Original Article

SEER-based machine learning prediction of bone metastasis in breast cancer: model development and validation

Ying Gao[1,2#], Lei Liu[3#], Shoujun Wang[1,4#], Weijie Tao[1], Jinmiao Wang[1,4], Ran Duan[1], Hai Xie[1], Hideaki Takahashi[5], Jie Hao[1], Ming Gao[1]

[1]Department of Breast and Thyroid Surgery, Tianjin Key Laboratory of General Surgery in Construction, Tianjin Union Medical Center, The First Affiliated Hospital of Nankai University, Tianjin, China; [2]Department of Thyroid and Neck Cancer, Tianjin Medical University Cancer Institute and Hospital, National Cancer Clinical Research Center, Tianjin Cancer Clinical Research Center, Tianjin Key Laboratory of Cancer Prevention and Treatment, Tianjin, China; [3]The Third Department of Breast Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Tianjin, China; [4]School of Medicine, Nankai University, Tianjin, China; [5]Department of Neurosurgery, Niigata Cancer Center Hospital, Niigata, Japan

Contributions: (I) Conception and design: Y Gao; (II) Administrative support: J Hao, M Gao; (III) Provision of study materials or patients: W Tao, J Wang; (IV) Collection and assembly of data: R Duan, H Xie; (V) Data analysis and interpretation: L Liu, S Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jie Hao, MD; Ming Gao, MD. Department of Breast and Thyroid Surgery, Tianjin Key Laboratory of General Surgery in Construction, Tianjin Union Medical Center, The First Affiliated Hospital of Nankai University, No. 190 Jieyuan Road, Hongqiao District, Tianjin 300121, China. Email: haojie1215@126.com; gaomingtmu@163.com.

Background: Breast cancer (BC) is the leading cancer in women. It often metastasizes to bone, worsening the prognosis. Diagnostic methods often fail to predict bone metastasis (BM). This study developed a machine learning (ML) model using the Surveillance, Epidemiology, and End Results (SEER) database for BM prediction, to refine treatments and improve outcomes.

Methods: Using SEER data, we studied 24,584 BC patients diagnosed 2010–2015 with radiologically confirmed BM. Tumor size, grade, tumor (T)/node (N) stages, and estrogen receptor (ER)/progesterone receptor (PR)/human epidermal growth factor receptor 2 (HER2) status were assessed. Stratified randomization divided the data into 70% training (n=18,438) and 30% validation (n=6,146). Six ML algorithms were developed, emphasizing random forest (RF). Receiver operating characteristic (ROC) curve analysis [area under the curve (AUC), sensitivity, specificity, negative predictive value (NPV)] assessed performance. The SHapley Additive exPlanations (SHAP) framework identified key BM predictors.

Results: Our analysis of 24,584 patients identified 1,298 (5.26%) patients with BM. Logistic regression (LR) provided the highest specificity [0.897, 95% confidence interval (CI): 0.889–0.905], contrasting with gradient boosting machine (GBM)'s maximal sensitivity (0.658, 95% CI: 0.609–0.707). With sensitivity at 0.658, better algorithms or multimodal methods are needed for case identification. The multilayer perceptron neural network (MLPNN) model demonstrated superior performance, with the highest AUC of 0.808 (95% CI: 0.798–0.818), surpassing the LR and adaptive boosting (AdaBoost) models, both with AUCs of 0.803 (95% CI: 0.793–0.813). The RF model was particularly adept at ruling out BM, with an NPV above 97%. The SHAP analysis identified tumor size, grade, T/N stages, ER/PR/HER2 status, and brain/liver/lung metastases as key predictors for risk stratification. Decision curve analysis showed RF's superior utility over the American Joint Committee on Cancer (AJCC) Staging System.

Conclusions: Our ML model demonstrates potential for predicting BM in patients with BC. It may serve as a clinical aid to identify at-risk individuals early. However, moderate sensitivity requires refinement for better case detection. This study supports integrating ML into clinical practice, advancing personalized oncology medicine.

Keywords: Breast cancer (BC); bone metastasis (BM); machine learning (ML); prediction model; prognosis

Submitted Jun 10, 2025. Accepted for publication Jul 23, 2025. Published online Jul 28, 2025.
doi: 10.21037/gs-2025-168
View this article at: https://dx.doi.org/10.21037/gs-2025-168

Introduction

Breast cancer (BC) has solidified its status as the most commonly diagnosed cancer among women globally (1) and is the leading cause of cancer-related death in women (1). This malignancy, driven by the unbridled growth of breast epithelial cells in response to a myriad of carcinogenic factors, poses a substantial threat to both quality of life and survival (2). Although the initial signs of BC can be as inconspicuous as a breast lump or other seemingly innocuous symptoms, the disease's ability to metastasize is a serious threat to health. In its later stages, BC can spread to distant sites within the body (3), resulting in multi-organ damage and significantly worsening the prognosis of those afflicted.

Bone metastasis (BM) is the most common diagnosis in patients with *de novo* stage IV BC, occurring in approximately 70% to 80% of patients (4,5). The development of BM is particularly debilitating, often resulting in severe pain, pathological fractures, and compromised quality of life. Moreover, it presents a significant clinical challenge due to the lack of effective early detection methods.

Traditional diagnostic modalities, including imaging techniques and histological assessments, have provided valuable insights into the diagnosis and management of BC. Medical imaging facilitates the early identification of suspicious breast lesions. In malignancy cases, clinicians develop integrated treatment and follow-up plans guided by pathology results. However, these methods are often unable to detect micrometastatic disease and are inherently subjective, limiting the efficacy of diagnosis (6). Although biomarkers have demonstrated a degree of value in aiding diagnosis, the inherent heterogeneity of BC precludes the development of a universally effective biomarker-based strategy (7). This complexity has prompted the exploration of alternative approaches for improving the prediction of metastasis.

Machine learning (ML) has transformed medical diagnostics by identifying patterns in complex datasets (8-10). However, its application to BM prediction in BC has been limited by small datasets or specific patient subsets. For example, Li *et al.* developed a deep learning model using magnetic resonance imaging (MRI)-derived radiomic features in 96 metastatic and 192 non-metastatic BC cases, achieving good area under the curve (AUC) but requiring specialized imaging (11). Thio *et al.* used multi-omics data (biochemical, hematological, and protease assays) to build a prognostic model, but its reliance on extensive laboratory data limits clinical applicability (12). These studies highlight the need for generalizable, interpretable models.

The complexity and heterogeneity of BC highlight the need for advanced predictive tools to identify patients at risk of BM. Existing models, limited by dataset size, specific imaging requirements, or lack of interpretability, underscore the demand for robust, generalizable approaches. This study aims to develop and validate ML models using a large, representative dataset to predict BM in BC patients, enhancing clinical decision-making through interpretable algorithms. We present this article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-2025-168/rc).

Highlight box

Key findings

- The machine learning (ML) models developed provide a method for predicting bone metastasis (BM) in patients with breast cancer (BC), with a high negative predictive value (NPV).

What is known and what is new?

- BM is the most common diagnosis in patients with *de novo* stage IV BC, but lacks effective early detection methods.
- This study introduces ML models using a large Surveillance, Epidemiology, and End Results (SEER) dataset, with interpretable predictions via the SHapley Additive exPlanations (SHAP) framework, improving clinical utility.

What is the implication, and what should change now?

- The models' high NPV can aid clinicians in ruling out BM, while moderate sensitivity suggests the need for further refinement. Integration of these models into clinical practice may enhance risk stratification and guide preemptive treatment strategies.

Methods

Study population

We analyzed data from the Surveillance, Epidemiology, and End Results (SEER) database. We extracted the data using SEER*Stat software (version 8.4.3). Our study included 24,584 BC patients. These patients were diagnosed between 2010 and 2015. We selected them to represent diverse geographic regions across the US. We collected various clinical and pathological variables. These included age, race, sex, and marital status. They also included laterality and histological grade. Tumor size was recorded. Estrogen receptor (ER) status, progesterone receptor (PR) status, and human epidermal growth factor receptor 2 (HER2) status were included. Tumor (T) stage and node (N) stage were also collected. The collected clinical data were suitable for ML applications. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

Diagnosis of BM

BM was diagnosed by imaging. Board-certified radiologists confirmed the diagnosis. They identified skeletal lesions visible on scans. These lesions showed typical features. Features included irregular margins. Cortical destruction was another feature. Pathological fractures were also present. Lesions appeared sclerotic, lytic, or mixed.

Data preparation and splitting

Age was divided into two groups according to a 60-year cutoff (<60 and ≥60 years), while tumor diameter was similarly divided according to a 2-cm cutoff (≤2 and >2 cm). Hormone receptor status (ER/PR) and HER2 expression were identified via immunohistochemical evaluation, with positivity thresholds (≥1%) determined as per the current American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines (13). For detailed information on variable coding, please refer to Table S1.

The cohort was divided into 70% training (n=18,438) and 30% validation (n=6,146) sets using stratified randomization (seed =42) to preserve the BM incidence rate (5.26%). This preserved the original BM incidence rate (5.26% across the full cohort) in both subsets, mitigating class imbalance bias during model training. The stratified sampling method ensured a proportional representation of positive cases in each split, enhancing reliability in imbalanced learning scenarios. To ensure consistent replicability, a predetermined randomization seed [42] was applied during dataset division. We acknowledged the presence of incomplete data as a limitation inherent to retrospective studies. Numerical variables underwent feature scaling through the Z-score standardization method, a process that centers variables by their mean and scales them by standard deviation. This normalization technique ensures consistent measurement units across features while mitigating distributional skewness to optimize model performance.

Screening for risk factors and model construction

To assess the importance of each feature in predicting BM, we employed the Boruta-Shap method, which consists of a two-stage process. Initially, this approach introduces random "shadow" features, akin to red herrings, and utilizes a classifier to rank all features based on their importance index, with default SHapley Additive exPlanations (SHAP) values being used. This ranking served as our initial lead. Subsequently, statistical tests were performed to evaluate each feature's significance until all features were distinctly categorized, allowing us to identify the most significant predictors.

ML model development

We developed ML models to predict BM, including logistic regression (LR) (14), random forest (RF) (15), support vector machine (SVM) (16,17), adaptive boosting (AdaBoost), multilayer perceptron neural network (MLPNN) (18), extreme gradient boosting (XGBoost) (19), and gradient boosting machine (GBM) (20). Model parameters were optimized to enhance predictive accuracy. The specific parameters of the relevant model can be found in Appendix 1.

LR was applied to estimate the survival probability based on patient characteristics via statistical methods. The RF models created multiple decision trees and aggregated their predictions. SVM was used due to its ability to process both continuous and categorical variables. SVM works by establishing a separating boundary between different classes of data. AdaBoost was effective in boosting the performance of basic classifiers, especially in tasks involving imbalanced datasets or those with a moderate level of complexity. Additionally, the MLPNN learned by adjusting the connections between its layers of neurons based on the

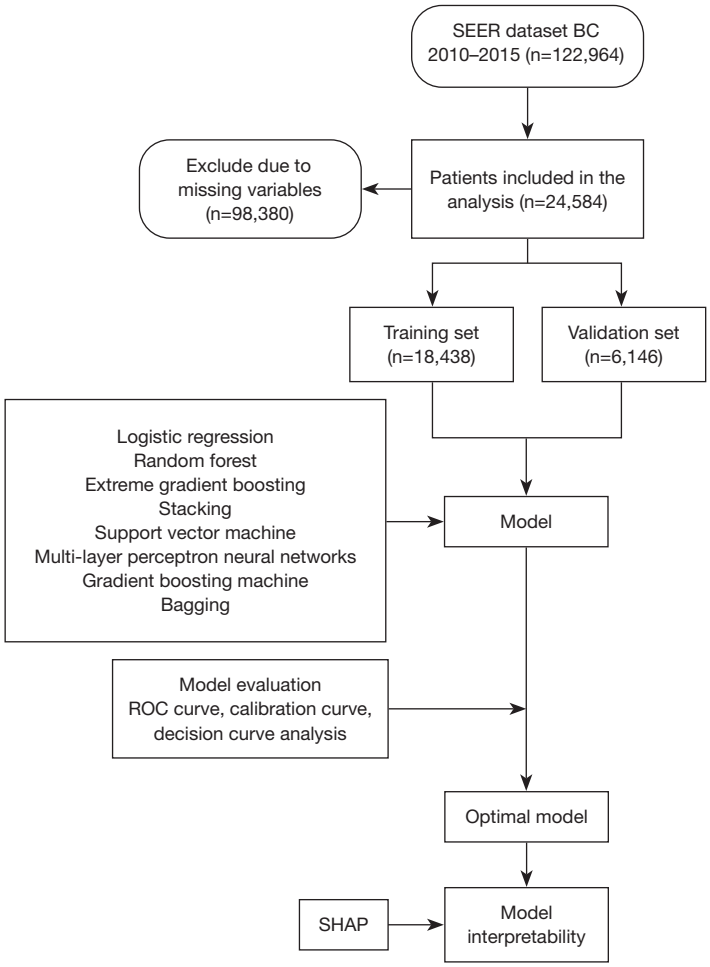

Figure 1 Flowchart participant inclusion. BC, breast cancer; ROC, receiver operating characteristic; SEER, Surveillance, Epidemiology, and End Results; SHAP, SHapley Additive exPlanations.

data it processes. The GBM constructed trees sequentially, with each new tree correcting the errors of its predecessor. Together, these models offered a deeper understanding of the complex relationships between patient characteristics and survival outcomes, providing valuable insights for prognosis. Appendix 2 provides a streamlined model to predict bone metastasis risk in specific individuals.

Statistical analysis

Model performance was evaluated using ROC curves, precision-recall curves, calibration curves, and decision curve analysis. Sensitivity was calculated as true positive (TP)/[TP + false negative (FN)]. Specificity was calculated as true negative (TN)/[TN + false positive (FP)]. Thresholds ≥0.80 for sensitivity and specificity indicate model robustness in oncology (21). All figures, tables, and underlying data were generated and cross-verified using SEER*Stat, Python, and R.

Results

Clinical and pathological characteristics

Data from a total of 122,964 BC cases were extracted from the SEER registry [2010–2015]. Following the exclusion of 98,380 cases with missing data, 24,584 patients met the inclusion criteria for final analysis (*Figure 1*). In our retrospective analysis, we examined the medical records of 24,584 patients with BC, identifying multiple clinical and pathological characteristics. Many of these patients had grade 2 tumors, which constituted 47.7% of our cohort.

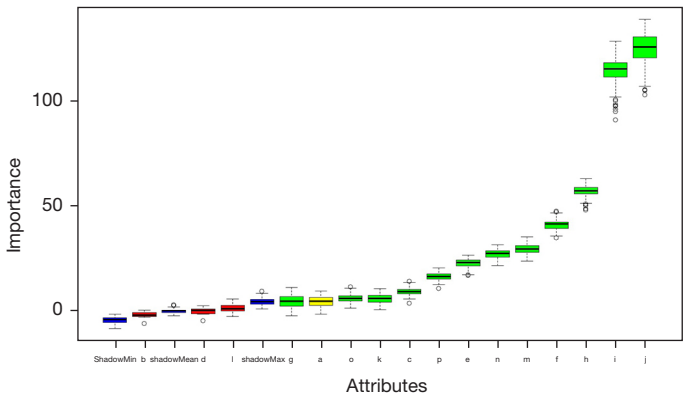

Figure 2 Significant features identified by boruta-shap algorithm for bone metastasis prediction in breast cancer. The x-axis lists the attributes, which include ShadowMin, ShadowMean, ShadowMax (reference or threshold value), a: age characteristics; b: race characteristics; c: patient sex characteristics; d: lateralization characteristics; e: tumor grade characteristics; f: T stage characteristics; g: N stage characteristics; h: brain metastasis characteristics; i: liver metastasis characteristics; j: lung metastasis characteristics; k: tumor multifocality characteristics; l: marital status characteristics; m: estrogen receptor status characteristics; n: progesterone receptor status characteristics; o: hormone receptor status characteristics; p: tumor size characteristics. The y-axis, labeled as a percentage, represents the importance of each predictor, with 100% indicating the maximum importance. Each color in the box plots signifies the effect size, where red indicates significantly lower importance compared to shadow features are systematically removed, blue represents shadow features (negative controls), green is accepted for significantly higher importance than shadow features—retain, yellow indicates features with importance close to shadow features, requiring further iteration. The box plots provide a visual summary of the distribution of importance scores for each attribute across the patient cohort, highlighting the relative influence of each clinical factor on the prediction of bone metastasis.

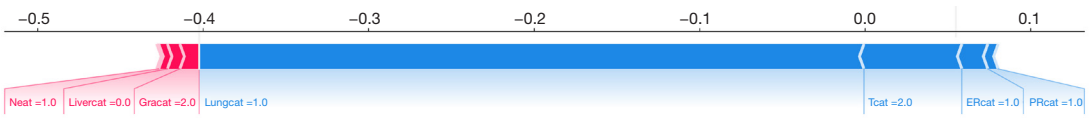

Figure 3 Force plot analysis of feature contributions to bone metastasis in breast cancer. Each color in the box plots signifies the effect size, where red indicates a negative effect and blue indicates a positive effect. N1 status, absence of liver metastases, and grade 2 classification were identified as risk factors promoting bone metastasis. In contrast, lung metastasis, T2 stage, ER positivity, and PR positivity acted as inhibitory factors against bone metastasis. ERcat, estrogen receptor status; Gracat, tumor grade; Livercat, liver metastasis; Lungcat, lung metastasis; Ncat, node stage; PRcat, progesterone receptor status; Tcat, tumor stage.

Distant metastasis occurred in 7.46% of the cohort, with BM occurring in 5.26%, lung metastasis in 2.81%, liver metastasis in 2.46%, and brain metastasis in 0.51%. These findings highlight the aggressive nature of BC and its propensity to metastasize.

Variable filtering

Our examination of variables that could predict BM in patients with BC led us to employ Boruta-Shap analysis. This analysis revealed that sex, tumor size, multifocality status, histological grade, T stage, N stage, ER status, PR status, HER2 status, and the presence of brain, liver, and lung metastasis were significant predictors for BM in patients with BC. *Figure 2* displays these predictors, where each color represents the effect size, with red indicating a negative effect and blue indicating a positive effect. *Figure 3* indicates that a histological grade of 2 was associated with a reduced likelihood of BM. To ensure that our models were trained and tested on distinct datasets, we divided our cohort into training (70%) and validation (30%) subsets. A statistical analysis revealed no significant differences in clinical or pathological characteristics between these two groups, as detailed in *Table 1*.

Table 1 The demographic and clinical characteristics of patients with breast cancer in the training and validation cohorts

Variable	Training cohort (n=18,438), n (%)	Validation cohort (n=6,146), n (%)	χ^2	P value
Age (years)			5.563	0.02
<60	10,233 (55.5)	3,517 (57.22)		
≥60	8,205 (44.5)	2,629 (42.78)		
Race			0.184	0.67
White	13,698 (74.29)	4,549 (74.02)		
Non-White	4,740 (25.71)	1,597 (25.98)		
Sex			0.765	0.38
Male	230 (1.25)	68 (1.11)		
Female	18,208 (98.75)	6,078 (98.89)		
Histological grade			0.110	0.74
1/2	9,657 (52.38)	3,204 (52.13)		
3	8,781 (47.62)	2,942 (47.87)		
T stage			0.087	0.77
1/2	15,067 (81.72)	5,012 (81.55)		
3/4	3,371 (18.28)	1,134 (18.45)		
N stage			0.323	0.57
1/2	16,885 (91.58)	5,614 (91.34)		
3	1,553 (8.42)	532 (8.66)		
Brain metastasis			0.603	0.44
No	18,348 (99.51)	6,111 (99.43)		
Yes	90 (0.49)	35 (0.57)		
Liver metastasis			0.733	0.39
No	17,994 (97.59)	5,986 (97.4)		
Yes	444 (2.41)	160 (2.6)		
Lung metastasis			0.679	0.41
No	17,929 (97.24)	5,964 (97.04)		
Yes	509 (2.76)	182 (2.96)		
Multifocality			1.865	0.17
No	14,537 (78.84)	4,795 (78.02)		
Yes	3,901 (21.16)	1,351 (21.98)		
ER status			0.648	0.42
Positive	14,424 (78.23)	4,838 (78.72)		
Negative	4,014 (21.77)	1,308 (21.28)		
PR status			0.085	0.77
Positive	12,455 (67.55)	4,164 (67.75)		
Negative	5,983 (32.45)	1,982 (32.25)		

Table 1 (*continued*)

Table 1 (*continued*)

Variable	Training cohort (n=18,438), n (%)	Validation cohort (n=6,146), n (%)	χ^2	P value
HER2 status			0.644	0.42
Positive	3,999 (21.69)	1,363 (22.18)		
Negative	14,439 (78.31)	4,783 (77.82)		
Tumor size (cm)			1.167	0.28
≤2	16,521 (89.6)	5,477 (89.11)		
>2	1,917 (10.4)	669 (10.89)		
BM			2.667	0.10
No	17,493 (94.87)	5,798 (94.34)		
Yes	945 (5.13)	348 (5.66)		

AJCC, American Joint Committee on Cancer; BM, bone metastasis; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; N, node; PR, progesterone receptor; T, tumor.

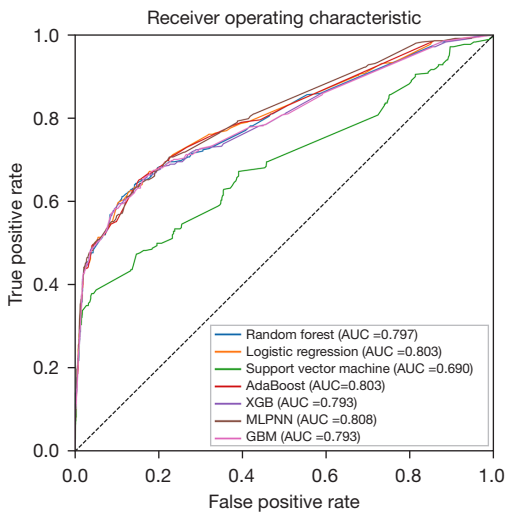

Figure 4 Receiver operating characteristic curves for the diagnostic performance of the machine learning model. AdaBoost, adaptive boosting; AUC, area under the curve; GBM, gradient boosting machine; MLPNN, multilayer perceptron neural network; XGBoost, extreme gradient boosting.

Predictive performance of ML models

In our evaluation of predictive performance, we analyzed six ML models using their receiver operating characteristic (ROC) curves, which are depicted in *Figure 4*. The MLPNN model demonstrated the highest performance, with an AUC of 0.808, surpassing the LR model, which had an AUC of 0.803. The RF model followed with an AUC of 0.797, while the SVM model showed a comparatively lower AUC of 0.690. Other models, such as AdaBoost and GBM, performed well with AUCs of 0.803 and 0.793, respectively, and the XGBoost model achieved an AUC of 0.793. These results highlight the competitive predictive performance of the MLPNN model. All models except for SVM exhibited a negative predictive value (NPV) above 97%. The SVM model had an NPV of 0.963, as summarized in *Table 2*. In summary, the ROC curves (*Figure 4*) present a competitive predictive performance landscape with several models showing strong discriminative abilities.

Precision-recall curves

Precision-recall curves (*Figure 5*) were employed to gain a nuanced view of the models' performance, particularly in the context of an imbalanced dataset. These curves highlighted the tradeoff between precision and recall for each model, with the RF model demonstrating a superior balance between these two metrics. The RF model was able to accurately identify patients with BM while minimizing false positives.

Calibration curves

To further assess the clinical utility of our models, we examined the calibration curves, as shown in *Figure 6*. These curves measured how well the predicted probabilities aligned with the actual outcomes, serving as a critical test of a model's reliability. The RF model once again stood out, showing a close alignment between predicted probabilities

Table 2 Algorithmic performance palette: a curated spectrum of ML algorithms' performances, showing their predictive capabilities in the context of BC

Model	LR	RF	XGBoost	SVM	MLPNN	GBM	AdaBoost
Cut-off	0.0456	0.0512	0.0502	0.0376	0.0385	0.0410	0.462
Training_AUC	0.818	0.845	0.837	0.715	0.815	0.836	0.821
Testing AUC	0.803	0.797	0.793	0.69	0.808	0.797	0.803
AUC (95% CI)	0.793–0.813	0.787–0.807	0.783–0.803	0.678–0.702	0.798–0.818	0.787–0.807	0.793–0.813
Specificity (95% CI)	0.897 (0.889–0.905)	0.887 (0.879–0.895)	0.838 (0.828–0.847)	0.959 (0.954–0.964)	0.857 (0.848–0.866)	0.833 (0.823–0.843)	0.846 (0.837–0.855)
Sensitivity (95% CI)	0.598 (0.546–0.649)	0.612 (0.56–0.664)	0.652 (0.601–0.701)	0.379 (0.328–0.431)	0.632 (0.58–0.681)	0.658 (0.609–0.707)	0.652 (0.601–0.701)
F1	0.36	0.35	0.299	0.369	0.314	0.296	0.309
Youden index	0.494	0.499	0.49	0.339	0.489	0.491	0.498
MCC	0.338	0.331	0.289	0.33	0.3	0.287	0.299
Kappa	0.305	0.293	0.233	0.33	0.251	0.229	0.244
NPV (95% CI)	0.974 (0.970–0.978)	0.974 (0.970–0.978)	0.976 (0.972–0.980)	0.963 (0.958–0.968)	0.975 (0.971–0.979)	0.976 (0.972–0.980)	0.976 (0.972–0.980)
PPV (95% CI)	0.257 (0.228–0.287)	0.245 (0.217–0.274)	0.194 (0.171–0.217)	0.359 (0.310–0.408)	0.209 (0.184–0.234)	0.191 (0.169–0.214)	0.203 (0.180–0.227)
PLR	5.776	5.418	4.019	9.319	4.406	3.941	4.24
NLR	0.449	0.437	0.415	0.647	0.429	0.41	0.411
mAP	0.683	0.683	0.682	0.626	0.688	0.682	0.677

This table features metrics (Train_AUC and Test_AUC), which show a model's ability to distinguish between cases with and without metastasis in the training and testing phases. AUC_CI indicates the reliability of these AUC scores. Specificity and sensitivity, along with their confidence intervals, measure a model's accuracy in identifying patients without and with BM, respectively. The F1 score and Youden index reflect the model's balance between precision and sensitivity, while MCC and Kappa assess the agreement between predicted and observed outcomes. NPV and PPV, with their intervals, show the model's effectiveness in ruling out and confirming metastasis. PLR and NLR explain how test results affect the likelihood of disease presence or absence. mAP provides an overall measure of precision across different stages. AdaBoost, adaptive boosting; AUC, area under the curve; BC, breast cancer; CI, confidence interval; GBM, gradient boosting machine; LR, logistic regression; mAP, mean average precision; MCC, Matthews correlation coefficient; ML, machine learning; MLPNN, multilayer perceptron neural network; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PPV, positive predictive value; RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

and observed outcomes, indicating its robustness in real-world settings.

Decision curve analysis

Decision curve analysis was conducted to clarify the clinical impact of the models, the results of which are depicted in *Figure* 7. We found that the RF model offered greater clinical value as compared to the traditional American Joint Committee on Cancer (AJCC) staging system, particularly in scenarios where the threshold probability for treatment was moderate to high. This not only validated our approach but also supports its potential to dramatically influence the nature of clinical decision-making.

Feature importance in ML models

Finally, we sought to determine the importance of each feature in our predictive model. The SHAP diagram, a visual representation of feature importance, revealed that nine variables significantly influenced the model's predictions. *Figure 8* illustrates how higher SHAP values correlate with an increased probability of BM in patients with BC; blue indicates low eigenvalues, purple indicates

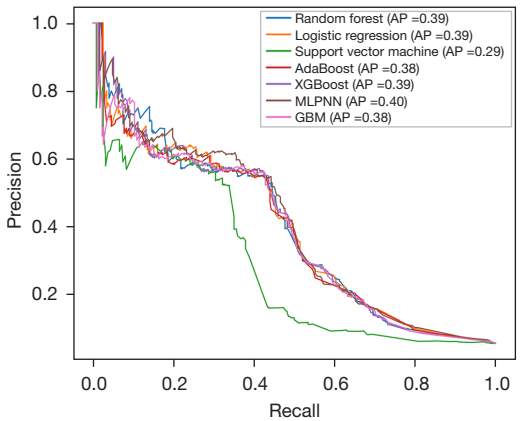

Figure 5 Precision-recall curves for the diagnostic models' performance in balancing true-positive rates against false-positive rates. AdaBoost, adaptive boosting; AP, average precision; GBM, gradient boosting machine; MLPNN, multilayer perceptron neural network; XGBoost, extreme gradient boosting.

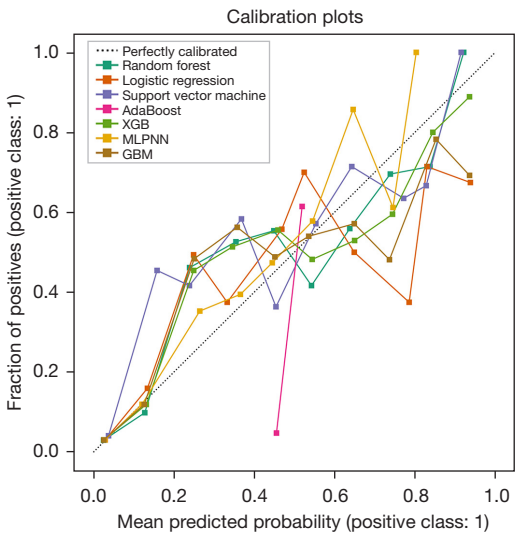

Figure 6 Calibration curves indicating the alignment of predicted probabilities with observed outcomes. AdaBoost, adaptive boosting; GBM, gradient boosting machine; MLPNN, multilayer perceptron neural network; XGBoost, extreme gradient boosting.

values near the mean, and red indicates high values. Notably, the presence of lung metastasis emerged as the most influential feature.

Discussion

BC is the most prevalent cancer among women globally

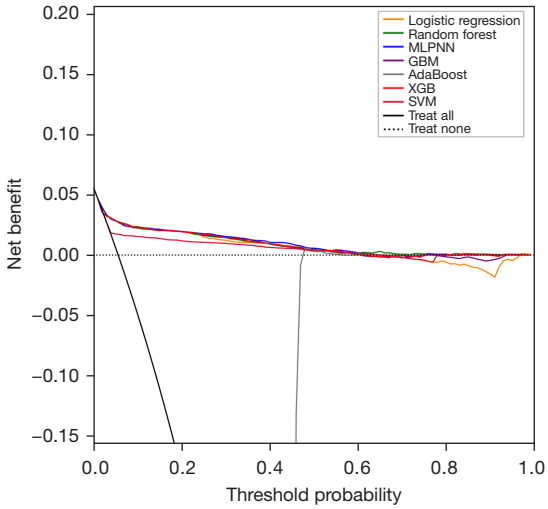

Figure 7 A decision curve analysis comparing the comparative clinical utility between the machine learning models and the traditional staging systems. AdaBoost, adaptive boosting; GBM, gradient boosting machine; MLPNN, multilayer perceptron neural network; SVM, support vector machine; XGBoost, extreme gradient boosting.

and poses a significant risk to public health. Despite remarkable progress in early detection and treatment, the threat of metastasis continues to undermine the optimism in the prognosis of those affected (1). Particularly alarming is the occurrence of BM, which impacts a substantial proportion of patients diagnosed with *de novo* stage IV BC. This condition, characterized by serious complications such as bone pain, pathological fractures, and spinal cord compression (collectively known as skeletal-related events), severely diminishes patients' quality of life and significantly lowers survival rates (22). Our study demonstrated that ML models, particularly the MLPNN and RF, can achieve high predictive accuracy for BM in patients with BC. The MLPNN model's AUC of 0.808 and the RF model's NPV >97% support their utility in clinical settings.

Additionally, our study's application of the Boruta-SHAP framework represents a significant leap forward in predictive analytics. This robust feature selection method can outperform traditional approaches, as it effectively sifts through a myriad of variables to identify key predictors of BM in BC. We found that sex, tumor size, multifocality status, histological grade, T stage, N stage, ER status, PR status, HER2 status, and the presence of distant metastases in other organs were significant risk factors for BM.

The association of lymph node involvement, increased

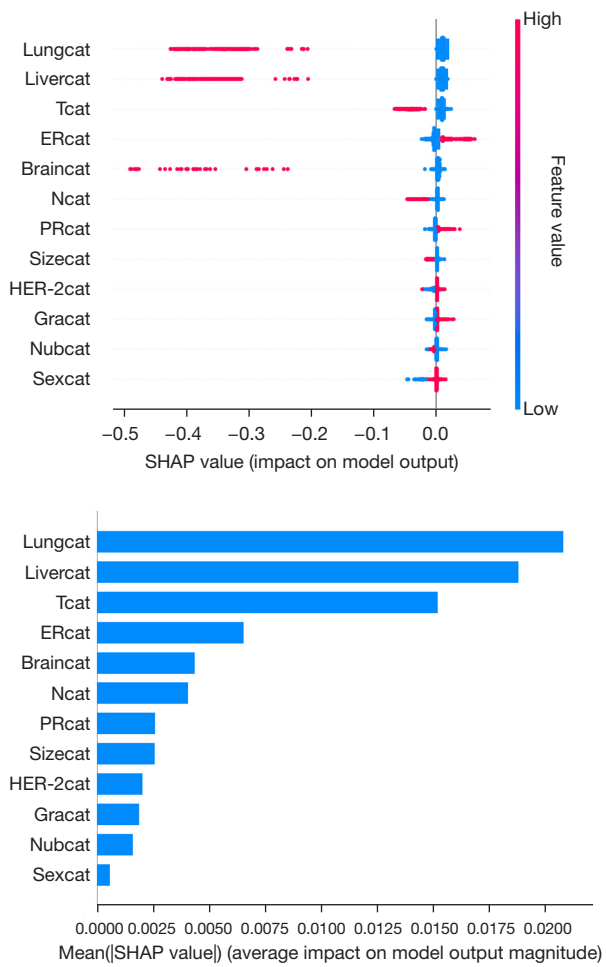

Figure 8 Feature importance of the XGBoost model for predicting bone metastasis in breast cancer. The upper panel presents the features according to the sum of SHAP values, with a greater intensity in color signifying greater predictive impact. The bar chart in the bottom panel ranks features by their average absolute SHAP values, highlighting the most influential predictors. The y-axis lists the attributes. Braincat, brain metastasis; ERcat, estrogen receptor status characteristics; Gracat, tumor grade characteristics; HER2cat, human epidermal growth factor receptor 2 status characteristics; Latcat, lateralization characteristics; Livercat, liver metastasis characteristics; Lungcat, lung metastasis characteristics; Marcat, marital status characteristics; Ncat, node stage characteristics; Nubcat, tumor multifocality characteristics; Tcat, tumor stage characteristics; PRcat, progesterone receptor status characteristics; Sexcat, patient sex characteristics; ShadowMin, reference or threshold value; Sizecat, tumor size characteristics. SHAP, SHapley Additive exPlanations; XGBoost, extreme gradient boosting.

tumor size, and reduced histopathological differentiation with the risk of metastasis is well-established in the literature (23). Our analysis supports these observations, aligning with the research conducted by Knapp *et al.* (24), which identified advanced T stage and N2 or N3 classifications as significant predictors of BM. Additionally, our study expands upon these findings by emphasizing the predictive significance of metastases to the brain, liver, and lungs, thereby illuminating the intricate relationships among these variables and their influence on the likelihood of BM. Moreover, our results are consistent with prior population-based studies that have indicated a higher occurrence of BM in patients diagnosed at more advanced cancer stages (25,26). These investigations, carried out in Denmark and the United Kingdom, underscore the importance of early and precise diagnosis in enhancing patient outcomes. By recognizing patients who are at an elevated risk of metastasis, earlier interventions can be implemented, facilitating a more tailored and proactive treatment strategy.

Using the Boruta-SHAP approach, we found that lung, liver, and brain metastases are highly significant predictors of the incidence of BM in patients with BC. These findings are especially noteworthy since they emphasize the importance of assessing several metastatic sites to assess the risk of BM. This discovery aligns with the molecular processes that underlie metastasis, such as the epithelial-mesenchymal transition (EMT) process, which entails the loss of intercellular cohesiveness and epithelial cell polarity, a crucial stage in the development of metastasis in BC (27). The complex character of BC is further highlighted by the role that immune cells and the tumor microenvironment play in promoting the spread of the illness. These mechanisms contribute to the varied metastatic patterns observed among different biological types of BC, further emphasizing the complexity of metastasis in BC.

The disease's heterogeneity makes identifying metastatic predictors for BC difficult. However, our study addressed this complication by demonstrating that the $HR^+/HER2^-$ subtype is most prevalent in patients with BM, aligning with evidence that luminal-type BC cells exhibit heightened metastatic propensity (28,29). Among BC subtypes, ER^+ tumors exhibit the highest propensity for BM. In patients with metastatic ER^+ BC, BM occur in up to 87% of cases, compared to <56% in those with ER^- disease (30,31). Interestingly, past research has also demonstrated that a collection of gene expression signals known as the Src-

responsive signature (SRS) is strongly linked to late-stage BM, indicating that luminal-type cancer cells may flourish in the bone microenvironment (32).

The advent of ML models has revolutionized the prediction of survival and lymph node metastasis in patients with BC (33-35). Our study diverges from the existing body of work in this field by illuminating the "black box" of conventional ML models (36,37) through the SHAP framework. This innovative approach enabled us to visualize the importance of features and clarify their influence on outcomes, thereby enhancing the interpretability and clinical applicability of our models.

Moreover, our introduction of ML models marks a significant advancement in predicting BM in patients with BC. Our analysis revealed that the MLPNN model demonstrated the highest predictive accuracy, yielding an AUC of 0.808. The RF model also showed strong performance with an AUC of 0.797, effectively identifying patients with BM. Although the SVM model had a comparatively lower AUC, it exhibited a high specificity of 0.959, demonstrating a markedly lower false-positive rate in detecting BM and minimizing overdiagnosis. These models offer actionable insights for clinicians, allowing for the early identification of high-risk patients and thus proactive interventions such as bone-protective therapies (e.g., denosumab) or intensified monitoring for skeletal-related events; meanwhile, the model's risk stratification can aid in personalized treatment planning, allowing clinicians to tailor follow-up imaging or systemic therapies based on individual risk. Finally, the SHAP analysis identified key predictors (e.g., tumor size and ER/PR/HER2 status) that clinicians can use to counsel patients regarding prognosis and management options, aligning with the growing emphasis on precision medicine in oncology.

Despite our study offering valuable insights into factors associated with distant metastasis in patients with BC, it is not without its limitations. We employed a retrospective design based primarily on the SEER database, which predominantly represents the US population. Moreover, we did not conduct external validation with an independent dataset. External validation is key to confirming how well predictive models perform in diverse, real-world settings. We fully acknowledge that the current model's sensitivity (0.65) has room for improvement, which will be a key focus of our future research. The suboptimal sensitivity may be attributed to the following factors: class imbalance: the training dataset contained only 15.53% positive samples, potentially biasing the model toward the negative class; insufficient feature representation: the current dataset, derived from the SEER database, may not fully capture critical biological signals associated with the disease. Future research should assess a variety of populations, thereby creating a more comprehensive and universally applicable survival prediction model. Finally, incorporating a broader array of predictive factors related to BM could further enhance the accuracy of future prediction models. We remain committed to refining and enhancing our model in subsequent studies.

Conclusions

Using SEER data and ML, this study explored key predictors of BM in BC patients. Several clinical factors and biomarkers emerged as meaningful indicators of metastasis risk. The RF model performed better than traditional approaches, suggesting potential for improved oncology predictions. This tool may help clinicians make more informed treatment decisions. To address data imbalances, Boruta-SHAP techniques enhanced model reliability. These findings support data-driven identification of high-risk patients, which could guide tailored therapies. Overall, our ML model offers valuable insights into BM prediction, presenting possibilities for clinical practice and future research.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-2025-168/rc

Peer Review File: Available at https://gs.amegroups.com/article/view/10.21037/gs-2025-168/prf

Funding: This study was supported by the Tianjin Health Science and Technology Project (No. TJWJ2024MS020), Tianjin Medical Key Discipline (Specialty) Construction Project (No. TJYXZDXK-058B), the Tianjin Municipal Science and Technology Project (No. 21JCZDJC00360), the Tianjin Health Research Project (No. TJWJ2022XK024), the Tianjin Education Commission Key Project (No. 2023YXZD05), and the National Natural Science

Foundation of China (No. 82372753).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-2025-168/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57-70.
3. Chambers AF, Groom AC, MacDonald IC. Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2002;2:563-72.
4. Kuchuk I, Hutton B, Moretto P, et al. Incidence, consequences and treatment of bone metastases in breast cancer patients-Experience from a single cancer centre. J Bone Oncol 2013;2:137-44.
5. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res 2006;12:6243s-9s.
6. Barrios CH. Global challenges in breast cancer detection and treatment. Breast 2022;62 Suppl 1:S3-s6.
7. Pasha N, Turner NC. Understanding and overcoming tumor heterogeneity in metastatic breast cancer treatment. Nat Cancer 2021;2:680-92.
8. Diz J, Marreiros G, Freitas A. Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. J Med Syst 2016;40:203.
9. Darcy AM, Louie AK, Roberts LW. Machine Learning and the Profession of Medicine. JAMA 2016;315:551-2.
10. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016;375:1216-9.
11. Li L, Tian H, Zhang B, et al. Prediction for Distant Metastasis of Breast Cancer Using Dynamic Contrast-Enhanced Magnetic Resonance Imaging Images under Deep Learning. Comput Intell Neurosci 2022;2022:6126061.
12. Thio Q, Karhade AV, Bindels BJJ, et al. Development and Internal Validation of Machine Learning Algorithms for Preoperative Survival Prediction of Extremity Metastatic Disease. Clin Orthop Relat Res 2020;478:322-33.
13. Wolff AC, Hammond MEH, Allison KH, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. J Clin Oncol 2018;36:2105-22.
14. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Ann Surg Oncol 2018;25:1783-5.
15. Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb) 2014;24:12-8.
16. Freitas P, Silva F, Sousa JV, et al. Machine learning-based approaches for cancer prediction using microbiome data. Sci Rep 2023;13:11821.
17. Huang S, Cai N, Pacheco PP, et al. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics 2018;15:41-51.
18. Hakami MA. Harnessing machine learning potential for personalised drug design and overcoming drug resistance. J Drug Target 2024;32:918-30.
19. Ramkumar M, Shanmugaraja P, Anusuya V, et al. Identifying cancer risks using spectral subset feature selection based on multi-layer perception neural network for premature treatment. Comput Methods Biomech Biomed Engin 2024;27:1804-16.
20. Dianati-Nasab M, Salimifard K, Mohammadi R, et al. Machine learning algorithms to uncover risk factors of breast cancer: insights from a large case-control study. Front Oncol 2023;13:1276232.
21. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453-73.
22. Zhang H, Zhu W, Biskup E, et al. Incidence, risk factors and prognostic characteristics of bone metastases and

skeletal-related events (SREs) in breast cancer patients: A systematic review of the real world data. J Bone Oncol 2018;11:38-50.
23. Gerratana L, Fanotto V, Bonotto M, et al. Pattern of metastasis and outcome in patients with breast cancer. Clin Exp Metastasis 2015;32:125-33.
24. Knapp BJ, Cittolin-Santos GF, Flanagan ME, et al. Incidence and risk factors for bone metastases at presentation in solid tumors. Front Oncol 2024;14:1392667.
25. Jensen A, Jacobsen JB, Nørgaard M, et al. Incidence of bone metastases and skeletal-related events in breast cancer patients: a population-based cohort study in Denmark. BMC Cancer 2011;11:29.
26. Hagberg KW, Taylor A, Hernandez RK, et al. Incidence of bone metastases in breast cancer patients in the United Kingdom: results of a multi-database linkage study using the general practice research database. Cancer Epidemiol 2013;37:240-6.
27. Brabletz T, Kalluri R, Nieto MA, et al. EMT in cancer. Nat Rev Cancer 2018;18:128-34.
28. Shi D, Bai J, Chen Y, et al. Predicting the Incidence and Prognosis of Bone Metastatic Breast Cancer: A SEER-Based Observational Study. Biomed Res Int 2020;2020:1068202.
29. Xiao W, Zheng S, Yang A, et al. Breast cancer subtypes and the risk of distant metastasis at initial diagnosis: a population-based study. Cancer Manag Res 2018;10:5329-38.
30. Savci-Heijink CD, Halfwerk H, Hooijer GK, et al. Retrospective analysis of metastatic behaviour of breast cancer subtypes. Breast Cancer Res Treat 2015;150:547-57.
31. Wu K, Feng J, Lyu F, et al. Exosomal miR-19a and IBSP cooperate to induce osteolytic bone metastasis of estrogen receptor-positive breast cancer. Nat Commun 2021;12:5196.
32. Clements ME, Johnson RW. Breast Cancer Dormancy in Bone. Curr Osteoporos Rep 2019;17:353-61.
33. Jiang C, Xiu Y, Qiao K, et al. Prediction of lymph node metastasis in patients with breast invasive micropapillary carcinoma based on machine learning and SHapley Additive exPlanations framework. Front Oncol 2022;12:981059.
34. Tahmassebi A, Wengert GJ, Helbich TH, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients. Invest Radiol 2019;54:110-7.
35.vLi J, Zhou Z, Dong J, et al. Predicting breast cancer 5-year survival using machine learning: A systematic review. PLoS One 2021;16:e0250370.
36. Zhou CM, Xue Q, Wang Y, et al. Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer. Surg Today 2021;51:756-63.
37. Lo Gullo R, Eskreis-Winkler S, Morris EA, et al. Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. Breast 2020;49:115-22.

Cite this article as: Gao Y, Liu L, Wang S, Tao W, Wang J, Duan R, Xie H, Takahashi H, Hao J, Gao M. SEER-based machine learning prediction of bone metastasis in breast cancer: model development and validation. Gland Surg 2025;14(7):1366-1378. doi: 10.21037/gs-2025-168