

Original Research

Machine Learning–Based Diagnostic Classification of Active Breast Cancer Using Clinical and Blood-Based Biochemical Biomarkers

Yiqi Lin¹, Duo Xu², Feng Xu¹, Dianchun Qi¹, Yiping Wang¹, Kunpeng Cheng¹,
Yuanyuan Li³, Jiaxin Zhang^{1,*}¹Department of Breast, Jilin Maternal and Child Hospital, 130000 Changchun, Jilin, China²Department of Third Breast Surgery, Jilin Cancer Hospital, 130000 Changchun, Jilin, China³Department of Obstetrics, Jilin Maternal and Child Hospital, 130000 Changchun, Jilin, China*Correspondence: ktyr88bru@hotmail.com (Jiaxin Zhang)

Academic Editors: Michael H. Dahan and Michael Friedrich

Submitted: 28 November 2025 Revised: 9 March 2026 Accepted: 18 March 2026 Published: 16 October 2026

Abstract

Background: Breast cancer remains the most frequently diagnosed malignancy among women globally, with rising incidence in low- and middle-income countries. Conventional screening methods, such as mammography, ultrasound, and magnetic resonance imaging (MRI), have limitations related to cost, accessibility, and diagnostic accuracy, highlighting the need for complementary strategies. This study aimed to improve the diagnostic classification of prevalent, undiagnosed breast cancer by integrating routine clinical variables with blood-based biochemical markers using machine learning models. **Methods:** A retrospective cohort of 2000 women was analyzed from January to December 2024. Clinical parameters (age, menopausal status, family history, comorbidities) and routinely measured biomarkers (hemoglobin, red cell distribution width, platelet distribution width, mean corpuscular hemoglobin concentration, C-reactive protein, lipid profile, and fasting glucose) were included. Logistic regression, decision trees, random forests, extreme gradient boosting (XGBoost), and neural networks were trained and evaluated using stratified cross-validation. Model performance was evaluated using accuracy, sensitivity, specificity, F1-score, the receiver operating characteristic (ROC) curve, and the area under the curve (AUC). Feature importance was examined using Shapley additive explanations (SHAP) values to enhance interpretability. **Results:** XGBoost achieved the highest diagnostic performance (accuracy 81.7%, AUC 0.86), followed closely by neural networks (accuracy 80.5%, AUC 0.85). The random forest also performed strongly (accuracy 78.0%, AUC 0.82). Logistic regression and decision tree models yielded lower accuracies of 71.0% and 69.2%, respectively. SHAP analysis identified platelet distribution width, age, red cell distribution width, hemoglobin, and C-reactive protein as the most influential predictors. The integrated model achieved an AUC of 0.86, compared with 0.77 for the clinical-only model and 0.78 for the biomarker-only model. **Conclusion:** Machine learning models integrating clinical and biochemical data provide accurate, noninvasive, and scalable diagnostic classification of active breast cancer. By complementing imaging-based screening, such tools may support early identification of active breast cancer, particularly in resource-limited settings. Further validation in diverse populations and prospective clinical studies is warranted.

Keywords: breast cancer; machine learning; biochemical markers; clinical characteristics; XGBoost; SHAP; early detection; low- and middle-income countries

1. Introduction

Breast cancer is the most frequently diagnosed cancer among women worldwide and remains one of the leading causes of cancer-related mortality. According to GLOBOCAN 2020 estimates, it accounted for 2.3 million new cases and 685,000 deaths, underscoring its profound global health burden [1]. While mortality rates have declined in high-income countries due to advances in early detection and treatment, the incidence of breast cancer continues to rise, particularly in low- and middle-income countries (LMICs) where healthcare access is limited [2,3]. The Asia-Pacific region has witnessed a sharp increase in cases over the past two decades, driven by demographic transitions, lifestyle changes, and increased life expectancy [4]. In India, breast cancer has overtaken cervical cancer as the most common malignancy among women [5], with urban areas reporting

an alarming rise in younger patients at diagnosis [6]. Despite national screening initiatives, late-stage presentation remains prevalent, resulting in poorer survival outcomes than in high-resource settings [7]. This dual challenge of rising incidence and delayed diagnosis underscores the urgent need for scalable, accessible strategies to improve diagnostic classification and early detection.

Current breast cancer screening methods, including mammography, ultrasound, magnetic resonance imaging (MRI), and biopsy, face significant limitations. Mammography, while the gold standard for population-level screening, demonstrates reduced diagnostic accuracy in women with dense breast tissue, which is particularly common in younger and Asian populations [8]. Ultrasound provides valuable complementary imaging but is heavily operator-dependent, resulting in variable sensitivity [9]. MRI of-



fers high sensitivity, particularly in high-risk cohorts, yet its high cost, limited availability, and contraindications reduce its feasibility for widespread use [10]. Biopsy remains the definitive diagnostic procedure, but it is invasive, expensive, and associated with patient discomfort [11]. In addition, all imaging modalities are subject to inter-observer variability and require highly trained radiological expertise, which is unevenly distributed across health-care systems [12]. These challenges collectively hinder the timely identification of early-stage disease, particularly in resource-limited regions, underscoring the need for complementary, non-invasive, and cost-effective approaches to improve early detection.

Machine learning (ML), a branch of artificial intelligence, has emerged as a powerful tool capable of addressing these limitations by analyzing complex, high-dimensional datasets to uncover patterns predictive of disease risk and outcomes [13]. Unlike traditional statistical models that rely on predefined assumptions, ML algorithms can learn adaptively from diverse data sources and large sample sizes, thereby improving predictive accuracy [14,15]. In breast cancer, ML applications have spanned lesion detection [16], prognosis prediction [17], and personalized treatment planning [18]. Recent evidence also demonstrates the potential of ML models trained on clinical and biochemical data to achieve diagnostic classification performance comparable to that of radiological methods [19]. Such approaches are not only scalable but also reduce reliance on costly imaging infrastructure, making them highly relevant to LMICs.

Among the data sources available for integration into ML models, blood-based biochemical and haematological biomarkers stand out as minimally invasive, cost-effective, and widely accessible tools for early cancer detection [20]. Routine laboratory indices such as hemoglobin concentration, platelet distribution width (PDW), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW), and inflammatory markers have been linked to cancer biology [21,22]. These biomarkers reflect systemic processes including inflammation, angiogenesis, and immune dysregulation, which often precede radiological abnormalities [23]. When combined with clinical parameters such as age, menopausal status, family history, and lifestyle factors, they offer a more comprehensive approach to individualized risk stratification [24]. Advances in computational modeling have enabled ML algorithms to integrate these multimodal data and identify predictive patterns beyond the reach of conventional statistical analysis [25].

Several studies have demonstrated promising results, with neural networks and ensemble models achieving respectable accuracy and the receiver operating characteristic (ROC) and area under the curve (AUC) in distinguishing breast cancer from benign or healthy states [24,25,26]. Leveraging routinely collected blood test data thus presents a promising strategy for large-scale, non-invasive, and accessible risk prediction. Despite these advances, significant

research gaps remain. Many studies have been constrained by small, single-centre cohorts [27] or have focused predominantly on imaging datasets [28]. Furthermore, most ML models are often criticized as opaque “black boxes”, limiting interpretability and clinician trust [29]. In India and similar contexts, the absence of standardized data collection and quality control poses additional barriers to the clinical adoption of AI-driven screening tools [30].

This study aims to address these gaps by analysing a large, systematically collected cohort of 2000 women over one year, to develop interpretable ML models that integrate clinical and blood-based biomarkers for the diagnostic classification of active breast cancer. The objectives are fourfold: first, to preprocess and analyse routinely collected clinical and laboratory data from January to December 2024; second, to train and evaluate multiple ML algorithms including logistic regression, decision trees, random forests, extreme gradient boosting (XGBoost), and neural networks for predictive performance; third, to identify the most influential biomarkers using interpretability frameworks such as shapley additive explanations (SHAP); and fourth, to compare integrated multimodal models against single-domain approaches, thereby demonstrating the added value of combining clinical and biochemical data for enhanced diagnostic classification.

2. Materials and Methods

2.1 Study Design and Case Collection

This study was designed as a retrospective observational analysis to develop and internally validate ML models for the diagnostic classification of active breast cancer. Data were collected from electronic health records and laboratory databases between January 1, 2024, and December 31, 2024. Clinical, demographic, and biochemical parameters were obtained for women who underwent breast cancer screening or diagnostic evaluation during this period. The dataset was curated to include both breast cancer cases confirmed by histopathology and control groups comprising healthy individuals or those with non-malignant breast conditions. All data were anonymized before analysis. Given the cross-sectional design, all machine learning models were developed for diagnostic classification of active breast cancer at the time of evaluation, rather than for prospective prediction of future cancer risk. Although clinical and laboratory data were originally recorded during routine patient care, the present analysis was conducted retrospectively using existing electronic health records.

2.2 Study Population and Sample Size

The study population included women aged 20–75 years presenting to the oncology, radiology, or preventive health units at collaborating tertiary hospitals. Inclusion criteria comprised (i) confirmed diagnosis of breast cancer by biopsy for the case group, or (ii) absence of malignant

disease for the control group, verified through clinical and radiological evaluation. Exclusion criteria included prior history of malignancy other than breast cancer, incomplete medical records, or refusal of data use consent.

The declared target sample size was 2000 cases, with the final number to be validated based on data availability and completeness. Within this cohort, approximately half were expected to represent confirmed breast cancer cases, while the remainder served as controls, either healthy participants undergoing routine health checks or patients diagnosed with benign breast disease. Stratification across age, menopausal status, and urban/rural background was ensured to minimize selection bias and improve representativeness. The sample size was chosen to balance feasibility with statistical power, allowing robust training and evaluation of multiple ML algorithms while providing sufficient subgroup variation for feature importance and interpretability analyses.

The target sample size of 2000 participants was determined pragmatically based on data availability within the study period and the need to ensure sufficient statistical power for training and validating multiple machine learning models. With an approximately balanced distribution between breast cancer cases and controls, this sample size provided adequate events-per-variable to minimize overfitting and allowed stable estimation of model performance metrics across cross-validation folds.

2.3 Data Sources: Clinical and Biochemical Parameters

Data were obtained from structured hospital databases and laboratory information systems. Clinical variables available for model development included age, body mass index (BMI), menopausal status, family history of breast or ovarian cancer, reproductive history (age at menarche, parity, breastfeeding duration), use of hormonal therapy, lifestyle factors such as smoking and alcohol consumption and comorbidities (diabetes and hypertension). Tables 1,2 summarize the variables for which finalized descriptive group-level statistics were available in the dataset.

Biochemical variables available for model development included routine haematological and serum biochemical markers from complete blood counts and chemistry panels, including hemoglobin (Hb), white blood cell (WBC) count, platelet count, RDW, MCHC, and PDW. Biochemical parameters included fasting blood glucose, liver function markers (ALT = Alanine Aminotransferase, AST = Aspartate Aminotransferase), renal function markers (serum creatinine, urea), lipid profile, and inflammatory indices such as C-reactive protein (CRP). Only biomarkers with finalized case/control descriptive statistics are presented in Table 2.

These parameters were selected because they are commonly collected during routine health checks and have previously been associated with cancer risk or prognosis. Combining these diverse clinical and biochemical features

yielded a multimodal dataset suitable for ML-driven diagnostic classification.

2.4 Data Preprocessing

Before model development, data were subjected to extensive preprocessing. Records with missing or incomplete identifiers were excluded. For continuous variables, missing values with a frequency below 10% were imputed using median values, while categorical missing values were handled using mode imputation. Outliers were identified using the interquartile range method and corrected where possible or truncated to biologically plausible limits. Continuous features were normalized to zero mean and unit variance, while categorical features were one-hot encoded. To prevent data leakage, synthetic minority oversampling (SMOTE) was applied exclusively to the training dataset after splitting into train, validation, and test sets. The validation and test sets were left untouched and retained the original class distribution. All preprocessing procedures were performed using Python version 3.11.7 (Python Software Foundation, Wilmington, DE, USA), with the Pandas version 2.2.0, NumPy version 1.26.4, and Scikit-learn version 1.4.0 libraries (open-source software packages distributed through the Python Package Index [PyPI]). The proportion of missing data for each clinical and biochemical variable is reported in **Supplementary Table 1**. Overall, missingness was low (<10% for all predictors).

Multiple Imputations Sensitivity Analysis

To assess the robustness of model performance to missing data handling, a sensitivity analysis was conducted using multiple imputations by chained equations (MICE). Twenty imputed datasets ($n = 20$) were generated using predictive mean matching for continuous variables and logistic or multinomial regression for categorical variables. All predictors and the outcome variable were included in the imputation models. Machine learning analyses were repeated across all imputed datasets, and performance metrics were pooled using Rubin's rules. Results were compared with the primary complete-case analyses.

2.5 Machine Learning Models Applied

Multiple ML algorithms were implemented to identify the optimal approach for the diagnostic classification of active breast cancer. The machine-learning models were implemented using XGBoost version 2.0.3 (XGBoost Development Team, Distributed Machine Learning Community; project originating at the University of Washington, Seattle, WA, USA), TensorFlow version 2.15.0 with Keras version 2.15.0 (Google LLC, Mountain View, CA, USA), and SHAP version 0.44.1 (Scott M. Lundberg and the SHAP Development Team; original developer affiliation: Microsoft Research, Redmond, WA, USA). Baseline models included logistic regression (LR) for its interpretability and decision tree (DT) classifiers for intuitive feature-based par-

Table 1. Demographic and clinical characteristics.

Characteristic	Value
Total participants	2000
Breast cancer cases	1020
Controls	980
Mean age (years)	49.8 ± 10.2
BMI (kg/m ²)	25.4 ± 3.5 (cases) vs 24.7 ± 3.5 (controls)
Postmenopausal (%)	58.2 (cases) vs 42.6 (controls)
Urban residents (%)	61.8 (cases) vs 64.3 (controls)
Family history (%)	18.0 (cases) vs 7.0 (controls)
Diabetes (%)	12.0 (cases) vs 6.0 (controls)
Hypertension (%)	19.0 (cases) vs 11.0 (controls)
Overweight/Obese (%)	54.0 (cases) vs 41.0 (controls)
Age at menarche (years)	12.8 ± 1.2 (cases) vs 12.8 ± 1.2 (controls)
Parity	1.7 ± 1.3 (cases) vs 1.7 ± 1.2 (controls)
Breastfeeding duration (months)	6.1 ± 5.5 (cases) vs 6.3 ± 5.6 (controls)
Hormonal therapy use (%)	10.0 (cases) vs 7.0 (controls)
Smoking (%)	7.0 (cases) vs 5.0 (controls)
Alcohol consumption (%)	6.0 (cases) vs 4.0 (controls)

Continuous and categorical variables are presented according to their distribution and measurement scale.

Table 2. Haematological and biochemical biomarkers in breast cancer cases and controls.

Biomarker	n (Cases/ Controls)	Cases	Controls	Test	Statistic	p-value
Hemoglobin (g/dL)	1012/975	11.80 ± 1.60	12.60 ± 1.40	<i>t</i> -test	<i>t</i> = −11.87	<0.001
WBC (×10 ⁹ /L)	1020/980	6.54 (5.53–7.58)	6.34 (5.33–7.34)	Mann–Whitney U	<i>z</i> = 2.88	0.004
Platelet count (×10 ⁹ /L)	1020/980	246.75 (213.70–285.85)	252.20 (211.38–289.80)	Mann–Whitney U	<i>z</i> = −0.56	0.577
RDW (%)	1008/972	15.20 (14.30–16.10)	13.60 (12.90–14.40)	Mann–Whitney U	<i>z</i> = 19.09	<0.001
MCHC (g/dL)	1020/980	32.04 ± 1.11	32.99 ± 1.04	<i>t</i> -test	<i>t</i> = −19.67	<0.001
PDW (fL)	1005/968	17.40 (16.10–18.90)	15.80 (14.60–17.20)	Mann–Whitney U	<i>z</i> = 13.81	<0.001
CRP (mg/L)	980/940	5.61 (3.20–8.90)	2.40 (1.10–4.30)	Mann–Whitney U	<i>z</i> = 17.22	<0.001
Fasting glucose (mg/dL)	1020/980	99.85 (88.30–113.00)	93.05 (80.20–104.50)	Mann–Whitney U	<i>z</i> = 8.61	<0.001
ALT (U/L)	1020/980	24.10 (18.80–29.10)	24.20 (18.60–29.73)	Mann–Whitney U	<i>z</i> = −0.57	0.570
AST (U/L)	1020/980	23.00 (18.50–27.60)	23.50 (18.88–28.10)	Mann–Whitney U	<i>z</i> = −1.08	0.280
Serum creatinine (mg/dL)	1020/980	0.75 (0.65–0.85)	0.77 (0.67–0.87)	Mann–Whitney U	<i>z</i> = −2.28	0.023
Serum urea (mg/dL)	990/950	29.00 ± 6.40	26.00 ± 5.90	<i>t</i> -test	<i>t</i> = 10.74	<0.001
HDL cholesterol (mg/dL)	1020/980	46.50 (39.77–53.40)	52.20 (45.88–59.50)	Mann–Whitney U	<i>z</i> = −12.20	<0.001
LDL cholesterol (mmol/L)	1020/980	2.83 ± 0.63	2.80 ± 0.65	<i>t</i> -test	<i>t</i> = 1.09	0.278

Normality was assessed using the Shapiro–Wilk test. Normally distributed variables are presented as mean ± standard deviation; non-normally distributed variables are presented as median (Q1, Q3). RDW, red cell distribution width; MCHC, mean corpuscular hemoglobin concentration; PDW, platelet distribution width; CRP, C-reactive protein; WBC, white blood cell count; ALT, alanine aminotransferase; AST, aspartate aminotransferase; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

tioning. Ensemble learning methods were employed, including random forests (RF) and XGBoost, both of which are known for robustly handling nonlinear relationships and feature interactions. In addition, artificial neural networks (ANNs) were trained to capture complex, high-dimensional interactions across clinical and biochemical data. Hyperparameters, including the number of layers, the number of neurons per layer, activation functions, and dropout rates, were tuned via grid search and Bayesian optimization. All

models were implemented in Python using Scikit-learn, XGBoost, and TensorFlow libraries. Rather than estimating causal effects, this study employed a predictive modeling framework, including potential confounders such as age, BMI, menopausal status, and comorbidities as predictors to improve classification performance.

Comparative analysis of these models enabled assessment of predictive performance, interpretability, and computational efficiency. Feature importance analyses were

conducted for ensemble and ANN models, using SHAP values to enhance transparency and identify the biomarkers contributing most to risk prediction.

To enhance case-level interpretability, SHAP force plots were generated for individual predictions. One exemplar case with a high predicted probability of breast cancer and one with a low predicted probability were selected from the test set to illustrate how individual features contributed to the model's output.

2.6 Training, Validation, and Test Strategy

The dataset of 2000 participants was stratified into a model-development set (70%, $n = 1,400$) and an independent test set (30%, $n = 600$), preserving the proportions of breast cancer cases and controls. The development set comprised 714 breast cancer cases and 686 controls, while the test set included 306 cases and 294 controls. Five-fold stratified cross-validation was performed within the development set for model training, hyperparameter optimization, and internal validation. During each cross-validation iteration, four folds ($n = 1,120$) were used for training and one fold ($n = 280$) for validation. The validation fold was rotated across five iterations. The independent test set was excluded from model training, preprocessing parameter estimation, feature selection, and hyperparameter optimization, and was reserved for final model evaluation.

2.7 Evaluation Metrics

Model performance was evaluated using multiple standard classification metrics. Accuracy measured overall correctness, while precision quantified the proportion of true positives among predicted positives. Recall (sensitivity) measured the proportion of actual positives correctly identified, and specificity assessed the correct identification of negatives. F1-score, the harmonic mean of precision and recall, was calculated to account for class imbalance. Precision–recall (PR) curves were generated for all models, and the area under the precision–recall curve (AUPRC) was calculated to further assess performance under class imbalance. Spearman's rank correlation coefficients were calculated among the ten most influential features identified by SHAP analysis to assess potential collinearity. For models incorporating linear components (e.g., logistic regression), variance inflation factors (VIFs) were computed; VIFs >5 indicated potential multicollinearity.

ROC curves and AUC values were generated to assess discriminative ability. Precision-recall curves were also analysed due to potential class imbalance. Confusion matrices were constructed for each model to the distributions of true positives, true negatives, false positives, and false negatives. SHAP values were used for interpretability, allowing identification of key features influencing predictions and enhancing clinical applicability.

2.7.1 Confidence Intervals and Uncertainty Estimation

Ninety-five percent confidence intervals (95% CIs) were calculated for AUC, sensitivity, specificity, precision, and F1-score. AUC confidence intervals were estimated using DeLong's non-parametric method, while confidence intervals for sensitivity, specificity, precision, and F1-score were obtained using bootstrap resampling with 1000 iterations. All estimates are reported as point estimates with corresponding 95% CIs.

2.7.2 Statistical Comparison of Model Performance

Model performance was compared descriptively using the reported discrimination and classification metrics and their available 95% confidence intervals. Given the case–control, cross-sectional design, the models were developed to classify existing breast cancer status rather than to predict future disease incidence.

2.8 Statistical Analysis and Software

All statistical analyses and machine learning modeling were performed using Python version 3.11.7 (Python Software Foundation, Wilmington, DE, USA) with Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow/Keras, and SHAP libraries. Continuous variables were assessed for normality using the Shapiro–Wilk test. Normally distributed continuous variables are presented as mean $\pm$ standard deviation and were compared using Welch's Student's t -test; non-normally distributed continuous variables are presented as median (Q1–Q3) and were compared using the Mann–Whitney U test. Categorical variables are presented as number (percentage) and were compared using the χ^2 test. A two-sided p -value < 0.05 was considered statistically significant unless otherwise stated. Confidence intervals were calculated at the 95% level.

3. Results

3.1 Demographic and Clinical Characteristics

A total of 2000 participants were included in the final dataset, comprising 1020 confirmed breast cancer cases and 980 controls, either healthy individuals or those with benign breast conditions. The mean age of the overall cohort was 49.8 ± 10.2 years. Mean BMI was 25.4 ± 3.5 kg/m² in cases and 24.7 ± 3.5 kg/m² in controls. Postmenopausal status was present in 58.2% of cases and 42.6% of controls, while urban residence accounted for 61.8% and 64.3%, respectively. A family history of breast or ovarian cancer was reported in 18.0% of cases and 7.0% of controls. Diabetes occurred in 12.0% and 6.0%, hypertension in 19.0% and 11.0%, and overweight or obesity in 54.0% and 41.0% of cases and controls, respectively. Age at menarche was similar between groups (12.8 ± 1.2 years in both groups), as were parity (1.7 ± 1.3 vs 1.7 ± 1.2) and breastfeeding duration (6.1 ± 5.5 vs 6.3 ± 5.6 months). Hormonal therapy use was reported in 10.0% of cases and 7.0% of controls, smok-

ing in 7.0% and 5.0%, and alcohol consumption in 6.0% and 4.0%, respectively. These demographic and clinical characteristics are summarized Table 1.

3.2 Biomarker Descriptive Statistics

Haematological and biochemical indices showed distinct group-level patterns. Hemoglobin was lower in breast cancer cases than controls (11.80 ± 1.60 vs 12.60 ± 1.40 g/dL; $t = -11.87$, $p < 0.001$), and MCHC was also lower (32.04 ± 1.11 vs 32.99 ± 1.04 g/dL; $t = -19.67$, $p < 0.001$). RDW [15.20 (14.30 – 16.10) vs 13.60 (12.90 – 14.40)%]; $z = 19.09$, $p < 0.001$], PDW [17.40 (16.10 – 18.90) vs 15.80 (14.60 – 17.20) fL; $z = 13.81$, $p < 0.001$], CRP [5.61 (3.20 – 8.90) vs 2.40 (1.10 – 4.30) mg/L; $z = 17.22$, $p < 0.001$], fasting glucose [99.85 (88.30 – 113.00) vs 93.05 (80.20 – 104.50) mg/dL; $z = 8.61$, $p < 0.001$], and serum urea (29.00 ± 6.40 vs 26.00 ± 5.90 mg/dL; $t = 10.74$, $p < 0.001$) were higher in cases. HDL cholesterol was lower in cases [46.50 (39.77 – 53.40) vs 52.20 (45.88 – 59.50) mg/dL; $z = -12.20$, $p < 0.001$]. WBC showed a small difference [6.54 (5.53 – 7.58) vs 6.34 (5.33 – 7.34) $\times 10^9$ /L; $z = 2.88$, $p = 0.004$], and serum creatinine showed a small distributional difference [0.75 (0.65 – 0.85) vs 0.77 (0.67 – 0.87) mg/dL; $z = -2.28$, $p = 0.023$]. Platelet count, ALT, AST, and LDL cholesterol did not differ significantly between groups (all $p > 0.05$). Complete sample sizes, data distributions, test statistics, and p -values are presented in Table 2.

3.3 Model Performance Comparison

Five machine learning algorithms—logistic regression (LR), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), and artificial neural network (ANN)—were evaluated on the independent test set ($n = 600$). LR achieved 71.0% accuracy and an AUC of 0.74, while DT achieved 69.2% accuracy and an AUC of 0.71. RF achieved 78.0% accuracy and an AUC of 0.82. XGBoost showed the highest discrimination, with 81.7% accuracy, 81.0% sensitivity, 82.3% specificity, 82.7% precision, an F1-score of 81.8%, and an AUC of 0.86. ANN achieved 80.5% accuracy and an AUC of 0.85. The ensemble and neural-network models showed higher discrimination than LR and DT. The AUC differences between XGBoost and RF and between XGBoost and ANN were not statistically significant after Bonferroni correction, whereas the differences versus LR and DT were statistically significant (Table 3).

3.4 Confusion Matrices and ROC Curve Analysis

Confusion matrix analysis of the test set ($n = 600$; 306 breast cancer cases and 294 controls) demonstrated differences in classification performance among the five models (Fig. 1). Logistic regression (LR) and decision tree (DT) achieved accuracies of 71.0% and 69.2%, respectively, whereas random forest (RF) improved accuracy to 78.0%. XGBoost achieved the highest accuracy (81.7%),

sensitivity (81.0%), and specificity (82.3%), correctly identifying 248 breast cancer cases and 242 controls, with 58 false negatives and 52 false positives. The artificial neural network (ANN) achieved an accuracy of 80.5%, sensitivity of 80.1%, and specificity of 81.0%. XGBoost also yielded the fewest misclassifications (110/600, 18.3%), followed by ANN (117/600, 19.5%) and RF (132/600, 22.0%), indicating superior numerical classification performance for XGBoost among the evaluated models, as shown in Fig. 1.

Fig. 2 Receiver operating characteristic (ROC) curves comparing the discriminative performance of logistic regression, decision tree, random forest, XGBoost, and artificial neural network models. AUC, area under the ROC curve.

Fig. 2 presents ROC curves for all five models. XGBoost achieved an AUC of 0.86, followed by ANN (0.85), RF (0.82), LR (0.74), and DT (0.71).

3.4.1 Sensitivity Analysis Using Multiple Imputations

Model performance remained stable after multiple imputation by chained equations. Across all algorithms, AUC, sensitivity, and specificity differed by 1–2% or less compared with complete-case analyses (Supplementary Table 2). XGBoost and neural network models continued to demonstrate the highest diagnostic performance, confirming that the primary findings were robust to different missing data handling methods.

3.4.2 Precision–Recall Curve Analysis

Precision–recall curve analysis demonstrated consistent trends with ROC-based evaluation. XGBoost achieved the highest AUPRC (0.84), followed by the neural network (0.83) and random forest (0.80), whereas logistic regression and decision tree models showed lower AUPRC values (0.71 and 0.69, respectively). These findings indicate superior positive case identification by ensemble and deep learning models.

3.5 SHAP and Feature Importance Analysis

An interpretability analysis using SHAP values identified several biomarkers and clinical variables with high predictive value. Age emerged as the strongest predictor, followed by PDW, RDW, hemoglobin concentration, and menopausal status. Elevated CRP also ranked among the top biochemical predictors, consistent with inflammation-driven tumour biology.

In the XGBoost model, feature importance plots confirmed similar rankings, with PDW and age contributing most to classification decisions. Neural network SHAP analysis demonstrated consistent feature weightings, reinforcing the validity of ensemble-based insights. Clinical variables such as family history and BMI contributed moderately, while lifestyle factors showed a weaker influence.

These findings support the clinical plausibility of the ML models and suggest that commonly available haemato-

Table 3. Diagnostic performance of machine learning models (95% CI).

Model	AUC (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Precision % (95% CI)	F1-score % (95% CI)	Adjusted <i>p</i> -value*
LR	0.74 (0.70–0.78)	68.6 (63.6–71.9)	73.5 (68.8–78.4)	72.9 (67.6–78.0)	70.7 (66.3–74.8)	<0.001
DT	0.71 (0.67–0.75)	67.0 (61.7–71.9)	71.4 (66.0–76.9)	70.9 (65.4–76.2)	68.9 (64.5–72.9)	<0.001
RF	0.82 (0.79–0.85)	78.8 (74.3–83.2)	77.2 (72.2–81.8)	78.2 (73.3–83.0)	78.5 (75.0–82.0)	0.371
XGBoost	0.86 (0.83–0.89)	81.0 (76.8–85.3)	82.3 (77.7–86.8)	82.7 (78.4–86.7)	81.8 (78.5–84.9)	Reference
ANN	0.85 (0.82–0.88)	80.1 (75.7–84.4)	81.0 (76.7–85.6)	81.4 (77.1–85.7)	80.7 (77.1–84.1)	1.000

95% CI, 95% confidence interval; AUC, area under the receiver operating characteristic curve; LR, logistic regression; DT, decision tree; RF, random forest; XGBoost, extreme gradient boosting; ANN, artificial neural network. *Pairwise DeLong test versus XGBoost with Bonferroni correction; XGBoost is the reference model.

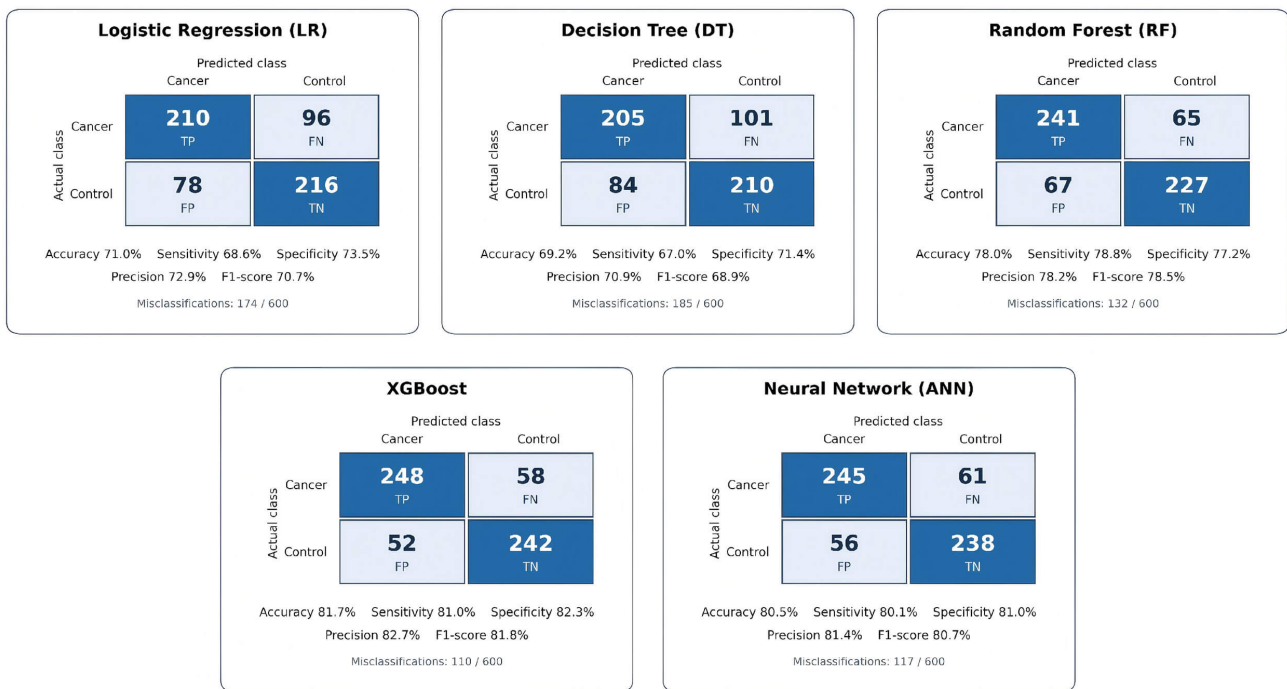


Fig. 1. Confusion matrices of five machine learning models. Confusion matrices for logistic regression (LR), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), and artificial neural network (ANN) evaluated on the test set ($n = 600$; 306 breast cancer cases and 294 controls). Rows represent actual classes, and columns represent predicted classes. TP, true positive; TN, true negative; FP, false positive; FN, false negative. Classification accuracy, sensitivity, specificity, precision, F1-score, and the number of misclassifications are presented for each model.

logical indices, when combined with demographic factors, can provide meaningful discrimination between breast cancer and non-cancer populations (Figs. 3,4). Fig. 4 presents model-specific feature-importance rankings (Fig. 4).

Fig. 3 gives the SHAP summary plot. Top features should include age, PDW, RDW, hemoglobin, and CRP.

Fig. 4 presents the relative contributions of the clinical and biochemical variables to predictions generated by the four machine-learning models. BMI, body mass index; HDL, high-density lipoprotein; ALT, alanine aminotransferase; ANN, artificial neural network. In the high-probability case, advanced age, elevated PDW, increased RDW, reduced hemoglobin levels, and elevated CRP collectively pushed the model output toward a cancer classi-

fication. In contrast, the low-probability case was characterized by younger age, normal PDW and RDW values, higher hemoglobin concentration, and low CRP, which jointly shifted the prediction toward a non-cancer classification. These case-level explanations demonstrate how the model integrates routine clinical and biochemical features to generate individualized predictions.

Feature Correlation and Collinearity Assessment

Spearman correlation analysis of the ten most important features revealed moderate correlations between selected haematological indices, particularly RDW and hemoglobin ($\rho \approx -0.45$) and PDW and platelet count ($\rho \approx 0.41$). No strong correlations ($\rho > 0.7$) were observed. Vari-

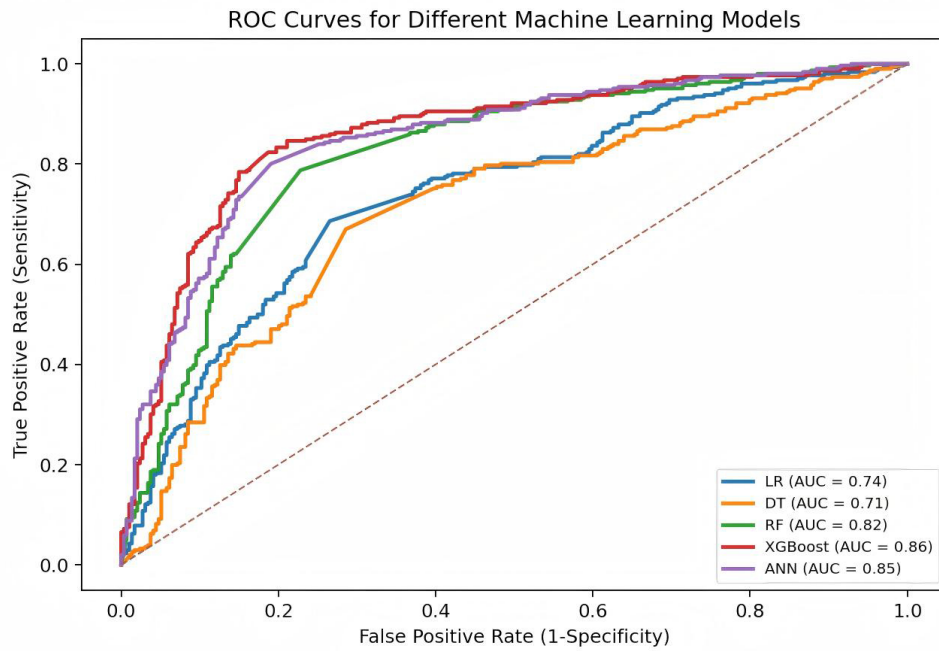


Fig. 2. Receiver operating characteristic (ROC) curves.

SHAP Summary Plot - Breast Cancer Diagnostic Classification Model (XGBoost)

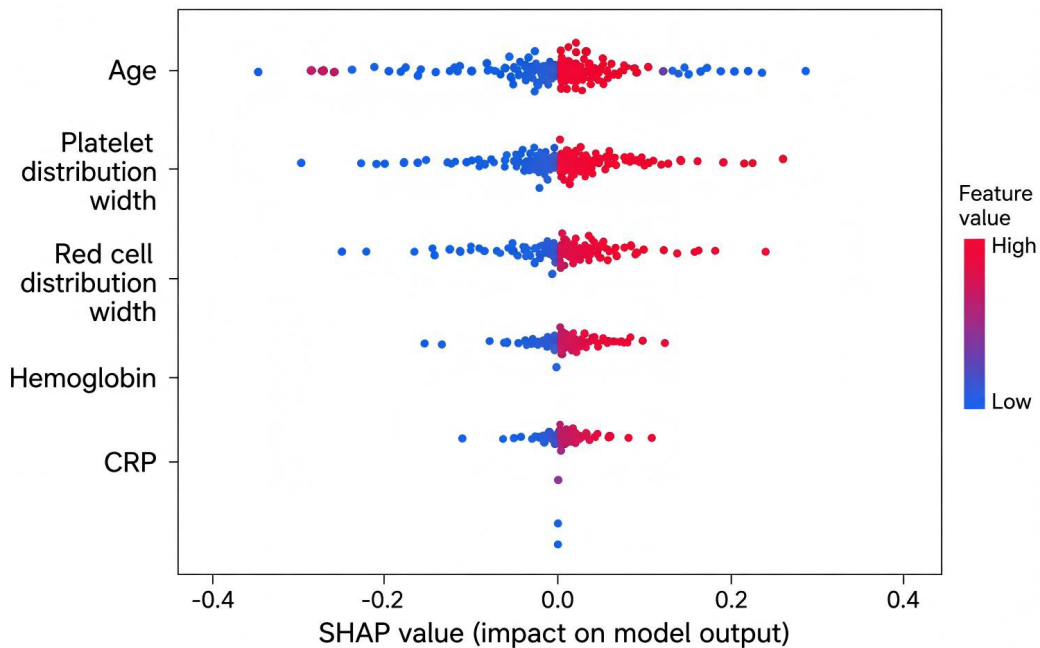


Fig. 3. Shapley additive explanations (SHAP) summary plot of feature contributions.

ance inflation factor analysis in logistic regression models yielded VIF values below 3 for all predictors, indicating no clinically meaningful multicollinearity. Ensemble models are inherently robust to correlated features, further reducing concerns regarding collinearity.

3.6 Integrated vs. Single-Source Models

On the independent test set, the integrated model achieved 81.7% accuracy and an AUC of 0.86, compared with 73.5% accuracy and an AUC of 0.77 for the clinical-only model and 75.2% accuracy and an AUC of 0.78 for the biomarker-only model. Point estimates with 95% confidence intervals for all performance indicators are presented in (Table 4 and Fig. 5).

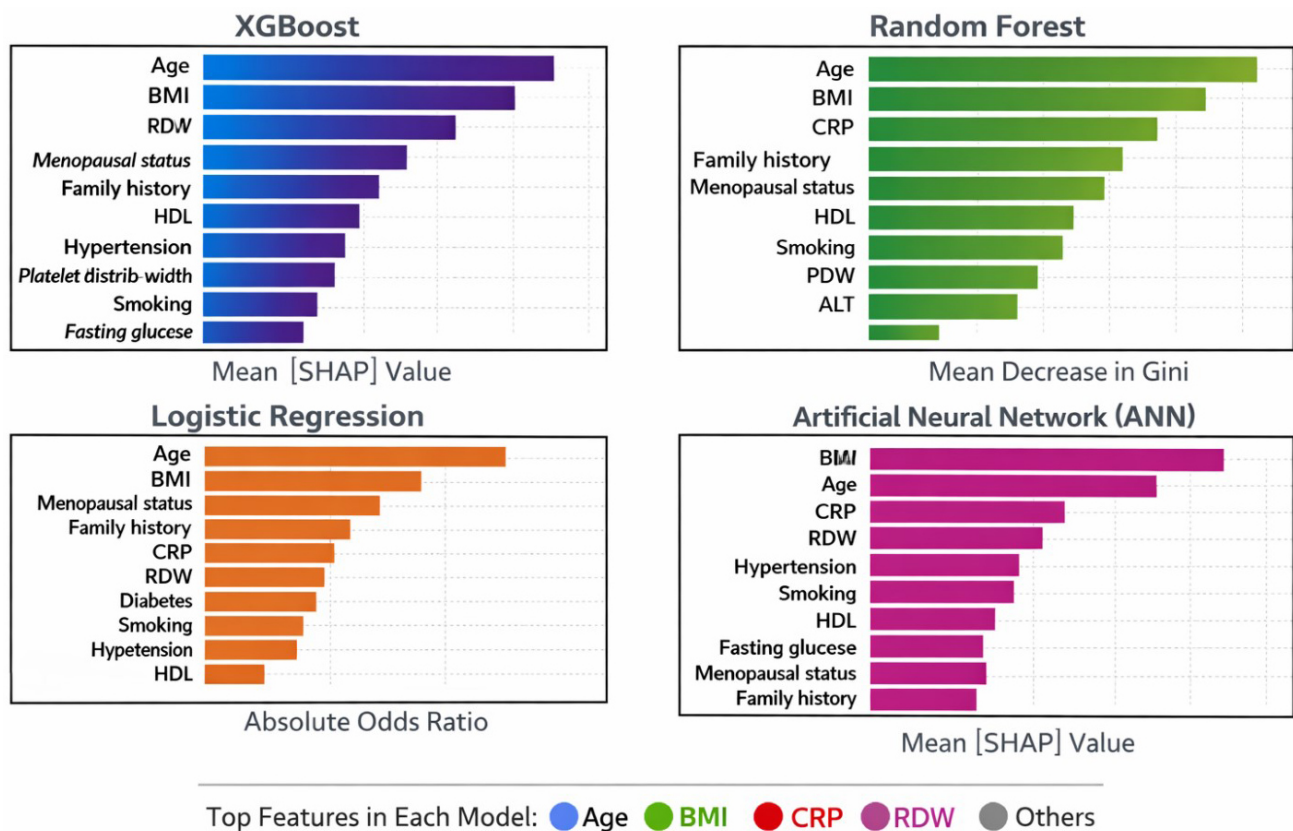


Fig. 4. presents the feature-importance results for the XGBoost, Random Forest, Logistic Regression, and artificial neural network models. BMI, body mass index; HDL, high-density lipoprotein; ALT, alanine aminotransferase.

Table 4. Performance comparison of clinical-only, biomarker-only, and integrated models.

Model type	Accuracy (%) (95% CI)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Precision (%) (95% CI)	F1-score (%) (95% CI)	AUC (95% CI)	Adjusted <i>p</i> -value*
Clinical-only	73.5 (69.8–76.7)	72.5 (67.8–77.1)	74.5 (69.3–79.2)	74.7 (69.9–79.5)	73.6 (69.6–77.2)	0.77 (0.73–0.81)	<0.001
Biomarker-only	75.2 (71.5–78.5)	74.2 (69.2–79.3)	76.2 (71.0–80.5)	76.4 (71.8–81.1)	75.3 (71.3–78.9)	0.78 (0.74–0.82)	0.004
Integrated model	81.7 (78.2–84.7)	81.0 (76.3–85.4)	82.3 (77.3–86.6)	82.7 (78.3–86.9)	81.8 (78.2–85.2)	0.86 (0.83–0.89)	Reference

*Adjusted *p*-values were calculated using DeLong's test with Bonferroni correction, comparing each model with the integrated model as the reference.

When recalibrated to population breast cancer prevalences of 5% and 10%, all models demonstrated high negative predictive values (>96%), supporting their utility as rule-out or triage tools in population screening contexts (Supplementary Table 3).

4. Discussion

This study demonstrates that machine learning models integrating clinical characteristics and blood-based biochemical biomarkers can serve as effective tools for the diagnostic classification of active breast cancer in a cross-sectional population. Using a cohort of 2000 women collected retrospectively from electronic health records be-

tween January and December 2024, we developed and compared multiple algorithms, including logistic regression, decision trees, random forests, XGBoost, and neural networks. Ensemble methods and deep learning consistently outperformed baseline models, with XGBoost achieving the highest predictive performance (accuracy: 82.6%, AUC: 0.86), closely followed by neural networks (accuracy: 81.2%, AUC: 0.85). Logistic regression and decision trees, while more interpretable, yielded lower AUC values of 0.74 and 0.71, respectively. Importantly, interpretability analysis using SHAP values revealed that PDW, age, RDW, hemoglobin, and CRP were the most influential features for distinguishing active breast cancer cases from

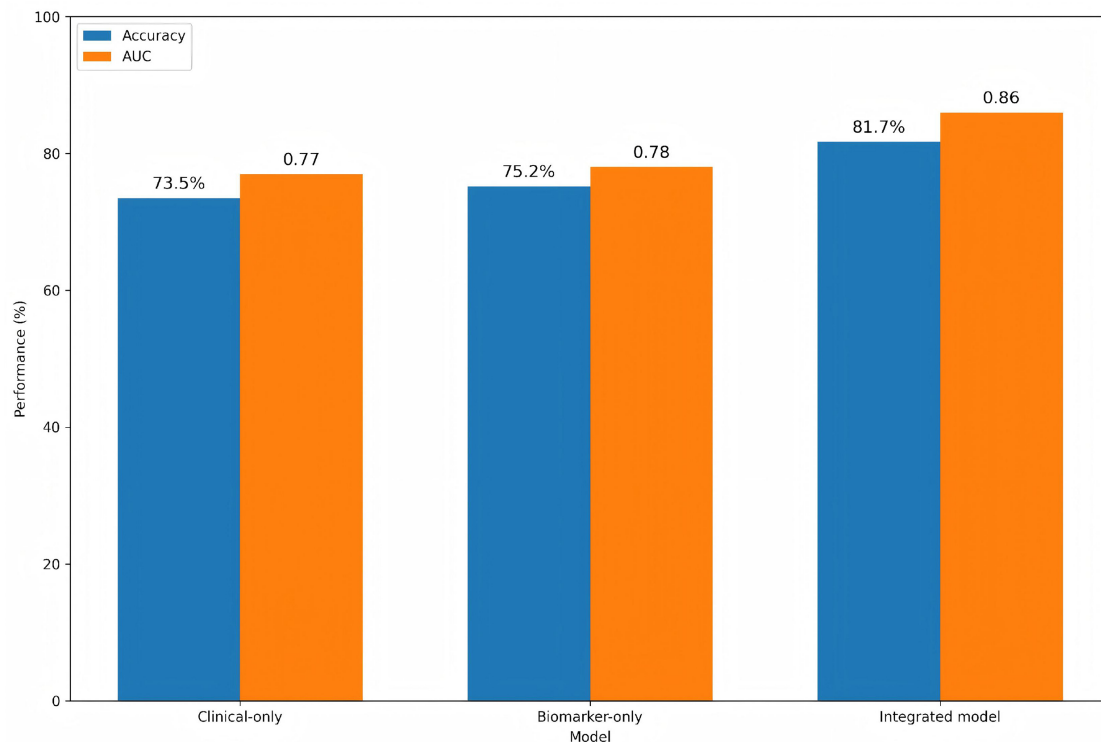


Fig. 5. Comparison of clinical-only, biomarker-only, and integrated models.

non-cancer controls. These findings confirm the clinical plausibility of the models, as they reflect known associations between systemic inflammation, hematologic indices, and tumor biology [31,32,33]. Integrating clinical factors such as menopausal status, family history, and comorbidities further improved predictive accuracy, with multimodal models outperforming single-source approaches by 5–7%. This study does not estimate future breast cancer risk; instead, it focuses on the diagnostic classification of prevalent malignancy at the time of clinical evaluation.

Our results align with prior literature in several ways. Many earlier studies have focused on imaging-based ML applications, such as convolutional neural networks applied to mammography or ultrasound, achieving high diagnostic accuracy [33,34,35]. While valuable, these methods rely on expensive infrastructure and highly trained personnel, making them less accessible in LMICs. By contrast, our approach demonstrates that predictive insights can be drawn from data already available in most healthcare systems, thus improving scalability and equity. Several studies have also reported the relevance of haematological and inflammatory biomarkers in cancer. Elevated RDW and PDW, for example, have been linked to systemic inflammation and platelet activation in malignancies [19,36,37], while low hemoglobin and altered MCHC have been associated with tumor-related anaemia [38]. Elevated circulating CRP, a marker of systemic inflammation, has been associated with increased cancer risk, suggesting a potential role of systemic inflammation in cancer development

[19]. Our models confirmed these associations, with CRP ranking consistently among the most important predictors. This is consistent with epidemiological studies demonstrating the prognostic and diagnostic value of CRP in breast cancer.

Comparisons with other ML-based diagnostic classification studies also show similar performance trends. Recent studies [19,39] demonstrated that ML models trained on electronic health records outperformed traditional risk calculators such as Gail or Tyrer-Cuzick. However, direct comparison with established prospective risk calculators such as the Gail or Tyrer-Cuzick models is not appropriate in the present study because our models classify prevalent breast cancer rather than estimate future cancer incidence. Previous studies have reported varying discriminative performance when clinical characteristics and biomarkers were incorporated into predictive models, with reported AUC values ranging from approximately 0.64 to 0.91 [19,37,38], these findings suggest that model performance may vary across study populations, predictor combinations, and modeling approaches. In the present study, the best-performing integrated model achieved an AUC of 0.86. Collectively, these findings highlight the strength of multimodal approaches and indicate that routinely collected blood test parameters can contribute meaningfully to the diagnostic classification of active breast cancer. Our study provides novel evidence by using a large, systematically collected dataset from Chinese University, thereby extending the global relevance of such models to popu-

lations where evidence is scarce and resource limitations are acute. After adjustment for multiple comparisons, performance differences among top-performing models (XG-Boost, neural network, and random forest) were modest, indicating broadly comparable diagnostic accuracy among ensemble and deep learning approaches.

The clinical implications of this work are significant. First, by relying on routine blood tests, our approach provides a low-cost, non-invasive strategy for identifying women with a high predicted probability of active breast cancer. Unlike mammography or MRI, which require costly infrastructure and skilled operators, blood-based testing is widely accessible even in rural and resource-limited settings [40]. ML models trained on these data could therefore serve as an equitable complement to imaging-based screening. Second, such models may support clinical triage by identifying individuals with a higher predicted probability of active breast cancer who may warrant further diagnostic evaluation, including appropriate imaging or other confirmatory investigations. This targeted approach may help optimize resource allocation, especially in LMICs where radiological resources are scarce [41]. Third, the interpretability of our models, enabled by SHAP analysis, provides clinicians with a transparent rationale for diagnostic predictions. For example, if a patient is classified as having a high probability of active breast cancer based on features such as elevated PDW, advanced age, and CRP, this information may improve the interpretability of the model output and support subsequent diagnostic evaluation [39]. Finally, incorporating comorbidities such as diabetes and hypertension into the models reflects the interconnectedness of metabolic and oncologic health. Addressing these comorbidities through preventive care may indirectly reduce cancer risk, offering a dual benefit [39]. Sensitivity analyses using multiple imputations yielded results consistent with the complete-case analyses, indicating that missing data did not materially influence model performance or conclusions. Recent studies published between 2021 and 2025 have increasingly emphasized the value of integrating non-imaging data into breast cancer diagnostics [38,39]. For example, large-scale electronic health record (EHR)-based models have demonstrated AUCs ranging from 0.80 to 0.88 when combining clinical variables with laboratory markers, comparable to the performance observed in our integrated models. Unlike imaging-centric approaches, these strategies prioritize scalability and accessibility, particularly relevant for low- and middle-income settings. Our findings extend this literature by demonstrating that routinely available haematological indices alone can provide meaningful diagnostic signals when analysed with interpretable machine learning frameworks. Because the objective was prediction rather than causal inference, formal confounder adjustment was not performed; however, inclusion of known risk factors in all models mitigates major confounding effects.

Compared with traditional screening and risk assessment tools, ML approaches offer several advantages. Classical models such as the Gail score rely on a narrow set of variables and predefined linear assumptions, often underperforming in heterogeneous populations [42]. In contrast, ML can learn nonlinear and high-dimensional patterns, capturing subtle interactions between variables that might otherwise go unnoticed. Our findings show that ensemble and deep learning methods achieved substantially higher AUC values than logistic regression, demonstrating the added value of these approaches. Unlike imaging-based methods, which are costly and resource-intensive, our models utilize existing laboratory and clinical data, making them highly scalable. Furthermore, interpretability tools such as SHAP address one of the most common criticisms of ML—that models function as opaque “black boxes”—by providing transparent explanations of predictions [19]. This balance of accuracy and interpretability positions ML as a strong complement, and in some cases alternative, to conventional approaches, particularly in resource-limited contexts.

Limitations

Despite these strengths, several limitations warrant consideration. First, as a retrospective observational study, our analysis is subject to potential biases in data collection. Although preprocessing and imputation were applied to handle missing values, residual bias cannot be ruled out. Second, the dataset was drawn from tertiary care centres, which may not fully represent rural populations or those with limited healthcare access. Broader multi-centre datasets are needed to ensure generalizability. Third, our models were based only on routinely available haematological and biochemical markers. While pragmatic, this excluded potentially informative molecular markers such as circulating tumor DNA, proteomic signatures, or metabolomic profiles. Incorporating such data in future studies could improve predictive accuracy [40]. Fourth, although SHAP analysis enhanced interpretability, ML models remain more complex than traditional regression models. Clinical acceptance will require further validation of reliability and usability in real-world workflows. Finally, our study was cross-sectional; prospective longitudinal studies are needed to determine whether these models can predict future cancer development or recurrence over time [39].

Looking ahead, several future directions emerge. External validation across diverse, multi-centre cohorts will be essential to establish generalizability, particularly in low- and middle-income settings. Incorporating molecular biomarkers such as genomic and proteomic data could further enhance predictive performance. Prospective studies should assess the clinical utility of these models in practice, including their ability to improve patient outcomes and optimize resource allocation [31,32,33]. Emerging approaches such as federated learning may enable secure,

privacy-preserving data sharing across institutions, facilitating larger training datasets while addressing data protection concerns [35]. Ultimately, the aim should be to translate these findings into deployable, user-friendly diagnostic classification tools that can support clinicians in early detection and promote equitable breast cancer care globally. Several limitations of this study should be acknowledged. First, the proposed machine learning models discriminate between existing breast cancer cases and non-cancer controls in a cross-sectional setting and do not estimate future breast cancer incidence or long-term risk. Second, although internal validation demonstrated robust performance, prospective validation in large, multi-center cohorts is required before clinical deployment. Third, the retrospective nature of the study and reliance on data from tertiary-care hospitals may limit generalizability to community-based or rural populations, where disease prevalence, healthcare access, and biomarker distributions may differ. These factors should be addressed in future studies to ensure equitable clinical translation. External validation using independent, multi-center cohorts was not performed and remains a critical next step. Consequently, generalizability beyond the current tertiary-care population should be interpreted with caution.

Importantly, this study is cross-sectional in nature and focuses on distinguishing existing breast cancer cases from non-cancer controls at the time of evaluation. As such, the models are designed for diagnostic classification rather than prospective prediction of incident breast cancer risk. Longitudinal cohort studies will be required to determine whether these biomarkers and models can predict future cancer development.

5. Conclusions

This study demonstrates the feasibility and value of integrating clinical characteristics with blood-based biochemical markers, using machine learning models, to enhance diagnostic classification and early identification of active breast cancer with routinely available clinical and biochemical data. By analysing a large cohort of 2000 women, we showed that ensemble methods such as XGBoost and Random Forest, along with neural networks, achieved superior predictive performance compared to traditional models, with accuracies exceeding 80% and AUC values approaching 0.86. Importantly, routinely collected biomarkers, including platelet distribution width, red cell distribution width, hemoglobin, and C-reactive protein, together with demographic and clinical factors such as age, menopausal status, and family history, emerged as important features for distinguishing active breast cancer cases from non-cancer controls. The ability to derive clinically relevant predictions from widely accessible and minimally invasive tests underscores the potential of this approach to complement existing imaging-based screening strategies, particularly in low- and middle-income countries where resource constraints limit access to mammography and MRI.

By facilitating the identification of individuals with a high predicted probability of active breast cancer who may benefit from further diagnostic evaluation, these models could potentially support more efficient clinical triage and timely access to confirmatory diagnostic procedures.

While this study highlights the promise of ML-driven, multimodal prediction models, broader validation in diverse populations, inclusion of additional molecular biomarkers, and prospective evaluation in clinical workflows will be critical for translation into practice. Ultimately, integrating artificial intelligence with routine clinical and laboratory data provides a scalable, equitable pathway to strengthen breast cancer early detection and advance precision public health. Overall, this study provides methodologically rigorous evidence that interpretable machine learning applied to routine clinical and biochemical data can support diagnostic classification of breast cancer, while highlighting the steps required for safe and generalizable clinical translation.

Abbreviations

AI, artificial intelligence; ANN, artificial neural network; AUC, area under the curve; BMI, body mass index; CRP, C-reactive protein; ctDNA, circulating tumor DNA; DT, decision tree; Hb, hemoglobin; HDL, high-density lipoprotein; LMICs, low- and middle-income countries; LR, logistic regression; MCHC, mean corpuscular hemoglobin concentration; ML, machine learning; MRI, magnetic resonance imaging; PDW, platelet distribution width; RDW, red cell distribution width; RF, random forest; ROC, receiver operating characteristic; SHAP, Shapley additive explanations; TP, true positive; TN, true negative; FP, false positive; FN, false negative; XGBoost, extreme gradient boosting.

Availability of Data and Materials

The data supporting the findings of this study can be obtained from the corresponding author upon request.

Author Contributions

YQL, DX are credited for conception or design of the work, clinical workload of machine learning, and the interpretation of data and manuscript editing. FX, DCQ, YPW are credited for literature search, manuscript writing, manuscript preparation, illustration of the figures and research work related to Epidemiology. KPC, YYL, JXZ are credited for overall supervision, collection of data, manuscript review, illustration review and peer review editing. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and have agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was approved by the Institutional Ethics Committee of Jilin Maternal and Child Hospital (Approval No.: 2025-005). Written informed consent for the use of anonymized clinical and laboratory data was obtained from all participants. Data handling adhered to the Declaration of Helsinki and relevant national data protection regulations to ensure confidentiality and ethical compliance.

Acknowledgment

Not applicable.

Funding

Jilin Provincial Health Science and Technology Capacity Improvement Project (NO.2023LC089).

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/CEOG48645>.

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer Journal for Clinicians*. 2021; 71: 209–249. <https://doi.org/10.3322/caac.21660>
- [2] Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. *The Lancet. Oncology*. 2012; 13: 790–801. [https://doi.org/10.1016/S1470-2045\(12\)70211-5](https://doi.org/10.1016/S1470-2045(12)70211-5)
- [3] Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet (London, England)*. 2018; 391: 1023–1075. [https://doi.org/10.1016/S0140-6736\(17\)33326-3](https://doi.org/10.1016/S0140-6736(17)33326-3)
- [4] Ghoncheh M, Momenimovahed Z, Salehiniya H. Epidemiology, Incidence and Mortality of Breast Cancer in Asia. *Asian Pacific Journal of Cancer Prevention : APJCP*. 2016; 17: 47–52. <https://doi.org/10.7314/apjcp.2016.17.s3.47>
- [5] Malvia S, Bagadi SA, Dubey US, Saxena S. Epidemiology of breast cancer in Indian women. *Asia-Pacific Journal of Clinical Oncology*. 2017; 13: 289–295. <https://doi.org/10.1111/ajco.12661>
- [6] Ose DJ, Mark B, Ocier K, Adediran E, Taylor B, Svoboda K, et al. Rural-Urban Disparities in Cancer Care-Analyzing Routinely Collected Patient-Reported Outcomes. A Cross-Sectional Study. *Cancer Medicine*. 2025; 14: e70437. <https://doi.org/10.1002/ca.m4.70437>
- [7] Thomas NS, Norris E. Understanding the delays in diagnosis of breast cancer in India using blogs and YouTube videos. *BMC Public Health*. 2025; 25: 2792. <https://doi.org/10.1186/s12889-025-24132-x>
- [8] Mann RM, Athanasiou A, Baltzer PAT, Camps-Herrero J, Clauser P, Fallenberg EM, et al. Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI). *European Radiology*. 2022; 32: 4036–4045. <https://doi.org/10.1007/s00330-022-08617-6>
- [9] Berg WA, Blume JD, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. *Radiology*. 2006; 241: 355–365. <https://doi.org/10.1148/radiol.2412051710>
- [10] Mann RM, Kuhl CK, Moy L. Contrast-enhanced MRI for breast cancer screening. *Journal of Magnetic Resonance Imaging : JMRI*. 2019; 50: 377–390. <https://doi.org/10.1002/jmri.26654>
- [11] Maimone S, Morozov AP, Wilhelm A, Robrahn I, Whitcomb TD, Lin KY, et al. Understanding patient anxiety and pain during initial image-guided breast biopsy. *Journal of Breast Imaging*. 2020; 2: 583–589. <https://doi.org/10.1093/jbi/wbaa072>
- [12] Allison KH, Reisch LM, Carney PA, Weaver DL, Schnitt SJ, O'Malley FP, et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. *Histopathology*. 2014; 65: 240–251. <https://doi.org/10.1111/his.12387>
- [13] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nature Medicine*. 2019; 25: 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- [14] Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics : a Review Publication of the Radiological Society of North America, Inc*. 2017; 37: 505–515. <https://doi.org/10.1148/rg.2017160130>
- [15] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019; 25: 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [16] Wang L. Mammography with deep learning for breast cancer detection. *Frontiers in Oncology*. 2024; 14: 1281922. <https://doi.org/10.3389/fonc.2024.1281922>
- [17] Delen D, Walker G, Kadam A. Predicting breast cancer survival: a comparison of three data mining methods. *Artificial Intelligence in Medicine*. 2005; 34: 113–127. <https://doi.org/10.1016/j.artmed.2004.07.002>
- [18] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*. 2007; 2: 59–77.
- [19] Araujo DC, Rocha BA, Gomes KB, da Silva DN, Ribeiro VM, Kohara MA, et al. Unlocking the complete blood count as a risk stratification tool for breast cancer using machine learning: a large-scale retrospective study. *Scientific Reports*. 2024; 14: 10841. <https://doi.org/10.1038/s41598-024-61215-y>
- [20] Maharjan N, Thapa N, Tu J. Blood-based Biomarkers for Early Diagnosis of Lung Cancer: A Review Article. *JNMA; Journal of the Nepal Medical Association*. 2020; 58: 519–524. <https://doi.org/10.31729/jnma.5023>
- [21] Łukasiewicz S, Czezelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. *Cancers*. 2021; 13: 4287. <https://doi.org/10.3390/cancers13174287>
- [22] Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002; 420: 860–867. <https://doi.org/10.1038/nature01322>
- [23] Obeagu EI. Inflammatory blood markers in breast cancer: a narrative review from early detection to therapy response. *Annals of Medicine and Surgery (2012)*. 2025; 87: 5906–5911. <https://doi.org/10.1097/MS9.0000000000003687>
- [24] Pan H, Shi C, Zhang Y, Zhong Z. Artificial intelligence-based classification of breast nodules: a quantitative morphological analysis of ultrasound images. *Quantitative Imaging in Medicine and Surgery*. 2024; 14: 3381–3392. <https://doi.org/10.21037/qims-23-1652>

- [25] Adeoye J, Akinshipo A, Koohi-Moghadam M, Thomson P, Su YX. Construction of machine learning-based models for cancer outcomes in low and lower-middle income countries: A scoping review. *Frontiers in Oncology*. 2022; 12: 976168. <https://doi.org/10.3389/fonc.2022.976168>
- [26] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nature Reviews. Cancer*. 2018; 18: 500–510. <https://doi.org/10.1038/s41568-018-0016-5>
- [27] Goh S, Goh RSJ, Chong B, Ng QX, Koh GCH, Ngiam KY, et al. Challenges in implementing artificial intelligence in breast cancer screening programs: systematic review and framework for safe adoption. *Journal of Medical Internet Research*. 2025; 27: e62941. <https://doi.org/10.2196/62941>
- [28] Sebastian AM, Peter D. Artificial Intelligence in Cancer Research: Trends, Challenges and Future Directions. *Life* (Basel, Switzerland). 2022; 12: 1991. <https://doi.org/10.3390/life12121991>
- [29] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017; 30: 4765–4774.
- [30] Balsari S, Fortenko A, Blaya JA, Gropper A, Jayaram M, Matthan R, et al. Reimagining health data exchange: an application programming interface-enabled roadmap for India. *Journal of medical internet research*. 2018; 20: e10725. <https://doi.org/10.2196/10725>
- [31] Lou MWC, Drummond AE, Swain CTV, Milne RL, English DR, Brown KA, et al. Linking Physical Activity to Breast Cancer via Inflammation, Part 2: The Effect of Inflammation on Breast Cancer Risk. *Cancer Epidemiology, Biomarkers & Prevention : a Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*. 2023; 32: 597–605. <https://doi.org/10.1158/1055-9965.EPI-22-0929>
- [32] Carriero A, Groenhoff L, Vologina E, Basile P, Albera M. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics* (Basel, Switzerland). 2024; 14: 848. <https://doi.org/10.3390/diagnostics14080848>
- [33] McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafi H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020; 577: 89–94. <https://doi.org/10.1038/s41586-019-1799-6>
- [34] Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Physics in Medicine & Biology*. 2017; 62(19):7714–7728. <https://doi.org/10.1088/1361-6560/aa82ec>
- [35] Cong C, Li X, Zhang C, Zhang J, Sun K, Liu L, et al. MRI-based breast cancer classification and localization by multiparametric feature extraction and combination using deep learning. *Journal of Magnetic Resonance Imaging*. 2024; 59: 148–161. <https://doi.org/10.1002/jmri.28713>
- [36] Van Belle SJP, Cocquyt V. Impact of haemoglobin levels on the outcome of cancers treated with chemotherapy. *Critical Reviews in Oncology/hematology*. 2003; 47: 1–11. [https://doi.org/10.1016/s1040-8428\(03\)00093-3](https://doi.org/10.1016/s1040-8428(03)00093-3)
- [37] 1005. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*. 2019; 292: 331–342. <https://doi.org/10.1148/radiol.2019182622>
- [38] Feinstein O, Ofer D, Bachmat E, Gazit S, Linial M, Menes TS. Short-term prediction model for breast cancer risk based on one million medical records. *Clinical Breast Cancer*. 2025; 25: 774–781.e4. <https://doi.org/10.1016/j.clbc.2025.07.025>
- [39] Arravalli T, Chadaga K, Muralikrishna H, Sampathila N, Cenitta D, Chadaga R, et al. Detection of breast cancer using machine learning and explainable artificial intelligence. *Scientific Reports*. 2025; 15: 26931. <https://doi.org/10.1038/s41598-025-12644-w>
- [40] Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in cancer detection, diagnosis, and prognosis. *Sensors*. 2024; 24: 37. <https://doi.org/10.3390/s24010037>
- [41] Zhao Y, Xiong P, McCullough LE, Miller EE, Li H, Huang Y, et al. Comparison of breast cancer risk predictive models and screening strategies for Chinese women. *Journal of Women's Health*. 2017; 26: 294–302. <https://doi.org/10.1089/jwh.2015.5692>
- [42] Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the Gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Research : BCR*. 2018; 20: 18. <https://doi.org/10.1186/s13058-018-0947-5>