Design on Early Warning System for Renal Cancer Recurrence Based on CNN-Based Internet of Things

DONG LIU1, JINKAI SHAO1, HONGGANG LIU1, AND WEI CHENG2,3

1Department of Urology, Shanxi Provincial People’s Hospital, Taiyuan 030012, China
2Department of Urology, Shanxi Bethune Hospital, Third Hospital of Shanxi Medical University, Taiyuan 030032, China
3Department of Urology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Corresponding author: Wei Cheng (chengwei508@126.com)

ABSTRACT Kidney cancer is a type of urinary system tumor. The incidence of kidney cancer, which is second only to bladder cancer, has shown an overall upward trend in recent years. However, the early judgment of kidney cancer is still in the imaging biomarker discovery stage. Early detection and treatment cannot be achieved. Based on the natural advantages of the Internet of Things in the medical field, we focused on an intelligent early warning model of renal cancer recurrence and built a renal cancer early warning system integrated with the Internet of Things. We integrated the influencing factors of renal cell carcinoma, constructed a sample set, conducted data analysis and optimized the dataset. Aiming at the instability of renal cancer recurrence, five supervised learning prediction algorithms, including multiple linear regression, Bayesian ridge regression, gradient boosting tree, support vector regression, and convolutional neural network were used to develop a renal cancer recurrence prediction model. The predictive performance of these five algorithms were compared and discussed. Finally, the best renal cancer recurrence prediction model was established by combining a convolutional neural network with an Internet of Things medical framework. This design provided an intelligent early warning system to predict the recurrence time of renal cancer patients. In addition, the warning prompts provided in accordance with the model results can assist doctors in making preliminary judgments of the patient’s condition and has a certain auxiliary effect on the clinical diagnosis and treatment of cancer and kidney cancers.

INDEX TERMS Kidney cancer, recurrence time, recurrence warning system, convolutional neural network.

I. INTRODUCTION

With the deployment of 5G networks on a global scale, the combination of artificial intelligence and the Internet of Things is becoming increasingly closer to being realized [1]. Especially in the context of big data, the Internet of Things has a wide range of applications in telemedicine, smart medical care, medical monitoring systems, medical equipment, and equipment management [2]. As an emerging field, the Internet of Things, combined with current cutting-edge sensor and network technologies, will greatly improve the lifestyles of people. In particular, the emergence of the Internet of Things technology has promoted further development of smart medical systems, enabling medical data to be maximized in diagnosis and treatment. As a result, the construction of a patient-centric smart medical platform has greater potential. At present, the risk of cancer in humans is constantly increasing, and the early diagnosis and warning of cancer have become the most concerning issues [3]. Therefore, recognizing that the incidence of renal cancer is increasing on a yearly basis, we build an intelligent early warning system based on a framework of the medical Internet of Things to intelligently predict the recurrence of renal cancer from a collection of raw data.

Since entering the 21st century, the application of information technology in all industries has promoted the rapid development of the information industry. The information industry has invested many social resources into providing...
a strong environment for technological development. The continuous development of information technology has not only played an absolute role in promoting the development of emerging industries but has also brought new development opportunities to traditional industries [4]. The utilization efficiency of big data in the medical and health fields has been greatly improved by new development opportunities in combination with information data storage, information transmission management and other traditional medical institution technologies. At the same time, the development of data mining technology has led to further developments in the medical industry, and the application of data mining technology in daily tasks has brought medical industry developments into an information age. The topic of cancer recovery can be traced back to 1971, when the United States first proposed the concept of cancer recovery in its National Cancer Plan [5]. Cenik defined cancer rehabilitation as “helping cancer patients maximize their physical, social, psychological and professional functions under the limitations of cancer itself and cancer treatment” [6]. The main purpose of cancer rehabilitation is to restore the patient’s psychological, physical, and physiological functions well. At the same time, physical function recovery, nutrition, exercise, and cancer pain recovery are also key areas of cancer rehabilitation.

In the past few years, researchers have made many efforts toward tumor cells detection and early cancer diagnosis. Early detection and treatment of cancer result in a higher survival rate, which greatly improves the quality of life of patients, but the factors affecting the postoperative recurrence of tumors in patients are very complex, and quantitative analysis is very challenging [7]. The postoperative recurrence time affects the patient’s quality of life and standard of living, and patient prognostic intervention has a crucial impact on the postoperative recurrence time. Personalized intervention for cancer patients with short recurrence times not only prolongs their survival and quality of life but also makes a certain contribution to the field of cancer rehabilitation [8].

Researchers in the field of cancer prognosis models generally use four machine learning models: logistic regression, neural networks, decision trees, and support vector machines. No researcher conducted a comprehensive study on the prognostic model of cancer patients using a variety of model algorithms. Some researchers used artificial neural network models to jointly detect a few indicators to establish a prognosis model for cancer patients and explore the survival rate of lung cancer patients, but there are shortcomings such as fewer input features and incomplete model algorithms. In this paper, we thoroughly evaluate the various indicators that affect the recurrence of kidney cancer, determine 53 basic impact indicators, and use five model algorithms to conduct research. The prognostic model of lung cancer patients is established by combining input features and changing model algorithm parameters. Section 2 of this article introduces the application of IoT technology in the medical field and the impact of data mining in the field of cancer research. Section 3 introduces the research framework of this article and the impact indicators that affect the recurrence of kidney cancer, related research algorithms, and the design of the final recurrence warning system. Section 4 studies the performance and stability of five machine learning algorithms by collecting data from kidney cancer patients, constructing a kidney cancer-based machine learning algorithm model, comparing and discussing this model, and obtaining the prediction result of the convolutional neural network model. The accuracy rate of this model is significantly higher than that of other models, and its stability is better. The results show that the convolutional neural network can improve the performance of renal cancer recurrence prediction and can introduce more flexible and effective solutions. Aiming at the problem of incomplete assessment of factors affecting renal cancer recurrence, we thoroughly study its pathogenesis and other related factors and finally determine 53 indicators to fully characterize the tendency of patients to relapse. In view of the current instability problems of individual algorithms, the integration of the Internet of Things technology and five algorithms is constructed through a system framework for predicting the tendency of renal cancer recurrence, and the evaluation has verified its advantages. Our system will have a certain auxiliary judgment for the treatment of cancer patients.

II. RELATED WORK

A. THE APPLICATION OF THE INTERNET OF THINGS IN THE MEDICAL FIELD

With the continuous development of science and technology, the internet in a general sense has been unable to meet the needs of the public, so the Internet of Things technology has emerged [9]. At present, the deep integration of the Internet of Things and traditional medicine, the transformation and upgrading of the Internet of Things technology, and the emergence of 5G technology have become important driving forces for the development of the Internet of Things in medical care [10] and a necessary technical system to support smart medical care. Generally, the Internet of Things architecture includes 3 layers, namely, an application layer, a network layer, and a perception layer. The application layer includes 3 layers, which are a domain application layer, a public service layer, and a middleware layer. The network layer is used to provide data transmission support, which is responsible for the network interconnection between things. The perception layer includes physical devices related to the Internet of Things technology. The IoT in medical care is achieved by embedding and equipping a variety of sensors in medical equipment. Then, the IoT integrates itself with the existing internet to achieve a full-scale three-dimensional interaction between hospitals, patients, and medical equipment.

The Internet of Things is a concept that was proposed at the end of the twentieth century. Its basic purpose is to connect to the internet in accordance with the agreed protocol through radio frequency identification (RFID), sensors, GPS, and other sensing devices to realize intelligent identification and management. There are two main types of applications of...
Internet of Things technology in health systems. One type is the use of sensing equipment, including electronic tags and radio frequency identification; the other is mobile medical applications based on mobile smart terminals such as mobile phones and tablets, which mainly includes neuromedicine for patients and the neurological management of things [11], supporting digital collections, processing, storing, transmitting, and sharing medical, equipment, hospital, personnel, and management information within a hospital to realize the visualization of material management, the digitization of medical and procedural information, the conscientization of medical procedures, and the humanization of service communication to meet the needs of medical and health information, medical equipment and supplies, and diagnostic management and monitoring of public health safety. We show the basic structure the medical IoT in Figure 1. Compared with traditional medicine, medical IoT has omnidirectional, multichannel, and full-cycle characteristics that are new and advantageous, and it has quickly become an important supporting technology in the medical field.

![Basic framework of medical IoT](image)

The Internet of Things services in medical and health services include the comprehensive use of optical technology, pressure-sensitive technology, RFID technology, and other technologies combined with a variety of sensors connected through a sensor network under a certain agreement and with the help of mobile terminals, embedded computing devices, and medical information processing. This platform exchanges information. We can divide its structure into four levels: a perception collection layer, transmission access layer, support management layer, and service application layer.

We divide the sensing collection layer into two sublayers: a data collection sublayer and an access sublayer. The data collection sublayer uses various technical means and equipment to transform the objects involved in the network into CPS nodes. The access sublayer transmits the collected data and accesses it through the network layer. Although there are various access methods for the access sublayer, it is necessary to select an appropriate access method according to the different needs of different objects. Cancer rehabilitation medical data include a large number of medical CT images, nutrition program information, and cancer tumor score monitoring levels that require a large amount of information transmissions, high-quality requirements, and fast response speeds. Coaxial or optical cable access is suitable for this type of application. Community health service centers usually manage common diseases. Therefore, their transmission data volume is small, and their place of use is essentially unchanged. Therefore, fixed broadband access is suitable for this type of application.

We divide the network layer into three sublayers: an active CPS network, middleware, and cloud data center. The active CPS network realizes the interconnection and intercommunication between active CPS nodes and other CPS nodes based on the communication network, information center, and network center while ensuring the reliability and accuracy of the signal. On the active CPS network system user management, node monitoring, information processing, and other functions of the active CPS node are designed and implemented. Then, an IoT middleware is established for a specific application. The middleware sublayer mainly realizes the unification of various data formats integrates information and builds a service support platform based on this foundation to provide open interfaces for various services at the application layer for third-party development and application.

We divide the service application layer into two layers: a cloud data center and application system. The cloud data center processes the data passed by the middleware and provides various services. On this basis, we have formed various applications to build an intelligent early warning system for renal cancer recurrence.

The specific application system is the core, which uses the cloud data center for data mining to provide information for cancer patients such as rehabilitation programs and medical decisions. On the one hand, the renal cancer recurrence early warning system based on the Internet of Things technology can update a patient’s condition data in real time; the early warning system tracks the condition of a patient’s tumor condition in real time and predicts the recurrence tendency based on the patient’s historical characteristic data. Ultimately, this early warning result can assist doctors in making early interventions.

The medical Internet of Things can assist in modern medical application scenarios with its excellent technological characteristics such as intelligent perception, service applications, precise control, real-time analysis, and adaptive regulation. Cancer is still a difficult problem that the medical field has not overcome, and the factors that affect cancer are numerous and complex. With the help of the Internet of Things technology, detailed data analysis results can be provided from the analysis of pathological characteristics and lesions, which can help cancer treatment and rehabilitation. Providing guidance can also establish a foundation overcoming cancer problems in the future. The perceptual control technology, network communication technology, and adaptive control technology in the intelligent renal cancer recurrence early warning system designed in this research has matured. However, the in-depth mining of rehabilitation medical data and processing technology of kidney cancer patient data still needs to be improved. Therefore, the study of a recurrence prediction model for renal cancer rehabilitation is the key to designing an intelligent early warning system.
B. CANCER DATA MINING RESEARCH
Cancer has become one of the major diseases affecting human health, and it is one of the main illness-related reasons that most families face poverty. In recent years, with the significant aging of the world’s population, coupled with the influence of factors such as environment and lifestyle, the overall global cancer incidence and mortality rate has increased year by year [12].

The continuous advancement of medical technology has gradually enriched tumor treatment methods, but current treatments cause certain side effects on a patient’s body. The World Health Organization has clearly pointed out that cancer itself is a lifestyle disease, the pathogenesis of cancer is very complicated, and multiple factors may cause cancer [13], [14]. Therefore, while treating cancer, the rehabilitation of patients is becoming increasingly important. Cancer rehabilitation is a newly developed branch of rehabilitation medicine and oncology. In recent years, with the increase in the proportion of cancer survivors, we aim to address the cancer itself. Additionally, rehabilitation needs resulting from the many problems caused by treatment are increasing [15]. In view of the severity of cancer, rehabilitation effects are also more complex, involving psychology, moral philosophy, sociology, nutrition, and other aspects [16]. In the past few decades, the survival rate of cancer patients has increased, and their lifespan has also been extended [17]. For a patient’s rehabilitation process to proceed smoothly, it is necessary to consider their physical, psychological, social, and work health among other factors. Therefore, in recent decades, research on tumor rehabilitation treatment has gradually become popular.

In the field of cancer-assisted diagnosis and treatment, a large number of researchers use data mining to study the prediction of cancer prognosis because it can identify corresponding patterns from complex cancer risk factors [18]. The cancer data mining process is shown in Fig. 2. If doctors can use data mining techniques to predict the relevant conditions of patients, it will be very helpful for patient rehabilitation. Montazeri et al. used naive Bayes, random forest, KNN, AdaBoost, SVM, and other machine learning techniques to predict the new rate of breast cancer and verified that the rule-based TRF classification model performed the best in breast cancer survival prediction accuracy [19]. Weiser et al. used the Cox regression to evaluate prognostic factors through multivariate analysis and used cubic splines to model nonlinear continuous variables. The model was bootstrapped internally, and the performance was evaluated through the consistency index and a calibration curve. The accuracy of colon cancer recurrence was predicted [20]. In 2018, Yue et al. outlined the application of several different machine learning algorithms on the Wisconsin Breast Cancer Database (WBCD) and studied the particle swarm algorithm and its improved algorithms to improve performance [21]. In addition, a large number of related studies on medical care have begun to develop on a continuous basis. Deepa et al. proposed an artificial intelligence system for medical care that used the Ridge-Adaline stochastic gradient descent classifier (RASGD) [22] for the early prediction of diseases and then use RASGD combined with ridge regression to improve the convergence speed of the classifier, thereby obtaining 92% classification accuracy. Abbas et al. proposed a new method called BCD-WERT to predict breast cancer. This approach utilized the extremely randomized tree and whale optimization algorithm (WOA) for efficient feature selection and classification [23]. Compared with other machine learning algorithms, the performance was improved with an accuracy rate of 98.6%. Kamel et al. proposed a gray wolf optimization (GWO) and support vector machine method for breast cancer diagnosis [24]. This method increased the diagnostic accuracy rate by 27.68%. Several researchers explored these types of problems in the past. Table 1 provides a comparison of the previous studies, the solutions proposed, and the associated limitations.

Due to the continuous development of science and technology and peoples’ growing medical needs, medical treatment will be the future direction of tumor treatment. From the accumulation of the most primitive data to the development of a set of clinical diagnosis methods for cancer, it takes a long period of time to complete research and clinical trials. This requires hospital to be in contact with patients to mobilize enthusiasm, participate in in-depth data mining, combine basic research and clinical trials, and make better use of data analysis in clinical diagnoses. Precision medicine will benefit the majority of patients [25], [26].

III. RESEARCH ON EARLY WARNING SYSTEM FOR RENAL CANCER RECURRENCE
A. RESEARCH FRAMEWORK
Fig. 3 shows the basic framework of this research. First, we collect initial renal cancer data, perform data cleaning, integration, selection, and conversion on the initial dataset through data preanalysis and extract meaningful features as the input of the model. Then, five supervised learning prediction algorithms are used to develop prediction models based on different feature sets, and the performance of the
TABLE 1. Literature review summary.

| Authors                  | Problem solved and limitations                                                                 |
|-------------------------|-------------------------------------------------------------------------------------------------|
| Montazeri M, et al. (2016) [19] | Breast cancer survival prediction using TRF. TRF cannot obtain good results when the amount of data are large. |
| Yue W, et al. (2018) [21]   | Breast cancer prediction by PSO and ML on the WDBC dataset. The PSO algorithm easily falls into the local minimum and cannot obtain correct results. |
| Zhou J, et al. (2015) [36]  | Liver tumor segmentation using a three-stage hybrid support vector machine. The method is too computationally expensive when the amount of data are very large. |
| Kamel S R, et al. (2019) [24] | Breast cancer diagnosis by using GWO and SVM. The GWO algorithm has poor global search capability. |
| Deepa N, et al. (2021) [22] | Earlier disease prediction using RASG. It is applicable to the two-class model and the RASGD data with lower feature dimensions. It is impossible to judge its stability on high-level data. |
| Abbas S, et al. (2021) [23] | Breast cancer prediction using a new method called BCD-WERT. The BCD-WERT algorithm may not be able to achieve fast calculation speeds for high-dimensional data. |

five algorithms in terms of prediction accuracy and stability is compared and discussed. This research requires us to comprehensively evaluate the factors that affect the recurrence of renal cancer, which has high value for precision cancer medicine research. The following subsections outline the evaluation indicators, prediction techniques, and preliminary design of the recurrence warning system used in this study.

B. COMPREHENSIVE ASSESSMENT OF INFLUENCING FACTORS OF KIDNEY CANCER

Based on long-term medical consensus, the ASC carcinogenic factor research report and TIES.IO cancer assessment data, we consider more than 125 factors that affect the recurrence of kidney cancer [27] and finally determine 53 factors that affect cancer recurrence through data preprocessing. Then, the analysis is performed. We classify seven main scores: basic score, tumor score, immune score, nutrition score, microenvironment score, psychological score, and aerobic activity score, and collect data from cancer patients and score them. We show the related items of each index score and its weights in Table 2. The weight parameters in the table are the results obtained based on the experience of doctors and experts.

C. PREDICTIVE ALGORITHM TECHNOLOGY

Ordinary linear regression analysis is used to analyze the interaction of multiple variables, and the linear model aims to determine the relationship between multiple explanatory variables and the explained variables [28]. In real life, when we perform statistical analysis on the dependent variable, there is often more than one independent variable that affects the dependent variable. We need to consider the relationship between the independent variable and the dependent variable and establish a regression equation.

$$y_i = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + u_i \quad (1)$$

In the formula, $b_0, b_1, \ldots, b_k$ are the regression coefficients to be estimated and $u_i$ is the random error.

Ridge regression is a modified ordinary least squares estimation [29] that was first proposed by Hoer in 1962. When there is multicollinearity between the independent variables, if a positive constant matrix is added to it, the degree of approaching singularity will be much smaller than that of $X’X$ approaching singularity. Therefore, the ordinary least squares estimation is performed to eliminate the effect of collinearity. The process of eliminating multicollinearity is actually a process of independent variable selection.

Gradient boosting, which was first proposed by Friedman in 2001 [30], is a supervised learning algorithm [31]. When the GBDT model makes a prediction, it first assigns an initial...
value to the predicted value of the sample and then traverses each decision tree. Each tree will adjust the predicted value. The final predicted result is the accumulated result of each decision tree. The basic steps of the algorithm are as follows:

1) Initialization $f_0(x) = 0, t = 1, 2, \ldots, T$;
2) Calculate the residual value:
   \[ r_{ti} = y_i - f_{t-1}(x_i), \quad i = 1, 2, 3, \ldots, m \]  
(2)
3) Fit the residual value $r_{ti}$ and obtain the regression tree through learning $h_t(x)$;
4) Update parameters $f_t(x) = f_{t-1}(x) + h_t(x)$;
5) Obtain regression problem boost tree:
   \[ f_T(x) = f_0(x) + \sum_{t=1}^{T} h_t(x) \]  
(3)

Support vector regression (SVR) is a new generation of machine learning methods. By introducing a kernel function to replace the inner product operation in the high-dimensional space, it solves the problem of nonlinear fitting [32]. In addition, this model is based on the principle of structural risk minimization. The global optimal solution can be obtained, and the nonlinear fitting solution can be prevented from appearing in the local optimal solution, ensuring high fitting accuracy and overcoming the shortcomings of traditional regression fitting methods.

SVR, as a type of model for SVM to address the fitting regression problem, predicts the to-be-predicted vector of the test data by establishing a nonlinear relationship between the to-be-predicted vector in the training data and the support vector [33]. Given a training data set, and $x_i \in X \subseteq \mathbb{R}^n$, $y_i \in Y \subseteq \mathbb{R}^m$, the number of training samples is. SVR transforms the input space into the high-dimensional feature space through the nonlinear transformation defined by the inner product kernel function and obtains the regression function in the high-dimensional feature space, as follows:

\[ f(x) = \omega \cdot \phi(x) + b \]  
(4)

In the formula, $\phi(x)$ is the nonlinear mapping, $\omega$ is the weight coefficient, and $b$ is the bias term.

A convolutional neural network was proposed by Hinton et al. in 2006. It has multiple hidden layer structures and can perform repeated multiple training on the input vector of the network to improve the accuracy of classification or prediction [34]. A convolutional neural network is one of the representative algorithms of deep supervised learning, and it is a neural network with a convolutional structure. A basic flowchart of this method is shown in Fig. 4.

The steps are summarized as follows:

1) The original dataset is obtained, and preliminary processing is performed on the dataset according to the pre-designed evaluation criteria to ensure the basic completeness and correctness of each data item in the dataset.
2) Data preprocessing is performed on the filtered original dataset, and the collected dataset is divided into a training set and a test set according to a certain proportion.
3) The preprocessed training samples are used to train the sparse autoencoder in an unsupervised manner, and batch training and gradient descent are used to obtain the extremely small optimal hidden layer parameters. As a result, a basic convolutional neural network model that achieves a certain prediction accuracy is obtained.
4) The optimal hidden layer parameters of the trained sparse autoencoder are used as the initial weight matrix and bias of the hidden layer of the deep neural network to complete the initialization of the convolutional neural network. Finally, the test set is used as the input dataset in the model, and the model prediction results are obtained.
5) All the results in the entire prediction model in the form of graph are counted and output in the form of multiple graphs.

**FIGURE 4. Convolutional neural network flow chart.**

**D. RECURRENCE WARNING SYSTEM DESIGN**

In this paper, by processing the initial patient dataset obtained from the hospital, we use machine learning algorithms to predict and evaluate the patient dataset, obtain the individual physical assessment of each patient, predict the patient's renal cancer recurrence results, and finally establish the patient's cancer recurrence prediction. As shown in Figure 5, the system includes four basic functional modules: a patient data acquisition module, a data sample evaluation index selection module, a renal cancer recurrence time prediction module, and a result output module.

The main function of the data acquisition module is to perform basic processing on the initial patient dataset obtained from the hospital, including missing value processing, discrete value processing, and outlier processing to ensure that the dataset can be reached as the input part of the system. In addition, the module can perform the basic detection of various data indicators on the dataset during operation, preventing data confusion and erroneous data caused by human factors, which will affect the prediction results of the system.

The main function of the data sample evaluation indicator selection module is to select the indicators that may have a greater impact on the prediction results from the indicators in the original dataset and improve the accuracy and efficiency of the system’s prediction of results by reducing the basic data items that need to be measured.

The renal cancer recurrence time prediction module contains five algorithms to predict the patient recurrence time.
The results are obtained by running Bayesian ridge regression, ordinary linear regression, support vector regression, gradient boosting tree, and convolutional neural network in parallel. The results of the algorithm are compared, analyzed, and evaluated. The best prediction result is selected and output.

In addition to the best prediction result of a patient’s renal cancer recurrence time, the result output module also contains a nutritional rehabilitation plan for the patient’s personal physical state, which helps the patient perform postoperative rehabilitation by analyzing several major nutritional factors that affect cancer.

**FIGURE 5.** System functional structure diagram.

### IV. EXPERIMENTAL ANALYSIS

#### A. DATA SET DESCRIPTION

This study collects data on kidney cancer patients. The recurrence time in the data ranged from 6 months to 60 months. Considering that the time of 6 months may be within the treatment period, the case of recurrence found within 6 months is considered invalid data. Seven influential factors, including basic score, tumor score, immune score, basic nutritional score, psychological score, microenvironment score, and exercise score are determined, and a recurrence warning model is established to predict the patient’s recurrence time.

According to the index scoring standard of renal cancer patients, the comprehensive score of the patient is combined with the early warning system for prediction and is provided to the doctor for reference to prolong the life of the patient. The recurrence data and index scores are shown in Figure 6.

#### B. DATA PREPROCESSING

The Pearson correlation, also known as the product difference correlation, is a method of calculating linear correlation proposed by British statistician Pearson in the 20th century. The Pearson correlation coefficient is often used in the following situations.

1. There is a linear relationship between the two variables, and both are continuous data.
2. The population of the two variables is normally distributed or a unimodal distribution close to normal.
3. The observations of the two variables are paired, and each pair of observations is independent of each other.

Twelve variables and recurrence times are entered for correlation analysis to reduce dimensionality and the correlation between each variable and recovery time is obtained.

Table 3 shows that for the twelve input indicators, the indicators and the recurrence time are all positively correlated, but the overall correlation coefficient is not high. Among them, the correlation coefficient between surgical age, sex, and recurrence time was extremely low.

After the correlation analysis between the input indicators and the recurrence time of renal cancer, it can be seen that age and sex have a minimal impact on the recurrence time of renal cancer, so the age and sex in the first selected input sample set are eliminated. The final input sample set is basic score, tumor score, immune score, psychological score, basic nutritional score, nutrition comparison score, safe intake score, total nutrition score, microenvironment score, and aerobic exercise score. The principle sample set is shown in Table 4.

### TABLE 3. Pearson correlation coefficient table of each sample input index and recurrence time.

| Index          | Tumor         | Base          | Nutrition     |
|----------------|---------------|---------------|---------------|
| Coefficient    | 0.163883      | 0.141286      | 0.130389      |
| Psychological  | 0.064263      | 0.050528      | 0.041019      |
| Immune         | 0.116865      | 0.113712      | 0.079997      |
| Microenvironment| 0.039811    | 0.011967      | 0.004773      |

The Spearman’s rank correlation coefficient is used to estimate the correlation between two variables $X$ and $Y$, where the correlation between variables can be described by a monotonic function \( [35] \). If the same two elements do not exist in the two sets of values of two variables, then when one of the variables can be expressed as a good monotonic function of the other variable (that is, the changing trends of the two variables are the same). The $\rho$ value between two variables can reach $+1$ or $-1$.

The Spearman rank correlation coefficient has less strict requirements on data conditions than the Pearson correlation coefficient, as long as the observations of the two variables are paired rank rating data or the observations of continuous variables.

From the comparison of Table 3 and Table 4, it can be seen that the trends of the Spearman correlation and Pearson correlation are approximately the same, but for the twelve input indicators, the overall characteristics are the same. From this, we can determine the ordering relationship between the
The scatter plots in Figure 6 show the correlation between the scoring index and recurrence time.

Twelve input indicators and the correlation before output; that is, the tumor score has the greatest impact on the recurrence time, while the age and sex have little effect. Therefore, if there are too many impact indicators in the follow-up research process, the age and sex of the operation can be ignored to improve the efficiency of the model to a certain extent.

After the correlation analysis between the input indicators and the recurrence time of renal cancer, it can be seen that age and sex have a minimal impact on the recurrence time.
of renal cancer. Therefore, the age and gender in the input sample set selected at the beginning were eliminated, and seven indicators are selected. The final input indicators are basic score, tumor score, immune score, psychological score, basic nutrition score, microenvironment score, and aerobic exercise score.

### C. RESULT ANALYSIS

For the collected kidney cancer patient dataset, the data preprocessing step is first performed. The recurrence prediction model is constructed based on five machine learning algorithms, and the model is further used to verify the performance of the model on the training data. The visualized prediction effect of the five model analyses is shown in Fig. 7.

From the perspective of the recurrence time of renal cancer, it is reasonable that the absolute error of the prediction value is less than 6 months. Figure 7 shows the prediction results of five algorithms on the kidney cancer test training set, including multiple linear regression (82.29%), support vector regression (87.85%), Bayesian ridge regression (85.46%), gradient boosting tree (88.58%), and convolutional neural network (92.35%). The results show that the accuracy rate is essentially stable. From the perspective of renal cancer recurrence time, these five algorithms can achieve at least 80% accuracy of recurrence time prediction. The average accuracy of the convolutional neural network reaches 92.35% and shows good performance and stability. Figure 8 shows a comprehensive comparison of the predictions of the five algorithms.

It can be seen from Fig. 8 that on the test set, except for the GBR and convolutional neural network algorithms, the other three algorithms overfit the data. Through the selection of the kernel function and the adjustment of the parameters, GBR shows good stability. In the actual debugging and testing process, the prediction result of the convolutional neural network can reach 96.1% at the highest and 82.3% at the lowest. The comparison and analysis of the results with the other four algorithms show that the convolutional neural network algorithm has the best effect, and the other algorithms have extreme conditions and instability. When analyzing the kidney cancer patient dataset, whether for existing patients, patients with similar data, or new patients, it is very important to choose a stable algorithm. Convolutional neural network algorithms are compared under the above conditions. The remaining 4 algorithms all have advantages.

### D. MODEL EVALUATION

Several types of evaluation methods are further introduced to evaluate the model. Table 2 below shows the error analysis details of the model evaluation. The MSE, MAE, EV, RMSE,
Table 5 shows that the model with the best effect is the fully convolutional neural network algorithm. When using the model to fit the training data, the effect of the Bayesian ridge is slightly worse. However, other regression algorithms represented by support vector regression have a serious problem; that is, when given specific data, overfitting will occur, which makes the prediction situation abnormal. The convolutional neural network algorithm has relatively large running time loss, but the overall prediction effect is the best.

However, the prediction of unknown data is what can truly reflect the regression prediction effect. From the actual effect, the convolutional neural network has advantages in stability and accuracy compared with the above algorithms.

Combined with the experimental results, as shown in Table 6, the convolutional neural network model shows better stability than the other machine learning regression algorithms used in this paper. When training the data, the prediction effect of the convolutional neural network is the most stable among the five models.

As shown in Table 6, the MAE, RMSE, MAPE, and accuracy of the convolutional neural network are 2.54, 10.35, 1.02, and 92.35%, respectively. On the test set, the convolutional neural network shows the lowest error and the highest accuracy among the five algorithms. Therefore, the CNN-based cancer recurrence time prediction model has excellent prediction performance and prediction accuracy on the renal cancer sample dataset. This system can help the prognosis of kidney cancer patients. It has certain guiding significance in the field of medical cancer rehabilitation.

V. CONCLUSION

In this paper, we collect the data of more than 700 renal cancer patients and analyze seven indicators of renal cancer from seven aspects: tumor module, basic module, microenvironment module, immune module, nutrition module, psychological module, and exercise module. We construct five learning algorithms for renal cancer recurrence prediction models to predict the time of renal cancer recurrence. Through model evaluation and comparison, it is found that the prediction accuracy of the convolutional neural network is 92.35%, which is significantly higher than other models, and the stability is higher. These results show that the renal cancer recurrence prediction model constructed in this paper is suitable for predicting the recurrence time of cancer patients. Based on the predictive model, an intelligent early warning system for renal cancer recurrence is designed to assist doctors in the initial diagnosis and improve the survival rate of patients through corresponding clinical interventions. In future work, the accuracy and stability of the algorithm model are still the focus of research.

ETHICS STATEMENT

The study was approved by the Shanxi Provincial People’s Hospital, Taiyuan, China.

Participants in this study were informed and consented.

REFERENCES

[1] A. Ghosh, D. Chakraborty, and A. Law, “Artificial intelligence in Internet of Things,” CAAI Trans. Intell. Technol., vol. 3, no. 4, pp. 208–218, Dec. 2018, doi: 10.1049/trit.2018.1008.

[2] C. González García, E. Núñez-Valdez, V. García-Díaz, C. P. G-Bustelo, and J. M. Cueva-Lovelle, “A review of artificial intelligence in the Internet of Things,” Int. J. Interact. Multimedia Artif. Intell., vol. 5, no. 4, p. 9, 2019, doi: 10.9781/ijima.2018.03.004.

[3] D. Metcaif, S. T. J. Milliard, M. Gomez, and M. Schwartz, “Wearables and the Internet of Things for health: Wearable, interconnected devices promise more efficient and comprehensive health care,” IEEE Pulse, vol. 7, no. 5, pp. 35–39, Sep. 2016, doi: 10.1109/MPUL.2016.2592260.

[4] D. Shi, J. Che, Y. Yan, B. Peng, X. Yao, and C. Guo, “Expression and clinical value of CD105 in renal cell carcinoma based on data mining in the cancer genome atlas,” Experim. Therapeutic Med., vol. 17, pp. 4499–4505, Apr. 2019, doi: 10.3892/etm.2019.7493.

[5] J. B. Fu, D. M. Molinares, S. Morishita, J. K. Silver, S. S. Dibaj, Y. Guo, and E. Brueger, “Retrospective analysis of acute rehabilitation outcomes of cancer inpatients with leptomeningeal disease,” PM&R, vol. 12, no. 3, pp. 263–270, Mar. 2020, doi: 10.1016/j.pmrj.12207.

[6] F. Cenik, B. Mähr, S. Palma, M. Keilani, T. Nowotny, and R. Crevenna, “Role of physical medicine for cancer rehabilitation and return to work under the premise of the ‘Wiedereingliederungsteilzeitgesetz,’” Wiener klinische Wochenschrift, vol. 131, nos. 19–20, pp. 455–461, Oct. 2019, doi: 10.1007/s00508-019-1504-7.

[7] A. G. Waks and E. P. Winer, “Breast cancer treatment: A review,” Jama, vol. 321, no. 3, pp. 288–300, Jan. 2019, doi: 10.1001/jama.2018.19233.

[8] A. Cheville, R. Crevenna, M. Korpan, and M. Quittan, “Cancer rehabilitation,” Brudonn’s Phys. Med. Rehabil., vol. 35, no. 4, pp. 568–593, Oct. 2021, doi: 10.1080/16501970310000511.

[9] F. Xia, L. T. Yang, L. Wang, and A. Vinel, “Internet of Things,” Int. J. Commun. Syst., vol. 25, no. 9, pp. 1101–1102, Sep. 2012, doi: 10.1002/dac.2417.
