Clinical Medicine Research
2024; 13(1): 6-12
http://www.sciencepublishinggroup.com/j/cmr
doi: 10.11648/j.cmr.20241301.12
ISSN: 2326-9049 (Print); ISSN: 2326-9057 (Online)

Combining Multiple Tumor Markers to Construct a Clinical Prediction Model for Breast Cancer

Zebin Liu, Lipeng Lin, Guosheng Zhu, Liying Qiu*

Laboratory Department, Shenzhen Futian District Maternity and Child Healthcare Hospital, Shenzhen, China

Email address:
qiuliying_ftbjy@163.com (Liying Qiu)

*Corresponding author

To cite this article:
Zebin Liu, Lipeng Lin, Guosheng Zhu, Liying Qiu. (2024). Combining Multiple Tumor Markers to Construct a Clinical Prediction Model for Breast Cancer. *Clinical Medicine Research*, *13*(1), 6-12. https://doi.org/10.11648/j.cmr.20241301.12

Received: December 17, 2023; **Accepted**: December 28, 2023; **Published**: January 8, 2024

Abstract: *Objective:* This study aims to explore early diagnostic methods for breast cancer, enhancing diagnostic sensitivity and specificity. *Methods:* A retrospective analysis was conducted, collecting data from 176 patients with breast mass treated at our institution. This cohort included 73 breast cancer patients and 103 patients with breast mass. Levels of four biomarkers - Ferritin (FER), Carcinoembryonic Antigen (CEA), Cancer Antigen 153 (CA153), and Cytokeratin 19 Fragment (CY211) - were measured. A clinical prediction model was constructed using multivariate Logistic regression, and its performance was evaluated through the Concordance Index (C-index), Receiver Operating Characteristic (ROC) curves, and calibration curves. Additionally, machine learning algorithms (K-Nearest Neighbors and Random Forest) were employed for validation, assessing the model's predictive capability through ROC curves, Precision−Recall Curves, True Negative Rate (TNR), True Positive Rate (TPR), F-measure, and 5-fold cross-validation. *Results:* Levels of FER, CEA, CA153, and CY211 in the breast cancer group were significantly higher than those in the breast mass group. The constructed prediction model demonstrated high predictive capability, with a C-index of 0.813 and an area under the ROC curve of 0.812. Clinical decision curve analysis indicated maximal net clinical benefit with a threshold probability range of 0.1 to 0.95, without compromising other patients' benefits. Machine learning validation confirmed the model's high accuracy and reliability. *Conclusion:* The clinical prediction model, utilizing a combination of FER, CEA, CA153, and CY211, effectively differentiates between benign breast mass and breast cancer, offering a novel perspective for early diagnosis.

Keywords: Breast Cancer, Machine Learning, FER, CEA, CA153, CY211, Clinical Prediction Model

1. Introduction

Breast cancer is one of the most common cancers among women worldwide and a leading cause of cancer-related deaths among females [1], with incidence and mortality rates varying significantly across different regions [2]. Advances in screening techniques and treatment methods have improved early diagnosis and treatment outcomes for breast cancer in recent years. However, the complexity and heterogeneity of breast cancer necessitate more precise diagnostic methods to guide treatment decisions.

Currently, breast cancer diagnosis primarily relies on imaging studies, clinical examination, and histopathological assessment. The application of biomarkers in recent years has provided new possibilities for early diagnosis and treatment of breast cancer [3-5]. Biomarkers, reflecting the biological characteristics of tumors, are crucial for prognosis assessment and treatment decision-making.

Ferritin (FER), Carcinoembryonic Antigen (CEA), Cancer Antigen 153 (CA153), and Cytokeratin 19 Fragment (CY211) are commonly used biomarkers in breast cancer diagnostics. Ferritin, a protein involved in iron storage and regulation, is associated with tumor development in certain cancers. CEA, widely used in tumor marker detection, especially in colorectal cancer, and CA153, commonly found in the serum of breast cancer patients, are closely related to disease recurrence and prognosis. CY211, a cytokeratin fragment, is associated with tumor aggressiveness and prognosis in breast cancer tissues.

Utilizing these biomarkers for combined diagnostic purposes could enhance the sensitivity and specificity of detecting breast cancer in patients with breast mass, providing more precise diagnostic information for clinical use. This

study collected FER, CEA, CA153, and CY211 data from patients with breast tumors treated at our hospital to construct a clinical prediction model, aiming to identify breast cancer in patients with breast tumors at an early stage, thereby facilitating early treatment and improving prognosis.

2. Materials and Methods

2.1. General Information

This retrospective study was approved by the hospital's ethics committee. Inclusion criteria: Patients with breast mass requiring further examination to clarify the nature of the tumor. Exclusion criteria: (1) Patients with other tumors or inflammatory diseases; (2) Pregnant or lactating women; (3) Patients with incomplete data.

From January 2018 to June 2023, 176 patients diagnosed with breast mass at our hospital were included in the study. Among them, 73 were diagnosed with breast cancer through histopathology, and 103 with benign breast mass. The age range in the breast cancer group was 32 to 85 years, with an average age of 51.31 ± 11.34 years; the benign breast mass group ranged from 18 to 71 years, with an average age of 39.20 ± 11.43 years. There was no statistical significance in age between the two groups (P>0.05).

2.2. Tumor Markers

Data for Ferritin (FER), Carcinoembryonic antigen (CEA), Cancer antigens 153 (CA153), and Cytokeratin 19 fragment antigen 21-1 (CY211) were collected from 176 breast mass patients. Reference values for FER were 4.63−204 ng/ml, for CEA 0−5 ng/ml, for CA153 0−31.3 U/ml, and for CY211 0−3.3 ng/ml.

2.3. Construction of Prediction Model

A predictive model was established using multivariate Logistic regression with the four tumor markers, visualized with nomograms. The predictive ability of the nomograms was differentiated using the C-index; the performance of the prediction model was evaluated by plotting ROC and calibration curves; clinical decision curves were plotted to assess the net benefit at different threshold probabilities to determine the clinical utility of the prediction model.

2.4. Machine Learning Validation Model

The model constructed in step 1.3 was validated using k-Nearest Neighbor (KNN) and Random Forest (RF) algorithms. The hyperparameter tuning was set to a random grid of 20. The model's predictive capacity was assessed through ROC (Receiver Operating Characteristic Curve), PR (Precision−Recall Curves), TNR, TPR, F-value, and 5-fold cross-validation. SHAP was used to evaluate the importance of each variable in the predictive model.

2.5. Statistical Methods

Data analysis was performed using SPSS 20.0 software. Data were presented as mean ± standard deviation. The t-test was used for comparisons between two groups, with P<0.05 considered statistically significant.

3. Results

3.1. Comparison of FER, CEA, CA153, CY211 Levels Between Benign Breast Mass Group and Breast Cancer Group

Results showed that the levels of FER, CEA, CA153, and CY211 were significantly higher in the breast cancer group compared to the breast mass group, with statistical significance (Table 1).

Table 1. *Comparison of Tumor Markers Between Groups.*

Group	FER	CEA	CA153	CY211
Breast mass Group	92.96±56.29	2.48±1.27	13.45±7.47	1.62±0.86
Breast Cancer Group	236.44±184.48	7.49±6.06	24.96±15.53	3.39±2.48
t	-7.43	-8.14	-6.54	-6.71
P	<0.01	<0.01	<0.01	<0.01

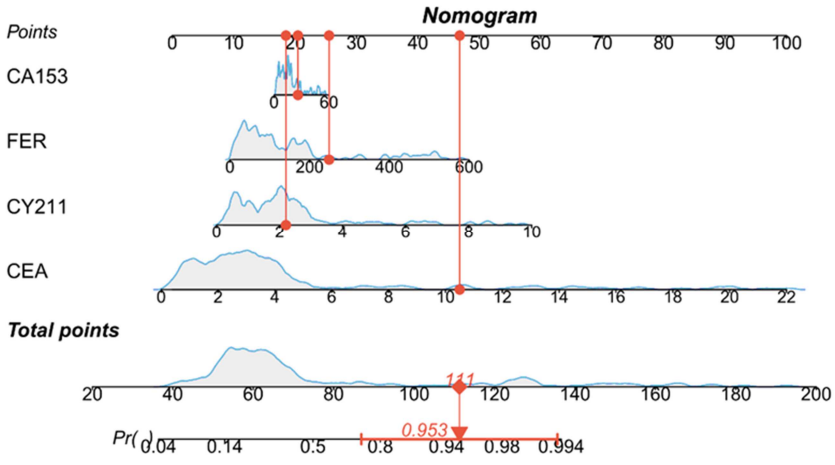

Figure 1. *Clinical Prediction Model Nomogram.*

3.2. Clinical Prediction Model

The four predictors were incorporated into a clinical prediction model using multivariable logistic regression, visualized through nomograms (Figure 1). The C-index of the model was 0.813. The ROC curve showed an area under the curve (AUC) of 0.812 (Figure 2A), and the calibration curve demonstrated good consistency (Figure 2B). The clinical decision curve indicated the highest net benefit at threshold probabilities ranging from 0.1 to 0.95 without compromising other patients' benefits (Figure 3).

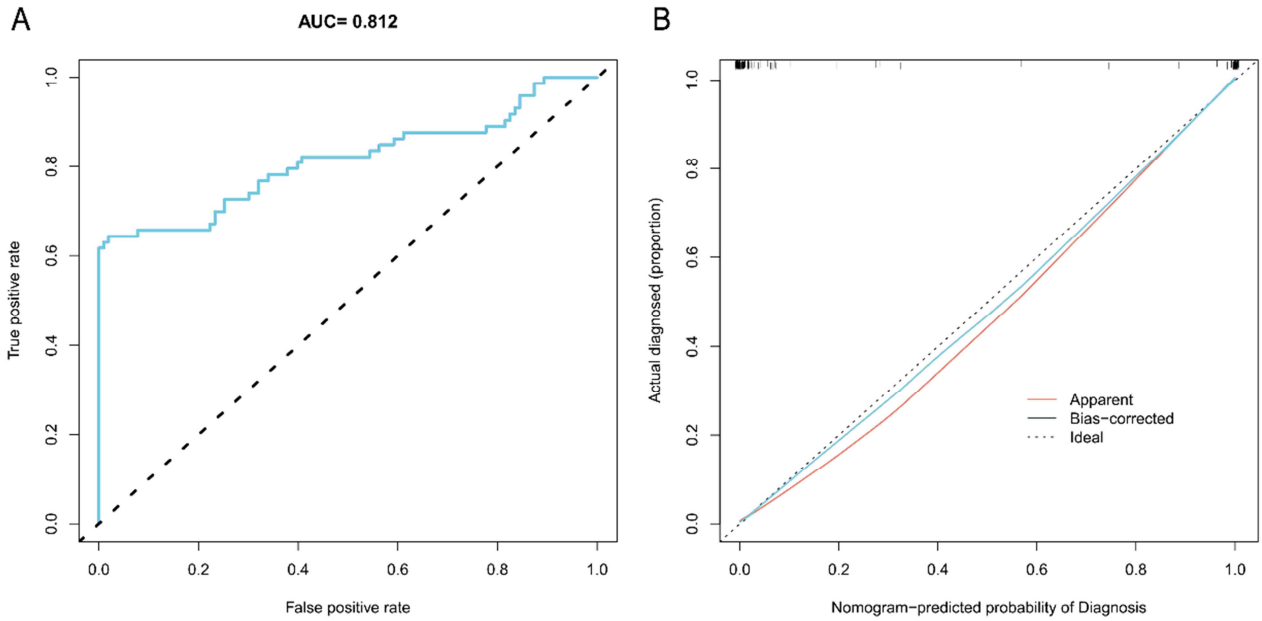

Figure 2. *Predictive Model ROC Curve and Calibration Curve (A: ROC Curve. B: Calibration Curve).*

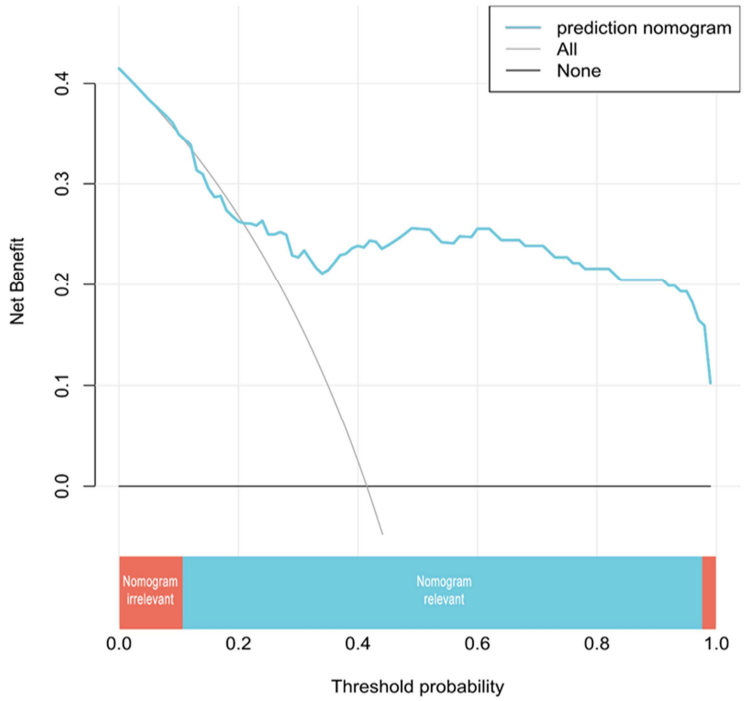

Figure 3. *Clinical Decision Curve.*

3.3. Machine Learning Validation Model

Results from KNN and RF showed that the area under the ROC curve was 0.9895 and 0.9840, respectively (Figure 4), and the area under the PR curve was 0.9877 and 0.9823, respectively (Figure 5). SHAP visualization showed that the importance of variables in the KNN algorithm was in the order of FER, CEA, CA153, CY211. In the RF algorithm, the order

was FER, CEA, CY211, CA153 (Figure 6), consistent with the clinical prediction model. The confusion matrix showed that the KNN prediction model had a TPR of 0.959, TNR of 0.980, accuracy of 0.72, and F1 value of 0.966; the RF prediction model had a TPR of 0.922, TNR of 0.979, accuracy of 0.954, and F1 value of 0.947 (Figure 7). 5-fold cross-validation showed that the ROC of both KNN and RF models was greater than 0.8 (Figure 8).

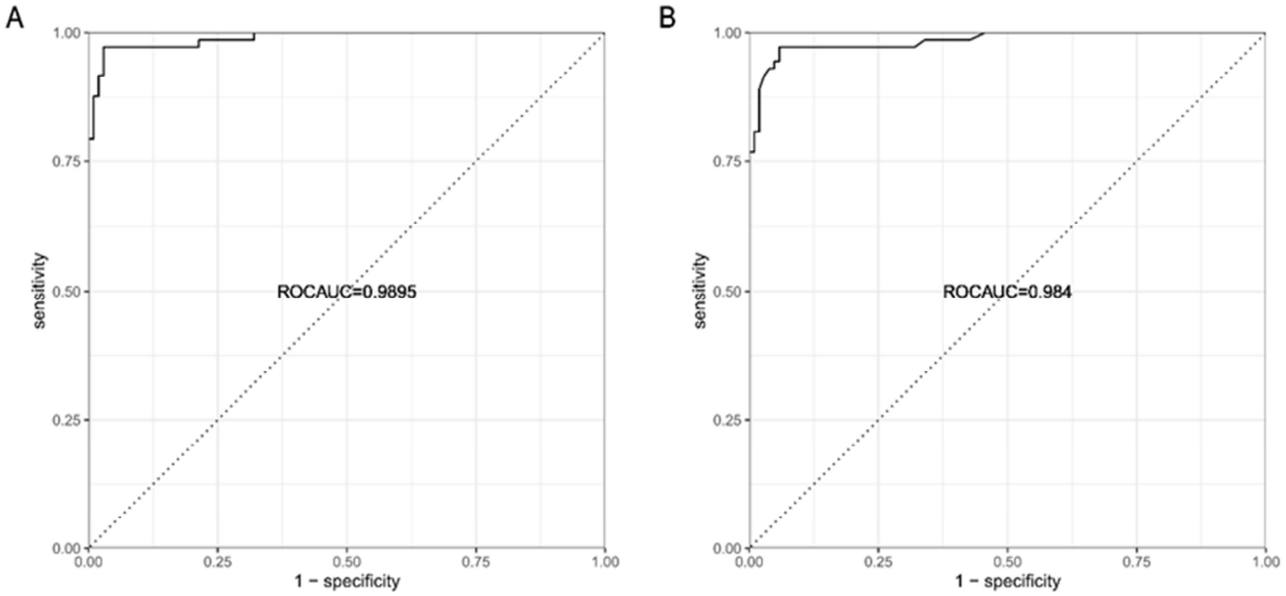

Figure 4. *ROC Curve (A. KNN B. RF).*

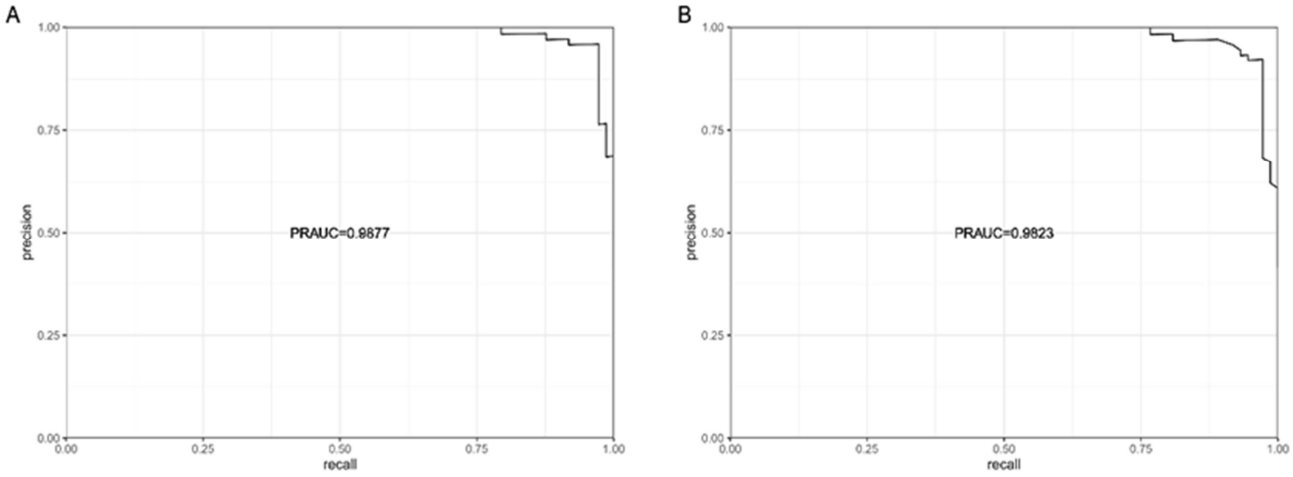

Figure 5. *PR Curve (A. KNN B. RF).*

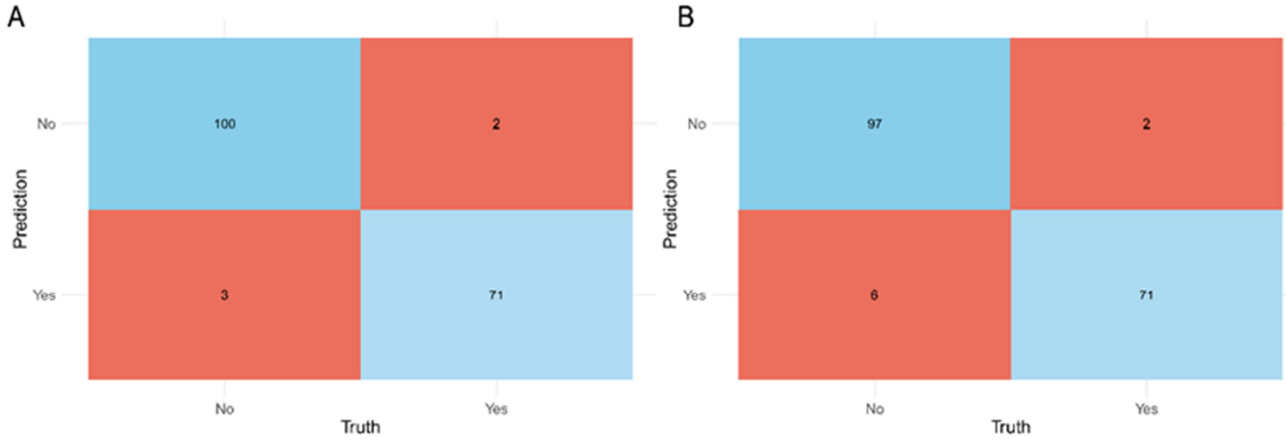

Figure 6. *SHAP Visualization (A. KNN B. RF).*

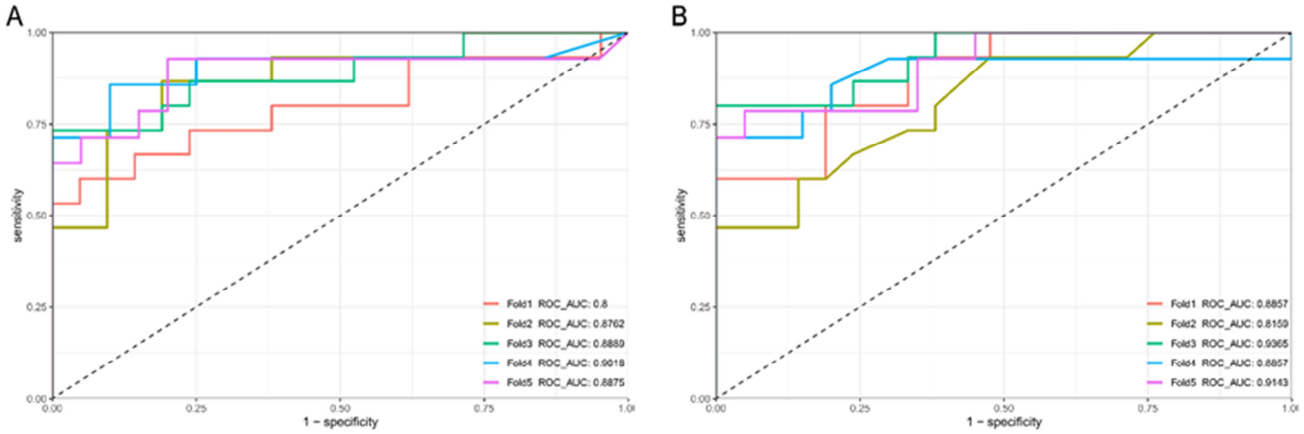

Figure 7. *Confusion Matrix (A. KNN B. RF).*

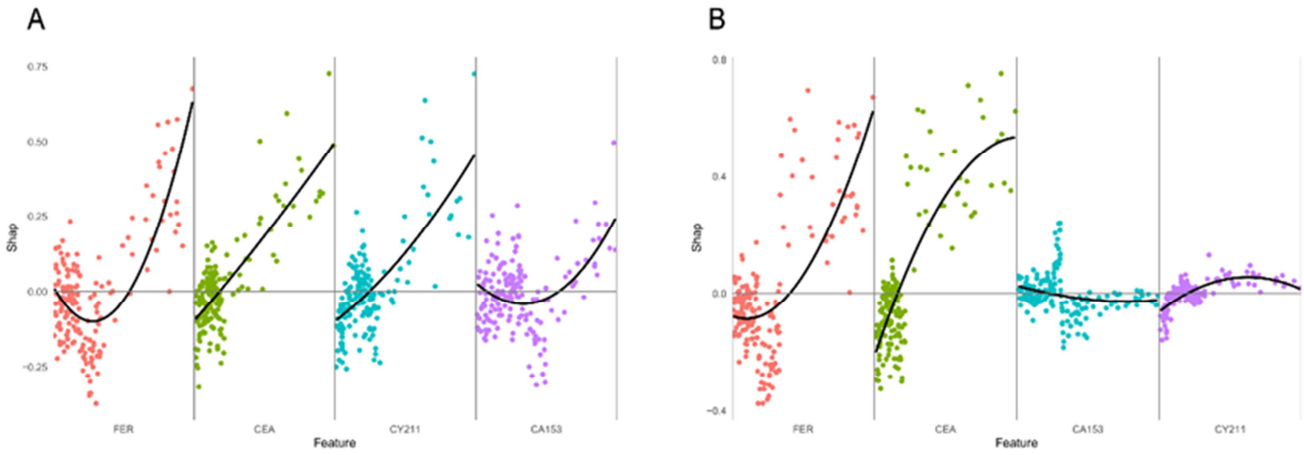

Figure 8. *5-Fold Cross-Validation (A. KNN B. RF).*

4. Discussion

Currently, the diagnosis of breast masses primarily relies on radiological imaging and histopathological examination. Although mammography and breast ultrasound are standard screening tools, they have limitations in early detection of breast cancer in women with dense breast tissue. Moreover, histopathological examination, while accurate, is invasive and can cause discomfort and risk to patients. Therefore, developing non-invasive methods with high sensitivity and specificity for early diagnosis is a critical direction in breast cancer research [6].

In recent years, machine learning has made significant advances in medical diagnostics, particularly in identifying and classifying breast masses. Machine learning algorithms can learn features from vast amounts of medical imaging and biomarker data, aiding in more precise diagnoses. In medical imaging, deep learning techniques have been widely applied to the automatic analysis of mammographic images, enhancing the detection rate of breast cancer. For instance, analyzing breast ultrasound images with convolutional neural networks (CNN) effectively differentiates between benign and malignant breast masses [7]. Additionally, MRI and PET scan image analysis have benefited from machine learning technology, increasing the diagnostic accuracy for breast cancer [8]. In the realm of biomarkers, machine learning plays a vital role in early detection and prognostic assessment of breast cancer. Analyzing the gene expression patterns of tumor tissues helps physicians select more appropriate treatment plans [9].

In our study, we developed a clinical prediction model incorporating four biomarkers: CEA, Ferritin, CA153, and CY211. The model's predictive capacity was evaluated using the C-index, ROC curves, calibration curves, and clinical decision curves. These evaluations demonstrated that the model possesses a high predictive ability, offering significant benefits to patients. Machine learning outcomes, including TNR, TPR, F-measure, and 5-fold cross-validation, indicated the model's efficacy in identifying patients with malignant breast tumors. These biomarkers play a significant role in the development of breast cancer. CEA, a widely used tumor marker glycoprotein, is closely associated with the invasiveness and prognosis of breast cancer in patients [10]. Another study also pointed out that CEA is one of the most valuable serum markers in breast cancer patients [11]. Elevated levels of Ferritin are related to the onset of breast cancer, with abnormal increases observed in breast cancer patients [12]. Studies have also found that increased levels of Ferritin in the blood and tumor tissues of breast cancer patients

are associated with poor prognosis and advanced histological grades [13]. CA153 is a specific marker for breast cancer, commonly used to monitor treatment effects and recurrence. Elevated levels of CA153 are typically associated with the presence and progression of breast cancer [14]. Notably, Wojtacki et al. demonstrated that in patients with metastatic breast cancer (MBC), an increase in CA15-3 levels of more than 30 u/mL could predict recurrence diagnosis 9 months in advance [15]. Tampellini et al. conducted a prospective trial on 526 MBC patients treated with anthracycline-based chemotherapy. They found a gradual shortening of median progression-free time from patients with normal CA15-3 levels (15.3 months) to those whose levels initially increased but decreased by 25% (11.7 months), to those with continually increasing levels (9.6 months), and finally to those with consistently high levels (8.6 months) [16]. It has been well-established that breast cancer cells express fragments of cytokeratin-19, one of the various cytokeratins that constitute the cell skeleton intermediate filaments. CY211, a tumor marker, can be detected using anti-CYFRA 21-1 antibodies for the serum fragments of cytokeratin-19. CY211 is a useful biomarker for detecting disease recurrence and assessing the treatment effect of breast cancer [17]. Another study showed that CY211 levels significantly decreased after two months of treatment with afatinib [18]. Our study demonstrates that the combined use of these biomarkers in early diagnosis of breast cancer shows high accuracy, indicating the combination's effectiveness in differentiating between benign breast masses and breast cancer patients. The advantage of a multi-biomarker combined diagnosis lies in providing more comprehensive tumor biological information, thus enhancing the sensitivity and specificity of the diagnosis. Additionally, the combination of multiple biomarkers can also aid in formulating personalized treatment plans, thereby improving patient prognosis.

5. Conclusion

In this study, a clinical prediction model for breast cancer was constructed by integrating multiple tumor biomarkers. This model demonstrates a high level of accuracy, offering a novel approach for the early diagnosis of breast cancer, and holds significance for the early detection and treatment of the disease. However, there are some limitations to our research. First, the levels of biomarkers are influenced by various factors and may exhibit individual variations. Second, our model is based on data from a specific population and may not be universally applicable. Overall, our research provides a fresh perspective on the early diagnosis of breast cancer, yet it necessitates further clinical data and in-depth studies for the validation and refinement of this model.

Author Contributions

Conceptualisation: Liying Qiu. Data curation: Zebin Liu, Lipeng Lin. Data analysis: Zebin Liu,, Guosheng Zhu. Writing-original draft: Zebin Liu, Liying Qiu. Approval of final manuscript: all authors.

ORCID

Zebin Liu: https://orcid.org/0009-0008-4993-0029
Lipeng Lin: https://orcid.org/0009-0004-2311-1333
Guosheng Zhu: https://orcid.org/0009-0009-5495-409X
Liying Qiu: https://orcid.org/0009-0006-0645-3256

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Wilkinson L, Gathani T. Understanding breast cancer as a global health concern. The British Journal of Radiology. 2021; 95. doi: 10.1259/bjr.20211033.

[2] Di Sibio A, Abriata G, Forman D, Sierra MS. Female breast cancer in Central and South America. Cancer Epidemiol. 2016; 44 Suppl 1: S110-S120. doi: 10.1016/j.canep.2016.08.010.

[3] Duffy MJ, et al. Tumor markers in breast cancer: Guidelines for use and interpretation in the clinical setting. European Journal of Cancer. 2005; 41(9): 1343-1357. doi: 10.1016/j.ejca.2005.02.025.

[4] K. Pantel, "Abstract PL01-01: Liquid biopsy: Novel technologies and clinical applications", Clinical Research (Excluding Clinical Trials), 2018. DOI: 10.1158/1538-7445. AM2018-PL01-01.

[5] M. Zubair, S. Wang, N. Ali, "Advanced Approaches to Breast Cancer Classification and Diagnosis", Frontiers in Pharmacology, 2021. DOI: 10.3389/fphar.2020.632079.

[6] Niell BL, Freer PE, Weinfurtner RJ, Arleo EK, Drukteinis JS. Screening for Breast Cancer. Radiol Clin North Am. 2017; 55(6): 1145-1162. doi: 10.1016/j.rcl.2017.06.004.

[7] Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016; 316(22): 2402-2410. doi: 10.1001/jama.2016.17216.

[8] Lakhani P, Prater AB, Hutson RK, et al. Machine Learning in Radiology: Applications Beyond Image Interpretation. J Am Coll Radiol. 2018; 15(2): 350-359. doi: 10.1016/j.jacr.2017.09.044.

[9] Savarese TM, Strohsnitter WC, Low HP, et al. Correlation of umbilical cord blood hormones and growth factors with stem cell potential: implications for the prenatal origin of breast cancer hypothesis. Breast Cancer Res. 2007; 9(3): R29. doi: 10.1186/bcr1674.

[10] Giovanella L, Ceriani L, Giardina G, et al. Serum cytokeratin fragment 21.1 (CYFRA 21.1) as tumour marker for breast cancer: comparison with carbohydrate antigen 15.3 (CA 15.3) and carcinoembryonic antigen (CEA). Clin Chem Lab Med. 2002; 40(3): 298-303. doi: 10.1515/CCLM.2002.047.

[11] Tarighati E, Keivan H, Mahani H. A review of prognostic and predictive biomarkers in breast cancer. Clin Exp Med. 2023; 23(1): 1-16. doi: 10.1007/s10238-021-00781-1.

[12] Fadavi P, Nafisi N, Hariri R, et al. Serum Ferritin, Vitamin D and Pathological Factors in Breast Cancer Patients. Med J Islam Repub Iran. 2021; 35: 162. Published 2021 Dec 6. doi: 10.47176/mjiri.35.162.

[13] Alkhateeb AA, Han B, Connor JR. Ferritin stimulates breast cancer cells through an iron-independent mechanism and is localized within tumor-associated macrophages. Breast Cancer Res Treat. 2013; 137(3): 733-744. doi: 10.1007/s10549-012-2405-x.

[14] Duffy MJ, Shering S, Sherry F, et al. CA 15-3: a prognostic marker in breast cancer. Int J Biol Markers. 2000; 15(4): 330-333. doi: 10.1177/172460080001500410.

[15] Wojtacki J, Kruszewski WJ, Sliwińska M, et al. Elevation of serum Ca 15-3 antigen: an early indicator of distant metastasis from breast cancer. Retrospective analysis of 733 cases. Przegl Lek. 2001; 58(6): 498-503.

[16] Tampellini M, Berruti A, Bitossi R, et al. Prognostic significance of changes in CA 15-3 serum levels during chemotherapy in metastatic breast cancer patients. Breast Cancer Res Treat. 2006; 98(3): 241-248. doi: 10.1007/s10549-005-9155-y

[17] Nakata B, Takashima T, Ogawa Y, et al. Serum CYFRA 21-1 (cytokeratin-19 fragments) is a useful tumour marker for detecting disease relapse and assessing treatment efficacy in breast cancer. Br J Cancer. 2004; 91(5): 873-878. doi: 10.1038/sj.bjc.6602074.

[18] Zhou L, Ren Y, Wang X, et al. Efficacy of afatinib in a HER2 amplification-positive endometrioid adenocarcinoma patient- a case report. Onco Targets Ther. 2019; 12: 5305-5309. Published 2019 Jul 4. doi: 10.2147/OTT.S206732.