AI-Driven Synthetic Biology for Non-Small Cell Lung Cancer Drug Effectiveness-Cost Analysis in Intelligent Assisted Medical Systems

Liu Chang, Jia Wu*, Nour Moustafa, Ali Kashif Bashir, Keping Yu

***Abstract*—According to statistics, in the 185 countries' 36 types of cancer, the morbidity and mortality of lung cancer take the first place, and non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer [1-2]. Significantly in many developing countries, limited medical resources and excess population seriously affect the diagnosis and treatment of lung cancer patients. The 21st century is an era of life medicine, big data, and information technology. Synthetic biology is known as the driving force of natural product innovation and research in this era. Based on the research of NSCLC targeted drugs, through the cross-fusion of synthetic biology and artificial intelligence, using the idea of bioengineering, we construct an artificial intelligence assisted medical system and propose a drug selection framework for the personalized selection of NSCLC patients. Under the premise of ensuring the efficacy, considering the economic cost of targeted drugs as an auxiliary decision-making factor, the system predicts the drug effectiveness-cost then. The experiment shows that our method can rely on the provided clinical data to screen drug treatment programs suitable for the patient's conditions and assist doctors in making an efficient diagnosis.**

***Index Terms*—non-small cell lung cancer (NSCLC), targeted drug therapy, synthetic biology, artificial intelligence medical system, drug effectiveness-cost analysis**

I. Introduction

RESEARCH shows that with the development of the social economy, people pay more and more attention to health issues, and the proportion of medical expenditure in the consumption structure is increasing gradually. Due to the specificity and scarcity of drugs used by cancer patients, the highest medical consumption is undoubtedly cancer patients [1]. According to the latest analysis report on the prevalence of malignant tumors in China conducted by the National Cancer Center in January 2019, China's annual medical expenses required exceed 220 billion yuan. According to the latest global cancer statistics published in the Journal of CA by the World Health Organization in 2018 [2], of the 36 types of cancers in 185 countries, among the estimated 18.1 million new cancer cases for both sexes combined, lung cancer is the most commonly diagnosed cancer (11.6% of the total cases) and the leading cause of cancer death (18.4% of the total cancer deaths) [2], both occupying the first place. Medical technology is a necessary technical means to ensure human health, and the rational allocation of medical resources for lung cancer patients has become a research hotspot [3].

Due to the aging population, the intensification of industrialization and urbanization, changes in lifestyle, and other reasons, global cancer morbidity and mortality have increased rapidly [5]. However, the differences between various countries or regions still exist, mainly due to the differences in social and economic changes. In most developing countries, due to the large population base, backward medical conditions, lack of medical resources, the ratio of doctors to patients is seriously unbalanced. According to the latest global cancer statistics released by the WHO in 2018, 57.3% of cancer deaths worldwide come from Asia [6-8]. Among them, lung cancer incidence is the first in 37 countries in East Asia, Southeast Asia, and North Africa.

According to the National Cancer Statistics Report released by the National Cancer Center in January 2019 [9], in 2015, malignant tumors in China were about 3.929 million, and the deaths were about 2.338 million. That is, an average of more than 10,000 people per day and 7.5 people per minute are diagnosed with cancer [10]. Lung cancer is cancer with the highest morbidity and mortality among cancer types. Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer, and the 5-year survival rate is only 15% [11-13]. The 2-year survival rate of small cell lung cancer (SCLC) is only 1%. This shows that lung cancer is very harmful to people's health. Under such circumstances, patients' diagnosis and treatment needs are challenging to meet [14]. Implementing systematic and professional solutions to assist doctors in diagnosis and decision-making effectively is an urgent problem to be solved.

According to statistics from the Chinese Ministry of Health in 2015, an average of more than 5,800 people can only share one doctor [15]. Doctors in big cities may need to treat more

This work was supported in The National Natural Science Foundation of China (61672540); Hunan Provincial Natural Science Foundation of China (2018JJ3299, 2018JJ3682)). (Corresponding author: Jia Wu)

Jia Wu and Liu Chang are with the School of Computer Science and Engineering, Central South University, China. Nour Moustafa, University of New South Wales Canberra, Australia; Ali Kashif Bashir is with Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom; Keping Yu is with Global Information and Telecommunication Institute, Waseda University, Japan. (e-mail: jiawu5110@163.com; alienliu@csu.edu.cn, nour.moustafa@unsw.edu.au, dr.alikashif.b@ieee.org, keping.yu@aoni.waseda.jp).

than 50 patients every day [16], which shows that doctors' ratio to patients is severely uneven. The cost of training a doctor in the imaging department who specializes in PET and CT is five times that of training a pulmonary doctor in developing countries. In hospitals with underdeveloped medical conditions, ask a PET-CT doctor with professional medical experience to mark pictures. Each picture needs to be marked at $10, and each patient needs at least $500 to mark. China's per capita medical expenditure is only 70 US dollars a year, and it cannot afford this economic pressure. In addition, it is particularly worth considering the medical cost of lung cancer patients during treatment. In the previous study of the auxiliary medical system in China, the differences in the economic capabilities and the individual requirements of different patients were not considered. Due to the lack of economic conditions, life cannot be protected by drugs. One of the results is that mild patients may suffer severe or even catastrophic lesions [17-18]. Most cancer patients lose their lives because they cannot receive timely and effective treatment. More than 50% of lung cancer patients in the world choose to abandon treatment or delay treatment because they cannot afford the high cost of medication [19]. Therefore, there is interest in balancing the cost and benefit of drugs.

As an interdisciplinary emerging in the new century, synthetic biology has attracted widespread attention from governments and scholars at home and abroad and was named "Ten Major Breakthroughs in the Field of Natural Sciences in the First Ten Years of the 21st Century" by "Science" magazine [20]. With the development of artificial intelligence, big data simulation computers and automated laboratory equipment are gradually being used in synthetic biology, greatly improving the efficiency of synthetic biology [21].

In this study, to solve the above problems, improve the medical environment of lung cancer patients and the rational allocation of medical resources. Based on the research of NSCLC targeted drugs, through the cross-fusion of synthetic biology and artificial intelligence, using the idea of bioengineering, we construct an artificial intelligence assisted medical system and propose a drug screening framework that proposes a method for predicting targeted drugs' efficacy based on patient characteristics. Considering the economic capacity and specific needs of different patients, we make the cost of drugs as auxiliary decision-making information to carry out drug effectiveness-cost prediction, achieving personalized treatment, which can realize and elevate the effectiveness and advantages of AI-driven synthetic biology for human wellbeing.

The main contributions of this article are as follows:

(1) A framework for targeted drug screening based on patient characteristics is proposed. Through in-depth extraction and training of patient characteristics, data processing of drug efficacy is completed. In addition, the drug cost is used as an auxiliary decision-making factor, and the ratio of the incremental effectiveness of the target drug to the incremental cost is analyzed to predict the best target drug treatment sequence.

(2) The model defines that all patients with NSCLC belong to one patient set, and patients who receive the same targeted drug treatment belong to the same unit set. From the perspective of the patient set and patient unit sets, choose the probability model to describe the effectiveness and cost of targeted drugs, making this method more universal and suitable for analyzing multiple drugs.

(3) In the process of drug efficacy prediction, the model uses a sparse autoencoder to process patient data to achieve multi-source data association and fusion. Then, through the deep neural network and softmax classifier, the effective and ineffective probability of the drug treatment is output, and the drug effect can be intuitively obtained.

II. Related Work

In today's world, cancer is a severe threat to human life and health. It is a common challenge facing medical research institutions in countries around the world. At present, many research methods of artificial computer intelligence have been widely used in cancer research in the medical field. As a meaningful way for humans to prevent and treat diseases, drugs have attracted many researchers to invest in related research.

In recent years, with the widespread use of artificial intelligence (AI) in clinical cancer research, the performance of assisted cancer prediction and diagnosis has reached a new height. AI can effectively help cancer diagnosis and prognosis [22], especially in terms of accuracy, which is even higher than general statistical applications in oncology.

In the medical IoT environment, the intelligent medical system has always been the research center. In order to promote the intelligent and efficient management of medical resources, researchers have developed an intelligent medical service system based on the Internet of Things environment [23], which improves the utilization of medical resources and helps the medical system develop towards digital, intelligent, and precise. The association analysis algorithm identifies the close connection between disease and treatment [24] and recommends medical methods based on a given patient examination report. This disease diagnosis and treatment recommendation system effectively obtains the required diagnosis and treatment plan and provides high-quality recommendations with a low-latency response. In addition, the use of backstepping technology and the general approximation nature of the fuzzy system [25], through an adaptive fuzzy back-stepping controller, to obtain a suitable drug treatment plan, the intelligent controller for the cancer immunotherapy system is also one of the research focuses.

In the 21st century, modern technology and bio-medicine have dramatically promoted comprehensive conceptual innovation and technological progress in the medical field. The development of various fields also reflects the common trend of pursuing and expressing "precision". Accurate drug selection for cancer patients is one of the leading research areas of precision medicine. The Learning-to-Rank (LTR) framework based on machine learning [26] is used to select and prioritize cancer-specific drugs. This method can accurately predict the best drug based on the data of a single cancer cell line and can significantly improve the success rate of cancer-specific drug selection compared with conventional methods for drug selection. In addition, a new pharmacodynamic modeling

method for drug interactions is established [27], and the best drug combination and drug dosage are determined by considering the side effects of drugs. The research uses the direct search optimization method to identify the best dose plan of the drug combination and then predicts the clinical utility of the drug combination through a simulation program. This method can help find the best drug dosage and dosage range effective and safe and shorten the drug development cycle.

Due to the large population base and lack of medical resources, cancer patients need to spend high medical costs in most developing countries, but most patients cannot afford it. Therefore, more and more researchers are working on medical cost-benefit research to treat cancer patients, hoping to maximize the benefits by spending the lowest cost. By studying the types, quantities, and combinations of medical resources that are feasible in various medical environments [28], a management system is developed around the concept of "integrated health care resource planning and management" to ensure the highest quality medical services while ensuring that patients need to pay the cost is within an affordable range. Some researchers have adopted the Markov model to evaluate the cost-benefit of first-line chemotherapy for patients with non-small cell lung cancer in Thailand from a social perspective and by estimating long-term costs and health status [29]. The study evaluates patients' treatment options and provides valid and reliable cost data for patients to choose chemotherapy options effectively.

The above methods do not consider the differences in the economic capacity and drug response of different groups of people and the particular requirements of different groups. In order to solve the deficiencies of the above methods, our study describes the effectiveness and cost of drugs based on probability analysis and drug decision making so that this method can be applied to the analysis of multiple drugs and can make a personalized drug decision plan to match the patient's financial ability and requirements.

III. System Design

With the popularization and development of artificial intelligence in the medical field, how to effectively assist doctors in diagnosis and treatment has become a research hotspot. Especially in most developing countries, where the population base is significant and doctors' ratio to patients is severely unbalanced, the demand is even more urgent [30-31]. Moreover, intelligent medical treatment can help shorten the diagnosis time, reduce the cost of treatment, and ensure efficacy.

A. Overall system architecture based on drug decision technology

The core of the medical assistance system mentioned in this article lies in a drug selection framework selected explicitly by NSCLC patients. The framework proposes an intelligent drug recommendation program that ensures efficacy, correlates the benefits of targeted drugs with economical costs, and makes drug effectiveness-cost predictions [32-33].

According to the carcinogenic sites that have been identified, the corresponding therapeutic drugs are designed. After the

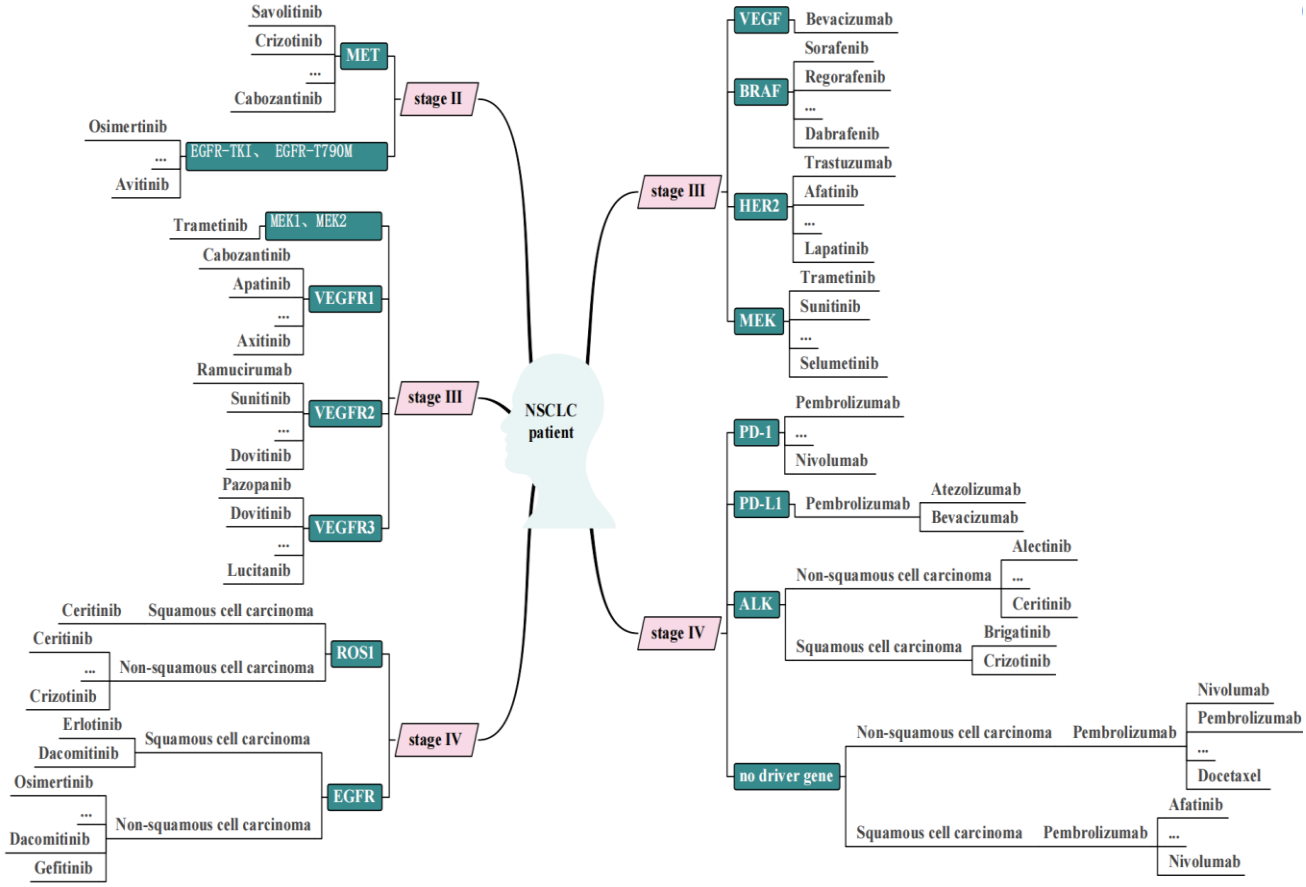

Fig. 1. Targeted drugs commonly used at home and abroad for patients with non-small cell lung cancer of different stages [3-7]

drugs enter the human body, they will accurately combine with the carcinogenic sites and act, allowing the tumor cells to die without affecting the function of healthy cells, tissues, or organs to improve efficacy. According to the target, this is the choice of medication, the treatment process of targeted drugs [34]. This study counted the targeted drugs commonly used by patients with non-small cell lung cancer in different stages and corresponding to the target. As shown in Figure 1, patients in Phase II, III, and IV have different commonly used targeted drugs such as EGFR, MET, and BRAF genes. The drugs will also overlap because the same drug may be effective against multiple targets.

The model considers the effectiveness and cost of targeted drugs from the perspective of the patient unit sets and the patient set and then analyzes the ratio of effectiveness increment to cost increment of targeted drugs from the perspective of probability to predict the optimal treatment targeted drug sequence, such as Figure 2 shows.

Specifically, it includes the following steps:

1) Predict the effectiveness of targeted drugs for the treatment of NSCLC patient unit sets and the patient set. It is defined that all patients with non-small cell lung cancer belong to one patient set, and patients who receive the same targeted drug therapy belong to the same unit set. Considering the changing degree of the tumor area and tumor markers after drug treatment, and digitally deal with the effectiveness of drug treatment. Then the patient's drug treatment data is input into the sparse self-encoding model for latent feature extraction and selection. The softmax classifier is used for feature training to output the drug's valid and invalid probability to predict the efficacy. After the patient unit set's effectiveness is obtained separately, analyze the effectiveness of the patient set.

2) Analyze the cost of targeted drugs to treat NSCLC patient unit sets and the patient set. From the probability model perspective, the lognormal distribution is selected to represent the cost of patient unit sets and the patient set.

3) Describe the effectiveness increment and cost increment of the patient unit set relative to the patient set.

4) Analyze the ratio of effectiveness increment to cost increment of targeted drugs from probability to predict the optimal treatment targeted drug sequence.

B. Drug decision-making steps

According to the overall drug decision-making framework of the medical assistance system we propose, its core lies in conducting drug screening and assisting doctors in drug decision-making to provide accurate medical services. Different patient groups have different responses to targeted drugs, so the probability model is used to describe the effectiveness and cost so that our method can be applied to the analysis of multiple targeted drugs. First, we analyze the effectiveness of targeted drugs used in patients with non-small cell lung cancer.

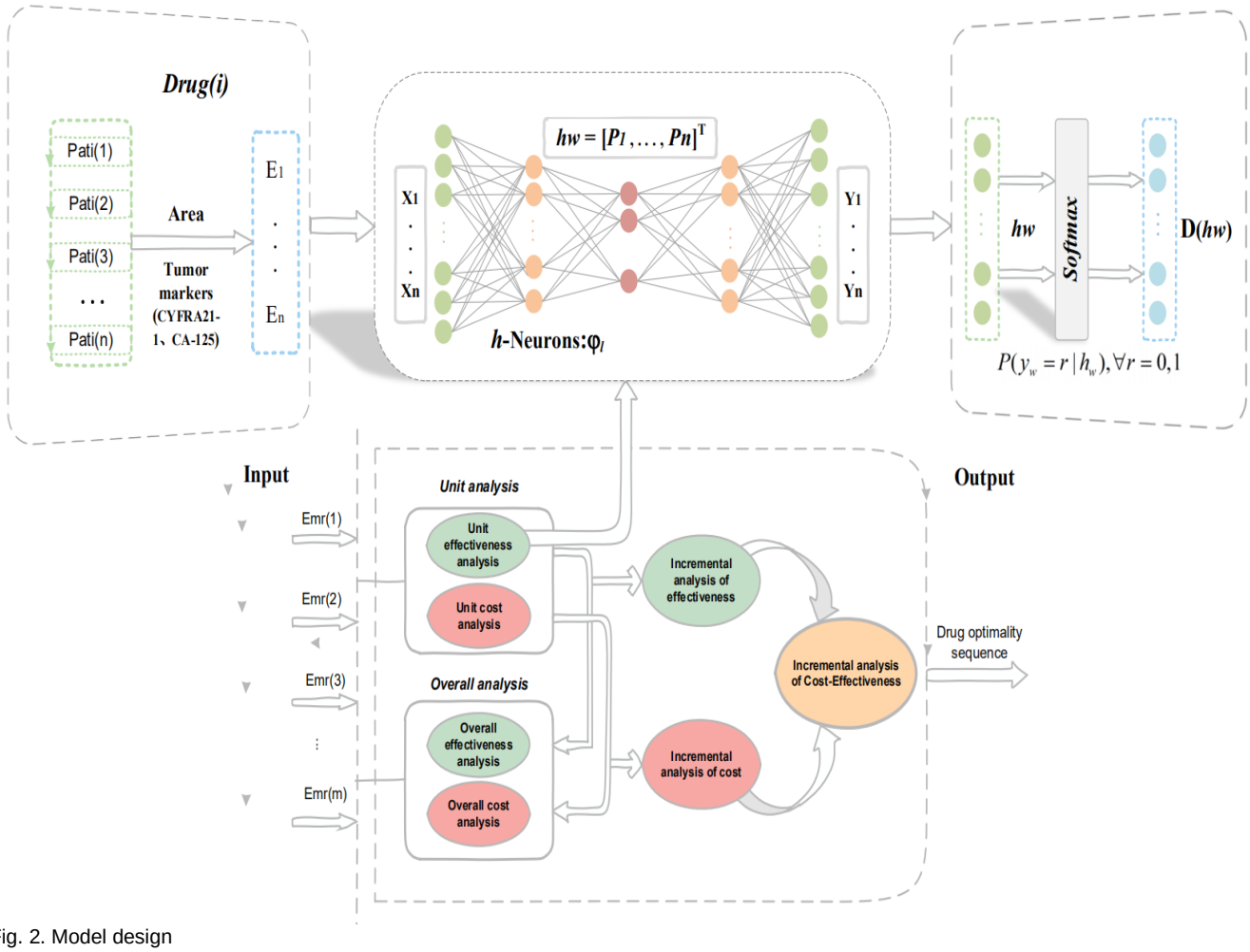

Fig. 2. Model design

1). Drug effectiveness analysis

In order to facilitate the patient's effectiveness analysis of the used drugs, all patients with non-small cell lung cancer are divided into different unit sets according to the targeted drugs used. That is, patients who receive the same drug for treatment belong to the same unit set [38]. To quantify the efficacy of each targeted drug, the unit set of patients receiving a single targeted drug is screened out separately, and compared with the patient set (including patients receiving single or multiple targeted drugs), so as to analyze relative benefits. All patient features in each unit set are extracted from the patient's electronic medical record, such as basic information, diagnosis and treatment records, and living habits, and convert statistical data into a set of binary eigenvector *X*. Then perform feature selection, remove irrelevant features and redundant features. To better obtain the latent features of the patients in the unit set, feature dimensionality reduction processing is performed. Each unit set uses the softmax classifier for latent feature training, and the drug treatment probability is obtained through the classifier training [35-36].

When the targeted drug therapy is sufficient, the patient's tumor area will be relatively small, and the level of serum tumor markers will be significantly reduced compared with before treatment. Therefore, we can use the changes in the tumor area and serum tumor marker levels to evaluate the treatment effect. The most specific tumor markers for the diagnosis of NSCLC include soluble fragments of cytokeratin (CYFRA21-1) and cancer antigens (CA-125), so we selected the changing degree of these two tumor markers for evaluation [3]. Define the tumor area (including spread area) before treatment as $Area_{bef}$. The total tumor area after treatment is $Area_{aft}$.

According to the change of tumor area before and after treatment to define the extent of area reduction:

$$L_{area} = \frac{Area_{bef} - Area_{aft}}{max\left\{Area_{bef}, Area_{aft}\right\}} \tag{1}$$

Therefore, the drug efficacy is defined as:

$$E_{drug_i} = \delta L_{area} + \alpha \sum_{a \in n} L_{CYF_a} + \eta \sum_{b \in n} L_{CA_b} \tag{2}$$

where L_{CYF_a} represents the changing degree of the tumor marker CYFRA21-1, and L_{CA_b} represents the changing degree of the tumor marker CA-125. δ, α, and η represent the impact factor of tumor area, tumor markers CYFRA21-1, tumor markers CA-125, and $\delta + \alpha + \eta = 1$.

Suppose the training data set is $\left\{\left(x_1, y_1\right), \cdots \left(x_n, y_n\right)\right\}$. x represents patient characteristics, and y represents effective output value after medication. It is input into a hidden layer with φ_l neurons, and after a nonlinear activation function, a visual layer with *n* neurons is output.

$$p_j^l = \sum_{i=1}^{\varphi_l} w_{ij}^l x_i + b_j^l \tag{3}$$

Where w_{ij}^l is the connection weight between the i-th neuron in layer l and the j-th neuron in layer $l+1$; b_j^l is the deviation of the j-th neuron in layer $l+1$; φ_l is the number of neurons in layer l, p_j^l is the activation value of the j-th neuron in layer l.

By adding sparse constraints based on self-encoding and adding additional penalties to the optimization target to impose sparse constraints on hidden units, hidden neurons can play a more significant role and learn exact features [12]. At this time, the sparse self-encoding reconstruction cost function of *n* training data is [8-9]:

$$\begin{aligned} F_{s\text{-s-coding}}\left(W, B\right) &= \lambda \sum_{j=1}^{\varphi_2} (\rho \ln \frac{\rho}{\hat{\rho}_j} + (1-\rho) \ln \frac{1-\rho}{1-\hat{\rho}_j}) \\ &+ \frac{1}{\varphi_l} \sum_{i=1}^{\varphi_l} \left(\frac{1}{2} \| h_{w,b}\left(x_i - x_i^{'}\right) \|^2 \right) + \frac{\gamma}{2} \|W\|_2^2 \end{aligned} \tag{4}$$

Where γ is a trade-off parameter, λ is the weight of the sparsity penalty factor, ρ is the average target activation value of the hidden layer, and $\hat{\rho}_j$ is the average activation value of the hidden layer node j. Then, we use the backpropagation algorithm to train the sparse self-encoding model to obtain the optimal weight matrix W and deviation vector B. The original data is input into the sparse self-encoding model, and the latent information is extracted as $\left\{\left(p_1, y_1\right), \cdots, \left(p_n, y_n\right)\right\}$. The latent information has good semantic representation ability.

Assume that the acquired patient historical data is H. There are W feature values set, so the latent feature extracted by the encoder is recorded as $h_w = [p_1, ..., p_n]^T$. The softmax classifier is used to model the patient's feature records, and the classifier can calculate the probability of the input feature data is effective or ineffective for the treatment, which is recorded as $P(y_w = r \mid h_w), \forall r = 0,1$. That is, r takes two types of labels 0 and 1, and y_w takes different values. We can get the probability that the drug is effective or ineffective by the output result. The category with the highest probability may be that the drug is the most effective for the patient unit set [37].

Therefore, the output of Softmax is a 2-dimensional vector. One element of the vector represents the probability value of the treatment is valid, and the other represents the probability value of the treatment is invalid, and the sum of the probability values is equal to 1 [6]. In order to obtain the sample's prediction, we need to use a mapping function to map the data into an interval of 0 to 1. This mapping is called a hypothesis function. We set the output hypothesis function to be:

$$D_\theta(h_w) = \begin{bmatrix} P(y_w = 0 \mid h_w, \theta) \\ P(y_w = 1 \mid h_w, \theta) \end{bmatrix} = \frac{1}{\sum_{c=0}^{1} e^{\theta_r^T h_w}} \begin{bmatrix} e^{\theta_0^T h_w} \\ e^{\theta_1^T h_w} \end{bmatrix} \tag{5}$$

Where $\theta = [\theta_0, \theta_1]$ is the k×n matrix of the training target, $\theta_0, \theta_1 \in R^{k \times 1}$ is the weight parameter of the classifier. Enter the patient data h_w, and record the probability of effective and ineffective as:

$$P(y_w = r \mid h_w, \theta) = \frac{e^{\theta_r^T h_w}}{\sum_{c=0}^{1} e^{\theta_r^T h_w}} \quad (6)$$

We can achieve the maximum likelihood of classifying the input patient data and determining whether the targeted drug is sufficient for the patient through the above analysis. At the same time, we need to train to modify the model parameters so that the model's predicted value continuously approaches the actual value. The cost function can describe the degree of approximation between the predicted value and the actual value. Function (ind) represents the indicator function, which enables the probability to be normalized. The predicted value y_w and the actual value E are normalized. The cost function is defined as follows:

$$cf = -\frac{1}{ind\{E\}}\left[\sum_{w=1}^{W}\sum_{r=0}^{1} ind\{y_w = r\} \ln \frac{e^{\theta_r^T h_w}}{\sum_{r=0}^{1} e^{\theta_r^T h_w}}\right] \quad (7)$$

The model's training process is the process of continuously adding new patient data and adjusting the model parameters so that the cost function becomes smaller and smaller. The gradient descent method is used to solve the cost function $cf(\cdot)$. Since it is strictly convex [8], the gradient descent method can guarantee convergence to the global optimum and obtain a unique solution. By differentiating the function, the gradient formula can be obtained as:

$$\nabla_{\theta_r} cf = -\frac{1}{ind\{E\}}\sum_{w=1}^{W}\left[h_w\left(ind\{y_w = r\} - p(y_w = r \mid h_w, \theta)\right)\right] + \lambda\theta_r \quad (8)$$

$$\theta^* = \arg\min_{\theta}\{-\frac{1}{W}\left[\sum_{w=1}^{W}\sum_{r=0}^{1} ind\{y_w = r\} \ln \frac{e^{\theta_r^T h_w}}{\sum_{r=0}^{1} e^{\theta_r^T h_w}}\right] + \frac{\delta}{2}\sum_{w=1}^{k}\sum_{r=0}^{1} E\theta_{wr}^2\} \quad (9)$$

Where $\nabla_{\theta_r} cf$ represents the partial derivative of $cf(\cdot)$ to θ_r. Using the above partial derivative formula and training the model parameters through the gradient descent method, the unique solution θ value that minimizes $cf(\cdot)$ can be obtained. The probability that the targeted drug therapy is effective and ineffective can be obtained by training the model. Based on this probability, we can predict the patient unit set and the effectiveness of the patient set, and use the beta distribution to describe the effectiveness of drug *i* against the unit set $U_{V_i}(v_i)$ and the effectiveness of all drugs against the patient set $U_V(v)$:

$$U_{V_i}(v_i) = \frac{1}{B(\xi_i, \omega_i)} v_i^{\xi_i - 1}(1 - v_i)^{\omega_i - 1} \quad (10)$$

$$U_V(v) = \frac{1}{B(\xi, \omega)} v^{\xi - 1}(1 - v)^{\omega - 1} \quad (11)$$

Among them, v_i and v are effectiveness variables, $B(\xi_i, \omega_i)$ is the beta function, and ξ and ω are the numbers of people who predict the treatment is effective and ineffective, respectively.

Based on the patient's medical record, the above analysis shows that the tumor area and the degree of tumor marker change after drug treatment are considered in this module, and the benefits of drug treatment are processed digitally. The sparse self-encoding model performs latent feature extraction on patient records [39]. The classifier trains the features to output the probability that the drug treatment is effective or ineffective, and the drug treatment effect is predicted, thereby realizing the prediction of the newly input patient set data.

2). Drug cost analysis

In order to consider the economic capacity of different patients and meet the needs of different patients, we consider the economic cost of targeted drugs as the model's assist decision-making information. Choose a lognormal distribution to describe the cost of different targeted drugs for NSCLC patient unit sets and the patient set. The process is described as follows:

Define $M_{drug(i)}[1, 2 \ldots\ldots n]$ represents the original drug cost of non-small cell lung cancer patients using the targeted drug *i*, n means n patients in the unit set *i*, in order to quantify the comparative cost better, normalize the original drug cost data, and $M'_{drug(i)}[1, 2 \ldots..n]$ represents the normalized cost data; Define $M_{drug}[1, 2 \ldots\ldots m]$ represents the original medication cost data of the NSCLC patient set, m represents the total number of patients is m, and $M'_{drug}[1, 2 \ldots m]$ represents the normalized data, that is:

$$M'_{drug(i)}[j] = \frac{M_{drug(i)}[j]}{Min(M_{drug(i)}) + Max(M_{drug(i)})} \quad (12)$$

$$M'_{drug}[j] = \frac{M_{drug}[j]}{Min(M_{drug}) + Max(M_{drug})} \quad (13)$$

where $M_{drug(i)}[j]$ represents the original drug cost of the targeted drug i for the patient number j, $M'_{drug(i)}[j]$ represents the cost after the normalized treatment; $M_{drug}[j]$ represents the original cost data of the patient number j of the NSCLC patients, $M'_{drug}[j]$ represents the cost corresponding to the normalized treatment; $Min(M_{drug(i)})$ represents the minimum value of the original medication cost data of the patient unit set *i*, $Max(M_{drug(i)})$ represents the maximum value of the original medication cost data of the patient unit set *i*; $Min(M_{drug})$ represents the minimum value of the original medication cost data of the patient set, and $Max(M_{drug})$ represents the maximum value of the original medication cost data of the patient set.

According to the unit cost data set $M'_{drug(i)}[1, 2 \ldots..n]$ and the total cost data set $M'_{drug}[1, 2 \ldots m]$ of patients with non-small cell lung cancer, the corresponding mean $\overline{X}(c)$ and standard deviation S(c) can be calculated; the parameters $\sigma_{\mathrm{d-cost}}$ and $\mu_{d-\mathrm{cost}}$ of the lognormal distribution are obtained:

$$\sigma_{\mathrm{d-cost}} = \sqrt{\ln\left(1+\frac{S(c)^2}{\overline{X}(c)^2}\right)} \tag{14}$$

$$\mu_{d-\cos t} = \mathrm{Ln}[\overline{X}(c)] - \frac{1}{2}\sigma^2 \tag{15}$$

Using a lognormal distribution model to represent the cost distribution of targeted drugs, the cost distribution of patient unit set *i* is $P_{M_i}\left(m_i;\mu_i,\sigma_i\right)$:

$$P_{M_i}\left(m_i;\mu_i,\sigma_i\right) = \frac{1}{\sqrt{2\pi}\sigma_i m_i} e^{-\frac{\left(\ln m_i-\mu_i\right)^2}{2\sigma_i^2}} \tag{16}$$

where m_i represents the cost variable of the unit set *i*, μ_i and σ_i are the parameters corresponding to the lognormal distribution;

The cost distribution of the patient set is $P_M(m;\mu,\sigma)$:

$$P_M(m;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma m} e^{-\frac{(\ln m-\mu)^2}{2\sigma^2}} \tag{17}$$

where *m* represents the overall cost variable, μ and σ represent the parameters corresponding to the lognormal distribution.

Through the analysis of this module, we can use a probabilistic model to describe the patient set and patient unit sets' cost of targeted drugs used to treat NSCLC.

3). Drug effectiveness-cost incremental ratio

In order to accurately assess the differences between different targeted drugs, the effectiveness and cost of targeted drugs are analyzed from the perspective of patient unit sets and patient sets, that is, analyze the effectiveness increment and cost increment of the patient unit set of the targeted drug *i* relative to the patient set.

First, find the probability density $Drug\text{-}U_{\Delta V_i}\left(\Delta v_i\right)$ of targeted drug *i*'s effectiveness relative to the overall effectiveness increase:

$$Drug\text{-}U_{\Delta V_i}\left(\Delta v_i\right) = \begin{cases} \int_0^{1-\Delta v} U_{\Delta V_i}\left(v, v+\Delta v_i\right) dv, & \Delta v_i > 0 \\ \int_0^{1+\Delta v} U_{\Delta V_i}\left(v_i-\Delta v_i, v_i\right) dv_i, & \Delta v_i \le 0 \end{cases} \tag{18}$$

Where Δv_i is the effectiveness increment

$$\begin{aligned} &Drug\text{-}U_{\Delta V}\left(v, v_i\right) = U_V(v) U_{V_i}\left(v_i\right) \\ &Drug\text{-}U_{\Delta V_i}\left(v, v_i\right) \\ &= \frac{1}{B(\xi,\omega)B\left(\xi_i,\omega_i\right)} v^{\xi-1}(1-v)^{\omega-1} v_i^{\xi_i-1}\left(1-v_i\right)^{\omega_i-1} \end{aligned} \tag{19}$$

The probability density of the patient unit cost of the targeted drug *i* relative to the increment of the patient cost is $Drug\text{-}P_{\Delta M_i}\left(\Delta m_i\right)$:

$$Drug\text{-}P_{\Delta M_i}\left(\Delta m_i\right) = \begin{cases} \int_0^{+\infty} P_{\Delta M_i}\left(m, m+\Delta m_i\right) dm, & \Delta m_i > 0 \\ \int_0^{+\infty} P_{\Delta M_i}\left(m_i-\Delta m_i, m_i\right) dm_i, & \Delta m_i \le 0 \end{cases} \tag{20}$$

Where Δm_i is the cost increment

$$\begin{aligned} &Drug\text{-}P_{\Delta m_i}\left(m, m_i\right) = P_M(m) P_{M_i}\left(m_i\right) \\ &Drug\text{-}P_{\Delta M_i}\left(m, m_i\right) \\ &= \frac{1}{2\pi\sigma\sigma_i m m_i} e^{\left(-\frac{(\ln m-\mu)^2}{2\sigma^2}-\frac{\left(\ln m_i-\mu_i\right)^2}{2\sigma_i^2}\right)} \end{aligned} \tag{21}$$

In order to show the difference in efficacy of different targeted drugs more clearly, and considering the economic cost of targeted drugs, we analyze the relationship of effectiveness increment to cost increment of targeted drugs, that is, the effectiveness-cost Incremental ratio. It can help predict the optimal treatment target drug sequence, thus assisting doctors in drug selection.

If the targeted drug *i* produces a positive effectiveness-benefit $\Delta v_i > 0$ and a positive cost-benefit $\Delta m_i > 0$ relative to the average cost of all drugs, the effectiveness-cost incremental ratio $f_1\left(\Delta v_i, \Delta m_i\right)$ is:

$$\begin{aligned} f_1\left(\Delta v_i, \Delta m_i\right) &= \int_0^{+\infty} P_{\Delta m_i}\left(m, m+\Delta m_i\right) dm \\ &* \int_0^{1-\Delta v} U_{\Delta v_i}\left(v, v+\Delta v_i\right) dv \end{aligned} \tag{22}$$

If the targeted drug *i* produces a negative effectiveness-benefit $\Delta v_i < 0$ and a positive cost-benefit $\Delta m_i > 0$ relative to the average cost of all drugs, the effectiveness-cost incremental ratio $f_2\left(\Delta v_i, \Delta m_i\right)$ is:

$$\begin{aligned} f_2\left(\Delta v_i, \Delta m_i\right) &= \int_0^{1+\Delta v} U_{\Delta v_i}\left(v_i-\Delta v_i, v_i\right) dv_i \\ &* \int_0^{+\infty} P_{\Delta m_i}\left(m, m+\Delta m_i\right) dm \end{aligned} \tag{23}$$

If the targeted drug *i* produces a positive effectiveness-benefit $\Delta v_i > 0$ and a negative cost-benefit $\Delta m_i < 0$ relative to the average cost of all drugs, the effectiveness-cost incremental ratio $f_3\left(\Delta v_i, \Delta m_i\right)$ is:

$$\begin{aligned} f_3\left(\Delta v_i, \Delta m_i\right) &= \int_0^{1+\Delta v} U_{\Delta v_i}\left(v_i-\Delta v_i, v_i\right) dv_i \\ &* \int_0^{+\infty} P_{\Delta m_i}\left(m_i-\Delta m_i, m_i\right) dm_i \end{aligned} \tag{24}$$

If the targeted drug *i* produces a negative effectiveness-benefit $\Delta v_i < 0$ and a negative cost-benefit $\Delta m_i < 0$ relative to the average cost of all drugs, the effectiveness-cost incremental ratio $f_4\left(\Delta v_i, \Delta m_i\right)$ is:

$$\begin{aligned} f_4\left(\Delta v_i, \Delta m_i\right) &= \int_0^{1-\Delta v} U_{\Delta v_i}\left(v, v+\Delta v_i\right) dv \\ &* \int_0^{+\infty} P_{\Delta m_i}\left(m_i-\Delta m_i, m_i\right) dm_i \end{aligned} \tag{25}$$

In summary, we can get the various targeted drugs' effectiveness-cost incremental ratio f, as a carrier of quick

diagnosis results, transmit to patients and doctors, thereby providing them Optimize the drug decision method of diagnosis.

4). The optimal choice of drugs

Based on the above analysis, the effectiveness-cost incremental ratio of the targeted drugs used to treat non-small cell lung cancer is obtained. According to the incremental ratio probability distribution, considering the specific needs of patients who are willing to increase the cost of drugs to improve the effectiveness, calculate the probability that each targeted drug can meet the requirements. Finally, according to the probability value of each targeted drug, decide which targeted drug is the best choice.

Precisely, according to the probability density of the four incremental ratios, the probability value of each targeted drug at the threshold t ($t = \Delta m_i / \Delta v_i$) that meets the patient's requirements is calculated:

If t≤1,

$$\begin{aligned} F_T(t_i) &= w_1 * \int_0^1 d\Delta v_i \int_0^{\Delta v_i * t_i} f_1(\Delta v_i, \Delta m_i)\, d\Delta m_i \\ &+ w_3 * \int_{-1}^0 d\Delta v_i \int_{-1}^{\Delta v_i * t_i} f_3(\Delta v_i, \Delta m_i)\, d\Delta m_i \\ &+ w_4 * \int_0^1 d\Delta v_i \int_{-1}^0 f_4(\Delta v_i, \Delta m_i)\, d\Delta m_i \end{aligned} \tag{26}$$

If t>1,

$$\begin{aligned} F_T(t_i) &= w_1 * \left(\begin{array}{l} \int_0^{1/t_i} d\Delta v_i \int_b^{\Delta v_i * t_i} f_1(\Delta v_i, \Delta m_i)\, d\Delta m_i \\ + \int_{1/t_i}^d d\Delta v_i \int_0^1 f_1(\Delta v_i, \Delta m_i)\, d\Delta m_i \end{array} \right) \\ &+ w_2 * \int_{-1}^0 d\Delta v_i \int_{-1}^{\Delta v_i * t_i} f_2(\Delta v_i, \Delta m_i)\, d\Delta m_i \\ &+ w_4 * \int_0^1 d\Delta v_i \int_{-1}^0 f_4(\Delta v_i, \Delta m_i)\, d\Delta m_i \end{aligned} \tag{27}$$

where *w*1, *w*2, *w*3, and *w*4 are weights given according to the significance of the effectiveness-cost incremental ratio.

Based on the above analysis, the probability value of each targeted drug finally obtained is sorted optimally, and the treatment drugs suitable for the patients are given for different patients with non-small cell lung cancer.

IV. Experiments

In this study, all relevant medical data comes from the mobile medical information records collected by the Ministry of Education-China Mobile Joint Laboratory and three Xiangya Hospitals of Central South University.

The diagnosis and treatment data of patients with non-small cell lung cancer are mainly obtained through the methods shown in Table 1 and Table 2. The three hospitals have different medical systems for data collection. Among them, Xiang'ya Hospital uses Hospital Information System (HIS) + Electronic Medical Record (EMR), and The Second Xiang'ya Hospital uses HIS + EMR + Laboratory Information System (LIS) + Radiological Information Management System (RIS). The Third Xiang'ya Hospital is HIS + EMR + EMR document base. We started to collect electronic records on January 4, 2002. The deadline is December 17, 2015. The records include patients' basic information, diagnosis information, drug records, etc., and all information is combined into a complete electronic medical record.

In system model design, we count the targeted drugs commonly used by NSCLC patients in different stages and corresponding to the target. In this experiment, six drugs are screened from various targeted drugs for different targets for comparison. Specifically, we select Gefitinib, Afatinib, Crizotinib, Apatinib, Erlotinib, Icotinib, six more commonly targeted drugs to verify.

TABLE 1
THREE HOSPITALS WITH DIFFERENT MEDICAL SYSTEMS FOR DATA ACQUISITION, WITH BEGINNING AND ENDING TIMES

Hospital name	System	Start time	Finish time
Xiang'ya Hospital	*HIS*	01-01-2011	07-07-2015
	EMR	12-01-2008	11-01-2015
The Second Xiang'ya Hospital	*HIS*	09-01-2009	11-05-2015
	EMR	09-25-2009	05-27-2015
	LIS	01-01-2002	05-31-2014
	RIS	02-01-2013	12-17-2015
The Third Xiang'ya Hospital	*HIS*	04-01-2002	12-05-2015
	EMR	04-01-2002	12-05-2015
	EMR document base	05-01-2014	12-09-2015

TABLE 2
DATA COLLECTION AND TYPE IN THREE HOSPITALS

Type	Number
Outpatient service	***968,545*** people
Patient information	***2,789,675*** items
Be hospitalized	***176,899*** people
Diagnosis records	***1,124,561*** items
Electronic medical records	***5,287,413*** items
Drug records	***90,631*** items
Routine inspection records	***24,287,612*** items
Medical laboratory records	***3,483,216*** iterms

In selecting a drug-assisted doctor for patient treatment, the system will select a drug suitable for the patient according to the trade-off between drug effectiveness and cost, and provide the doctor with an auxiliary diagnosis. To evaluate the effectiveness of various targeted drugs, this experiment uses a 10-fold cross-validation method because this method can provide a sufficiently accurate estimate of the correct error rate. We randomly select 100 patients with six different drugs for treatment from the data set and randomly divided all the data into ten equal subsets. On the one hand, six subsets are used as the training set; on the other hand, the remaining four subsets are used as the test set. Three hospitals randomly select the same number of data sets and randomly divide them into training and test data. After that, calculate the average error of all ten experiments separately. In this way, the test and training will be repeated ten times to ensure the assessment's accuracy.

We set some fundamental indicators to evaluate the performance of the power prediction classification [40-41].

TP (True Positive): positive samples predicted to be positive by the model.

TN (True Negative): negative samples predicted to be negative by the model.

FP (False Positive): negative samples predicted to be positive by the model.

FN (False Negative): positive samples predicted to be negative by the model.

By displaying these four indicators together in a graph, we can obtain such a confusion matrix and use the confusion matrix to visualize the performance of classification, as shown in Figure 3.

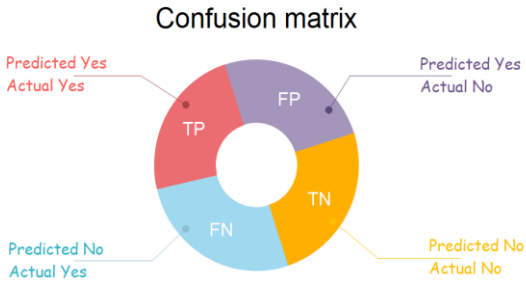

Fig. 3. Confusion matrix

Define the following measures as the assessment of the prediction results:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$SEN = \frac{TP}{TP + FN} \quad SPEC = \frac{TN}{TN + FP}$$

By defining the above measures, the number of instances can be converted to a ratio between 0 and 1 to promote standardized metrics.

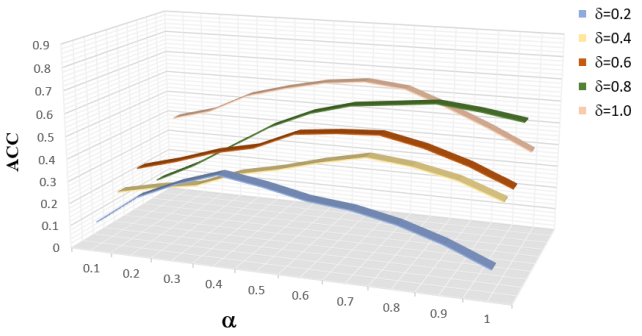

Fig. 4. The trade-off relationship between the effectiveness evaluation value parameters α, δ and the evaluation measure ACC

Figure 4 reflects the trade-off relationship between the effectiveness evaluation value parameters α, δ, and the evaluation measure ACC. When the proportion of parameter α value gradually increases in the calculation of the evaluation, the accuracy will be relatively improved, and the impact on the prediction effect will be higher. This may be because the change in tumor markers has gradually been relatively stable. When the weight ratio interval is between 0.5-0.6, the accuracy is the highest. Another set of parameter δ is the weight of the change in the tumor area on the training model drug cure of sufficient information. It can be seen that as the δ weight increases from 0.2 to 1, the accuracy increases. The auxiliary information on drug effectiveness helps to improve the accuracy of the classification prediction of drug effectiveness.

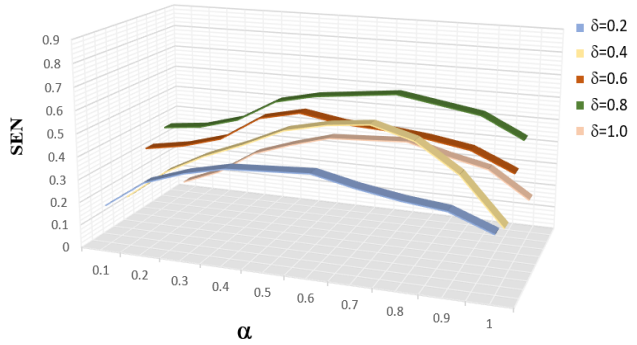

Fig. 5. The trade-off relationship between the effectiveness evaluation value parameters α, δ and the evaluation measure SEN

Figure 5 shows the trade-off relationship between the validity evaluation value parameters α, δ, and the evaluation measure SEN. As the weight of α increases, the sensitivity parameter first increases and then decreases, and there is a distinct difference between the overall increase and decrease. When the value of a is 0.5, the sensitivity parameter reaches the highest peak as a whole because the marker's sensitivity is very strong. The other parameter δ, with the increase of its weight, the overall degree of change in tumor area in the sufficient information of drug cure has little difference in sensitivity. The average minimum sensitivity appears when the δ weight is the lowest, which may be because the relative changes in patient tumor markers can be better considered.

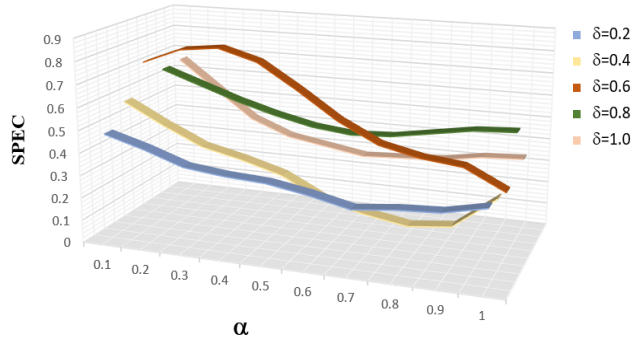

Fig. 6. The trade-off relationship between the effectiveness evaluation value parameters α, δ and the evaluation measure SPEC

Figure 6 shows the trade-off relationship between the validity evaluation value parameters α, δ, and the evaluation measure SPEC. As the weight of the tumor marker α increases, the initial specificity has steadily decreased. Contrary to the accuracy, there is no specificity reduction while achieving a substantial increase in accuracy. When parameter δ's weight rises, overall, the specificity is better at its high weight, and the more significant the impact of the change in tumor area on the prediction effect. When the δ value is 0.6, the specificity is significantly better than the other values. It may be that the area

is more specific for the evaluation method, and the specificity of the tumor marker is limited.

The model assumes that the initial weights of the three diagnostic parameters each account for one-third, and then they are multiplied by the optimal weights obtained from the simulation experiments to obtain the ratio between the three to allocate. After calculation, to achieve the best prediction effect, the model sets the weights of the three diagnostic parameters α, δ, and η to be 0.35, 0.4, and 0.25, respectively.

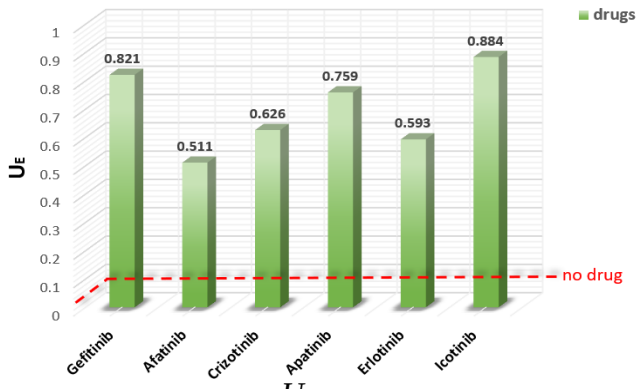

Fig. 7. The effectiveness value U_E of six drugs use independently

We can get the effectiveness of 6 targeted drugs (Gefitinib, Afatinib, Crizotinib, Apatinib, Erlotinib, Icotinib) for NSCLC patients, as shown in Figure 7. As shown in the figure, the performance of the three targeted drugs Icotinib, Gefitinib, and Apatinib, is relatively superior. Considering the existence of the body's immune system, patients who have not received drug treatment will also show a specific effect, so we have added a comparison line in the figure.

This study considers drug cost information as part of the decision-making process. The cost estimate only includes medical and health costs directly related to clinical trials, including drug treatment and significant adverse reaction treatment costs. Among them, the cost of adverse reactions is based on the opinions of clinical experts. Only the adverse reactions that teach grades 3 and 4 are used for estimation, so we only added liver protection and diarrhea costs. Taking China's drug cost as an example for analysis, the specific unit price details of the six drugs and the cost of significant adverse reactions are shown in Table 3. All the data in the cost statistics table comes from the China Medical Price Publicity Network.

TABLE 3
MEDICAL COST

	parameters	unit	average price(¥)
Drug cost	*Gefitinib*	250mg/tablet	498/10tablets
	Afatinib	30mg/tablet	988/7tablets
	Crizotinib	250mg/capsule	25154.81/60capsules
	Apatinib	250mg/tablet	1007.04/10tablets
	Erlotinib	150mg/tablet	1752.34/7tablets
	Icotinib	125mg/tablet	757.73/21 tablets
Adverse reaction cost	Liver protection cost	/person	636
	Diarrhea cost	/person	54

Randomly select 100 patients with six different drugs from the data set and divide them into six treatment groups. Based on the consideration of clinical trials, each treatment group's medical cost in 10 cycles (one month is a cycle) is counted, including the cost of drugs, the unit cost of diarrhea, and liver protection multiplied by the corresponding incidence of adverse reaction costs.

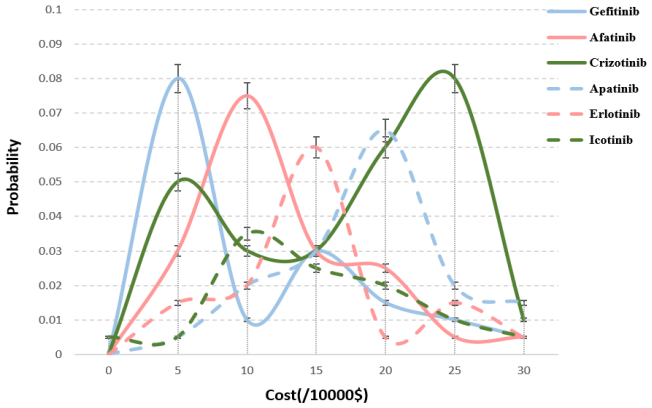

Fig. 8. Probability distribution chart of medical cost in ten cycles

Through Monte Carlo simulation, we can get the probability distribution map of simulation cost. As shown in Figure 8, there is an absolute difference in the six drugs' peak cost. Crizotinib has the highest proportion of high costs, and the proportion of expenses reaching 250,000 is as high as 8%. The second is Apatinib, with the highest cost accounting for 200,000, more than six percentage points, and the proportion of people with the highest cost is higher than the other five drugs. In addition, the cost values of the high probability intervals of Erlotinib, Afatinib, and Gefitinib sequentially decrease. Among them, it is worth noting that except for Icotinib, the probability of the cost of the other five drugs being 0 is 0. Icotinib has a low probability of cost 0, because some patients are new drug testers, and the pharmaceutical company does not charge any fees.

After describing the costs of various targeted drugs for the treatment of non-small cell lung cancer, combined with the targeted drugs' effectiveness, this study can evaluate the cost-effectiveness of the drugs to make optimal predictions. Accurately, it is reflected in the cost-effectiveness ratio *f* of the medicine, as shown in Figure 9.

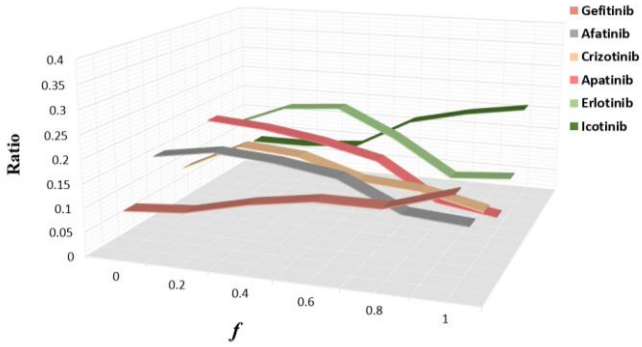

Fig. 9. cost-effectiveness ratio of six targeted drugs

It can be seen from the figure that the *f* of the targeted drug-Icotinib is the highest overall value because its higher value accounts for a higher proportion. That is, the treatment of NSCLC has better effects and is relatively low cost. The *f* of

Gefitinib ranks second overall because of its lower cost and good therapeutic effect. The high *f*-values of Apatinib, Crizotinib, and Erlotinib are concentrated in the low ratio range of 0.10-0.20. The cost-effectiveness difference is not massive, and it is in the middle of the six drugs. Afatinib's poor cost-effectiveness performance is mainly due to its relatively poorest treatment of NSCLC. Therefore, the targeted drugs Icotinib and Gefitinib are the most cost-effective to treat NSCLC. To sum up, we can conclude that after capturing characteristics of the patient's actual efficacy and drug cost, the system can make the appropriate drug choice based on the trade-off between drug effectiveness and cost.

Besides, according to the model's incremental ratio probability distribution, we can get the patient's requirements to improve unit effectiveness at the overall level and increase medical expenses, that is, to increase the value of willingness to pay (WTP: percentage of original cost) The schematic diagram of the probability value F is satisfied, as shown in Figure. 10. It can be seen from the figure that with the growth of WTP, the probability values of the five targeted drugs Gefitinib, Icotinib, Crizotinib, Erlotinib, and Apatinib all increase to a certain extent, of which Gefitinib has the most apparent increase rate, that is, cost-effectiveness the highest and best meet the needs of patients. When the WTP was increased by 25%, Gefitinib's F even reached 0.88, far exceeding other drugs. However, Afatinib did not produce enough positive effect gains than the other five drugs when increasing WTP.

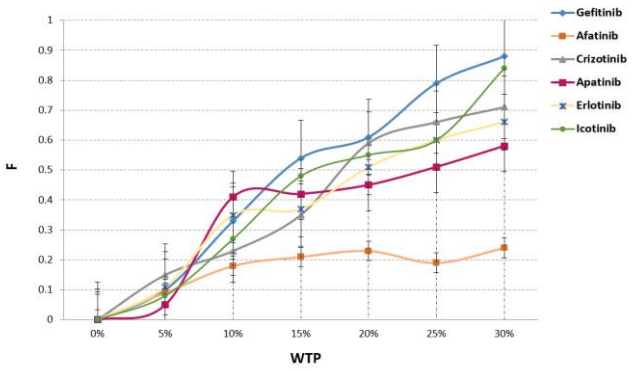

Fig. 10. Comparison of F when increasing WTP

In summary, in treating patients with non-small cell lung cancer, considering the effectiveness and economic cost of targeted drug therapy, in contrast, Icotinib, Gefitinib, Crizotinib, three drugs meet the requirements.

However, the intelligent medical system can only play a role in assisting decision-making and cannot replace doctors to make the final decision on the medication of patients with NSCLC. Doctors can use the diagnosis opinions given by the system to help them choose more suitable personalized medicines for patients, make a second diagnosis, and improve work efficiency.

V. Conclusions

Synthetic biology is the introduction of engineering design concepts based on biology, rational design, and transformation of organisms, and providing new solutions to the challenges faced by humankind, which is of great significance to the improvement of human wellbeing. Based on the study of targeted drugs for non-small cell lung cancer, and through the cross-fusion of synthetic biology and artificial intelligence, we construct an artificial intelligence medical assistance system and propose a drug screening framework. From the patient unit sets and patient set perspectives, extract the latent feature of the patient through the latent autoencoder, train the feature through the classifier, and data processing of drug efficacy to predict the effectiveness of the targeted drug. At the same time, considering the drug cost as an auxiliary decision-making factor, analyze the ratio of effectiveness increment to cost increment of targeted drugs to predict the optimal targeted drug treatment sequence. Experiments show that, based on probabilistic analysis and drug decision-making, this method can rely on the clinical data provided to decide on a medication treatment plan suitable for the patient's condition and assist the doctor in making an efficient diagnosis.

During treatment, NSCLC patients must take various medicines containing antibiotics, vitamins, and so on to maintain the normal functioning of the body's functions. Generally, the side effects of targeted drugs taken by cancer patients during treatment are independent and inevitable, including systemic reactions, skin toxicity, and digestive tract toxicity. In addition, with long-term use of the same targeted drug, patients are likely to develop resistance to this drug. In recent years, studies have found that combined drugs can effectively reduce the side effects of drugs. With the development and research of anti-tumor drugs, drug combinations for non-small cell lung cancer treatment are updated continuously and have become a research hotspot. Therefore, in future work, we can consider mining the characteristics of patients and drugs, exploring suitable decision-making methods for patients with balanced efficacy and cost.

References

[1] International Agency for Research on Cancer. "Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018." International Agency for Research on Cancer: Lyon, France (2018).

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

[3] A. G. Pallis, "A review of treatment in non-small-cell lung cancer, " *European Respiratory Disease*, 8(4), 208-212, 2012.

[4] R. L. Siegel, K. D. Miller, and A. Jemal. "Cancer statistics, 2018, " *CA: a cancer journal for clinicians*, 68(1), 7,2018.

[5] L. Peng, B. Liao, W. Zhu, Z. Li and K. Li, "Predicting Drug–Target Interactions With Multi-Information Fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 561-572, March 2017.

[6] N. Kureshi, S. S. R. Abidi and C. Blouin, "A Predictive Model for Personalized Therapeutic Interventions in Non-small Cell Lung Cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 424-431, Jan. 2016.

[7] J. Wu, Y. L. Tan, Z. G. Chen and M. Zhao, "Decision based on big data research for non-small cell lung cancer in medical artificial system in developing country", *Computer Methods and Programs in Biomedicine*. 159C (2018) pp. 87-101.

[8] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Bio-technology Journal*, vol. 13, pp. 8-17, 2015.

[9] E. Deluche, E. Onesti, and F. Andre, "Precision Medicine for Metastatic Breast Cancer," *American Society of Clinical Oncology Educational Book*, no. 35, pp. e2-e7, 2015.

[10] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," *Cancer Letters*, vol. 471, pp. 61-71, 2020.
[11] J. Wu, Y. L. Tan, et al., "Data Decision and Drug Therapy Based on Non-Small Cell Lung Cancer in a Big Data Medical System in Developing Countries", *Symmetry*, 2018, 10(5), 152.
[12] P. Ding, R. Yin, J. Luo and C. K. Kwoh, "Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological, and Network Knowledge," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1336-1345, May 2019.
[13] Rebecca L. Siegel, Kimberly D. Miller, Ahmedin Jemal DVM, et al. "Cancer statistics, 2019. " *CA Cancer J Clin*, 2019.
[14] J. Sheng, F. Li and S. T. C. Wong, "Optimal Drug Prediction From Personal Genomics Profiles," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1264-1270, July 2015.
[15] X. Li, et al., "Characterization of Drug Efficacy Regions Based on Dosage and Frequency Schedules," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 488-498, March 2011
[16] N. Kureshi, et al., "A Predictive Model for Personalized Therapeutic Interventions in Non-small Cell Lung Cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 424-431, Jan. 2016
[17] LE Schnipper, NE Davidson, DS Wollins, et al. "American Society of Oncology Statement: a conceptual framework to assess the value of cancer treatment options", *Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology*, pp. 2563-2577, 2015.
[18] S. Daoud, A. Mdhaffar, M. Jmaiel and B. Freisleben, "Q-Rank: Reinforcement Learning for Recommending Algorithms to Predict Drug Sensitivity to Cancer Therapy," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 3154-3161, Nov. 2020.
[19] Y. He, J. Liu, and X. Ning, "Drug Selection via Joint Push and Learning to Rank," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 1, pp. 110-123, 2020.
[20] 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, 2012.
[21] Kerstin N Vokinger, Thomas J Hwang, Thomas Grischott, Sophie Reichert, Ariadna Tibau, Thomas Rosemann, Aaron S Kesselheim, "Prices and clinical benefit of cancer drugs in the USA and Europe: a cost–benefit analysis, " *The Lancet Oncology*, vol. 21, no. 5, pp. 664-670, 2020.
[22] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges." *Cancer Letters* (2020): 61-71.
[23] S. Lu, A. Wang, S. Jing, T. Shan, X. Zhang, Y. Guo, and Y. Liu, "A study on service-oriented smart medical systems combined with key algorithms in the IoT environment," *China Communications*, vol. 16, no. 9, pp. 235-249, Sept. 2019.
[24] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, "A disease diagnosis and treatment recommendation system based on big data mining and cloud computing." *Information Sciences* (2018): 124-149.
[25] H. Nasiri, and A. A. Kalat. "Adaptive fuzzy back-stepping control of drug dosage regimen in cancer treatment." *Biomedical Signal Processing and Control* (2018): 267-276.
[26] A. Rahangdale and S. Raut, "Gene-Expression Based Predictor for Drug Selection and Prioritization Using Learning-to-Rank," 2018 International Conference on Bioinformatics and Systems Biology (BSB), Allahabad, India, 2018, pp. 1-6
[27] E. Zanderigo, V. Sartori, G. Sveticic, et al., "A New Model for Drug Interactions and Optimal Drug Dosing," 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, 2005, pp. 2353-2356.
[28] M. Poluta and E. Nunziata, "Integrated Healthcare Resource Planning and Management," 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, 2006, pp. 5444-5446
[29] U. Permsuwan, S. Thongprasert, and B. Sirichanchuen, "Cost-Utility Analysis of First-Line Pemetrexed Plus Cisplatin in Non-Small Cell Lung Cancer in Thailand." Value in health regional issues (2020): 9-16.
[30] Y. Zhang, J. Liu, L. Shi, M. Chen, and J. Liu, "β1-Adrenoceptor antibodies induce PPCM via inhibition of PGC-1α related pathway." *Scandinavian Cardiovascular Journal* 5(2021):1-10.
[31] L. Shi, Y. Zhang, J. Zhang, Y. Gao, and X. Yang. "MiR-339 is a potential biomarker of coronary heart disease to aggravate oxidative stress through Nrf2/FOXO3 targeting Sirt2." *Annals of Palliative Medicine* 10.3(2021):2596-2609.
[32] Z. Lv and H. Song, "Mobile Internet of Things Under Data Physical Fusion Technology," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4616-4624, May 2020.
[33] Q. Du, H. Song and X. Zhu, "Social-Feature Enabled Communications Among Devices Toward the Smart IoT Community," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 130-137, January 2019.
[34] X. Liu, H. Song and A. Liu, "Intelligent UAVs Trajectory Optimization From Space-Time for Data Collection in Social Networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 853-864, 1 April-June 2021.
[35] L. Shi, Y. Zhang, J. Zhang, Y. Gao, and X. Yang, "Application of blood pre-albumin and NT-pro BNP levels in evaluating prognosis of elderly chronic heart failure patients." *Experimental and therapeutic medicine* 20.2(2020).
[36] A. S. Lima, J. N. de Souza, J. A. B. Moura, and I. P. da Silva, "A consensus-based multicriteria group decision model for information technology management committees." *IEEE Transactions on Engineering Management*, vol. 65, no. 2, pp: 276-292, 2018.
[37] J. N. de Souza, and R. Boutaba, "Erratum to: Managing QoS in Multimedia Networks and Services," *Managing QoS in Multimedia Networks and Services*, Springer, Boston, MA, 2017.
[38] H. P. Borges, J. N. de Souza, B. Schulze, and A. R. Mury, "Automatic services instantiation based on a process specification, " *Journal of network and computer applications*, 39, 1-16, 2014.
[39] J. Li, G. Deng, W. Wei, H. Wang and Z. Ming, "Design of a Real-Time ECG Filter for Portable Mobile Medical Systems," *IEEE Access*, vol. 5, pp. 696-704, 2017.
[40] J. Q. Li, R. Li, Z. Z. Chen, et al., "Design of a Continuous Blood Pressure Measurement System Based on Pulse Wave and ECG Signals," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1-14, 2018, Art no. 1900114.
[41] J. Q. Li, Z. Z. Chen, L. Huang, M. Fang, B. Li, X. Fu, and Q. Zhao, "Automatic Classification of Fetal Heart Rate Based on Convolutional Neural Network," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1394-1401, April 2019.

Liu Chang is currently pursuing the master's degree with School of Computer Science and Engineering, Central South University, China. Her research interests include wireless network, big data research, and medical informatics.

Jia Wu received the Ph.D. Degrees in software engineering Central South University, Chang-sha, Hunan, P.R.China, in 2016. He is associate professor in School of computer and engineering, Central South University. Since 2010, he has been Algorithm engineer in IBM company in Seoul, Republic of Korea and in Shang-hai, P. R. China. He is a senior member of CCF (China Computer Federation), a member of IEEE and ACM. His research interests include wireless network, big data research, and medical Informatics.

Nour Moustafa, University of New South Wales Canberra, Australia.

Ali Kashif Bashir is with Department of Computing and Mathematics, E-154, John Dolton, Chester Street, M15 6H, Manchester Metropolitan University, Manchester, United Kingdom.

Keping Yu is with Global Information and Telecommunication Institute, Waseda University, Shinjuku, Tokyo 169-8050, Japan.