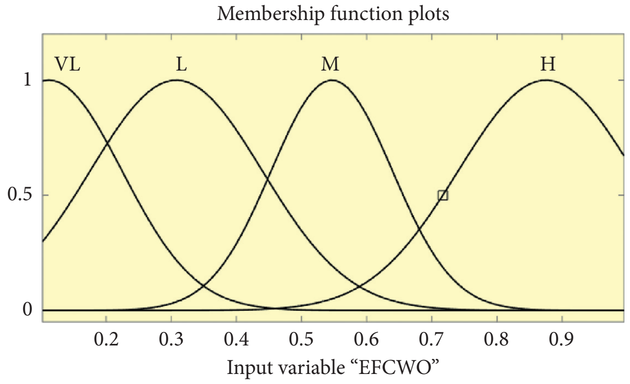


Figure 8: Initial EFCWO membership function distribution.

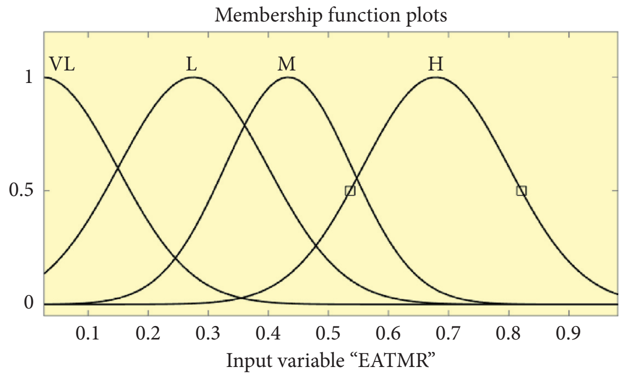


Figure 9: Initial distribution of the EATMR membership function.

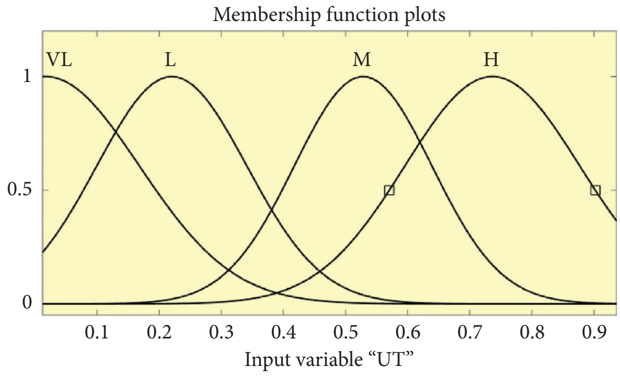


Figure 10: Initial UT membership function distribution.

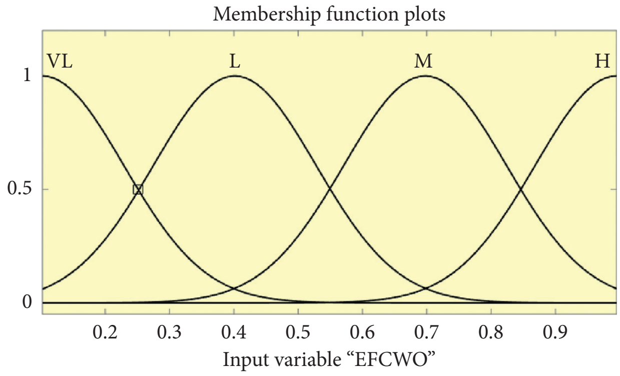


Figure 11: Distribution of the membership function of EFCWO after training.

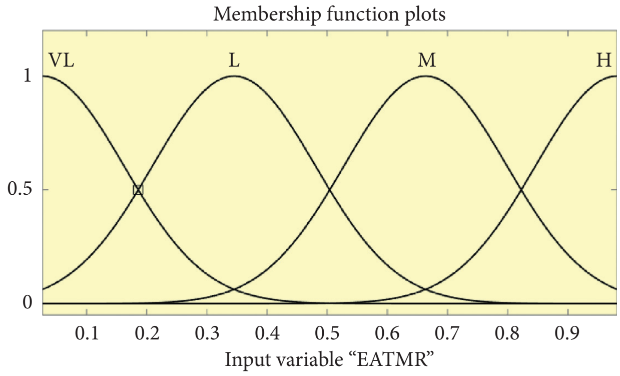


Figure 12: Distribution of the membership function of EATMR after training.

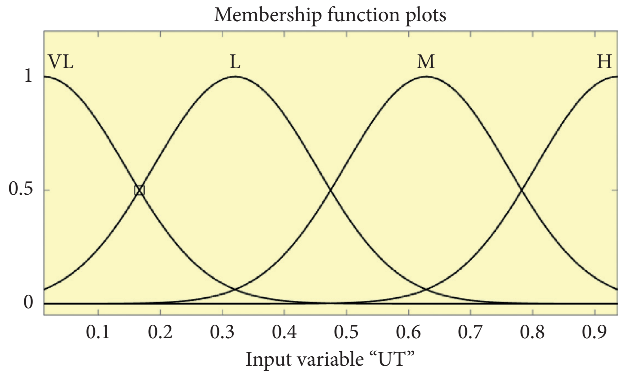


Figure 13: Distribution of the membership function of UT after training.

Table 2: Parameters of the membership function for each input variable.

Input	Category	Symbol	Parameters
EFCWO	Very Low	VL	[0.1263, 0.1026]
	Low	L	[0.1301, 0.4]
	Middle	M	[0.1306, 0.6974]
	High	H	[0.1361, 0.9948]
EATMR	Very Low	VL	[0.1352, 0.02671]
	Low	L	[0.1355, 0.345]
	Middle	M	[0.1372, 0.6633]
	High	H	[0.1379, 0.9816]
UT	Very Low	VL	[0.1308, 0.01257]
	Low	L	[0.1308, 0.3205]
	Middle	M	[0.1312, 0.6284]
	High	H	[0.1317, 0.9363]

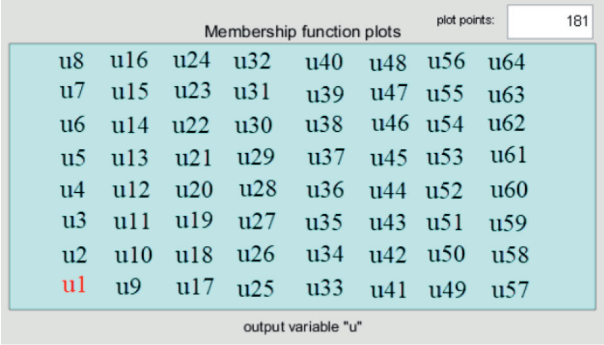

Membership function plots
plot points:
181
u8 u16 u24 u32 u40 u48 u56 u64
u7 u15 u23 u31 u39 u47 u55 u63
u6 u14 u22 u30 u38 u46 u54 u62
u5 u13 u21 u29 u37 u45 u53 u61
u4 u12 u20 u28 u36 u44 u52 u60
u3 u11 u19 u27 u35 u43 u51 u59
u2 u10 u18 u26 u34 u42 u50 u58
u1 u9 u17 u25 u33 u41 u49 u57
output variable "u"

Figure 14: Output function after training.

UT. At the same time, in order to more intuitively represent the impact of changes in each indicator variable on the final risk, this article analyzes the influence of the other two indicators on the final risk by fixing one of the indicator variables and generates a three-dimensional perspective, as shown in Figures 15–17.

When a user requests access to medical information, the system firstly fuzzy processes the indicators EFCWO, EATMR, and UT to calculate the membership of each indicator. Then, all possible output results are listed according to the rule base, and finally, rule aggregation and defuzzification are performed to obtain the final risk value. Assuming that the corresponding three index values of a user are [0.549, 0.504, 0.474], the final output risk value is 0.432, as shown in Figure 18. If the risk value is within the range that the system can tolerate, the system will allow the user to access; otherwise, it will deny its access or pass a risk reduction policy until it meets the system tolerable range.

Table 3: Parameters corresponding to each output function.

Output	Parameters
u_1	[0.03051 0.01467 0.01683 0.12]
u_2	[0.3774 0.1372–0.01363 0.02312]
u_3	[0.443 0.4179–0.1277 -0.01897]
u_4	[-0.03896 -0.0089 -0.03655 -0.03707]
u_5	[0.485 0.2496–0.05756 0.07634]
u_6	[0.5079 0.464–0.1796 0.006043]
u_7	[0.5119 0.4787–0.1898 -0.001796]
u_8	[0.2428 0.2227–0.2206 0.1825]
u_9	[0.2414 0.2065–0.3251 0.2686]
u_{10}	[0.5012 0.4815–0.1948 0.01095]
u_{11}	[0.5215 0.4604–0.1868 0.01171]
u_{12}	[0.467 0.3887–0.06568 -0.02206]
u_{13}	[0.04532 0.1187 0.01707 0.1451]
u_{14}	[0.2684 0.3431–0.1505 0.1398]
u_{15}	[0.1355 0.2624 0.0368 0.1517]
u_{16}	[0.004216 0.04936 0.09015 0.09087]
u_{17}	[0.2995 0.4212–0.1542 0.08838]
u_{18}	[0.489 0.5101–0.1965 0.002561]
u_{19}	[0.4994 0.5166–0.1884 -0.00816]
u_{20}	[0.2702 0.3696–0.1533 0.07767]
u_{21}	[0.5031 0.482–0.2098 0.006902]
u_{22}	[0.5007 0.5–0.2028 0.0005266]
u_{23}	[0.5002 0.5053–0.203 -0.0004162]
u_{24}	[0.482 0.4641–0.2138 0.03419]
u_{25}	[0.4556 0.4782–0.1971 0.03202]
u_{26}	[0.5022 0.4975–0.2002 0.0005379]
u_{27}	[0.5049 0.5074–0.2001 -0.006636]
u_{28}	[0.4858 0.4522–0.2018 0.03658]
u_{29}	[0.0756 0.2806–0.06298 0.3443]
u_{30}	[0.3788 0.4849–0.2036 0.06709]
u_{31}	[0.4328 0.5311–0.1983 0.001309]
u_{32}	[0.2548 0.3029–0.1321 0.2265]
u_{33}	[0.3366 0.5211–0.1764 0.1085]
u_{34}	[0.494 0.4967–0.1854 -0.0004179]
u_{35}	[0.4941 0.5023–0.1875 -0.004024]
u_{36}	[0.3509 0.3769–0.1338 0.06765]
u_{37}	[0.5019 0.5005–0.1969 -0.002493]
u_{38}	[0.5009 0.5012–0.1993 -0.0011]
u_{39}	[0.4995 0.5–0.199 -0.0003297]
u_{40}	[0.4913 0.4742–0.1929 0.01018]
u_{41}	[0.4564 0.4871–0.2081 0.03637]
u_{42}	[0.502 0.5049–0.2009 -0.003726]
u_{43}	[0.5042 0.5–0.2006 -0.002425]
u_{44}	[0.4889 0.4682–0.2055 0.03072]
u_{45}	[0.3961 0.3408 0.03439 0.2079]
u_{46}	[0.3897 0.5465–0.1283 0.007818]
u_{47}	[0.4311 0.5097–0.1533 0.01045]
u_{48}	[0.2762 0.249 0.0003293 0.1953
u_{49}	[0.2775 0.01391 0.002306 0.2623]
u_{50}	[0.4887 0.4925–0.1719 0.000453]
u_{51}	[0.4912 0.4522–0.1474 -0.01305]
u_{52}	[0.07887 0.01718 0.1357 0.1378]
u_{53}	[0.4754 0.3644–0.2464 0.07633]
u_{54}	[0.5056 0.492–0.2099 0.001454]
u_{55}	[0.4954 0.4885–0.2105 0.01425]
u_{56}	[0.5082 0.2743–0.2392 0.1156]
u_{57}	[0.3531 0.364–0.1186 0.2152]
u_{58}	[0.5056 0.4799–0.1823 -0.0004327]
u_{59}	[0.5099 0.4738–0.1813 -0.005195]
u_{60}	[0.4547 0.3052–0.1262 0.1013]
u_{61}	[0.2633 0.2822 0.07883 0.3343]
u_{62}	[0.2627 0.3892–0.02458 0.2768]
u_{63}	[0.2915 0.2224–0.001363 0.2944]
u_{64}	[0.2022 0.1827 0.2395 0.268]

4.4. Performance Analysis. Firstly, the overall effect of the model is evaluated using testing data 1, and the degree of agreement between the output of the model and the actual output is analyzed through comparative experiments. Figure 19 shows a partial screenshot of the experimental work area in this paper, where the variable ANFIS is the Adaptive Neuro-Fuzzy Inference System after training. From the foregoing, the system has three input variables and one output variable. Therefore, the input variable of testing data 1 is named testing data 1_input and the output variable is named testing data 1_output to generate 150 ∗ 3 and 150 ∗ 1 data structures, respectively.

Then, the comparative analysis results, as shown in Figure 20, can be achieved through the following code:

```
x = (1 : 1 : 150);


y = evalfis (Testingdata1_input, ANFIS); y1 = plot (x,
Testingdata1_output, "or")
hold on;
y2 = plot(x, y, "+k");
legend([y1, y2], "Actual output,""ANFIS output")
```

From the comparison results in Figure 20, it can be clearly seen that the ANFIS output results after training are basically consistent with the actual output results. There is no very obvious error, and the sum of error squares is $7.53521e-06$. The same method was used to perform the remaining nine experiments in sequence, and the results are shown in Table 5.

According to Table 5, the final error value of 10-fold cross validation is $7.0159e-6$, which is less than $1e-5$. Therefore, the model in this paper is feasible in predicting the risk of medical big data privacy disclosure.

Next, the accuracy rate, recall rate, and F1 value of the model will be specifically analyzed under the conditions of different proportions of illegal users. In order to facilitate comparative analysis, this article will refer to the experimental methods of Hui et al. [31] and Wang and Hong [28] to evaluate the performance of the medical big data security and privacy protection model based on risk access control proposed in this paper. It is known from the foregoing that illegal users have more diversity and instability in selecting work goals and accessing medical records, and their risk value is higher than that of legitimate users. Therefore, when the users with higher risk values are illegal users and the risk values of illegal users are higher, the model is considered to be effective. This article will set up 6 groups of experiments, each of which will generate 600 users' EFCWO, EATMR, and UT values and, then, calculate their

Table 4: Rule base.

	Fuzzy rules
1	If (a is VL) and (b is VL) and (c is VL) then (risk is $u1$)
2	If (a is VL) and (b is VL) and (c is L) then (risk is $u2$)
3	If (a is VL) and (b is VL) and (c is M) then (risk is $u3$)
4	If (a is VL) and (b is VL) and (c is H) then (risk is $u4$)
5	If (a is VL) and (b is L) and (c is VL) then (risk is $u5$)
6	If (a is VL) and (b is L) and (c is L) then (risk is $u6$)
7	If (a is VL) and (b is L) and (c is M) then (risk is $u7$)
8	If (a is VL) and (b is L) and (c is H) then (risk is $u8$)
9	If (a is VL) and (b is M) and (c is VL) then (risk is $u9$)
10	If (a is VL) and (b is M) and (c is L) then (risk is $u10$)
11	If (a is VL) and (b is M) and (c is M) then (risk is $u11$)
12	If (a is VL) and (b is M) and (c is H) then (risk is $u12$)
13	If (a is VL) and (b is H) and (c is VL) then (risk is $u13$)
14	If (a is VL) and (b is H) and (c is L) then (risk is $u14$)
15	If (a is VL) and (b is H) and (c is M) then (risk is $u15$)
16	If (a is VL) and (b is H) and (c is H) then (risk is $u16$)
17	If (a is L) and (b is VL) and (c is VL) then (risk is $u17$)
18	If (a is L) and (b is VL) and (c is L) then (risk is $u18$)
19	If (a is L) and (b is VL) and (c is M) then (risk is $u19$)
20	If (a is L) and (b is VL) and (c is H) then (risk is $u20$)
21	If (a is L) and (b is L) and (c is VL) then (risk is $u21$)
22	If (a is L) and (b is L) and (c is L) then (risk is $u22$)
23	If (a is L) and (b is L) and (c is M) then (risk is $u23$)
24	If (a is L) and (b is L) and (c is H) then (risk is $u24$)
25	If (a is L) and (b is M) and (c is VL) then (risk is $u25$)
26	If (a is L) and (b is M) and (c is L) then (risk is $u26$)
27	If (a is L) and (b is M) and (c is M) then (risk is $u27$)
28	If (a is L) and (b is M) and (c is H) then (risk is $u28$)
29	If (a is L) and (b is H) and (c is VL) then (risk is $u29$)
30	If (a is L) and (b is H) and (c is L) then (risk is $u30$)
31	If (a is L) and (b is H) and (c is M) then (risk is $u31$)
32	If (a is L) and (b is H) and (c is H) then (risk is $u32$)
33	If (a is M) and (b is VL) and (c is VL) then (risk is $u33$)
34	If (a is M) and (b is VL) and (c is L) then (risk is $u34$)
35	If (a is M) and (b is VL) and (c is M) then (risk is $u35$)
36	If (a is M) and (b is VL) and (c is H) then (risk is $u36$)
37	If (a is M) and (b is L) and (c is VL) then (risk is $u37$)
38	If (a is M) and (b is L) and (c is L) then (risk is $u38$)
39	If (a is M) and (b is L) and (c is M) then (risk is $u39$)
40	If (a is M) and (b is L) and (c is H) then (risk is $u40$)
41	If (a is M) and (b is M) and (c is VL) then (risk is $u41$)
42	If (a is M) and (b is M) and (c is L) then (risk is $u42$)
43	If (a is M) and (b is M) and (c is M) then (risk is $u43$)
44	If (a is M) and (b is M) and (c is H) then (risk is $u44$)
45	If (a is M) and (b is H) and (c is VL) then (risk is $u45$)
46	If (a is M) and (b is H) and (c is L) then (risk is $u46$)
47	If (a is M) and (b is H) and (c is M) then (risk is $u47$)
48	If (a is M) and (b is H) and (c is H) then (risk is $u48$)
49	If (a is H) and (b is VL) and (c is VL) then (risk is $u49$)
50	If (a is H) and (b is VL) and (c is L) then (risk is $u50$)
51	If (a is H) and (b is VL) and (c is M) then (risk is $u51$)
52	If (a is H) and (b is VL) and (c is H) then (risk is $u52$)
53	If (a is H) and (b is L) and (c is VL) then (risk is $u53$)
54	If (a is H) and (b is L) and (c is L) then (risk is $u54$)
55	If (a is H) and (b is L) and (c is M) then (risk is $u55$)
56	If (a is H) and (b is L) and (c is H) then (risk is $u56$)
57	If (a is H) and (b is M) and (c is VL) then (risk is $u57$)
58	If (a is H) and (b is M) and (c is L) then (risk is $u58$)
59	If (a is H) and (b is M) and (c is M) then (risk is $u59$)
60	If (a is H) and (b is M) and (c is H) then (risk is $u60$)
61	If (a is H) and (b is H) and (c is VL) then (risk is $u61$)
62	If (a is H) and (b is H) and (c is L) then (risk is $u62$)
63	If (a is H) and (b is H) and (c is M) then (risk is $u63$)
64	If (a is H) and (b is H) and (c is H) then (risk is $u64$)

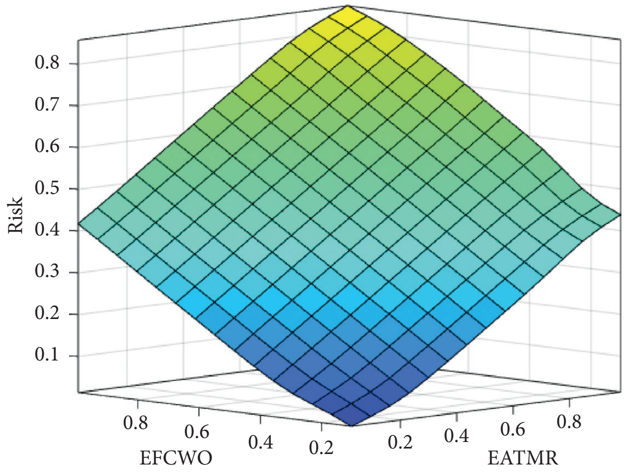


Figure 15: Rules interface between EFCWO, EATMR, and risk.

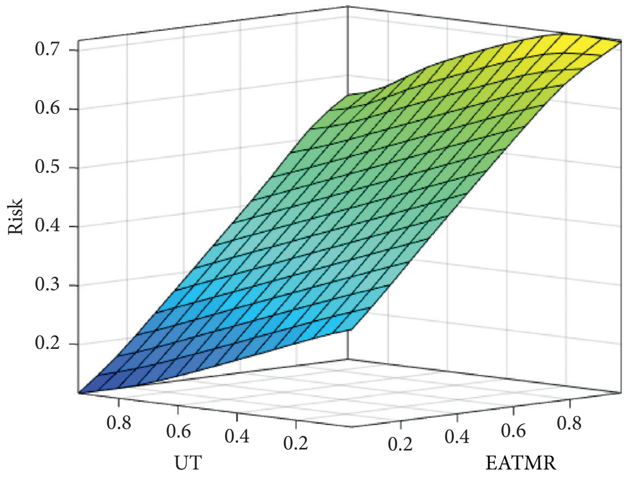


Figure 16: Rules interface between UT, EATMR, and risk.

corresponding risk values through the risk quantification model based on the adaptive neural fuzzy theory. The number of curious doctors accounts for 2.5%, 5%, 7.5%, 10%, 12.5%, and 15%, then each group of data is sorted according to the magnitude of the risk value from high to low, and finally, the experimental results are calculated, as shown in Table 6.

The experimental results show that the performance of the model in this paper improves with the increase of the number of illegal users. This is because when the number of users is constant, the more the number of illegal users, the less the number of legal users and the larger the

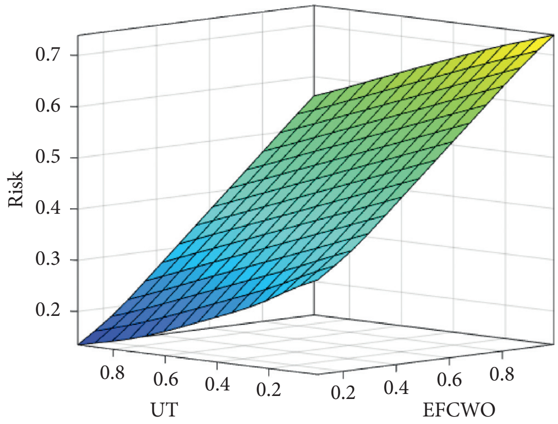
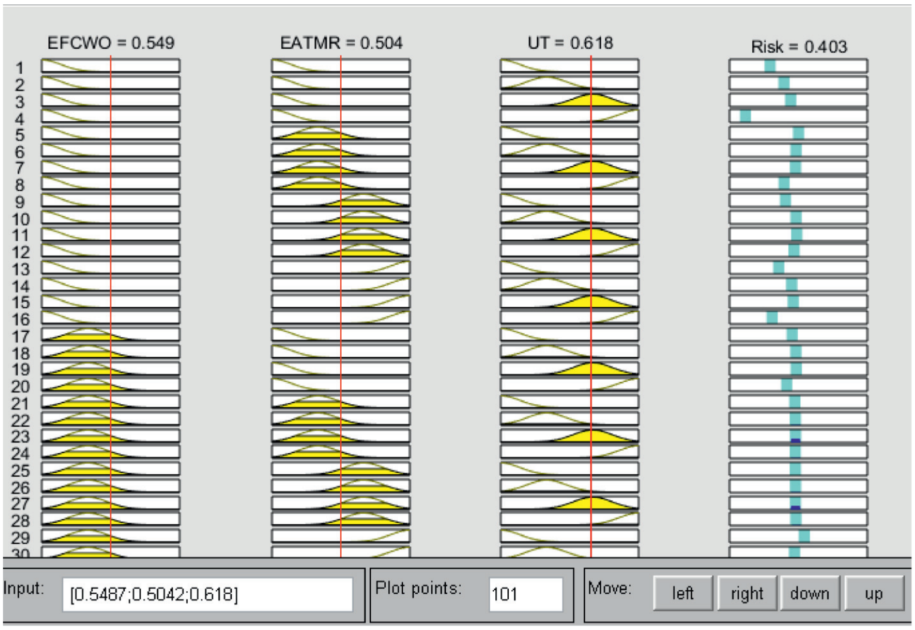


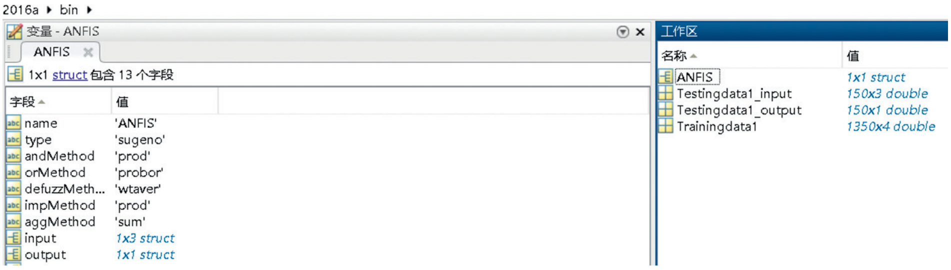
Figure 17: Rules interface between UT, EFCWO, and risk.



Input: [0.5487;0.5042;0.618]

Plot points: 101

Figure 18: Risk output result.



1x1 struct 包含 13 个字段

字段	值
name	'ANFIS'
type	'sugeno'
andMethod	'prod'
orMethod	'probor'
defuzzMeth...	'wtaver'
impMethod	'prod'
aggMethod	'sum'
input	1x3 struct
output	1x1 struct

名称	值
ANFIS	1x1 struct
Testingdata1_input	150x3 double
Testingdata1_output	150x1 double
Trainingdata1	1350x4 double

Figure 19: Partial screenshot of the experimental workspace.

proportion of illegal users among high-risk users. In addition, through a comparative analysis with the methods of Hui et al. [31] and Wang and Hong [28], it is found that when the number of illegal users reaches 15%, the performance of the model does not change significantly, but when the number of illegal users is less than 15%, the performance of the model is significantly superior to the methods of Hui et al. [31] and Wang et al. [28] because this

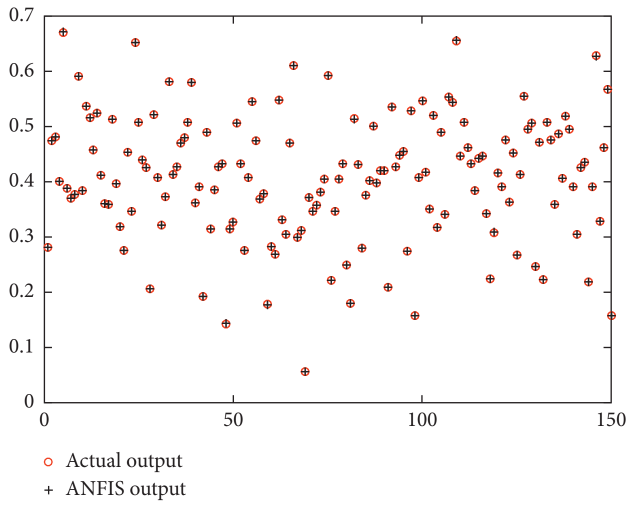


FIGURE 20: Feasibility analysis results of the model.

TABLE 5: Result of 10-fold cross validation.

Number	Training data set	Testing data set	Sum of squared errors	The average
1	Training data1	Testing data1	7.53521*e* – 06	7.0159*e* – 6
2	Training data2	Testing data2	7.12167*e* – 06	
3	Training data3	Testing data3	6.59361*e* – 06	
4	Training data4	Testing data4	6.23479*e* – 06	
5	Training data5	Testing data5	7.98632*e* – 06	
6	Training data6	Testing data6	6.63364*e* – 06	
7	Training data7	Testing data7	7. 38572*e* – 06	
8	Training data8	Testing data8	6.72831*e* – 06	
9	Training data9	Testing data9	7.84101*e* – 06	
10	Training data10	Testing data10	6.09874*e* – 06	

TABLE 6: Impact of different proportions of illegal users on the model performance.

Proportion of illegal users (%)	*N* (our model)	Accuracy (our model, Hui et al.'s model, and Wang and Hong's model)			Recall (Hui et al.'s, Wang and Hong's, and our model)			F1 value (Hui et al.'s and Wang and Hong's model)		
2.5	15	0.82	0.70	0.69	0.17	0.12	0.12	0.28	0.20	0.20
	30	0.81	0.65	0.64	0.22	0.22	0.21	0.35	0.33	0.32
	45	0.76	0.66	0.63	0.34	0.31	0.28	0.47	0.42	0.39
	60	0.74	0.63	0.61	0.47	0.44	0.34	0.57	0.52	0.44
	75	0.71	0.56	0.55	0.52	0.46	0.41	0.60	0.51	0.47
5	15	0.95	0.90	0.87	0.21	0.17	0.16	0.34	0.29	0.27
	30	0.94	0.89	0.85	0.37	0.33	0.27	0.53	0.48	0.41
	45	0.93	0.87	0.82	0.51	0.47	0.33	0.65	0.61	0.47
	60	0.91	0.85	0.81	0.61	0.56	0.41	0.73	0.68	0.54
	75	0.89	0.83	0.74	0.72	0.67	0.48	0.80	0.74	0.58
7.5	15	0.98	0.95	0.91	0.27	0.19	0.17	0.42	0.32	0.29
	30	0.96	0.93	0.89	0.39	0.32	0.29	0.55	0.48	0.44
	45	0.95	0.92	0.87	0.51	0.46	0.38	0.66	0.61	0.53
	60	0.92	0.90	0.85	0.69	0.64	0.49	0.79	0.75	0.62
	75	0.91	0.89	0.84	0.81	0.79	0.67	0.86	0.84	0.75

Table 6: Continued.

Proportion of illegal users (%)	N (our model)	Accuracy (our model, Hui et al.'s model, and Wang and Hong's model)			Recall (Hui et al.'s, Wang and Hong's, and our model)			F1 value (Hui et al.'s and Wang and Hong's model)		
10	15	1.00	1.00	0.97	0.27	0.21	0.21	0.43	0.35	0.35
	30	1.00	1.00	0.96	0.39	0.33	0.33	0.56	0.50	0.49
	45	1.00	0.97	0.94	0.59	0.55	0.54	0.74	0.70	0.69
	60	1.00	0.96	0.92	0.72	0.67	0.65	0.84	0.79	0.76
	75	0.97	0.95	0.90	0.83	0.81	0.77	0.89	0.87	0.83
12.5	15	1.00	1.00	1.00	0.26	0.21	0.21	0.41	0.35	0.35
	30	1.00	1.00	1.00	0.43	0.41	0.38	0.60	0.58	0.55
	45	1.00	1.00	0.98	0.60	0.56	0.52	0.75	0.72	0.68
	60	1.00	1.00	0.97	0.76	0.69	0.66	0.86	0.82	0.79
	75	1.00	0.98	0.94	0.83	0.81	0.78	0.91	0.89	0.85
15	15	1.00	1.00	1.00	0.27	0.23	0.22	0.43	0.37	0.36
	30	1.00	1.00	1.00	0.49	0.45	0.39	0.66	0.62	0.56
	45	1.00	1.00	1.00	0.64	0.59	0.54	0.78	0.74	0.70
	60	1.00	1.00	0.98	0.78	0.72	0.68	0.88	0.84	0.80
	75	1.00	1.00	0.95	0.89	0.82	0.76	0.94	0.90	0.84

paper considers the user's historical trust value on the basis of both, and the possibility of system misjudgment is reduced to some extent.

5. Conclusions

This article proposes a risk assessment model for the medical field. This model not only considers the risks that users may bring when choosing work objectives and accessing medical records but also considers the user's trust and reduces the misjudgment rate of the system on legitimate users under special circumstances. In our model, when a user requests access to medical information, the system can evaluate the risk of the model based on the membership function, output function, and rule base after training and decide whether to grant access based on the size of the risk, which can not only prevent illegal doctors' excessive access but also will not affect the normal work of the legitimate doctors. In addition, it is proved by comparison experiments that the evaluation output of the model is basically consistent with the actual output, and the recall and accuracy methods are superior to the existing models.

Data Availability

The original data of this article have been signed in a confidentiality agreement with the hospital and are temporarily unavailable, but the processed data (data used to support the research in this article) can be shared publicly and submitted with the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 71972165, 61763048, 61263022, and 61303234), National Social Science Foundation of China (No. 12XTQ012), Innovation and Promotion of Education Foundation Project of Science and Technology Development Center of Ministry of Education (No. 2018A01042), Science and Technology Foundation of Yunnan Province (Nos. 2017FB095 and 201901S070110), and the 18th Yunnan Young and Middle-aged Academic and Technical Leaders Reserve Personnel Training Program (No. 2015HB038).

References

[1] X. T. Jin, *Big Data of Health Care*, People's Medical Publishing House, Beijing, China, 2018, in Chinese.

[2] S. E. Fienberg, A. Slavkovic, and C. Uhler, "Privacy preserving GWAS data sharing," in *Proceedings of the International Conference on Data Mining Workshops*, pp. 628–635, IEEE Computer Society, Vancouver, Canada, December 2011.

[3] J. L. Raisaro, C. Choi, and S. Pradervand, "Protecting Privacy and Security of Genomic Data in i2b2 With Homomorphic Encryption and Differential Privacy," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 5, pp. 1413–1426, 2018.

[4] C. Lin, Z. Song, and H. Song, "Differential privacy preserving in big data analytics for connected health," *Journal of Medical Systems*, vol. 40, no. 4, p. 97, 2016.

[5] C. Lin, P. Wang, H. Song, Y. Zhou, Q. Liu, and G. Wu, "A differential privacy protection scheme for sensitive big data in body sensor networks," *Annals of Telecommunications*, vol. 71, no. 9-10, pp. 465–475, 2016.

[6] Y. Li, W. Wen, and G. Q. Xie, "Review of differential privacy protection research," *Journal of Computer Applications*, vol. 29, no. 9, pp. 3201–3205, 2014, in Chinese.

[7] P. Xiong, T. Q. Zhu, and X. F. Wang, "Differential privacy protection and application," *Journal of Computers*, vol. 37, no. 1, pp. 101–122, 2014, in Chinese.

[8] Z. Z. Xian and Q. L. Li, "Application of differential privacy protection in recommendation system," *Journal of Computer Applications*, no. 5, pp. 1549–1553, 2016, in Chinese.

[9] X. T. Wu, *Research on Privacy Protection and its Key Technologies in Big Data Environment*, Nanjing University, Nanjing, China, 2017, in Chinese.

[10] Y. L. Bai, “Application of differential privacy protection in medical big data,” *Electronic Technology and Software Engineering*, no. 24, pp. 163, 2017, in Chinese.

[11] J. Zhang and H. L. Zhang, “Research on security and protection system of big data based on cryptography,” *Information Security Research*, vol. 3, no. 7, pp. 652–656, 2017, in Chinese.

[12] Y. Tian, Y. B. Peng, and Y. L. Yang, “Attribute-based data access control scheme in wireless body area network,” *Journal of Computer Applications*, vol. 32, no. 7, pp. 2163–2167, 2015, in Chinese.

[13] D. He, S. Zeadally, and N. Kumar, “Anonymous authentication for wireless body area networks with provable security,” *IEEE Systems Journal*, vol. 11, no. 4, pp. 2590–2601, 2016.

[14] H. Xing, *Research on Privacy Protection Technology of Wireless Mobile Medical Monitoring Network*, Shanghai Jiaotong University, Shanghai, China, 2014, in Chinese.

[15] L. Wei and F. Xu, “Research on privacy protection technology of medical grid,” *Computer Technology and Development*, vol. 5, pp. 254–257, 2012, in Chinese.

[16] H. H. Wang, X. Wu, and H. Wang, “Research on regional health medical data sharing and privacy protection strategy,” *Technology Innovation and Application*, no. 31, pp. 181-182, 2017, in Chinese.

[17] H. Q. Chen, “Challenges to privacy protection of medical data in the big data environment and related technologies,” *Electronic Technology and Software Engineering*, no. 16, pp. 51–53, 2014, in Chinese.

[18] B. Mounia and C. Habiba, “Big data privacy in healthcare Moroccan context,” *Procedia Computer Science*, vol. 63, pp. 575–580, 2015.

[19] H. S. Gao and Z. Q. Sang, “Life cycle and governance of big data in medical industry,” *Journal of Medical Information*, vol. 34, no. 9, pp. 7–11, 2013, in Chinese.

[20] X. Liu and X. M. Wang, “Ethical issues in the construction of medical big data,” *Ethics Research*, no. 6, pp. 119–122, 2015, in Chinese.

[21] JASON Report: JSR-04-132, *Broader Access Models for Realizing Information DomiCorporation*, MITRE Corporation, McLean, VA, USA, 2004.

[22] S. Kandala, R. Sandhu, and V. Bhamidipati, “An attribute based framework for risk-adaptive access control models,” in *Proceedings of the 6th International Conference on Availability and Security*, pp. 236–241, IEEE, Vienna, Austria, August 2011.

[23] Q. Ni, E. Bertino, and J. Lobo, “Risk-based access control systems built on fuzzy inferences,” in *Proceedings of the 5th ACM Symposium on Information,Computer and Communications Security Bing: ASIACCS*, pp. 250–260, Beijing, China, April 2010.

[24] P.-C. Cheng, P. Rohatgi, and C. Keser, “Fuzzy multi level security an experiment on QuantifiedRisk adaptive access control,” in *Proceedings of the IEEE Symposium on Security and Privacy: S&P*, pp. 222–230, Berkeley, CA, USA, May 2007.

[25] L. Chen and J. Crampton, “Risk aware role-based access control,” *Security and Trust Management*, vol. 7071, pp. 140–156, 2011.

[26] N. Diep, L. X. Hung, Y. Zhung, and S. Lee, “Enforcing access control using risk assessment,” *Universal Multiservice Networks*, vol. 2, pp. 419–424, 2007.

[27] M. Sharma, Y. Bai, S. Chung, and L. Dai, “Using risk in access control for cloud-accessed eHealth,” in *Proceedings of the 2012 IEEE 14th International Conference on High Performance Computing and Communications, HPCC*, pp. 1047–1052, Liverpool, UK, June 2012.

[28] Q. Wang and J. Hong, “Quantified risk-adaptive access control for patient privacy protection in health information systems,” in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security: ASIACCS*, pp. 406–410, Hong Kong, China, March 2011.

[29] R. A. Shaikh, K. Adi, and L. Logrippo, “Dynamic risk-based decision methods for access control systems,” *Computers & Security*, vol. 3, no. 31, pp. 447–464, 2012.

[30] D. Choi, D. Kim, and S. Park, “A framework for context sensitive risk-based access control in medical information systems,” *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 265132, 9 pages, 2015.

[31] Z. Hui, H. Li, M. Zhang, and D. G. Feng, “Risk-adaptive access control model for medical big data,” *Journal of Communications*, vol. 36, no. 12, pp. 190–199, 2015.

[32] J. Zhang, *Medical Big Data Privacy Security Risk Assessment in Cloud Environment*, Yunnan University of Finance and Economics, Kunming, China, 2018, in Chinese.

[33] R. Jiang, M. Y. Shi, and W. Zhou, “A privacy security risk analysis method for medical big data in urban computing,” *IEEE Access*, vol. 7, pp. 143841–143854, 2019.

[34] J. Li, Y. Bai, and N. Zaman, “A fuzzy modeling approach for risk-based access control in eHealth cloud,” in *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, vol. 2103, pp. 17–21, Melbourne, Australia, July 2013.

[35] E. Smith and J. Eloff, “Cognitive fuzzy modeling for enhanced risk assessment in a health care institution,” *IEEE Intelligent Systems*, vol. 5, no. 4, pp. 69–74, 2000.

[36] M. Y. Shi, R. Jiang, X. H. Hu, and J. W. Shang, “A privacy protection method for health care big data management based on risk access control,” *Health Care Management Science*, vol. 23, no. 3, pp. 427–442, 2019.

[37] X. Q. Lian, *Fuzzy Control Technology*, China Electric Power Press, Beijing, China, 2003, in Chinese.

[38] W. Shang, T. Gong, C. Chen, J. Hou, and P. Zeng, “Information security risk assessment method for ship control system based on fuzzy sets and attack trees,” *Security and Communication Networks*, vol. 2019, Article ID 3574675, 11 pages, 2019.

[39] M. Fayaz, S. Ahmad, L. Hang, and D. Kim, “Water supply pipeline risk index assessment based on cohesive hierarchical fuzzy inference system,” *Processes*, vol. 7, no. 128, pp. 1–15, 2019.

[40] X. M. Shi and Z. Q. Hao, *Fuzzy Control and MATLAB Simulation*, Tsinghua University Press, Beijing Jiaotong University Press, Beijing, China, 2018, in Chinese.

[41] Z. L. Jiang, *Introduction to Artificial Neural Networks*, Higher Education Press, Beijing, China, 2001, in Chinese.

[42] X. L. Wu and Z. H. Lin, *MATLAB-aided Fuzzy System Design*, Xidian University Press, Xi’an, China, 2002, in Chinese.

[43] G. Y. Li and L. J. Yang, *Neural, Fuzzy, Predictive Control and Their MATLAB Implementation*, Electronic Industry Press, Beijing, China, 2018, in Chinese.

[44] H. H. Niu and L. X. Liu, “Application research of neural network in information security risk assessment,” *Computer Simulation*, vol. 28, no. 6, pp. 117–120, 2011, in Chinese.

[45] Y. B. Liu, W. F. Zhang, and X. M. Wang, “Access control scheme based on multi-attribute fuzzy trust evaluation in

cloud manufacturing environment," *Computer Integrated Manufacturing Systems*, vol. 24, no. 2, 2018, in Chinese.

[46] X. J. Shi and W. H. Yu, "Quantitative method of access control risk based on fuzzy neural network," *Intelligent Computer and Application*, vol. 8, no. 1, pp. 1–4, 2018, in Chinese.