Progressive Breast Cancer Diagnosis Model Based on Multi-classifier and Multi-modal Fusion

Jiyun Li, Chenxi Jia, and Chen Qian

***Abstract*—The clinical diagnosis of breast cancer in real life is a comprehensive process which needs to consider different sources of information and use different medical examination methods according to different stages of the disease. First, routine and more economical medical examination should be carried out according to the needs of the disease, and then more accurate but expensive examination should be carried out according to the condition. When the data is seriously missing while the required features are selected, it will seriously affect the accuracy of the traditional comprehensive diagnosis model. A large amount of data is missing due to partial inspections that have not been performed within a certain period of time. At this time, the accuracy of traditional model will be greatly reduced. In order to solve this problem, this paper proposes a progressive breast cancer diagnosis strategy using multi-criteria and multi-classifier fusion that realizes the development according to the course of disease and continuously supplements the examination information to achieve a progressive comprehensive diagnosis of breast cancer. The architecture also has good scalability, which can be extended to more types of classifiers and input information of different modes, so as to achieve multi-criteria and multi-source comprehensive decision. Compared with the traditional multi-source breast cancer comprehensive diagnosis strategy, the experimental results show that the progressive breast cancer comprehensive diagnosis strategy has better predictive performance and clinical practicability.**

***Index Terms*—Breast cancer, multi-classifier fusion, multi-modal fusion, progressive diagnosis.**

I. Introduction

Breast cancer has become the most common cancer and the leading cause of cancer death among women in China [1]. The accurate diagnosis of breast cancer involves multiple data, wide dimensions and strong heterogeneity of physical and chemical indicators. Due to the limited medical resources and the further pursuit of improving the accuracy of diagnosis and treatment, artificial intelligence has broad application prospects in the classification, diagnosis and prognosis prediction of breast cancer. In the actual clinical diagnosis of breast cancer, the patient is firstly given some economic and routine tests such as blood routine and mammary molybdenum target for preliminary screening. Secondly, according to the different conditions of the patient, the doctor decides to perform more accurate special tests such as blood tumor markers examination, breast MRI, etc. Finally, according to the preliminary examination, if the condition requires, more accurate traumatic examinations such as needle biopsy of the breast and immunohistochemistry will be performed [2]. Different types of clinical examinations have different data characteristics, evaluation standards and indicators with different weight. Therefore, when using artificial intelligence methods for breast cancer auxiliary diagnosis, individual models are suitable for different types of examinations. Whether these data with distinct characteristics should be treated independently or regarded as a complete problem is a key to modeling. From the perspective of clinical medicine, effective breast cancer diagnosis must be derived from multi-source data. Multi-modal features as input make it possible to combine features with classifiers. For example, whether a modal feature should be input into a classifier or all features should be input into one classifier further increases the burden and complexity of this problem. Since different classifiers may classify different information, people want to obtain a more reliable model by maximizing the use of this information instead of choosing the best information from the available classifiers. The lack of standardized assessment techniques for classifier performance, such as repeatability and clinical applicability also complicates the decision-making process. However, exploring an effective method to manage complicated clinical information and selecting an appropriate classifier for predictive modeling still requires continuous research and verification in the actual clinical environment. At the same time, it is a very realistic problem which is how to ensure the validity of models due to different inspection items under different time dimensions. This is the premise that the comprehensive diagnosis model of breast cancer has good universality and can be applied to clinical diagnosis.

Modal refers to the way in which something happened or experienced [3]. Each source or form of information can be regarded as a modal. If a research problem contains multiple modes, it is called multi-modal. A model to integrate multi-source and multi-modal clinical data for effective breast cancer diagnosis must involves multi-modal fusion. Multi-modal fusion methods can be divided into three categories: feature-based fusion, decision-based fusion, and hybrid fusion [4]. Feature-based fusion refers to the fusion of data in different modalities directly at the level of modal features [5]. Due to the difference in the amount of different modal data, feature distribution and representation methods, in order to better reduce data imbalances, reduce feature redundancy and mine data associations, a variety of optimization algorithms based on feature fusion have been generated. Bishop *et al.* [6] proposed controlling redundant features by adding sparse regularized silver to the objective

Manuscript received March 25, 2020; revised February 19, 2021. This work was supported by the Science and Technology Development Foundation of Shanghai (18511102703, 16JC1400802, 16JC1400803), the Special Fund of Shanghai Municipal Commission of Economy and Informatization (RX-RJJC-08-16-0483, 2017-RGZN-01004, XX-XXFZ-02-18-2666, XX-XXFZ-01-18-2604).

The authors are with the College of Computer Science and Technology, Donghua University, Shanghai, China (e-mail: jyli@dhu.edu.cn, 2181829@mail.dhu.edu.cn, chen.qian@dhu.edu.cn).

doi: 10.18178/ijmlc.2021.11.6.1066

function. Ngiam *et al*. [7] proposed a coupled modal deep autoencoder (DAE), which learns high-dimensional abstract features through deep neural networks. These features have better expression capabilities. In terms of clinical application, Viceconti *et al*. [8] proposed a multi-dimensional comprehensive medical information management and analysis system that comprehensive biomedical information such as biology, omics, physiology and including space, time to integrate these highly heterogeneous biological data. Decision-based fusion uses different models for different feature patterns and the combination of decision values uses a fusion mechanism, such as averaging, voting or using a learning model. It allows predictions without one or more eigenmodes or parallel data, but ignores low-order interactions between eigenmodes. For example, Emaminejad *et al*. [9] trained a naive Bayesian network classifier using eight age features and trained a multi-layer perceptron classifier using two genomic biomarkers to predict the risk of cancer recurrence that applied several fusion methods to combine the predicted risk scores generated by these two classifiers. Hybrid fusion is a strategy to try to use the advantages of the two methods in a common framework. For example, Shoshtari *et al*. [10] predicts the course of multiple sclerosis based on the fusion of information from myelinated water imaging (MWI), diffusion tensor imaging (DTI) and resting state functional magnetic resonance imaging (RSFMRI).

II. Model

A. Model Overview

In this study, a progressive breast cancer diagnosis model based on multi-modal and multi-classifier fusion is proposed. It is hoped that the comprehensive diagnosis of multimodal breast cancer will be more practical and effective in clinical practice. Feature classifier fusion is essentially a Multi-criteria Decision Making (MCDM) problem, which is used to deal with situations where a set of variable factors exists. MCDM attempts to find the best alternative (or assign weight) among a set of alternatives by considering certain criteria for alternatives. Generally speaking, the MCDM program first determines a set of standards through which all optional qualities are evaluated and a series of judgments are made. Then an aggregation process transforms the evaluation matrix into a vector to represent each alternative result [11]. For the current research of feature classifier, the prediction output of different classifiers can be used as an alternative scheme of MCDM.

Feature classifier fusion is essentially a decision-making process which involves quantifying the contribution of each alternative output with specific criteria to measure the quality of the output, such as accuracy, AUC, sensitivity and specificity. The whole scheme consists of five process parts: feature extraction, data preprocessing, feature selection, classifier-level fusion and modal-level fusion. The overall process is shown in Fig. 1. At the same time, in order to realize progressive diagnosis, the mechanism of missing modal information is added to make the application of the model more flexible and more practical. From the perspective of mode, this model integrates three kinds of data information: blood routine, blood tumor markers and immunohistochemistry, which can not only complement each other, but also facilitate the actual clinical situation of progressive diagnosis and treatment.

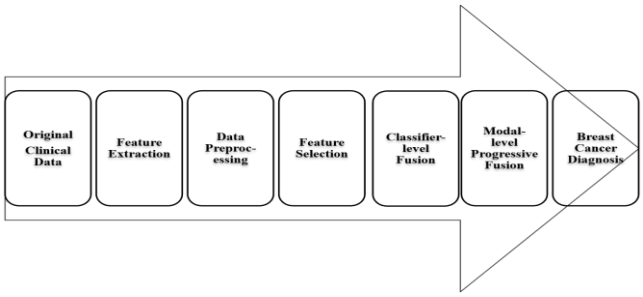

Fig. 1. Schematic diagram of the overall scheme.

The model integrates six classifier information: random forest (RF) [12], decision tree (DT) [13], k-nearest-neighbor (KNN) [14], support vector machine (SVM) [15], logistic regression (LG) [16], Long Short Term Memory networks (LSTM) [17]. RF is a kind of statistical learning theory. It uses bootstrap resampling method to extract multiple samples from the original samples, modeling the decision tree of each bootstrap sample, and then combining the prediction of multiple decision trees to get the final prediction results by voting. DT adopts the top-down recursive method, and its basic idea is to construct a tree with the fastest entropy decline by taking the information entropy as the measurement. At the leaf node, the entropy is 0. It has the advantages of readability and fast classification. KNN algorithm is a basic classification and regression algorithm. Its basic implementation uses a lazy learning process of majority voting, that is, it is actually a memory-based learning method. It simply counts the maximum number of labels in the K nodes closest to the target point to give the target point.

SVM is a two-class classification model. Its basic model is the linear classifier with the largest interval defined in the feature space. The learning algorithm of SVM is an optimization algorithm for solving convex quadratic programming. L R is a machine learning method for solving binary (0 or 1) problems, used to estimate the likelihood of something. Logistic regression assumes that the dependent variable y follows a Bernoulli distribution and linear regression assumes that the dependent variable y follows a Gaussian distribution. Therefore, there are many similarities with linear regression. LSTM is a special RNN that can learn long dependency. It was improved, popularized by many people, mainly to solve the problem of gradient disappearance and gradient explosion in the process of long sequence training.

In the fusion strategy part, this paper proposes a two-level fusion scheme, namely classifier-level fusion and modal-level fusion. The two-level fusion strategy can ensure the classification accuracy of single-mode data while ensuring the interpretability of the fusion between multi-mode data and the accuracy of classification results, reducing the coupling between different modes, and facilitating the increase and decrease of modal types. Different classifier information complements each other can improve the accuracy and avoid the selection of the best classifier at the same time. Modal level fusion has good universality in the face of multi-source data with different characteristics. The model framework is shown in Fig. 2.

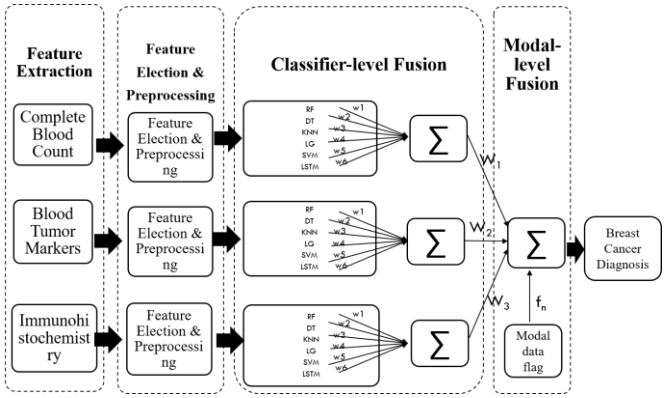

Fig. 2. Model frame diagram.

B. Feature Extraction and Selection

The data comes from real medical data of a medical institution in Shanghai. Breast cancer-related examination data are selected into three categories: complete blood count, blood tumor markers, and immunohistochemistry. These three categories are also typical indicators for different stages of the breast cancer diagnosis. At the same time, age and gender were also selected as supplementary characteristics. Since these different kinds of data are derived from real data, they are stored in different databases, and some of them come from free text. Therefore, a series of preprocessing work is required. Firstly, the data of breast cancer and healthy patients were screened from the overall database. Secondly, data splicing is performed, and the data in different databases are connected according to attributes such as the medical card number, the consultation serial number, and the report number to form a comprehensive attribute sub-table. Thirdly, data cleaning is performed to remove redundant, blank and other impurity data. Finally, we use the regular matching method to extract the physical and chemical inspection information from the free text in the sub table and build a new sub table.

Feature selection is a process which helps to select a small number of explanatory features for model construction. It has been proved to help reduce the chance of over fitting and make the model more consistent with clinical practice with interpretability [18]. In this paper, lasso and random forest with reduced mean accuracy were used for feature selection. The Complete Blood Count is selected eleven items: BA # BA% LY # LY% MCH MCHC MCV MPV NE # NE% RBC. The neutrophil percentage (NE%) of patients in the breast cancer group was (63.62 ±7.54) %, the number of basophils (BA#) was (0.04 ± 0.039) × 109 / L, the percentage of basophils ((BA%) is (0.50 ±0.317)%, the mean corpuscular volume (MCV) was (87.85 ±5.71) fL, the mean corpuscular hemoglobin (MCH) was (28.98 ± 2.76) pg, the mean corpuscular hemoglobin concentration (MCHC) was (331.2 ±17.46) The mean platelet volume (MPV) was (8.66 ±0.97) fL. The above index values in patients with benign breast group were (61.16 ±7.47) %, (0.03 ±0.011) ×109 / L, (0.43 ± 0.213) %, (85.46 ± 7.49) fL, (27.96 ± 3.51) pg, (326.0 ± 18.78) g / L, (8.43 ±0.83) fL. Comparing the two groups, the breast cancer group was higher than the benign breast group and the differences were statistically significant (P <0.05). The number of lymphocytes in patients with breast cancer (LY #) It is (1.85 ± 0.49) × 109 / L, the percentage of lymphocytes (LY%) is (27.46 ±6.78) %, and the number of red blood cells (RBC) is (4.59 ±0.42) ×1012 / L. The above indexes of patients in the benign breast group were (1.98 ± 0.49) ×109 / L, (29.48 ±6.83) %, and (4.70 ±0.39) ×1012 / L. Compared with the two groups, the breast cancer group was lower than the benign breast lesion group. The difference was statistically significant (P <0.05). It shows that the data is reasonable from the selected indicators in the importance of the indicators and the statistical differences of the indicators and the selected indicators have clinical diagnostic significance in combination with the verification of clinical related literatures [19].

TABLE I: DATA MAPPING TABLE

Indicator Name	Raw Data	Map Data
Sex	Female	1
	Male	0
ER PR Ki67 HER2	(-)	1
	(+)	2
	(+-++)	2.5
	(++)	3
	(++-+++)	3.5
	(+++)	4

TABLE II: THE SUMMARY OF EXPERIMENTAL DATA

Data Set \ Quantity	Breast Cancer	Healthy
Complete Blood Count	1931	2031
Blood Tumor Markers	1519	906
Immunohistochemistry	1013	501

TABLE III: THE SUMMARY OF DATA SET CHARACTERISTICS

Data Set	Feature Name
Complete Blood Count	BA# BA% LY# LY% MCH MCHC MCV MPV NE# NE% RBC
Blood Tumor Markers	CA125 CA153 CA199 CEA
Immunohistochemistry	ER PR Ki-67 HER2(CerbB-2)

Four blood tumor markers were selected: carbohydrate antigen (CA) 153, CAl25, CAl99, carcinoembryonic antigen (CEA). The CA125 of patients in the breast cancer group was (43.6 ±20.7) U / ml, the CA153 was (34.8 ±15.2) U / ml, the CA199 was (40.4 ±11.2) U / ml and the CEA was (6.1 ±2.5) ng / ml. The above indexes of benign breast cancer group were (15.4 ±7.9) U / ml, (8.2 ±2.9) U / ml, (17.8 ±5.9) U / ml and (2.4 ± 1.4) ng / ml. Compared between the two groups, the breast cancer group was higher than the benign breast group, and the differences were statistically significant (P <0.05). Consult the medical literature to prove that the selected indicators have clinical diagnostic significance [20]. Four immunohistochemical indicators were selected: Progesterone receptor (PR), estrogen receptor (ER), human epidermal growth factor receptor-2 (HER-2/CerbB-2), nuclcar-associated antigen Ki- 67 (Ki-67). HER-2 and CerbB-2 are different representations of the same indicator. Immunohistochemistry is divided into six grades from (-) to (+++). The specific mapping table is shown in the Table I. Referring to the medical literature, these indexes have the significance of medical clinical diagnosis [21].

Table II shows the data types and quantities of the data. Table III shows the characteristics of clinical indicators that selected for each modal.

C. Classifier Fusion Based on MCDM

According to the performance of training prediction, this model estimates the weight of the output scores of each classifier. Specifically, n classifiers are defined, each classifier is C_i (i = 1, 2, …, n), and the corresponding prediction probability is represented by P_i. The prediction performance of each classifier C_i in the training verification stage can be quantified by evaluation criteria M, such as accuracy, AUC, specificity and sensitivity. The evaluation matrix D is formed by $d_{i,j}$ (I = 1, 2, …, N, j = 1, 2, …, M) the number of rows N represents different classifiers and the number of columns M represents each evaluation indexes. The weight of each evaluation index is expressed as a_j (j = 1, 2, …, M). Here we simply set a_j to 1 / M.

Input an $N * C$ evaluation matrix, the evaluation index weight a_j, and the evaluation matrix is normalized as in

$$d' = \frac{d_{i,j}}{\sqrt{\sum_{k=1}^{N} d_{i,j}^2}}, i = 1,2,...,N,\ j = 1,2,...,M \cdot \quad (1)$$

Evaluation matrix multiplied by index weight as in

$$d''_{i,j} = d'_{i,j} \times a_j, i = 1,2...,N,\ j = 1,2,...,M \cdot \quad (2)$$

Define the maximum and minimum values for each evaluation index as in

$$\overline{\mathrm{EC}}^{\max} = \left\{ d_{i,j}^{'} \middle| i = 1,2.....N) \middle| \ j = 1,...2....M \right\} = \left\{ ec_j^{\max} \right\} \middle| j = 1,2.....,M \} \quad (3)$$
$$\overline{\mathrm{EC}}^{\min} = \left\{ d_{i,j}^{'} \middle| i = 1,2.....N) \middle| \ j = 1,...2....M \right\} = \left\{ ec_j^{\min} \right\} \middle| j = 1,2.....,M \}$$

Calculate the distance from each evaluation object to the maximum and minimum values as in

$$\mathrm{D}is_{i,\max} = \sqrt{\sum_{j=1}^{M} (d_{i,j} - ec_j^{\max})^2}, i = 1,2...,N$$
$$\mathrm{D}is_{i,\min} = \sqrt{\sum_{j=1}^{M} (d_{i,j} - ec_j^{\min})^2}, i = 1,2...,N \quad (4)$$

Assign weights to each evaluation object as in

$$w'_i = Dis_{i,\min} / (Dis_{i,\min} + Dis_{i,\max}) \cdot \quad (5)$$

Weight normalization as in

$$w_i = \frac{w'_i}{\sum_{k=1}^{N} w_k}, i = 1,2,...,N \cdot \quad (6)$$

D. Classifier-Level Fusion

The purpose of classifier-level fusion is to have better classification results for different data, while avoiding the problem of optimal classifier selection. Classifier fusion refers to the weighted summation of each classifier according to the evaluation matrix. In a single mode, a k-fold cross-validation is applied to the training data set to obtain the prediction ability $p_{i,j}$ (i=1, 2, …, N, j=1, 2, …, k) of each classifier.

The prediction ability is calculated from the evaluation matrix to calculate its accuracy, area under the curve (AUC), specificity, sensitivity and then construct the corresponding evaluation matrix $D^k_{N,M}$. Since the weight of each evaluation index is set equal here, $p_{i,j}$ is the average of the evaluation matrix column direction. The prediction probability of each classifier is the weighted average of each classifier, as in

$$P_j = \sum_{j=1}^{N} p_{i,j*} W_m \cdot \quad (7)$$

E. Modal-level Progressive Fusion

The purpose of mode level fusion is to consider the information of different modes and various factors. This is more in line with the real-world requirements for breast cancer diagnosis. Its core idea is to give different weights to different modes, which means that different modes have different influence on the final results, and the results of each mode classification are weighted and averaged. There are S modalities, and the modal data flags are f_n (n = 1,2, ..., S). Using classifier fusion P_j, similar to classifier fusion, Ws can be obtained according to Section B.

In order to realize the gradual mechanism, when calculating the modal prediction probability, only the modal with f_n = 1 is considered. When f_n = 1, $\underline{W_j}$ = 0. Calculate the final output probability:

$$P = (\sum_{j}^{S} P_j * W_j) / \sum_{j}^{S} W_j \cdot \quad (8)$$

F. Model Training and Testing Process

After the above theoretical derivation, according to the above ideas to build the model. The idea of model training is to modify the performance of the weight parameter optimization model after the initial model has been iterated for several rounds, and then stop the iteration after reaching the set optimization goal to get the final model with better performance.

- The training process is as follows:

1) Extract the characteristics of breast cancer clinical indicators from real world clinical databases including demographics, blood routine, blood tumor markers, and immunohistochemistry. Each category contains data on breast cancer patients and healthy people.
2) Feature extraction using LASSO and random forest with reduced average accuracy.
3) The selected features in each mode are input into six classifiers: RF, DT, KNN, SVM, LG and LSTM.
4) Calculate the weight w_m (m = 1, 2, ..., 6) of each classifier according to the evaluation matrix.
5) According to the results of the fourth step of different modalities, the modal weights W_n (n = 1, 2, 3) were calculated according to the evaluation matrix.

The trained model is obtained through the above process. In order to test the model performance and later practical application, the following describes the test process

- The test process is as follows:

1) Extract the clinical information of breast cancer from the patients to be tested.
2) Select the features selected during training and fill in the corresponding modal data flag f_n (n = 1, 2, 3). If some checks have not been performed, the modal is set to 0, otherwise it is set to 1.
3) Input the selected features into the trained RF, DT, KNN, SVM, LG, LSTM and perform classifier decision fusion based on the w_m (m = 1, 2, ..., 6) obtained through training

4) Perform modal-level decision fusion based on modal weights W_n (n = 1, 2, 3), modal data flags f_n and then obtain prediction results.

III. Experiment

A. Comparing Classifiers in Single Modal

The experimental results are as shown in the Table IV.

TABLE IV: The Comparison Results of Single Classifier and Classifier Fusion in Single Modal

Accuracy / Classifier	Complete Blood Count	Blood Tumor Markers	Immunohisto-chemistry
RF	0.7959	0.8223	0.8392
DT	0.7577	0.7754	0.8233
KNN	0.7627	0.8292	0.8051
SVM	0.7361	0.7937	0.8418
LG	0.7253	0.8128	0.8137
LSTM	0.7926	0.8061	0.8225
Classifier Fusion	0.8161	0.8325	0.8639

To verify the validity of classifier-level fusion, the accuracy lists of RF, DT, KNN, SVM, LG, LSTM in complete blood count, blood tumor markers and immunohistochemistry are validated. The accuracy after classifier fusion is also listed in Table III. The optimal blood routine single classifier is RF that accuracy is 79.59%, the optimal model of blood tumor marker single classifier is KNN that accuracy is 82.92%, and the optimal model of immunohistochemical single classifier is SVM that accuracy is 84.18%. It can be seen that different classifiers are suitable for data with different data characteristics. In order to obtain better accuracy, different classifiers need to be selected according to different data. At the same time, experiments also prove that the performance of the classifier fusion method based on multi-criteria decision fusion is better than the strategy using a single classifier.

TABLE V: The Comparison of Different Fusion Methods

Evaluation / Fusion	Accuracy	AUC	Specificity	Sensitivity
MV	0.8725	0.9186	0.8909	0.8381
WAF	0.8916	0.9302	0.9150	0.8726
Proposed Model	0.9137	0.9346	0.9225	0.8862

B. Comparing Classifiers in Single Modal

The experimental results are shown in Table V. To verify the effectiveness of the different fusion methods, the majority voting (MV) [22], the weighted average fusion (WAF) [23] and the proposed model are compared and analyzed. Majority voting is a simple method of decision level fusion. This method is similar to the voting process in the election process. It uses a single classifier to output categories for a given test sample, and then synthesizes the classification results of multiple classifiers to divide the final categories of the test sample into a class with the same decision of most classifiers. Weighted average fusion is a conventional fusion method. Its idea is that different data or classifiers have different importance, so different classifiers are given different weights and the classification results of different classifiers are weighted average to get the final classification results. Pre-processed Complete Blood Count, blood tumor markers and immunohistochemical data were simultaneously input into the model. The results show that the proposed model is superior to the other two fusion methods in terms of accuracy, AUC, specificity and sensitivity. It is proved that the fusion strategy used in this paper is effective.

C. Comparison with Non-progressive Model

The experimental results are shown in Table VI. In order to prove that the progressive model is more consistent with the actual clinical diagnosis, the progressive model and the non-progressive model were compared.

TABLE VI: The Comparison of Different Fusion Methods

Evaluation / Test Conditions	Accuracy	AUC	Specificity	Sensitivity
Non-progressive				
Remove Complete Blood Count	0.7052	0.7783	0.7453	0.6461
Remove Blood Tumor Markers	0.6976	0.7901	0.7356	0.6250
Remove Immunohistochemistry	0.6519	0.7588	0.7072	0.6122
Progressive				
Remove Complete Blood Count	0.8829	0.9236	0.8957	0.8634
Remove Blood Tumor Markers	0.8760	0.9026	0.8711	0.8527
Remove Immunohistochemistry	0.8406	0.8986	0.8516	0.8126

Both experiments use the model proposed in this paper, and their difference is only in whether to add a progressive mechanism. Model performance was compared when blood routine, blood tumor markers and immunohistochemical data were removed. Compare the performance of progressive and non-progressive models from four aspects: Accuracy, AUC, Specificity, and Sensitivity. The experimental results show that when the data is missing, the accuracy of the non-progressive model is greatly reduced and the performance of the model using the progressive mechanism is slightly better than the model without the progressive mechanism. Therefore, it is necessary to adopt a progressive model, which is also in line with the actual situation.

IV. Conclusion

In this paper, we discuss the integration of multi-source and multi-modal data in the clinical diagnosis and treatment of breast cancer due to the lack of data due to different stages of diagnosis. Firstly, a two-step fusion strategy is proposed for multi-source and multi-modal data fusion. The first step is to fuse the classifiers which improves the generality of the model and avoids the problem of optimal classifier selection. The second step of multi-modal fusion solves the problem of information fusion of different data sources with different data characteristics, and has good scalability to easily add new data types to the original model. Secondly, according to different actual clinical diagnosis and treatment due to the lack of examination items due to different stages of the disease, the accuracy of the model is seriously affected by the lack of data. Modal data flags have been added to indicate

missing data and adjusted accordingly in the final modal level fusion.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Prof. Li help provide experimental raw medical data. Prof. Li and Dr. Qian guide ideas and solve difficult problems encountered during the experiments. Chenxi Jia conducts experiments and writes this paper under the guidance of prof. Li and Dr. Chen.

References

[1] W. Zhe *et al*., "Research on prognostic factors of breast cancer based on machine learning methods," *China Digital Medicine*, vol. 14, no. 1, pp. 18-20 + 46, 2019.

[2] J. B. Li and Z. F. Jiang, "Interpretation of the updated points of CSCO BC guidelines in 2019," *Chinese Journal of Oncology*, vol. 3, 2019.

[3] T. Baltrusaitis, C. Ahuja, and L. P. Morency, "Multi-modal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1-1, 2018.

[4] Q. He *et al*., "Feasibility study of a multi-criteria decision-making based hierarchical model for multi-modality feature and multi-classifier fusion: Applications in medical prognosis prediction," *Information Fusion*, 2020, vol. 55, pp. 207-219.

[5] Z. Liang, *Research on Multi-modal Data Fusion Algorithm*, 2018.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics*, Springer-Verlag New York, Inc. 2006.

[7] J. Ngiam, A. Khosla, M. Kim *et al*., "Multimodal deep learning," in *Proc. the 28th International Conference on Machine Learning*, Bellevue: ACM, 2011, pp. 689-696.

[8] M. Viceconti, G. Clapworthy, and D. Testi *et al*., "Multimodal fusion of biomedical data at different temporal and dimensional scales," *Computer Methods and Programs in Biomedicine*, 2011, vol. 102, no. 3, pp. 227-237.

[9] N. Emaminejad, W. Qian, Y. Guan, M. Tan, Y. Qiu, H. Liu, and B. Zheng, "Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients," 2015.

[10] S. A. Sheikhi *et al*., "A Multimodal data fusion approach efficiently predicts disease duration in multiple sclerosis," in *Proc. 2016 International Joint Conference on Neural Networks (IJCNN)*, 2016.

[11] E. Cables, M. T. Lamata, and J. L. Verdegay, "RIM-reference ideal method in multicriteria decision making," *Inf. Sciences*, pp. 337-338, pp. 1–10, 2016.

[12] S. V. Random, "Forest: A classification and regression tool for compound classification and QSAR modeling," *Journal of Chemical Information & Computer Sciences*, vol. 43, 2003.

[13] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems Man & Cybernetics*, vol. 21, no. 3, pp. 660-674.

[14] L E. Peterson, *K-nearest Neighbor*, vol. 4, no. 2, p. 1883, 2009.

[15] C. Saunders, M. O. Stitson, J. Weston *et al*., *Support Vector Machine*, vol. 1, no. 4, pp. 1-28, 2002.

[16] X. Xin and E. Frank, "Logistic regression and boosting for labeled bags of instances," in *Proc. 8th Pacific/Asia Conference on Advances in Knowledge Discovery and Data*, CiteSeer, 2004.

[17] A. Suzuki, H. Fukushima, and T. Mukawa, "Upregulation of CREB-mediated transcription enhances both short- and long-term memory," *Journal of Neuroscience*, vol. 31, no. 24, pp. 8786-8802.

[18] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[19] L. L. Zhang, L. Yun, W. B. Dua *et al*., "Changes in peripheral blood indexes and diagnostic value of tumor markers in patients with breast cancer," *Chinese Medicine*, vol. 3, pp. 421-425.

[20] Z. Y. Qiu, W. L. Xu, Y. Yan *et al*., "Diagnostic value of protein chip technology in detecting tumor markers for breast cancer," *Journal of Practical Cancer*, vol. 24, no. 5, pp. 486-490, 2009.

[21] Y. F. Guo, H. Y. Zhou, and F. Y. Han, "Analysis of molecular biological indexes and clinicopathological characteristics of breast cancer," *Chinese Medical Guide*, vol. 14, no. 32, pp. 162-163, 2016.

[22] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information Fusion*, vol. 6, no. 1, pp. 63-81.

[23] Z. G. Liu and P. Q, J. Dezert *et al*., "Combination of classifiers with optimal weight based on evidential reasoning," *IEEE Trans Fuzzy Sys*, vol. 26, no. 3, pp. 1217-1230, 2018.

Jiyun Li received the MS and PhD degrees from Donghua University in 1996 and 2003, respectively. Currently she is a professor of computer science and technology in Donghua University. Her research interests include data engineering, machine learning, etc.

She was a research fellow at the Hong Kong Polytechnic University from 2003 to 2004. She was a visiting scholar in the Department of Psychology and Brain Sciences at Indiana University, Burmingtonv from 2008 to 200

Prof. Li is a member of the committee on Distributed and Parallel Computing of the Chinese Computer Society, a member of the China Artificial Intelligence and Artificial Psychology committee, and a member of the Chinese Society of Graphic and Imaging.

Chenxi Jia was born in Hohhot, Inner Mongolia, China, in 1995. She received the BEng degree from Nanjing University of Posts and Telecommunications in 2016. She is currently studying for a master's degree in software engineering at Donghua University. Her main research interests are artificial intelligence, deep learning, intelligent medical, etc.

She worked as a mobile driver software developer at Huaqin Communications from 2016 to 2017. She worked as a system software development engineer at Ambarella in 2018.

Chen Qian was born in Nanjing, China, in 1988. He received the B.E degree from Shanghai Jiao Tong University in 2010, the MSc and Ph.D. degrees from The University of Manchester in 2014 and 2019, respectively. His current research interests include artificial intelligence, blockchain, function-as-a-service, software product line engineering, component-based development and software architecture modeling.

From 2010 to 2013, he devoted himself to industry as a software and website developer, and from 2014 to 2019, he was a member of the Component-based Software Development Research Group in The University of Manchester. After the doctoral graduation, he joined the School of Computer Science and Technology in Donghua University and became a lecturer to date.