IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 19, 2018, accepted December 6, 2018, date of publication December 12, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886198

Big Medical Data Decision-Making Intelligent System Exploiting Fuzzy Inference Logic for Prostate Cancer in Developing Countries

KANGHUAI LIU[1,2], ZHIGANG CHEN[1,2], (Member, IEEE), JIA WU[1,2], (Member, IEEE), YANLIN TAN[3], LEILEI WANG[1,2], YEQING YAN[1,2], HENG ZHANG[1,2], AND JUNYAO LONG[1,2]

[1]School of Software, 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 of Central South University, Changsha 410083, China

Corresponding authors: Zhigang Chen (czg@csu.edu.cn) and Jia Wu (jiawu5110@163.com)

This work was supported in part by the Major Program of National Natural Science Foundation of China under Grant 71633006, in part by the National Natural Science Foundation of China under Grant 61672540, in part by the China Postdoctoral Science Foundation funded project under Grant 2017M612586, in part by the Postdoctoral Science Foundation of Central South University under Grant 185684, in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2018zzts615, and in part by the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory.

ABSTRACT In most developing countries, it has become a severe challenge for the limited medical resources and outdated healthcare technology to meet the high demand of large population. From the perspective of social development, this unbalanced healthcare system in developing counties has also exacerbated the contradiction between physicians and patients, particularly those suffering from malignant diseases (such as prostate cancer). Rapid improvements in artificial intelligence, computing power, parallel operation, and data storage management have contributed significantly to a credible medical data decision-making on the detection, diagnosis, treatment, and prognosis of malignant diseases. Consequently, to address these existing problems in the current healthcare field of developing countries, this paper proposes a novel big medical data decision-making model exploiting fuzzy inference logic for prostate cancer in developing countries, constructing an intelligent medical system for disease detection, medical data analysis and fusion, treatment recommendations, and risk management. Based on 1 933 535 items of hospitalization information from over 8000 prostate cancer cases in China, the experimental results demonstrate that the intelligent medical system could be adopted to assist physicians and medical specialists in coming up with a more dependable diagnosis scheme.

INDEX TERMS Prostate cancer, fuzzy inference logic, intelligent medical system, big medical data decision-making model, fusion of multimodal medical data, machine-assisted diagnosis.

I. INTRODUCTION

Prostate cancer (PCa) has become the second commonest malignant tumor and the fifth leading cause of high morbidity and mortality in males, and it poses a rising public threat to human beings all over the world [1]. To be specific, in 2012, more than 1.1 million males worldwide were diagnosed with prostate cancer [2]. Additionally, the average morbidity rate of prostate cancer is approximately 11% over a male's lifetime, while the mortality rate from the disease is about 4% [2]. In 2018, people suffering from prostate cancer in Asia account for the half of 18.1 million new cancer cases all over the world [3]. Meanwhile, the risk of prostate cancer has constantly risen to 13.5% which stands at the second highest among all cancers in male [4].

Some developing countries in Asia and Africa, such as China, India, and South Africa, may face the same challenge in social healthcare field that is a deep contradiction between huge population and scarce medical resource (undeveloped medical technology and insufficient public healthcare services) [5]. In particular, the repeatability and complexity of medical diagnosis program, a massive influx of multimodal medical data, and behindhand medical equipment will lead to a high misdiagnosis rate and relatively low diagnosis efficiency of medical staff eventually.

In Beijing, one of China's metropolis, there are over 20 million workers and 10 million children and the aged. However, in such a densely populated city, only no more than 3,000 healthcare personnel can provide healthcare service [6]. Moreover, these healthcare personnel also must deal with hundreds of pathological reports from remote areas with behindhand medical system, because advanced medical devices and excellent physicians are centralized in big cities and first-class hospitals. What's more crucial is that heavy work and mental stress have already constituted a huge burden to those physicians, while thousands of patients are still waiting for malignant tumor diagnosis [7], [8]. Proportionately, in China, over 5000 patients must share one physician, and one general practitioner needs to take care of at least 70 prostate cancer patients a day [9], [10]. Over the whole diagnosis cycle, a mass of multimodal medical data associated with prostate cancer will be extracted, but only 30% of these statistics can be useful for the detection, diagnosis, and prognosis of the disease [11].

Accordingly, there may also exist the same plight in most hospitals in developing countries [12]:

- Due to the existence of a large number of diseased population in most developing countries, some complicated and repetitive work, such as analysis of similar medical reports, images, and biochemical indices, may cause a high misdiagnosis rate and relatively low working efficiency of medical staff.
- In many developing countries, such as China, a high misdiagnosis rate from healthcare personnel has resulted in a serious social contradiction between physicians and patients, particularly those with malignant diseases. Many relatives of patients start doubting excessive consumption coming from physicians' treatment decisions.
- In the current healthcare field of developing countries, there is a crisis of confidence between physicians and patients due to physicians' fruitless efforts to convince the sick people and patients' little knowledge of complicated physiological indices.

To solve these problems, intelligent medical system (expert system) [13]–[15] is gradually employed to improve the healthcare situation in developing countries. With the assistance of intelligent medical system [14], [16], physicians or medical experts are able to combine the auxiliary therapy proposals from machine-assisted model with their own diagnosis experience to make a more reliable treatment decision [17]. In fact, it is a tedious and repetitive task for physicians to assess and determine the clinical stage of prostate cancer by the statistics related to tumor markers in pathological reports. Therefore, these assignments can be accomplished in a machine-assisted system.

Moreover, in the light of machine-assisted system, IoT (internet to things) [18]–[22] can be applied in the healthcare field of developing countries. Hospitals, patients, and physicians would establish an timely and effective medical communication, in which physicians are able to duly submit meaningful diagnosis information and patients can acquire real-time medical reports [23]. As a consequence, intelligent medical system could effectively improve the discordant social relationship between physicians and patients [24], [25].

In this work, we propose a big medical data decision-making intelligent system exploiting fuzzy inference logic [26], [27] for prostate cancer in developing countries. Through manual and automatic weight adjustment, fuzzy inference system employs the fusion of medical information associated with various disease indicators (tumor markers) to analyze and determine the clinical stage of prostate cancer, recommending the appropriate treatment strategies to physicians or medical experts. The combination of machine-assisted diagnosis and artificial judgment could contribute significantly to the final treatment decision of medical staff. To observe the curative powers of the deterministic treatment method on patients in real time, risk management model can be applied to monitor the changes in patients' physiological indicators. Meanwhile, physicians are able to make timely treatment adjustments for prostate cancer, which effectively ensures that the disease remains under control. In conclusion, the contributions of this study are summarized as follows:

(1) A novel prostate cancer detection model based on Mamdani fuzzy inference system is constructed to judge whether a patient suffers from prostate cancer and to determine the clinical stage of the disease.

(2) Combining statistical analysis and medical data decision-making, machine-assisted system is able to provide physicians with the fast and accurate curative options automatically.

(3) According to the physiological index comparison between different diagnostic intervals, the progression of prostate cancer can be monitored in real time, and physicians are able to assess the curative effect of the deterministic therapeutic scheme on patients.

(4) To ensure the feasibility and accuracy of big medical data decision-making, the experiment in this study is based on 1,933,535 items of medical Hospitalization information from three hospitals in China. Experimental results demonstrate that intelligent medical system could effectively enhance physicians' working efficiency and reduce their misdiagnosis rate.

The remainder of this paper is organized as follows: In Section II, we present the state of the art related to our works; Moreover, big medical data decision-making model will be proposed and constructed in Section III; In Section IV, the detailed description of experimental performance is provided; Finally, the conclusion of this paper is shown in the last section.

II. RELATED WORKS

In recent years, many intelligent medical systems are gradually applied in the auxiliary detection, diagnosis, treatment, and prognosis of a disease, which shows that big medical data decision-making model has also become a hot research issue in healthcare application and intelligent system fields.

Next, we will give a detailed introduction to the research status related to our works.

The use of fuzzy inference logic for disease detection and diagnosis has become increasing popular. However, most existing disease detection methods based on fuzzy inference logic present a lack of sufficient medical data information and do not assign reasonable weights to evaluation indexes [24], [31], [32] Specifically, Caifeng and Deng [24] proposed an innovative intelligent diagnosis system using a fuzzy concept lattice. In the system, the process of retrieving medical information and diagnosing disease is performed using fuzzy concept lattices based on rules of clinical diagnosis. By extracting symptoms and the extents of diseases, and computing the largest similarity of the fuzzy sets of patients' symptoms, a fuzzy formal context can be adopted to determine the matching disease patients are most likely to suffer from.

Tsipouras *et al.* [31] presented a novel framework for Cardiovascular diseases on the basis of the beat classification and fuzzy inference logic. In the fuzzy expert system, defined by medical experts, the initial set of rules for Cardiovascular diseases is automatically transformed into a fuzzy inference model. Then, using the methodology of data mining, the effective information can be used to determine the basic setting of fuzzy model. Additionally, an efficient optimization strategy in the system is formed based on the first derivate information and the gradient descent method. Obajemu *et al.* [32] recommended a novel fuzzy modeling framework for risk grouping of the patients with bladder cancer by collecting and assessing the censored survival database. Based on the type-2 fuzzy inference modeling method and tow databases of bladder cancer patients (real life and manually produced datasbases), the risk score and prognostic indicators of the disease can be precisely predicted, and then the system automatically provides doctors with therapeutic recommendations and effective risk management of the disease at different stages.

Moreover, medical assist system based on machine learning model [17], [33], [34] is also widely adopted to strengthen the control and management of high-risk diseases. Compared with these existing schemes based on machine learning, this work creatively adopts iterative computation of Mamdani minimizing operation to implement the risk management of prostate cancer. Litjens *et al.* [17] presented a fully automated machine-assisted detection methodology for Prostate cancer in magnetic resonance imaging (MRI). This is the first Prostate cancer MRI system that is dedicated to compare the expected performance of radiologists with magnetic resonance images of each patient, and then the intelligent system employs the ROC (response receiver operating characteristic curve) and FROC (free-response receiver operating characteristic curve) of per-patient to verify the status of 347 Prostate cancer patients, so as to show the real significance of the two-phase scheme including both voxel and candidate classification.

Su *et al.* [33] recommended a medical diagnosis system based on a jacobian-matrix machine learning model (JMLM). In the system, the approximation features of the expansion form of linear Taylor equation is the basis of obtaining learning ability by JMLM model. Furthermore, the machine learning structure in JMLM can be gradually established when it satisfies the performance criterion of making the accurate disease detection and diagnosis. Zhennao *et al.* [34] proposed a novel prediction framework model for parkinson's disease (PD) base on an optimal support vector machine (SVM). To determine the optimal parameter indicators for the diagnosis process of parkinsonrs disease, a new-style swarm intelligence method can be used to maximize the generalization ability of SVM classifier in the system, whereupon the proposed framework model could be regarded as a feasible support instrument for PD decisions through sustaining optimization of disease parameters.

Additionally, for high-risk disease like prostate cancer and lung disease, in order to guarantee the accuracy and effectiveness of the diagnostic strategy provided by machine-assisted system, a large of volume of multi-format medical data from examination reports, disease imaging, doctors' records and wearable devices, have become the most valuable evidence for social healthcare [23], [35]–[38]. Sahoo *et al.* [35] presented a predication approach for future health development based on medical big data. A novel patient data collection mechanism is established and the corresponding correlation evaluation is processed using an analytic application of cloud-enable diagnostic big data, and then the future health condition of patients can be used to construct the stochastic prediction model in the cloud environment. Han *et al.* [36] proposed a public database of lung imaging signs for disease detection, diagnostic methods, and medical treatments (LISS). Corresponding to 9 types of CT scanning images of common lung disease, the publicly available database consists of 271 CT scanning images, where 677 abnormal areas are assessed, tested and recorded. According to the theory of radiation medicine, the 2-D and 3-D scanning patterns may be applied in different categories of CT imaging sign collection of abnormal areas. On the basis of the big data cloud built by LISS, the detection, treatment and prediction of lung disease are able to obtain an impeccable document and a reasonable reference system.

Furthermore, the application of medical big data also involves the appropriate classification of medical performance and the security protection of private medical information. Wang *et al.* [37] recommended an assignment strategy of diagnosis code exploiting sparsity-based disease correlation embedding. According to an open source large-scale database from patients in intensive care unit, the medical information is extracted to establish a bag-of-words model in the system, and then the proposed sparsity regulation algorithm could be adopted to acquire and construct disease correlations. Ultimately, the classification process of multi-label property is improved significantly by obtaining the disease correlations and globally optimizing the framework.

Hamid *et al.* [38] presented a security model for the private information of medical big data based on a fog computing

mechanism in healthcare data environment. This mechanism generates two different photo galleries via the multimedia data of users within the healthcare cloud. Any access by the user is automatically verified through paring cryptography, and when some unauthorized behaviors are discovered, the user cannot continue to execute the operations related to private diagnosis information, which makes the original multimedia data more secure and invisible in the medical big data environment. Therefore, to ensure the feasibility and accuracy of medical data decision-making and analysis, the experiment in our works is performed in a big medical data environment. When the size of data samples is large enough, big medical data decision-making model shows a relatively high degree of accuracy in the detection of prostate cancer.

III. SYSTEM MODEL DESIGN

Intelligent medical system, a novel applied technique with its foundation in machine learning, is emerging in recent years as an auxiliary diagnosis application for medical information fusion, promising to ameliorate the current healthcare situation of developing countries. With the rapid development of artificial intelligence technology, machine-assisted system is gradually employed to organize, mine, extract, and fuse high and diverse amount of medical information in many developing countries [14], [16]. This effectively assists physicians or medical experts in understanding the state of an illness and making a more accurate and reasonable treatment decision. Therefore, we establish a novel decision-making model exploiting fuzzy inference logic in big medical data environment for prostate cancer in developing countries.

A. THE OVERALL FRAMEWORK OF MEDICAL DECISION-MAKING MODEL FOR PROSTATE CANCER IN BIG DATA ENVIRONMENT

In big medical data environment, the overall flow chart of medical data analysis and decision-making can be established through five phases: tumor marker selection, data preprocessing, machine learning model (MLMs) reconstitution, data training, and data decision-making. Because of the independence and concurrency of the operational process in intelligent hospital system, multiple patients could be simultaneously detected and diagnosed by the system. Moreover, the detailed description of each phase (shown as Fig. 1) in the system is demonstrated as follows:

The detection and diagnosis of prostate cancer is commonly first suspected when a screening test such as a digital rectal examination (DRE) or PSA testing are abnormal. Moreover, further diagnosis tests, such as a magnetic resonance imaging (MRI) fusion biopsy, PCA3 RNA test, computed tomography (CT) scan, random 12-core biopsy, and bone scan can be used to accurately assess the aggressiveness of the tumor and stage the disease [5]. Among the data obtained from these detection tests, tumor marker must be the most valuable signal for the diagnosis, treatment, and prognosis of the disease. Therefore, as shown in Fig.1, in the first phase of machine-assisted system, six different disease indicators (tumor markers) could be adopted to evaluate the clinical stage of prostate cancer: Prostatic Acid Phosphatase (PAP), Prostate Membrane Antigen (PSMA), Total Prostate-Specific Antigen (TPSA), Free Prostate-Specific Antigen (FPSA), Hemoglobin (Hb), and Red Blood Cell (RBC) [8]. Since each disease indicator plays a different role in the diagnosis of prostate cancer, the data associate with disease indicators are reasonably divided into two categories: Key Parameters (KP) and General Parameters (GP). A huge fluctuation in KP means the underlying major changes in patients' condition, while GP can only be regarded as an auxiliary reference for medical data analysis and decision-making of the disease.

Furthermore, in the second phase, the digital data, obtained from prostate cancer patients, will be properly preprocessed from irregular, unreasonable and invalid to standard and legal formats, which can be recognized by MLMs. After that, KP and GP are respectively loaded into different data models $M(t)$ and $C(t)$, which are two mathematical sets used to store digitized medical information.

Reasonable weight evaluation and allocation for disease indicators are implemented in the third phase. The weights w_{TPSA}, w_{FPSA}, w_{RBC}, w_{Hb}, w_{PAP}, and w_{PSMA} shown as Fig.1 can be determined after automatical and artificial parameter tuning. After that, the primary and secondary disease correlations $PC(t)$ and $SC(t)$ are used as the fuzzy input of machine-assisted system, which demonstrates how much disease indicators affect the clinical staging of prostate cancer.

Moreover, medical data fusion for prostate cancer is processed in the fourth phase. The staging of prostate cancer provides significant signals about the extent of the disease in the body and the selection of curative options. According to the guidelines of Tumor Node Metastasis (TNM), there are four main stages (I, II, III, IV) in prostate cancer: preliminary, early, middle, and terminal [2]. The staging of prostate cancer given by intelligent medical system is only based on the statistic data associated with disease indicators. By defining membership functions and fuzzy sets, six different membership degrees are obtained from the primary and secondary disease correlations $PC(t)$ and $SC(t)$, and each of these degrees corresponds a membership subset (*Low*, *Medium*, or *High*). As shown in Fig.1, nine different fuzzy combinations could be generated from six different membership degrees in the system. According to 'If-Then' rules, each fuzzy combination between membership degrees $D(PC(t))$ and $D(SC(t))$ is applied to comprehensively evaluate the clinical stage of prostate cancer.

Eventually, as shown in Fig. 1, medical data decision-making and risk management for prostate cancer are performed in the fifth phase of the system. According to the iterative calculation of Mamdani minimizing operation [26], machine-assisted system makes adjustments to PEV_{PCa} and treatment recommendations at different diagnosis intervals. By comprehensively evaluating and analyzing

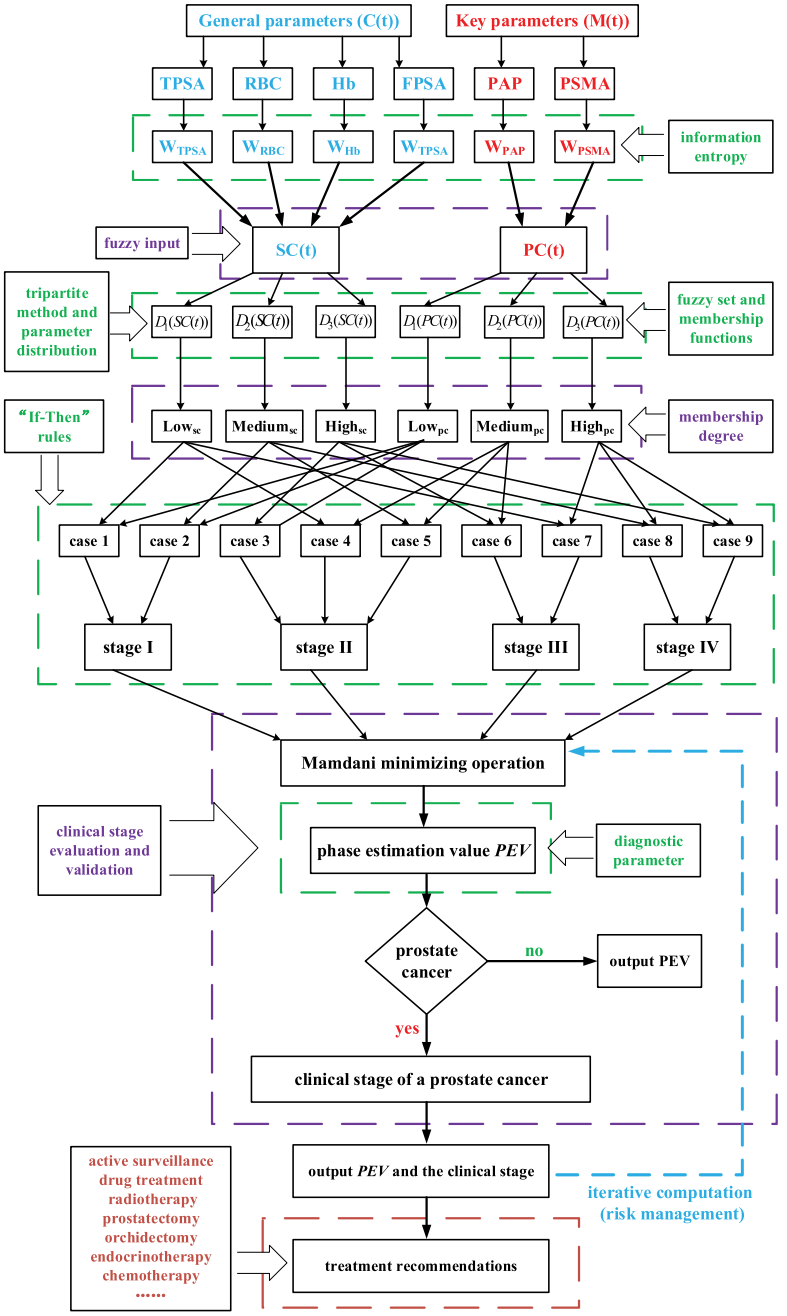


FIGURE 1. The overall structure of big medical data decision-making model in intelligent medical system.

medical reports, scanning images, clinical symptoms of patients, the auxiliary diagnosis result from machine-assisted system, etc, physicians and medical experts will determine the stage of prostate cancer and work on an more accurate therapeutic strategy.

B. THE MEDICAL DECISION-MAKING MODEL OF MACHINE-ASSISTED SYSTEM FOR PROSTATE CANCER IN BIG DATA ENVIRONMENT

There is no doubt that the treatment strategy developed by physicians is of great significance for the rehabilitation of prostate cancer patients. However, due to the serious contradiction between large population and limited medical resources, a diagnosis scheme, which is devised for prostate cancer patients, sometimes is not therapeutic enough in most developing countries [5], [6]. Consequently, it is indispensable for physicians to improve their diagnosis efficiency and accuracy with the assistance of artificial intelligence system.

In medical decision-making model, to come up with an effective diagnosis scheme and grasp the optimal treatment period, we define t and $F_{PCa}(t)$ as the current diagnosis interval and auxiliary diagnosis function for prostate cancer,

respectively. To quantify the whole process of auxiliary diagnosis in machine-assisted system, we formalize $F_{PCa}(t)$ as

$$F_{PCa}(t) = FIL(C(t), M(t)) \tag{1}$$

where $C(t)$ and $M(t)$ represent the corresponding model input of GP and KP, respectively. Moreover, FIL denotes the process of Mamdani fuzzy inference for $PC(t)$ and $SC(t)$. According to different performances of disease correlations on medical data decision-making of prostate cancer, weight allocation and machine learning model binning in intelligent medical system mainly contains two portions: $C(t)$ and $M(t)$ (shown as Fig. 1). Therefore, the corresponding data models $C(t)$ and $M(t)$ are defined as

$$\begin{cases} C(t) = \{TPSA, FPSA, Hb, RBC\} \\ M(t) = \{PAP, PSMA\} \end{cases} \tag{2}$$

With the change of two parameter models $C(t)$ and $M(t)$, a fluctuation in the auxiliary diagnostic function $F_{PCa}(\varepsilon)$ may imply different clinical stages of prostate cancer. To prevent the deterioration of prostate cancer over the whole diagnosis cycle, the related mathematical inference process in machine-assisted system is shown as

$$\begin{cases} F_{PCa}(\varepsilon) \geq F_{PCa}(\varepsilon + 1) \\ \qquad\qquad\qquad\quad \varepsilon \in [0, T] \\ F_{PCa}(\varepsilon) \geq F_{PCa}(\varepsilon - 1) \end{cases} \tag{3}$$

where T represents the entire diagnosis cycle for prostate cancer and ε indicates the most severe period of the disease. In addition, intelligent medical system can automatically provide different treatment recommendations for physicians based on the clinical stage of prostate cancer it determines. Meanwhile, risk management mechanism will be adopted by physicians to measure the curative effect of the treatment method on prostate cancer.

For the sake of rational parameter allocation and indicator scheduling, the primary and secondary disease correlations $PC(t)$ and $SC(t)$, defined by six disease indicators (tumor markers), can be respectively formalized as

$$\begin{cases} SC(t) = W_{GP}C(t) \\ PC(t) = W_{KP}M(t) \end{cases} \tag{4}$$

where W_{GP} and W_{KP} represent the weight adjusters for general and key parameters, respectively. Moreover, when prostate cancer is at different stages, patients may represent different clinical manifestations, thus the system voluntarily triggers the weight adjusters W_{GP} and W_{KP} to rationalize disease correlations on the basis of the application of information entropy method.

As a consequence, the primary and secondary disease correlations $PC(t)$ and $SC(t)$ can be expanded as Eq.(5), as shown at the bottom of this page, where $\sum_{i=1}^{\partial} \frac{\delta(i)}{\partial}$ represents the average value of each disease indicator in the past ∂ years and δ is the current value of each disease indicator from prostate cancer patients. Besides, w means the weight value that set by weight adjusters W_{GP} and W_{KP} for different disease indicators.

C. REASONABLE WEIGHT EVALUATION AND ALLOCATION FOR DISEASE INDICATORS

Throughout the whole process of auxiliary diagnosis and risk management, the purpose of weight adjustment and allocation is to match each disease indicator with a reasonable weight. Based on physicians' diagnosis experience and data training results, machine-assisted system automatically triggers the weight adjusters W_{GP} and W_{KP} to allocate suitable weights to tumor markers so that a more effective curative proposal can be recommended to physicians or medical specialists.

As the above Eq.(5) shows, the fused medical information between w and $\delta / \sum_{i=1}^{\partial} \frac{w(i)}{\partial}$ can be regarded as a disease indicator item $w \times \delta / \sum_{i=1}^{\partial} \frac{w(i)}{\partial}$. To determine each disease indicator item for prostate cancer, an improved entropy assessment method [28], [29] is employed to calculate weight of each disease indicator item. Firstly, we construct a disease indicator array TM as

$$\begin{aligned} TM &= \{C(t), M(t)\} \\ &= [TPSA, FPSA, Hb, RBC, PAP, PSMA] \end{aligned} \tag{6}$$

Then, utilizing the legitimate data after the preprocessing, the weight measurement matrix W^*_{TM} can be expressed as Eq.(7), as shown at the top of the next page, where $\delta_{(Pa_m,TPSA_m)}$, $\delta_{(Pa_m,FPSA_m)}$, $\delta_{(Pa_m,RBC_m)}$, $\delta_{(Pa_m,Hb_m)}$, $\delta_{(Pa_m,PAP_m)}$, and $\delta_{(Pa_m,PSMA_m)}$ denote the average contribution degree (average value) of TPSA, FPSA, RBC, Hb, PAP, and PSMA in the $m-th$ patient Pa_m during the whole diagnosis cycle, respectively. Additionally, m is the total number of the prostate cancer patients who have been diagnosed by machine-assisted system during the past ∂ years. Therefore, the solo contribution degree $CD_{(Pa_i,TM(j))}$ of the $j-th$ disease

$$\begin{cases} SC(t) = w_{TPSA} \times \dfrac{\delta_{TPSA}}{\sum_{i=1}^{\partial} \frac{\delta_{TPSA}(i)}{\partial}} + w_{RBC} \times \dfrac{\delta_{RBC}}{\sum_{i=1}^{\partial} \frac{\delta_{RBC}(i)}{\partial}} + w_{Hb} \times \dfrac{\delta_{Hb}}{\sum_{i=1}^{\partial} \frac{\delta_{Hb}(i)}{\partial}} + w_{FPSA} \times \dfrac{\delta_{FPSA}}{\sum_{i=1}^{\partial} \frac{\delta_{FPSA}(i)}{\partial}} \\ PC(t) = w_{PAP} \times \dfrac{\delta_{PAP}}{\sum_{i=1}^{\partial} \frac{\delta_{PAP}(i)}{\partial}} + w_{PSMA} \times \dfrac{\delta_{PSMA}}{\sum_{i=1}^{\partial} \frac{\delta_{PSMA}(i)}{\partial}} \end{cases} \tag{5}$$

$$W^*_{TM} = \begin{bmatrix} \delta_{(Pa_1,TPSA_1)} & \delta_{(Pa_1,FPSA_1)} & \delta_{(Pa_1,RBC_1)} & \delta_{(Pa_1,Hb_1)} & \delta_{(Pa_1,PAP_1)} & \delta_{(Pa_1,PSMA_1)} \\ \delta_{(Pa_2,TPSA_2)} & \delta_{(Pa_2,FPSA_2)} & \delta_{(Pa_2,RBC_2)} & \delta_{(Pa_2,Hb_2)} & \delta_{(Pa_2,PAP_2)} & \delta_{(Pa_2,PSMA_2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \delta_{(Pa_m,TPSA_m)} & \delta_{(Pa_m,FPSA_m)} & \delta_{(Pa_m,RBC_m)} & \delta_{(Pa_m,Hb_m)} & \delta_{(Pa_m,PAP_m)} & \delta_{(Pa_m,PSMA_m)} \end{bmatrix} \tag{7}$$

indicator of the $i-th$ patient can be computed by

$$CD_{(Pa_i,TM(j))} = \frac{\delta_{(Pa_i,TM(j))}}{\sum_{i=1}^{m} \delta_{(Pa_i,TM(j))}} \tag{8}$$

To synthetically determine the impact of the $i-th$ disease indicator on the m patients who have been diagnosed by intelligent medical system in the past ∂ years, we rigorously define $E_{TM(j)}$ as the total contribution degree of the $j-th$ disease indicator of the m patients, and $E_{TM(j)}$ can be calculated by

$$E_{TM(j)} = -\frac{1}{\ln(m)} \sum_{i=1}^{m} CD_{(Pa_i,TM(j))} \ln(CD_{(Pa_i,TM(j))}) \tag{9}$$

Apparently, the range of $E_{TM(j)}$ is from 0 to 1. In general, when $E_{TM(j)}$ is relatively large, physiological indicators in patients are disordered and prostate cancer is at more serious stages. On the contrary, if the value of $E_{TM(j)}$ is small and stabilized, which implies the disease may be in a recovery period and patients' condition tends to be normal.

Consequently, the differences of contribution degrees between different clinical stages could be adopted to evaluate the weight of each disease indicator. To measure the impact of each disease indicator on the clinical staging of prostate cancer, we first define $d_{TM(j)}$ as the consistent contribution degree of each patient on the $j-th$ disease indicator, which is expressed as $d_{TM(j)} = 1 - E_{TM(j)}$. On the basis of the evaluation and allocation of weight adjusters W_{GP} and W_{KP}, the weight value of the $j-th$ disease indicator can be expressed as

$$w_{TM(j)} = \frac{d_{TM(j)}}{d_{TPSA} + d_{FPSA} + d_{RBC} + d_{Hb} + d_{PAP} + d_{PSMA}} \tag{10}$$

where $d_{TPSA} + d_{FPSA} + d_{RBC} + d_{Hb} + d_{PAP} + d_{PSMA}$ is the total consistent contribution degree of six disease indicators of m patients during the past ∂ years.

Sometimes, the weight value set by machine-assisted system for disease indicators may not be reliable enough. Consequently, based on physicians' diagnosis experience and medical information analysis, some subjective corrective parameters will be used to correct the weight of each disease indicator. According to mathematical theory, the determinate weight value $W^*_{TM(j)}$ is strictly formalized as Eq.(11), as shown at the bottom of the next page, where $\lambda_{TM(j)}$, λ_{TPSA}, λ_{FPSA}, λ_{RBC}, λ_{Hb}, λ_{PAP}, and λ_{PSMA} are defined as the manually amendment variables of $TM(j)$, TPSA, FPSA, RBC, Hb, PAP and PSMA, respectively.

After comprehensively taking into consideration physicians' diagnosis settings and data training results, the system automatically and reasonably matches each disease indicator to its corresponding weight value. Furthermore, weight error analysis could be used to assess the impact of weight values on medical data decision-making of prostate cancer, thereby further correcting the weight of each disease indicator reasonably. Eventually, the weights $w_{TPSA}, w_{FPSA}, w_{RBC}, w_{Hb}, w_{PAP}$, and w_{PSMA} in Eq.(5) can be determined after automatical and artificial parameter tuning.

D. BIG MEDICAL DATA FUSION AND DECISION-MAKING FOR PROSTATE CANCER IN INTELLIGENT FUZZY INFERENCE SYSTEM

Low-grade prostate cancer commonly don't present any clinical symptoms, so most prostate cancer cases have already been in a high-grade once the disease is detected. It's quite likely that the data associated with tumor markers show a normal distribution [5], [8], [11]. Specifically, the vast majority of patients are in the stage III or IV of prostate cancer, while the number of patients in the stage I or II of the disease are relatively small. In addition, disease indicators are constantly changing during the whole diagnostic cycle, hence it may be inaccurate to determine the clinical stage of prostate cancer via the fixed value of tumor marker at a certain time. This also leads to a large deviation in auxiliary diagnosis decision-making for prostate cancer. Consequently, by analyzing the distribution of the statistics related to tumor markers and determining the relationship between clinical staging and disease indictors, fuzzy inference logic [26], [27] could be applied to assess and validate the clinical stage of prostate cancer.

Derived from the mathematical theory of engineering control, fuzzy inference logic can demonstrate the vague clinical staging of prostate cancer by transforming piecewise functions into curves. In intelligent medical system, Mamdani fuzzy model [27], a widely used technique in fuzzy inference logic, could be used as a medical data decision-making tool for prostate cancer in big data environment due to its extensive applicability [26], [27]. In general, a Mamdani fuzzy inference model can be implemented using three interlocking components: Fuzzifier, Fuzzy Inference, and Defuzzifier. In the following sections, we will give a detailed description for the assessment and confirmation of clinical staging of prostate cancer.

1) DETERMINING MEMBERSHIP FUNCTIONS AND MEMBERSHIP DEGREES FOR DISEASE CORRELATIONS *PC*(*T*) AND *SC*(*T*)

In the estimation process of clinical stage of prostate cancer, $PC(t)$ and $SC(t)$ are regarded as the fuzzy input of Mamdani fuzzy inference system (shown as Fig. 1). These two disease correlations are included into an array $AD_{PCa}(i)$. In addition, we simultaneously define $MF_{PCa}(i)$ as the domain of discourse for fuzzy sets (*Low*, *Medium*, and *High*), thus the related initialization definitions can be formalized as

$$AD_{PCa}(i) = \begin{cases} PC(t) & i = 1 \\ SC(t) & i = 2 \end{cases} \tag{12}$$

$$\begin{aligned} MF_{PCa}(AD_{PCa}(i)) = \{&D_{Low}(AD_{PCa}(i)), \\ &D_{Medium}(AD_{PCa}(i)), \\ &D_{High}(AD_{PCa}(i))\} \end{aligned} \tag{13}$$

where $D_{Low}(AD_{PCa}(i))$, $D_{Medium}(AD_{PCa}(i))$, $D_{High}(AD_{PCa}(i))$ corresponding to three levels of fuzzy sets (*Low*, *Medium*, and *High*) are membership degrees for disease correlations. In other words, each disease correlation corresponds three different levels of fuzzy sets, and each fuzzy set in the fuzzier component can be computed by its corresponding membership function.

Consequently, it is necessary for us to define three reasonable membership functions for disease correlations in intelligent medical system. In general, there are many membership function paradigms in engineering control mathematics, such as orthogon, trapezoidal, triangular, or inverse [26], [27]. Because of the normal distribution of the statistics related to disease indicators [5], tripartite method [8], [11] can be applied to define three different membership functions, which aims to clarify the random interval of disease indicators.

In the system, $D_{Low}(AD_{PCa}(i))$, $D_{Medium}(AD_{PCa}(i))$, $D_{High}(AD_{PCa}(i))$ respectively represent *low*, *medium*, and *high* correlations between clinical staging and disease indicators. According to the concept of the tripartite method, the geometric graph of the three membership functions is shown as Fig.(2). Because patients in various clinical stages of prostate cancer may show different physiological indicators, each data training can determine a partition combination that consists of two different demarcation values (μ, η), where μ is the demarcation point between *low* and *medium* disease correlations and η represents the demarcation point between *medium* and *high* disease correlations. Then, the partition combination (μ, η) can be expressed as

$$\begin{cases} \mu \sim N(d_1, \sigma_1^2) \\ \eta \sim N(d_2, \sigma_2^2) \end{cases} \tag{14}$$

where d_1 and d_2 are two critical parameters between different disease correlation degrees, and σ_1 and σ_2 are scale

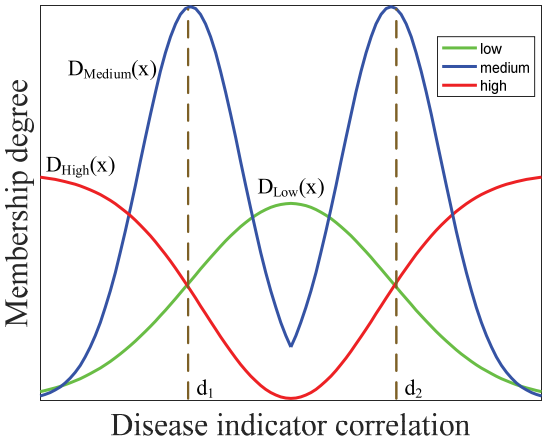


FIGURE 2. Three different membership functions for disease correlations.

parameters for the changes in disease indicators. Moreover, the mathematical mapping Eq.(15) and the mathematical expression of membership degree Eq.(16) can be defined by two demarcation values (μ, η).

$$E_{\mu,\eta}(AD_{PCa}(i)) : U \to \{D_{Low}, D_{Medium}, D_{High}\} \tag{15}$$

$$\begin{aligned} &E_{\mu,\eta}(AD_{PCa}(i)) \\ &= \begin{cases} D_{Low}(AD_{PCa}(i)) & AD_{PCa}(i) \le \mu \\ D_{Medium}(AD_{PCa}(i)) & \mu < AD_{PCa}(i) < \eta \\ D_{High}(AD_{PCa}(i)) & AD_{PCa}(i) > \eta \end{cases} \end{aligned} \tag{16}$$

The probability of each disease indicator falling into a random interval within its corresponding fluctuation range is equal due to the independence between tumor markers. Moreover, if fuzzy inputs (disease correlations) gradually increase from normal to abnormal ranges, the corresponding changes in membership degrees may signify a stage transition of prostate cancer, which also reveals a benign or malignant development of the disease. Therefore, we assume that $P_\mu(x)$ and $P_\eta(x)$ represent the probability density of the random variables μ and η, respectively. The distribution function of disease correlations can be shown as

$$\begin{cases} D_{Low}(AD_{PCa}(i)) = P\{AD_{PCa}(i) \le \mu\} \\ \qquad\qquad = \int_{AD_{PCa}(i)}^{+\infty} P_\mu(x)dx \\ D_{Medium}(AD_{PCa}(i)) = 1 - D_{Low}(AD_{PCa}(i)) \\ \qquad\qquad - D_{High}(AD_{PCa}(i)) \\ D_{High}(AD_{PCa}(i)) = P\{\eta < AD_{PCa}(i)\} \\ \qquad\qquad = \int_{-\infty}^{AD_{PCa}(i)} P_\eta(x)dx \end{cases} \tag{17}$$

Through rigorous mathematical manipulation, the above Eq.(17) can be transformed to a relatively simple form, which

$$w^*_{TM(j)} = \frac{\lambda_{TM(j)} w_{TM(j)}}{\lambda_{TPSA} w_{TPSA} + \lambda_{FPSA} w_{FPSA} + \lambda_{RBC} w_{RBC} + \lambda_{Hb} w_{Hb} + \lambda_{PAP} w_{PAP} + \lambda_{PSMA} w_{PSMA}} \tag{11}$$

TABLE 1. 'If-Then' rules for staging assessment of prostate cancer.

Rule No.	If (And)		Then (clinical stage)
	$D(PC(t))$	$D(SC(t))$	
1	Low	Low	stage I (preliminary stage)
2	Low	Medium	
3	Low	High	stage II (early stage)
4	Medium	Low	
5	Medium	Medium	
6	Medium	High	stage III (middle stage)
7	High	Low	
8	High	Medium	stage IV (terminal stage)
9	High	High	

is expressed as

$$\begin{cases} D_{Low}(AD_{PCa}(i)) = 1 - \Phi(\dfrac{AD_{PCa}(i) - a_1}{\sigma_1}) \\ D_{Medium}(AD_{PCa}(i)) = \Phi(\dfrac{AD_{PCa}(i) - a_1}{\sigma_1}) \\ \qquad\qquad -\Phi(\dfrac{AD_{PCa}(i) - a_2}{\sigma_2}) \\ D_{High}(AD_{PCa}(i)) = \Phi(\dfrac{AD_{PCa}(i) - a_2}{\sigma_2}) \end{cases} \tag{18}$$

where $\Phi(x)$ is a probability density function of the normal distribution for disease indicators and it can be calculated by

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-(\frac{t^2}{2})} dt \tag{19}$$

In the fuzzifier component, according to the membership function set by manual experience and data training, the system allocates three different membership degrees (*Low*, *Medium*, and *High*) for each disease correlation ($PC(t)$ or $SC(t)$). After multiple data training and parameter adjustment, the critical and scale parameters can be tuned to the optimum values, which allows the random interval of disease indicators to clearly reflect the current clinical stage of prostate cancer.

2) FUZZY INFERENCE OF CLINICAL STAGING FOR PROSTATE CANCER BASED ON 'IF-THEN' RULES

In Mamdani fuzzy inference system, by defining membership functions and fuzzy sets, six different membership degrees are obtained from the primary and secondary disease correlations $PC(t)$ and $SC(t)$, and each of those degrees corresponds a membership subset (*Low*, *Medium*, or *High*). As shown in Table 1, nine different fuzzy combinations could be generated from six different membership degrees in the system. According to 'If-Then' rules, each fuzzy combination between membership degrees $D(PC(t))$ and $D(SC(t))$ is applied to comprehensively evaluate the clinical stage of prostate cancer. Therefore, various fuzzy combinations between membership degrees $D(PC(t))$ and $D(SC(t))$ roughly reflect different clinical stages of prostate cancer (shown as Fig.(1)).

During machine-assisted diagnosis process, the primary correlation $PC(t)$, defined by disease indicators PAP and PSMA, dominates the entire diagnostic cycle of prostate cancer. In contrast, the secondary correlation $SC(t)$, including disease indicators TPSA, FPSA, RBC, and Hb, is just an auxiliary reference for the estimation of clinical staging of the disease. With the increase of membership degrees of disease correlations, patients' physiological indicators may be disordered and their condition is more deteriorated. Consequently, fuzzy output of the system is also a fuzzy set that belongs to a special domain of discourse $OM_{PCa}(x)$, which can be formulated as

$$OM_{PCa}(x) = \{stageI, stageII, stageIII, stageIV\} \tag{20}$$

Therefore, according to the transformation of fuzzy relation R_{PCa}, the stage estimation process of prostate cancer in machine-assisted system is expressed as

$$R_{PCa} = (A \cap E) \longrightarrow B \tag{21}$$

If $PC(t)$ is equal to A and $SC(t)$ is E, then $OM_{PCa}(x)$ is equivalent to B. Because each disease indicator obtains different normal ranges in clinical characterization, the primary and secondary disease correlations belong to different domains of discourse in fuzzy inference system. To consider all of the pairing between $PC(t)$ and $SC(t)$, the fuzzy relation R_{PCa} can be calculated by

$$R_{PCa} = \overrightarrow{(A^T \wedge E)} \wedge B \tag{22}$$

where $\overrightarrow{(A^T \wedge E)}$ stands for arranging $A^T \wedge E$ line by line as a vector. Through testing and training large amounts of data from patients, A and E become a vector including much information associated with disease indicators, hence the relationship between disease correlations and phase judgment is determined through fuzzy inference logic after data training and manual adjustment.

3) MAMDANI MINIMIZING OPERATION FOR THE EVALUATION AND VALIDATION OF STAGING OF PROSTATE CANCER

Ordinarily, the application of 'If-Then' rules just provides a preliminary and roughly stage judgment for prostate cancer. Consequently, the purpose of defuzzifier component is to convert the fuzzy phase judgment to a specific evaluation value. Traditional Mamdani minimizing operation [26], [27], a commonly used method in the defuzzifier component, consists of two specific steps: AND and OR operations. The purpose of OR operation is to magnify the impact of each disease correlation ($PC(t)$ or $SC(t)$) on clinical staging of prostate cancer. In contract, the AND operation is aiming to minimize the fused impact of each disease indicator's maximum influence on clinical staging of prostate cancer.

To be specific, in the process of OR operation, the primary and secondary disease correlations respectively correspond three different membership degrees (*Low*, *Medium*, and *High*), each of which can be maximized to a shaded area that is constructed by graph and coordinate axes. This means six different maximum shadow areas could be generated in

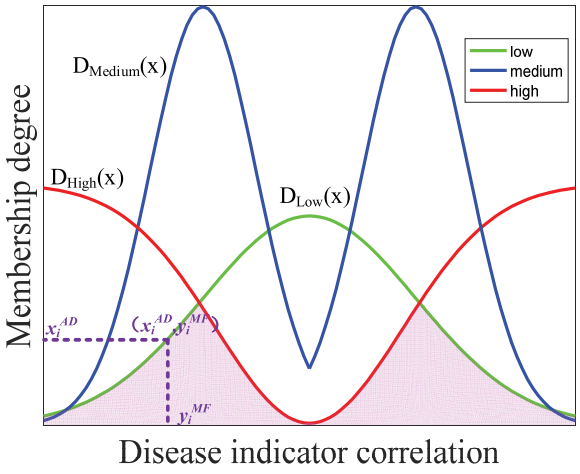


FIGURE 3. The controlling result of three membership functions for disease correlations.

the system. Moreover, to organically combine the effects of multiple disease indicators, the AND operation executes the process of overlapping these six maximum shadow areas to obtain the final controlling result CR_{PCa} for the validation of current clinical stage of prostate cancer.

In realistic disease diagnosis, there is no doubt that digital expression is the most accurate reference for the staging of prostate cancer. Therefore, in order to digitize the final controlling result, the centroid method can be adopted to gain phase estimation value of prostate cancer PEV_{PCa}. Firstly, according to the formalized definition of theory of control mathematics, the whole fuzzy inference process for stage estimation of prostate cancer can be expressed as

$$OM_{PCa}(x) = AD_{PCa}(i) \circ R_{PCa} \tag{23}$$

Then, the final controlling result CR_{PCa} of three membership functions for disease correlations can be calculated by Mamdani minimizing operation, and it is formalized as

$$CR_{PCa} = \int_{AD_{PCa} \times MF} \frac{AD_{PCa}(i) \wedge E_{\mu,\eta}(AD_{PCa}(i))}{(m_{AD}, n_{MF})} \tag{24}$$

where (m_{AD}, n_{MF}) is the ordered pair set between disease correlations and membership functions. For example, if the overlap area of six different maximum shadow is shown as the pink area in Fig.(3), then the final controlling result is evolved into the centroid of the pink area reasonable. The phase estimation value PEV_{PCa} for prostate cancer, also called diagnostic parameter, can be computed by

$$PEV_{PCa} = F_{PCa}(t) = \frac{\sum_{i=1}^{n} x_i^{AD} \times y_i^{MF}}{\sum_{i=1}^{n} y_i^{MF}} \tag{25}$$

where n and i respectively represent the total number and ordinal of the coordinates on the boundary of the pink area (overlap area) in Fig.(3). Besides, x_i^{AD} is the value of disease indicator correlation and y_i^{MF} represents the membership degree of x_i^{AD}. Moreover, $F_{PCa}(t)$, shown in Eq.(1), is the auxiliary diagnosis function and t is the current diagnosis interval for prostate cancer. Meanwhile, by synthetically assessing clinical stage judgement and phase estimation value, machine-assisted system is able to recommend different therapeutic proposals to physicians, such as active surveillance, drug treatment, excision, endocrinotherapy, radiotherapy, chemotherapy, etc.

E. TREATMENT ADJUSTMENT AND RISK MANAGEMENT FOR PROSTATE CANCER IN INTELLIGENT HOSPITAL SYSTEM

In theory, an accurate therapeutic schedule is particularly important for the treatment and rehabilitation of prostate cancer patients due to the high morbidity and mortality from the disease. Moreover, the underlying development trend of prostate cancer and the clinical manifestations from patients are two most uncontrollable risk factors over the whole diagnosis interval, hence only a timely adjustment to treatment recommendations can effectively prevent the deterioration of the disease.

As shown in Fig.(1), after the normal value of β has been determined by machine-assisted system, the process of treatment adjustment and risk management will be performed via the iterative calculation of Mamdani minimizing operation, which can be demonstrated as follows:

If $F_{PCa}(t+1) > F_{PCa}(t) > \beta$ existed, it is a sign that the condition in patients is getting worse and the drugs for them may be noneffective, thus physicians must immediately change their treatment scheme so that the disease is under control.

If $F_{PCa}(t) > F_{PCa}(t+1) > \beta$ existed, this demonstrates that the condition of patients takes a favorable turn and the deterministic treatment method is beneficial to their recovery, so physicians may keep the original treatment method or adopt a better curative strategy.

If $F_{PCa}(t) > \beta > F_{PCa}(t+1)$ existed, which means that prostate cancer patients show a signal of gradual recovery, therefore physicians may provide them with some precautionary measures or controller medications.

On the whole, the purpose of treatment adjustment and risk management is to monitor prostate cancer in real time and simultaneously provide the most effective treatment method for prostate cancer patients as fully as possible.

F. ALGORITHM COMPLEXITY ANALYSIS

For the sake of the readability of medical data decision-making in intelligent medical system, we rigorously establish detailed pseudocode to explain the extensibility and applicability of the algorithm. To be specific, in the progress of information entropy method, each disease indicator can be assigned an appropriate weight value through n times of iterative data training, therefore the time complexity of the process of weight adjustment is $O(n)$. Furthermore, according to the fuzzy inference logic, three different levels of fuzzy sets can be allocated to each disease indicator, and the system accurately gives a phase estimation value for prostate cancer,

TABLE 2. Medical data collection with start and finish time from the three hospitals in China.

Hospital	System	Start time	Finish time
Xiangya Hospital	HIS	01-01-2011	07-07-2015
	EMR	12-01-2008	11-01-2015
The Second Xiangya Hospital	HIS	09-01-2009	11-05-2015
	EMR	09-25-2009	05-27-2015
	EMR document file	01-01-2011	05-10-2015
	LIS	01-01-2002	05-31-2014
	RIS	02-01-2013	12-17-2015
	PACS	01-01-2012	12-18-2015
The Third Xiangya Hospital	HIS	04-01-2002	12-05-2015
	EMR	04-01-2002	12-05-2015
	EMR document base	05-01-2014	12-09-2015

Algorithm 1 Big Medical Data Decision-Making Model

Input: FPAP, PSMA, TPSA, FPSA, RBC, Hb

Output: $F_{PCa}(t), f(x)$, treatment recommendations

Begin
Obtaining data associated with disease indicators;
Preprocessing and binning modules $M(t)$ and $C(t)$;
Establishing disease correlations $PC(t)$ and $SC(t)$;
Defining diagnosis function $F_{PCa}(t)$;
for (each item x_j of disease indicators) **do**;
  if (doctor.set(λ_j)) **then**
    $w_{TM(j)} = w^*_{TM(j)}$;
  else
    $w_{TM(j)} = w_{TM(j)}$;
  end if
end for
for (each item i of disease indicators) **do**
  $MF_{PCa}(AD_{PCa}(i))$=$D_{Low}(AD_{PCa}(i))$, $D_{Medium}(AD_{PCa}(i))$,$D_{High}(AD_{PCa}(i))$;
  Defining function $E_{\mu,\eta}(AD_{PCa}(i))$;
  Computing different membership degrees;
end for
$F_{PCa}(t) = FIL(AD_{PCa}(i))$;
Giving the clinical stage of prostate cancer based on "If-Then" rules;
for (each minimizing operation t in the whole diagnostic interval) **do**
  if $(F_{PCa}(t+1) > F_{PCa}(t) > \beta)$ **then**
    Adjusting or keeping treatment proposals;
  else if $(F_{PCa}(t) > F_{PCa}(t+1) > \beta)$ **then**
    Adopting more effective curative schemes;
  else$(F_{PCa}(t) > \beta > F_{PCa}(t+1))$
    Giving some controller medications;
  end if
end for
Output $F_{PCa}(t), f(x)$, curative recommendations;
End

so the time complexity of fuzzy inference logic is $O(n)$. Ultimately, the time complexity of the process of treatment adjustment and risk management is the number of diagnostic intervals $O(i)$. As a consequence, the overall time complexity of medical data decision-making in intelligent medical system can be computed by $O(n + n + i) = O(n)$.

TABLE 3. Medical data classification for prostate cancer from the three hospitals in China.

Data category	Amount
Medical information	1,933,535 items
Outpatient service	691,238 people
Doctors' device in outpatient	24,021,298 items
Be hospitalized	1,149,187 people
Diagnosis	1,089,327 items
Electronic medical records	4,855,619 items
Doctors' device in clinical	25,757,699 items
Inspection records	157,426 items
Medical laboratory records	8,725,586 items
Routine inspection records	22,358,881 items
Operation records	318,022 items
Drug records	120,546 items

IV. EXPERIMENT PERFORMANCE

A. DATA ACQUISITION, CLASSIFICATION, AND PREPROCESSING

In our works, medical hospitalization information used in this experiment is collected from three first-class hospitals in China: Xiangya Hospital, the second Xiangya Hospital, and the third Xiangya Hospital. Information recording center collects, classifies, preprocesses, and integrates various categories of medical data associated with prostate cancer based on different systems of the three hospitals. These statistics mainly reflect the curative effect of therapeutic selection and the changes in patients' physiological indicators over the whole diagnosis cycle. In addition, Table 2 exhibits the start and finish time of data acquisition from various systems of the three hospitals.

As shown in Table 3, large amounts of medical data associated with prostate cancer from 2011 to 2015 are strictly recorded, preprocessed, and classified by different systems in the three hospitals. Moreover, in order to ensure the accuracy and rationality of the experiment, we extract 23658 items of structured and recognizable medical information from more

TABLE 4. The normal range of each disease indicator in healthy people.

Disease indicator	Normal range
Total Prostate-Specific Antigen (TPSA)	0-4.0 ng/ml
Free Prostate-Specific Antigen (FPSA)	4-20 ęlg/L
Hemoglobi (Hb)	120-165 g/L
Red Blood Cell (RBC)	12-15 g/100ml
Acid phosphatase (PAP)	0-3.5 ng/ml
Prostate-specific membrane antigen (PSMA)	0-4 ng/ml

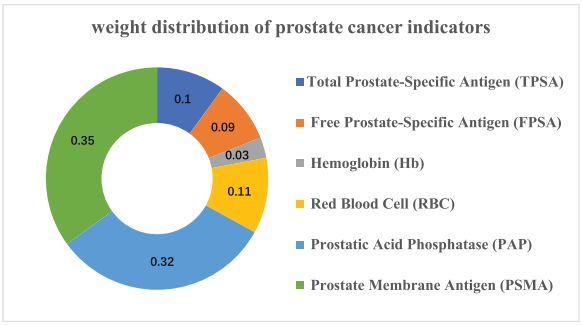


FIGURE 4. Reasonable weight allocation for each disease indicator of prostate cancer.

TABLE 5. Clinical staging of PCa by diagnostic parameter (PEV_{PCa}).

Clinical stage of prostate cancer	PEV_{PCa}
stage I (preliminary period)	16-37
stage II (early period)	38-97
stage III (middle period)	98-209
stage IV (terminal period)	>209

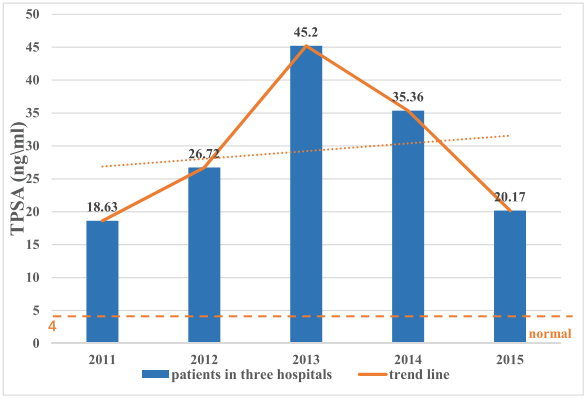


FIGURE 5. The average performance of TPSA from prostate cancer cases in the three hospitals between 2011 and 2015.

than 8000 patients who had been diagnose as prostate cancer in the three hospitals since 2011.

B. SETTING OF EXPERIMENTAL ENVIRONMENT VARIABLES

In accordance with the standard of clinical medicine, the normal range of each disease indicator is rigorously demonstrated in Table 4. In this experiment, 80% of 23658 items of structured and recognizable medical information is used as training set, while the remaining data is regard as test set for medical data analysis and decision-making. After data training and manual debugging, the scale and critical parameters of membership functions in machine-assisted system will be legitimately formalized as $d_1 = 0.3$, $d_2 = 0.7$, and $\sigma_1^2 = \sigma_2^2 = 0.1$.

Moreover, during the process of data training, the weight of each disease indicator is automatically evaluated and adjusted by machine-assisted system on the basis of physicians' diagnosis experience. Furthermore, the detailed weight distribution for disease indicators is shown as Fig. 4, in which key parameters (PAP and PSMA) are the most important reference in auxiliary diagnosis and general parameters (TPSA, FPSA, Hb, and RBC) are of secondary importance in medical data decision-making for prostate cancer. At various clinical stages of prostate cancer, each tumor maker may show different reference value for the detection, diagnosis, and prognosis of the disease. Therefore, these parameters need to be optimized continuously so that machine-assisted system could put forward more therapeutic proposals.

To accurately identify patients' condition, the system has made a specific digital division for four clinical stages of prostate cancer. As shown in Table 5, fluctuations of phase estimation value (PEV_{PCa}) in different intervals may reveal different clinical stages of prostate cancer. Furthermore, the increase of PEV_{PCa} very likely imply the deterioration of prostate cancer.

C. MEDICAL DATA ANALYSIS AND DECISION MAKING FOR PROSTATE CANCER IN BIG DATA ENVIRONMENT

As shown in Fig. 5, between 2011 and 2015, the average performance of TPSA slowly increased from 18.63 to 20.17 ng/ml and it reached the maximum value of 45.2 ng/ml in 2013, which demonstrates that prostate cancer cases' condition has been controlled effectively during the last 5 years. However, because the normal range of TPSA is between 0 and 4 ng/ml, those cases, who had been diagnosed with prostate cancer in the three hospitals in the last 5 years, were still in a state of physiologic derangement. In theory, it's more than likely that people with a value of TPSA above 10 ng/ml will suffer from prostate cancer. Besides, patients definitely have developed prostate cancer when the average performance of TSPA exceeds 50 ng/ml. On the whole, a tendency to first rise and then fall in the trend line suggests that these cases had made a gradual recovery from physicians' treatment decisions during the last 5 years.

Additionally, the performance of FPSA/TPSA is another significant basis for the detection, diagnosis, and treatment of prostate cancer in clinical medicine. The normal range of FPSA/TPSA is theoretically equal to or greater than 0.25. The morbidity rate of prostate cancer must be above 56% when the average performance of FPSA/TPSA is below 0.1.

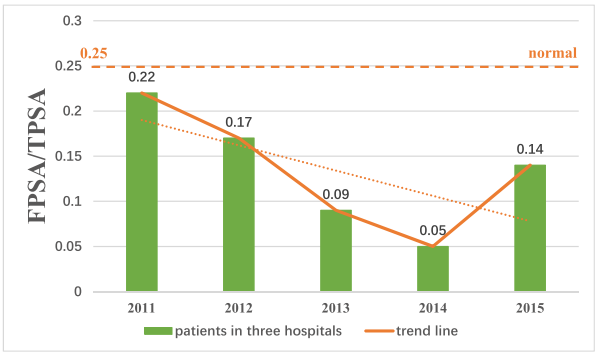


FIGURE 6. The average performance of FPSA/TPSA from prostate cancer cases in the three hospitals between 2011 and 2015.

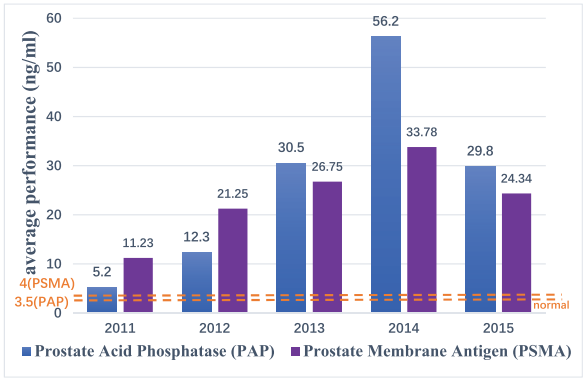


FIGURE 7. The average performance of key parameters from prostate cancer cases in the three hospitals between 2011 and 2015.

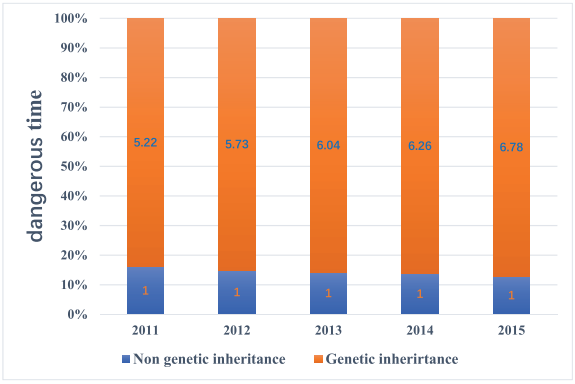


FIGURE 8. The impact of genetic inheritance on medical data decision-making of prostate cancer.

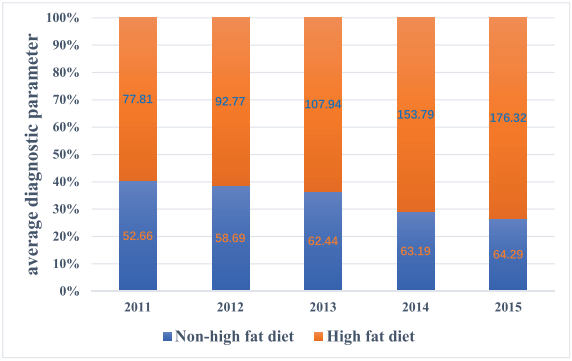


FIGURE 9. The impact of dietary-habit on medical data decision-making of prostate cancer.

In this experiment, the average performance of FPSA/TPSA from prostate cancer cases in the three hospitals between 2011 and 2015 is demonstrated in Fig. 6. According to the statistical medical data between 2011 and 2014, we can see that the average performance of FPSA/TPSA dramatically declined from 0.22 to 0.05, which obviously reveals that most prostate cancer cases' condition was constantly deteriorating. Moreover, these structured medical data further illustrate that most cases diagnosed by the three hospitals in the last 5 years were in the clinical stages III or IV of prostate cancer (middle or terminal stage). Fortunately, state of health of these prostate cancer cases began to gradually improve from 2014 to 2015, mostly because physicians had taken some therapeutic measures for them, such as drug treatment, excision, radiotherapy, chemotherapy, etc.

As the most important guide to the detection, diagnosis, and prognosis of prostate cancer, key paraments (PSMA and PAP) should be analyzed and assessed throughout the whole diagnosis cycle so that machine-assisted system could accurately make therapeutic decisions. In addition, the concept of clinical medicine indicates that the normal ranges of PSMA and PAP are below 4 and 3.5 ng/ml, respectively. Fig. 7 reflects the average performance of key parameters from prostate cancer cases in the three hospitals between 2011 and 2015. Overall, there is been a tendency to first rapidly rise and then slowly fall in the statistical data related to key paraments in the last 5 years. Moreover, it should be noted that PAP and PSMA reached the maximum values of 56.2 and 33.78 ng/ml in 2014, respectively. This demonstrates that symptoms in most cases were still getting worse from 2011 to 2014, and they had begun to recover gradually since 2014. From the perspective of medical data decision-making, physicians may develop more effective therapeutic methods to control patients' condition after consultation of specialists.

So far, what exactly causes prostate cancer is unknown, but several risk factors for the disease have been identified. In fact, saturated fatty acid and genetic inheritance have been regarded as the two main inducements of the symptoms of prostate cancer. Specifically, men who have a family history of prostate cancer will have a higher risk of developing the disease, and about 10% of prostate cancer patients have a family history. Additionally, it appears that a diet rich in red meats and dairy products, as well as high in calcium, may be related to an increased risk of prostate cancer.

By mining and extracting data associated with dietary-habit and genetic inheritance from 23658 items of medical

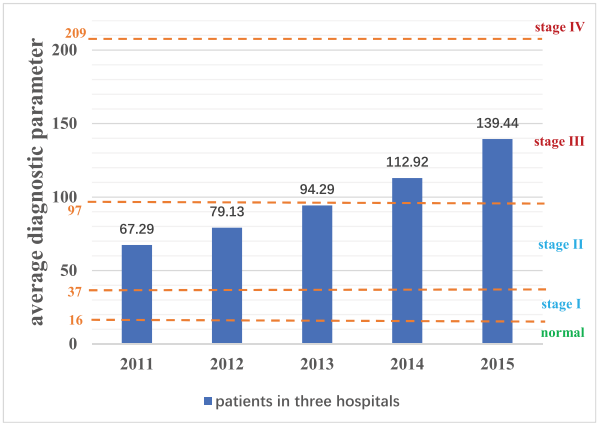


FIGURE 10. Medical data analysis and decision-making for prostate cancer staging in machine-assisted system.

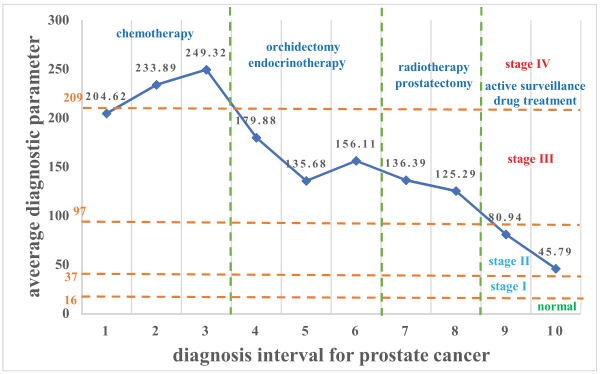


FIGURE 11. Treatment recommendations and risk management for prostate cancer in machine-assisted system.

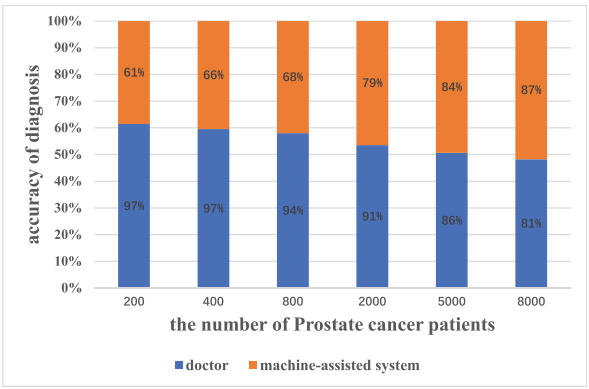


FIGURE 12. The diagnostic accuracy comparison between physicians and machine-assisted system.

information, we have explored the impact of these two risk factors on medical data decision-making of prostate cancer. As shown in Fig. 8, the risk coefficient (average diagnostic parameter) from the prostate cancer cases without genetic inheritance is initialized to 1. Besides, the average dangerous coefficient from the prostate cancer cases with genetic inheritance is approximately 6 times that from those cases without genetic inheritance. In another Fig. 9, from 2011 to 2015, the average diagnostic parameter from the prostate cancer cases without a high-fat diet declined from 52.66 to 64.29, while that from those cases with a high-fat diet constantly went up from 77.81 to 176.32. Overall, the ratio of average diagnostic parameter between the cases with a high-fat diet and those without a high-fat diet slowly rose from 1.477 to 2.743 during the last 5 years. Consequently, physicians could develop some precautionary measures for males with family history of prostate cancer, and prostate cancer cases with a high-fat diet must reasonably reduce their daily intake of fat.

The purpose of the staging of prostate cancer is to describe how serious the cancer is and how far cancer cells have spread in the body. Therefore knowing whether a prostate cancer is at stage I, II, III, or IV is significant in making the best choices about treatment. According to the digital staging for prostate cancer in Table 4, Fig. 10 exhibits the average diagnostic parameter PEV_{PCa} associated with the stage of the disease from the prostate cancer cases in the three hospitals during the last 5 years. Increasing by nearly 2.1 times from 2011 to 2015, the average PEV_{PCa} in one year had gradually rose from 67.29 to 139.44, and the overall average PEV_{PCa} in the last 5 years had reached approximately 100, which reveals that most cases' prostate cancer were in intermittent-grade or high-grade (at stage III or IV). Additionally, the 23658 items of medical information show that most of the 8000 cases from the three hospitals had already developed middle or terminal prostate cancer when the cancer was detected, which is almost consistent with the testing results in machine-assisted system.

The curative options for prostate cancer are based on many factors, including the aggressiveness of the tumor, age and general health, the stage of the disease, patients' preferences, etc. The curative recommendations for prostate cancer from machine-assisted system mainly relies on the clinical stage of the disease that is determined by the system. Fig. 11 reveals treatment recommendations and risk management for prostate cancer in different diagnosis intervals. From the diagnosis intervals 1 to 3, most of the 8000 cases had been in the stage III or IV of prostate cancer when the disease were just detected, hence chemotherapy could be the main curative recommendations from the system. After that, from the diagnosis intervals 4 to 8, the cyclic assessment of the staging of prostate cancer shows that average diagnostic parameter continuously declined from 179.88 to 125.29, thus many alternative therapies are recommended to physicians during this diagnostic period, mainly including orchidectomy, endocrinotherapy, radiotherapy, or prostatectomy. As the improvement of prostate cancer cases' symptoms, machine-assisted system will give preference to active surveillance and drug treatment during the diagnostic intervals 9 and 10.

On the whole, most of the 8000 cancer cases in the three hospitals during the last 5 years had improved significantly with physicians' therapies. Unlike many cancers, terminal prostate cancer can often be controlled for a long period

of time with appropriate treatments. Consequently, it's quite likely that physicians will make a more accurate treatment decision by comprehensive considering their own diagnostic experience and a curative recommendation from intelligent medical system, which could effectively enhances the survival rate of prostate cancer patients.

In this experiment, it is worth mentioning that diagnostic accuracy is rigorously defined as the probability that patients are diagnosed with prostate cancer in their first detection. As shown in Fig. 12, with the number of cases continuously increases from 200 to 8000, the diagnostic accuracy from physicians kept going down from 97% to 81%, while that from machine-assisted system gradually rose from 61% to 87%. In particular, when the size of total samples reaches 8000, the diagnostic accuracy from machine-assisted system exceeds that from physicians for the first time. In addition, the diagnostic accuracy from machine-assisted system has increased by about 42.6% when expanding the size of data samples from 200 to 8000.

V. CONCLUSION

In this study, to address the severe contradiction between large population and scarce medical resources in most developing countries, we propose a medical data decision-making model exploiting fuzzy inference logic for prostate cancer in big medical data environment. Although the great enhancement of diagnostic accuracy from intelligent medical system could be benefit for multimodal medical data fusion and decision-making of prostate cancer, the system is just an auxiliary diagnosis tool for the detection, diagnosis, treatment, and prognosis of the disease. The system cannot replace physicians in choosing the final curative options for patients due to the high mortality and morbidity from malignant tumor. However, intelligent medical system, as an applied technology that is dedicated to adjuvant therapy, still could partly enhance the diagnosis efficiency and accuracy of medical staff in developing countries.

In the future work, we will further implement the information fusion between more multimodal medical data so that some more reliable medical data analysis and decision-making can be provided for prostate cancer in big medical data environment.

ACKNOWLEDGEMENT

Availability of data and materials: all medical data collected from Xiangya hospital, the second Xiangya hospital, and the third Xiangya hospital of Central South University (http://www.xiangya.com.cn/english/)

REFERENCES

[1] M. C. S. Wong *et al.*, "Global incidence and mortality for prostate cancer: Analysis of temporal patterns and trends in 36 countries," *Eur. Urol.* vol. 70, no. 5, pp. 862–874, 2016.

[2] M. K. K. Niaziin *et al.*, "Visually meaningful histopathological features for automatic grading of prostate cancer," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 1027–1038, Jul. 2016.

[3] F. Bray *et al.*, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[4] L. Ospina-Pinillos *et al.*, "Using new and innovative technologies to assess clinical stage in early intervention youth mental health services: Evaluation study," *J. Med. Internet Res.*, vol. 20, no. 9, p. e259, 2018.

[5] A. P. Danso, "The problems of prostate cancer in developing countries—An African perspective," in *Prostate Cancer—Clinical and Scientific Aspects: Bridging the Gap*. London, U.K.: World Scientific, 2015, pp. 1123–1130.

[6] J. Lin *et al.*, "High incidence of incidental prostate cancer in transurethral resection of prostate specimens in China. The value of pathologic review," *Anal. Quant. Cytopathol. Histopathol.*, vol. 38, no. 1, pp. 31–37, 2016.

[7] D. He *et al.*, "Real-world use of docetaxel for metastatic castration-resistant prostate cancer in China: Results from a large observational study," *Ann. Oncol.*, vol. 28, no. 5, pp. 1–2, 2017.

[8] Y. Pan *et al.*, "Characteristics of prostate cancer detection rate (PCDR) in Chinese Han population under different prostate biopsy methods," *Oncotarget*, vol. 8, no. 20, pp. 32930–32936, 2017.

[9] Y. S. Wu *et al.*, "Evaluation of PSA-age volume score in predicting prostate cancer in Chinese population," *Asian J. Androl.*, vol. 20, no. 4, pp. 324–329, 2018.

[10] E. P. H. Choi *et al.*, "Health-related quality of life of Chinese patients with prostate cancer in comparison to general population and other cancer populations," *Supportive Care Cancer*, vol. 24, no. 4, pp. 1849–1856, 2016.

[11] D. P. Kash, M. Lal, A. H. Hashmi, and M. Mubarak, "Utility of digital rectal examination, serum prostate specific antigen, and transrectal ultrasound in the detection of prostate cancer: A developing country perspective," *Asian Pacific J. Cancer Prevention* vol. 15, no. 7, pp. 3087–3091, 2014.

[12] R. Chen *et al.*, "Prostate specific antigen and prostate cancer in Chinese men undergoing initial prostate biopsies compared with Western cohorts," *J. Urol.*, vol. 197, no. 1, pp. 90–96, 2016.

[13] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[14] D. Raví *et al.*, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[15] G. C. Gutierrez-Tobal, D. Alvarez, A. Crespo, F. D. Campo, and R. Hornero, "Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings," *IEEE J. Biomed. Health Inform.*, to be published, doi: 10.1109/JBHI.2018.2823384.

[16] Z. Yu, "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 874–885, May 2018.

[17] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in MRI," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1083–1092, May 2014.

[18] L. L. Wang, Z.-G. Chen, and J. Wu, "Vehicle trajectory prediction algorithm in vehicular network," *Wireless Netw.*, vol. 3, pp. 1–14, Jul. 2018.

[19] Y. Deng, Z. Chen, D. Zhang, and M. Zhao, "Workload scheduling toward worst-case delay and optimal utility for single-hop fog-IoT architecture," *IET Commun.*, vol. 12, no. 17, pp. 2164–2173, Oct. 2018, doi: 10.1049/iet-com.2018.5077.

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

[21] 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.

[22] J. Wu, Z. Chen, and M. Zhao, "Information cache management and data transmission algorithm in opportunistic social networks," *Wireless Netw.*, vol. 8, pp. 1–12, Feb. 2018.

[23] 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.

[24] C. Zou and H. Deng, "Using fuzzy concept lattice for intelligent disease diagnosis," *IEEE Access*, vol. 5, pp. 236–242, 2017.

[25] A. Cameron, F. Khalvati, M. A. Haider, and A. Wong, "MAPS: A quantitative radiomics approach for prostate cancer detection," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 6, pp. 1145–1156, Jun. 2015.

[26] K. Liu, Z. Chen, J. Wu, and L. Wang, "FCNS: A fuzzy routing-forwarding algorithm exploiting comprehensive node similarity in opportunistic social networks," *Symmetry*, vol. 10, p. 338, Aug. 2018.

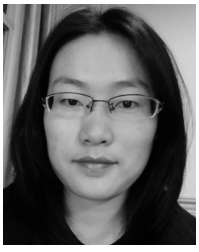
[27] L. C. Dutu, G. Mauris, and P. Bolon, "A fast and accurate rule-base generation method for Mamdani fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 715–733, Apr. 2018.
[28] A. V. Makkuva and Y. Wu, "Equivalence of additive-combinatorial linear inequalities for Shannon entropy and differential entropy," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3579–3589, May 2018.
[29] B. Arras and Y. Swan, "IT formulae for gamma target: Mutual information and relative entropy," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1083–1091, Feb. 2018.
[30] R. Kicsiny, "Black-box model for solar storage tanks based on multiple linear regression," *Renew. Energy*, vol. 125, pp. 857–865, Sep. 2018.
[31] M. G. Tsipouras, C. Voglis, and D. I. Fotiadis, "A framework for fuzzy expert system creation—Application to cardiovascular diseases," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 11, pp. 2089–2105, Nov. 2007.
[32] O. Obajemu, M. Mahfouf, and J. W. F. Catto, "A new fuzzy modeling framework for integrated risk prognosis and therapy of bladder cancer patients," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1565–1577, Jun. 2018.
[33] M.-C. Su, Y.-Z. Hsieh, C.-H. Wang, and P.-C. Wang, "A jacobian matrix-based learning machine and its applications in medical diagnosis," *IEEE Access*, vol. 5, pp. 20036–20045, 2017.
[34] Z. Cai, J. Gu, and H.-L. Chen, "A new hybrid intelligent framework for predicting Parkinson's disease," *IEEE Access*, vol. 5, pp. 17188–17200, 2017.
[35] P. K. Sahoo, S. K. Mohapatra, and S. L. Wu, "Analyzing healthcare big data with prediction for future health condition," *IEEE Access*, vol. 4, pp. 9786–9799, 2017.
[36] G. Han *et al.*, "The LISS—A public database of common imaging signs of lung diseases for computer-aided detection and diagnosis research and medical education," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 648–656, Feb. 2015.
[37] S. Wang, X. Chang, X. Li, G. Long, L. Yao, and Q. Sheng, "Diagnosis code assignment using sparsity-based disease correlation embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3191–3202, Dec. 2016.
[38] H. A. A. Hamid, S. M. M. Rahman, M. S. Hossain, A. Almogren, and A. Alamri, "A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography," *IEEE Access*, vol. 5, pp. 22313–22328, 2017.

KANGHUAI LIU received the B.E. degree from Nanchang University, China. He is currently pursuing the master's degree with the School of software, Central South University. He is also a Researcher with the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory. His research interests include wireless communications and networking, algorithm design, opportunistic networks, medical data decision-making, big data, machine learning, deep learning, and data mining.

ZHIGANG CHEN received the B.E., M.S., and Ph.D. degrees from Central South University, China, in 1984, 1987, and 1998, respectively. He is currently a Professor, a Ph.D. Supervisor, and the Dean of School of Software, Central South University. He is also the Director and an advanced Member of the China Computer Federation, where he is also a member of the Pervasive Computing Committee. His research interests cover the general area of cluster computing, parallel and distributed system, computer security, wireless networks, and big data.

JIA WU was born in 1983. He is currently pursuing the master's degree with the School of Software, Central South University. He is also a Researcher with the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory. His research interests include wireless communications and networking, wireless network, opportunistic networks, medical decision-making, big data, machine learning, deep learning, and data mining.

YANLIN TAN received the Ph.D. degree from The Second Xiangya Hospital of Central South University, China, in 2013. He is currently an Attending Doctor with the PET CCT Center, The Second Xiangya Hospital of Central South University. He received financial support from the National Natural Science Foundation of China, in 2014. His research interests include early diagnosis, clinical staging and therapeutic effect evaluation of common malignant tumors.



LEILEI WANG was born in 1992. She is currently pursuing the M.S. degree with Central South University. Her research interests include opportunistic network, vehicular network, big data, and machine learning.

YEQING YAN is currently pursuing the master's degree with the School of Software, Central South University. She is also a Researcher with the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory. Her research interests include wireless communications and networking, wireless network, opportunistic networks, and medical decision-making.

HENG ZHANG is currently pursuing the master's degree with the School of Software, Central South University. She is also a Researcher with the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory. Her research interests include wireless communications and networking, wireless network, opportunistic networks, medical decision-making, and edge computing.

JUNYAO LONG is currently pursuing the master's degree candidate with the School of Software, Central South University. He is also a Researcher with the "Mobile Health" Ministry of Education–China Mobile Joint Laboratory. His research interests include wireless communications and networking, wireless network, opportunistic networks, medical decision-making, and sensor networks.