Nopour *BMC Cancer* (2025) 25:940
https://doi.org/10.1186/s12885-025-14369-5

BMC Cancer

RESEARCH Open Access

Prediction of one-year recurrence among breast cancer patients undergone surgery using artificial intelligence-based algorithms: a retrospective study on prognostic factors

Raoof Nopour[1]*

Abstract

Background and aim Breast cancer is highly prevalent, with an increasing trend in women globally. Although the survival of breast cancer is relatively high, the recurrence rate is also high, demanding effective predictive solutions to breast cancer prognosis among post-operative patients. So far, Artificial intelligence algorithms integrated with various clinical data have demonstrated potential predictive capability regarding breast cancer recurrence.

Objective This study aims to specifically conduct a predictive analysis of one-year recurrence of breast cancer by comparing and analyzing different machine learning and deep learning algorithms trained by structural prognostic data.

Materials and methods This retrospective study was carried out using one database, including 1156 post-operative breast cancer data from 30 January 2020 to 30 December 2022, in three clinical centers in Tehran City. The inclusion criteria were patients who had undergone at least one surgery, had at least one year of medical records, and did not have other conditions. The patients who were diagnosed with malignant BC and had undergone adjuvant therapies without surgery were excluded from the study. Twenty-three prognostic factors were utilized to train algorithms to establish prediction models for the one-year recurrence of breast cancer. The data were analyzed using univariate and adjusted correlation-based methods and chosen machine learning and deep learning algorithms. The discrimination, calibration, and clinical utility were leveraged to assess the algorithms' performance efficiency. The SHapley Additive exPlanations plot was generated to identify the prominent prognostic factors affecting the one-year recurrence of breast cancer.

Results Totally, 445 relapsed and 711 non-relapsed cases were utilized in this study. Our empirical study showed that the random forest with a positive predictive value of 0.96, negative predictive value of 0.92, sensitivity of 0.92, specificity of 0.96, accuracy of 0.94, F-score of 0.94, area under the receiver operator characteristics curve of 0.919 was the best-performing model for predicting the breast cancer recurrence. As the analysis of SHapley Additive

*Correspondence:
Raoof Nopour
nopour.r70@gmail.com; nopour.r@iums.ac.ir

Full list of author information is available at the end of the article

exPlanations indicated, the tumor grade, HER-2, and the number of lymph nodes involved were more significant predictors.

Conclusion The current study demonstrated the potential predictive power of the random forest for early predicting tumors among breast cancer patients who have undergone surgery and its utility in enhancing decision-making in clinical environments. It is crucial in promoting the prognosis, more effectively choosing therapies, augmenting post-operative breast cancer patients' survival, and controlling the limited healthcare resources.

Clinical trial number Not applicable.

Keywords Artificial intelligence, Tumor recurrence, Breast cancer, Prognosis, Clinical utility

Introduction

Breast cancer (BC) is one of the most common cancers among the female population worldwide, with the highest mortality rate compared to other malignancies, and is continuously increasing globally [1]. It is estimated that about 2.3 million new cases are identified annually regarding BC [2]. This malignancy is the fifth most common cause of cancer death, accounting for 6.6% of all cancer deaths worldwide [3]. As a significant public health dilemma across all countries, this malignancy is more prevalent in Western nations than in Eastern ones [4]. However, women in low-income and middle-income countries are experiencing a sharply rising trend due to adopting Westernized dietary practices, increased longevity, and changing lifestyles [5]. Global trends regarding BC incidence and death across 185 countries have indicated that BC will experience a 38% and 68% increase associated with new cases and mortality rates, respectively, by 2050 [6].

Early diagnosis and efficient prognosis of BC significantly improve the prognosis and increase the likelihood of recovery; thus, it will help choose a more effective treatment strategy and decrease mortality rates [7, 8]. Also, manual prognosis and diagnosis of BC take many hours, even for a clinical specialist, so establishing one model with better predictive strength is advantageous for patients. Moreover, it is beneficial in eliminating the toxic side effects and additional medical costs associated with unnecessary treatments [9].

Although the five-year BC survival is relatively high, the recurrence rate is also high (roughly 20–30%, depending on the stage) [10]. Approximately 30% of patients who initially manifest the early stage of BC will experience a recurrence following ten years [11]. Accurate risk assessment of BC recurrence in post-operative patients can significantly enhance their prognosis by choosing appropriate treatment, planning follow-up measures, promoting patient health, and using medical resources [10, 12].

Artificial intelligence (AI) has exhibited potential predictive efficiencies in oncology, such as risk evaluation, timely diagnosis, patient prognosis accuracy, and treatment plans based on deep knowledge obtained from the data [13]. AI-based algorithms, including machine learning (ML) and deep learning (DL), are crucial in BC diagnosis and prognosis through developing prediction models in various research tasks in BC, including risk forecasting, probability of tumor malignancy, detecting subtypes or staging, response to adjuvant treatments, and patient progression-free survival (PFS) or overall survival (OS) [14]. More specifically, they have appeared in different applications in BC detection with satisfactory predictive accuracy using biosensors [15], routine clinical [16], multi-omics [17], and image data [18]. AI models aiming to forecast BC recurrence were recognized as a supportive tool in clinical environments for assessing follow-up visits to timely interventions and enhanced therapies to lower BC mortality [19–23].

To our knowledge, limited studies have specifically examined the one-year BC recurrence among patients who have undergone surgery by leveraging ML and DL models combined with structural data on prognostic factors. This study aims to establish prediction models for the one-year BC recurrence by comparing and analyzing various ML and DL models to achieve novel predictive insight into BC prognosis in this period.

Methods

The current research was according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) AI statement [24] (Supplementary A).

Study population and database

This retrospective study aimed to establish prediction models using AI algorithms for one-year tumor recurrence among BC patients who had undergone surgery. To achieve this, we leveraged one electronic Excel data sheet (.xlsx format), including the data of 1156 post-operative BC patients from 30 January 2020 to 30 December 2022, belonging to three clinical centers in Tehran City. Seven hundred and eleven and 445 cases were associated with non-relapsed and relapsed BC cases following one year of performing the surgery, respectively. The inclusion criteria were patients who had undergone at least one surgery to clear tissue from tumors, had a minimum of one year of medical records in those centers, and did not have other conditions affecting BC recurrence. Conversely, the patients who were diagnosed with malignant BC and had undergone adjuvant therapies without surgery were excluded from the study. The process of arriving at the study sample size leveraged based on the eligibility criteria is presented in Fig. 1.

Input features and outcome variable

The input features of the current database leveraged for the model construction included age, smoking, alcohol, residence, ethnicity, obesity, use of contraceptive pills, physical activity, number of other chronic conditions, premature menstruation, absence of pregnancy, personal history of BC, family history of BC, family history of different cancers, tumor size, estrogen receptor, progesterone receptor, HER-2, number of lymph nodes involved, tumor grade, surgery type, radiotherapy, chemotherapy, hormonotherapy, targeted therapy. The outcome variable was the status of BC recurrence after one year among patients who had undergone surgery. This variable assesses the relapse status among these patients, including relapsed tumor and non-relapsed conditions, which were allotted with 1 and 0 codes in the database, respectively.

Preprocess of dataset

We reviewed the current database in multiple steps to enhance its quality for analysis. First, we investigated the database regarding the redundancy. The redundancy in this step was defined as having more than one row allocated to one patient, and it was registered by a similar ID. We removed all redundancies affecting the database quality and might increase analysis bias. Second, any issues regarding noisy data were solved using statistical methods. Also, the invalid values for qualitative data were eliminated from the database by finding the actual data from the record as possible; otherwise, we removed this value and embedded it using the mode of values in the corresponding feature. Third, we handled the missing data in two conditions. If the missing values constituted more than 10% of the corresponding record's data, we removed them from further analysis; otherwise, in the condition that we couldn't find the actual value, we imputed them using the average and mode values of the corresponding feature considering the quantitative and qualitative data, respectively.

Feature selection

Feature selection (FS) is crucial in decreasing the data dimensionality and improving framework performance in ML [25]. Some objectives of FS techniques include eliminating irrelevant and noisy features, lowering the computational time, and enhancing the sophistication of training and testing classifiers, which results in more cost-effective models [26, 27]. It also improves algorithms' performance in learning, avoids overfitting, and

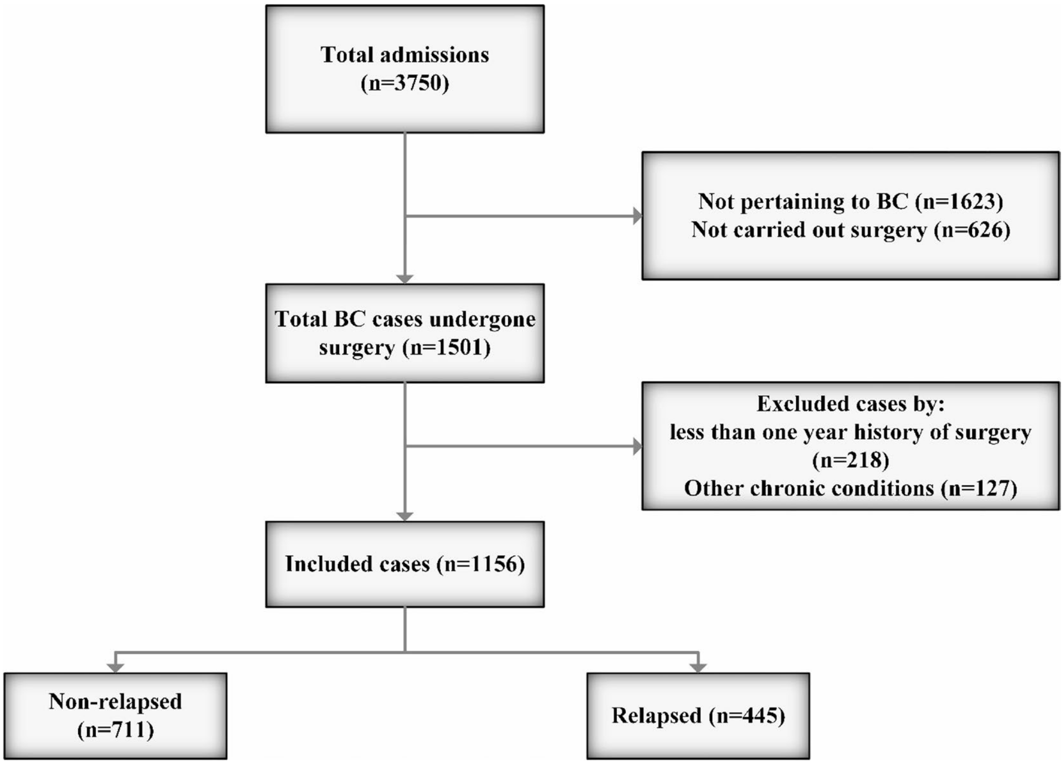

Fig. 1 The flowchart of the sample selection process

assists in establishing more generalized models [27]. To achieve this aim in the current study, we leveraged univariate (eta and Chi-square) and adjusted (logistic regression) correlation-based analysis. The adjusted regression analysis implies the estimation of the outcome variable through the input variable with the presence of at least one covariate. The variables obtained significant correlation with the outcome class (one-year BC recurrence) in univariate and adjusted analyses were considered for model construction. $P < 0.05$ was considered a statistically significant level to select more competent prognostic factors regarding one-year BC recurrence.

Data balance solving

This study utilized 1156 post-operative patients, including 711 and 445 BC cases about non-recurrent and recurrent cases, respectively. This proportion of cases belonging to different class labels causes an imbalance during training models, resulting in biased accuracy. In this condition, the classifier would forecast the new cases favoring the majority class (as the non-recurrent cases). To address this challenge, we leveraged the Synthetic Minority Oversampling Technique (SMOTE). This technique has been specially developed to tackle the imbalanced dataset using synthetic samples from minority cases. The fundamental idea behind the SMOTE as an oversampling technique is to generate synthetic samples in the feature space formed by the corresponding sample and its K-nearest neighbors due to its potential to impede overfitting and assist the classifier in finding decision thresholds between classes [28]. By mitigating bias and achieving highlighted characteristics of the minority class, SMOTE leads to more accurate forecasting performance in ML models [29]. It should be noted that before more profound analysis in the FS and establishing prediction models for predicting one-year BC recurrence, the number of minority cases was augmented to 702 cases (number of non-relapsed cases after preprocessing) by using the SMOTE technique to mitigate the risk of bias in the performance results. This way, the number of recurrent cases has risen from 438 to 702.

Parameter adjustment, hold-out, and K fold cross-validation

We employed GridSearch Cross-validation (CV) to establish prediction models. It is considered a robust and fine-tuning technique for the algorithms' training and evaluation, which plays a crucial role in model optimization. As an automatic hyperparameter adjustment, GridSearch CV allows a systematic exploration through prespecified hyperparameter values to identify the more competent hyperparameter combinations, contributing to a more favorable forecasting capability [30, 31].

The hold-out method randomly divides the data into two independent parts: the train and test sets. Typically, two-thirds or 70 of the original data is used for the algorithm's training, and one-third or 30% of the data is considered to assess the accuracy of the trained algorithm [32, 33]. Although the estimate is pessimistic in this situation, with one portion of the initial data to establish the model, training algorithms with K-fold CV for hyperparameter tuning would enhance the performance efficiency.

Also, this study used the K-fold CV method to assess the performance during model establishment. This way, all the dataset samples are randomly divided into K equal sections, namely folds. (K-1) folds are leveraged to train algorithms and establishment purposes, and one fold is used to test the algorithms' performance. This process would be iterated in K epochs, and the total performance of algorithms is calculated as the mean of performance obtained in K times training of the algorithms. The current study applied ten-fold CV as an extensively used and superior validation technique in ML research due to the distinguished and efficient characteristics compared to the conventional dividing sample method. This method facilitates mitigating bias in forecasting errors and enhances data leverage in training and validation processes without overfitting or intervening in testing data with validation data. It also advocates against the theory of testing established by arbitrarily split data [34].

Model development

Seven chosen ML and DL algorithms were employed to establish prediction models for the one-year BC recurrence. ML algorithms leveraged included Random Forest (RF), eXtreme Gradient Boosting (XGB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). In contrast, DL algorithms consist of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Deep Neural Networks (DNN). The ML model would provide a more favorable predictive performance efficiency and enhanced explainability, while DL models favor eliciting more sophisticated patterns from the tabular data. The proposed algorithms utilized simultaneously give us more predictive insights and the most profound knowledge on one-year BC relapse. For example, some research indicated that although CNNs are efficient in favorable prediction efficiency when leveraging image data, they were effective on tabular data by learning local dependencies and patterns in tabular datasets, similar to identifying image attributes [35]. In more detail, the specific architecture of CNN, namely 1D-CNN, can provide favorable predictive efficiency when we use tabular data for the training algorithm, despite 2D-CNN converting the input data shape, which hinders its efficiency regarding the tabular data. So, 2D-CNN requires some steps to

Table 1 Confusion matrix: TP: true positive (relapsed cases correctly classified as relapsed by model), TN: true negative (non-relapsed cases correctly classified by model), FN: false negative: relapsed cases incorrectly classified as non-relapsed cases by model), FP: false positive: non-relapsed cases incorrectly classified as relapsed cases by model)

Output		**Predicted values**	
		Relapsed (+)	**Non-relapsed (-)**
Real values	(+) Relapsed	TP	FN
	(-) Non-relapsed	FP	TN

Table 2 Performance criteria of AI models

Assessment metrics	**Formula**
PPV	$\frac{TP}{TP+FP}$
NPV	$\frac{TN}{TN+FN}$
Specificity	$\frac{TN}{TN+FP}$
Sensitivity	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+FN+FP+TN}$
F-Score	$\frac{TP}{TP+\frac{1}{2}(FN+FP)}$

transform data when we deal with input data of tabular datasets. Leveraging 1D-CNN with LIME or SHAP values would give robust predictive performance XAI when we utilize tabular datasets [36]. Although LSTM is most commonly used for time-series data, they have indicated efficacy in achieving sequential patterns from the tabular data. It optimizes predictive performance, especially for medical data, by targeting features such as tumor recurrence, which are not dynamic data but possess intuitive consecutive associations [37, 38]. Moreover, in other research, some specific configurations of LSTM algorithms, such as Stacked Bidirectional LSTM, obtained satisfactory performance when the tabular datasets were considered for training the algorithm. This way, the minimal conversion of the tabular data is needed for this architecture, and it requires some characteristics such as batch size and features to train the algorithm [37].

Performance analysis of established models

We used a variety of performance criteria to comprehensively assess the predictive efficiency of AI-based algorithms to predict the one-year BC recurrence. The confusion-based (Tables 1 and 2) metrics, including Positive Predictive Value (PPV), Negative Predictive Value (NPV), sensitivity, specificity, accuracy, and F-score, were employed to analyze the capability of models' performance. Moreover, the Receiver Operator Characteristics (ROC) of models were assessed to gauge the models' potentiality to predict the positive and negative cases against each other in various thresholds. The calibration of AI-based models was investigated using the calibration curve plot to evaluate the reliability of the estimated risks. Decision-analytic matrices are set up to abridge the model's performance in underpinning decision-making. To assess the clinical utility of the AI-based models established for clinical decision-making, we generated the Decision Curve Analysis (DCA) plot. It consisted of Net Benefit (NB) across clinically sensible risk thresholds, displaying the oscillations between benefit (true positives) and cost (false positives) in various threshold probabilities among a range of reasonable thresholds [39].

Explainability of AI-based models

Although AI algorithms have been seen in many tasks with remarkable performance efficiency, the inherent sophistication of the algorithms puts AI in the black box. Indeed, it hinders the extensive adoption of healthcare decision-making due to the lack of transparent, accountable, and understandable algorithms for humans [40, 41]. Explainable Artificial Intelligence (XAI) is incrementally cited in scientific publications and is essential in improving trust in AI models established in medicine and healthcare [42, 43]. They assist clinical professionals in adopting the advantages of AI in their tasks without any previous bias occurring due to the lack of interpretability [44].

As a game-theory approach, SHAP (SHapley Additive exPlanations) by Lundberg and Lee (2017) is a state-of-the-art FS mechanism. It interprets each model's prediction by introducing individual features' contribution to the model's outcome and estimates the Shapley values [45, 46]. In the current study, we leveraged the SHAP values to indicate the XAI and predict the one-year BC recurrence. It exhibits more transparent ML models with robust predictive performance. Furthermore, integrating the DL models with SHAP would provide a more profound knowledge of the data with interpretability.

Ethical considerations

The ethical committee board of Tehran University of Medical Sciences (TUMS) approved the study (Ethics code: IR.TUMS.SPH.REC.1398.401). To protect the privacy and confidentiality of patients, the unique identification information of patients was concealed during all steps of the study. All methods of the present study followed the relevant guidelines and regulations. Informed consent was obtained from all subjects and/or their legal guardian(s). Also, there was no patient or public involvement during the study's design, conduct, reporting, interpretation, or dissemination.

Results

Preprocess of database and sample characteristics

After investigating the database regarding duplicates, five cases, including three relapsed and two non-relapsed duplicate cases, were identified and excluded. Thirty-five instances, consisting of 10 noisy and 25 missing data (low rate), were handled using statistical methods and utilized for analysis. Eleven cases, including four relapsed and seven non-relapsed cases, with more than 10% missing values, were excluded from the further analysis. In total, 1140 cases, consisting of 702 non-relapsed and 438 relapsed cases, remained for analysis. The characteristics of the data leveraged in the study for relapsed and non-relapsed cases are presented in Table 3.

Feature selection

The results of the univariate and adjusted correlation-based analyses for one-year relapse are presented in Table 4. As shown in Table 4, age, smoking, alcohol, obesity, use of the contraceptive pill, physical activity, number of other chronic conditions, premature menstruation, absence of pregnancy, personal history of BC, family history of BC, family history of different cancers, tumor size, estrogen receptor, progesterone receptor, HER-2, number of lymph nodes involved, tumor grade, surgery type, radiotherapy, chemotherapy, hormonotherapy, and targeted therapy were considered more significant prognostic factors for one-year BC relapse.

Optimal parameters using hyperparameter adjustment

Table 5 displays the AI-based algorithms with optimized parameters from the hyperparameter adjustment using the Gridsearch CV. Moreover, the ranges of hyperparameter space search to achieve the higher accuracy algorithms are presented in Table 6.

Performance evaluation of AI-based models

Cross-validation

Table 7 indicates the performance efficiency of the chosen AI-based models for predicting the one-year BC recurrence using various performance metrics.

As given in Table 7, CNN, with a PPV of 0.80, NPV of 0.82, sensitivity of 0.82, specificity of 0.80, accuracy of 0.81, and F-score of 0.81, exhibited more predictive performance for one-year BC relapse compared to all DL models. Moreover, RF with a PPV of 0.96, NPV of 0.92, sensitivity of 0.92, specificity of 0.96, accuracy of 0.94, and F-score of 0.94 was superior to other ML models based on all performance criteria investigated. Comparing the DL and ML models indicated that RF, as an ensemble model, attained more performance competency than others in predicting the one-year BC relapse. DL models with performance metrics values between 0.7 and 0.82 yielded relatively satisfactory prediction performance. Conversely, base ML models, including KNN and SVM, obtained performance criteria between 0.61 and 0.72, indicating the lowest performance than other ML and DL models. Despite RF being considered the best-performing model based on these performance criteria, the DL models exhibited relatively good to favorable prediction performance efficiency with less variability than ML models.

Hold-out method

The performance measurements of AI algorithms in train and test modes by the hold-out method are presented in Tables 8 and 9, respectively.

As shown in Tables 8 and 9, RF with a PPV of 0.98, NPV of 0.96, sensitivity of 0.96, specificity of 0.98, accuracy of 0.97, and F-score of 0.97 in train mode and with a PPV of 0.89, NPV of 0.88, sensitivity of 0.88, specificity of 0.89, accuracy of 0.88, and F-score of 0.88 in test mode was considered the best-performing model in two modes. Conversely, the SVM with a PPV of 0.66, NPV of 0.69, sensitivity of 0.71, specificity of 0.64, accuracy of 0.67, and F-score of 0.68 in train mode, and with a PPV of 0.58, NPV of 0.58, sensitivity of 0,59, specificity of 0.57, accuracy of 0.58, and F-score of 0.58 in test mode attained lower performance than other AI algorithms. Except for SVM, other algorithms obtained favorable prediction performance (all criteria more than 0.7). With all performance criteria of 0.76, the CNN algorithm obtained satisfactory predictive performance for BC recurrence in the test mode. In general, all algorithms have experienced less than 10% performance decrease in the test rather than train mode, indicating a low risk of overfitting in training algorithms.

The ROC curve plots of the chosen ML and DL models for predicting one-year tumor recurrence are illustrated in Figs. 2 and 3, respectively. (horizontal and vertical vertices indicate 1-specificity and sensitivity, respectively). Based on the ROC plots, among all ML models, RF attained higher AUC, indicating higher prediction performance regarding relapsed and non-relapsed cases in various discrimination thresholds. In contrast, the CNN obtained higher performance competency across all DL models. Comparing the AU-ROC of ML and DL models demonstrated the eligibility of the RF model as the best-performing model for predicting tumor recurrence among BC patients who have undergone surgery. Also, investigating Figs. 4 and 5 confirms the more reliable prediction performance attributed to the RF model based on the calibration curve plot (more similar to the ideal curve).

The NB curves of the DCA for the AI-based models are displayed in Fig. 6. In line with the ROC plot analysis, RF has superior NB compared to other DL and ML models, showing a more competent clinical utility than others.

Table 3 The characteristics of the samples used for analysis

Feature	Value	Total cases ($n = 1140$)	Non-relapsed ($n = 702$)	Relapsed ($n = 438$)	$P_{-\text{value}}$
Age (year)	< 40	214	128	86	< 0.001
	40–50	323	241	82	
	50–60	346	215	131	
	> 60	257	118	139	
Residence	Rural	559	320	239	0.036
	Urban	581	382	199	
Ethnicity	Persian	798	463	335	0.014
	Non-Persian	342	239	103	
Smoking	Have	58	15	43	< 0.001
	Haven't	1082	687	395	
Alcohol	Have	106	48	58	< 0.001
	Haven't	1034	654	380	
Obesity	Have	429	271	158	0.089
	Haven't	711	431	280	
Use of contraceptive pills	Have	157	68	89	< 0.001
	Haven't	983	634	349	
Physical activity	Have	483	224	259	< 0.001
	Haven't	657	478	179	
Number of other chronic conditions	0	389	253	136	0.052
	1	421	276	145	
	> 1	330	173	157	
Premature menstruation	Have	237	98	139	< 0.001
	Haven't	903	604	299	
Absence of pregnancy	Have	56	24	32	0.047
	Haven't	1084	678	406	
Personal history of BC	Have	25	7	18	< 0.001
	Haven't	1115	695	420	
Family history of BC	Have	194	92	102	< 0.001
	Haven't	946	610	336	
Family history of other cancers	Have	517	265	252	0.023
	Haven't	623	437	186	
Tumor size (cm)	< 2	285	196	89	< 0.001
	2–5	570	425	145	
	> 5	222	53	169	
	Invasion to the skin and chest	63	28	35	
Estrogen receptor	Positive	650	366	284	0.017
	Negative	490	336	154	
Progesterone receptor	Positive	763	439	324	0.001
	Negative	377	263	114	
HER-2	Positive	380	135	245	< 0.001
	Negative	760	567	193	
Number of lymph nodes involved	0	364	282	82	< 0.001
	< 3	410	288	122	
	3–10	262	83	179	
	> 10	104	49	55	
Tumor grade	I	170	138	32	0.036
	II	627	388	239	
	III	343	176	167	
Surgery type	Lumpectomy	513	384	129	< 0.001
	Mastectomy	627	318	309	
Radiotherapy	Have	785	533	252	< 0.001
	Haven't	355	169	186	
Chemotherapy	Have	724	450	274	0.027
	Haven't	416	252	164	
Hormonotherapy	Have	785	456	329	< 0.001
	Haven't	355	246	109	
Targeted therapy	Have	719	424	295	0.001
	Haven't	421	278	143	

Table 4 The univariate and adjusted analyses of predictors

Feature	Univariate	Adjusted	Feature	Univariate	Adjusted
	$P_{-\text{value}}$	$P_{-\text{value}}$		$P_{-\text{value}}$	$P_{-\text{value}}$
Age	0.013	0.034	Family history of other cancers	0.049	0.001
Residence	0.115	0.103	Tumor size	< 0.001	< 0.001
Ethnicity	0.216	0.197	Estrogen receptor	< 0.001	< 0.001
Smoking	0.035	0.017	Progesterone receptor	< 0.001	0.001
Alcohol	0.046	0.05	HER-2	< 0.001	< 0.001
Obesity	0.001	< 0.001	Number of lymph nodes involved	< 0.001	< 0.001
Use of contraceptive pills	0.001	0.013	Tumor grade	< 0.001	< 0.001
Physical activity	< 0.001	0.015	Surgery type	< 0.001	< 0.001
Number of other chronic conditions	< 0.001	0.001	Radiotherapy	0.001	0.001
Premature menstruation	0.001	0.034	Chemotherapy	< 0.001	0.001
Absence of pregnancy	0.001	< 0.001	Hormonotherapy	0.016	0.035
Personal history of BC	< 0.001	< 0.001	Targeted therapy	0.023	0.015
Family history of BC	< 0.001	0.001			

Table 5 The algorithms' hyperparameter tuning

Algorithm	Parameters
CNN	activation: selu, epochs: 20, learning_rate: 0.03, neurons: 25, optimizer: 0.8, batch_size: 60.
RF	batch_size = 45, max_depth = 7, feature_num = 5, epoch = 50, cal_out_bag = false, seed = 3, base_classifier = DT.
SVM	Kernel = RBF, C = 10, precision = 0.1, RBF gamma = 0.3, stopping criteria = 10E-3.
DNN	activation(relu), dropout(0.5), Dense(2), pool_size=(3), kernel size = 7, filters_num:40.
XGB	max_depth = 10, min_child_weight = 2, learning_rate = 0.5, colsample_bytree = 0.75, subsample = 0.2.
LSTM	Dense (5), learning rate = 0.3, momentum = 0.65, epochs_num = 50, batch size = 50, activation = sigmoid.
KNN	K = 5, batch size = 35, CV = true, search algorithm = Linear NN search, distance weighting = None.

Table 6 The ranges of algorithms' hyperparameters

Algorithm	Parameters
CNN	Epochs (10, 15, 20, 25, 30, 40, 50), learning rate (0.02–0.05), neurons (10, 15, 20, 25, 30, 35, 40, 45, 50), optimizer (0.6, 0.7, 0.8, 0.9)
RF	Max depth (6, 7, 8, 9, 10, 12, 15), feature_num (4, 5, 6, 7, 8, 9, 10), epoch (20, 30, 40, 50, 70, 100),
SVM	C (10,15), RBF gamma (0.1, 0.2, 0.3, 0.4, 0.5), stopping criteria (10E-1, 10E-2, 10E-3)
DNN	Dense (2, 3, 4, 5), pool-size (2, 3, 4, 5, 6, 7), Kernel size (5, 7, 8, 10), filters_sum (10, 20, 30, 40, 50)
XGB	Maz_depth (5, 6, 7, 8, 9, 10, 15), learning_rate (0.3, 0.5, 0.7, 0.8, 0.9)
LSTM	Dense (2, 3, 5, 7), learning rate (0.1, 0.2, 0.3, 0.5), epoch_num (20, 30, 40, 50, 70, 100)
KNN	K (3, 5, 7, 9, 11, 13, 15), search algorithm (Linear NN search, KD-Tree)

Table 7 The performance of AI-based models

Algorithm	PPV	NPV	Sensitivity	Specificity	Accuracy	F-Score
CNN	0.80	0.82	0.82	0.80	0.81	0.81
RF	0.96	0.92	0.92	0.96	0.94	0.94
SVM	0.61	0.62	0.62	0.61	0.62	0.62
DNN	0.71	0.72	0.73	0.70	0.71	0.72
XGB	0.72	0.74	0.75	0.72	0.73	0.73
LSTM	0.76	0.77	0.77	0.76	0.76	0.76
KNN	0.71	0.69	0.68	0.72	0.70	0.70

SHAP summary plot

As previously stated, the RF algorithm outperformed others in predicting the BC one-year recurrence regarding PPV, NPV, sensitivity, specificity, accuracy, F-Score, and ROC analysis. Accordingly, the SHAP plot (Fig. 7) was specially generated for this algorithm. As shown in this figure, the tumor grade, number of lymph nodes involved, and HER-2 were the more significant predictors in predicting recurrence.

Discussion

This study aims to predict the one-year recurrence of BC using AI-based algorithms and prognostic factors. We applied seven chosen ML and DL algorithms, including RF, XGB, SVM, KNN, CNN, LSTM, and DNN, to

Table 8 The performance of algorithms in train mode

Algorithm	PPV	NPV	Sensitivity	Specificity	Accuracy	F-Score
CNN	0.84	0.85	0.86	0.84	0.85	0.85
RF	0.98	0.96	0.96	0.98	0.97	0.97
SVM	0.66	0.69	0.71	0.64	0.67	0.68
DNN	0.75	0.80	0.82	0.73	0.77	0.78
XGB	0.79	0.80	0.81	0.78	0.79	0.80
LSTM	0.82	0.81	0.80	0.82	0.81	0.81
KNN	0.75	0.73	0.71	0.76	0.74	0.73

Table 9 The performance of algorithms in test mode

Algorithm	PPV	NPV	Sensitivity	Specificity	Accuracy	F-Score
CNN	0.76	0.76	0.76	0.76	0.76	0.76
RF	0.89	0.88	0.88	0.89	0.88	0.88
SVM	0.58	0.58	0.59	0.57	0.58	0.58
DNN	0.67	0.68	0.69	0.66	0.67	0.68
XGB	0.68	0.69	0.71	0.66	0.68	0.69
LSTM	0.69	0.70	0.71	0.68	0.69	0.70
KNN	0.66	0.63	0.60	0.68	0.64	0.63

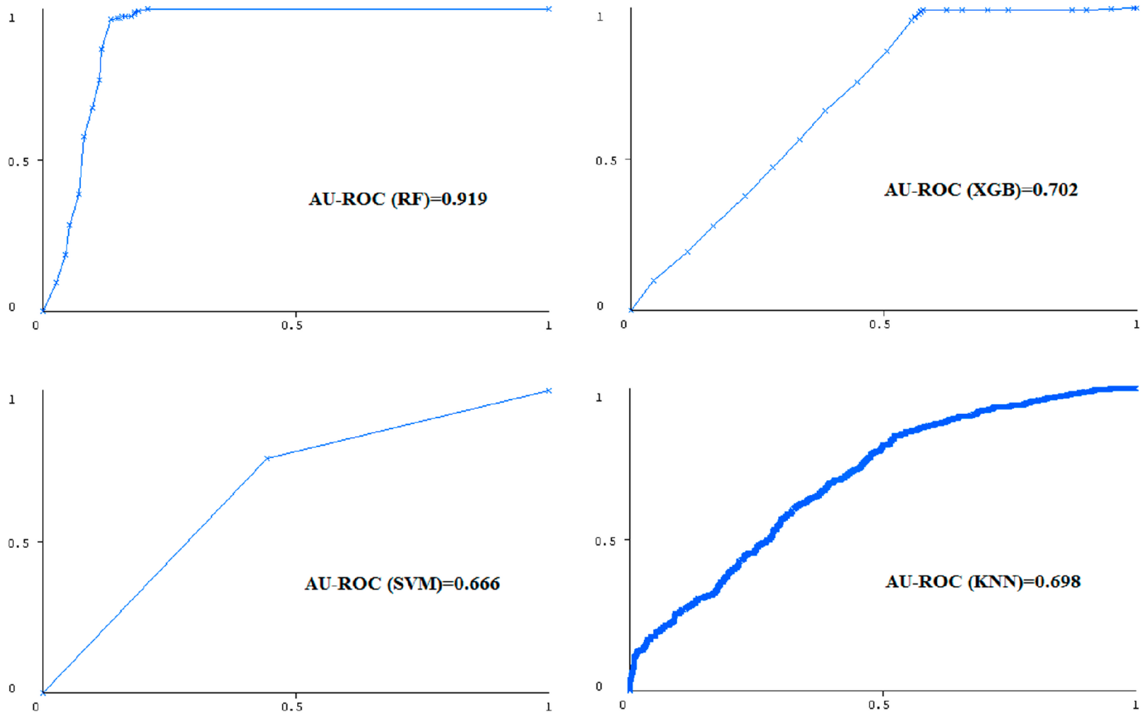

Fig. 2 The ROC curve plots for ML models

establish prediction models to this end. The comprehensive performance criteria, including PPV, NPV, sensitivity, specificity, accuracy, F-score, AU-ROC, calibration, and DCA curves, were measured to assess models' discrimination, reliability, and clinical utility for tumor recurrence. We generated the SHAP summary plot to indicate the XAI and show the interpretability of the AI-based models. The current empirical results showed that RF, with a PPV of 0.96, NPV of 0.92, sensitivity of 0.92, specificity of 0.96, accuracy of 0.94, F-score of 0.94, AU-ROC of 0.919, more calibration, and NB was an outstanding model in predicting one-year recurrence of BC. Also, the SHAP summary plot for the RF model exhibited tumor grade, number of lymph nodes involved, and HER-2 as the more competent descriptors for predicting the one-year recurrence of BC.

So far, several efforts have been made to establish a prediction model for BC recurrence using DL and ML

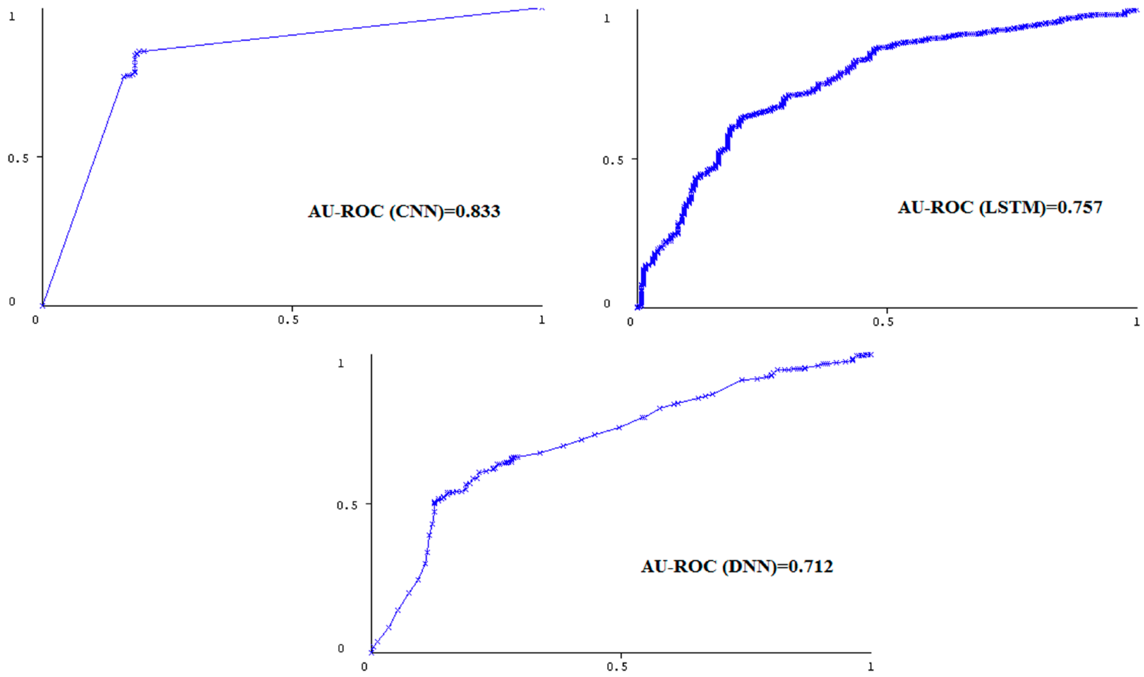

Fig. 3 The ROC curve plots for DL models

algorithms, and we have pointed out some of these studies. LG's analysis has focused on the ML models for predicting BC recurrence by utilizing 1189 records and 22 predictors. They indicated that SVM is an outstanding model with an accuracy of 0.957 in predicting recurrence. Also, HER2, tumor grade, and the number of lymph nodes involved were prominent prognostic factors for the two-year BC recurrence [47]. In the current study, the ensemble model, such as RF, revealed more predictive performance efficiency than other DL and ML models. Also, the aforementioned prognostic factors were considered more significant for predicting the one-year recurrence of BC, similar to LS's study.

Despite DL algorithms obtaining higher predictive performance when dealing with more complicated data such as images, text, and others, the current study showed higher performance efficiency for RF compared to the DL algorithms. Indeed, some ML models, such as RF as one hybrid ML algorithm, usually provide satisfactory performance efficiency when we leverage simpler data, such as tabular datasets, to establish prediction models. The RF model has also obtained satisfactory predictive performance in other studies leveraging tabular data; for example, it has attained an excellent predictive performance with an accuracy of 0.97 [48]. The superiority of ML models over DL is not always confirmed in all empirical studies using tabular data, and this superiority varies depending on the data used. In other words, the DL models may obtain higher predictive performance than ML ones in some research. In one study by Cai et al., the DL model with an accuracy of 0.89 was considered a higher-performing model compared to ML models [49]. Indeed, in some scenarios, the DL models may provide more performance efficiency. So, leveraging DL models may assist us in obtaining higher predictive performance when we use tabular data.

Zuo et al. examined BC recurrence using the prognostic factors obtained by AdaBoost as the best classifier for predicting recurrence [50]. Alzubiet al. leveraged ML and natural language processing algorithms to extract essential features from the medical records and developed the prediction model for BC recurrence. OneR algorithm was considered the best model on this topic. Alzubiet's study determined tumor grade as a top-ranking predictor for BC recurrence, consistent with our experimental study [21]. Lou et al. investigated ten-year BC recurrence using ML models and registry data on prognostic factors. Artificial Neural Network (ANN) with an AU-ROC of 97.62% and 95% CI of 96.83-98.41% in validation data was recognized as a more eligible model to enhance clinical decision-making. Based on the ANN model, tumor stage was the best predictor for the prognosis of ten-year BC recurrence, which was in line with our result [51]. Liu et al. indicated the competency of XGB with the AU-ROC of 0.75 and 0.72 in internal and external datasets for predicting BC recurrence and metastasis using the clinical data extracted from the histopathological images [22]. Compared to the study of Liu, where ML models trained by the extracted features of image data showed higher performance, in the present study, DL models achieved higher performance than some ML models when using routine clinical data.

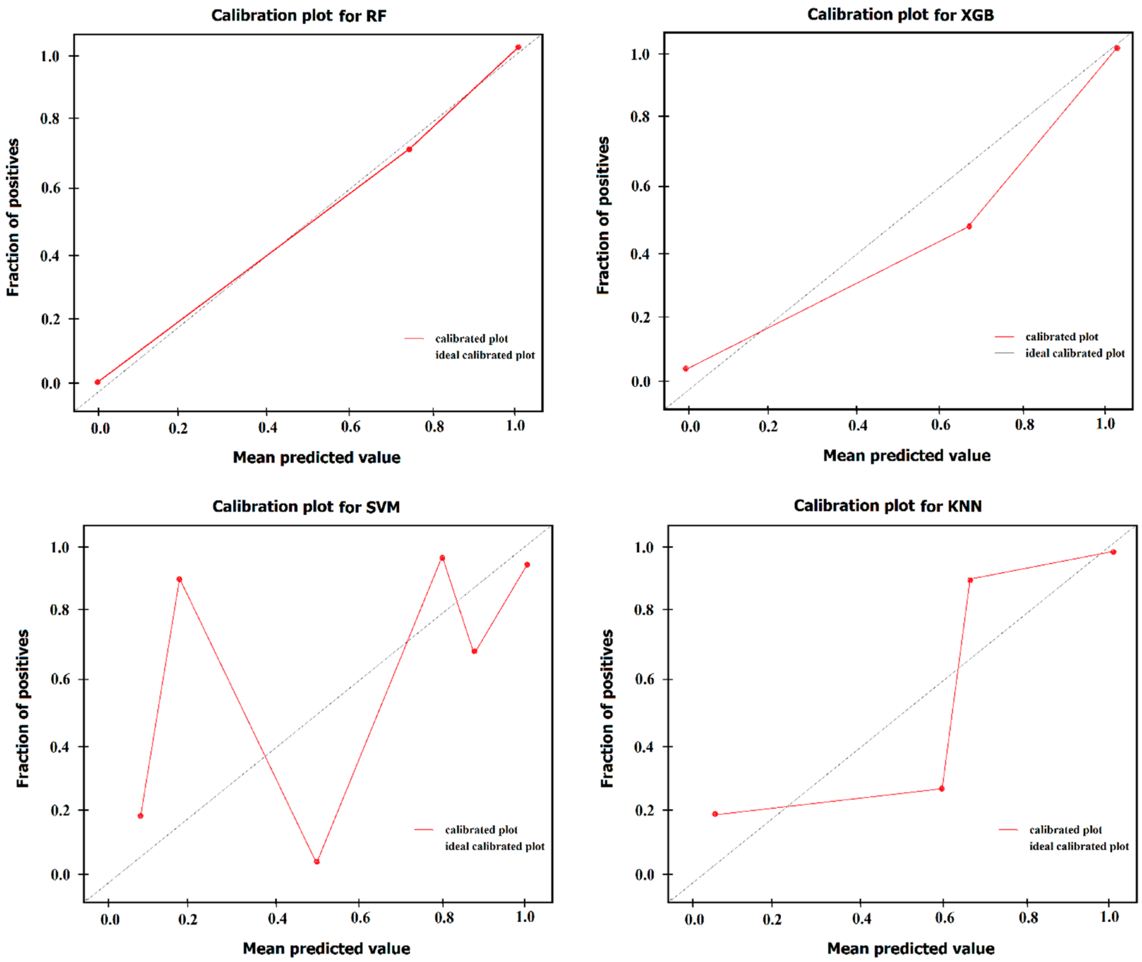

Fig. 4 Calibration plots for ML models

Kim et al. leveraged SVM to construct prediction models for the five-year BC metastasis after surgery, which resulted in an AU-ROC of 0.85 and 95% CI of 0.79–0.91 [10]. The current research resulted in higher performance of DL and ensemble ML models for predicting one-year BC recurrence despite Kim's study that has obtained the SVM as a higher-performing model for the five-year BC relapse.

Despite the current study that revealed that the RF model obtained higher predictive results than other AI algorithms, other studies have shown that the base algorithms, such as SVM or simple NNs, are more efficient in predicting BC recurrence. The superiority of algorithms' performance entirely depends on the experimental conditions with different datasets, even among structural types. Although ensemble algorithms such as RF usually get more performance than simple ones, this subject cannot be extended to all experimental results. As indicated in the previous studies, the simple algorithms are more efficient for predicting BC relapse.

Regarding the model selection and prognostic factors, we found one study conducted by Tsenga et al., which focused on serum biomarkers and clinicopathological data and resulted in the superiority of RF with an AU-ROC of 0.75 for predicting BC recurrence. Moreover, this study obtained HER2 as a critical predictor of prognostication [52]. The LN involvement rate and HER2 value were two significant indicators of BC recurrence in Mosayebi's study, revealing their potential influence in enhancing the prognosis of BC relapse [53]. DL models were known as efficient algorithms for predicting BC recurrence. One study by Kim et al. revealed that the Weibull Time To Event Recurrent Neural Network (WTTE-RNN) trained by clinical data and follow-up clinical information entailing laboratory and image data, with the AU-ROC between 0.91 and 0.92, is satisfactory to predict 2–7 years of BC recurrence [54]. Howard's study utilized the DL approach and clinical features on pathological images to predict this aim [55]. Druker et al. leveraged DL with MRI data to predict at least 2 (or 5) years of BC recurrence-free follow-up and gained an AU-ROC of 0.8 [56]. Su et al. focused on BC recurrence using DL trained by histopathological data and obtained the overall AU-ROC of 0.775 and 0.811 from two separate

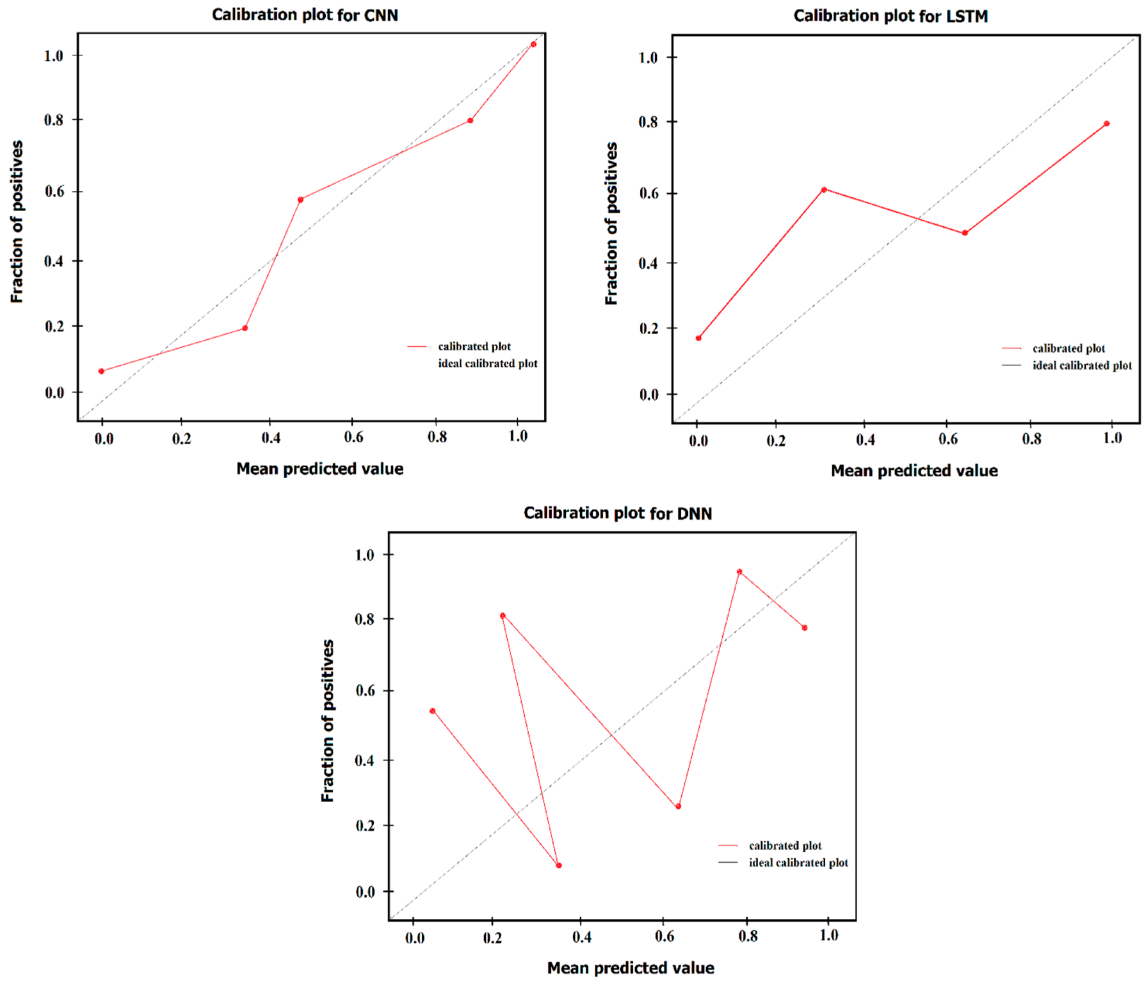

Fig. 5 Calibration plots for DL models

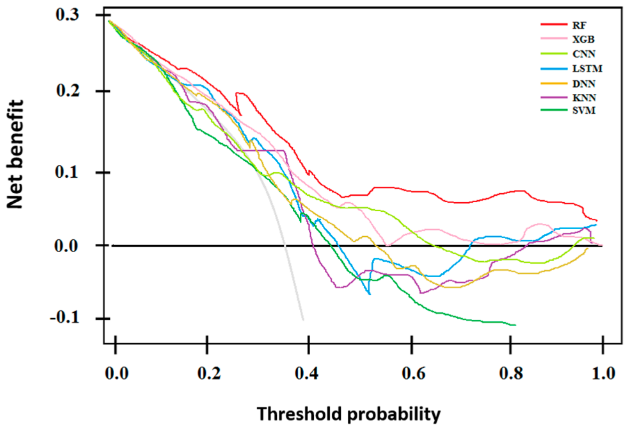

Fig. 6 The DCA of AI-based models. The NB curves for the chosen prognostic factors are indicated. The X and Y axes indicate the risk thresholds for predicting the one-year BC recurrence and NB, respectively

datasets [57]. Phan et al. utilized deep convolutional neural networks and 233 complete slide images belonging to 138 BC cases to predict BC recurrence, resulting in an accuracy of 0.91.

As previously mentioned, more studies focusing on BC recurrence were on investigating the patients' follow-up more than two years after BC surgery. However, the patient's follow-up and identification after performing one year after surgery is crucial in enhancing the patients' prognosis and choosing more efficient therapy measures to increase patients' survival. Hence, we endeavored to utilize AI-based algorithms to predict the one-year BC recurrence. In addition, other studies revealed that the DL algorithms obtain favorable performance efficiency when trained by tabular and structural data. By achieving this approach, we demonstrated the efficacy of DL algorithms for predicting the one-year BC recurrence. Furthermore, the XAI demonstrated the predictive competency of prognostic descriptors for one-year BC recurrence, highlighting their potential utility in decision-making in clinical environments. It is similar to other studies that prognosticated the BC recurrence in two to ten years after performing surgery.

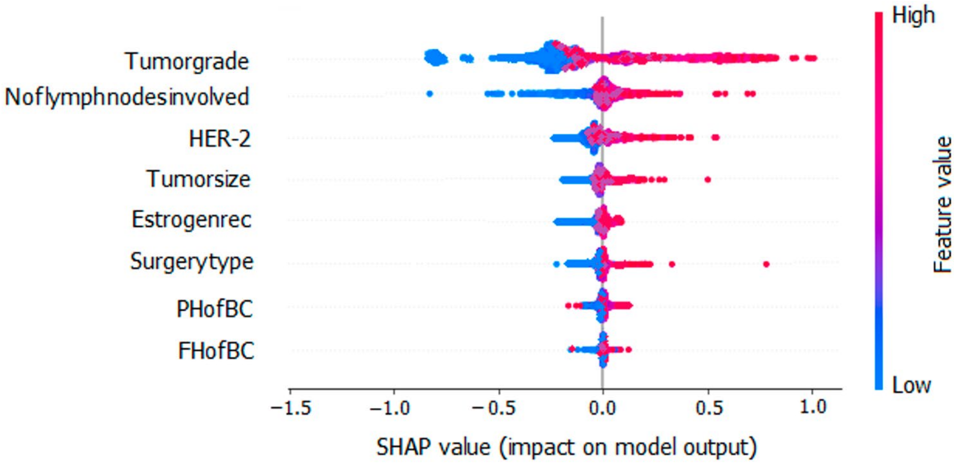

Fig. 7 The SHAP summary plot for one-year BC recurrence. It put together feature importance and feature influence. Each point on the summary plot illustrates the Shapley value associated with one instance. The positions on the y and x axes represent the feature and Shapley values, respectively. The blue and red colors show the value of the individual features from low to high, respectively

Limitations and future trends

Despite the benefits of the current study, we faced some limitations that should be considered. Firstly, we used the 1156 data of BC cases belonging to three clinical centers, which may influence the performance of the current established models. We recommend conducting a prospective study and gathering more clinical data from more clinical centers to increase the performance and comprehensiveness of the models. Secondly, we used preprocessing steps, such as the imputation process, to fill in some missing data, which may impact the data quality. For this subject, we suggest finding real value from patients' records or contacting patients if possible to increase the chance of preserving the characteristics of the original data. Thirdly, this study failed to gauge the external validity of models, and we suggest performing this step to assess the generalizability of the models for predicting the one-year BC recurrence. Another limitation of this study was the lack of evaluating the current prognostic models with new cases and comparing the performance with other existing tools such as Oncotype DX, Nottingham Prognostic Index, or AJCC staging to assess the performance ability of the model regarding other prognostic tools. For future studies, we recommend one cohort study to compare the prognostic models constructed by the AI-based algorithms with other existing tools to achieve better clinical utility of the established prediction models.

Conclusion

The current study specifically conducted a predictive analysis of the ML and DL algorithms to predict the one-year BC recurrence. Our empirical results demonstrated the superiority of the RF model over the other ML and DL models in terms of prediction performance. XAI showed that tumor grade, number of lymph nodes involved, and HER-2 were more eligible to enhance the clinical utility of the model established. The current empirical results indicate that RF is a potential technology in the early prediction of BC among post-operative patients. Doctors can use the knowledge extracted from the model and effective prognostic factors to make more efficient clinical decisions in healthcare settings. Considering the performance obtained in the current research, this model can improve the prognosis and therapy, control costs and other limited resources in healthcare, and increase survival by enhancing the efficiency of clinical decision-making by healthcare providers.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-14369-5.

Supplementary Material 1

Acknowledgements

We thank all the people who assisted us in all steps of this study.

Author contributions

R.N. conducted the writing, review, and editing of this manuscript.

Funding

There was no funding for this manuscript.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the ethics committee of Tehran University of Medical Sciences (IR.TUMS.SPH.REC.1398.401). The study was performed in compliance with this institutional guideline, ethical guidelines for clinical research of the Iranian government, and the Declaration of Helsinki. The database leveraged was anonymous (did not include names or any identification detail), and confidentiality of information was assured. All methods were performed in accordance with the relevant guidelines and regulations by the Declaration of Helsinki. Informed consent was obtained from all subjects and/or their legal guardian(s).

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Author details

[1]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Received: 13 March 2025 / Accepted: 20 May 2025

References

1. Smolarz B, Nowak AZ, Romanowicz H. Breast cancer—epidemiology, classification, pathogenesis and treatment (Review of Literature). Cancers [Internet]. 2022; 14(10).
2. Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A, Breast, Cancer—Epidemiology. Risk factors, classification, prognostic markers, and current treatment Strategies—An. Updated Rev Cancers [Internet]. 2021; 13(17).
3. Huang J, Chan PS, Lok V, Chen X, Ding H, Jin Y, et al. Global incidence and mortality of breast cancer: a trend analysis. Aging. 2021;13(4):5748–803.
4. Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun. 2021;41(11):1183–94.
5. Roheel A, Khan A, Anwar F, Akbar Z, Akhtar MF, Imran Khan M et al. Global epidemiology of breast cancer based on risk factors: a systematic review. Front Oncol. 2023;13.
6. Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025;31.
7. Naseem U, Rashid J, Ali L, Kim J, Haq QEU, Awan MJ, et al. An automatic detection of breast Cancer diagnosis and prognosis based on machine learning using ensemble of classifiers. IEEE Access. 2022;10:78242–52.
8. Ying Y, Yang M, Chen J, Yao C, Bian W, Wang C, et al. Identification and evaluation of a risk model predicting the prognosis of breast cancer based on characteristic signatures. Translational cancer Res. 2023;12(6):1441–51.
9. Arya N, Saha S. Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowl Based Syst. 2021;221:106965.
10. Kim W, Kim KS, Lee JE, Noh D-Y, Kim S-W, Jung YS, et al. Development of novel breast Cancer recurrence prediction model using support vector machine. Jbc. 2012;15(2):230–8.
11. Mazo C, Aura C, Rahman A, Gallagher WM, Mooney C. Application of artificial intelligence techniques to predict risk of recurrence of breast cancer: A systematic review. J Personalized Med [Internet]. 2022; 12(9).
12. Zeng L, Liu L, Chen D, Lu H, Xue Y, Bi H et al. The innovative model based on artificial intelligence algorithms to predict recurrence risk of patients with post-operative breast cancer. Front Oncol. 2023;13.
13. Zhang B, Shi H, Wang H. Machine learning and AI in Cancer prognosis, prediction, and treatment selection: A critical approach. J Multidisciplinary Healthc. 2023;16(null):1779–91.
14. Jones MA, Islam W, Faiz R, Chen X, Zheng B. Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction. Front Oncol. 2022;12.
15. Amethiya Y, Pipariya P, Patel S, Shah M. Comparative analysis of breast cancer detection using machine learning and biosensors. Intell Med. 2022;2(2):69–81.
16. Shanbehzadeh M, Nopour R, Erfannia L, Amraei M, Mehrabi N, Mashoufi M. Comparing data mining algorithms for breast Cancer diagnosis. Shiraz E-Medical J. 2022;23(7):e120140.
17. Sammut S-J, Crispin-Ortuzar M, Chin S-F, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022;601(7894):623–9.
18. Godishala AK, Yassin H, Veena R, Lai DTC, editors. Breast Cancer Tumor Image Classification Using Deep Learning Image Data Generator. 2022 7th International Conference on Image, Vision and Computing (ICIVC); 2022 26–28 July 2022.
19. Yang P-T, Wu W-S, Wu C-C, Shih Y-N, Hsieh C-H, Hsu J-L. Breast cancer recurrence prediction with ensemble methods and cost-sensitive learning. Open Med. 2021;16(1):754–68.
20. Kim J-Y, Lee YS, Yu J, Park Y, Lee SK, Lee M et al. Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary Cancer center registry. Front Oncol. 2021;11.
21. Alzu'bi A, Najadat H, Doulat W, Al-Shari O, Zhou L. Predicting the recurrence of breast cancer using machine learning algorithms. Multimedia Tools Appl. 2021;80(9):13787–800.
22. Liu X, Yuan P, Li R, Zhang D, An J, Ju J, et al. Predicting breast cancer recurrence and metastasis risk by integrating color and texture features of histopathological images and machine learning technologies. Comput Biol Med. 2022;146:105569.
23. Mikhailova V, Anbarjafari G. Comparative analysis of classification algorithms on the breast cancer recurrence using machine learning. Med Biol Eng Comput. 2022;60(9):2589–600.
24. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ (Clinical Res ed). 2024;385:e078378.
25. Dhal P, Azad C. A comprehensive survey on feature selection in the various fields of machine learning. Appl Intell. 2022;52(4):4543–81.
26. Singh LK, Khanna M, Singh R. Efficient feature selection for breast cancer classification using soft computing approach: A novel clinical decision support system. Multimedia Tools Appl. 2024;83(14):43223–76.
27. Effrosynidis D, Arampatzis A. An evaluation of feature selection methods for environmental data. Ecol Inf. 2021;61:101224.
28. Pradipta GA, Wardoyo R, Musdholifah A, Sanjaya INH, Ismail M, editors. SMOTE for handling imbalanced data problem: a review. 2021 Sixth International Conference on Informatics and Computing (ICIC). IEEE; 2021. pp. 3–4.
29. Soomro AA, Mokhtar AA, Muhammad MB, Saad MHM, Lashari N, Hussain M, et al. Data augmentation using SMOTE technique: application for prediction of burst pressure of hydrocarbons pipeline using supervised machine learning models. Results Eng. 2024;24:103233.
30. Ahmed AHA, Jin W, Ali MAH. Comparative analysis of intelligent models for predicting compressive strength in recycled aggregate concrete. Model Earth Syst Environ. 2024;10(4):5273–91.
31. Ranjan GSK, Verma AK, Radhika S, editors. K-nearest neighbors and grid search CV based real time fault monitoring system for industries. 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). March 2019; 29–31.
32. Mezzadri G, Laloë T, Mathy F, Reynaud-Bouret P. Hold-out strategy for selecting learning models: application to categorization subjected to presentation orders. J Math Psychol. 2022;109:102691.
33. Mohamed AMA. Post-pruning decision tree algorithms based on bayes minimum risk and loss minimization to combat over-fitting problem. 2018.
34. Wu C-C, Yeh W-C, Hsu W-D, Islam MM, Nguyen PA, Poly TN, et al. Prediction of fatty liver disease using machine learning algorithms. Comput Methods Programs Biomed. 2019;170:23–9.
35. Buturović L, Miljković DJB. A novel method for classification of tabular data using convolutional neural networks. 2020;2020.05. 02.074203.
36. Begum M, Shuvo MH, Nasir M, Hossain A, Hossain M, Ashraf I, et al. LCNN: lightweight CNN architecture for software defect feature identification using explainable AI. IEEE Access. 2024;PP:1.
37. Gicic A, Đonko D, Subasi A. Time sequence deep learning model for ubiquitous tabular data with unique 3D tensors manipulation. Entropy [Internet]. 2024; 26(9).
38. Alvi RH, Rahman MH, Khan AAS, Rahman RM. Deep learning approach on tabular data to predict early-onset neonatal sepsis. J Inform Telecommunication. 2021;5(2):226–46.
39. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol. 2018;74(6):796–804.
40. Linardatos P, Papastefanopoulos V, Kotsiantis S, Explainable AI. A review of machine learning interpretability methods. Entropy [Internet]. 2021; 23(1).

41. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. J Artif Intell Res. 2021;70:245–317.
42. Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI [Internet]. 2023; 4(3):[652– 66 pp.].
43. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inform Fusion. 2023;99:101805.
44. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inform Fusion. 2020;58:82–115.
45. Marcílio WE, Eler DM, editors. From explanations to feature selection: assessing SHAP values as feature selection mechanism. 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Nov 2020; 7–10.
46. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.
47. Ahmad LG, Eshlaghy A, Poorebrahimi A, Ebrahimi M, Razavi A. Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inf. 2013;4(124):3.
48. Ahmadi M, Nopour R, Nasiri S. Developing a prediction model for successful aging among the elderly using machine learning algorithms. Digit HEALTH. 2023;9:20552076231178425.
49. Cai T, Long J, Kuang J, You F, Zou T, Wu L. Applying machine learning methods to develop a successful aging maintenance prediction model based on physical fitness tests. Geriatr Gerontol Int. 2020;20(6):637–42.
50. Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. BMC Med Inf Decis Mak. 2023;23(1):276.
51. Lou S-J, Hou M-F, Chang H-T, Chiu C-C, Lee H-H, Yeh S-CJ et al. Machine learning algorithms to predict recurrence within 10 years after breast Cancer surgery: A prospective cohort study. Cancers [Internet]. 2020; 12(12).
52. Tseng Y-J, Huang C-E, Wen C-N, Lai P-Y, Wu M-H, Sun Y-C, et al. Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. Int J Med Informatics. 2019;128:79–86.
53. Mosayebi A, Mojaradi B, Bonyadi Naeini A, Khodadad Hosseini SH. Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. PLoS ONE. 2020;15(10):e0237658.
54. Kim J-Y, Lee YS, Yu J, Park Y, Lee SK, Lee M et al. Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary Cancer center registry. Front Oncol. 2021;Volume 11–2021.
55. Howard FM, Dolezal J, Kochanny S, Khramtsova G, Vickery J, Srisuwananukorn A, et al. Integration of clinical features and deep learning on pathology for the prediction of breast cancer recurrence assays and risk of recurrence. Npj Breast Cancer. 2023;9(1):25.
56. Drukker K, Edwards A, Papaioannou J, Giger M, editors. Deep learning predicts breast cancer recurrence in analysis of consecutive MRIs acquired during the course of neoadjuvant chemotherapy. Medical Imaging 2020: Computer-Aided Diagnosis. 2020; SPIE.
57. Su Z, Niazi MKK, Tavolara TE, Niu S, Tozbikian GH, Wesolowski R, et al. BCR-Net: A deep learning framework to predict breast cancer recurrence from histopathology images. PLoS ONE. 2023;18(4):e0283562.

Publisher's note