

ORIGINAL RESEARCH

Early warning model of recurrence site for breast cancer rehabilitation patients based on adaptive mesh optimization XGBoost

Aimin Yang^{1,2,3,4,5}, Zezhong Ma^{2,3,4,5,*}, Jian Wang^{2,5}, Shujuan Yuan^{1,3,4,5}, Zunqian Zhang^{1,3,4,6}, Dianbo Hua⁷

¹Hebei Engineering Research Center for the Intelligentization of Iron Ore Optimization and Ironmaking Raw Materials Preparation Processes, North China University of Science and Technology, 063210 Tangshan, Hebei, China

²Hebei Key Laboratory of Data Science and Application, North China University of Science and Technology, 063210 Tangshan, Hebei, China

³The Key Laboratory of Engineering Computing in Tangshan City, North China University of Science and Technology, 063210 Tangshan, Hebei, China

⁴Tangshan Intelligent Industry and Image Processing Technology Innovation Center, North China University of Science and Technology, 063210 Tangshan, Hebei, China

⁵College of Science, North China University of Science and Technology, 063210 Tangshan, Hebei, China

⁶School of Metallurgy and Energy, North China University of Science and Technology, 063210 Tangshan, Hebei, China

⁷Beijing Sitairui Cancer Data Analysis Joint Laboratory, 101100 Beijing, China

*Correspondence

Mazz@stu.ncst.edu.cn
(Zezhong Ma)

1. Introduction

With the continuous improvement of living standards, the environment and lifestyle of humans have also changed and improved accordingly, but the ensuing cancer issue still remains one of the most serious problems threatening human health. The “Global Cancer Incidence and Mortality in 2018” (GLOBOCAN 2018) [1, 2] estimated that there were approximately 18.1 million new cancer cases and 9.6 million cancer deaths worldwide in 2018, with a high incidence of cancer in both developed and developing countries. According to data released by the World Health Organization’s International Center for Research on Cancer, cancer accounts for ~13% of all deaths every year. Although there have been important breakthroughs in cancer treatment and related technology, the results are still far from expectations because most cancer cases are still incurable.

Abstract

Objective: Breast cancer is a malignant disease with a high mortality rate. Using postoperative rehabilitation data of breast cancer patients, this study explored the effects of immune, tumor, microenvironmental, psychological, nutritional, aerobic exercise and advanced work indexes on the rehabilitation of breast cancer patients. To determine the weight of the impact of different indications on cancer recovery. By combining the adaptive grid optimization algorithm with the XGBoost (Extreme Gradient Boosting) algorithm, an intelligent prediction model for breast cancer rehabilitation was constructed using patients indexes as input and the recurrence location as the output. Our results showed that the model constructed in this study could effectively predict cancer cell metastasis during breast cancer recurrence in recovered patients. Compared with artificial intelligence algorithm models such as neural network algorithm, support vector machine algorithm, gradient boosting tree algorithm and Adaboost, the model demonstrated a forecast accuracy rate of >93%. The model established in this study could effectively predict the recurrence position of breast cancer and provide an auxiliary reference for doctors to treat breast cancer patients more effectively.

Keywords

Breast cancer; Adaptive mesh optimization; XGBoost; Location of recurrence

Cancer cells proliferate via uncontrolled rapid mitosis [3, 4], during which the cells secrete specific substances that dissolve and destroy surrounding tissues. Researchers at The Ohio State University have found that cancer cells still survived and grew in tumor tissues with low oxygen levels (hypoxic conditions), providing a basis for developing new agents able to reverse hypoxia-related pathways to inhibit tumor growth. However, the main practical issue is evaluating their effectiveness in clinical trials based on traditional therapeutic efficacy indicators such as tumor size, cell proliferation and apoptosis markers, which are sometimes of little significance when assessing tumor invasion. Since the movement of cancer cells cannot be measured directly in a patient, the movement inferred from histological analyses might not be reliable, considering that the metastatic evolution of cancer is often a very indirect indicator of cell invasion.

In 1971, the United States first proposed the concept of “tumor rehabilitation” [5–7] with the main purpose of helping cancer patients restore their psychological and physical functions despite the disease conditions and limited treatment. Cancer recovery mainly depends on the nature of the tumor and its stage [8, 9]. Early detection and rehabilitation of cancer have greatly improved patients’ survival rates and quality of life. However, the factors affecting postoperative recurrence in cancer patients are complex, making it very challenging to predict their postoperative condition. Based on evaluation data from cancer patients, scholars have classified and predicted whether a tumor is benign or malignant, postoperative recurrence time and the type of tumor. The development and implementation of machine learning and artificial intelligence is the current research trend in medicine [10, 11] and are slowly becoming integral in cancer-assisted diagnosis and treatment [12, 13]. It is reported that ~90% of cancer-related deaths are due to cancer cell metastasis, making it the most difficult problem to overcome, for which post-treatment rehabilitation is gaining more and more attention as an attempt to improve the patients’ quality of life.

With the gradual increase in cancer incidence, predicting cancer recovery with traditional methods and doctors’ personal experience can no longer meet patients’ needs. Presently, data mining, artificial intelligence, deep learning and intelligent prediction models are being used, often in combination, to more accurately estimate patients’ cancer recovery and predict potential recurrence [14, 15].

Based on the existing unsolved issues and patients’ needs, this paper proposes a recurrence prediction model for cancer patients’ rehabilitation based on the adaptive grid optimization algorithm combined with the XGBoost algorithm. Section 2 introduces the relevant information on sample set construction, section 3 introduces the proposed algorithm and model, section 4 discusses the results and compares the advantages of the proposed model with other algorithms, and section 5 provides a summary of this study and an outlook for the future.

2. Sample set introduction

Based on the long-term medical consensus, the ASCO (American Society of Clinical Oncology) carcinogenic factor research report and the cancer assessment data of TIES.IO (Beijing Stairui Health Technology Co., Ltd.) selected the following 12 factors affecting cancer recurrence: gender, age, basic score, tumor score, immune score, basic nutrition score, nutritional comparison score, safe intake score, total nutrition score, microenvironment score, psychological score, aerobic activity score. These data were collected, and the Pearson’s correlation coefficient, Spearman correlation coefficient and other analytic methods were used to determine the correlation of the sample input indicators, which finally identified the following 6 factors, tumor score, immune score, basic nutrition score, psychological score, microenvironment score, aerobic exercise and advanced behavior, as being the most influential factors.

Through data collection and correlation analysis, it was observed that the cancer conditions of patients were related to their own physical condition, daily activities, dietary intake,

environmental conditions and psychological factors, while there was no significant difference between gender and other factors [16]. We collected the postoperative follow-up data of a large cohort of breast cancer patients. In this dataset, tumor module, immune module, nutrition module, psychological module, microenvironment module, aerobic exercise and advanced homework all had multiple associated items. The numbers in parentheses indicated the weights of the associated items, which were based on the experience of six cancer specialists. The weight of each module represented the association between the module and liver cancer recurrence, and the weight was scored from 0–10 points, where 0 points indicate no correlation with tumor recurrence, and 10 points indicate a direct correlation. The breast cancer patient index evaluation criteria are shown in Table 1.

The age at surgery of cancer patients in the dataset varied considerably. Considering the significant deviation in the model’s prediction for patients under the age of 25 and the presence of other systemic diseases for those over the age of 65 [17], only cancer patients between the age of 25 and 65 were considered in this study. As shown in Table 2, the influencing factors of cancer recurrence were divided into 6 first-level and 45 second-level indicators. The weight of each second-level indicator was 10 at the highest and 1 at the lowest. They were estimated using the following equation:

$$I = \sum_{i=1}^n x_i w_i / \sum_{i=1}^n w_i \quad (1)$$

Here, x_i represents the i -th value of the second-level indicator under the first-level indicator, and w_i represents the weight of the i -th second-level indicator under the first-level indicator.

3. XGBoost algorithm based on adaptive grid parameter optimization

Compared with Gradient Boost Decision Tree (GBDT), the XGBoost algorithm had optimization improvements in two key positions, making it more suitable for classification scenarios [18, 19], with the following two advantages:

(1) The XGBoost algorithm adds a regularization term to the objective function, making overfitting less likely to occur during model training;

(2) GBDT uses the first-order derivative in the optimization process, which defines loss function more accurately than XGBoost performs the second-order Taylor expansion on the objective function.

Based on these improvements, the XGBoost model performs better than GBDT in breast cancer classification. The core idea of the XGBoost algorithm is to continuously add a decision tree and perform feature splitting to form a split tree. Each time a new decision tree is added, a new function is learned to fit the residuals of the previous prediction round.

As the ensemble model of the decision tree, the predicted value of each tree in the total number of t trees for the sample is used to predict the sample in the XGBoost system. The objective function in the prediction model of cancer patient recurrence location is defined as:

TABLE 1. Breast cancer patient index evaluation criteria.

First-Level Indicator	Related terms and Weights
Immune Score	CD3 + CD4 + CD8 + /CD45 + (4); CD3 + CD4 + /CD45 + (8); CD4 + /CD8 + (10); CD3 + CD16 + CD56 + /CD45 + (6); CD3 - CD56 + (5); CD4 + CD25 + (1); Exercise ECG (X ± SD) (2); Sports Leather (X ± SD) (2).
Tumor Score	Size (10); Placeholder (10); Violate the relationship (10); Angiogenesis (10); Pathological typing (3); CTC value (9); Differentiation (10); Mutation target (1).
Basic Nutrition Score	Total nutrition (6); Balanced nutrition (3); Nutrition safety assessment (5); Cancer cell proliferation (10); Immune cell proliferation (10); Angiogenesis (8); Amino acid evaluation (5); Proteomics evaluation (10).
Psychological Score	Life event scale (1); Cornell Medical Index (2); Self-rating anxiety scale (5); Self-rating depression scale (5); Baker Anxiety Scale (5); Baker Depression Questionnaire (5); Pittsburgh sleep Quality index (4); Texas Social Behavior Questionnaire (3); Family function assessment (1); Exercise ECG (X ± SD) (2); Sports Leather (X ± SD) (2);
Microenvironment Score	O ₂ (3); pH value (4); Interstitial pressure (2); Inflammatory response (7); Vascular permeability (6); CTC value (9); Proteomic analysis (8).
Exercise and Advanced Work	Aerobic exercise (4); Advanced social work (3); Texas Social Behavior Questionnaire (3).

CD: Cluster of Differentiation, ECG: electrocardiogram, SD: Stable Disease; CTC: Circulating Tumor Cell.

$$Obj(\theta) = \sum_{i=1}^n l(y_i, y_l) + \sum_{k=1}^K \omega(f_k) \quad (2)$$

Here, \hat{y}_i represents the predicted value of the model, y_i represents the category label of the i sample, and f_k represents the k tree model. Formula (2) mainly consists of two parts, whereby one part is the loss function, and the other part is the regularization term. The regularization term $\omega(f_k)$ is given by equation (3):

$$\omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3)$$

Here, γ represents the penalty coefficient, T represents the number of leaf nodes of each tree, and W_j represents the set of scores of the leaf nodes of each tree. In the algorithm, the regularization term is mainly used to control the complexity of the model. On the premise of ensuring the accuracy of the training samples, it can reduce the complexity of the model, avoid overfitting, and improve generalization ability.

The loss function is mainly used to measure the difference between the predicted score of the algorithm and the true score. The following steps are used to train the data and identify parameters minimizing the objective function. Here, the objective function is optimized using the additive training distribution. The steps are to optimize in order, from the first tree to the k -th tree:

$$y_i^{(t)} = \sum_{k=1}^t f_k(x_i) = y_i^{(t-1)} + f_t(x_i) \quad (4)$$

After the last tree is $f_t(x_i)$ optimized, an optimal tree is

obtained, which is the tree that can minimize the set objective function on the basis of $f_{t-1}(x_i)$. The objective function of Taylor expansion can be defined as:

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^n \omega(f_i) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \omega(f_t) + constant \quad (5)$$

Here, constant represents the complexity of the previous tree. When MSE is used as the loss function, the formula can be transformed into:

$$Obj^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) + constant \quad (6)$$

Here,

$$g_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(\hat{y}_i^{(t-1)}) \quad (7)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(\hat{y}_i^{(t-1)}) \quad (8)$$

The goal of the XGBoost algorithm is to minimize the objective function. Thus, the constant term can be removed to obtain:

$$\sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t) \quad (9)$$

After the loss function is set, another definition of the tree

can be made with the following formula:

$$f_t(x) = \omega_{q(x)}, \omega \in R^T, q(x) : R^d \rightarrow \{1, 2, \dots, T\} \quad (10)$$

Here, a tree has T leaf nodes, and the values on these nodes form a T -dimensional vector ω , with $q(x)$ as a map. The sample to be tested is then divided into a certain leaf node, making $\omega_{q(x)}$ the tree to the sample. Smaller values of γ and λ in the regularization formula (2) lead to a more complex model. The formula can be defined as follows:

$$\begin{aligned} Obj^{(t)} &\approx \sum_{i=1}^n \left[g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2 \right] + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \\ &= \sum_{j=1}^T \left[\left(\sum_{I \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{I \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \end{aligned} \quad (11)$$

Upon simplifying the formula, we obtain:

$$Obj^{(t)} = \sum_{j=1}^T \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma \quad (12)$$

After the structure of the t tree is determined, the values of G_j and H_j can be determined. At this time, the optimal value of each leaf node and the value of the corresponding objective function can be calculated using the following formulas:

$$\omega_j^* = \frac{G_j}{H_j + \lambda} \quad (13)$$

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (14)$$

Based on the XGBoost algorithm, we adopted the adaptive mesh optimization method, a semi-automatic parameter optimization method, which can effectively find the corresponding adaptation parameters for different data [20, 21], using the following steps:

- (1) First grid searches the algorithm framework parameters, while the other parameters take fixed values;
- (2) Obtain the optimization result in the above step, add the parameter setting value, and grid search the minimum subtree weight sum and maximum depth parameter of the tree;
- (3) Take the optimization results in the previous two steps, add the parameter values, and grid search for the gamma node parameters;
- (4) Take the optimization results in the above steps, add the parameter setting value, and grid search the sampling rate and feature sampling rate parameters of each tree to the sample;
- (5) Get the optimization result in the above-mentioned steps, add the parameter setting value, and grid search the regularization term parameter;
- (6) Take the optimization results in (1) to (5), add the parameter values, and grid search for the learning rate parameter.

By combining the adaptive mesh optimization algorithm and the XGBoost algorithm, we constructed an XGBoost early warning model for locating recurrence in cancer rehabilitation patients based on adaptive grid parameter optimization. The model structure diagram is shown in Fig. 1.

4. Analysis of experimental results

4.1 Experimental results of XGBoost based on adaptive mesh optimization

According to the evaluation criteria of each indication described above and the constructed prediction model, the data of 11,470 breast cancer patients were selected for model training and testing. The sample input indicators consisted of the following six indicators: immune score, tumor score, microenvironment score, psychological score, nutrition score, aerobic exercise and advanced score. Then, the breast cancer cells were transferred as the output indicator, and the constructed adaptive mesh optimization algorithm combined with the XGBoost prediction model was used for training. The parameter optimization process is shown in Fig. 2.

Fig. 2 shows that the model accuracy was effectively improved after optimizing the seven key parameters of the XGBoost algorithm through an adaptive mesh optimization algorithm.

After searching through the adaptive mesh optimization algorithm, we determined the optimal parameters of the model as: n-estimators = 100, max-depth = 6, min-child-weight = 2, gamma = 0.0, reg-alpha = 0.0001, reg-lambda = 0.0001, colsample-bytree = 0.7, subsample = 0.8, and learning = 0.3.

Then, 1170 patients were selected as the model test data of this experiment to test the established AMO-XGBoost (Adaptive Mesh Optimization - XGBoost) model. The test results are shown in Fig. 3.

Fig. 3 shows that the established AMO-XGBoost early warning model had better effects than the basic XGBoost. The comparison results of the two models are shown in Table 2.

TABLE 2. Comparison of prediction model indexes between the AMO-XGBoost algorithm and XGBoost algorithm.

Algorithmic model	AMO-XGBoost	XGBoost
Accuracy	93.07%	92.58%
F1	96.22%	95.60%
Precision	93.69%	92.54%
Recall	98.89%	98.87%

AMO-XGBoost: Adaptive Mesh Optimization - XGBoost.

As can be seen from Table 2, the AMO-XGBoost prediction model could effectively predict the recurrence location of breast cancer for 1170 patients, further improving the accuracy of the XGBoost algorithm model and effectively predicting the recurrence location of cancer patients for rehabilitation, and could be used as a reference by doctors during patients' rehabilitation.

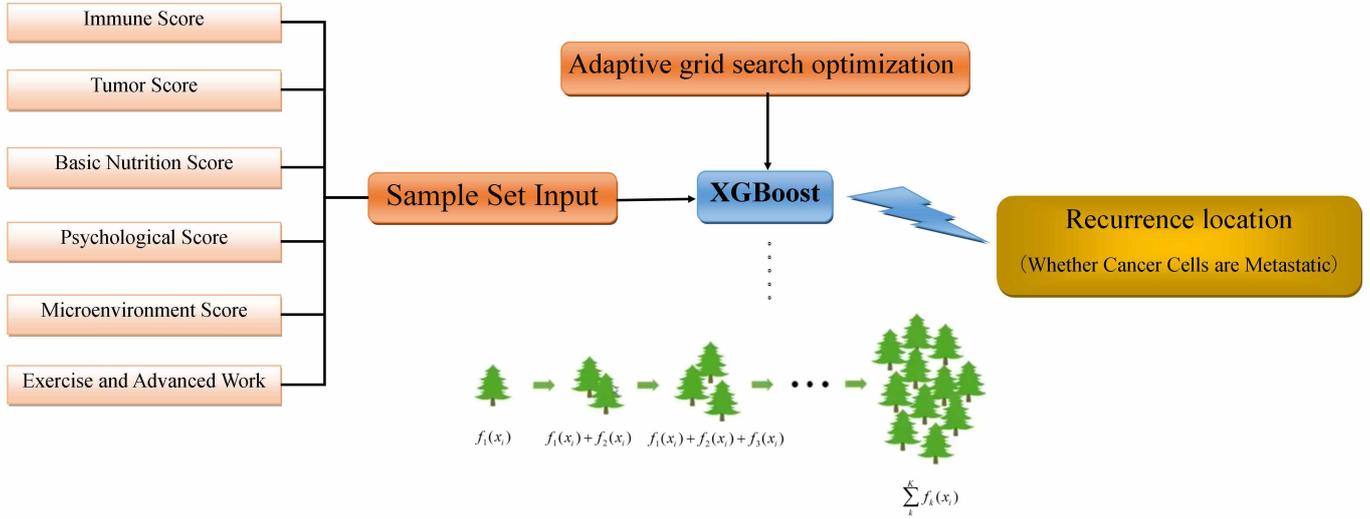


FIGURE 1. Structure of prediction model for recurrence location of breast cancer.

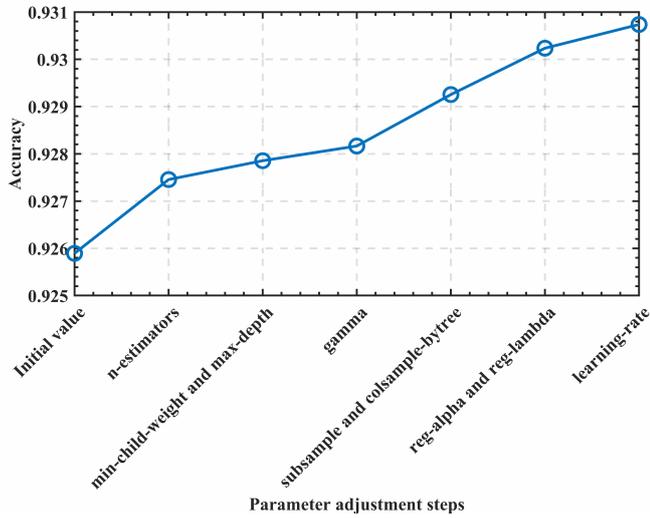


FIGURE 2. Parameter optimization process of adaptive online course search algorithm.

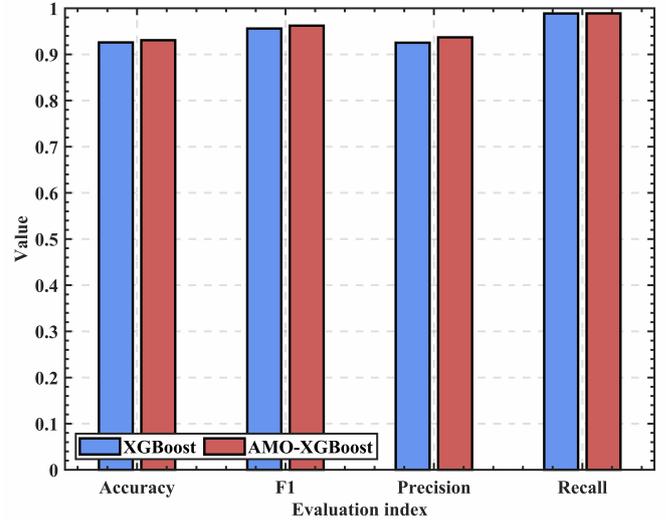


FIGURE 3. Results of XGBoost prediction model based on adaptive mesh optimization (AMO-XGBoost).

4.2 Comparative analysis of experimental models

Apart from comparing the AMO-XGBoost algorithm model with the XGBoost algorithm model, it was also compared with more machine learning algorithm models, such as the Support Vector Machine(SVM), Gradient Boosting Decision Tree (GBDT), Multilayer Perceptron (MLP) and AdaBoost algorithm models, to determine their accuracy in recurrence location. The accuracy comparison of the six algorithms is shown in Fig. 4.

Fig. 4 shows that our proposed breast cancer patient rehabilitation recurrence location model had the highest accuracy compared with other machine learning models. Only the XGBoost basic algorithm had almost similar accuracy to our proposed algorithm, while the other algorithms were significantly inferior. Next, we compared the precision of various models (Fig. 5).

Fig. 5 shows that the proposed AMO-XGBoost algorithm had obvious precision advantages, effectively confirming the

application value of the prediction model. Next, we compared the F1 and Recall values of the models. The visualization results are shown in Fig. 6 (a) and (b).

Fig. 6 shows that the visualization results of the F1 and Recall value confirmed the validity of the proposed model. Compared with other common machine learning models, the proposed model can be used in auxiliary diagnosis and prediction of recurrence location of breast cancer patients. It had better precision, higher Recall and F1 value. Its Precision value is shown in Table 3 below.

The Area Under Curve and Percision-Recall Curves curves also confirmed that the AMO-XGBoost breast cancer rehabilitation prediction model was superior to the basic artificial intelligence algorithm, to a certain extent, and could effectively assist doctors during clinical diagnosis and treatment. AUC curve and PR curve are shown in Fig. 7 (a) and (b). The AUC value is shown in Table 4.

The 9-fold cross-validation was adopted to further verify the validity of the proposed model. According to our verification

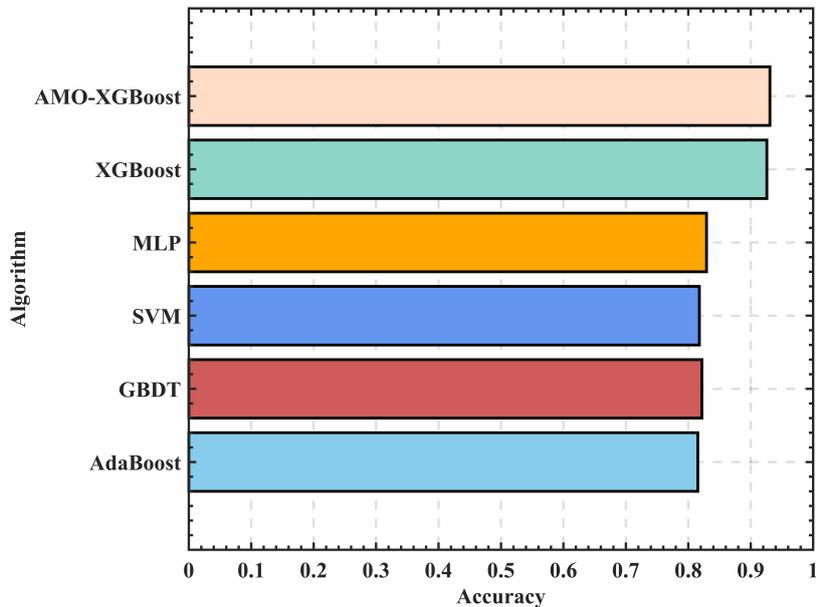


FIGURE 4. Accuracy comparison between the AMO-XGBoost algorithm and other machine learning algorithms. AMO-XGBoost: Adaptive Mesh Optimization-XGBoost; MLP: Multilayer Perceptron; SVM: Support Vector Machine; GBDT: Gradient Boosting Decision Tree.

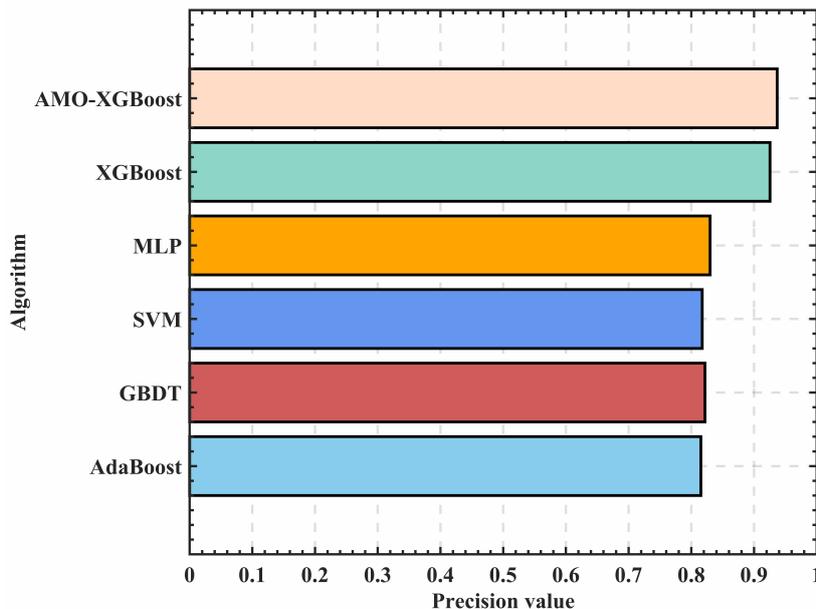


FIGURE 5. Precision comparison between the AMO-XGBoost algorithm and other machine learning algorithms.

results, the accuracy of the AMO-XGBoost algorithm was $>90\%$ effective, indicating that the model could effectively perform with the same type of external data. The 9-fold cross-validation results are shown in Table 5.

Through model construction and independent validations, we determined the performance of the AMO-XGBoost model for predicting the recurrence location of breast cancer patients, which was further confirmed by comparative experiments. Overall, the results showed that the proposed model could effectively help diagnose cancer recovery patients. The probability of cancer cell metastasis when cancer recurred using the model could reach a prediction accuracy of $>93\%$, suggesting

great potential in prolonging the patients' survival. After model testing, we conceived the interface for the forecast model.

4.3 Conception of prediction model of breast cancer recurrence location

Considering the difficulty of applying artificial intelligence in medicine, it is important to inform doctors about the issues of artificial intelligence diagnostic-assisted decisions through models. We conceived an auxiliary diagnosis interface based on the proposed model (Fig. 8), which can effectively help doctors understand this experimental model and assist doctors

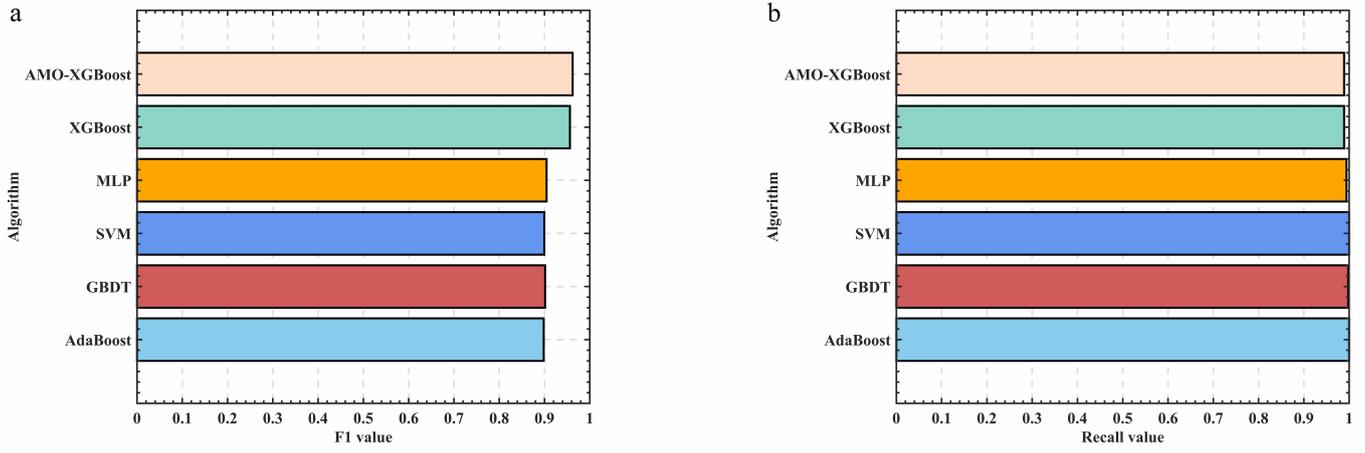


FIGURE 6. Comparison of F1 (a) and recall (b) value between the AMO-XGBoost algorithm and other machine learning algorithms.

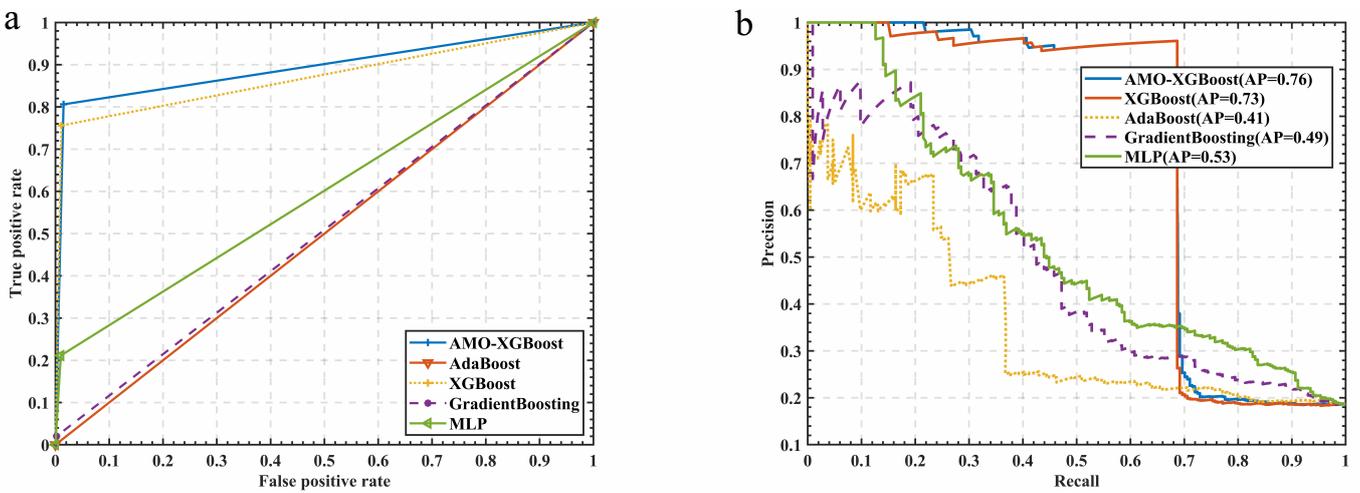


FIGURE 7. AUC and PR visualization results. (a) AUC Curve graph. (b) PR Curve graph.

Visual Interface for Auxiliary Diagnosis Model of Breast Cancer Recurrence Location

Enter the index:

Immune Score:

Tumor Score:

Basic Nutrition Score:

Psychological Score:

Microenvironment Score:

Exercise and Advanced Work:

Forecast Result

Yes/No

Forecast Button

FIGURE 8. Interactive interface of breast cancer recurrence location prediction model.

TABLE 3. Comparison of prediction model indexes between the AMO-XGBoost algorithm and other algorithm models.

Algorithmic model	AMO-XGBoost	XGBoost	MLP	SVM	GBDT	Adaboost
Accuracy	93.07%	92.58%	82.91%	81.78%	82.17%	81.51%
F1	96.22%	95.60%	90.46%	89.94%	90.12%	89.81%
Precision	93.69%	92.54%	82.99%	81.73%	82.17%	81.52%
Recall	98.89%	98.87%	99.41%	100%	99.78%	100%

TABLE 4. AUC value results.

Algorithm	AMO-XGBoost	XGBoost	AdaBoost	GBDT	MLP
Auc Value	0.891	0.872	0.5	0.509	0.562

AUC: Area under the ROC Curve.

TABLE 5. 9-fold cross-validation results.

Algorithm	Accuracy								
AMO-XGBoost	0.9093	0.9364	0.9276	0.9416	0.9425	0.9364	0.9372	0.9303	0.9233
XGBoost	0.9015	0.9233	0.9254	0.9329	0.9359	0.9233	0.9251	0.9296	0.9266
AdaBoost	0.8003	0.8003	0.7995	0.7995	0.7995	0.7995	0.7995	0.7995	0.7995
GBDT	0.803	0.8073	0.8047	0.8099	0.8065	0.8047	0.8047	0.803	0.8082
MLP	0.8134	0.8309	0.8152	0.8291	0.8178	0.8204	0.8134	0.8082	0.8108

in diagnosing patients.

Fig. 8 shows the concept of the interactive interface of the proposed prediction model. First, the doctor fills in the relevant indicators of the patient, and then through the prediction and evaluation of the artificial intelligence algorithm, the interface would show the results of whether the recurrent cancer cells in the patient have metastasized, which can assist doctors in diagnosing and formulating a more individualized treatment plan with the aim to improve the patients' treatment outcomes and prolonging their survival.

5. Conclusions

The score of each index of the proposed model was obtained by evaluating each patient's physical health (medical evaluation), after which the index set affecting cancer recurrence was used to construct the model. The proposed prediction model demonstrated promising accuracy in predicting cancer recurrence location. The model combined with artificial intelligence algorithms could predict patients' cancer recurrence location with an 80% accuracy, which was based on the adaptive grid optimization XGBoost, which achieved an accuracy rate of more than 93%. In addition, the proposed model could be used in treatment planning by combining the prediction with corresponding clinical intervention measures, which could effectively improve the survival rate of patients.

The limitation of this model are few interactions with doctors and that its efficacy could be affected by patients and disease heterogeneity. We are also actively conducting clinical tests. Thus, its purpose is mainly to be used as a reference when making clinical decisions [22, 23]. At present, the use of artificial intelligence in medicine is still in its infancy. In the future, with more breakthroughs in science and technology,

we believe that artificial intelligence would have greater contributions to auxiliary diagnosis and treatment and precision medicine.

AUTHOR CONTRIBUTIONS

In this research, the first author, AMY—was responsible for organizing the paper and revising the overall framework. The second author, ZZM—was responsible for the idea of the paper, the overall paper writing, algorithm fusion research, and experimental analysis. The third author, JW—was responsible for the operation of the experimental part of the paper, the fourth author SJY, DBH—are responsible for data analysis and sorting.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Shanxi Provincial People's Hospital, Taiyuan, China. All patients provided written informed consent. The study was approved by The Fourth Affiliated Hospital of Hebei Medical University, Shijiazhuang, China. All patients provided written informed consent.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research supported by the Natural Science Foundation of Hebei Province (NO. E2021209024).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Kumar A, Ambade B, Sankar TK, Sethi SS, Kurwadkar S. Source identification and health risk assessment of atmospheric PM_{2.5}-bound polycyclic aromatic hydrocarbons in Jamshedpur, India. *Sustainable Cities and Society*. 2020; 52: 101801.
- [2] Ma P, Zhang Z, Wang J, Zhang W, Liu J, Lu Q, *et al*. Review on the application of metalearning in artificial intelligence. *Computational Intelligence and Neuroscience*. 2021; 2021: 1–12.
- [3] Yang A, Zhang W, Wang J, Yang K, Han Y, Zhang L. Review on the application of machine learning algorithms in the sequence data mining of DNA. *Frontiers in Bioengineering and Biotechnology*, 2020, 8: 1032.
- [4] Nabizadeh R, Sorooshian A, Delikhoon M, Baghani AN, Golbaz S, Aghaei M, *et al*. Characteristics and health effects of volatile organic compound emissions during paper and cardboard recycling. *Sustainable Cities and Society*. 2020; 56: 102005.
- [5] Qian Y, Kent EE. Gender differences in the association between unmet support service needs and mental health among American cancer caregivers. *Supportive Care in Cancer*. 2022; 30: 5469–5480.
- [6] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians*. 2018; 68: 394–424.
- [7] Zhang Y, Zhou T, Han S, Chang J, Jiang W, Wang Z, *et al*. Development and external validation of a nomogram for predicting the effect of tumor size on cancer-specific survival of resected gallbladder cancer: a population-based study. *International Journal of Clinical Oncology*. 2021; 26: 1120–1129.
- [8] Fu JB, Molinares DM, Morishita S, Silver JK, Dibaj SS, Guo Y, *et al*. Retrospective analysis of acute rehabilitation outcomes of cancer inpatients with leptomenigeal disease. *PM & R*. 2020; 12: 263–270.
- [9] Cenik F, Mähr B, Palma S, Keilani M, Nowotny T, Crevenna R. Role of physical medicine for cancer rehabilitation and return to work under the premise of the “Wiedereingliederungsteilzeitgesetz”. *Wiener Klinische Wochenschrift*. 2019; 131: 455–461.
- [10] Yu G, Chen Z, Wu J, Tan Y. A diagnostic prediction framework on auxiliary medical system for breast cancer in developing countries. *Knowledge-Based Systems*. 2021; 232: 107459.
- [11] Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. *Technology and Health Care*. 2016; 24: 31–42.
- [12] Asri H, Mousannif H, Moatassime HA, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*. 2016; 83: 1064–1069.
- [13] Agarap AFM. ‘On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset’. *ICMLSC '18: Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*. Association for Computing Machinery: New York, USA. 2018.
- [14] Rosario-Concepción RA, Calderín YB, Aponte CL, López-Acevedo CE, Sepúlveda-Irrizarry FL. Oncologists’ attitude and knowledge about cancer rehabilitation. *PM & R*. 2021; 13: 1357–1361.
- [15] Yang A, Han Y, Liu C, Wu J, Hua D. D-TSVR recurrence prediction driven by medical big data in cancer. *IEEE Transactions on Industrial Informatics*. 2021; 17: 3508–3517.
- [16] Paul K, Buschbacher R. Cancer rehabilitation. *American Journal of Physical Medicine & Rehabilitation*. 2011; 90: S1–S4.
- [17] Yancik R. Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older. *JAMA*. 2001; 285: 885.
- [18] Sheng X, Huo W, Zhang C, Zhang X, Han Y. A paper quality and comment consistency detection model based on feature dimensionality reduction. *Alexandria Engineering Journal*. 2022; 61: 10395–10405.
- [19] Chen T, Guestrin C. XGBoost. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM: New York. 2016.
- [20] Bennett JA, Botkin ME. Structural shape optimization with geometric description and adaptive mesh refinement. *AIAA Journal*. 1985; 23: 458–464.
- [21] Hoppe H, DeRose T, Duchamp T, McDonald J, Stuetzle W. ‘Mesh optimization’. *SIGGRAPH '93: Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. Association for Computing Machinery: New York, USA. 1993.
- [22] Shiraishi J, Li Q, Appelbaum D, Doi K. Computer-aided diagnosis and artificial intelligence in clinical imaging. *Seminars in Nuclear Medicine*. 2011; 41: 449–462.
- [23] Shao D, Dai Y, Li N, Cao X, Zhao W, Cheng L, *et al*. Artificial intelligence in clinical research of cancers. *Briefings in Bioinformatics*. 2022; 23: bbab523.

How to cite this article: Aimin Yang, Zezhong Ma, Jian Wang, Shujuan Yuan, Zunqian Zhang, Dianbo Hua. Early warning model of recurrence site for breast cancer rehabilitation patients based on adaptive mesh optimization XGBoost. *European Journal of Gynaecological Oncology*. 2022; 43(4): 40-48. doi: 10.22514/ejgo.2022.031.