

Model-Independent Evaluation of Tumor Markers and a Logistic-Tree Approach to Diagnostic Decision Support

Weizeng Ni¹, Samuel H. Huang^{1*}, Qiang Su² and Jinghua Shi³

¹*Department of Mechanical Engineering, University of Cincinnati
Cincinnati, OH, USA.*

²*School of Economics & Management, Tongji University
Shanghai, P. R. China.*

³*Department of Industrial Engineering and Logistics Management, Shanghai
Jiaotong University, Shanghai, P. R. China.*

ABSTRACT

Sensitivity and specificity of using individual tumor markers hardly meet the clinical requirement. This challenge gave rise to many efforts, e.g., combing multiple tumor markers and employing machine learning algorithms. However, results from different studies are often inconsistent, which are partially attributed to the use of different evaluation criteria. Also, the wide use of model-dependent validation leads to high possibility of data overfitting when complex models are used for diagnosis. We propose two model-independent criteria, namely, area under the curve (AUC) and Relief to evaluate the diagnostic values of individual and multiple tumor markers, respectively. For diagnostic decision support, we propose the use of logistic-tree which combines decision tree and logistic regression. Application on a colorectal cancer dataset shows that the proposed evaluation criteria produce results that are consistent with current knowledge. Furthermore, the simple and highly interpretable logistic-tree has diagnostic performance that is competitive with other complex models.

Keywords: cancer diagnosis, decision support, machine learning, tumor marker

1. INTRODUCTION

Serum-based tumor markers are claimed to be “most ideal for early diagnosis of cancer” [1]. As such, researchers have applied a wide array of machine learning methods and statistical tools on multiple tumor markers in pursuit of accurately predicting tumor types. Although promising outcomes have been achieved, results from different studies are often inconsistent. In this paper, we investigate the origins of the inconsistency in

*Corresponding author: Samuel H. Huang, Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH 45221. Phone: (513) 556-1154. Fax: (513) 556-3390. E-mail: sam.huag@uc.edu. Other authors: niwz@mail.uc.edu; suq@tongji.edu.cn; cherry_shi44@sjtu.edu.cn.

three aspects, namely, evaluation criteria, evaluation procedures, and predictive models. Based on the findings, we propose proper criteria to evaluate the diagnostic value of tumor markers and a new predictive model with the objective of providing an unbiased, effective, and interpretable tumor marker-based cancer diagnosis procedure.

Early diagnosis will significantly increase the chance of successful cancer treatment. Conventional computed tomography (CT) and transabdominal ultrasonography are often employed for initial imaging in early diagnosis [2]. However, CT devices and operations are too costly to be deployed for large population, and it is not applicable to tumors with small undetectable size. A great deal of efforts have been focused on serum-based tumor markers due to their ease of operation and cost advantage, which have the potential of application on asymptomatic population of large scale. In the past decades, an increasing number of tumor makers have been discovered and reported to be related to specific cancers. For instance, Ni *et al.* [3] reported more than 10 tumor markers for pancreatic cancer. Sun *et al.* [4] evaluated 12 tumor markers on hundreds of patients with diverse cancer types including lung cancer, breast cancer, and liver cancer. The use of individual tumor markers for cancer diagnosis, however, is limited since the resulting sensitivity and specificity hardly meet the strict clinical requirement. For example, the 10 reported markers in [3] are found to lack sufficient sensitivity and specificity for diagnosis. Therefore, researchers resorted to combining multiple tumor markers and employing machine learning algorithms to improve diagnosis accuracy. However, inconsistency of outcomes across different studies was found. Tumor markers initially reported to be useful may fail to yield consistently good outcomes in subsequent studies, and different studies drew contradictory conclusions of optimal tumor markers for the same type of cancer. This inconsistency could be explained by a number of factors such as poor study design, assays that are not standardized, and inappropriate or misleading statistical analyses [5]. The problems of evaluation criteria, evaluation procedures, and predictive models could potentially be the cause of the inconsistency, which will be looked into and discussed as follows.

In many earlier studies, the individual diagnostic value of tumor marker is determined by sensitivity and specificity that are calculated by labeling patients as elevated or un-elevated based on suggested cutoff points. Unfortunately, individual tumor marker is rarely found to be clinically acceptable due to the low sensitivity, which stems from the fact that elevation of tumor marker is frequently found in benign disorder and normal individual [6]. Researchers also combined multiple tumor markers by parallel test [2, 3] and serial test [7]. In a parallel test, patients are labeled as elevated only when all tumor markers levels are above the corresponding cutoff points; while in a serial test, patients are labeled as elevated as long as any one of the serum levels is above the cutoff point. These two procedures have been widely used due to its simplicity and interpretability. However, strictly adhering to the suggested cutoff points has two drawbacks: 1) they are determined for different patient groups in different studies thus may represent different percentiles of benign controls [8]; and 2) referring to only one fixed cutoff point does not fully take advantage of diagnostic information conveyed by the continuous values. The former drawback can be illustrated by the following example. Based on the same suggested cutoff point, Moghimi *et al.* [2] calculated the sensitivity and specificity of tumor marker CEA for pancreatic cancer as

83.78% and 69.44%, respectively, whereas Ni *et al.* [3] achieved results of 45% for sensitivity and 75% for specificity.

From the standpoint of machine learning, assessing the diagnostic value of tumor markers can be viewed as the procedure of building a predictive model and evaluating the discriminative power of independent variables (e.g., tumor markers, sex, and age) with respect to the response variable (e.g., tumor type) based on the prediction accuracy. Parallel and serial tests can be viewed as simple predictive models taking specificity and sensitivity as the criteria of prediction performance. Their prediction procedures seem to be “rough” in combining multiple independent variables and are not suited for accommodating the contradictory effect of sensitivity and specificity. By this rationale, several models originated from machine learning such as support vector machine (SVM) [4], classification tree [9], evolutionary algorithm [10-12] and neural networks [9, 13-17] have been widely applied for both evaluation of diagnostic values and diagnostic decision support. Sun *et al.* [4] applied SVM on a sample set with 12 tumor markers, and it was found that compared with parallel test, SVM achieved more specificity increase at the expense of less sensitivity decrease. Poon *et al.* [9] applied classification tree and neural networks for the diagnosis of the hepatocellular carcinoma subgroup without elevated tumor marker AFP, and it was concluded that neural networks appeared to outperform classification tree and its potential of recognizing subtle diagnostic pattern for multiple tumor markers was demonstrated. Flores-Fernández *et al.* [17] selected CEA, CA125, and CRP from 14 biomarkers using neural networks to constitute an optimized set of biomarkers for lung cancer diagnosis. However, complex models such as SVM and neural networks may not be the best choices for capturing the simple dependency underlying tumor markers and tumor type. Inappropriate evaluation procedure that lacks independent model validation is another origin of misleading results. Carpelan-Holmström *et al.* [8] took AUC derived from the probabilities returned by logistic regression model which was trained on the whole dataset as the evaluation criterion, ignoring the model bias toward overoptimistic result. This could be avoided by training a model on a training dataset and then evaluating it on an independent validation dataset.

In light of these observations and analyses, we realize that a good evaluation criterion for diagnostic value of tumor markers should be model-independent, that serves universally under different scenarios and considers multivariate diagnostic value. We propose area under the curve (AUC) and Relief weights as the evaluation criteria for univariate and multivariate diagnostic values, respectively. A simple model termed *logistic-tree* is developed specifically for diagnostic decision support. Combining the advantages of classification tree and logistic regression, a logistic-tree is able to present understandable diagnosis procedure, explore multi-variable interaction effect, and output continuous probabilities.

2. METHODS

2.1. Univariate Evaluation Criterion

AUC has been applied to cancer diagnosis in several studies [18-21]. In this paper, AUC serves as the evaluation criterion for both diagnostic values of individual markers and performance of predictive models. The advantages of ROC curve and AUC for cancer

diagnosis can be summarized as follows: 1) ROC curve allows clinicians to arbitrarily choose a cutoff point by which a certain requirement for sensitivity (or specificity) can be guaranteed, 2) the diagnostic information conveyed by continuous values can be fully utilized, and 3) it is also applicable to continuous probabilities returned by predictive models.

2.2. Multivariate Evaluation Criterion

Multivariate dependency represents the discriminative power in the context of multiple variables with respect to the response variable. Other than individual discriminative power, it includes two more factors of dependency: interaction effect and redundancy. Interaction effect in cancer diagnosis can be explained by the phenomenon that high diagnosis accuracy is achieved by considering two tumor markers together, although each one has weak diagnostic value. Redundancy is explained as follows: when the information carried by two tumor markers are highly correlated, combining these two tumor markers will not improve the diagnosis accuracy compared to the case when only one of the tumor marker is used. It is preferable to remove redundant variables for the purpose of identifying an optimal subset of discriminative variables.

Good at detecting interaction effect, Relief algorithm [22] is a variable estimation algorithm with a natural interpretation and many successful applications. Its pseudo code is presented as follows:

```

Input: training dataset comprised of  $m$  variables
      and a response variable with two classes.
Output: Relief weights  $W[m]$  for the  $m$  variables.
Initialize the vector  $W[m] = 0$ ;
do  $i = 1$  to  $n$ 
  randomly select a sample  $S_i$ ;
  find the nearest hit  $H$  and nearest miss  $M$ ;
  do  $j = 1$  to  $m$ 
     $Diff. H = diff(j, S_i, H)$ ;
     $Diff. M = diff(j, S_i, M)$ ;
     $W[j] = W[j] - Diff. H / n + Diff. M / n$ ;
  end do
end do

```

The input to Relief algorithm is a dataset characterized by m variables and a binary response variable. The variables are firstly standardized to jointly construct a m -dimensional variable space. In each iteration, a sample is randomly selected. Then the nearest hit can be identified which is defined as the sample whose position in the variable space is nearest to the selected sample and also has the same class. Analogously, the nearest miss is defined as the sample whose position in the variable space is nearest to the selected sample but has the different class. Intuitively, the variables whose values differ more between the selected sample and its nearest miss are thought to contribute more to differentiate them. Therefore, the difference of variable values between the selected sample and its nearest miss is referred to as having positive discriminative power. Likewise, for the selected sample and the nearest hit, discriminative variables are supposed to have close values [23]. Difference

of variable values in this situation is called negative discriminative power. Following this rationale, Relief values are updated by adding the weighted positive discriminative power and subtracting the weighted negative discriminative power. In this manner, Relief values gradually evolve by iterating this procedure.

It should be noted that Relief algorithm is incapable of detecting redundant variables. Since the number of tumor markers is low in cancer diagnosis application and they are rarely correlated perfectly in reality, the benefits of the diagnosis information that redundant variables bring probably will not be offset by the redundancy effect. Therefore, redundant variables are unlikely to degrade the efficiency or effectiveness of predictive models. On the other hand, the redundancy between tumor markers merits analysis with domain knowledge and it could be helpful if used wisely. For example, two tumor markers are found to be related, so their relationship may help reveal some valuable biomedical information of the interaction between two cells producing these two tumor markers. For practical cancer diagnosis, domain knowledge can help choose the tumor marker whose level is easier to be measured in order to lower the experiment cost.

2.3. Logistic-tree Model

Many seemingly “sophisticated” models such as SVM and neural networks may not be suitable for cancer diagnosis. The predominant reason lies in the assumption of simple structure underlying the tumor markers and tumor type. Normally the number of tumor markers found to be informative is less than a dozen. There is a low probability of complicated dependency between tumor markers and tumor type, so models designed to discover complex structure have higher probability of overfitting dataset than simple models. The second reason is that the complicated prediction procedures, often described as “black box,” are uninterpretable for domain experts as the process and results cannot be explained and validated by domain knowledge. This hinders the acceptance of these models by clinicians. Logistic regression model and classification tree can overcome this problem. With sound statistical justification and straightforward prediction procedure, “what” they predict and “how” they predict can be easily interpreted by domain knowledge and validated by statistical methods.

In cancer diagnosis application, predictive models normally act as decision support, so the continuous probabilities of tumor type are favored for the follow-up diagnosis and interpretation. Logistic regression model typically produces continuous probabilities ranging from 0 to 1, while classification tree provides discrete probabilities since all instances residing in the same terminal node share same probabilities irrespective of the different continuous variable values. On the other hand, the prediction mechanism of classification tree helps detect interaction effect of multiple independent variables which happens to be the weakness of logistic regression model. To combine their respective strengths, we design a model termed *logistic-tree* by merging logistic regression model and classification tree in the following way. A logistic regression model is developed on each terminal node after the pruned tree has been built. In contrast to all other models, the advantages of logistic-tree can be characterized as 1) presenting understandable prediction procedure, 2) detecting interaction effect, and (3) providing continuous probabilities. Its diagnosis power will be assessed by comparing it with other models in the experimental study.

Table 1. Percentages of missing values for all independent variables

Variable names	CEA	Age	CA199	AFP	CA125	CA50	CA724	CA153	CA211	CA242
Percentage of missing value (%)	0.00	1.56	10.21	23.53	27.34	36.33	78.03	88.24	88.75	91.52

3. RESULTS

3.1. Dataset

The dataset of colorectal cancer [24] was provided by Renji Hospital, a university hospital affiliated with Shanghai Jiaotong University, Shanghai, China, and the study was approved by the same university. The raw dataset is composed of data of 578 patients characterized by measurements of 9 serum-based tumor markers and age. Each patient has colorectal tumor either benign or malignant. These 9 tumor markers are not all measured and recorded for all patients, resulting in large amount of missing values. Table 1 lists these 9 tumor markers and age, along with their corresponding percentages of missing values in increasing order. For simplicity, the tumor markers whose percentages of missing values greater than 50% are excluded and only the cases having no missing values for all the remaining variables are retained for analysis. Thus, the final dataset is reduced to a dataset of 253 patient samples and 6 independent variables (5 tumor markers and age). Among the 253 patients, 111 patients are diagnosed with malignant tumors and 142 patients as having benign tumors. Note that the dataset is transformed by applying natural logarithmic scale on each tumor marker, and the tumor marker values presented below are the transformed values.

An experiment is conducted to classify these patients as having either benign or malignant tumors based on the tumor marker values. The methodologies described in the previous section are tested with the data. The study is implemented on the statistical computing environment of R, which has modeling packages including SVM, ANNs, random forest, and others. A brief summary of the experimental procedure can be found in Figure 1.

3.2. Diagnostic Value Evaluation

Figure 2 shows the ROC curves of 5 tumor markers and age, and the corresponding AUC values are presented in Table 2. CEA has the highest AUC value with significant margin in comparison with others. AFP and CA125 have AUC values quite close to 0.5, so they can be viewed as having nearly no individual diagnostic value. Age, CA50, and CA199 have AUC values ranging from 0.57 to 0.62, showing weak individual dependency on tumor type, but their interaction effect can be further explored by multivariate diagnostic value analysis.

As shown in Table 3, age is ranked second on Relief measurements while it shows nearly no individual diagnostic value with AUC close to 0.5. Therefore, it is reasonable to assume that age may have interaction effect with other tumor markers on tumor type. The scatter plot of CEA vs. age is shown in Figure 3 for further investigation. A red vertical line and blue horizontal line are manually added to delineate their joint effect

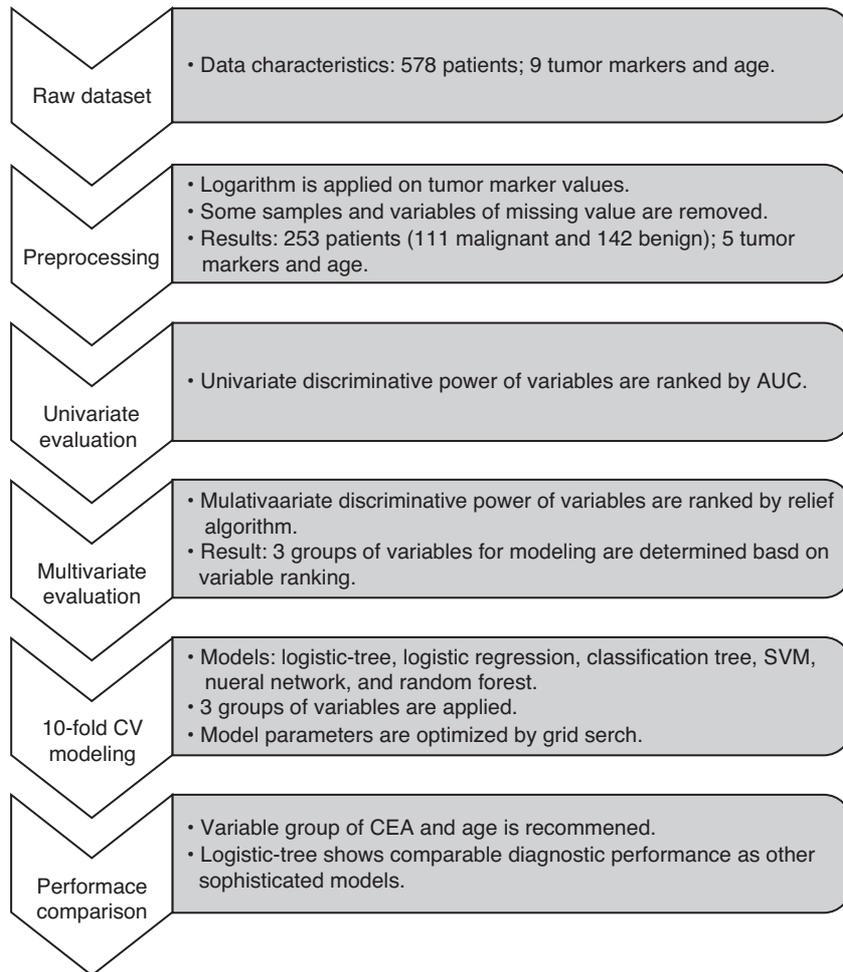


Figure 1. Flow chart of experimental procedure.

on tumor type. Evidently, the red vertical line which denotes a cutoff value of around 1.0 on CEA distinguishes two tumor types the best. In the section to the left of the red vertical line, we find the blue horizontal line reflecting the cutoff point of around 55 on age best distinguishes two tumor types. Without considering CEA, age indeed does not exhibit much discriminative power. Only when CEA values are less than 1.0 does age play a role in diagnosis. This two-line separation mechanism results in one mixed section and two other sections where high prediction accuracy can be achieved. In cancer diagnosis, the mixed section deserves more detailed clinical analysis and its final diagnosis can be made by referring to other factors.

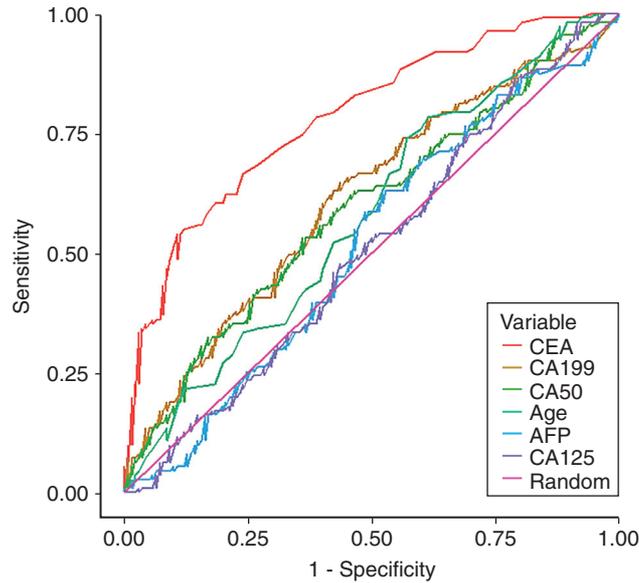


Figure 2. ROC curves of 5 tumor markers and age.

Table 2. AUC for five tumor markers and age

Variables names	CEA	CA199	CA50	Age	AFP	CA125
AUC	0.7844	0.6135	0.5931	0.5785	0.5184	0.5092

Table 3. Relief measurements for five tumor markers and age

Variable names	CEA	Age	CA199	AFP	CA50	CA125
Relief measurement	0.09074	0.02252	0.01225	-0.00094	-0.00464	-0.01406

3.3. Diagnostic Performance Comparison

Logistic-tree, logistic regression model, classification tree, SVM, neural networks, and random forest are tested with the same dataset for comparison of diagnostic performance. To verify the effectiveness of Relief algorithm when working with diverse models, three variable sets are considered: 1) CEA and age, which are found to have interaction effect on tumor type and rank highest on Relief measurements, 2) CEA, age and CA199, whose Relief measurements are above 0, and 3) all independent variables. As all the models' outputs are continuous, AUC is chosen as the performance evaluation criterion. Classification accuracy is also provided using a cutoff point of 0.5. Throughout the experiment, the procedure of 5 repetitions of cross-validation is utilized to achieve unbiased assessment. In n -fold cross-validation, the whole dataset is split into n folds, each

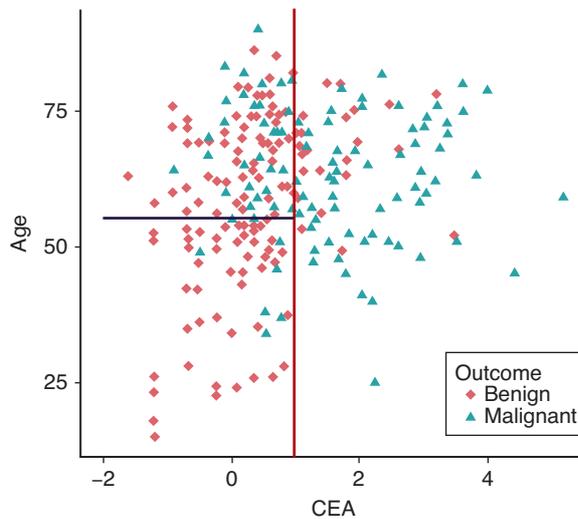


Figure 3. Scatter plot of CEA vs. age.

having roughly the same sample size. Every fold is subsequently used to test the model that is trained on the other $n-1$ folds. Thus, n -fold cross-validation will produce n models and n corresponding evaluation results. When n is set to be equal to the sample size N , it becomes the special version named leave-one-out cross-validation (LOOCV). LOOCV method has low bias but high variance. To accommodate the trade-off between reducing measurement variance and fully utilizing the samples for model training, we choose 10-folds cross-validation in the experiment. Additionally, parameters for SVM, neural networks, and random forest are tuned by means of grid search. For SVM, radial basis kernel is chosen, and the kernel coefficient γ and $cost$ (C-constant of the regularization term in the Lagrange formulation) are tuned. The tuned parameters for neural networks are $size$ and $decay$ which are, respectively, number of units in the hidden layer and coefficient for weight decay. For random forest, the tuned parameter $mtry$ represents the number of variables randomly sampled as candidates at each split. In a grid search, a vector of values for each parameter is manually defined, and the modeling is performed based on each combination of parameters values. The combination of parameters' values leading to the highest performance is considered as optimized parameters.

The summary of performance comparison in the procedure of 5 repetitions of 10-fold cross validation is presented in Table 4, and the result of each 10-fold cross validation can be found in the Appendix. Totally 3 scenarios of variable sets and 6 models are considered and they form 18 combinations. AUC averaged over 5 repetitions of 10-fold cross validation are obtained as the primary indicator of predictive performance.

Classification rates in terms of cutoff point 0.5 are also averaged over 5 repetitions of 10-fold cross validation for reference. The corresponding tuned parameters for SVM, neural networks, and random forest are also shown. For each scenario of variable set,

Table 4. Comparison of discriminative performance of different models and variable sets.

	Model	Logistic tree	Logistic regression	Classification tree	SVM		Neural networks		Random forest
	Optimized parameters				gamma	cost	size	decay	mtry
CEA	AUC	0.8038	0.7945	0.7611	0.032	8	6.8	0.006	1
Age	p-value of t-test (AUC)	*	0.3112	0.0134	0.3896		0.4751		0.0002
	Classification rate (%)	72.98	73.23	73.94	72.24		72.43		65.49
CEA	Optimized parameters				gamma	cost	size	decay	mtry
CEA	AUC	0.8075	0.8030	0.7590	0.002	11.2	15.2	0.1	1.4
Age	p-value of t-test (AUC)	0.4335	0.3324	0.0033	0.3759		*		0.0211
CA199	Classification rate (%)	72.99	72.16	73.87	72.52		71.14		68.23
CEA	Optimized parameters				gamma	cost	size	decay	mtry
CEA	AUC	0.7692	0.7872	0.7692	0.002	8	2.4	0.026	5.2
Age	p-value of t-test (AUC)	0.2957	0.3378	0.1815	*		0.1978		0.2522
CA199	Classification rate (%)	69.09	70.89	73.66	71.27		69.21		71.86
CA125									
CA50									
AFP									

Note: AUC and Classification rate are averaged over 5 repetitions of 10-fold cross validation. In every scenario of variable set, the maximum averaged AUC is highlighted in bold, and it is compared with AUC of other models based on the p -value of t-test. The optimized parameters averaged over 5 repetitions for SVM, Neural network, and Random forest are also shown.

the significance of the superiority of the best model (with the highest mean of AUC) over the others is indicated by the p -values of t-test (the p -value for the best model itself is denoted by *). Logistic-tree, neural networks, and SVM achieve the highest AUC means for the 3 scenarios of variable sets respectively (highlighted in bold in Table 4). From the p -value of t-test, one can see that these three models perform similarly well across all the 3 scenarios. For instance, in the case of CEA and age, logistic-tree achieves the highest performance, but the p -values for SVM, neural networks, and logistic regression are around 0.4. It shows that logistic-tree is not performing significantly better than these three models in this case. In terms of classification rate,

Table 5. Comparison of sensitivity and specificity (cutoff Point of 0.5) on training and test data of 10-fold cross validation (average over 5 repetitions).

		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Average	AUC
Training	Spec	0.89	0.87	0.82	0.89	0.84	0.89	0.85	0.88	0.89	0.90	0.87	0.8109
	Sens	0.57	0.53	0.56	0.53	0.60	0.56	0.58	0.54	0.55	0.55	0.56	
Test	Spec	0.88	0.91	0.89	0.73	0.87	0.86	0.68	0.82	0.82	0.82	0.83	0.8038
	Sens	0.43	0.75	0.86	0.63	0.47	0.43	0.63	0.80	0.50	0.55	0.60	

classification tree constantly outperforms other models when a cutoff point of 0.5 is used. Again, this is an indication that sophisticated models do not necessarily lead to better classification performance (classification tree is much simpler than SVM and neural networks).

In summary, the performance comparison demonstrated that, either for criterion of AUC or classification rate, logistic-tree shows competitive diagnostic performance compared with other parameter-optimized sophisticated models in the case of discriminative variables being selected by Relief algorithm. Considering that the AUC mean achieved by applying logistic-tree on the variable set of CEA and age is comparable to other highest AUC means, we recommend using the variable set of CEA and age for this application of colorectal cancer diagnosis. Table 5 shows the sensitivity and specificity on training and test data (10-fold cross validation) achieved using logistic-tree with CEA and age. It can be seen that the model performance is quite stable across different folds.

4. DISCUSSION

In the diagnostic value evaluation, Relief successfully identifies the interaction effects of CEA and age on colorectal cancer diagnosis. It further reveals how CEA and age are interacting in differentiating two tumor types as illustrated in Figure 3. As logistic-tree achieves highest AUC for the variable set of CEA and age, its prediction process in terms of tree structure is investigated. For the variable set of CEA and age, 8 out of 10 logistic-trees built and pruned in 10-fold cross-validation procedure have the same tree structure as that presented in Figure 4, with slightly different splitting values on the nodes. The pair of numbers in each terminal node represents the proportions of patient tumors labeled as benign and malignant. The classification mechanism of the tree structure as shown in Figure 4 is consistent with the partition mechanism illustrated by Figure 3. As stated before, Relief has the potential of exploring the interaction effect of multiple independent variables on the response variable, which makes the informative variables recommended by Relief algorithm especially suited for models with classification mechanism of a tree structure. As expected, neural networks, SVM, and random forest which are renowned for discovering complex dependency structure are not superior to logistic-tree with simple modeling mechanism in this case.

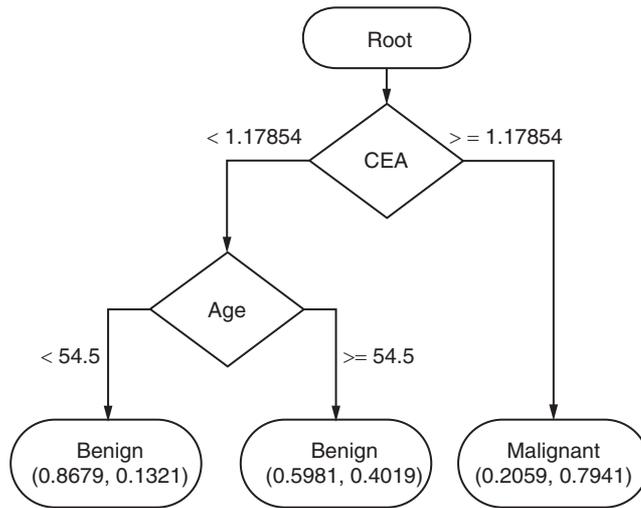


Figure 4. Structure of the logistic-tree.

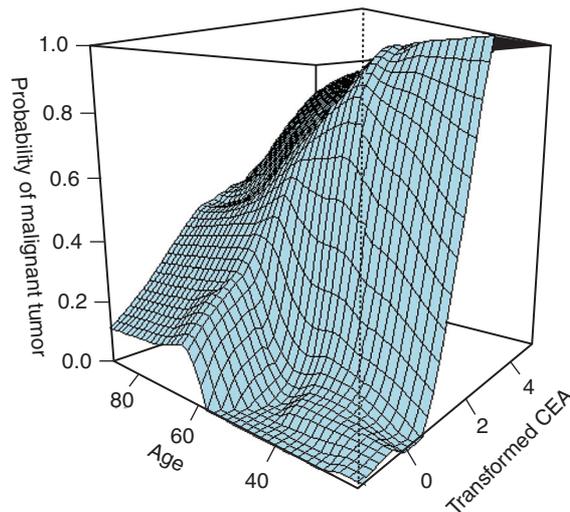


Figure 5. Response surface of the logistic-tree.

The response surface of logistic-tree in terms of the probability of having malignant tumor against CEA and age is presented in Figure 5. It can be seen that CEA is highly indicative of the status of colorectal tumor. This is consistent with current knowledge, although CEA alone is not recommended to be used as a screen test for colorectal cancer [25]. As for age, when it is over approximately 55, even if the CEA value is low, there is a drastic increase of the probability of malignant tumor. This is also consistent

with the recommendation by the US Preventive Services Task Force [26] of starting regular screening of colorectal cancer at the age of 50.

5. CONCLUSION

Evaluation of the diagnostic values of tumor markers for decision support in the context of cancer diagnosis is not an ordinary data analysis problem. In this type of problem, multivariate dependency, interpretability of diagnosis procedure and issue of overfitting need to be carefully addressed. The drawbacks of some extant approaches and their unintended consequences are pointed out and illustrated, which makes us realize the need of evaluation criteria that are model-independent. The effectiveness of Relief algorithm for multivariate diagnostic value evaluation is validated both theoretically and experimentally. We also demonstrate the necessity of using simple model for clinical decision support due to high interpretability. Accordingly, logistic-tree is developed which combines the advantages of classification tree and logistic regression. Its diagnostic performance is also proved to be highly competitive with other complex models in the colorectal dataset application. This cancer diagnosis procedure is evaluated on only a colorectal cancer dataset; its effectiveness need to be further validated using more datasets of different cancers in the future.

ACKNOWLEDGEMENT

This research was supported in part by National Science Foundation of China (NSFC) projects (71090404, 71072026).

CONFLICT OF INTEREST

The authors indicated no potential conflicts of interest.

REFERENCES

- [1] Özkan H, Kaya M, Cengiz A. Comparison of tumor marker CA 242 with CA 19-9 and carcinoembryonic antigen (CEA) in pancreatic cancer. *Hepato-Gastroenterology*. 2003, 50:1669–74.
- [2] Moghimi M, Ghodosi I. Clinical efficacy of combined use of CEA-CA19-9 in patients with pancreatic malignancy. *Journal of Medical Sciences*. 2007, 7:625–9.
- [3] Ni XG, Bai XF, Mao YL, Shao YF, Wu JX, Shan Y, et al. The clinical value of serum CEA, CA19-9, and CA242 in the diagnosis and prognosis of pancreatic cancer. *European Journal of Surgical Oncology*. 2005, 31:164–9.
- [4] Sun Z, Fu X, Zhang L, Yang X, Liu F, Hu G. A Protein Chip System for Parallel Analysis of Multi-tumor Markers and its Application in Cancer Detection. *Anticancer Research*. 2004, 24:1159–65.
- [5] McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies. *Journal of Clinical Oncology*. 2005, 23:9067–72.
- [6] Marrelli D, Caruso S, Neri A, Pedrazzani C, Capuano L, Mazzei MA, et al. Clinical utility of serum tumor markers in the diagnosis of malignant intestinal occlusion. A prospective observational study. *International Journal of Biological Markers*. 2011, 26:58–64.
- [7] Qin XL, Wang ZR, Shi JS, Lu M, Wang L, He QR. Utility of serum CA19-9 in diagnosis of cholangiocarcinoma: In comparison with CEA. *World Journal of Gastroenterology*. 2004, 10:427–32.
- [8] Carpelan-Holmström M, Louhimo J, Stenman UH, Alfthan H, Järvinen H, Haglund C. Estimating the probability of cancer with several tumor markers in patients with colorectal disease. *Oncology*. 2004, 66:296–302.

- [9] Poon TCW, Chan ATC, Zee B, Ho SKW, Mok TSK, Leung TWT, et al. Application of classification tree and neural network algorithms to the identification of serological liver marker profiles for the diagnosis of hepatocellular carcinoma. *Oncology*. 2001, 61:275–83.
- [10] Song HJ, Kim YS, Choi KW, Kim JD, Park CY, Lee E, et al. Building Biomarker Combinations for Korean Ovarian Cancer Screening Using Statistics and Machine Learning. *Applied Mechanics and Materials*. 2012, 145:229–33.
- [11] Winkler SM, Affenzeller M, Jacak W, Stekel H. Identification of cancer diagnosis estimation models using evolutionary algorithms: a case study for breast cancer, melanoma, and cancer in the respiratory system. *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation: ACM*, 2011. p. 503–10.
- [12] Winkler* SM, Affenzeller M, Kronberger G, Kommenda M, Wagner S, Dorfer V, et al. On the use of estimated tumour marker classifications in tumour diagnosis prediction—a case study for breast cancer. *International Journal of Simulation and Process Modelling*. 2013, 8:29–41.
- [13] Bottaci L, Drew PJ, Hartley JE, Hadfield MB, Farouk R, Lee PWR, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet*. 1997, 350:469–72.
- [14] Naguib RNG, Robinson MC, Neal DE, Hamdy FC. Neural network analysis of combined conventional and experimental prognostic markers in prostate cancer: A pilot study. *British Journal of Cancer*. 1998, 78:246–50.
- [15] Wu Y, Wu Y, Wang J, Yan Z, Qu L, Xiang B, et al. An optimal tumor marker group-coupled artificial neural network for diagnosis of lung cancer. *Expert Systems with Applications*. 2011, 38:11329–34.
- [16] Flores J, Herrera E, Leal G, González M, Sánchez F, Rojas A, et al. Artificial neural network-based serum biomarkers analysis improves sensitivity in the diagnosis of lung cancer. V Latin American Congress on Biomedical Engineering CLAIB 2011 May 16-21, 2011, Habana, Cuba: Springer, 2013. p. 882–5.
- [17] Flores-Fernández JM, Herrera-López EJ, Sánchez-Llamas F, Rojas-Calvillo A, Cabrera-Galeana PA, Leal-Pacheco G, et al. Development of an optimized multi-biomarker panel for the detection of lung cancer based on principal component analysis and artificial neural network modeling. *Expert Systems with Applications*. 2012, 39:10851–6.
- [18] Casas Pina T, Tovar Zapata I, Bermejo López J, Latour Pérez J, Paricio PP, Martínez Hernández P. Tumor markers in lung cancer: Does the method of obtaining the cut-off point and reference population influence diagnostic yield? *Clinical Biochemistry*. 1999, 32:467–72.
- [19] Körner H, Søreide K, Stokkeland PJ, Søreide JA. Diagnostic accuracy of serum-carcinoembryonic antigen in recurrent colorectal cancer: A receiver operating characteristic curve analysis. *Annals of Surgical Oncology*. 2007, 14:417–23.
- [20] Weiss HL, Niwas S, Grizzle WE, Piyathilake C. Receiver operating characteristic (ROC) to determine cut-off points of biomarkers in lung cancer patients. *Disease Markers*. 2003, 19:273–8.
- [21] Song H-J, Yim J-K, Park C-Y, Kim Y-S, Kim J-D. Comparison of the Diagnostic Performances of CA125 and Multiple Marker Combinations for Early Diagnosis of Ovarian Cancer. *International Journal of Bio-Science & Bio-Technology*. 2013, 5.
- [22] Kira K, Rendell LA. A Practical Approach to Feature Selection. *Proceedings of the Ninth International Workshop on Machine Learning: Morgan Kaufmann Publishers Inc.*, 1992. p. 249–56.
- [23] Robnik-Šikonja M, Kononenko I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*. 2003, 53:23–69.
- [24] Shi J, Su Q, Zhang C, Huang G, Zhu Y. An intelligent decision support algorithm for diagnosis of colorectal cancer through serum tumor markers. *Computer Methods and Programs in Biomedicine*. 2010, 100:97–107.
- [25] Bast R.C. Jr., Ravdin P, Hayes DF, Bates S, Fritsche H, Jr., Jessup JM, et al. 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology. *Journal of Clinical Oncology*. 2001, 19:1865–78.
- [26] Calonge N, Petitti DB, DeWitt TG, Dietrich AJ, Gregory KD, Harris R, et al. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine*. 2008, 149:627-37.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

