IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 12, 2019, accepted April 3, 2019, date of publication April 10, 2019, date of current version April 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909538

Diagnosis and Data Probability Decision Based on Non-Small Cell Lung Cancer in Medical System

JIA WU[1,2], (Member, IEEE), PEIYUAN GUAN[1,2], (Student Member, IEEE), AND YANLIN TAN[2,3]

[1]School of Computer Science and Engineering, Central South University, Changsha 410075, China
[2]Mobile Health Ministry of Education-China Mobile Joint Laboratory, Changsha 410083, China
[3]PET-CT Center, The Second Xiangya Hospital, Central South University, Changsha 410083, China

Corresponding authors: Jia Wu (jiawu5110@163.com) and Peiyuan Guan (guanpeiyuan@csu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672540, and in part by the Hunan Provincial Natural Science Foundation of China under Grant 2018JJ3299 and Grant 2018JJ3682. The work of Y. Tan was supported by the National Natural Science Foundation of China.

ABSTRACT There are many factors affecting the survival of people in developing countries, such as the tremendous number of population, nonuniform medical resources, and the threatening of malignant diseases. The improvements in medical information system in developing countries may lead to a bright future. By using effect medical resources and utilizing the information coming from the medical system, the doctors could come to a diagnosis with analysis. The probability of getting sick is very useful information which assists doctors to improve the accuracy of disease diagnosis, shortening treatment time, and reducing the incidence of misdiagnosis. This paper aims to build a model, considering not only probability analysis but also decision making, which can play a crucial role to figure out the probability of non-small lung cancer transitions in four different stages. In each process of the model, selecting effective parameters with big data are adopted for finding maximum effect with the top three high relevancy diagnose and decision data. With effective treatment methods that improve the relevancy diagnose data, the probability of malignant disease development will decrease. It is proved by the statistical analysis of clinical data that the model provides clinical data fast with enough accuracy.

INDEX TERMS Non-small cell lung cancer(NSCLC), data probability decision, state transition, relevancy data.

I. INTRODUCTION

Many developing countries in Asia and Africa, health of people may not be protected by medicine, because of huge population and underdevelopment medical technology [1], [2]. The SARS virus and the Ebola virus broke out in Asia and Africa, more than thousands of people lose their life. If patients can obtain treatment and epidemic disease can be controlled in those areas, more people can keep their life [13]–[17].

Some developing countries such as China and India which have over one billion people may faced the biggest problem is population and medical resource imbalance. In China, a super city such as Beijing, has over 20 million workers and 10 million children and the aged. Only no more than 3,000 doctors can afford medical health service. Unfortunately, those doctors must solve hundreds of pathological reports which come from thousands of kilometers, because advanced medical devices and excellent doctors are centralized in super cities and advanced hospitals. Over more, some reports and pathological sections may be a big challenge, the question 'malignant tumor or not' may wait for advanced hospitals and doctors answered [18]–[22].

In many malignant tumors, lung cancer is high incidence rate tumor in developing countries [3]. Non-small cell lung cancer (NSCLC) in China may account for 85% with lung cancer. Over 70% of patients with lung cancer present with locally advanced. It was a difficult condition to manage due to the lack of effective treatments [4]. However, the 5-year survival rate may rise to 80% when a patient can be diagnosingd in the early stage [5].

In fact, many judgments and diagnosis reports for NSCLC could be analyzed by medical devices. Doctors can make a decision by their experience after reports supported. However, in developing counties, we must face some embarrassed.

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

(1) high misdiagnosis rate between doctors and patients carry social contradiction. Many families start doubting excessive consumption coming from doctors' conclusion.

(2) many repetitive works may affect doctors, such as similarity diagnosis reports, images, biochemical indices, can increase high misdiagnosis rate by doctors.

(3) no effective evidences or conclusions can acquire confidence between patients and doctors, because patients can not know complicated indices.

How to give an effective evidence and reduce high misdiagnosis rate in developing countries when we faced big data population.

The development of AI medical system in developing countries may create good results [24], [25]. With medical resources and AI medical system, doctors can make decision analysis. The probability of acquiring sick assists doctors or decision making. In the light of decision system in medical treatment, the features can be applied in IoT system, hospitals, patients, and doctor would combine a communication medical system. This system does not only quickly provide messages to patients, but also reduces the pressure of obtaining resources [26]–[30], [32]. According to AI medical system, patients could know the conclusions by medical system. It could improve many social contradiction between doctors and patients

In this work, we establish a probability analysis and decision making model and design a method to analyze NSCLC. The contributions in this study are:

(1) we found a model to discuss the evolution stages of NSCLC, and then divide disease diagnosis parameters evolution process.

(2) according to big data research, we establish the three effective parameters and evaluate the selection stage with NSCLC.

(3) in the light of the probability of malignant disease development, we combine effective treatment methods in AI system in hospital.

(4) to combine clinical data statistical analysis and probability decision, system may provide accurate and fast analysis clinical data and decision-making advice.

II. RELATED WORKS

Computer-controlled medical decision making and data trans Computer-controlled medical system has become research hotspots in the medical treatment field. Plenty of works have entered clinical application.

Literature [7] suggests a system data model in medical data and IOT system. The resource-based data accessing method (UDA-IoT) can be designed to acquire the IoT data and then improve data resources. This medical system may explain IoT-based system in emergency medical services and demonstrate how to collect or interoperate IoT data and then support to emergency medical services. The result explains resource-based IoT data system can be effective in heterogeneous data environment and support data accessing timely in medical system. However, many developing countries can not afford high cost in IoT data system. Because emergency medical services need to many high quality devices in system.

Literature [8] shows a knowledge-based medical system. It can conclude two parts. One is the heuristic approach to enhance. It used to pseudo relevance method for more effective query expansion, even if expanding queries boosting the similarity score. The other is how to improve the retrieval performance with knowledge-based. It can explain a relevance model based on tensor factorization which can identify semantic association patterns in sparse settings. All data and patterns are used as inference paths in knowledge-based query expansion and copy to medical information retrieval. Knowledge-based in many developing countries is a big challenge, how to collect those data is hard to completed.

Literature [9] proposed a framework in medical system. It used natural language processing and then analyzed clinical notes readmission. In this system, many methods can be adopted in the field of data mining and machine learning. This framework is created and selected the best components while maintaining fast computational times in medical environment. However in hospitals, data mining and machine learning is not a good method, because it may cause data divulged. It is very insecurity in medical system.

Literature [10] discusses the discrete event system specification in medical system engineering. It can develop coordination models in transactions that involve multiple disparate activities and then it needs to be selectively sequenced to coordinated care interventions. This system shows how such coordination concepts provide the layers and support the proposed information in healthcare as a learning collaborative system. In literature [11] addresses an organ-centric compositional medical system. In this model, medical devices can be composed into clusters and then created organ-specific physiology in system manner. The organ-centric brings device in many patterns of sensing and then control human physiology. The organ-centric architectural patterns can enable rapid and composition of supervisory controllers in medical scenarios. In the literature [12] proposes a low-complexity coordination protocol for networked in medical system. The suggestion architecture organizes in a hierarchical and then reminder the manner in accordance to human physiology. It avoids potential conflicts and useless controls, when efficient concurrent in medical devices. Those methods can solve high time complexity in medical system, we can use those methods in application.

This article will analyze the condition of disease stage, the combination of effective selection and the tracking of associated data, with considerations about the design of effective treatment decision-making model.

III. MODEL DESIGN

A. ILLNESS DEVELOPMENT STAGES ASSOCIATED WITH DIAGNOSTIC PARAMETER ANALYSIS OF THE MODEL

In the study of modern medicine, an intelligent diagnosis assists doctors in condition analysis and judgment, which can

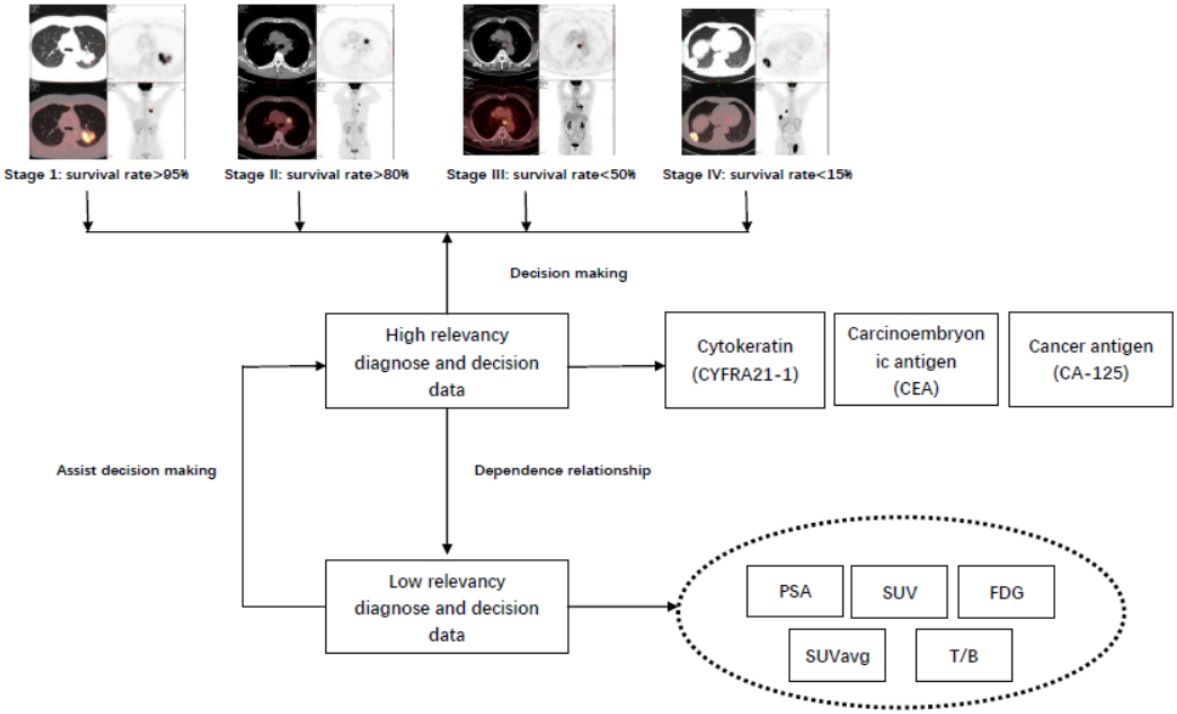


FIGURE 1. Development stage of NSCLC patients and relationships between the correlation data.

effectively shorten diagnosis time and reduce the probability of misdiagnosis. The model proposed in this study provides a bright diagnosis solution. At the same time, doctors can also obtain auxiliary diagnostic conclusions through the system. The system not only provides a comprehensive analysis of the patient's condition, but also gives an auxiliary diagnosis and offers suggested treatments.

In this section, we will discuss the decision-making system, diagnosis process in hospital system and decision process in history information system.

Figure 1 shows the four stages of development of NSCLC and reveals the interrelationships between the data. By considering the tumor markers and PET/CT scan results of NSCLC, patients may obtain early and accurate diagnosis conclusions about lung cancer. Many patients who are in stage I and II of NSCLC survive for many years due to the surgical resection. If the patients are more advanced illness, such as stage III or IV, they are not suitable for surgical resection. Chemotherapy has positive effect for palliation if the situation is locally advanced or the tumor cells have been metastatic.

There are some low relevancy diagnosis data between some decision parameters. Those data belong to the CT scan and do not use contrast injection which are causing less damage by human. If high relevancy data are less than the threshold values, the patient's condition is likely to enter the next phase. Low relevancy data are useful if they are used in a proper way. By considering them associated with high relevancy data, it is possible to obtain a common decision. Furthermore, doctors can determine the patient is in which stage of the disease.

The stage division in NSCLC may adopt the values of machine scanning to identify which stage the patient is in. It assumes the stage decision $V_{stage}(t)$, which indicates the correlation between the relevancy diagnose and decision data diagnosis. The decision-making value can be judged by value $V_{stage}(t)$ as follows:

$$V_{stage}(t) = \alpha \times V_{high}(t) + \beta \times V_{low}(t) \tag{1}$$

In equation (1), t represents the current time. $V_{high}(t)$ is the value of high relevancy data. On the contrast, the low relevancy data could be presented by $V_{low}(t)$. Parameters α and β are the influence coefficients.

In NSCLC, $V_{stage}(t)$ is consist of seven critical regions, $V_{stage}(t) \subseteq [\phi(i), \phi(j)]$. Figure 1 explains the different survival rates in four stages. The range of $V_{stage}(t)$ may stand for time t in the stage of disease.

The most important parameters in NSCLC conclude the Cytokeratin (CYFRA21-1), Carcinoembryonic antigen (CEA), and Cancer antigen (CA)-125.If a patient has significantly high results of the tests on those three parameters, it may cause the deterioration. In this situation, it is a common phenomenon that the survival rate decreases dramatically.

The high relevancy data $V_{high}(t)$ conclude three parts. Those namely, Cytokeratin (CYFRA21-1) value of $V_{high-CYF}(t)$, Carcinoembryonic antigen (CEA) value of $V_{high-CEA}(t)$, and Cancer antigen (CA)-125 of $V_{high-CA}(t)$. The high

relevancy data $V_{high}(t)$ can explain as:

$$V_{high}(t) = \alpha_i \times \frac{V_{high-CYF}(t)}{V_{high-CYF}(t)(\overline{A_{age}(y)})} + \alpha_j \times \frac{V_{high-CEA}(t)}{V_{high-CEA}(t)(\overline{A_{age}(y)})} + \alpha_k \times \frac{V_{high-CA}(t)}{V_{high-CA}(t)(\overline{A_{age}(y)})} \quad (2)$$

Moreover, CYFRA21-1, CEA, and CA-125 values, age average, and the range of normal parameter values of the region are determined.

Once the patient may choose the parameters they want to consider, system can judge the value of $V_{high}(t)$. After that, we can come to the conclusion about the stage of NSCLC.

(1) $V_{high}(t) \geq \varepsilon, \varepsilon$ is a threshold. It may control and determine the development for NSCLC in the third or fourth stage. We set the influence factor as $\alpha = 1, \beta = 0$. The determination value of $V_{stage}(t)$ can represent as:

$$V_{stage}(t) = V_{high}(t) = \alpha_i \times \frac{V_{high-CYF}(t)}{V_{high-CYF}(t)(\overline{A_{age}(y)})} + \alpha_j \times \frac{V_{high-CEA}(t)}{V_{high-CEA}(t)(\overline{A_{age}(y)})} + \alpha_k \times \frac{V_{high-CA}(t)}{V_{high-CA}(t)(\overline{A_{age}(y)})} \quad (3)$$

Equation (3) shows the development of illness, which indicates the illness has already been the late stage. For a patient, the possibility for survival is quite low in this period.

For most patients in NSCLC, it is difficult to diagnose when the illness is in the first stage. Most of the diagnosis is confirmed when the illness is in the late stage with a low probability of survival.

(2) $V_{high}(t) < \varepsilon$, it explains the patient is the development of NSCLC in the first or in the second stage. For low relevancy data $V_{low}(t)$, this parameter can contain detection in PET-CT and CT. And then the $V_{low}(t)$ can be defined as:

$$V_{low}(t) = \frac{1}{n}\sum_{i=1}^{n} \beta_i \times v_{low}^{(i)}(t) \quad (4)$$

$v_{low}^{(i)}(t)$ explains the low relevancy step in the diagnosis in the i measures of detection, β_i explains the parameter i factor, $0 < \beta_i < 1$. In NSCLC, parameter i concludes T/B, SUVmax, and PSA on the inspection with project indicators. It can check project up to eight, $0 \leq i \leq 8$.

From Equations (2)-(4), the decision value $V_{stage}(t)$ can be shown as follows:

$$V_{stage}(t) = \begin{cases} V_{high}(t), & V_{high}(t) \geq \varepsilon \\ \alpha \times V_{high}(t) + \beta \times V_{low}(t), & V_{high}(t) < \varepsilon \end{cases}$$

$$= \begin{cases} \alpha_i \times \frac{V_{high-CYF}(t)}{V_{high-CYF}(t)(\overline{A_{age}(y)})} + \alpha_j \times \frac{V_{high-CEA}(t)}{V_{high-CEA}(t)(\overline{A_{age}(y)})} + \alpha_k \times \frac{V_{high-CA}(t)}{V_{high-CA}(t)(\overline{A_{age}(y)})} & V_{high}(t) \geq \varepsilon \\ \frac{1}{n}\sum_{i=1}^{n} \beta_i \times v_{low}^{(i)}(t) + \alpha_i \times \frac{V_{high-CYF}(t)}{V_{high-CYF}(t)(\overline{A_{age}(y)})} + \alpha_j \times \frac{V_{high-CEA}(t)}{V_{high-CEA}(t)(\overline{A_{age}(y)})} + \alpha_k \times \frac{V_{high-CA}(t)}{V_{high-CA}(t)(\overline{A_{age}(y)})}, & V_{high}(t) < \varepsilon \end{cases} \quad (5)$$

According to judge the value of $V_{stage}(t)$, patient and doctor can access in the stage of NSCLC. Many works in decision-making and treatment analysis method also can be adopted in time.

In NSCLC treatment, doctors may focus on CYFRA21-1, CEA, and CA-125 values when they make a decision. In order to accomplish accuracy prediction, we can select 5 to 8 low relevancy parameters joining in auxiliary diagnosis. Such as T/B, SUVmax, PSA and other parameters can be adopted in different step in NSCLC.

B. DIAGNOSTIC PARAMETERS AND TREATMENT DECISION-MAKING ANALYSIS OF THE RELATIONAL MODEL

In the medical diagnosis decision-making, diagnostic parameters directly affect how the doctors identify at which stage of the disease the patient is in, especially high relation data directly affect patient survival. Drug, physical, and surgical therapies can change the patient's vital signs and promote the evolution of physiological parameters to a benign illness. At the same time, drug or treatment of postoperative changes in lifestyle, such as drinking and smoking habits, sleep, and exercise, is also an important means to improve from the disease.

In the treatment of NSCLC, stages I and II treatments mainly adopted the drug approach to improve the way of life and to control the condition. For stages III and IV, in addition to medication and lifestyle adjustments, surgery, chemotherapy, and laser treatment are necessary methods.

Each patient adopted different treatments; the parameters of the improvement effect and the improvement of the illness condition are also different.

Figure 2 shows the treatment processes of NSCLC diagnosis parameters and the diagnosis treatment decision-making scheme. For the treatment of NSCLC, the main treatment methods are mainly used in the first stage. Twelve kinds of treatments are used in stage I and six of that in stage II. Furthermore, stages III and IV increased number of treatments to six more. The four main lifestyles are drinking, smoking, sleep, and exercise. We can set ω_i and ϕ_j treatments and lifestyle data packets, $1 \leq i \leq 12$, $1 \leq j \leq 4$. By combining

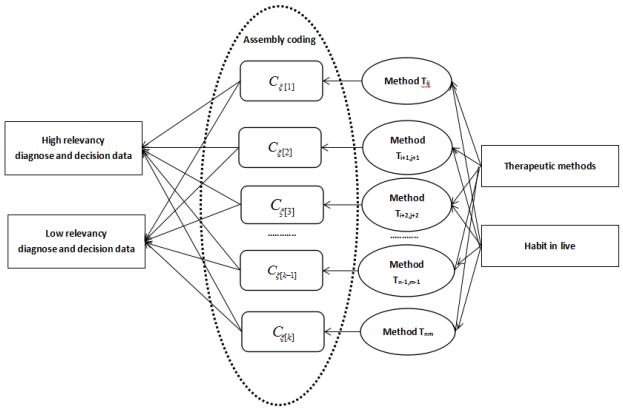


FIGURE 2. Diagnosis parameters and diagnosis and treatment decision-making scheme.

the two ways, methodologies T_{ij} are obtained as:

$$T_{ij} = (\omega_i, \phi_j), \quad 1 \le i \le 12, 1 \le j \le 4 \tag{6}$$

T_{ij} is group of users, with 16-bit coding vector. $C_{\xi[k]}$ representation is shown below:

$$C_{\xi[k]} = \{\underbrace{1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0}_{\textbf{Therapeuticmethods}} | \underbrace{0, 1, 1, 1}_{\textbf{Habitinlive}}\}$$

In 16-bit coding, 1 to 12 are treatments, and 13 to 16 bits are encoded lifestyles. 1 adopted the packets, 0 means no use.

In the treatment of NSCLC, a large number of patients have multiple diagnoses in the hospital, with most of the patient's condition in the first stage and in the second phase of the treatment process. To reduce the data storage space in the recording process, the code is simplified as follows:

$$C_{\xi[k]} = \begin{cases} \{\underbrace{1, 0, 1, 0, 1, 1}_{\textbf{Therapeuticmethods}} | \underbrace{0, 1, 1, 1}_{\textbf{Habitinlive}}\} \\ \{\underbrace{\underbrace{1, 0, 1, 0, 1, 1}_{Stage1,2}, \underbrace{0, 0, 0, 0, 0, 0}_{Stage3,4}}_{\textbf{Therapeuticmethods}} | \underbrace{0, 1, 1, 1}_{\textbf{Habitinlive}}\} \end{cases}$$

In this method, statistics for each treatment group for correlation parameters can be changed. Especially, in the hospital HIS system, each patient will be assigned a storage record HIS: dig/IDXXX/code, ID number, patient's code, and program code. Record of the current treatment facilitates HIS record.

In NSCLC treatment, we usually watch for the high and the low relevancy data to determine the efficacy of treatment. If a treatment effect is obvious, we tend to continue in this way and continue to the next treatment. If the treatment method does not significantly improve, then we use another kind of treatment method.

We can set up the Tr_{sur} explains the time period of t as diagnosis parameter range. Suppose that we currently selected $C_{\xi[i]}$ scheme as the treatment method, the changes can be calculated the diagnosis value of $V_{stage}(t)$ according to Equation (5):

$$V_{stage}(t) = V_{stage}(t) - V_{stage}(t+1) \tag{7}$$

Here, we can analyze the change of $V_{stage}(t)$:

1) $V_{stage}(t) < 0$, the current treatment, is not an effective treatment choice and worsens the illness. The doctors and patients give up the current treatment options.

2) $V_{stage}(t) \ge Tr_{sur}$, the current treatment plan, has considerable improvement, is an excellent treatment

3) $0 < V_{stage}(t) < Tr_{sur}$, the current treatment, affects the disease, but the treatment effect is not obvious; at this time, we should add other

treatments $C_{\xi[i+1]}$, continue to treat the disease, and calculate the next treatment period $V_{stage}(t)$:

$$V_{stage}(t) = V_{stage}(t) - V_{stage}(t+1) - V_{stage}(t+2) \tag{8}$$

Obtain with $V_{stage}(t)$ and comparison the diagnosis change value Tr_{sur}. If $0 < V_{stage}(t) < Tr_{sur}$, continue to add new method for treatment. Diagnosis and treatment after N time is derived to obtain a polynomial calculation formula as follows:

$$V_{stage}(t) = V_{stage}(t) - \sum_{i=1}^{N} V_{stage}(t+i) \tag{9}$$

After N treatment, $V_{stage}(t) \ge Tr_{sur}$ indicated that the treatment exhibited obvious changes. In medical HIS system, each article of selected medical record, will be recorded by HIS system. Record format is: His: dec/ID/code, ID is the patient's code, which is the corresponding decision call number for many times.

C. PATIENT'S DIAGNOSIS DECISION CALLS AND DIAGNOSIS PROCESS

The illness development stages are associated with diagnostic parameter of the model analysis, parameters of diagnosis, and treatment decisions model analysis. The relationship between hospital diagnosis systems for the following process can be designed; the design model is shown in Figure 3.

Figure 3 shows a patient in a hospital environment with several diagnosis records. HIS system of operation steps are as follows:

1) The patient gets into the hospital for the first time, is assigned as ID327; 327 is the patient in HIS system of tags. At the same time, the system allocates a time mark His: day/ID327/20170102.

2) By the doctor's orders, a doctor's advice records His: doctor's advice, patients by PET-CT, TC devices, such as multiple diagnosis results are obtained, and the four diagnoses are: His: dig/ID327/18546; His: dig/ID327/18442; His:: dig/ID327/18891; and His:: dig/ID327/18112. With four protocols, the two current two suitable decisions record His: dec/ID327/557; His:: dec/ID327/512. Patients and doctors can find this item.

3) Patients, after a time interval, go to the hospital again. At this time, HIS allocation for the second time tag His:: day/ID327/20170104.

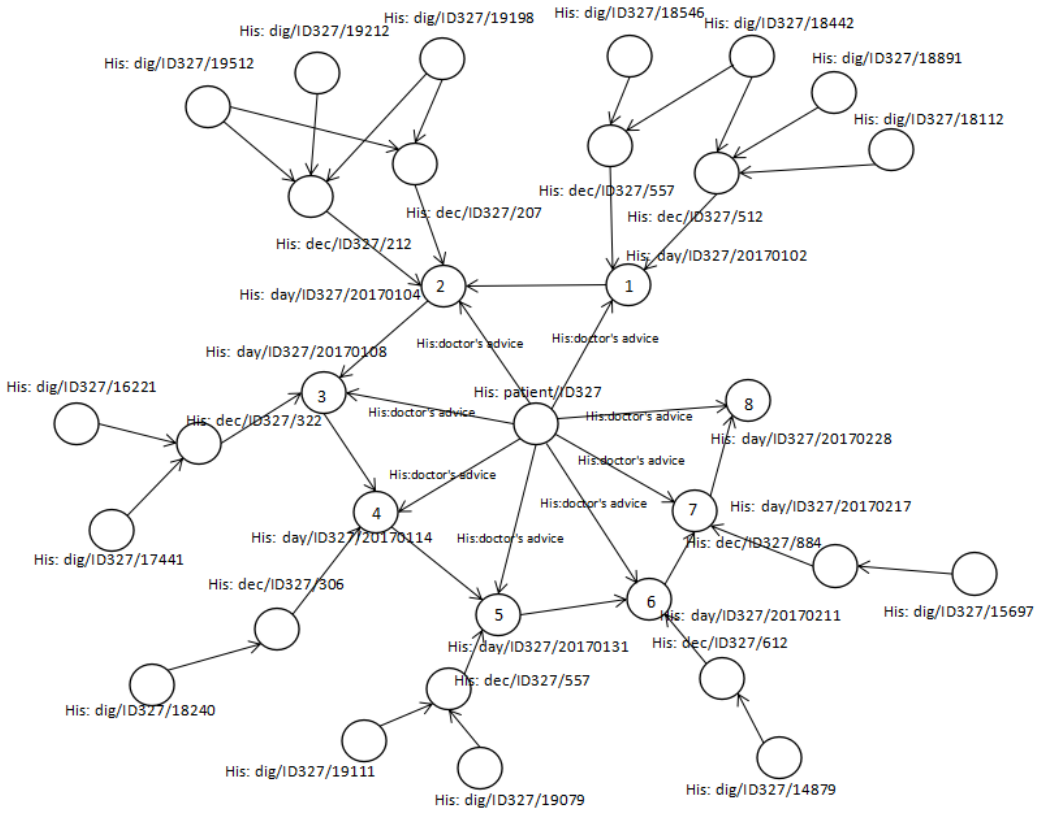


FIGURE 3. Diagnosis process in hospital system.

4) Through the doctor's orders His: doctor's advice, in the process of three diagnoses: His: dig/ID327/19512; His:dig/ID327/19212; His: dig/ID327/19198; Two suitable decision records: His: dec/ID327/212;andHis: dec/ID327/207, were formed.

5) The above operation is repeated, after three to seven treatment processes.

6) When the patient goes to the hospital for the eighth treatment, the system allocates a time mark His: day/ID327/20170228. The patient's physiological indexes have reached the rehabilitation conditions; the patient can no longer receive hospital treatment; and the treatment process is over.

Through the doctor's orders, record, and decision of the diagnosis application of the system, diagnosis and treatment of speed and degree of alternative are improved. The doctor can recommend items based on the system chosen for the patient's condition to achieve the effect of rapid diagnosis.

IV. EXPERIMENTAL DESIGN

In this paper, all records come from the mobile health information of the Ministry of Education-China Mobile Joint Laboratory.

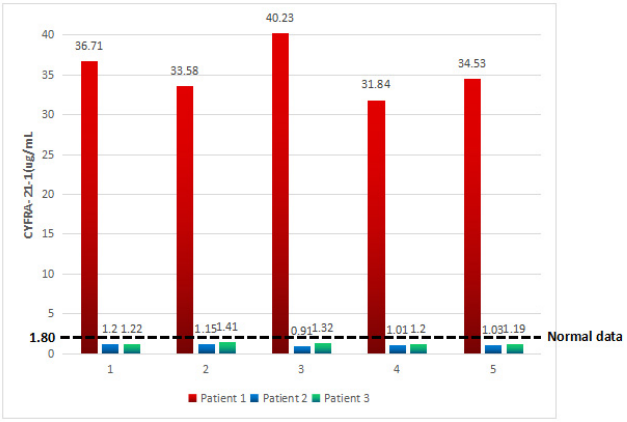
NSCLC data in three hospitals from 2011 to 2015 are classified. Various consolidated NSCLC data cases are shown in Table 1.

TABLE 1. Data collection and type for NSCLC in three hospitals.

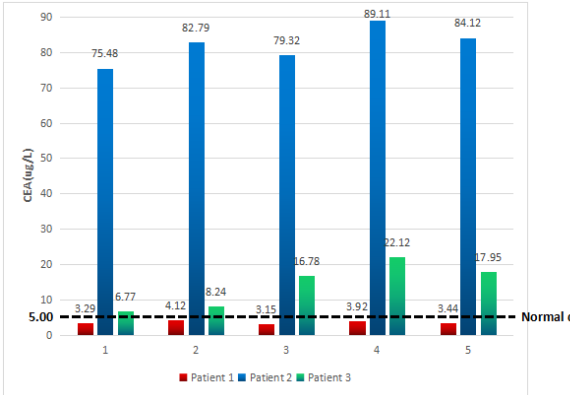
TYPE	NUMBER
Patient information	2,789,675 items
Outpatient service	968,545 people
Doctors' device in outpatient	28,554,590 items
Be hospitalized	1,676,899 people
Diagnosis	1,124,561 items
Electronic medical records	5,287,413 items
Doctors' device in clinical	31,427,790 items
Inspection records	179,712 items
Medical laboratory records	9,483,216 items
Routine inspection records	24,287,612 items
Operation records	393,218 items
Drug records	90,631 items

The experiment is described as follows. Table 2 and 3 records the normal data area in NSCLC.

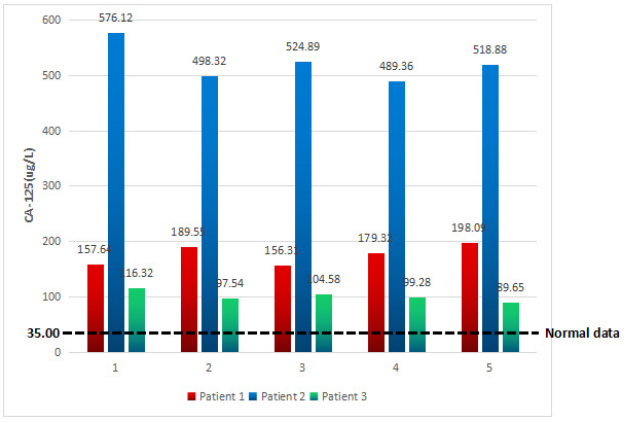
Table 4 shows the various parameter test results of the three patients in the hospital. Every patient after five diagnostic parameter collections contains three groups with high relevancy associated diagnostic indicators and five groups with low relevancy index. In this experimental design, the age for all patients are 45-60 years old. Normal data in high parameters on average are: 1.8(CYFRA-21-1), CEA(5.0), CA-125(35.0).



(a)



(b)



(c)

FIGURE 4. (a) Three patients CYFRA-21-1 performance (b) Three patients with performance of the CEA (c) Performance of the three patients CA-125.

Figure 4a depicts the comparison of three patients about the results of CYFRA-21-1. Obviously, we can come to the conclusion that the CYFRA-21-1 is normal if the parameter is in the range of[0, 1.8].Patients 2 and 3 showed sampling results that were within the normal range, according to their physical sign CYFRA-21-1. Five sampling results of patient1 are over

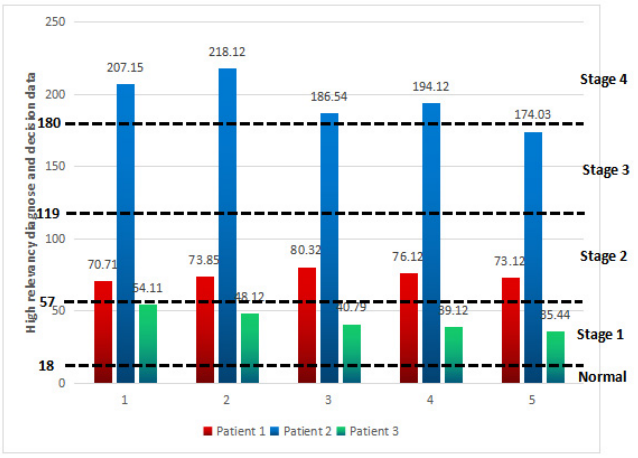


FIGURE 5. high relevancy data in different phase contrast (three strong correlation parameters with similar weighting factors).

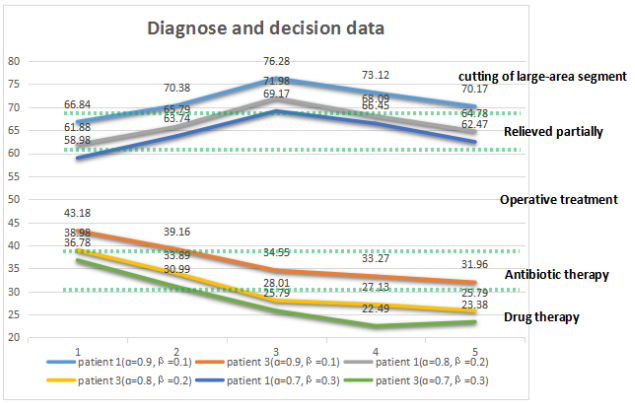


FIGURE 6. NSCLC disease stages 1 and 2 of the decision analysis.

30 which is much larger than 1.8.The average level of patient 1is about 35. This parameter indicates the patient is in the danger of illness.

Figure 4b shows the analyses of three patients in the CEA. From Figure 4 b, it is a consequence the normal range of CEA should be[0, 5.0]. One patient showed five sampling results within the normalrange.CEA signs show the normal. Patients 2 and 3have five abnormal sampling results which are larger than 5.0; particularly in patient 2, sign parameters are 16 times greater than the normal, showing the CEA in the abnormal state with patients.

Figure 4c shows the analysis of the three patients in the CA-125. We can see that the CA-125 in the normal range is between 0 and 35.0, with the five sampling results of all the patients larger than normal and showing all the CA-125 in the abnormal state of patients.

Based on the analysis of three patients' strong relational parameters and Equation (2), it is possible for us to calculate V_{high} (t) as their decision value. Assuming the weight of three strong correlation parameters is the same, the distribution of patients who have high values of correlation parameter is shown in Figure 5.

TABLE 2. highrelevancy diagnose and decision data with normal data in NSCLC.

high relevancy data and normal area			
Parameter	CYFRA-21-1(ug/mL)	CEA(ug/L)	CA-125(KU/L)
Normal data area	0~1.80	0~5.00	0~35.00

TABLE 3. low relevancy diagnose and decision data with normal data in NSCLC.

low relevancy data and normal area					
Parameter	NSE(ug/mL)	CA242(KU/L)	PSA(ug/mL)	HGH(ug/mL)	Free-PSA(ug/mL)
Normal data area	0~13.00	0~20.00	0~5.00	0~7.50	0~1.00

TABLE 4. Physiological indexes for patients in collection.

		CYFRA-21-1 (UG/ML)	CEA (UG/L)	CA-125 (KU/L)	NSE (UG/ML)	CA242 (KU/L)	PSA(UG /ML)	HGH(U G/ML)	FREE-PSA(UG /ML)	FERRITI N(KU/L)
PATIENT1	1	36.71	3.29	157.64	21	31	0.81	0.51	1.88	154.2
	2	33.58	4.12	189.55	16	24	1.01	0.82	1.45	189.6
	3	40.23	3.15	156.31	27	32	0.95	0.77	1.78	175.8
	4	31.84	3.92	179.32	22	28	1.45	0.48	0.81	193.7
	5	34.53	3.44	198.09	19	31	0.98	0.89	0.57	173.8
PATIENT2	1	1.20	75.48	576.12	33	9	1.22	11.25	21.88	935.7
	2	1.15	82.79	498.32	37	5	1.48	22.82	28.74	854.1
	3	0.91	79.32	524.89	22	6	1.88	19.85	24.32	718.2
	4	1.01	89.11	489.36	24	7	0.99	23.58	26.81	921.5
	5	1.03	84.12	518.88	27	4	1.57	18.78	37.58	814.6
PATIENT3	1	1.22	6.77	116.32	31	21	7.22	6.51	0.12	258.9
	2	1.41	8.24	97.54	36	26	7.52	5.32	0.55	322.7
	3	1.32	16.78	104.58	35	29	7.14	4.87	0.17	278.9
	4	1.20	22.12	99.28	28	24	8.56	5.99	0.45	341.8
	5	1.19	17.95	89.65	21	22	8.47	6.02	0.67	304.8

According to the result of Equation (2), we have the analysis of three patient diagnosis decision-making five times. In the whole process, we set three strong correlation parameters with similar weighting factors, namely $\alpha_i = \alpha_j = \alpha_k = \frac{1}{3}$. Thus, we can calculate the different decision high relevancy data decision values of diagnosis for each patient. In the figure, patient 1 has five high relevancy data, with determined values between 57 and 119. The patient is in the second stage of NSCLC concurrent and belongs to the mild NSCLC. Patient 2 from first to fourth diagnostic process have high relevancy data decision values greater than 180, indicating that the patient is in the fourth stage. In the fifth diagnosis, high relevancy data decision values were between 119 and 180, indicating that the patient's condition is at stage III to IV. Patient 3 has strong correlation parameters decision values between 18 and 57, and has complications by the first stage.

Using the strong correlation parameters for determining value, the system can rapidly predict in which stage or severity of the disease the patients are in. Thus, the doctor can facilitate quickly the choice of treatment plan,, thereby reducing the error caused by the doctor's personal judgment.

Patients 1 and 2 are in NSCLC disease stages 1 and 2. To accurately determine the choice of treatment plan, we can combine the strong correlation parameters to determining value method and data analysis for the two patients. The analysis procedure is discussed below.

Patients 1 and 2 are in NSCLC disease stages 1 and 2. To accurately determine the choice of treatment plan, we can combine the strong correlation parameters to determining value method and data analysis for the two patients. The analysis procedure is discussed below.

According to Equation (5), two patients have five different diagnosis results. The analysis diagram is shown in Figure 6.

TABLE 5. Physiological indexes for a patients in collection.

		CYFRA-21-1 (UG/ML)	CEA (UG/L)	CA-125 (KU/L)	NSE (UG/ML)	CA242 (KU/L)	PSA(UG/ML)	HGH(UG/ML)	FREE-PSA(UG/ML)	FERRITIN(KU/L)
PATIENT	1	4.16	285.41	711.01	34	8	1.12	10.25	58.81	835.7
	2	5.57	277.99	688.81	36	5	1.42	11.21	49.71	754.1
	3	3.55	257.15	521.42	27	6	1.86	12.15	48.22	738.2
	4	4.28	231.44	461.56	25	6	1.29	11.20	47.55	622.1
	5	3.47	184.88	408.18	36	5	1.54	15.71	38.51	422.6
	6	4.84	128.11	321.88	27	7	1.68	13.88	35.12	351.8
	7	5.17	62.89	295.1	38	6	1.71	12.51	11.6	211.1
	8	3.89	21.17	178.2	21	5	1.55	13.61	7.1	209.7

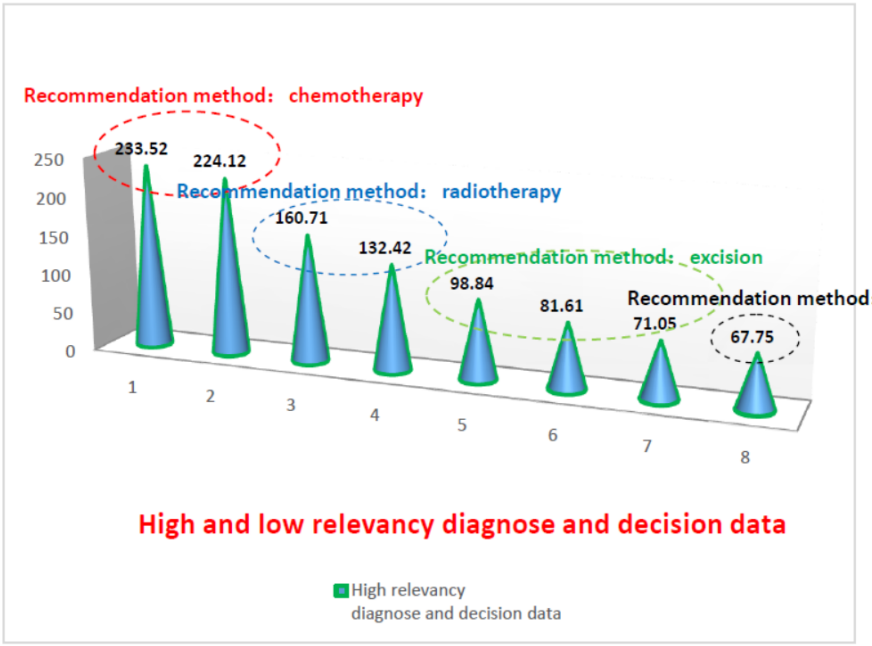


FIGURE 7. Altered values of the patient's high relevancy data decision (HRDD).

Figure 6 shows the combined high relevancy data determined values and the low relevancy data for determining value method of patients 1 and 3. By changing the influence factors, we choose different treatment methods. We can see from Figure 6 that with the impact factor, $\alpha = 0.9$ and $\beta = 0.1$, in patient 1, large-area segment is cut off by the high relevancy data to determine value correction, which can improve the treatment of patients with complications. Complications are reduced or relieved partially, which can help reduce the operation risk of patients, reduce the trauma of patients, and accelerate postoperative healing. When α is reduced from 0.9 to 0.7, β increases from 0.1 to 0.3, and three patients exhibited these obvious changes. The need for surgical treatment from the beginning is gradually adjusted to the need of antibiotic therapy. Finally, high relevancy data for determining value correction can be used for normal medication treatment of NSCLC. Judging from this, adopting the high relevancy data decision value can accurately adjust the parameters of each stage of NSCLC; thus, improving the methods of treatment.

Table 5 represents major patients in a hospital for the whole process of diagnosis and treatment. The patient's high relevancy data decision values can alter the curve.

Figure 7 indicates that the patient's HRDD value in the early time is pretty high. By using the first chemotherapy method, the HRDD values began to decrease. The system would rather recommend radiotherapy methods than chemotherapy, because it found that chemotherapy is more

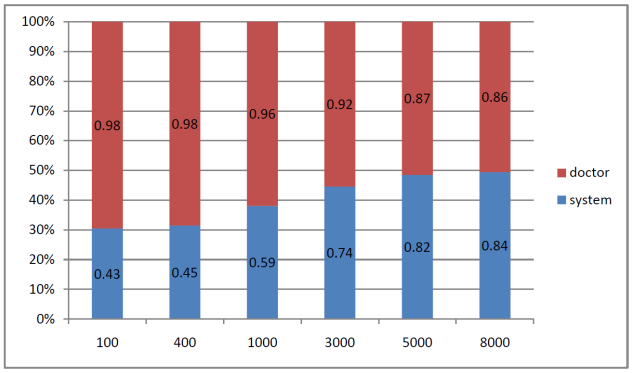


FIGURE 8. The accuracy of the diagnostic auxiliary system.

dangerous to the human body. With two process of radiotherapy treatment, the values of parameter are between 180 to 120. Radiotherapy is an effective treatment which relieved the high mortality rate of NSCLC.

If a patient's HRDD value comes to 71.05, the patient is likely to suffer from the NSCLC phase II. By the surgical method, the focus could be removed directly in this phase. In the figure 7, after resection of the lesion site, HRDD value decreases to 67.25.Inthis situation, the treatment of patient switches into drug.

The calculation of correlation parameter helps doctors to adjust their strategies. Thus, it is possible for medical staff reducing the suffering of patients.

Figure 8 shows the accuracy of the diagnostic auxiliary system. From history data, we want to know whether a patient has NSCLC or not. From this figure, the decisions by doctors are very accurately. With small samples (100-500), the accuracy reaches 98%. In big data samples (over 1,000), the accuracy also reaches 88%.

Diagnostic auxiliary system in small samples displays inaccuracy. The accuracy rate is only 43-59%. If there are not enough training data storing in data base, the result is not assisted by doctors. In big data samples, training data are also increased. The accuracy has improved to over 80% when the diagnosis data reach 5,000. The accuracy has approach for doctor.

However, diagnostic system is only an auxiliary system, it not replaces doctors to make accuracy decisions in NSCLC. Even if we want to system merely judges 'have' or 'not'. But we can adopt diagnostic auxiliary system to assist doctor decrease workload while the training data ever-increasing and the accuracy improving continually.

V. CONCLUSION

This article proposes the foundation of a medical system model considering both probability analysis and decision-making. It is an efficient method to calculate the transition probability of NSCLC in four different stages. In each of the evolutionary process, an effective parameter selecting method based on a set of large data is adopted at the purpose of mining maximum effect of three kinds of correlation parameters. Combined with effective treatment methods to improve the correlation parameters, the probability of malignant disease development will effectively reduce. Statistically, results of experiments proves that the proposed model can offer clinical data, which is fast and accurate for decision-making advice.

In the future, by perfecting various treatment methods and collecting diagnoses, the patient's diagnosis is helpful for deep learning and data mining. They will improve the effect of calculation and provide doctors with suitable diagnostic methods fast. Moreover, probability analysis and decision-making in NSCLC can be adopted to non-small cell carcinoma by computation migration. It is good for doctors to improve diagnostic test accuracy studies.

REFERENCES

[1] J. Wu, Y. Tan, Z. Chen, and M. Zhao, "Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country," *Comput. Methods Programs Biomed.*, vol. 159, pp. 87–101, Jun. 2018. doi: 10.1016/j.cmpb.2018.03.004.

[2] J. Wu, Y. Tan, Z. Chen, and M. Zhao, "Data decision and drug therapy based on non-small cell lung cancer in a big data medical system in developing countries," *Symmetry*, vol. 10, no. 5, p. 152, 2018. doi: 10.3390/sym10050152.

[3] J. L. Sullivan and D. C. Clinical, "Lung cancer screening," *New England J. Med.*, vol. 353, no. 20, pp. 2194–2195, 2005.

[4] F. L. Greene. *American Joint Committee on Cancer, American Cancer Society. AJCC Cancer Staging Manual*. 6th ed. New York, NY, USA: Springer Verlag, 2002.

[5] D. Moro, D. Villemain, J. P. Vuillez, C. A. Delord, and C. Brambilla, "CEA, CYFRA21-1 and SCC in non-small cell lung cancer," *Lung Cancer*, vol. 13, no. 2, pp. 169–176, 1995.

[6] N. Reinmuth *et al.*, "Prognostic impact of Cyfra21-1 and other serum markers in completely resected non-small cell lung cancer," *Lung Cancer*, vol. 36, no. 3, pp. 265–270, 2002.

[7] B. Xu, L. D. Xu, H. Cai, C. Xie, J. Hu, and F. Bu, "Ubiquitous data accessing method in IoT-based information system for emergency medical services," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1578–1586, May 2014.

[8] H. Wang, Q. Zhang, and J. Yuan, "Semantically enhanced medical information retrieval system: A tensor factorization based approach," *IEEE Access*, vol. 5, no. 99, pp. 7584–7593, 2017.

[9] D. G. Pfister *et al.*, "American Society of Clinical Oncology treatment of unresectable non–small-cell lung cancer guideline: Update 2003," *J. Clin. Oncol.*, vol. 22, no. 2, pp. 330–353, Jan. 2004.

[10] Y. Tan *et al.*, "The effect of corticosteroid administration on soft-tissue inflammation associated with rhBMP-2 use in a rodent model of inflammation," *Spine*, vol. 38, no. 10, pp. 806–813, May 2013.

[11] A. Agarwal, C. Baechle, R. Behara, and X. Zhu, "A natural language processing framework for assessing hospital readmissions for patients with COPD," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 588–596, Mar. 2018.

[12] H. Greenspan, B. V. Ginneken, R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[13] J. Wu and Z. Chen, "Sensor communication area and node extend routing algorithm in opportunistic networks," *Peer-Peer Netw. Appl.*, vol. 11, no. 1, pp. 90–100, Jan. 2018. doi: 10.1007/s12083-016-0526-4.

[14] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[15] J. Wu, Z. Chen, and M. Zhao, "Information cache management and data transmission algorithm in opportunistic social networks," in *Proc. Wireless Netw.*, 2018, pp. 1–12. doi: 10.1007/s11276-018-1691-6.

[16] S. Han, W. K. Ng, L. Wan, and V. C. S. Lee, "Privacy-preserving gradient-descent methods," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 6, pp. 884–899, Jun. 2010.

[17] C. Phua, K. Smith-Miles, V. Lee, and R. Gayler, "Resilient identity crime detection," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 533–546, Mar. 2012.

[18] L. W. C. Chan, T. P. Chan, B. T. F. Cheung, K. Mo, K. K. L. Fung, "Simulation, visualization and dosimetric validation of scatter radiation distribution under fluoroscopy settings," *J. Biomed. Eng. Inform.*, vol. 1, no. 1, pp. 93–102, 2015.

[19] L. W. C. Chan *et al.*, "Novel structural co-expression analysis linking the NPM1-associated ribosomal biogenesis network to chronic myelogenous leukemia," *Sci. Rep.*, vol. 5, Jul. 2015, Art. no. 10973.

[20] W. U. Jia, Z. Chen, and M. Zhao, "Effective information transmission based on socialization nodes in opportunistic networks," *Comput. Netw.*, vol. 129, pp. 297–305, Sep. 2017.

[21] J. Wu, Z. Chen, and M. Zhao, "SECM: Status estimation and cache management algorithm in opportunistic networks," in *The Journal of Supercomputing*, 2018, pp. 1–19. doi: 10.1007/s11227-018-2675-0.1

[22] Z. Cheng, M. Subbarayan, X. Chen, and S. S. Gambhir, "Synthesis of (4-[^{18}F]fluorophenyl)triphenylphosphonium as a potential imaging agent for mitochondrial dysfunction," *J. Int. Isotope Soc.*, vol. 48, no. 2, pp. 131–137, Feb. 2005.

[23] J. Wu, Z. Chen, and M. Zhao, "Weight distribution and community reconstitution based on communities communications in social opportunistic networks," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 1, pp. 158–166, Jan. 2019. doi: 10.1007/s12083-018-0649-x.

[24] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, and S. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 714–722, Mar. 2019.

[25] W. J. Allen, R. E. Gabr, G. B. Tefera, A. S. Pednekar, M. W. Vaughn, and P. A. Narayana, "Platform for automated real-time high performance analytics on medical image data," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 318–324, Mar. 2018.

[26] P. Guan and J. Wu, "Effective data communication based on social community in social opportunistic networks," *IEEE Access*, vol. 7, pp. 12405–12414, 2019. doi: 10.1109/ACCESS.2019.2893308.

[27] M. Chen and Y. Hao, "Task offloading for mobile edge computing in software defined ultra-dense network," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 587–597, Mar. 2018.

[28] M. Chen, Y. Hao, K. Lin, L. Hu, and Z. Yuan, "Label-less learning for traffic control in an edge network," *IEEE Netw.*, vol. 32, no. 6, pp. 8–14, Nov. 2018.

[29] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani, "Deep features learning for medical image analysis with convolutional autoencoder neural network," *IEEE Trans. Big Data*, to be published. doi: 10.1109/TBDATA.2017.2717439.

[30] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[31] K. Hwang and M. Chen, *Big-Data Analytics for Cloud, IoT and Cognitive Computing*. London, U.K.: Wiley, 2017.

[32] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, and C.-H. Youn, "5G-smart diabetes: Towards personalized diabetes diagnosis with healthcare big data clouds," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 16–23, Apr. 2018.

JIA WU received the Ph.D. degree in software engineering from Central South University, Changsha, Hunan, China, in 2016, where he is currently an Associate Professor with the School of Software. Since 2010, he has been an Algorithm Engineer with IBM Company, Seoul, South Korea, and also in Shanghai, China. His research interests include wireless communications and networking, wireless networks, big data research, and mobile health in network communication. He is a Senior Member of China Computer Federation, and a member of ACM.

PEIYUAN GUAN received the master's degree from the Software College, Central South University, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree with the Information Science and Engineering College. His research interests include edge computing, the Internet of Things, and game theory. He is a Student Member of CCF.

YANLIN TAN received the Ph.D. degree from The Second Xiangya Hospital, Central South University, China, in 2013. He is currently an Attending Doctor with the PET-CT Center, The Second Xiangya Hospital. His research interests include early diagnosis, clinical staging, and therapeutic effect evaluation of common malignant tumors.