



Adoption of combined detection technology of tumor markers via deep learning algorithm in diagnosis and prognosis of gallbladder carcinoma

Yigang Chang^{1,2} · Qian Wu³ · Limin Chi⁴ · Huaying Huo² · Qiang Li¹

Accepted: 23 April 2021 / Published online: 17 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

This study was to explore the application value of back propagation (BP) neural network (BPNN) and genetic algorithm (GA) in the combined detection and prognosis of tumor markers in patients with gallbladder cancer. 446 patients with gallbladder cancer were included in the experimental group, 279 patients with benign gallbladder disease were included in the control group, and 188 healthy people were selected and included in the blank group. Serum tumor markers (CA242, CA199, CEA, and CA125) of the three groups were detected by electrochemical luminescent immune analyzer, and follow-up data for 5 years after surgery were collected. Based on BPNN and GA, an optimization algorithm for multi-tumor markers was constructed and applied to the combined detection of tumor markers in patients. The artificial neural network (ANN), dynamic network biomarker (DNB), auxiliary diagnosis algorithm of the support vector machine (SVM) based on the particle swarm optimization (PSO) (PSO-SVM), matched-pairs feature selection (MPFS) based on the machine learning, and the BPNN were introduced to compare with the algorithm constructed. The diagnostic performances of the algorithms were evaluated with the fivefold cross-validation method. The results showed that the levels of CanAg (CA) 242, carcinoma embryonic antigen (CEA), CA199, and CA125 and positive rates in the experimental group were significantly higher than those in the control group and the blank group ($P < 0.05$); but the differences between control group and blank group were not visible ($P > 0.05$). The sensitivity (91.72%) and specificity (87.49%) in detecting CA242 and CA199 based on the proposed algorithm were the highest; the sensitivity (0.9186), specificity (0.8622), and accuracy (94.94%) of the proposed algorithm were higher than those of the conventional algorithms. The postoperative follow-up survival rate of patients in the experimental group was reduced from 41.72% in the first year to 4.28% in the fifth year; tumor node metastasis (TNM) stage IV, neck gallbladder cancer, and CA199 were significantly correlated with the survival rate of patients in the experimental group ($P < 0.05$). In summary, the combined detection technology of multiple tumor markers based on deep learning

algorithms showed excellent diagnostic and prognostic performance for gallbladder cancer. The occurrence of gallbladder cancer was related to the tumor markers CA242, CA199, CEA, and CA125, showing better detection effects by combination of CA242 and CA199. The TNM stage IV, neck gallbladder cancer, and CA199 were independent risk factors for the decrease in survival rate of patients with gallbladder cancer.

Keywords Gallbladder cancer · Back propagation neural network · Genetic algorithm · Multi-tumor markers · CA242

1 Introduction

Gallbladder cancer is a relatively rare but highly malignant tumor. When some cells of the gallbladder undergo the gene mutations, they lose their normal growth regulating ability and become malignantly proliferating cancer cells, forming gallbladder cancer [1–3]. Moreover, the symptoms of early gallbladder cancer are not obvious or atypical, and are mainly manifested as cholecystitis or gallstones, so the clinical diagnosis of gallbladder cancer is very difficult. It is often found in the middle and late stages, which seriously affects the treatment and prognosis effect [4, 5]. Tumor markers refer to substances that are characteristically present in malignant tumor cells, or abnormally produced by malignant tumor cells, or generated by the body due to stimulus response to tumors. It can reflect the occurrence and development of tumors [6]. At present, there are two main tumor markers for clinical diagnosis of gallbladder cancer, carcinoembryonic antigen (CEA) and carbohydrate antigen 199 (CA199). Unfortunately, these two indicators have no high specificity. Therefore, some auxiliary examinations have to be performed based on the symptoms of the patient for a comprehensive judgment clinically, and it cannot be determined solely by the increase in CEA and CA199 [7, 8].

BPNN is a multi-layer feed-forward neural network trained according to the error back propagation algorithm, and is also the most widely used deep learning neural network [9]. After decades of development and growth, neural networks have been widely used in industrial control, medical diagnosis, artificial intelligence, and other fields [10, 11]. Lasheras et al. [12] used the multivariate adaptive regression splines (MARS) and ANN technology to classify and model the genetic diagnosis of prostate cancer, and it was verified that the model showed higher accuracy and sensitivity in diagnosing the prostate cancer in contrast to the model based on the MARS. Shakeel et al. [13] used the discrete AdaBoost optimization integrated learning generalized neural network to analyze the selected lung features, successfully analyzed the biomedical data of lung, and effectively classified the normal and abnormal features of patients with lung cancer. GA is a method to search for the optimal solution by simulating the natural selection of Darwin's biological evolution theory and the biological evolution process of genetic mechanism, and it also has been extensively applied in medical diagnosis [14]. Chen et al. (2020) [15] applied GA to the classification of 801 gene expression samples of 5 kinds of cancers, and found that

the average diagnostic accuracy of GA was 98.81%, and the standard deviation was 0.0174. Huang et al. (2018) [16] discussed the application of SVM in cancer genome and tumor marker selection, indicating that SVM was a powerful classification tool that can be used in cancer diagnosis. Dochez et al. (2019) [17] compared the application of human epididymal protein 4 (HE4), CA125, malignancy risk index (RMI), and risk of ovarian malignancy algorithm (ROMA) in the diagnosis of ovarian cancer; IT was found that the sensitivity of serum CA125 was low in the early stage of cancer, and the level of HE4 was overexpressed in ovarian tumors. The specificity was 94%, which was not affected by endometriotic cysts. The area under the curve (AUC) of CA125 and HE4 was as high as 0.96, the AUC area of RMI was lower (0.86), and the specificity of ROMA was lower (84%). Therefore, the BPNN and GA were combined in this study to construct an optimization algorithm for multi-tumor marker to analyze the diagnosis and prognosis of patients with gallbladder cancer.

In summary, there are currently many studies to evaluate the malignant tumors using the deep learning and GA, but there is no report on the study of the joint application of both. Based on this, an optimization algorithm for multi-tumor marker was constructed based on BPNN and GA and applied to the combined detection of tumor markers. The series, eXGBoost based on the decision tree, PSO-SVM, MPFS based on the machine learning, and the BPNN were introduced to compare with the algorithm constructed, aiming to provide effective assistance for the early clinical diagnosis of gallbladder cancer.

2 Materials and methodologies

2.1 Research objects

In this study, 446 patients with gallbladder cancer who were admitted to the Shanxi Provincial People`s Hospital from January 30, 2013 to January 30, 2014 were included in the experimental group, 279 patients with benign gallbladder lesions (cholecystolithiasis, cholecystitis) were included in the control group, and 188 healthy people who underwent physical examination during the same period in the same hospital were selected in the blank group. The study complied with the Declaration of Helsinki and had been approved by the Medical Ethics Committee of Shanxi Provincial People`s Hospital. Patients and their families had understood the advantages and disadvantages of participating in this study, were willing to cooperate with various examinations required in the study protocol, and had sign the informed consents.

Inclusion criteria patients aged over 18 years, patients without medical treatment such as radiotherapy and chemotherapy before detection,; patients with complete clinical data, and patients diagnosed with gallbladder cancer by pathology.

Exclusion criteria patients with other malignant tumors, patients with post-operative death due to reasons other than gallbladder cancer, and patients with unclear consciousness and poor compliance.

2.2 Acquisition of related data

The serum tumor markers were detected as follows. An automatic electrochemical luminescent immune analyzer produced by Shanghai Hegao Biotechnology Co., Ltd. was used to detect the serum tumor markers in patients in three groups. At 8 o'clock in the morning, 6 mL of fasting cubital vein blood was collected from all patients and centrifuged at 5000 r/m to collect the supernatant. Then the supernatant was stored in a refrigerator at $-80\text{ }^{\circ}\text{C}$. The levels of serum CA242, CA199, CEA, and CA125 of the subjects were detected with a fully automatic electrochemical luminescent immune analyzer one day before surgery, 7 days after surgery, 60 days after surgery, 120 days after surgery, and 360 days after surgery. It was deemed as positive if $\text{CA242} > 15\text{ U/mL}$, $\text{CA199} > 39\text{ U/mL}$, $\text{CEA} > 10\text{ g/L}$, and $\text{CA125} > 35\text{ U/mL}$.

The basic data were acquired as follows. In this study, the basic data of all subjects (age, gender, occupation, weight, height, smoking history, drinking history, clinical TNM stage, biliary comorbidity, surgical method, family medical history, complete resection (RO) or not, pathogenic site, histological classification, and CT image) were acquired from database of Shanxi Provincial People's Hospital.

2.3 Optimization algorithm for multi-tumor marker

In this study, the BPNN and GA were used jointly to construct an optimization algorithm for multi-tumor marker of gallbladder cancer. Before construction, the model parameters were encoded using the BPNN. The transfer function among the hidden layers of the four-layer BPNN was set to the Sigmoid function, then the form of the transfer function can be expressed as Eq. (1):

$$f(x) = 2 \cdot \frac{1 - e^{-\beta x}}{\beta(1 + e^{-\beta x})} \quad (1)$$

In above Eq. (1), β represented the shape parameter of the transfer function between hidden layers, and $\beta > 0$. In addition, the connection weight values between the nodes among four layers under the results of neural network were set to $W1$, $W2$, and $W3$, respectively. The threshold of the first and the second node in the hidden layer was set to $B1$ and $B2$, respectively, and the threshold of the node in the output layer was set to $B3$. These parameters were encoded with real values, number of the real values were limited to m . At this time, the four-layer BPNN random coding string shown in Fig. 1 can be obtained. $n1$ represented the product of the number of nodes in the input layer and the first hidden layer; $n2$ represented the product of the number of nodes in the first and the second hidden layer; $n3$ represented the product of the number of nodes in the second hidden

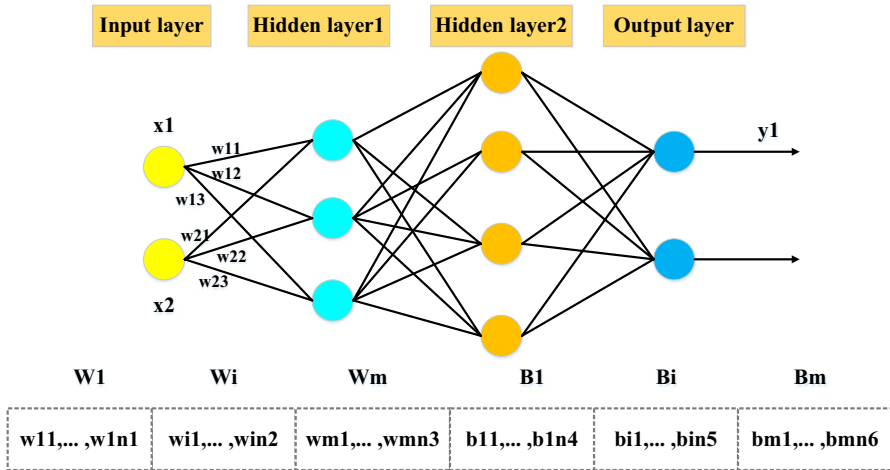


Fig. 1 The proposed four-layer BPNN and corresponding random coding strings

layer and the output layer; n_4 represented the number of nodes in the first hidden layer; n_5 indicated the number of nodes in the second hidden layer; and n_6 indicated the number of nodes in the output layer.

Then, the GA was introduced to create the initial population based on the coding string. It was assumed that each of the four tumor markers was set with 2 segmentation points, and the chromosome would be composed of 8 segmentation points. The segmentation points randomly appeared within the range of the tumor marker values, thus Eq. (2) could be obtained:

$$Chr = [4.14 \cdot 5.33 \cdot 41.82 \cdot 5.82 \cdot 61.62 \cdot 180.42 \cdot 6.31 \cdot 112.73] \tag{2}$$

$$FieldDR = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 400 & 500 & 450 & 200 & 250 & 350 & 400 & 500 \end{bmatrix} \tag{3}$$

In above Eqs. (2) and (3), Chr represented the segmentation point set, and $FieldDR$ represented the range of values. The information entropy [18] was introduced to calculate the information gain of each segmentation point in the segmentation point set Chr , the segmentation point with the highest gain was included in the candidate segmentation point set A_1 , and then the segmentation point was divided into a left and a right section. The calculation was repeated in the left and right section all the time, and the acquired segmentation point with the highest gain was included in the set A_1 . Due to the limitation of candidate segmentation points, the parameter $\frac{N_1}{\alpha}$ was introduced to control, where N_1 represented the number of segmentation points in the set Chr , α was a controllable integer, and $1 < \alpha \leq N_1$. The value of α was determined based on the actual number of segmentation points.

According to the processing of dividing points, the formed initial population was [300, 8] if the number of the population was set to 300. The adaptive value

proportional selection method was adopted to select the individual with higher adaptive value from the current population so as to produce the mating pool, then it could be concluded as:

$$G = \{x_1, x_2, \dots, x_m\} \tag{4}$$

In above Eq. (4), G represented the population, and m represented the size of the population. It was assumed that the crossover probability of the population is P_c , and the new chromosomes $Chr1^*$ and $Chr2^*$ could be acquired through the crossover operation of chromosome $Chr1 = [4.14 \cdot 5.33 \cdot 41.82 \cdot 5.82 \cdot 61.62 \cdot 180.42 \cdot 6.31 \cdot 112.73]$ and $Chr2 = [5.71 \cdot 7.32 \cdot 17.99 \cdot 35.17 \cdot 21.55 \cdot 51.28 \cdot 10.16 \cdot 42.88]$:

$$Chr1^* = [4.14 \cdot 5.33 \cdot 17.99 \cdot 35.17 \cdot 21.55 \cdot 51.28 \cdot 6.31 \cdot 112.73] \tag{5}$$

$$Chr2^* = [5.71 \cdot 7.32 \cdot 41.82 \cdot 5.82 \cdot 61.62 \cdot 180.42 \cdot 10.16 \cdot 42.88] \tag{6}$$

Considering that the randomness of selection and crossover operation may destroy the individual with the highest adaptive value in the current population, the optimal segmentation preservation strategy was introduced in this study [19]. Firstly, it had to find the individual with the highest adaptive value in the population; secondly, it had to find the individual with the lowest adaptive value in the current population; thirdly, the individual with the lowest adaptive value was replaced by the individual with the highest adaptive value. The calculation had to be stopped when the genetic algebra reached the maximum evolutionary algebra. The calculation steps of this algorithm were summarized as Fig. 2.

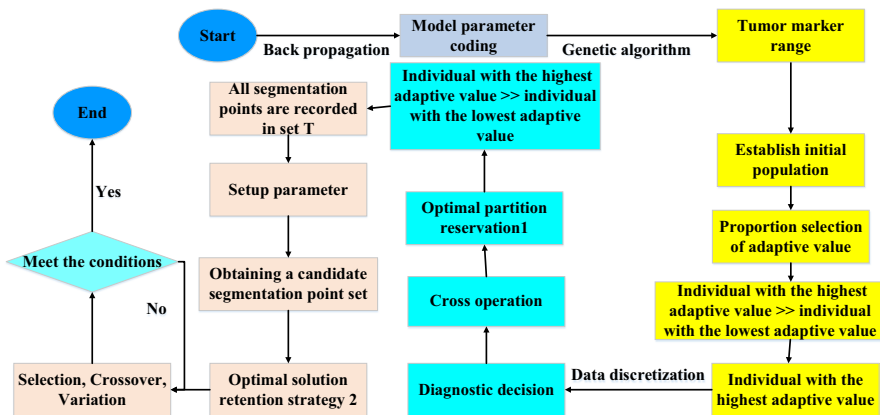


Fig. 2 The constructed optimization algorithm for multi-tumor marker

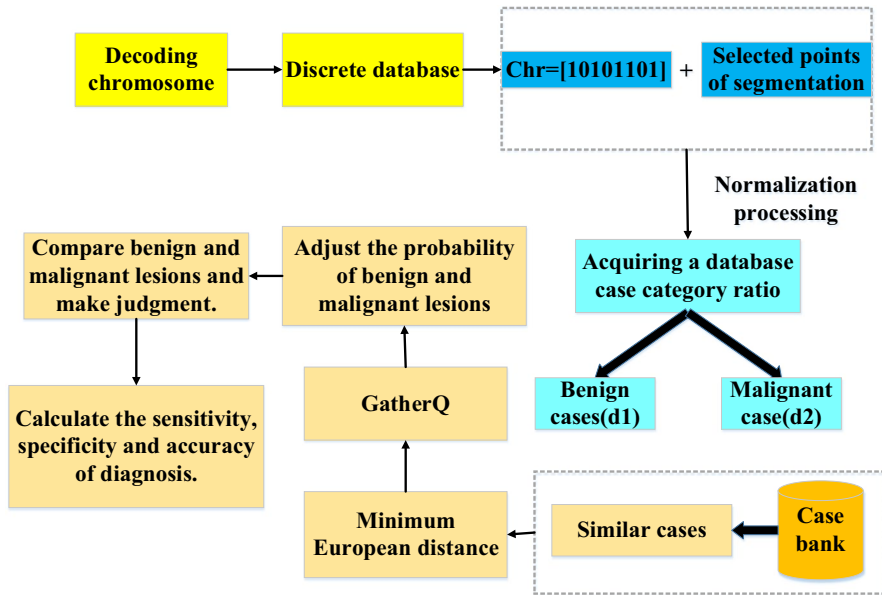


Fig. 3 Schematic diagram for case-based reasoning diagnosis of the optimization algorithm for multi-tumor marker

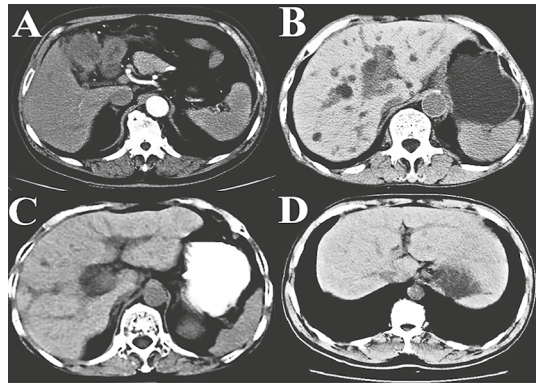
2.4 Case-based reasoning diagnosis of the optimization algorithm for multi-tumor marker

The case-based reasoning method was introduced in this study to apply the algorithm to actual diagnosis [20]. The case-based reasoning could judge the current real-time data, and then select the case that was most similar to the current data from the case database as the diagnosis basis of this data. As shown in Fig. 3, the case-based reasoning was used in this study to diagnose the patients, and the calculated corresponding sensitivity and specificity of chromosome were given. The specific steps were as follows. Firstly, each chromosome in GA could correspondingly indicate the combination of several tumor markers and the selected segmentation site. Therefore, decoding these chromosomes can obtain a discretized tumor marker database. Secondly, the indicators with discretization characteristics were normalized. Thirdly, the number of categories in the tumor marker database was counted and recorded, including the number of benign and malignant cases. Fourthly, the remaining cases were compared with the cases in the database, and Euclidean distance of each case was calculated. The cases with the smallest Euclidean distances could be recorded as a set. Fifthly, there are often more benign lesions than malignant lesions in actual, so there was a big difference in the number of categories in the tumor marker database, which affected the unfairness of the comparison of cases. Thus, the probability of benign and malignant cases should be further adjusted. Sixthly, the diagnosis sensitivity, specificity, and accuracy of algorithm could be calculated based on the final results.

Table 1 Basic information of subjects in three groups

Indicator	Experimental group	Control group	Blank group	X ²	P value
Age (years old)	63.18 ± 10.86	62.22 ± 9.47	63.05 ± 9.64	2.478	0.067
Weight (kg)	61.94 ± 7.88	62.55 ± 5.93	61.57 ± 7.69	1.529	0.182
Male ratio (%)	47.15	45.82	46.33	1.921	0.083
Female ratio (%)	52.85	54.18	53.67	3.557	0.218
Average disease course (months)	8.77 ± 2.71	9.04 ± 2.66	8.33 ± 2.53	2.052	0.256
Height (cm)	161.67 ± 13.58	162.33 ± 15.66	163.03 ± 15.82	1.639	0.077

Fig. 4 CT images of abdomens of patients in three groups. Note: **a** and **b** gave CT images of patients in experimental group, **c** was the CT image of patient in control group, and **d** was the CT image of healthy subject in blank group



2.5 Statistical analysis

The data in this study were analyzed and processed by SPSS19.0 version statistical software, and the measurement data were expressed in the form of mean ± standard deviation ($\bar{x} \pm s$). The counting data were given in the form of percentage (%). The basic data (age, height, weight, male proportion, female proportion, and average course of disease) of the patients in experimental group, control group, and blank group were compared by paired t test. The clinical TNM staging, occupation, smoking history, drinking history, biliary comorbidities, surgical methods, family history, RO resection, pathogen location, and histological classification of the patients in experimental group were compared by one-way analysis of variance. The Cox proportional hazards model was adopted to analyze the relevant risk factors affecting the 5-year survival rate of patients with gallbladder cancer. Fivefold cross-validation was applied to analyze the diagnostic sensitivity, specificity, and accuracy of the proposed algorithm and ANN, DNB, PSO-SVM, MPFS, and BPNN. The difference was statistically significant at $P < 0.05$. Origin 8.0 was adopted to draw the figures.

3 Results and discussion

3.1 Comparison on basic information of subjects in three groups

As shown in Table 1, there was no significant difference in age, height, weight, male ratio, female ratio, and average disease course in the three groups ($P > 0.05$), so they were comparable.

Figure 4a showed that the intrahepatic bile duct was dilated, and there was a round low-density shadow in the posterior lower lobe of the liver without enhancement, the gallbladder was enlarged, and nodular wall was thickened locally. After enhancement, it was slightly strengthened, the middle and upper bile ducts were dilated, and there was a local soft tissue density shadow. Thus, it was judged as the gallbladder cancer in lower part of the common bile duct. Figure 4b showed that the obstruction plane was determined as the pancreatic head plane, and the bile duct dilated in the lower part of the common bile duct was interrupted. The surrounding soft tissue was thickened, the enhancement was lack of blood vessel enhancement, the gallbladder was severely dilated, and the wall of the cyst wall showed obvious enhanced wall nodules. Thus, it was determined as the advanced gallbladder cancer. The expansion of the gallbladder, thickening of the wall, turbidity and stones in the cavity of the gallbladder were visible in Fig. 4c. There was no solid mass in the enhanced scanning cavity without metastatic lesions in the liver, so it was judged to be gangrenous cholecystitis and malignant changes (low-grade malignancy) of gallbladder myoma. Figure 4d was a CT scanning image of the abdomen of a normal person in the blank group. It showed that the edges were smooth, and the shape of each layer was different. The liver blood vessels were in the form of strips, branches, or dots with low density, and the closer to the inferior vena cava area, the thicker. In addition, the intrahepatic bile duct was invisible.

3.2 Descriptive statistics of pathological characteristics of patients in the experimental group

Table 2 disclosed that the male ratio and female ratio in the experimental group were similar; the clinical TNM stage III accounted for the largest proportion (45.02%); the proportion of citizens in occupations (58.81%) was higher than that of farmers (41.19%); the proportion of patients with smoking history was similar; the proportion of patients without drinking history (60.28%) was significantly higher than that of patients without drinking history (39.72%); the proportion of patients with biliary complication (65.28%) was significantly higher than that of patients without biliary comorbidity; the proportion of patients undergoing radical surgery (82.11%) was higher than that of palliative surgery (17.89%); the proportion of patients with surgery duration longer than or equal to 3 h (38.55%) was significantly less than that with surgery duration shorter than 3 h (61.45%); the proportion of patients without no family medical history (79.44%) was significantly higher than that with a family medical history (20.56%); the proportion of patients with RO resection (43.62%) was less than the patients with non-RO resection (56.38%); the patients

Table 2 Descriptive statistics of pathological characteristics of patients in the experimental group

Variable	Classification	Quantity of samples	Proportion
Gender	Male	210	47.15%
	Female	236	52.85%
Clinical TNM stage	II	117	25.07%
	III	163	45.02%
	IV	186	29.91%
Occupation	Farmer	192	41.19%
	Citizen	274	58.81%
Smoking history	Yes	246	52.73%
	No	220	47.27%
Drinking history	Yes	185	39.72%
	No	281	60.28%
Biliary complication	Yes	304	65.28%
	No	162	34.72%
Surgery method	Palliative	83	17.89%
	Radical	383	82.11%
Surgery duration	≥ 3 h	180	38.55%
	< 3 h	286	61.45%
Family medical history	Yes	96	20.56%
	No	370	79.44%
RO resection	Yes	203	43.62%
	No	263	56.38%
Pathogenic site	Body	75	16.06%
	Neck	241	51.76%
	Gallbladder bottom	150	32.18%
Histological classification	High differentiation	164	35.17%
	Moderate differentiation	202	43.28%
	Low differentiation	100	21.55%

with pathogenic site in the neck accounted for the largest proportion (51.76%), followed pathogenic site in the gallbladder bottom (32.18%); the proportion of patients with moderate differentiation was the highest (43.28%), followed by patients with high differentiation (35.17%). Above analysis suggested that clinical TNM stage, drinking history, biliary comorbidities, surgical methods, family medical history, RO, pathogenic site, and histological classification may be related to postoperative recovery and recurrence.

3.3 Comparison on levels of serum tumor markers of subjects in three groups

As shown in Fig. 5, it gave the levels of serum tumor markers of patients in the three groups one day before surgery (T1), 7 days after surgery (T2), 60 days after

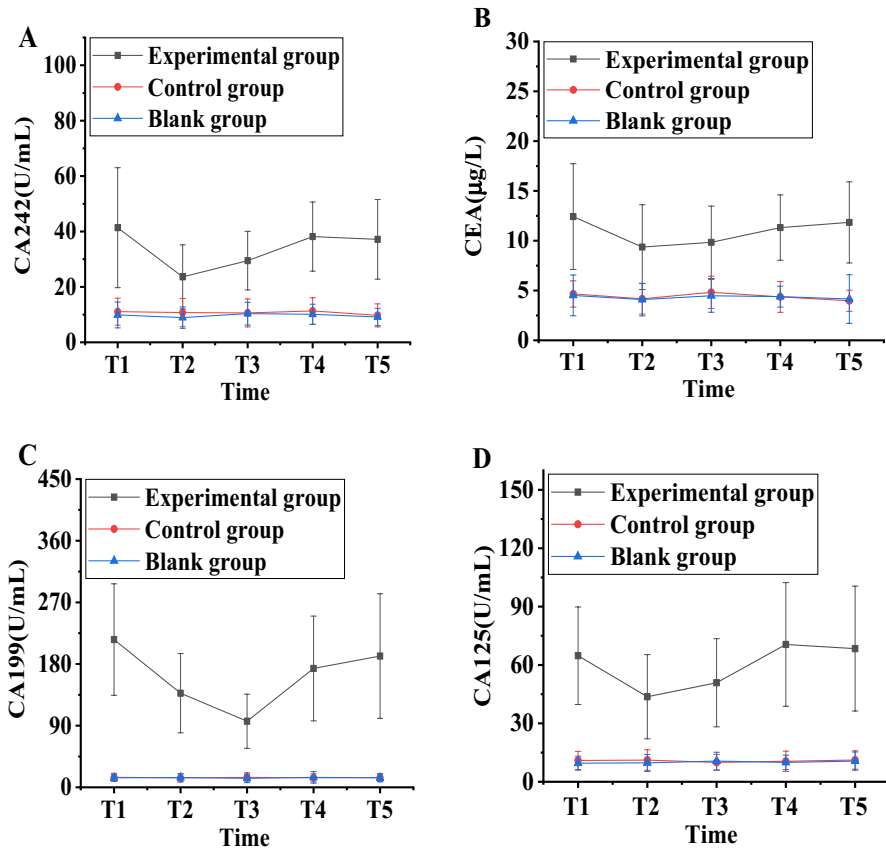


Fig. 5 Comparison levels of on serum tumor markers of subjects in three groups. Note: **a** referred to the level of CA242, **b** referred to the level of CEA, **c** referred to the level of CA199, and **d** referred to the level of CA125. * indicated that the difference was statistically significant compared with the experimental group ($P < 0.05$)

surgery (T3), 120 days after surgery (T4), and 360 days after surgery (T5). The levels of CA242, CEA, CA199, and CA125 in the experimental group decreased at the beginning and increased over time. The reason may be that most patients were in the middle and late stages of gallbladder cancer, and relapsed after radical resection [21]. The levels of CA242, CEA, CA199, and CA125 in the experimental group were always significantly higher than those in the other two groups ($P < 0.05$); the levels of CA242, CEA, CA199, and CA125 in the control group were always slightly higher than those in the blank group, but the difference was not statistically significant ($P > 0.05$). Such results were similar to the findings of Wu et al. (2018) [22], indicating that the occurrence and development of gallbladder cancer might be related to the CA242, CEA, CA199, and CA125. The higher the level of the tumor marker is, the higher possibility of gallbladder cancer. Therefore, the four tumor markers will be further explored in future.

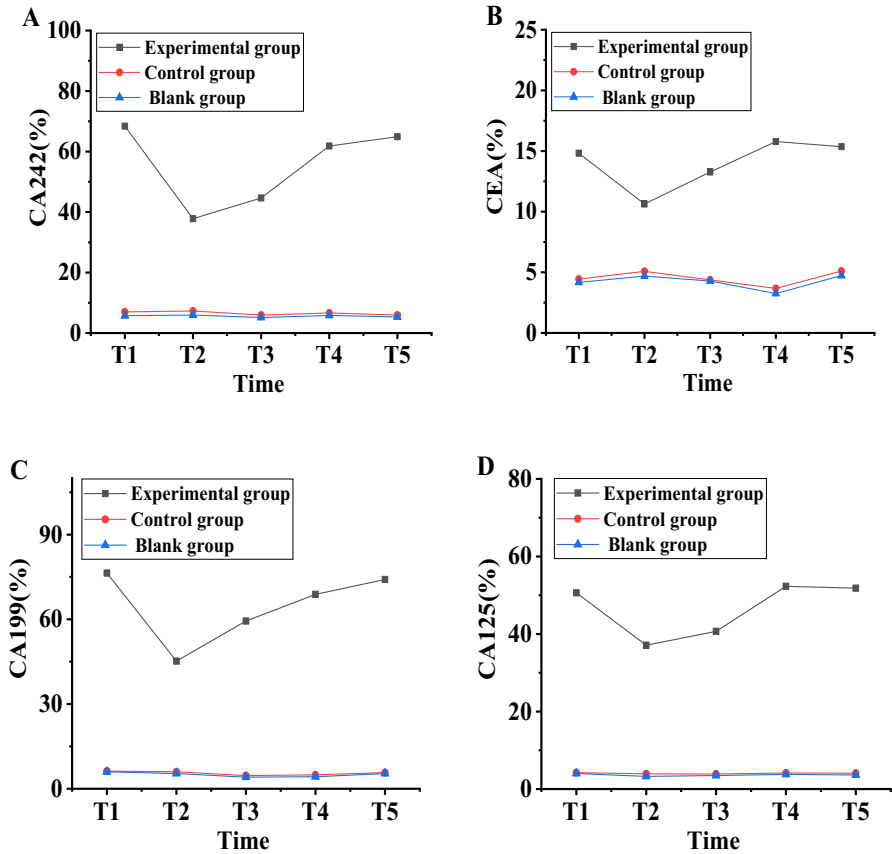


Fig. 6 Comparison on positive expression rates of serum tumor markers for subjects in three groups. Note: **a** referred to the level of CA242, **b** referred to the level of CEA, **c** referred to the level of CA199, and **d** referred to the level of CA125. * indicated that the difference was statistically significant compared with the experimental group ($P < 0.05$)

3.4 Comparison on positive expression rates of serum tumor markers for subjects in three groups

Figure 6 illustrated the positive expression rates of serum tumor markers of patients in the three groups at T1, T2, T3, T4, and T5. The positive expression rates of CA242, CEA, CA199, and CA125 in the experimental group decreased at the beginning and increased over time. The reason was the same as above that most patients might relapse after treatment [23]. The positive expression rates of CA242, CEA, CA199, CA125 in the experimental group were always higher than those in the control group and the blank group ($P < 0.05$); and the positive expression rates of CA242, CEA, CA199, and CA125 in the control group and the blank group had no significant difference ($P > 0.05$). Results here were similar to the findings of Pérez-Palma et al. (2020) [24],

suggesting that the increase in the positive expression rate of CA242, CEA, CA199, and CA125 was related to the development of gallbladder cancer.

3.5 Diagnosis results of patients with gallbladder cancer by the proposed algorithm

Before diagnosis, the candidate segmentation point set of each tumor marker has to be determined, so the control parameter α of the candidate segmentation point set was set to 2, 3, 4, 5, 6, 7, 8, 9, and 10, respectively, then the two quantitative indicators (candidate segmentation point value and standard deviation) of the parameter α were shown in Fig. 7. When the parameter α was 6, the candidate segmentation point values of the four tumor markers CA242, CEA, CA199, and CA125 were 11, 10, 10, and 11, respectively, and the standard deviation was 0.811. In this case, it could meet the requirement that the values of the candidate segmentation points were greater than or equal to 10 and the standard deviation was minimal. α value of other parameters failed to satisfy the requirements. Too many or too few segmentation points could affect the performance of the algorithm. It was easily drag down the effectiveness and simplicity of the algorithm when the segmentation points were too many, and the search range of the algorithm was affected with too less segmentation points. Based on previous experience, the parameter α value was set to 6, the population size, crossover rate, and mutation rate were determined as 280, 85%, and 15%, respectively.

The diagnosis results of the patients with gallbladder cancer using the proposed algorithm were given in Table 3. It revealed that CA242+CA199 showed the highest sensitivity (0.938), specificity (0.886), and accuracy (94.18%), which was the same as the research results of Balakrishnan et al. (2020) [25]. It indicated that CA242 combined with CA199 was the best combination for the diagnosis of gallbladder cancer. Both sensitivity and specificity were high, so the clinical promotion value was high. The sensitivity, specificity, and accuracy were 0.895, 0.851, and 90.54%, respectively, for CA242+CEA+CA199. Compared with the ROC curve analysis results, the sensitivity, specificity, and accuracy of the proposed algorithm on the CA242+CA199 combined detection were significantly higher than those of the traditional medical diagnosis methods. It meant that the algorithm proposed in this study could significantly improve the final diagnostic accuracy [26].

Fig. 7 Candidate segmentation point values and standard deviations of tumor markers with different α values

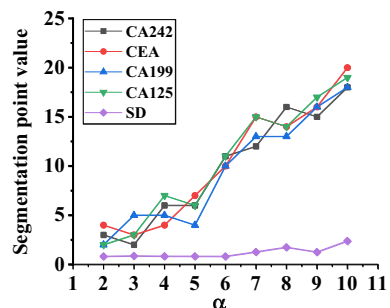


Table 3 Diagnosis results of patients with gallbladder cancer by the proposed algorithm

Indicator	Sensitivity	Specificity	Accuracy
CA242+CEA	0.832	0.769	83.26%
CA242+CA199	0.938	0.886	94.18%
CA242+CA125	0.805	0.789	84.22%
CEA+CA199	0.847	0.802	84.17%
CA199+CA125	0.839	0.819	82.68%
CA242+CEA+CA199	0.895	0.851	90.54
CA242+CEA+CA125	0.796	0.817	83.48%
CEA+CA199+CA125	0.825	0.743	81.79%
CA242+CA199+CA125	0.759	0.768	78.92%
CA242+CEA+CA199+CA125	0.817	0.832	83.69%

Above table showed that CA242+CA199 was the optimal combination, which was verified with the case-based reasoning diagnosis. As shown in Table 4, the discrete values of CA242 and CA199 obtained for each patient sample were different, and the malignant probability of diagnosis results was above 50%.

The first patient sample was undertaken as an example. The dispersion value of CA242 and CA199 was 4 and 9, respectively. After the other two tumor markers (CEA and CA125) were added separately, the results shown in Fig. 8 could be obtained. It indicated that when the dispersion value was 4, the corresponding malignancy probability of CEA and CA125 was the highest. This result was consistent with the medical common sense that the malignant probability of cases increased with the increase in tumor markers, which further verified that CA242+CA199 was the optimal combination.

3.6 Comparison on effects of proposed algorithm and traditional algorithm

In this study, the diagnosis effect of the proposed algorithm was compared with the ANN [27], DNB [28], PSO-SVM [29], MPFS based on the machine learning [30], and the BPNN [31]. The fivefold cross-validation analysis was performed for each algorithm, and the results were shown in Fig. 9. The diagnostic sensitivity and accuracy of the proposed algorithm were always higher than those of SP, MPFS, XGBoost, PSO-SVM, and BPNN algorithm. The specificity was similar to that of SP algorithm and was always higher than that of the MPFS, XGBoost, PSO-SVM, and BPNN algorithm. For further comparison, the five-step verification results were averaged, as shown in Fig. 10. The diagnostic sensitivity of the proposed algorithm (0.9186) was significantly higher than that of the ANN, MPFS, DNB, PSO-SVM, and BPNN ($P < 0.05$), which was similar to the research results of Wang et al. [32]. Thus, the tumor marker optimization algorithm based on BPNN and GA showed high sensitivity in the diagnosis of gallbladder cancer, and was superior to other artificial intelligence diagnosis algorithms based on machine learning or deep learning. The diagnostic specificity of the proposed algorithm (0.8622) was not significantly

Table 4 Case-based reasoning diagnosis of patients with gallbladder cancer based on the proposed algorithm

Sample	CA242		CA199		Diagnosis results		Proportion (%)
	Original value	Discrete value	Original value	Discrete value	Detected result	Actual result	
	1	33.18	4	153.96	9	Malignant	
2	29.64	2	187.84	7	Malignant	Malignant	72
3	42.12	4	139.66	3	Benign	Malignant	69
4	29.67	3	180.56	7	Malignant	Malignant	77
5	38.82	3	152.39	2	Benign	Malignant	59
6	51.34	4	171.56	4	Malignant	Malignant	64
...
446	32.68	3	153.77	6	Malignant	Malignant	75

Fig. 8 The malignant probability corresponding to different discrete values after the addition of CEA and CA125

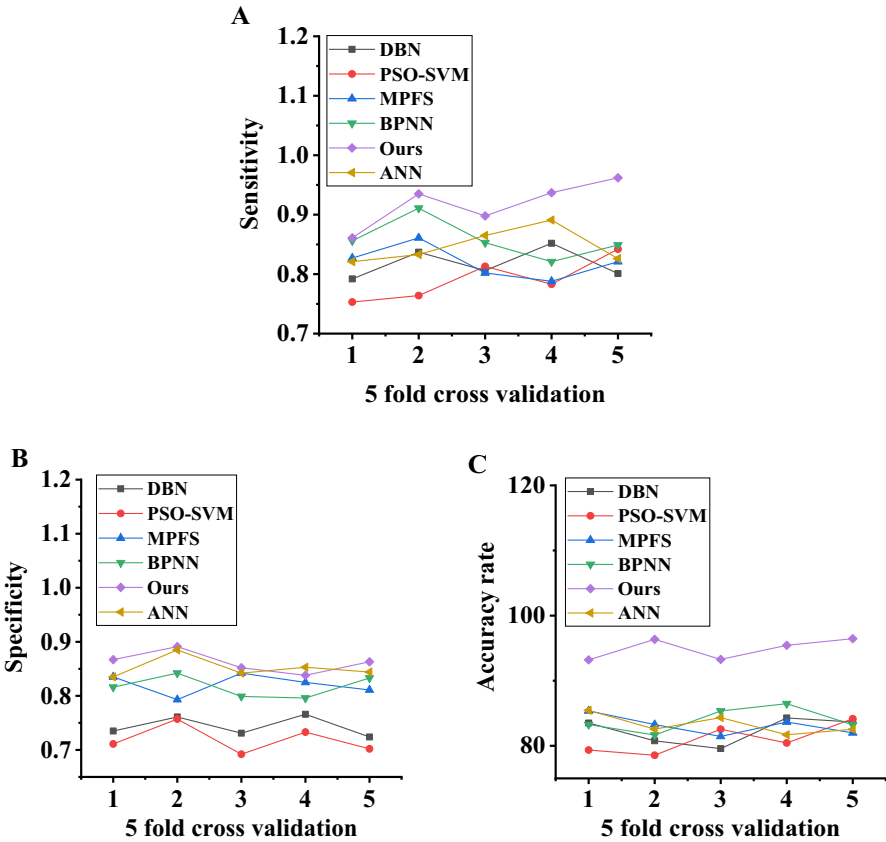
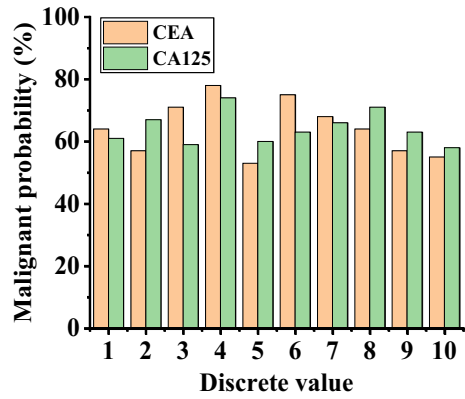


Fig. 9 Effects of proposed algorithm and traditional algorithm. Note: **a** referred to the diagnostic sensitivity, **b** indicated the diagnostic specificity, and **c** represented the diagnostic accuracy

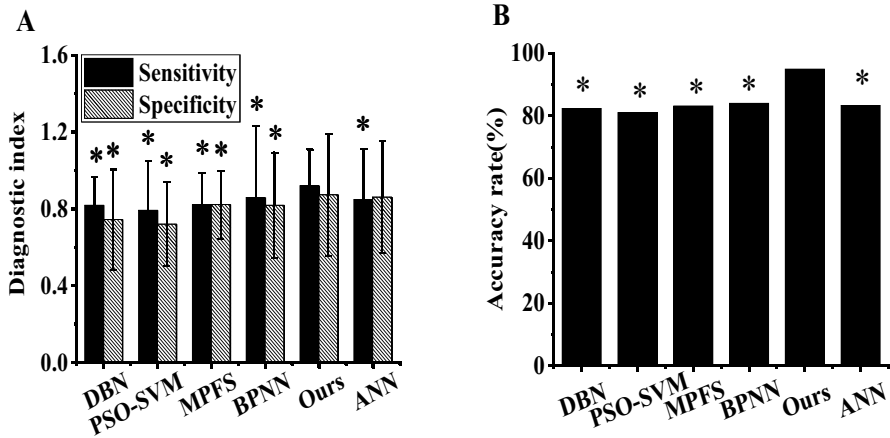
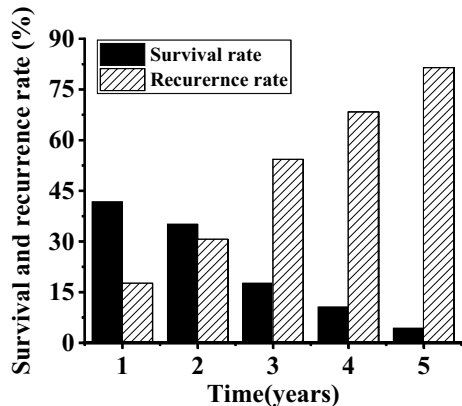


Fig. 10 Comparison on average values for diagnosis effects between the proposed algorithm and the traditional algorithms. Note: **a** referred to the average values of diagnostic specificity and sensitivity, **b** indicated the average value of diagnostic accuracy. *Indicated that the difference was statistically significant compared with the proposed algorithm ($P < 0.05$)

different with that of ANN ($P > 0.05$), but it was significantly higher than that of MPFS, DNB, PSO-SVM, and BPNN algorithms ($P < 0.05$). It suggested that diagnostic specificity of the proposed algorithm was better than that of the traditional algorithms. In addition, the diagnostic accuracy of the proposed algorithm (94.944%) was significantly higher than that of other algorithms, indicating that the proposed algorithm could maintain high sensitivity, high specificity, and high accuracy. It revealed that the optimization algorithm for multi-tumor marker constructed in this study showed high sensitivity, high specificity, and high accuracy.

Fig. 11 Survival rate and recurrence rate in 5 years after surgery of patients in experimental group



3.7 Postoperative follow-up results and related risk factors

As shown in Fig. 11, the survival rate of patients in 5 years after surgery in the experimental group decreased from 41.72% in the first year to 4.28% in the fifth year, and the recurrence rate increased from 17.58% in the first year to 81.44% in the fifth year. Thus, the survival rate and the recurrence rate were in the opposite trend, which was different from the follow-up results of Liska et al. [33]. The reason may be that the surgical treatment of patients with early gallbladder cancer had the best effect, but most of the patients with gallbladder cancer were in the middle and late stage. The palliative and radical surgery only could remove the tumor visible to the naked eye and were impossible to eliminate the remaining cancer cells in the body completely, so the survival period of the patient was prolonged transitorily.

The Cox proportional hazard model (PHM) was used to further explore the relevant risk factors affecting the survival rate of patients with gallbladder cancer in 5 years after surgery. The independent variables included gender (male was set to 1, and female was set to 0), clinical TNM staging (II was set to 2, III was set to 1, and IV was set to 0), occupation (citizen was set to 0, and farmer was set to 1), smoking history (yes was set to 1, and no was set to 0), drinking history (yes was set to 1, and 0 was for no), biliary complications (yes was set to 1, and 0 was for no), surgery method (palliative was set to 0, and radical was set to 1), operation duration (≥ 3 h was set to 0, and < 3 h was set to 1), family medical history (yes was set to 1, and 0 was for no), RO (yes was set to 1, and 0 was for no), pathogenic site (neck was set to 0, the gallbladder bottom was set to 1, and body was set to 2), histological

Table 5 Related risk factors of survival rate in 5 years after surgery for patient with gallbladder cancer

Factor	X ²	P value	Parameter	Risk ratio	Standard derivation
Gender	0.276	0.168	-0.052	0.866	0.481
Clinical TNM staging	8.582	0.016	-0.072	1.144	0.927
Occupation	0.338	0.421	0.019	0.937	0.528
Smoking history	0.258	0.093	-0.077	0.869	0.818
Drinking history	1.555	0.268	0.044	0.961	0.482
Biliary complications	1.639	0.055	0.051	0.892	0.461
Surgery method	0.330	0.063	0.038	0.962	0.522
Surgery duration	1.155	0.082	-0.057	0.792	0.441
Family medical history	1.428	0.177	0.062	0.467	0.274
RO resection	0.557	0.067	-0.039	0.897	0.511
Pathogenic site	10.429	0.005	-0.052	0.961	0.391
Histological classification	0.316	0.050	-0.061	1.052	0.717
CA242	0.831	0.073	0.488	1.153	0.841
CEA	1.289	0.052	0.362	1.042	0.671
CA199	15.482	0.002	0.017	2.370	0.427
CA125	0.738	0.057	0.022	0.945	0.388

classification (high differentiation was set to 0, mediate differentiation was set to 1, and low differentiation was set to 2), and levels of four tumor markers (CA242, CEA, CA199, and CA125). The survival conditions of patients in 5 years after surgery were taken as the dependent variables.

The results were shown in Table 5. It revealed that there was a significant negative correlation between clinical TNM staging, pathogenic site, and survival rate in 5 years after surgery ($P < 0.05$). There was a significant positive correlation between level of CA199 and survival rate in 5 years after surgery ($P < 0.05$), which suggested that later clinical TNM staging, pathogen parts in the neck, and higher level of serum CA199 could cause the lower survival rate of patients in 5 years after surgery. Thus, TNM stage IV, gallbladder cancer in neck, and high level of serum CA199 were independent risk factors affecting the survival rate of patients in 5 years after surgery.

4 Conclusion

In this study, an optimization algorithm for multi-tumor marker was constructed based on the BPNN and GA, and was applied in the diagnosis of tumor markers in patients with gallbladder cancer. The results showed that the combined detection technology of multiple tumor markers based on deep learning algorithms showed excellent diagnostic and prognostic performance for gallbladder cancer. The occurrence of gallbladder cancer was related to the tumor markers CA242, CA199, CEA, and CA125, showing better detection effects by combination of CA242 and CA199. The TNM stage IV, neck gallbladder cancer, and CA199 were independent risk factors for the decrease in survival rate of patients with gallbladder cancer. However, the optimization algorithm for multi-tumor marker constructed in this study had only been used in the diagnosis of gallbladder cancer, and other malignancies had not been popularized yet. Therefore, it was considered to extend the algorithm to the evaluation of patients with more types of malignant tumors in the future, so as to provide assistance for the early diagnosis of tumors. In conclusion, the results of this study provided an effective experimental basis for the clinical diagnosis and prognosis of malignant tumors.

References

1. Erlangga Z, Wolff K, Poth T, Peltzer A, Nahnsen S, Spielberg S, Timrott K, Woller N, Kühnel F, Manns MP, Saborowski A, Vogel A, Saborowski M (2019) Potent antitumor activity of liposomal irinotecan in an organoid-and CRISPR-Cas9-based murine model of gallbladder cancer. *Cancers* 11(12):1904
2. Liang Y, Li Q, Chen P, Xu L, Li J (2019) Comparative study of back propagation artificial neural networks and logistic regression model in predicting poor prognosis after acute ischemic stroke. *Open Medicine* 14(1):324–330
3. Mochizuki K, Hata H, Naitou K, Motosugi U, Kondo T (2020) Carcinosarcoma (adenocarcinoma, neuroendocrine carcinoma, undifferentiated carcinoma and chondrosarcoma) of the gallbladder. *Clin J Gastroenterol* 13(1):110–115

4. Geng ZM, Cai ZQ, Zhang Z, Tang ZH, Xue F, Chen C, Zhang D, Li Q, Zhang R, Li WZ, Wang L, Si SB (2019) Estimating survival benefit of adjuvant therapy based on a Bayesian network prediction model in curatively resected advanced gallbladder adenocarcinoma. *World J Gastroenterol* 25(37):5655–5666
5. Diana M, Schiraldi L, Liu YY, Memeo R, Mutter D, Pessaux P, Marescaux J (2016) High intensity focused ultrasound (HIFU) applied to hepato-bilio-pancreatic and the digestive system-current state of the art and future perspectives”. *Hepatobiliary Surg Nutr* 5(4):329–344
6. Zhang W, Tang G, Zhou S, Niu Y (2019) LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination. *BMC Genomics* 20(11):1–12
7. Liu J, Li X, Huang J, Liu Y (2019) Matrix metalloproteinase 2 knockdown suppresses the proliferation of HepG2 and Huh7 cells and enhances the cisplatin effect. *Open Medicine* 14(1):384–391
8. Elbadawy M, Abugomaa A, Yamawaki H, Usui T, Sasaki K (2020) Development of prostate cancer organoid culture models in basic medicine and translational research. *Cancers* 12(4):777
9. Abugomaa A, Elbadawy M, Yamawaki H, Usui T, Sasaki K (2020) Emerging roles of cancer stem cells in bladder cancer progression, tumorigenesis, and resistance to chemotherapy: a potential therapeutic target for bladder cancer. *Cells* 9(1):235
10. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA, Erickson BJ (2019) A Survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow. *J Am Coll Radiol* 16(9):1318–1328
11. Chuang CY, Chang TT, Li DK, Chiu CM, Fang CW, Ng HP, Huang CM, Chen JJ (2020) Colectomy influences the radial pulse parameters of traditional Chinese medicine pulse diagnosis in patients with colorectal cancer. *Eur J Integr Med* 2020(35):101067
12. Lasheras JES, Donquiles CG, Nieto PJG, Moleon JJJ, Salas D, Gómez SLS, de la Torre AJM, González-Nuevo J, Bonavera L, Landeira JC, de Cos Juez FJ (2020) A methodology for detecting relevant single nucleotide polymorphism in prostate cancer with multivariate adaptive regression splines and backpropagation artificial neural networks. *Neural Comput Appl* 32(5):1231–1238
13. Shakeel PM, Tolba A, Al-Makhadmeh Z, Jaber MM (2019) Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks. *Neural Comput Appl* 32(3):777–790
14. García-Díaz P, Sánchez-Berriel I, Martínez-Rojas JA, Díez-Pascual AM (2020) Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-Seq data. *Genomics* 112(2):1916–1925
15. Chen ZY, Fu S, Li M, Li MH, Zhang W, Ou HB (2020) Exploring artificial neural network combined with laser-induced auto-fluorescence technology for noninvasive in vivo upper gastrointestinal tract cancer early diagnosis. *IJS Oncology* 5(1):83
16. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W (2018) Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 15(1):41–51
17. Dochez V, Caillon H, Vaucel E, Dimet J, Winer N, Ducarme G (2019) Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *J Ovarian Res* 12(1):28
18. van Erp LW, Cunningham M, Narasimman M, Ali HA, Jhaveri K, Drenth JPH, Janssen HLA, Levy C, Hirschfield GM, Hansen BE, Gulamhusein AF (2020) Risk of gallbladder cancer in patients with primary sclerosing cholangitis and radiographically detected gallbladder polyps. *Liver Int* 40(2):382–392
19. Kumar M, Alshehri M, AlGhamdi R, Purushottam S, Vikas D (2020) A DE-ANN inspired skin cancer detection approach using fuzzy C-means clustering. *Mobile Netw Appl* 25:1319–1329
20. Zhang ZG, Chen Y, Ji R, Zhao YJ, Wang J, Robinson L, Chen XP, Zhang L (2018) Synchronous cancers of gallbladder carcinoma and combined hepatocellular cholangiocarcinoma: an unusual case and literature review. *BMC Cancer* 18(1):1046
21. Kabakov A, Yakimova A, Matchuk O (2020) Molecular chaperones in cancer stem cells: determinants of stemness and potential targets for antitumor therapy. *Cells* 9(4):892
22. Wu CH, Luo Y, Fei X, Chou YH, Chiou HJ, Wang HK, Lai YC, Lin YH, Tiu CM, Wang J (2018 Apr) Algorithmic approaches to the diagnosis of gallbladder intraluminal lesions on ultrasonography. *J Chin Med Assoc* 81(4):297–304
23. Zhang W, Shi J, Li R, Han Z, Li L, Li GH, Yang B, Yin Q, Wang YY, Ke Y, Li Q (2020) Effectiveness of olaparib treatment in a patient with gallbladder cancer with an ATM-inactivating mutation. *Oncologist* 25(5):375

24. Pérez-Palma E, Bustos BI, Lal D, Buch S, Azocar L, Toliat MR, Lieb W, Franke A, Hinz S, Burmeister G, von Schönfels W, Schafmayer C, Ahnert P, Völzke H, Völker U, Homuth G, Lerch MM, Puschel K, Gutiérrez RA, Hampe J, Nürnberg P, Miquel JF, De Ferrari GV (2020) Copy number variants in lipid metabolism genes are associated with gallstones disease in men. *Eur J Hum Genet* 28(2):264–273
25. Balakrishnan S, Ibrahim KS, Duraisamy S, Sivaji I, Kandasamy S, Kumarasamy A, Kumar NS (2020) Antiquorum sensing and antibiofilm potential of biosynthesized silver nanoparticles of *Myristica fragrans* seed extract against MDR *Salmonella enterica* serovar Typhi isolates from asymptomatic typhoid carriers and typhoid patients. *Environ Sci Pollut Res* 27(3):2844–2856
26. Yuza K, Sakata J, Prasoon P, Hirose Y, Ohashi T, Toge K, Miura K, Nagahashi M, Kobayashi T, Wakai T (2019) Long-term outcomes of surgical resection for T1b gallbladder cancer: an institutional evaluation. *BMC Cancer* 20(1):1–9
27. Mcgee EE, Koshiol J (2020) Comments on "Should we consider gallbladder cancer a new smoking related cancer? A comprehensive meta-analysis focused on dose-response relationships. *Int J Cancer* 147(2):593–594
28. Jin Y, Li F (2020) Comment on: Comparison of oncological outcomes after open and laparoscopic re-resection of incidental gallbladder cancer. *Br J Surg* 107(6):769–769
29. Goyal V, Singh G, Tiwari O, Punia S, Kumar M (2019) Intelligent skin cancer detection mobile application using convolution neural network. *J Adv Res Dyn Control Syst* 11(7):253–259
30. Wang H, Dong F, Wang Y, Wang X, Hong D, Liu Y, Zhou J (2020) Betulinic acid induces apoptosis of gallbladder cancer cells via repressing SCD1. *Acta Biochim Biophys Sin* 50(2):200–206
31. Vega EA, Conrad C (2020) Author response to: is laparoscopic re-resection of incidental gallbladder cancer really non-inferior to the open approach? *Br J Surg* 107(6):767–768
32. Wang H, Zhan M, Liu Q, Wang J (2020) Glycochenodeoxycholate promotes the metastasis of gallbladder cancer cells by inducing epithelial to mesenchymal transition via activation of SOCS3/JAK2/STAT3 signaling pathway. *J Cell Physiol* 235(2):1615–1623
33. Liska V, Treska V, Skalicky T, Fichtl J, Bruha J, Vycital O, Topolcan O, Palek R, Rosendorf J, Polivka J, Holubec L (2017) Evaluation of tumor markers and their impact on prognosis in gallbladder, bile duct and cholangiocellular carcinomas—a pilot study. *Anticancer Res* 37(4):2003–2009

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Yigang Chang^{1,2} · Qian Wu³ · Limin Chi⁴ · Huaying Huo² · Qiang Li¹

✉ Qiang Li
liqiang4016@yahoo.com

¹ Tianjin Medical University Cancer Institute and Hospital and National Clinical Research Center for Cancer, Tianjin 300060, China

² Shanxi Provincial People'S Hospital, Taiyuan 030012, China

³ Shanxi Tumor Hospital, Taiyuan 030013, China

⁴ First Hospital of Shanxi Medical University, Taiyuan 030001, China