

RESEARCH ARTICLE

Identification Tool for Gastric Cancer Based on Integration of 33 Clinical Available Blood Indices Through Deep Learning

BOWEN ZHANG^{1,2}, LONG CHENG³, YUZHEN NIU⁴, AMING WANG¹, PENGYI ZHANG²,
TIANIAN SHEN⁵, LILI XI⁶, DEKUI ZHANG³, AND SHUYAN LI¹

¹School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou 221004, China

²Department of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou 730000, China

³Department of Gastroenterology, Lanzhou University Second Hospital, Lanzhou 730000, China

⁴School of Life Science, Shandong University of Technology, Zibo 250100, China

⁵The First School of Clinical Medicine, Lanzhou University, Lanzhou 730000, China

⁶Institution of Drug Clinical Trial, The First Hospital of Lanzhou University, Lanzhou 730000, China

Corresponding authors: Shuyan Li (lishuyan@xzhmu.edu.cn) and Dekui Zhang (sczdk1972@163.com)

The work of Dekui Zhang was supported in part by the National Natural Science Foundation of China under Grant 81770525, in part by the Industrial Support and Guidance Project of Colleges and Universities in Gansu Province under Grant 2019C-21, and in part by the Cuiying Science and Technology Project of the Second Hospital of Lanzhou University under Grant CY2018-ZD01. The work of Lili Xi was supported in part by the Key Research and Development Project of Gansu Province under Grant 20JR5RA370.

ABSTRACT Gastric cancer (GC) is one of the most common cancers in the world. In cancer detection, liquid biopsy, as a noninvasive and rapid method, is growing in importance. Different from traditional liquid biopsy using a single biomarker, this study integrated a variety of blood biochemical indices and established an identification system by means of deep learning under the H2O framework method. Based on data from 2951 samples, 58 routine blood biochemical indices, age and gender were collected as comprehensive indices to establish the identification model. Then, the number of indices was reduced to simplify the model, and 33 indices were utilized to build the final identification tool. A tenfold cross-validation technique was used to evaluate the performance of the proposed method. The sensitivity, specificity, accuracy, and area under the ROC curve on the cross-validation set were 85.44%, 83.82%, 84.54% and 0.9165, respectively. The identification tool is built free online at <http://www.cppdd.cn/GC2>. The proposed system provides a new approach to identify GC with advantages of being efficient, noninvasive and economical. The deep learning of the integration of these blood biochemical indices will bring insights into the comprehensive understanding of GC pathology, as well as the prevention, screening, diagnosis, and prognosis of GC.

INDEX TERMS Clinical available blood indices, deep learning, gastric cancer, identification, online server.

I. INTRODUCTION

Gastric cancer (GC) was responsible for over 1.08 million new cancer cases in 2020 and more than 768,000 deaths worldwide [1]. Although the incidence and mortality of GC have declined, it is still the fifth most commonly diagnosed cancer and ranks fourth in the mortality rate globally [2]. The high prevalence of GC, poor prognosis [3], and limited treatment options have resulted in a heavy medical burden [4].

The development of GC is a multistage, multistep, and multimechanism process that is highly heterogeneous in

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

terms of structural growth, cell differentiation, molecular pathogenesis, and stages [4]. Therefore, the diagnosis of GC is difficult, especially early GC. Meanwhile, the 5-year survival rate of patients with early GC after undergoing surgery exceeds 90%. In contrast, the 5-year survival rate of patients with stage IV GC is less than 5% [5]. The earlier a patient with GC is diagnosed, the easier it can be cured [6]. However, due to nonspecific symptoms of early patients, less than 25% of patients with GC are detected at an early stage [3]. Therefore, a fast and low-cost detection tool is urgently needed to detect GC.

Endoscopy and image examination are the gold standard diagnostic methods for GC [7]. However, the detection

accuracy of conventional endoscopy is only 69% to 79% [8], and image examination has difficulty identifying early lesions with high accuracy [9]. In addition, endoscopy may cause gastric bleeding, gastric perforation, and bacterial infections [10]. Imaging examination should be combined with pathological examination to make a final diagnosis [11]. Doctors are required to have experience in endoscopy and image examination [8]. There is a severe shortage of endoscopists and large-scale endoscopy centers in poverty-stricken areas [12]. Therefore, endoscopy and imaging examinations are not suitable for large-scale screening.

As a minimally invasive method, liquid biopsy can provide information for cancer diagnosis, tumor monitoring, and clinical prognosis [13]. Here, blood is the most important body fluid [14], and plasma is excellent at reflecting gastrointestinal diseases [15]. Experts believe that GC tissue releases tumor markers into the peripheral blood, such as circulating tumor DNA (ctDNA), cell-free DNA (cfDNA), tumor-associated RNA, protein, and circulating tumor cells (CTCs) [16]. With the progress of genomics and proteomics and the development of PCR detection technology, an increasing number of blood biomarkers have been found and applied to the diagnosis and prognosis of cancer [17]. Traditional serum markers, such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9), have been applied to diagnose GC [18]. Studies have shown that the comprehensive sensitivity of blood DNA methylation as a diagnostic marker of GC is 57% (95% CI, 50-63%) [19]. CtDNA is affected by tumor type and stage, which needs further study [13]. CfDNA is more sensitive than conventional tumor markers, but it is insufficient to distinguish other diseases, such as inflammatory diseases and infection [16]. MicroRNA (miRNA) [20], long noncoding RNA (lncRNA) [6], and circular RNA (circRNA) [21] play a key role in the occurrence and development of tumors. However, the chemical instability of RNA prevents it from becoming a good biomarker [19]. The count of circulating tumor cells is considered to be less than 5 in 7.5 ml of blood, which makes detection difficult [13]. The combined detection of multiple biomarkers can avoid the interference of blood mutation template molecules and reduce sampling deviation and individual differences [17], [22]. Various organs, tissues, blood exchange substances, and blood indices are in a state of dynamic balance. This balance is likely to be disrupted due to body diseases, resulting in abnormal blood indices [23]. Wu *et al.* showed that the neutrophil-lymphocyte ratio (NLR) and platelet-lymphocyte ratio (PLR) were significantly different between GC patients with different stages. The combination of PLR and CEA is better than CEA alone (AUC=0.671) for the diagnosis of gastric cancer (AUC=0.780) [10]. The deep learning method with excellent self-learning ability can deal with multidimensional nonlinear statistical relationships better than traditional methods [24]. The Clinical available blood indices based on deep learning provide a new idea for the differentiation of GC.

In this work, an identification model of GC was established based on blood indices through a deep learning algorithm. GC and other diseases could be effectively distinguished in this model.

II. MATERIALS AND METHODS

A. SOURCE OF MATERIALS

The total number of samples was 2951, including 1836 negative samples and 1115 positive samples. A total of 2682 samples were collected for cross-validation, and 269 samples were collected for external validation. The positive samples were from GC patients, while the negative samples were from 559 healthy people, 229 gastric ulcer patients, 158 gastric polyps patients, 174 gastritis patients, 124 liver cancer patients, 203 lung cancer patients, 124 pancreatic cancer patients, 136 breast cancer patients, and 129 colon cancer patients. To avoid diagnostic errors, the diagnostic results of the above samples were confirmed by several experts. Each sample contains age, gender, and 26 routine blood indices from Sysmex XE-5000 and 32 biochemical indices from Ortho VITROS 5600. The details of the dataset are listed in Table 1, and detailed information can be shown in supplementary table S1. This study met the ethical requirements with the consent of all patients and healthy volunteers and was reviewed and approved by the ethics committee of the Second Hospital of Lanzhou University.

B. MACHINE LEARNING METHOD

Deep learning utilizes multiple neural network layers to obtain more important information from multidimensional data [8]. The artificial neural network can effectively reflect the nonlinear process of tumor development and metastasis [24]. The H2O framework includes an advanced artificial neural network [25]. Deep learning algorithms are embedded in the H2O framework. The deep learning algorithm is similar to the classical multilayer perceptron (MLP) [26], which is optimized through continuous iteration. Stochastic gradient descent was used for training the model, and backpropagation was used for optimization. The performance of the identification model can be further optimized by adjusting the hyperparameters of the neural network.

The principle of layered sampling was followed to divide the total data into the training set (2682 samples) and test set (269 samples). The initial H2O classification prediction model was evaluated by tenfold cross-validation on the training set, and the test set was used for external validation and did not participate in the construction of the model. The first was the choice of the number of hidden layers. In the initial experiment, we set the number of neurons in each hidden layer to 50 and then increased the number of hidden layers to obtain the AUC and logloss of the training set and tenfold cross-validation, as shown in Fig. 1. As shown in Fig. 1a, with the increase in the hidden layer, the AUC of the training set continued to increase, but the AUC of the tenfold cross-validation increased first and then oscillated. The logloss of

TABLE 1. Details of the cross-validation sets and external validation sets.

	Sample Number of CV set	Sample Number of EV set	Age range of CV set (mean \pm SD)	Age range of EV set (mean \pm SD)	Male/Female of CV set	Male/Female of EV set
GC	1014	101	16-86(56.2 \pm 11.5)	27-82(55.5 \pm 11.1)	739/275	78/23
Healthy	508	51	0-88(46.0 \pm 13.7)	24-76(46.4 \pm 13.4)	267/241	29/22
Gastric ulcer	208	21	2-83(55.7 \pm 14.9)	16-73(52.4 \pm 13.8)	140/68	18/3
Gastric polyps	144	14	15-81(56.7 \pm 12.2)	34-79(49.9 \pm 15.6)	54/90	8/6
Gastritis	158	16	10-90(55.5 \pm 14.7)	38-76(58.0 \pm 12.1)	72/86	4/12
Liver cancer	113	11	19-86(56.0 \pm 10.9)	51-67(56.6 \pm 5.6)	78/35	9/2
Lung cancer	185	18	27-81(59.5 \pm 11.0)	30-78(59.3 \pm 11.9)	137/48	11/7
Pancreatic cancer	112	12	26-87(60.3 \pm 12.1)	35-73(56.7 \pm 11.6)	59/53	9/3
Breast cancer	123	13	24-80(49.4 \pm 11.3)	36-52(46.7 \pm 4.8)	3/120	0/13
Gastric ulcer	208	21	2-83(55.7 \pm 14.9)	16-73(52.4 \pm 13.8)	140/68	18/3

tenfold cross-validation tended to increase as the hidden layer increased, as shown in Fig. 1b. Experiments showed that as the number of hidden layers increased, the performance of the model improved, but tenfold cross-validation showed that the model tended to fall into overfitting. To avoid overfitting and underfitting, three hidden layers were selected as the basic structure of the neural network. Then, the grid search method was applied to adjust the hyperparameters of the h2o.deeplearning function, including “input dropout ratio”, “activation”, “initial weight distribution”, “loss”, “distribution”, etc. Multiple hyperparameters were combined to determine the best solution. The specific settings are shown in Table 2. Finally, the number of iterations and the size of the hidden layer were adjusted. The final iterations were set to 1000, and the three hidden layers were 80, 80, and 110. At first, all the features, including age, gender, 26 routine blood indices and 32 biochemical indices were used as the input layer of the neural network, constructing model-1. To simplify the blood detection process and reduce the model noise, the built-in function of the H2O package (h2o.varimp) was applied to calculate the importance of indices. The value of the calculated importance were listed in table 3 as importance percentage. Then, the number of indices was reduced to establish the final model (model-2).

The deep learning neural network was executed by the H2O package (version 3.32.0.1) in R (version 3.2.4). The H2O package is a parallel machine learning package that provides fast, scalable machine learning algorithms.

C. ASSESSMENT METHOD

The generalization capability of the model can be evaluated by cross-validation, thereby avoiding excessive fit [27]. Tenfold cross-validation divided the training set into 10 parts, 9 of which were used to build the model, and the remaining one was used as the internal test set to verify the performance of the model. Ten models were established by repeating 10 times to detect the accuracy and reliability of the model. The external test set did not participate in the establishment

of the model and was only for verifying the performance of the model.

The visual effect of the model is presented with a confusion matrix [28]. Sensitivity (Sens, (1)), specificity (Spec, (2)), and accuracy (ACC, (3)) were calculated by true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN).

$$Spec = \frac{TN}{TN + FP} \quad (1)$$

$$Spec = \frac{TN}{TN + FP} \quad (2)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

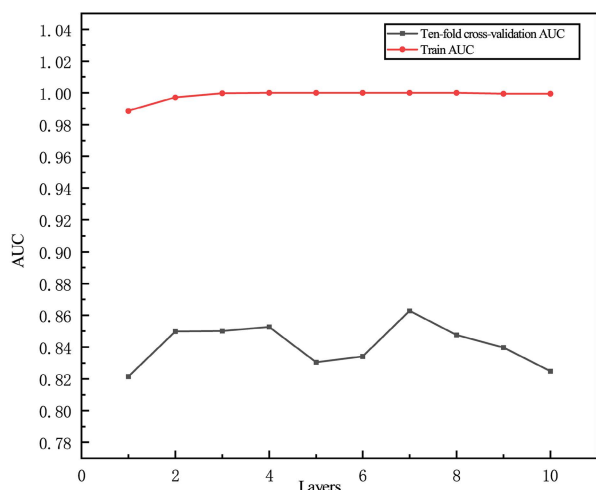
The receiver operating characteristic curve (ROC) was drawn by using sensitivity as the y-axis and 1-specificity as the x-axis. The area under the curve (AUC) was obtained by calculating the area under the ROC curve. The closer the AUC is to 1, the better classification of the model.

III. RESULTS

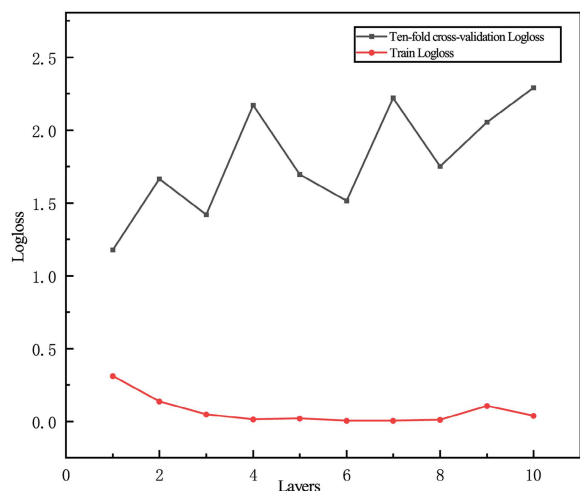
A. GOOD IDENTIFICATION PERFORMANCE

Model-1 is based on 60 features, with excellent classification and prediction performance. For the external test set, the AUC of model-1 was 0.9152, the sensitivity was 80.20%, the specificity was 91.07%, and the accuracy was 86.99%. The ROC curve of the external test set is shown in Fig. 2a.

As shown in Fig. 3, with the increase in features involved in modeling, the performance of the model increases. Finally, we selected 33 features to build the final model (model-2). Model-2 has good performance, with a sensitivity, specificity, total accuracy and area under the curve of 85.44%, 83.82%, 84.54% and 0.9165 for the cross-validation set, respectively. The selected features are shown in Table 3. For the external test set, the sensitivity, specificity, and accuracy of model-2 were 85.15%, 81.55%, and 82.90%, respectively. The ROC curve of the test set is shown in Fig. 2b, and the AUC was 0.9126.



(a)



(b)

FIGURE 1. Changes in AUC and logloss values for the training set and tenfold cross-validation on the training set as the number of hidden layers increased. (a) The AUC of the training set increased first and then flattened, while the AUC of the tenfold cross-validation first increased and then oscillated. (b) The logloss of the training set first decreased and then oscillated, and the logloss of the tenfold cross-validation tended to increase.

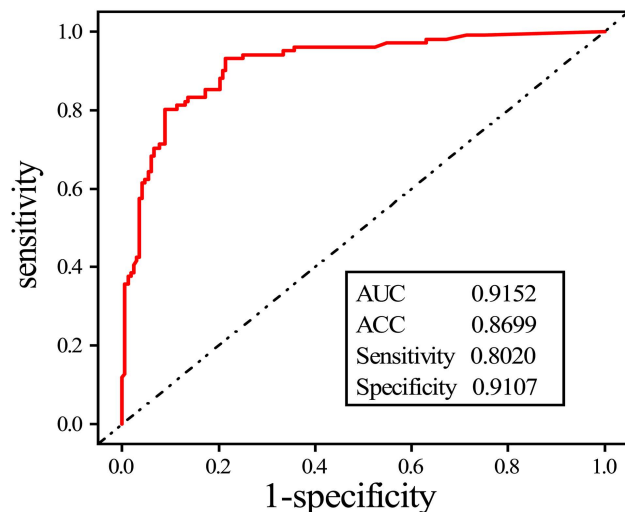
B. USER-ORIENTED ONLINE DIAGNOSTIC TOOL

We have designed an online identification website for patients and medical workers at <http://www.cppdd.cn/GC2>. The user can input the corresponding blood test data in the text box according to the prompt and obtain the corresponding identification results. The homepage of this website is shown in Fig. 4. Users should use the same equipment as the blood testing equipment used in this study. The system deviation caused by using other blood testing equipment might affect the accuracy of this identification.

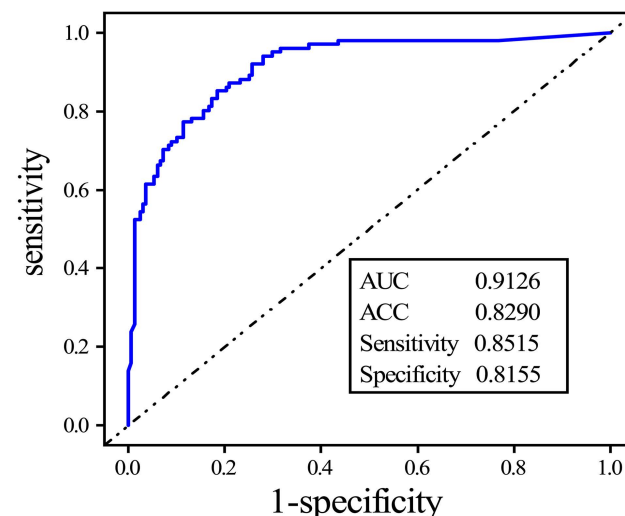
IV. DISCUSSION

A. ADVANTAGES OF THIS STUDY

A comprehensive comparison of previous works and our method is given in Table 4. In general, the identification



(a)



(b)

FIGURE 2. Identification performance of this method. (a) ROC curve and identification performance of the external validation set of model-1. (b) ROC curve and identification performance of the external validation set of model-2.

method for GC presented in this study is suitable for the screening of GC on a large scale compared with previous methods, especially in regions that lack medical resources, with the advantages listed below.

First, the blood indices on which this method were built were all collected from a Clinical available blood biochemical detector. These data are inexpensive and easy to obtain. Traditional methods for the detection of gastric cancer always rely on the detection of characteristics in gastric tissue. For example, Lu *et al.* detected the content of hsa_circ_0005758 (circRNA) in gastric tissue, and the sensitivity and specificity were 75.0% and 67.7%, respectively [30]. Pang *et al.* detected the content of LINC00152 (lncRNA) in gastric tissue; the sensitivity was 62.5%, and the specificity was 68.1% [31].

TABLE 2. Neural network hyperparameter settings.

Hyperparameters	Adjustment method	Hyperparameter settings
input dropout ratio	adjusted from 0 to 0.5 with a step of 0.1	0.2
activation	Tanh; TanhWithDropout; Rectifier; RectifierWithDropout; Maxout; MaxoutWithDropout	RectifierWithDropout
initial weight distribution	UniformAdaptive; Uniform; Normal	UniformAdaptive
loss	Automatic; CrossEntropy; Quadratic	Quadratic
distribution	AUTO; Bernoulli; Multinomial	Bernoulli
epochs	adjusted from 500 to 10000 with a step of 100	1000
hidden layer 1	adjusted from 30 to 120 with a step of 10	80
hidden layer 2	adjusted from 30 to 120 with a step of 10	80
hidden layer 3	adjusted from 30 to 120 with a step of 10	110

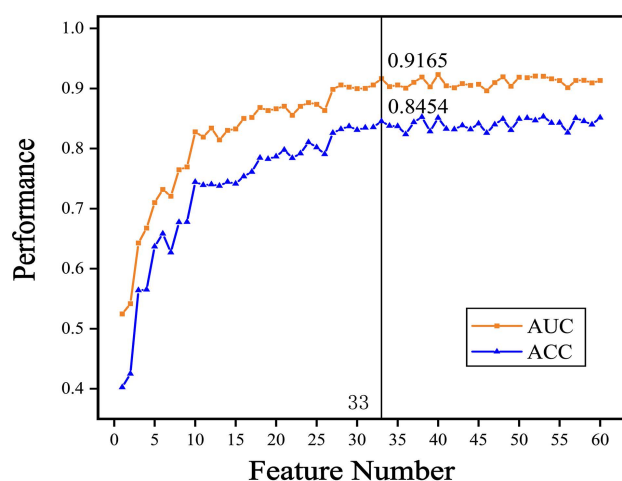


FIGURE 3. The performance of the model varies with the top-ranking number of features involved.

In addition, spectral-based methods on tissue images were also discovered and achieved outstanding results. For example, Li *et al.* analyzed the spectral characteristics of gastric mucosa tissue to diagnose GC through a deep learning method and achieved good results, with a specificity of 96.7% and a sensitivity of 96.6% [9]. The above studies using gastric tissue require experienced gastroscopic doctors to accurately extract gastric mucosa samples, which is not suitable for screening GC on a large scale. The sensitivity of the traditional GC biomarker pepsinogen was 69% [32]. Lin *et al.* used urinary surface-enhanced Raman spectroscopy (SERS) to diagnose GC from healthy sample based on gold nanoparticles, with a sensitivity of 90% and specificity of 93.8% [33]. However, Lin’s study was not efficient at distinguishing between GC and breast cancer, with specificity of 81.4% and a lower sensitivity of 62.0%. Cui *et al.* measured the content of mir-106a in gastric juice samples, and the sensitivity and specificity were 73.8% and 89.3%, respectively [20]. The sampling of gastric juice is more complicated than that of blood. Research on blood markers is extensive. Hu *et al.*

conducted a meta-analysis of DNA methylation in serum and plasma, and the results showed high specificity and low sensitivity [19]. Li *et al.* and Zhao *et al.* analyzed the content of circular RNA in plasma, and the results were not ideal [15], [18]. The use of plasma markers, such as protein P08493, MYC (cfDNA), and miR-20a (MiRNA), did not achieve good results [5], [34], [35]. Comprehensive multiple indices can achieve better results in the diagnosis of GC. Similar to this work, Zhu *et al.* used the gradient boosting decision tree (GBDT) method to analyze clinical blood indices, such as hemoglobin, and biomarkers, such as carcinoembryonic antigen (CEA), to establish a gastric cancer identification model and achieved good results [36]. Su *et al.* used machine learning to analyze the mass spectra of various proteins in serum and obtained good results [37]. Mass spectrometers are costly and have high requirements for operators, so this method is not suitable for large-scale screening.

Second, this method was developed based on a large and comprehensive dataset, including not only gastric cancer samples and healthy samples but also other major cancer samples and other gastric diseases with symptoms similar to gastric cancer. These samples included samples that are likely to be confused in mass screening of GC. Most of the previous work was only deduced on a small sample size [9], [33], [37]. Only a few works were studied on a large dataset with thousands of samples, as in this work, such as in works by Hu *et al.* [19] and Huang *et al.* [32].

Third, this method presents a better identification performance than previous works, especially for the test set, which implies that this method has a better generalization ability and stronger robustness, even than our previous work [29], with a better AUC value.

B. ANALYSIS OF KEY FEATURES

Cancer is associated with dysregulation of multiple biological processes, and the analysis of key features facilitates the discovery and validation of biomarkers that contribute to understanding cancer pathogenesis and developing drugs [38].

TABLE 3. Ranking of important features and results of correlation test.


Rank	Features (Abbreviation)	Importance percentage	<i>P</i> -value*	<i>P</i> value ¹	<i>P</i> value ²	<i>P</i> value ³
1	Creatine kinase isoenzymes (CK-MB)	0.0494	1.89E-14	1.70E-01	2.39E-02	7.99E-47
2	Magnesium (Mg)	0.0469	8.59E-01	4.63E-05	1.02E-06	1.98E-01
3	Immature granulocytes (IG#)	0.0452	8.89E-19	9.04E-29	1.67E-92	1.91E-11
4	Immature granulocyte ratio (IG%)	0.0431	1.20E-31	5.21E-01	3.37E-76	3.82E-08
5	Plateletcrit (PCT)	0.0390	1.72E-08	8.61E-24	3.00E-01	2.28E-02
6	Potassium (K)	0.0384	5.33E-03	5.00E-53	2.58E-03	4.60E-07
7	Gamma Glutamyl transferase (γ -GGT)	0.0288	3.00E-38	4.84E-27	1.17E-46	7.75E-04
8	Basophil count (BASO#)	0.0251	9.06E-06	1.45E-17	4.98E-01	3.24E-03
9	Calcium (Ca)	0.0232	7.74E-11	5.30E-73	3.90E-05	1.43E-02
10	Monocyte ratio (MO%)	0.0223	1.08E-26	2.65E-189	8.42E-01	5.52E-04
11	Direct bilirubin (DBil)	0.0213	7.52E-03	5.14E-01	1.44E-04	2.02E-01
12	Phosphorus (P)	0.0212	3.16E-03	2.11E-01	1.50E-02	4.79E-08
13	Lactate dehydrogenase (LD)	0.0205	5.04E-02	6.59E-01	1.37E-08	5.67E-28
14	Amylase (AMY)	0.0199	1.31E-10	7.42E-06	3.26E-13	1.59E-02
15	Basophil ratio (BASO%)	0.0197	3.07E-29	7.08E-218	1.16E-04	5.92E-02
16	Low-density lipoprotein (LDL)	0.0196	1.07E-16	1.66E-54	4.90E-01	1.65E-07
17	Red blood cell distribution width-CV (RDW-CV)	0.0188	1.27E-49	2.18E-73	1.10E-05	1.24E-34
18	Creatine kinase (CK)	0.0185	5.64E-22	3.35E-53	2.50E-01	9.63E-11
19	Blood urea nitrogen/Creatinine (BUN/Cre)	0.0184	2.00E-10	4.28E-21	3.27E-03	1.72E-02
20	Total bilirubin (TBil)	0.0167	9.96E-01	2.51E-01	2.47E-02	1.29E-01
21	Mean platelet volume (MPV)	0.0165	1.98E-01	6.21E-02	7.67E-01	1.12E-07
22	Alkaline phosphatase (ALP)	0.0161	1.80E-01	2.72E-04	1.00E-14	2.21E-02
23	Eosinophil ratio (EO%)	0.0151	2.97E-40	8.43E-194	1.04E-04	6.44E-02
24	Alanine transaminase (ALT)	0.0150	9.53E-11	2.88E-05	9.16E-11	2.02E-04
25	Triglyceride (TG)	0.0147	1.30E-15	4.86E-21	3.24E-01	5.27E-18
26	Gender	0.0143	1.95E-28	1.55E-16	8.61E-20	7.77E-17
27	Aspartate aminotransferase (AST)	0.0140	3.39E-01	1.11E-03	1.42E-08	2.51E-01
28	High-density lipoprotein (HDL)	0.0137	3.61E-01	1.25E-15	2.12E-06	8.52E-01
29	Red blood cell distribution width-SD (RDW-SD)	0.0133	3.17E-62	2.16E-112	6.91E-15	1.16E-16
30	Eosinophil cells (EO)	0.0133	1.08E-02	7.06E-08	1.29E-02	2.35E-04
31	Large platelet ratio (P-LCR)	0.0131	1.11E-01	1.19E-13	2.74E-04	3.47E-15
32	Indirect bilirubin (I-Bil)	0.0129	2.16E-01	7.65E-01	9.91E-01	7.28E-03
33	Albumin/Globulin (ALB/GLB)	0.0126	1.48E-05	4.76E-19	9.41E-48	3.14E-02

Note: *The statistical significant difference of features between gastric cancer (positive sample group) and all negative sample group, including three 3 sub-groups (normal group, other cancer group and other gastric disease group).¹The statistical significant difference of features between gastric cancer group and normal sub-group. ²The statistical significant difference of features between gastric cancer group and other cancer sub-group.³The statistical significant difference of features between gastric cancer group and other gastric disease sub-group. The detailed statistical method about that how the significant differences were calculated was shown in supplementary table S2.

To discover insights into key features in different populations, all samples were categorized into four population groups: the GC group, normal group, other cancer group, and other gastric disease group. We selected features with significant differences and drew a boxplot between groups, as shown in Fig. 5. In Fig. 5a, the levels of PCT in the normal group

were higher than those in the GC group. Current studies have shown that tumor cells can indirectly promote platelet production and activation. Activated platelets can also protect and even promote the growth and metastasis of tumor cells [39], [40]. Our statistics may show that the hematopoietic function of bone marrow in some patients with gastric

Online Identification of Gastric Cancer



- [Home](#)
- [Document](#)
- [Research](#)
- [About Us](#)
- [Contact Us](#)

[Xuzhou Medical University](#)

Please input single sample data in the text form:

<p>Monocyte ratio (MO%) <input type="text" value="0.09"/></p> <p>Amylase (AMY) <input type="text" value="72"/> U/L</p> <p>Immature granulocytes (IG#) <input type="text" value="0.02"/> $10^9/L$</p> <p>Creatine kinase (CK) <input type="text" value="75"/> U/L</p> <p>Calcium (Ca) <input type="text" value="1.97"/> mmol/L</p> <p>Creatine kinase isoenzymes (CK-MB) <input type="text" value="26"/> U/L</p> <p>Direct bilirubin (DBil) <input type="text" value="16"/> $\mu\text{mol}/L$</p> <p>Gamma-glutamyl transferase (γ-GGT) <input type="text" value="10"/> U/L</p> <p>Lactate dehydrogenase (LD) <input type="text" value="217"/> U/L</p> <p>Total bilirubin (TBil) <input type="text" value="18.3"/> $\mu\text{mol}/L$</p> <p>Red blood cell distribution width-CV (RDW-CV) <input type="text" value="15.5"/> %</p> <p>Mean platelet volume (MPV) <input type="text" value="8.1"/> fl</p> <p>Eosinophil ratio (EO%) <input type="text" value="0.03"/> %</p> <p>Triglyceride (TG) <input type="text" value="1.24"/> mmol/L</p> <p>High-density lipoprotein (HDL) <input type="text" value="1.06"/> mmol/L</p> <p>Eosinophil cells (EO) <input type="text" value="0.15"/> $10^9/L$</p> <p>ALB/GLB <input type="text" value="0.87"/></p>	<p>Magnesium (Mg) <input type="text" value="0.90"/> mmol/L</p> <p>Plateletcrit (PCT) <input type="text" value="0.21"/> %</p> <p>Immature granulocyte ratio (IG%) <input type="text" value="0.2"/> %</p> <p>Kalium (K) <input type="text" value="3.89"/> mmol/L</p> <p>Large platelet ratio (P-LCR) <input type="text" value="35.5"/> %</p> <p>Basophil count (BASO#) <input type="text" value="0.01"/> $10^9/L$</p> <p>Phosphorus (P) <input type="text" value="1.19"/> mmol/L</p> <p>Basophil ratio (BASO%) <input type="text" value="0.01"/> %</p> <p>Low-density lipoprotein (LDL) <input type="text" value="2.25"/> $\mu\text{mol}/L$</p> <p>Aspartate aminotransferase (AST) <input type="text" value="21"/> U/L</p> <p>Blood urea nitrogen Creatinine (BUN/Cre) <input type="text" value="0.04"/></p> <p>Alkaline phosphatase (ALP) <input type="text" value="73"/> U/L</p> <p>Alanine transaminase (ALT) <input type="text" value="17"/> U/L</p> <p>Sex <input type="text" value="1"/></p> <p>Red blood cell distribution width-SD (RDW-SD) <input type="text" value="52"/> fl</p> <p>Indirect bilirubin (I-Bil) <input type="text" value="2.3"/> $\mu\text{mol}/L$</p>
---	---

Copyright © 2021 - CPPDD · All Rights Reserved

FIGURE 4. Main web page of this GC identification tool.

cancer is affected by chemotherapy drugs or radiotherapy, resulting in the suppression of hematopoietic function and thrombocytopenia [41], [42].

As shown in Fig. 5b, the levels of potassium in the normal group were higher than those in other groups. Cancer and gastric diseases may lead to disordered homeostasis and potassium ion balance. Wu *et al.* showed that the potassium

content in GC tissue was higher than that in normal tissue [43]. Ding *et al.* proved that the Eag1 potassium channel was overexpressed in GC tissue [44]. In this experiment, the low levels of potassium in the blood of GC patients may be related to this, which requires further study.

In Fig. 5c, the basophil count of the normal group is shown to be higher and more concentrated than that of the other

TABLE 4. Performance comparison of this work with other methods for identifying GC. (N/A: not applicable).

Identification method	Sample number	Sample type	Sens (%)	Spec (%)	AUC
This work (cross-validation set)	2682	Blood indices	85.44	83.82	0.9165
This work (testing set)	269	Blood indices	85.15	81.55	0.9126
Our previous work [29]	322	Blood indices	82.52	86.30	0.9092
spectral-spatial [9]	120	Gastric tissue	96.6	96.7	N/A
hsa_circ_0005758(circRNA) [30]	263	Gastric tissue	75.0	67.7	0.721
LINC00152(lncRNA) [31]	142	Gastric tissue	62.5	68.1	0.645
pepsinogen [32]	3785	Serum	69	73	0.76
SERS [33]	98	Urine	90	93.8	0.981
miR-106a (MiRNA) [20]	141	Gastric juice	73.8	89.3	0.871
DNA methylation [19]	4172	Plasma	71	89	0.87
DNA methylation [19]	4172	Serum	50	98	0.91
hsa_circ_0001017 (circRNA) [15]	242	Plasma	67.6	89.7	0.851
hsa_circ_0000181 (circRNA) [18]	207	Plasma	99.0	20.6	0.582
Protein P08493 [5]	30	Plasma	73.3	80.0	0.813
MYC(cfDNA) [34]	136	Plasma	75.44	75.95	0.8155
miR-20a(MiRNA) [35]	120	Plasma	75	88	0.8593
Blood routine, CEA and CA [36]	213	Whole blood	87.0	84.1	0.86
mass spectral [37]	150	Serum	85.3	88.0	N/A

groups, which is also confirmed by the higher basophil ratio in Fig. 5d. Tumor-infiltrating basophils are considered to be a poor prognostic factor for GC, but there is no significant correlation between blood basophils and tumor-infiltrating basophils [45]. The higher percentage of basophils in the normal group may be due to the increase in white blood cells in the other groups.

As shown in Fig. 5e, the calcium content in the normal group was higher than that in the GC group. Studies have shown that calcium may inhibit the damage of salt to gastric mucosa, thereby reducing the risk of GC [46]. However, the work of Xie *et al.* indicated that serum calcium concentration was positively correlated with the expression of calcium-sensing receptor (CaSR). Calcium can activate the overexpression of CaSR to promote the proliferation of GC cells [47]. These contradictory conclusions need further study.

Monocytes (Mo) are believed to promote tumor growth and proliferation [48]. High levels of monocytes are thought to confer poor prognosis in GC patients [49]. Figure 5f shows that the Mo% of the normal group is significantly higher than that of the other groups, but there is no significant difference in the Mo content among the groups. The results show that the count of white blood cells in all groups except the normal group increased significantly.

A low concentration of creatine kinase (CK) in GC tissue has been demonstrated [50]. This is similar to the results shown in Fig. 5g. CK is very important for ATP homeostasis in cells. Abnormal CK levels may lead to apoptosis [51].

The increase in eosinophils in the tissue can improve the prognosis of patients with GC. It has been proven that eosinophils have a good anticancer effect on colorectal cancer [52]. In Fig. 5h, the EO% of the normal group is significantly lower than that of the GC group. Eosinophils in cancer patients may increase due to autoimmune responses.

The red blood cell distribution width-SD (RDW-SD) in the normal group was lower than that in the other groups, as shown in Fig. 5i. Wei *et al.* showed that the red blood cell distribution width of GC patients is significantly higher than that of the normal group and further increases with the development of tumors. This phenomenon may be related to the decrease in hemoglobin caused by malnutrition in GC patients and may be related to inflammation [53]. Chang *et al.* showed that the ratio of CK-MB to total CK in hematological malignancies was higher in patients with colorectal cancer, lung cancer, and hepatocellular carcinoma [54]. This conclusion is roughly the same as the distribution of CK-MB and CK in this study and provides a direction for distinguishing patients with gastric diseases from patients with gastric cancer.

One should note that there were some features that are not statistically significant (with high p-value) between gastric cancer sample group and whole negative sample group, such as Mg and DBil in table 3, but they were at least had significant difference (p-value <0.05) with one of the 3 negative sub-groups (normal group, other cancer group and other gastric cancer group). This is just the

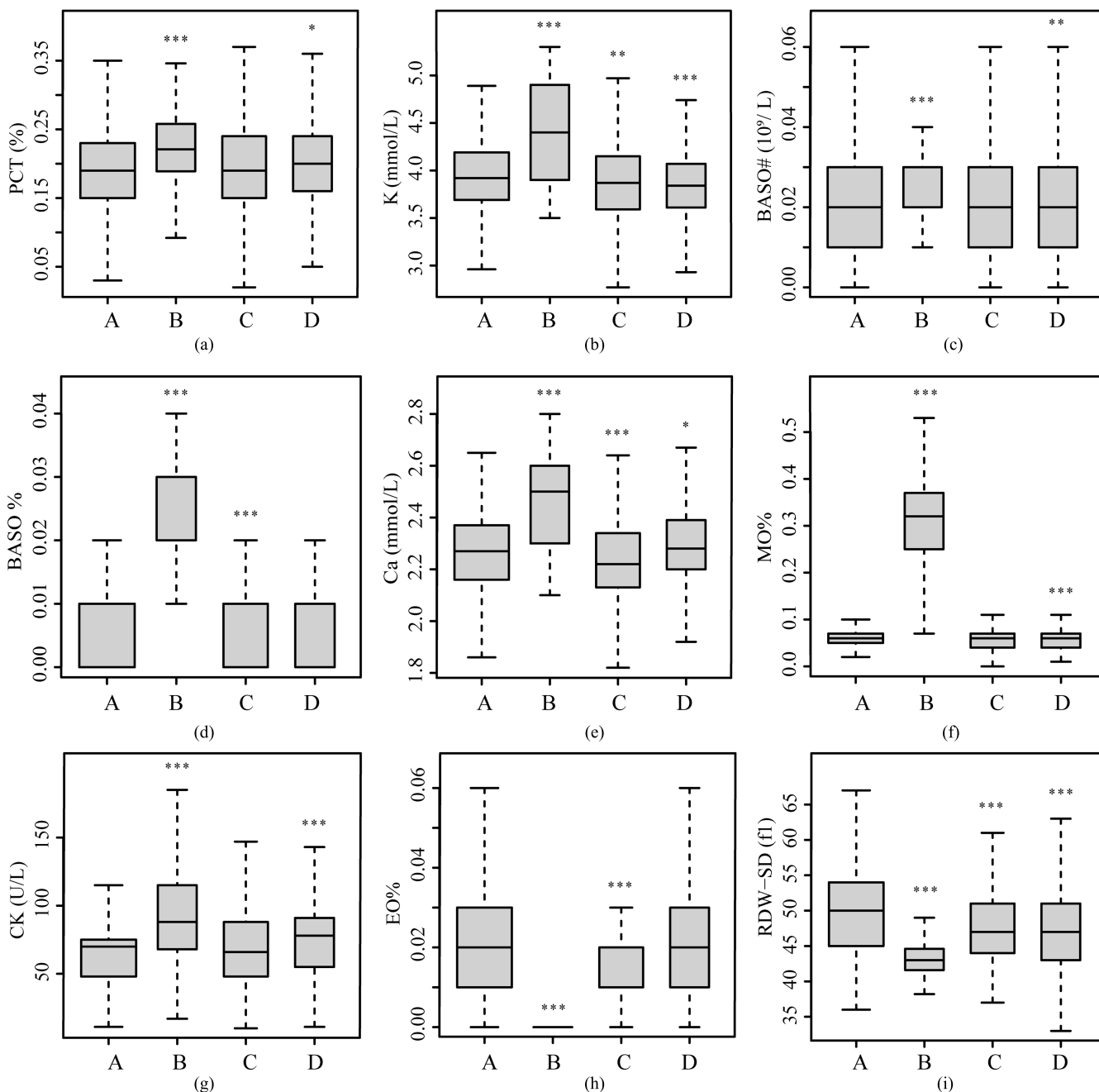


FIGURE 5. Features with significant differences in different patient groups. In the figure, A, B, C, and D represent the GC group, normal group, other cancer group, and other gastric disease group, respectively. PCT(a), plateletcrit; K(b), potassium; BASO#(c), basophil count; BASO%(d), basophil ratio; Ca(e), calcium; MO%(f), monocyte ratio; CK(g), creatine kinase; EO%(h), eosinophil ratio; RDW-SD(i), red blood cell distribution width-SD. (p-value=0.05-0.01, one star; p-value=0.01-0.001, two stars; p-value<0.001, three stars.)

reason why that deep learning is utilized here. Deep learning methods with excellent self-learning ability can handle multi-dimensional nonlinear statistical relationships better than traditional statistical methods [24]. Therefore, these features were also included as key features in the final model since their relationship with gastric cancer is in a multi-dimensional nonlinear way instead of linear statistical way.

V. CONCLUSION

We utilized an artificial neural network under the H2O framework method to process routine blood data and blood biochemical data, extracted 33 important indices as biomarkers for the identification of GC, and established an identification model and user-friendly web testing terminal. These biomarkers are considered to have great value in the physiological research of GC. Compared with conventional

diagnostic methods, the diagnostic model established in this study has the advantages of higher accuracy, noninvasiveness, and inexpensive detection. It has the potential to screen a wide range of people and effectively reduces the pressure on the medical system. With the further collection and learning of blood test data, the performance of this diagnostic model can be further improved.

APPENDIX

The original data of the samples for modeling and validation are provided in supplementary table S1.

The detailed statistical method and results for difference of selected features between gastric cancer group and other sub-groups are provided in supplementary table S2.

The box plots for all 33 selected features are provided in supplementary file S3.

ACKNOWLEDGMENT

(Bowen Zhang and Long Cheng contributed equally to this work.)

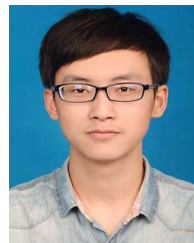
REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.
- [3] Z. H. Zhang, M. M. Dou, X. F. Yao, H. Tang, Z. B. Li, and X. Y. Zhao, "Potential biomarkers in diagnosis of human gastric cancer," *Cancer Invest.*, vol. 34, no. 3, pp. 22–115, 2016.
- [4] E. Van Cutsem, X. Sagaert, B. Topal, K. Haustermans, and H. Prenen, "Gastric cancer," *Lancet*, vol. 388, no. 10060, pp. 2654–2664, 2016.
- [5] B. Zhou, Z. Zhou, Y. Chen, H. Deng, Y. Cai, X. Rao, Y. Yin, and L. Rong, "Plasma proteomics-based identification of novel biomarkers in early gastric cancer," *Clin. Biochem.*, vol. 76, pp. 5–10, Feb. 2020.
- [6] E. Delshad, F. T. Shamsabadi, S. Bahramian, F. Mehravar, H. Maghsoudi, and M. Shafiee, "In silico identification of novel lncRNAs with a potential role in diagnosis of gastric cancer," *J. Biomol. Struct. Dyn.*, vol. 38, no. 7, pp. 1954–1962, May 2020.
- [7] Y. Huang, J. Zhu, W. Li, Z. Zhang, P. Xiong, H. Wang, and J. Zhang, "Serum microRNA panel excavated by machine learning as a potential biomarker for the detection of gastric cancer," *Oncol. Rep.*, vol. 39, pp. 1338–1346, Dec. 2017.
- [8] P.-H. Niu, L.-L. Zhao, H.-L. Wu, D.-B. Zhao, and Y.-T. Chen, "Artificial intelligence in gastric cancer: Application and future perspectives," *World J. Gastroenterol.*, vol. 26, no. 36, pp. 5408–5419, Sep. 2020.
- [9] Y. P. Li, L. Deng, X. Yang, Z. Liu, X. Zhao, F. Huang, S. Zhu, X. Chen, Z. Chen, and W. Zhang, "Early diagnosis of gastric cancer based on deep learning combined with the spectral-spatial classification method," *Biomed. Opt. Exp.*, vol. 10, no. 10, pp. 4999–5014, Oct. 2019.
- [10] Y. Wu, M. Jiang, Y. Qin, F. Lin, and M. Lai, "Single and combined use of neutrophil-lymphocyte ratio, platelet-lymphocyte ratio and carcinoembryonic antigen in diagnosing gastric cancer," *Clinica Chim. Acta*, vol. 481, pp. 20–24, Jun. 2018.
- [11] W. Ge, L. M. Zheng, and G. Chen, "The combination of seven preoperative markers for predicting patients with gastric cancer to be either stage IV or non-stage IV," *Gastroenterol Res. Pract.*, vol. 2018, Jun. 2018, Art. no. 3450981.
- [12] L. M. Zhang, Y. Zhang, L. Wang, J. Y. Wang, and Y. L. Liu, "Diagnosis of gastric lesions through a deep convolutional neural network," *Dig. Endosc.*, vol. 33, no. 5, pp. 788–796, Sep. 2020.
- [13] C. T. U. Guimarães, N. N. F. Martins, K. C. da Silva Oliveira, C. M. Almeida, T. M. Pinheiro, C. O. Gigeck, S. R. de Araújo Cavallero, P. P. Assumpção, M. A. C. Smith, R. R. Burbano, and D. Q. Calcagno, "Liquid biopsy provides new insights into gastric cancer," *Oncotarget*, vol. 9, no. 19, pp. 15144–15156, Mar. 2018.
- [14] M. Bahreini, A. Hosseinzadegan, A. Rashidi, S. R. Miri, H. R. Mirzaei, and P. Hajian, "A Raman-based serum constituents' analysis for gastric cancer diagnosis: *In vitro* study," *Talanta*, vol. 204, pp. 826–832, Nov. 2019.
- [15] T. Li, Y. Shao, L. Fu, Y. I. Xie, L. Zhu, W. Sun, R. Yu, B. Xiao, and J. Guo, "Plasma circular RNA profiling of patients with gastric cancer and their droplet digital RT-PCR detection," *J. Molecular Med.*, vol. 96, no. 1, pp. 85–96, 2018.
- [16] C.-T. Liu, C.-Q. Hong, X.-C. Huang, E.-M. Li, Y.-W. Xu, and Y.-H. Peng, "Blood-based markers in the prognostic prediction of esophagogastric junction cancer," *J. Cancer*, vol. 11, no. 15, pp. 4332–4342, 2020.
- [17] Z.-B. Huang, H.-T. Zhang, B. Yu, and D.-H. Yu, "Cell-free DNA as a liquid biopsy for early detection of gastric cancer (review)," *Oncol. Lett.*, vol. 21, no. 1, p. 1, Nov. 2020.
- [18] Q. Zhao, S. Chen, T. Li, B. Xiao, and X. Zhang, "Clinical values of circular RNA 0000181 in the screening of gastric cancer," *J. Clin. Lab. Anal.*, vol. 32, no. 4, May 2018, Art. no. e22333.
- [19] W. Hu, W. Zheng, Q. Liu, H. Chu, S. Chen, J. J. Kim, J. Wu, and J. Si, "Diagnostic accuracy of DNA methylation in detection of gastric cancer: A meta-analysis," *Oncotarget*, vol. 8, no. 68, pp. 113142–113152, Dec. 2017.
- [20] L. Cui, X. Zhang, G. Ye, T. Zheng, H. Song, H. Deng, B. Xiao, T. Xia, X. Yu, Y. Le, and J. Guo, "Gastric juice MicroRNAs as potential biomarkers for the screening of gastric cancer," *Cancer*, vol. 119, no. 9, pp. 1618–1626, May 2013.
- [21] Y. Xie, Y. Shao, W. Sun, G. Ye, X. Zhang, B. Xiao, and J. Guo, "Downregulated expression of hsa_circ_0074362 in gastric cancer and its potential diagnostic values," *Biomarkers Med.*, vol. 12, no. 1, pp. 11–20, Jan. 2018.
- [22] H. Sun, B. He, Z. Nie, Y. Pan, K. Lin, H. Peng, T. Xu, X. Chen, X. Hu, Z. Wu, D. Wu, and S. Wang, "A nomogram based on serum bilirubin and albumin levels predicts survival in gastric cancer patients," *Oncotarget*, vol. 8, no. 25, pp. 41305–41318, Jun. 2017.
- [23] P. A. Weiss and P. A. Weiss, "Can cancer be diagnosed with a blood test during routine examinations?" *Clin. J. Oncol. Nursing*, vol. 11, no. 6, pp. 1–940, 2007.
- [24] S.-J. Que, Q.-Y. Chen, Qing-Zhong, Z.-Y. Liu, J.-B. Wang, J.-X. Lin, J. Lu, L.-L. Cao, M. Lin, R.-H. Tu, Z.-N. Huang, J.-L. Lin, H.-L. Zheng, P. Li, C.-H. Zheng, C.-M. Huang, and J.-W. Xie, "Application of preoperative artificial neural network based on blood biomarkers and clinicopathological parameters for predicting long-term survival of patients with gastric cancer," *World J. Gastroenterol.*, vol. 25, no. 43, pp. 6451–6464, Nov. 2019.
- [25] A. S. Elsayad, A. I. E. Desouky, M. M. Salem, and M. Badawy, "A deep learning H₂O framework for emergency prediction in biomedical big data," *IEEE Access*, vol. 8, pp. 97231–97242, 2020.
- [26] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 97–683, Sep. 1992.
- [27] G. Jiang and W. Wang, "Error estimation based on variance analysis of k-fold cross-validation," *Pattern Recognit.*, vol. 69, pp. 94–106, Sep. 2017.
- [28] I. Duntsch and G. Gediga, "Indices for rough set approximation and the application to confusion matrices," *Int. J. Approx. Reasoning*, vol. 118, pp. 155–172, Mar. 2020.
- [29] J. Wu, Y. Yang, L. Cheng, J. Wu, L. Xi, Y. Ma, P. Zhang, X. Xu, D. Zhang, and S. Li, "GCdiscrimination: Identification of gastric cancer based on a milliliter of blood," *Briefings Bioinf.*, vol. 22, no. 1, pp. 536–544, Jan. 2021.
- [30] R. D. Lu, Y. F. Shao, X. P. Tao, G. L. Ye, B. X. Xiao, and J. M. Guo, "Clinical significances of hsa_circ_0067582 and hsa_circ_0005758 in gastric cancer tissues," *J. Clin. Lab. Anal.*, vol. 33, no. 9, Nov. 2019, Art. no. e22984.
- [31] Q. Pang, J. Ge, Y. Shao, W. Sun, H. Song, T. Xia, B. Xiao, and J. Guo, "Increased expression of long intergenic non-coding RNA LINC00152 in gastric cancer and its clinical significance," *Tumor Biol.*, vol. 35, no. 6, pp. 5441–5447, Jun. 2014.
- [32] Y.-K. Huang, J.-C. Yu, W.-M. Kang, Z.-Q. Ma, X. Ye, S.-B. Tian, and C. Yan, "Significance of serum pepsinogens as a biomarker for gastric cancer and atrophic gastritis screening: A systematic review and meta-analysis," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0142080.

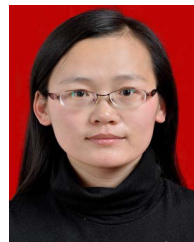
- [33] X. Lin, Q. Wu, J. Lin, X. Liu, X. Jia, M. Peng, D. Lin, S. Qiu, and S. Feng, "Label-free liquid biopsy based on urine analysis using surface-enhanced Raman spectroscopy for noninvasive gastric and breast cancer detection," *J. Raman Spectrosc.*, vol. 51, no. 11, pp. 2245–2254, Nov. 2020.
- [34] K. U. Park, H. E. Lee, D. J. Park, E. J. Jung, J. Song, H.-H. Kim, G. Choe, W. H. Kim, and H. S. Lee, "MYC quantitation in cell-free plasma DNA by real-time PCR for gastric cancer diagnosis," *Clin. Chem. Lab. Med.*, vol. 47, no. 5, pp. 6–530, Jan. 2009.
- [35] H. Cai, Y. Yuan, Y.-F. Hao, T.-K. Guo, X. Wei, and Y.-M. Zhang, "Plasma microRNAs serve as novel potential biomarkers for early detection of gastric cancer," *Med. Oncol.*, vol. 30, no. 1, p. 452, Mar. 2013.
- [36] S.-L. Zhu, J. Dong, C. Zhang, Y.-B. Huang, and W. Pan, "Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics," *PLoS ONE*, vol. 15, no. 12, Dec. 2020, Art. no. e0244869.
- [37] Y. Su, J. Shen, H. Qian, H. Ma, J. Ji, H. Ma, L. Ma, W. Zhang, L. Meng, Z. Li, J. Wu, G. Jin, J. Zhang, and C. Shou, "Diagnosis of gastric cancer using decision tree classification of mass spectral data," *Cancer Sci.*, vol. 98, no. 1, pp. 37–43, Jan. 2007.
- [38] J. Fu, Y. Zhang, J. Liu, X. Lian, J. Tang, and F. Zhu, "Pharmacometabonomics: Data processing and statistical analysis," *Briefings Bioinf.*, vol. 22, no. 5, p. 25, Sep. 2021.
- [39] X. Zhu, Y. Cao, P. Lu, Y. Kang, Z. Lin, T. Hao, and Y. Song, "Evaluation of platelet indices as diagnostic biomarkers for colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, Aug. 2018, Art. no. 11814.
- [40] W. Qian, X. Ge, J. Wu, F. Gong, M. Wu, M. Xu, L. Lian, W. Wang, W. Li, and M. Tao, "Prognostic evaluation of resectable colorectal cancer using platelet-associated indicators," *Oncol. Lett.*, vol. 18, pp. 571–580, May 2019.
- [41] E. D. Hager, H. Dziambor, D. Hohmann, P. Winkler, and H. Strama, "Effects of lithium on thrombopoiesis in patients with low platelet cell counts following chemotherapy or radiotherapy," *Biol. Trace Elem. Res., Article*, vol. 83, no. 2, pp. 139–148, Nov. 2001.
- [42] F. Heubel-Moenen, Y. M. C. Henskens, P. W. M. Verhezen, R. J. H. Wetzels, H. C. Schouten, and E. A. M. Beckers, "Fibrinolysis in patients with chemotherapy-induced thrombocytopenia and the effect of platelet transfusion," *J. Thrombosis Haemostasis*, vol. 17, no. 7, pp. 1073–1084, Jul. 2019.
- [43] C.-W. Wu, Y.-Y. Wei, C.-W. Chi, W.-Y. Lui, F.-K. P'eng, and C. Chung, "Tissue potassium, selenium, and iron levels associated with gastric cancer progression," *Digestive Diseases Sci.*, vol. 41, no. 1, pp. 119–125, Jan. 1996.
- [44] X. W. Ding, H. S. Luo, X. Jin, J. J. Yan, and Y. W. Ai, "Aberrant expression of Eag1 potassium channels in gastric cancer patients and cell lines," *Med Oncol.*, vol. 24, no. 3, pp. 50–345, 2007.
- [45] X. He, Y. Cao, Y. Gu, H. Fang, J. Wang, X. Liu, K. Lv, K. Yu, Y. Fei, C. Lin, H. Liu, H. Zhang, H. Li, J. Xu, R. Li, and H. He, "Clinical outcomes and immune metrics in intratumoral basophil-enriched gastric cancer patients," *Ann. Surgical Oncol.*, vol. 28, no. 11, pp. 6439–6450, Mar. 2021.
- [46] C.-Y. Yang, M.-F. Cheng, S.-S. Tsai, and Y.-L. Hsieh, "Calcium, magnesium, and nitrate in drinking water and gastric cancer mortality," *Japanese J. Cancer Res.*, vol. 89, no. 2, pp. 124–130, Feb. 1998.
- [47] R. Xie, J. Xu, Y. Xiao, J. Wu, H. Wan, B. Tang, J. Liu, Y. Fan, S. Wang, Y. Wu, T. X. Dong, M. X. Zhu, J. M. Carethers, H. Dong, and S. Yang, "Calcium promotes human gastric cancer via a novel coupling of calcium-sensing receptor and TRPV4 channel," *Cancer Res.*, vol. 77, no. 23, pp. 6499–6512, Dec. 2017.
- [48] Y.-L. Wang, X.-X. Ge, Y. Wang, M.-D. Xu, F.-R. Gong, M. Tao, W.-J. Wang, L.-M. Shou, K. Chen, M.-Y. Wu, and W. Li, "The values of applying classification and counts of white blood cells to the prognostic evaluation of resectable gastric cancers," *BMC Gastroenterol.*, vol. 18, no. 1, p. 99, Jun. 2018.
- [49] Y.-C. Pan, Z.-F. Jia, D.-H. Cao, Y.-H. Wu, J. Jiang, S.-M. Wen, D. Zhao, S.-L. Zhang, and X.-Y. Cao, "Preoperative lymphocyte-to-monocyte ratio (LMR) could independently predict overall survival of resectable gastric cancer patients," *Medicine*, vol. 97, no. 52, Dec. 2018, Art. no. e13896.
- [50] R. D. Hirata, M. H. Hirata, B. Strufaldi, R. A. Possik, and M. Asai, "Creatine kinase and lactate dehydrogenase isoenzymes in serum and tissues of patients with stomach adenocarcinoma," *Clin. Chem.*, vol. 35, no. 7, pp. 1385–1389, Jul. 1989.
- [51] Y.-B. Yan, "Creatine kinase in cell cycle regulation and cancer," *Amino Acids*, vol. 48, no. 8, pp. 1775–1784, Aug. 2016.
- [52] H. Reichman, M. Itan, P. Rozenberg, T. Yarmolovski, E. Brazowski, C. Varol, N. Gluck, S. Shapira, N. Arber, U. Qimron, D. Karo-Atar, J. J. Lee, and A. Munitz, "Activated eosinophils exert antitumorogenic activities in colorectal cancer," *Cancer Immunol. Res.*, vol. 7, no. 3, pp. 388–400, Mar. 2019.
- [53] T.-T. Wei, L.-L. Wang, J.-R. Yin, Y.-T. Liu, B.-D. Qin, J.-Y. Li, X. Yin, L. Zhou, and R.-Q. Zhong, "Relationship between red blood cell distribution width, bilirubin, and clinical characteristics of patients with gastric cancer," *Int. J. Lab. Hematol.*, vol. 39, no. 5, pp. 497–501, Oct. 2017.
- [54] C.-C. Chang, C. B. Liou, M. J. Su, Y. C. Lee, C. T. Liang, J. L. Ho, H. W. Tsai, T. H. Yen, and F. Y. Chu, "Creatine kinase (CK)-MB-to-total-CK ratio: A laboratory indicator for primary cancer screening," *Asian Pacific J. Cancer Prevention*, vol. 16, no. 15, pp. 6599–6603, 2015.



BOWEN ZHANG is currently pursuing the M.D. degree with the College of Chemistry and Chemical Engineering, Lanzhou University. His current research interest includes artificial intelligence in clinical diagnosis systems.



LONG CHENG is currently pursuing the Ph.D. degree with the Department of Gastroenterology, Lanzhou University Second Hospital. His current research interest includes digestive system diseases, such as ulcerative colitis.



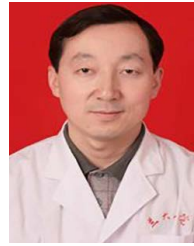
YUZHEN NIU received the Ph.D. degree from the School of Chemistry and Chemical Engineering, Lanzhou University, in June 2017. Then, she works with the School of Life Sciences and Medicine, Shandong University of Technology. Her current research interest includes computer-aided drug design.



AMING WANG is currently a Professor with the School of Medical Information and Engineering, Xuzhou Medical University. His research interest includes application of artificial neural network model in medical information processing.



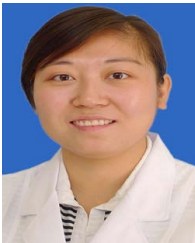
PENGYI ZHANG received the M.D. degree from the College of Chemistry and Chemical Engineering, Lanzhou University, in June 2021. His current research interests include artificial intelligence in clinical diagnosis systems and toxoplasmosis drug development.



DEKUI ZHANG is currently a Chief with the Department of Gastroenterology, Lanzhou University Second Hospital. His current research interests include diagnosis and treatment of esophagus and gastropathy, inflammatory bowel disease and digestive tract tumors, and endoscopic minimally invasive treatment of digestive diseases.



TIANTIAN SHEN is currently pursuing the M.D. degree with the First School of Clinical Medical, Lanzhou University. Her current research interests include virtual drug screening and biological activity verification.



LILI XI received the Ph.D. degree from the School of Chemistry and Chemical Engineering, Lanzhou University, in June 2010. Since July 2010, she has been working with the Institution of Drug Clinical Trial, The First Hospital of Lanzhou University. Her current research interests include computer-aided drug molecular design and computer-aided diagnosis models.



SHUYAN LI is currently a Professor with the College of Medical Information and Engineering, Xuzhou Medical University. Her current research interests include application of artificial intelligence in clinical diagnosis systems and computer-aided drug design.

...