SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(5), 1692.
https://doi.org/10.62617/mcb1692

Article

Deep neural network-based interpretable prediction model for survival outcomes in female breast cancer patients: Integrating biomechanical perspectives with clinicopathological features

Yichen Zhang

National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing 102206, China; m202321289@xs.ustb.edu.cn

CITATION

Zhang Y. Deep neural network-based interpretable prediction model for survival outcomes in female breast cancer patients: Integrating biomechanical perspectives with clinicopathological features. Molecular & Cellular Biomechanics. 2025; 22(5): 1692. https://doi.org/10.62617/mcb1692

ARTICLE INFO

Received: 25 February 2025
Accepted: 10 March 2025
Available online: 24 March 2025

Abstract: Background: This study integrates biomechanical perspectives with clinicopathological data to develop a DNN model for survival prediction. By linking tumor size and lymph node status to biomechanical drivers such as solid stress and cell migration forces, we aim to uncover the mechanobiological mechanisms underlying prognosis heterogeneity. **Methods:** We analyzed data from 37,917 patients in the SEER database, encompassing clinical characteristics, pathological features, and treatment details. The DNN, featuring an attention mechanism, was evaluated using metrics such as accuracy, precision, recall, F1 score, and Area Under Curve (AUC). Interpretability techniques were applied to identify prognostic factors. **Results:** The DNN model achieved F1 scores of 0.928 and 0.935 for validation and test sets, respectively, with an AUC of 0.96, surpassing traditional models. Key factors identified included regional lymph node positivity, tumor size, and tumor grade, with a notable negative correlation between regional lymph node positivity and survival. **Conclusions:** DNN models with attention mechanisms demonstrate superior predictive performance and valuable interpretability in identifying critical prognostic factors.

Keywords: breast cancer; deep neural networks; machine learning; survival prediction; biomechanics; interpretability analysis

1. Introduction

In 2022, the most common female malignancy in 157 out of 185 countries was breast cancer, which still has the highest incidence of female cancers and highly heterogeneous tumor characteristics and clinicopathological features [1–3]. Female patients have a 10%–15% recurrence rate within 5 years of breast cancer diagnosis. In 2022, 2.3 million women worldwide were diagnosed with breast cancer, and 670,000 are expected to die from the disease. Given that breast cancer has a very high mortality and recurrence rate, there is an urgent need for universally applicable and accurate methods to identify patients at high or low risk of mortality. Prompt identification of patients will facilitate personalized treatment decisions and enable precision therapy [4,5].

It is well known that the interplay between patient characteristics, clinicopathological features, tumor characteristics, and other variables complicates the identification of independent risk factors for predictable outcomes. However, machine learning (ML) provides vital support to address this problem, as it can effectively recognize complex relationships between variables and handle complex datasets [6–8]. ML models algorithmically learn patterns from massive patient data covering demographic information, histopathological characteristics, and treatment options.

Hence, ML-based schemes are expected to facilitate accurate predictive models based on the individual patient's case, thereby providing personalized medicine to breast cancer patients.

Biomechanical properties of the tumor microenvironment play a key role in breast cancer progression. It has been shown that increased extracellular matrix (ECM) stiffness promotes cancer cell invasion and metastasis through activation of the integrin-FAK-YAP/TAZ signaling pathway [40]. In addition, the accumulation of solid stress within the tumor compresses blood vessels, induces hypoxia and enhances chemoresistance. In this study, tumor size not only reflects morphological features, but may also serve as a proxy indicator of the heterogeneity of the tumor mechanical microenvironment—larger tumors may be accompanied by higher internal stress and ECM fibrosis, and regional lymph node-positive status may suggest an enhanced metastatic capacity of the cancer cells through mechanosensitive pathways (e.g., Rho-ROCK-mediated cellular migration) achieving enhanced metastatic capacity [41]. Integration of these biomechanically relevant features by machine learning models may provide a more comprehensive mechanistic explanation for prognostic prediction [42].

In the medical field, neural networks are increasingly used. For example, Zhang et al. developed a deep learning model based on multimodal ultrasound images to predict breast cancer molecular staging with an AUC of 0.93, but did not integrate biomechanical features [43]. In addition, Cheng et al. combined tumor microenvironment genomic data to construct a survival prediction model and found that ECM fibrosis-related genes were significantly associated with prognosis, but their model did not take into account the effect of mechanical stress on tumor development [45]. These studies indicate that existing studies are deficient in the joint analysis of biomechanical and clinical features, and the innovation of this study is to fill this gap.

As an important branch of machine learning, neural networks have the advantages of ML in dealing with complex data, better recognizing nonlinear relationships, and dealing with large amounts of high-dimensional data. By simulating the hierarchical structure of the human brain, neural networks can automatically extract features from data and perform deeper learning and reasoning. This advantage has spurred the widespread use of neural networks in medical image analysis and genomics research. For instance, Zhou et al. [9] made significant progress in preoperative breast cancer molecular staging prediction by combining multimodal ultrasound imaging techniques with convolutional neural networks (CNN). Zheng et al. [10] utilized a deep learning radiomics technique using conventional ultrasound and shear wave elastography in combination with clinical parameters to predict axillary lymph node status preoperatively. Wang et al. [11] developed the DeepGrade model to improve risk stratification of Nottingham histological Grade-2 breast cancer patients using digital pathology images and neural networks. Stashko et al. [12] introduced STIFMap to measure the hardness heterogeneity of breast tumors using CNN. These studies have demonstrated the great potential and clinical value of neural networks in individualized therapy [13,14]. Recently, deep neural networks (DNN) have been at the forefront of neural network research due to their multilevel feature extraction and powerful learning capabilities, providing prediction and classification capabilities with higher accuracy [15].

When introducing ML and neural network models, the interpretability of the models becomes a key topic. Although these models are excellent at handling complex data and making predictions, their black-box nature prohibits clinicians and researchers from understanding the models' decision-making process [16]. Therefore, improving model interpretability is critical to ensuring their trust and transparency in clinical applications. Researchers can learn which features significantly impact model predictions through interpretability techniques, allowing a better understanding of disease mechanisms and patient prognosis. Model interpretability will improve their credibility and provide a scientific and reliable basis for individualized treatment [17,18].

This study aimed to identify the principal risk factors and forecast patient survival from a substantial number of clinicopathologic, therapeutic, and oncologic conditions across diverse racial, age, and marital status groups, employing ML and DNN models. The ML methods evaluated were Support Vector Machine (SVM), Naive Bayes, Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), and Multilayer Perceptron (MLP). Furthermore, interpretability analysis on the developed model based on extensive Vivid and SHAP analysis evaluated the intermediate model processes to ascertain the extent to which the model can be trusted. Conducting interpretable analyses on the models aimed at identifying the primary factors influencing patient prognosis, thereby offering a valuable reference for clinicians.

The remainder of this study is organized as follows. Section 2 outlines the materials and methods employed, including the data sources, variable collection, statistical analysis, and selection of machine learning algorithms. Section 3 presents the experimental results, including sample characteristics, model comparison, and interpretability analysis. Section 4 discusses the findings and their clinical significance in-depth, identifies the limitations of this study, and suggests future research directions. Finally, Section 5 concludes this paper.

2. Materials and methods

2.1. Data sources

This study used data from the SEER database, which includes data from 17 registries collected by the SEER program for cancer diagnosis and survival outcomes. The database covers approximately 26.5% of the U.S. population (based on the 2020 Census) [19]. The SEER database is known for its high-quality cancer information, is open-source, and can be accessed through the official website (www.seer.cancer.gov).

The patient cohort for this study comprised patients diagnosed with breast cancer in the SEER database between 2000 and 2021 (International Classification of Disease for Oncology, 3rd Edition ICD-0-3 codes C50.0-C50.9). A total of 181 data items from 1,294,351 patients were extracted by SEER*Stat (version 8.4.3), a software officially provided by SEER. In this study, the collected data were meticulously cleaned and sampled, ultimately having 37,917 female breast cancer patients of all ages that formed a large-sample dataset. These data were divided into training, validation, and testing sets for our medical artificial intelligence study. **Figure 1** illustrates the detailed data screening process.

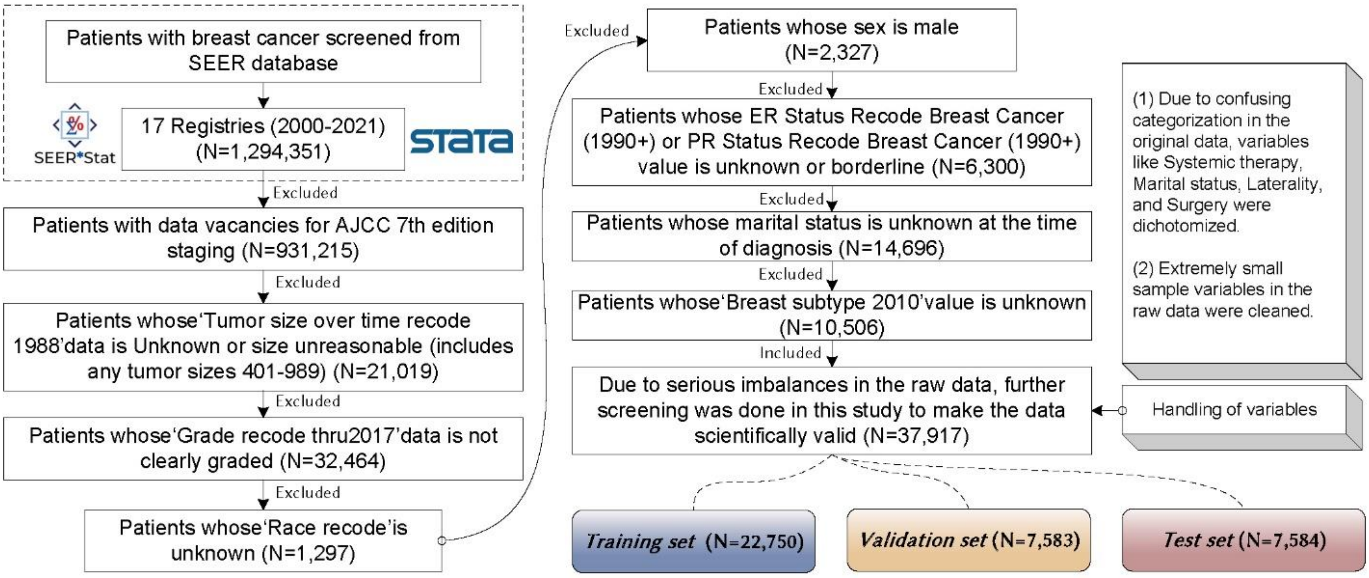
Patients with breast cancer screened from SEER database
SEER*Stat
17 Registries (2000-2021) (N=1,294,351)
STATA
Excluded
Patients with data vacancies for AJCC 7th edition staging (N=931,215)
Excluded
Patients whose'Tumor size over time recode 1988'data is Unknown or size unreasonable (includes any tumor sizes 401-989) (N=21,019)
Excluded
Patients whose'Grade recode thru2017'data is not clearly graded (N=32,464)
Excluded
Patients whose'Race recode'is unknown (N=1,297)
Excluded
Patients whose sex is male (N=2,327)
Excluded
Patients whose ER Status Recode Breast Cancer (1990+) or PR Status Recode Breast Cancer (1990+) value is unknown or borderline (N=6,300)
Excluded
Patients whose marital status is unknown at the time of diagnosis (N=14,696)
Excluded
Patients whose'Breast subtype 2010'value is unknown (N=10,506)
Included
Due to serious imbalances in the raw data, further screening was done in this study to make the data scientifically valid (N=37,917)
(1) Due to confusing categorization in the original data, variables like Systemic therapy, Marital status, Laterality, and Surgery were dichotomized.
(2) Extremely small sample variables in the raw data were cleaned.
Handling of variables
Training set (N=22,750)
Validation set (N=7,583)
Test set (N=7,584)

Figure 1. Data screening workflow.

2.2. Variables

This study included the following variables: Age, tumor size, regional nodes positive, race, primary site, grade, laterality, histology recode (based on Histologic Type ICD-O-3), therapeutic information (surgery, radiotherapy, chemotherapy, and systemic therapy), breast subtype, ER, PR, marital status and survival ending. Tumor size: As a potential indicator of heterogeneity in the tumor mechanical microenvironment, larger tumors may be accompanied by higher internal solid stress and ECM fibrosis. Regional nodes positive: reflects the ability of cancer cells to migrate and may be associated with activation of the Rho GTPase-mediated mechanosignaling pathway.

Regional nodes positive: In addition to reflecting the ability of cancer cells to migrate (e.g., Rho-ROCK pathway activation), this variable is significantly associated with response to treatment. Clinical studies have shown that patients with positive lymph nodes are less sensitive to adjuvant chemotherapy (HR = 1.32, 95% CI: 1.15–1.52) and have a more significant decrease in the physiological functioning dimension of the 5-year postoperative quality of life score (EORTC QLQ-C30) ($\beta = -12.4$, $p <$ 0.001). Thus, the inclusion of this variable not only associates biomechanical mechanisms but also predicts treatment tolerance and long-term quality of survival.

2.3. Statistical analysis

This research conducted statistical analysis using the R software (version 4.3.0), where continuous data with normal distribution were expressed as mean ± SD, and the comparisons between the two groups relied on the *t*-test of two independent samples. Moreover, continuous data with skewed distribution were expressed as M (Q_1, Q_3), and the rank-sum test of two independent samples was used. Besides, categorical data were expressed as *n* (%), and the chi-square test or Fisher probability method was used. The differences between groups were considered statistically significant at $P <$ 0.05 [20].

2.4. ML algorithms

Several ML algorithms were used for classification tasks, including SVM [21], Naive Bayes [22], Logistic Regression [23], Decision Tree [24], Random Forest [25], KNN [26], and MLP [27]. These ML algorithms are chosen to build classification prediction models as they have demonstrated appealing results in paramedical tasks [28]. All ML models were implemented in Python (version 3.9.18). We present the operating principles of all competitor ML methods for completeness.

2.4.1. SVM algorithm

SVM is a supervised learning model widely used in classification and regression analysis. Its core idea is to maximize the boundaries between categories by constructing an optimal hyperplane in high-dimensional space to achieve effective data classification, i.e., separate the samples of different categories. The boundary is a straight line for two-dimensional data, while the boundary is a plane for three-dimensional data. In higher-dimensional spaces, the goal is a hyperplane. Suppose we have a set of training samples (x_1, y_1), (x_2, y_2), …, (x_n, y_n), where x_i denotes the feature vector, and $y_i \in \{1, -1\}$ is the category label. The optimal hyperplane is expressed as follows:

$$\omega x + b = 0 \tag{1}$$

where w is the normal vector, which determines the direction of the hyperplane, and b is the bias, which determines the distance of the hyperplane. For the optimal hyperplane, SVM maximizes the interval between the two categories, i.e., maximizes the interval distance, by solving the following optimization problem:

$$\min_{w,b} \frac{1}{2} \parallel w \parallel^2 \qquad y_i(w \cdot x_i + b) \geq 1, \forall i \tag{2}$$

For linearly indistinguishable data, SVM introduces a slack variable, $\xi_i \geq 0$, which allows some data points to fall within the classification interval or be misclassified. At this point, the optimization objective becomes:

$$\min_{w,b,\xi} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n} \xi_i \qquad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \forall i \tag{3}$$

where C is a regularization parameter that balances the weight between the interval maximization and misclassification penalties. For this study, we set C = 2.0, which allows some misclassification, but the interval is still as large as possible.

2.4.2. Naive Bayes algorithm

Naive Bayes is a simple yet effective probabilistic classifier based on Bayes' theorem, particularly suited to high-dimensional data and classification problems. Despite its straightforward assumptions (i.e., features are independent of each other), it demonstrates robust performance in numerous real-world applications. Naive Bayes classifiers are founded upon Bayes' theorem and employ a posterior probability approach to classify data based on feature conditions. The theorem is expressed as:

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)} \tag{4}$$

where $P(y|X)$ is the posterior probability of category y given feature X, $P(X|y)$ is the likelihood probability of feature X given category y, $P(y)$ is the prior probability of category y, and $P(X)$ is the marginal probability of feature X.

2.4.3. Logistic regression

Logistic regression is a widely used linear classification model that suits binary classification problems. It categorizes data points by estimating the probability of belonging to a specific category. Specifically, the logistic regression model represents the category probability using a linear transformation of log odds. Let a set of feature vectors be $X = (x_1, x_2, \ldots, x_n)$, and the corresponding category label is $y \in \{0, 1\}$. The logistic regression model is mathematically formulated as follows:

$$P(y = 1|X) = \frac{1}{1 + \exp\big(-(w \cdot X + b)\big)} \tag{5}$$

where $P(y = 1|X)$ denotes the probability that a given feature X belongs to category 1, w is the weight vector, b is the bias term, and exp denotes the exponential function. By maximizing the log-likelihood function, w and b can be estimated. The log-likelihood function is defined as follows:

$$\mathcal{L}(w, b) = \sum_{i=1}^{n} \big[y_i \log P(y_i = 1|X_i) + (1 - y_i) \log\big(1 - P(y_i = 1|X_i)\big)\big] \tag{6}$$

2.4.4. Decision tree

A decision tree is a tree-structured ML model that builds classifiers by recursively dividing the dataset into smaller subsets. The dataset is divided by selecting the optimal features and split points so that the purity of the subset (i.e., the proportion of similar samples) after each division is as high as possible. Commonly used segmentation criteria include Information Gain based and Gini Index. This study adopts the Information Gain based segmentation criterion.

2.4.5. Random forest

Random Forest is an integrated learning method that improves classification and robustness by constructing multiple decision trees and combining their outputs. Random forest generates multiple sub-datasets by sampling the original dataset multiple times (with put-back sampling) and trains a decision tree on each sub-dataset. The voting results of all decision trees determine the final classification result. Suppose we have B decision trees $h_1(x)$, $h_2(x)$, …, $h_B(x)$ for input sample x. The final output of the random forest is:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_B(x)\} \tag{7}$$

where mode denotes the result of taking a majority vote.

2.4.6. KNN algorithm

KNN is an instance-based learning method that classifies different samples by measuring their distances from each other. KNN selects the k samples with the closest distances by calculating the distances between the samples to be classified and all the samples in the training set. The majority class of these k samples then determines the class of the samples to be classified. Precisely, for an input sample x, KNN identifies the k nearest samples in the training set, x_1, x_2, …, x_k, and the category of x is determined based on the majority voting principle.

2.4.7. MLP method

MLP is a feed-forward neural network that classifies data by learning its complex nonlinear relationships. It comprises an input layer, one or more hidden layers, and an output layer.

2.5. DNN algorithms

This study developed a DNN and combined it with an attention mechanism for data classification. DNN is a multilayer neural network that performs classification by learning complex nonlinear relationships between data points. Subsequently, the attention mechanism augments the model's focus on salient features, enhancing its classification performance. The attention mechanism assigns varying weights to the input features, enabling the model to prioritize the most pertinent features for the task. Its process is described as follows. (1) The attentional weights are calculated by transforming the input features using a hyperbolic tangent (tanh) activation function. Subsequently, the SoftMax activation function converts these transformed values into weights, representing each feature's importance in the current task. (2) The new feature representation is then calculated by multiplying the input features with their corresponding attentional weights, thus obtaining the weighted feature representation. This step enhances the relative importance of the salient features in the new representation.

The attention mechanisms layer pays particular attention to biomechanically relevant features such as tumor size and lymph node status, capturing the nonlinear effects of these variables on mechanosensitive pathways by assigning them higher weights.

Figure 2 illustrates the developed DNN model, which comprises four layers, i.e., an input layer, a hidden layer, an attention layer, and an output layer. The input layer receives the feature data, which is normalized to ensure each feature is at the same scale. The hidden layer comprises two internal layers, where the initial layer is a fully connected layer containing a specific number of neurons with a ReLU activation function. A dropout layer is added after the fully connected layer to prevent overfitting. The second layer comprises an additional fully connected layer containing fewer neurons. These layers utilize the ReLU activation function and are followed by a Dropout layer. The attention layer is incorporated after the second fully connected layer, enhancing the model's focus on salient features. This is achieved by calculating the attention weights and generating new feature representations. The final layer is the output layer, which employs a sigmoid activation function appropriate for binary classification tasks to generate the final classification results. It should be noted that the learning rate was adjusted using the Adam optimizer, which accelerated convergence dynamically during training. Finally, the binary cross-entropy loss function was selected to optimize the binary classification task.

The model uses the Adam optimizer to dynamically adjust the learning rate. Experiments show that Adam outperforms other optimizers in terms of convergence speed and generalization performance: compared with SGD (validation set loss value of 0.35) and Adagrad (validation set loss value of 0.33), Adam has a stable validation set loss value of 0.24 after 50 rounds of training (Supplementary **Figure S1**). Its adaptive learning rate mechanism effectively alleviates the gradient sparsity problem and accelerates model convergence.

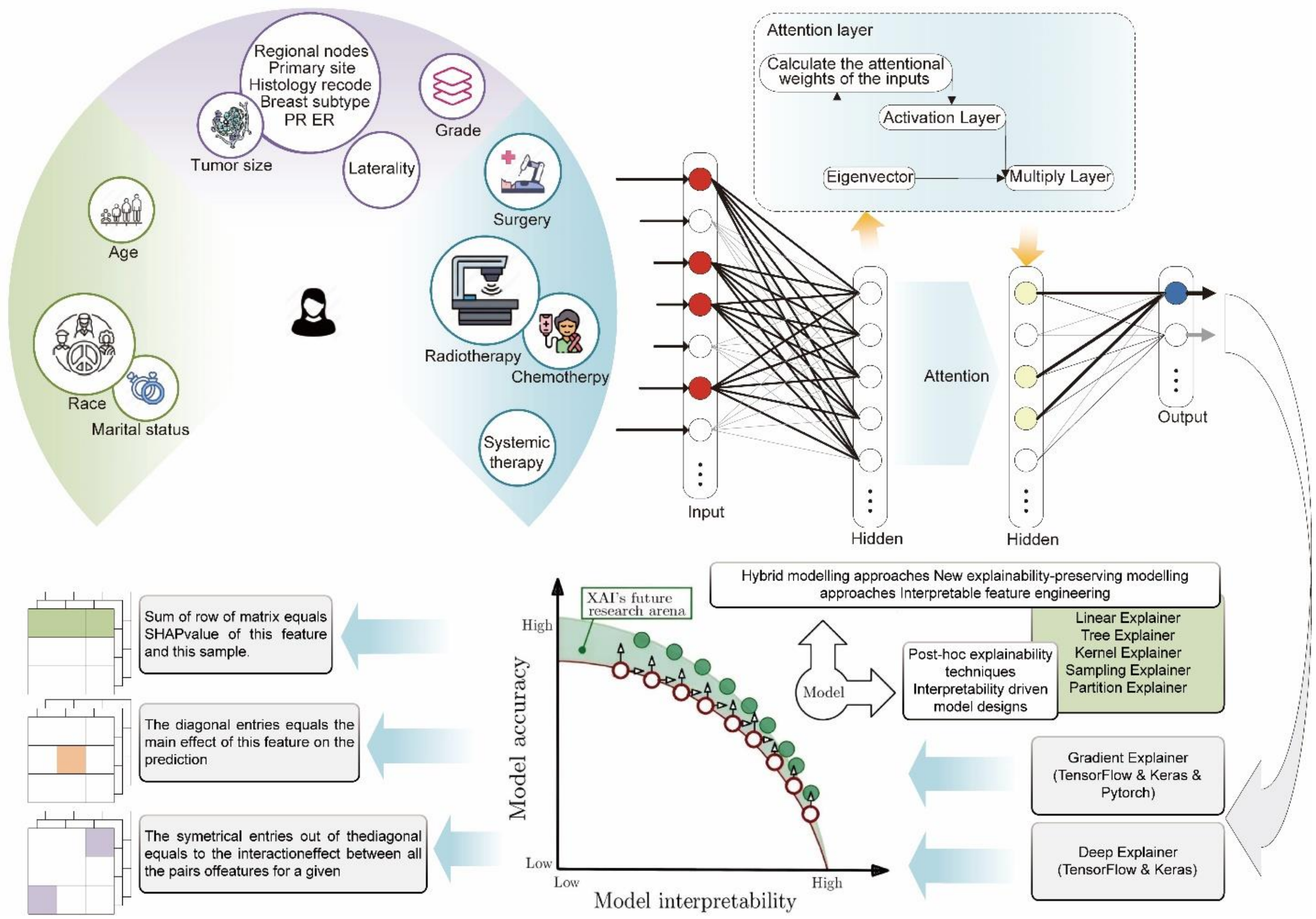
Regional nodes
Primary site
Histology recode
Breast subtype
PR ER
Tumor size
Laterality
Grade
Age
Surgery
Radiotherapy
Chemotherpy
Race
Marital status
Systemic therapy
Attention layer
Calculate the attentional weights of the inputs
Activation Layer
Eigenvector
Multiply Layer
Input
Hidden
Attention
Hidden
Output
Sum of row of matrix equals SHAPvalue of this feature and this sample.
The diagonal entries equals the main effect of this feature on the prediction
The symetrical entries out of thediagonal equals to the interactioneffect between all the pairs offeatures for a given
XAI's future research arena
High
Low
Model accuracy
Low
High
Model interpretability
Hybrid modelling approaches New explainability-preserving modelling approaches Interpretable feature engineering
Model
Post-hoc explainability techniques
Interpretability driven model designs
Linear Explainer
Tree Explainer
Kernel Explainer
Sampling Explainer
Partition Explainer
Gradient Explainer (TensorFlow & Keras & Pytorch)
Deep Explainer (TensorFlow & Keras)

Figure 2. Proposed DNN model architecture utilizing an attention mechanism. The left part shows the features that affect the model's input, and the right part describes the internal structure of the model, including the input layer, hidden layer, and output layer. The attention layer calculates the weights of the input features to enhance the influence of important information, followed by the activation and multiply layers to generate the final output.

The DNN's structure and its mathematical description are presented below. For a feature vector $x \in R^n$, where n is the number of features, placed at the input layer, the first hidden layer outputs h_1, expressed as follows:

$$h_1 = \mathrm{ReLU}(W_1 x + b_1) \tag{8}$$

where W_1 denotes a weight matrix of the first hidden layer, and b_1 is a bias vector of the first hidden layer.

Dropout is applied to the first hidden layer with a dropout rate of 0.5, adopting Equation (9). The output h_2 at the second hidden layer is expressed as:

$$h_1' = \mathrm{Dropout}(h_1, 0.5) \tag{9}$$

$$h_2 = \mathrm{ReLU}(W_2 h_1' + b_2) \tag{10}$$

The attention mechanism added in the proposed network has weights calculated using Equation (11), and feature weighting is performed using Equation (12).

$$A = \mathrm{softmax}(\tan h(W_a h_2 + b_a)) \tag{10}$$

where W_a is the weight matrix of the attention weighting computation layer, b_a is the bias vector of the attention weighting computation layer, and a is the attention weight vector. The feature representation weighted by attention is formulated as follows:

$$h_{\text{attention}} = h_2 \odot a \tag{12}$$

where $\odot$ denotes element-by-element multiplication.

The output layer also applies dropout with a 0.5 rate, with the computations of the output layer using Equation (13).

$$\hat{y} = \sigma(W_3 h'_{\text{attention}} + b_3) \tag{13}$$

where σ is the sigmoid activation function, w_3 denotes the weight matrix of the output layer, b_3 is the bias vector of the output layer, and $\hat{y}$ is the model's predicted output.

The loss function in the model compilation process uses the following binary cross-entropy.

$$\mathcal{L} = -\frac{1}{m}\sum_{i=1}^{m} [y^{(i)}\log(\hat{y}^{(i)}) + (1 - y^{(i)})\log(1 - \hat{y}^{(i)})] \tag{14}$$

where L is the loss function value, m is the number of samples, and $y(i)$ and $\hat{y}(i)$ denote the true and the predicted labels of the i-th sample. The model is trained for 50 epochs using small batch stochastic gradient descent with a batch size of 32.

2.6. Interpretability analysis

2.6.1. Vivid analysis

Variable importance (VImp), variable interaction measures (VInt), and partial dependence plots (PDPs) are some of the more important concepts in Vivid analysis. VImp defines the contribution of each independent variable to the model's predictive performance, determining which variables have the most significant impact. VInt assesses the impact of interactions between two or more variables on the model predictions, which helps identify complex dependencies. PDPs reveal the dependence between one or more independent variables and the target variable, excluding the influence of other variables. Specifically, PDPs plot the variable effect on the predicted outcome while keeping other variables constant and gradually changing the value of the target variable [29].

Embedded methods integrate VImp into machine learning algorithms. For instance, the random forest approach employs its tree-based architecture to assess model performance. Vivid analysis utilizes the minimum depth to quantify the variables' significance and interaction strength based on their position within the random tree to elucidate the internal mechanisms of random forests. Furthermore, Vivid analysis calculates importance scores through conditional inference on random forests. Regarding VInt, Friedman and Popescu [30] introduced the *H*-statistic, which is a model-independent measure that utilizes partial dependence to quantify interactions. It compares the joint effect of a pair of variables with the sum of their marginal effects and is defined as follows:

$$H_{jk}^2 = \frac{\sum_{i=1}^{n} [f_{jk}(x_{ij}, x_{ik}) - f_j(x_{ij}) - f_k(x_{ik})]^2}{\sum_{i=1}^{n} f_{jk}^2(x_{ij}, x_{ik})} \tag{15}$$

Friedman [31] developed the concept of PDPs as a model-independent method for visualizing the relationship between selected predictors and model results while averaging the effects of other predictors. Similarly, Goldstein et al. [32] proposed the Individual Conditional Expectation (ICE) curve, which describes the relationship between a specific predictor and the model results by setting the other predictors at a particular observation level. Essentially, PDPs represent the mean result of all ICE curves within the dataset. The partial dependence of the model fit function g on the predictor variables S is:

$$f_S(x_S) = \frac{1}{n}\sum_{i=1}^{n} g(x_s, x_{C_i}) \tag{16}$$

where S is a subset of the p predictor variables, C denotes the predictor other than S, x_{C1}, x_{C2}, …, x_{Cn} are the x_C values occurring in the n observations of the training set, and $g()$ is the predicted values of the machine learning model. The local dependence function $f_S(x_S)$ can be plotted for one or two variables to reveal the marginal fit.

2.6.2. SHAP analysis

SHAP is an explainable artificial intelligence technique that mathematically assigns a weight called a Shapley value to each feature of the training model. In cooperative game theory, the Shapley value was initially proposed to ensure fair gains distribution among features [33]. The Shapley value ϕ_i for a given feature i is calculated as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}[v(S \cup \{i\}) - v(S)] \tag{17}$$

where N represents the set of all features, S is the subset of features excluding feature i, and $v(S)$ is the model's predicted value when only the subset S features are included.

The deep SHAP method combines deep learning models and Shapley values to conduct a feature contribution analysis. This is achieved through the following steps. The initial step is to select an appropriate background data set and utilize it to model the feature distribution. Subsequently, an approximation is calculated using the properties of the deep model, whereby the marginal contribution of features is determined through a reasonable approximation method. Ultimately, the model is decomposed based on the hierarchical structure of the deep model, with the contribution of each feature calculated and accumulated layer by layer. All possible subsets of features that include and exclude the feature are considered for each feature. Notably, the discrepancy in model predictions, when a feature is included or excluded, indicates the marginal contribution of that feature. Deep SHAP values synthesize the following advantages: if the marginal contribution of a feature to the prediction increases, its SHAP value rises accordingly. The sum of the SHAP values of all features equals the model's predicted output, thereby ensuring the completeness and additivity of the interpretation.

Although calculating directly the Shapley values is highly complex, using Deep SHAP approximation methods allows for the efficient computation of SHAP values in deep learning models, thereby maintaining high computational efficiency. Using Deep SHAP facilitates a more transparent interpretation of the prediction outcomes of deep learning models, thereby facilitating a more comprehensive understanding of feature importance and model behavior.

3. Results

3.1. Sample characteristics analysis

Table 1 summarizes the patients stratified by survival ending. A total of 37,917 patients were enrolled in this study, of which 16,681 (43.99%) had a survival ending of dead and 21,236 (56.01%) had an alive ending. The patients were grouped based on age, tumor size, regional nodes positive, race, and grade. For the trials, surgery, radiotherapy, chemotherapy, breast subtype, ER, PR, primary site, histology recode, and marital status differences were statistically significant ($P < 0.05$), while laterality and systemic therapy differences were statistically insignificant ($P > 0.05$).

Table 1. Patient profile by survival status.

Variables	**Total (*n* = 37917)**	**Dead (*n* = 16681)**	**Alive (*n* = 21236)**	***P***
Age, Mean ± SD	60.97 ± 13.43	61.52 ± 14.90	60.53 ± 12.14	< 0.001
Tumor size, M (Q_1, Q_3)	20.0 (12.0, 35.0)	20.0 (12.0, 35.0)	20.0 (12.0, 35.0)	< 0.001
Regional nodes positive, M (Q_1, Q_3)	0.0 (0.0, 2.0)	2.00 (1.0, 7.0)	0.0 (0.0, 0.0)	< 0.001
Race, *n*(%)				< 0.001
Black	4497 (11.86)	2891 (17.33)	1606 (7.56)	
White	29,974 (79.05)	12,503 (74.95)	17,471 (82.27)	
Other	3446 (9.09)	1287 (7.72)	2159 (10.17)	
Primary site, n(%)				< 0.001
C50.0-Nipple	117 (0.31)	63 (0.38)	54 (0.25)	
C50.1-Central portion of breast	1958 (5.16)	1097 (6.58)	861 (4.05)	
C50.2-Upper-inner quadrant of breast	4437 (11.70)	1530 (9.17)	2907 (13.69)	
C50.3-Lower-inner quadrant of breast	2103 (5.55)	779 (4.67)	1324 (6.23)	
C50.4-Upper-outer quadrant of breast	12,773 (33.69)	5251 (31.48)	7522 (35.42)	
C50.5-Lower-outer quadrant of breast	2847 (7.51)	1220 (7.31)	1627 (7.66)	
C50.6-Axillary tail of breast	190 (0.50)	108 (0.65)	82 (0.39)	
C50.8-Overlapping lesion of breast	8828 (23.28)	3825 (22.93)	5003 (23.56)	
C50.9-Breast, NOS	4664 (12.30)	2808 (16.83)	1856 (8.74)	
Grade, *n*(%)				< 0.001
Well differentiated; Grade I	7468 (19.70)	735 (4.41)	6733 (31.71)	
Moderately differentiated; Grade II	15,518 (40.93)	5174 (31.02)	10,344 (48.71)	
Poorly differentiated; Grade III	14,900 (39.30)	10,748 (64.43)	4152 (19.55)	
Undifferentiated; anaplastic; Grade IV	31 (0.08)	24 (0.14)	7 (0.03)	

Table 1. (*Continued*).

Variables	**Total (*n* = 37917)**	**Dead (*n* = 16681)**	**Alive (*n* = 21236)**	***P***
Laterality, *n*(%)				0.684
Right	18,627 (49.13)	8175 (49.01)	10,452 (49.22)	
Left	19,290 (50.87)	8506 (50.99)	10,784 (50.78)	
Histology recode, *n*(%)				< 0.001
8560-8579: complex epithelial neoplasms	284 (0.75)	236 (1.41)	48 (0.23)	
8010-8049: epithelial neoplasms, NOS	148 (0.39)	101 (0.61)	47 (0.22)	
8390-8429: adnexal and skin appendage neoplasms	67 (0.18)	41 (0.25)	26 (0.12)	
8500-8549: ductal and lobular neoplasms	36,533 (96.35)	16,153 (96.83)	20,380 (95.97)	
8140-8389: adenomas and adenocarcinomas	344 (0.91)	86 (0.52)	258 (1.21)	
8440-8499: cystic, mucinous and serous neoplasms	541 (1.43)	64 (0.38)	477 (2.25)	
Surgery, *n*(%)				< 0.001
No	706 (1.86)	684 (4.10)	22 (0.10)	
Yes	37,211 (98.14)	15,997 (95.90)	21,214 (99.90)	
Radiotherapy, *n*(%)				< 0.001
No	1,7898 (47.20)	8486 (50.87)	9412 (44.32)	
Yes	20,019 (52.80)	8195 (49.13)	11,824 (55.68)	
Chemotherapy, *n*(%)				< 0.001
No	20,088 (52.98)	5665 (33.96)	14,423 (67.92)	
Yes	17,829 (47.02)	11,016 (66.04)	6813 (32.08)	
Systemic therapy, *n*(%)				0.433
No	8204 (21.64)	3578 (21.45)	4626 (21.78)	
Yes	29,713 (78.36)	13,103 (78.55)	16,610 (78.22)	
Breast subtype, *n*(%)				< 0.001
HR-/HER2-	5893 (15.54)	4486 (26.89)	1407 (6.63)	
HR-/HER2+	1710 (4.51)	1152 (6.91)	558 (2.63)	
HR+/HER2+	3682 (9.71)	1791 (10.74)	1891 (8.90)	
HR+/HER2-	26,632 (70.24)	9252 (55.46)	17,380 (81.84)	
ER, *n*(%)				< 0.001
Negative	8052 (21.24)	5955 (35.70)	2097 (9.87)	
Positive	29,865 (78.76)	10,726 (64.30)	19,139 (90.13)	
PR, *n*(%)				< 0.001
Negative	12,509 (32.99)	8481 (50.84)	4028 (18.97)	
Positive	25,408 (67.01)	8200 (49.16)	17,208 (81.03)	
Marital status, *n*(%)				< 0.001
Separated	436 (1.15)	239 (1.43)	197 (0.93)	
Widowed	5417 (14.29)	3023 (18.12)	2394 (11.27)	
Single (never married)	6032 (15.91)	3135 (18.79)	2897 (13.64)	
Divorced	4420 (11.66)	2033 (12.19)	2387 (11.24)	
Unmarried or domestic partner	127 (0.33)	53 (0.32)	74 (0.35)	
Married (including common law)	21,485 (56.66)	8198 (49.15)	13,287 (62.57)	

SD: Standard deviation, M: Median, Q_1: 1st Quartile, Q_3: 3st Quartile.

There was a significant interaction effect of race variables with tumor characteristics. Black patients had a higher rate of positive regional lymph nodes (17.33%) than whites (7.56%) and other races (10.17%), and the percentage of tumors graded as poorly differentiated (Grade III) was 64.43% (whites: 19.55%). In addition, the mean tumor size was significantly larger in black patients (25.0 mm, Q1–Q3: 15.0–40.0) than in whites (18.0 mm, Q1–Q3: 10.0–30.0, $p < 0.001$). SHAP analysis revealed that the negative contribution of the black race to model output (SHAP value = −0.15) was consistent with its higher lymph node metastasis rate and tumor malignancy, suggesting that racial differences may indirectly influence prognosis through the biomechanical microenvironment (e.g., degree of ECM fibrosis).

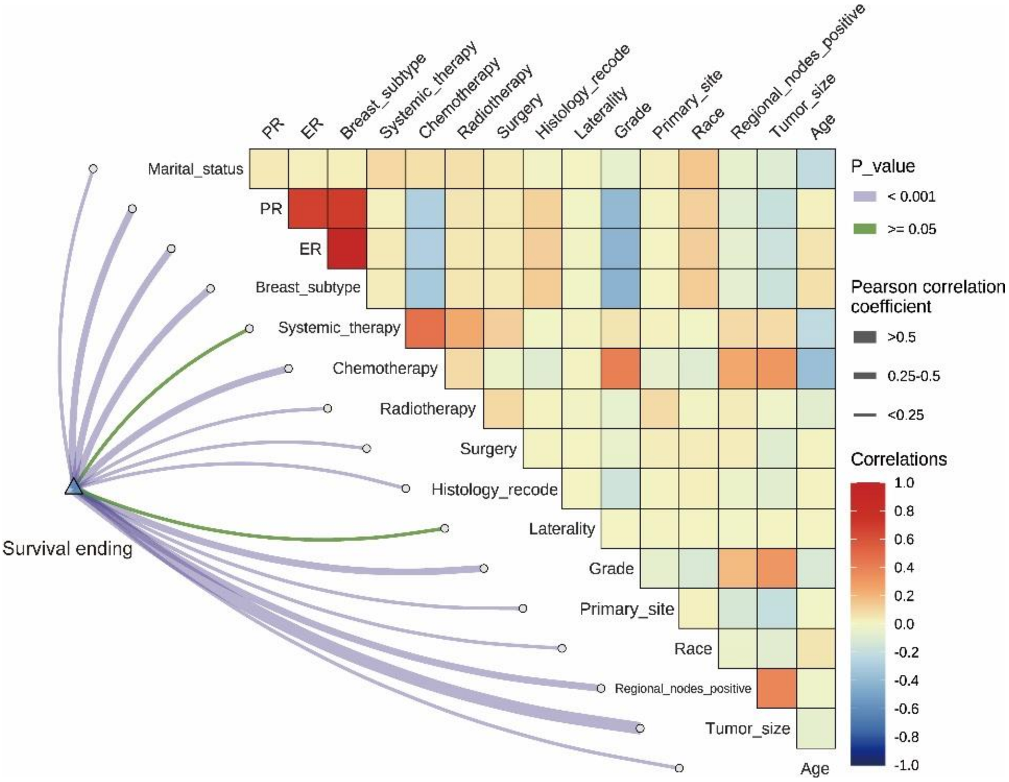

Figure 3. Pearson correlation heatmaps and p-value network diagrams.

This study employed Pearson's correlation coefficient to analyze the relationship between clinical and demographic factors and survival outcomes in breast cancer patients. **Figure 3** illustrates the correlation matrix, elucidating the significant associations between disparate variables. Notably, several factors demonstrated robust correlations, with tumor size and survival having a negative correlation, indicating that larger tumors were associated with poorer survival outcomes ($r < -0.5$, $p < 0.001$). Furthermore, positive regional lymph nodes demonstrated a negative correlation with survival, reinforcing their role as a prognostic indicator. Additional analysis revealed a positive correlation between tumor grade and survival outcome, indicating that lower-graded tumors may be associated with superior survival outcomes ($r > 0.5$). Other notable correlations included the relationship between age and survival, with results indicating that the older group exhibited poorer survival outcomes.

3.2. Model comparison

The performance of the ML and DNN models was evaluated on the validation and test sets using accuracy, precision, recall, and F1 scores. **Table 2** summarizes the corresponding results. Precisely, DNN demonstrated the highest performance across all metrics on the validation and test sets, with F1 scores of 0.928 and 0.935, respectively. The MLP model also exhibited strong performance, closely following DNN. Classical ML algorithms, such as SVM and Random Forest, demonstrated superior performance, although inferior to the DNN-based models. Decision Tree and Naive Bayes exhibited suboptimal performance compared to the other models.

Table 2. Performance comparison of different models on validation and test sets.

Dataset	Model	Accuracy	Precision	Recall	F1 score
Validation	SVM	0.899	0.899	0.922	0.910
Validation	Naive Bayes	0.816	0.800	0.894	0.844
Validation	Logistic Regression	0.896	0.894	0.923	0.908
Validation	Decision Tree	0.797	0.815	0.823	0.819
Validation	Random Forest	0.909	0.922	0.913	0.918
Validation	KNN	0.860	0.862	0.892	0.876
Validation	MLP	0.915	0.913	0.937	0.925
Validation	DNN	0.919	0.916	0.942	0.928
Test	SVM	0.903	0.901	0.929	0.915
Test	Naive Bayes	0.819	0.807	0.890	0.847
Test	Logistic Regression	0.898	0.893	0.930	0.911
Test	Decision Tree	0.797	0.815	0.825	0.820
Test	Random Forest	0.913	0.925	0.920	0.923
Test	KNN	0.868	0.865	0.907	0.885
Test	MLP	0.916	0.914	0.940	0.926
Test	DNN	0.926	0.920	0.950	0.935

The dynamics of loss and accuracy during the training process of DNN are provided in Supplementary **Figure S1**, which reveals that the training loss decreases significantly as the epochs increase, with an initial value of about 0.36 and stabilizing after 30 epochs to 0.24. This indicates that the model fits well in the training set. The validation loss also shows a decreasing trend in the initial phase, but its decrease is slightly smaller and finally stabilizes at 0.24, implying that the model has good generalization ability. The training accuracy increases rapidly in the initial stage and converges to 0.91 after 50 epochs, indicating that the model's performance on the training set reaches a high level. The validation accuracy follows a similar trend to the training accuracy and finally reaches 0.92, showing the model's strong prediction ability on the validation set. Regarding performance metrics, DNN performs the best but has a marginal advantage due to the relatively large sample size of the dataset. In Supplementary **Figure S2**, the overall value of correct sample classification prediction can be observed from the validation and test sets, demonstrating the DNN's advantage.

Figure 4 depicts the receiver operating characteristic (ROC) curve of all competitor methods and the confusion matrix of the DNN model on the validation and test sets. The DNN model exhibits the highest performance on both sets, with an area under the curve (AUC) of up to 0.96, which is markedly superior to the average of the other ML models. The decision tree model exhibits suboptimal performance, with an AUC of 0.88 (**Figure 4A**). It should be noted that the modeling strategy employed demonstrated DNN's optimal performance for categorical prediction of the patients' survival outcomes. Indeed, from the 7583 patients in the validation set, 3965 were correctly identified as alive, 3028 were correctly identified as deceased, and 326 and 264 were misclassified in the false positive (FP) and false negative (FN) groups, respectively. In the test set involving 7584 patients, 4040 patients were correctly predicted as alive, 3002 patients as dead, and 325 and 217 patients were misclassified as FP and FN, respectively (**Figure 4B**).

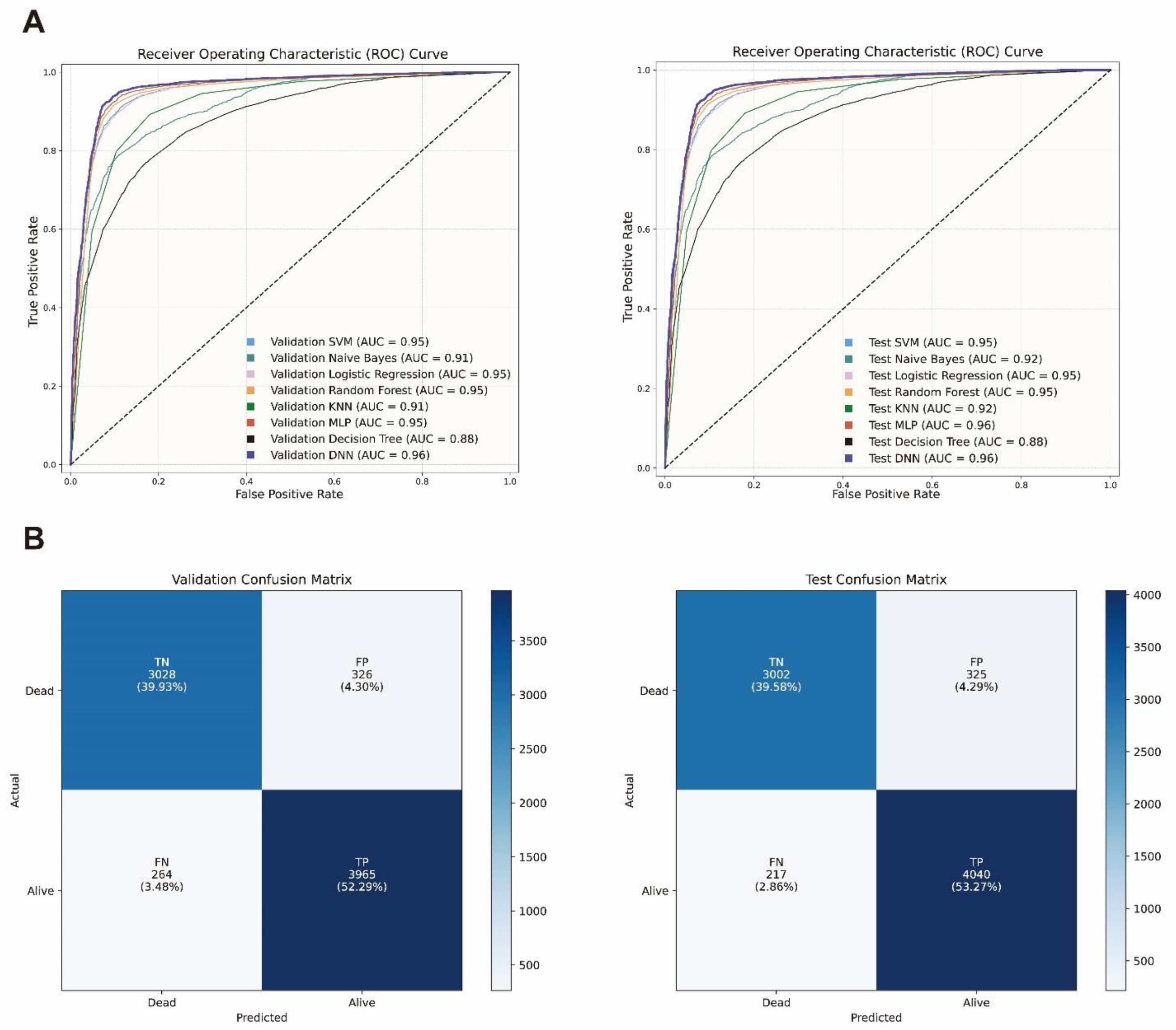

Figure 4. ROC curves for the model and confusion matrices for the DNN model, **(A)** ROC curves of the model on the validation and test sets with the AUC of the model labeled; **(B)** confusion matrix of DNN models on validation and test sets.

TN, true negative; FP, false positive; FN, false negative; TP, true positive.

Supplementary **Figure S3** compares all models, suggesting that the Decision Tree and Naive Bayes models have relatively low performance and are located in the lower correlation coefficient and standard deviation region, implying poor prediction performance. The KNN and Random Forest perform relatively well and are in the higher region of the correlation coefficient. The Random Forest model performs well in both standard deviation and correlation coefficient, suggesting high accuracy and stability in prediction. Logistic regression, MLP, and DNN show excellent performance in the graph, located in the region of high correlation coefficient and slight standard deviation. This indicates that these models are more closely related to the actual values in terms of predicted values.

3.3. Model interpretability analysis

This study conducted interpretability analysis on DNN, the core model of this study, and the Random Forest model, which is the best-performing ML model. DNN was analyzed using the SHAP method, and the Random Forest model was analyzed using the Vivid method. This decision was made because the random forest model has a more explicit tree structure, and the Vivid method can effectively capture and demonstrate the global importance of features within the tree model and the intricate interactions between features. Furthermore, the Vivid method exhibits high computational efficiency and intuitive visualization capabilities in interpreting the tree model. Concerning the DNN, which has a complex model structure and nonlinear solid relationships, the SHAP method provides a unified framework based on game theory and thus accurately assigns the contribution of each feature to the prediction results. This method can comprehensively explain the prediction mechanism of DNN models.

Figure 5 presents the results of the Vivid analysis on the Random Forest model, including a heatmap, network diagram, and generalized biased dependency pair diagram. The heatmap and network diagram highlight that regional nodes positive, tumor size, grade, PR, age, and ER are the six most important variables in the prediction process of the random forest model. Notably, regional nodes positive, tumor size, and grade show a strong interaction in the random forest model, and regional nodes positive show a moderate interaction with PR, age, ER, and radiotherapy (**Figure 5A,B**). The generalized partial dependency plot depicts the six variables with the highest order of importance. The univariate partial dependence plot with ICE curves is presented on the diagonal of this figure, and the rest of the figure is a bivariate partial dependence plot. From the univariate and bivariate partial dependence coefficients, it is evident that regional nodes positive, tumor size and grade significantly affect the response. Moreover, PR, ER, and age have less effect on the response (**Figure 5C**). Partial dependency plots for another 10 variables are depicted in Supplementary **Figure S4**.

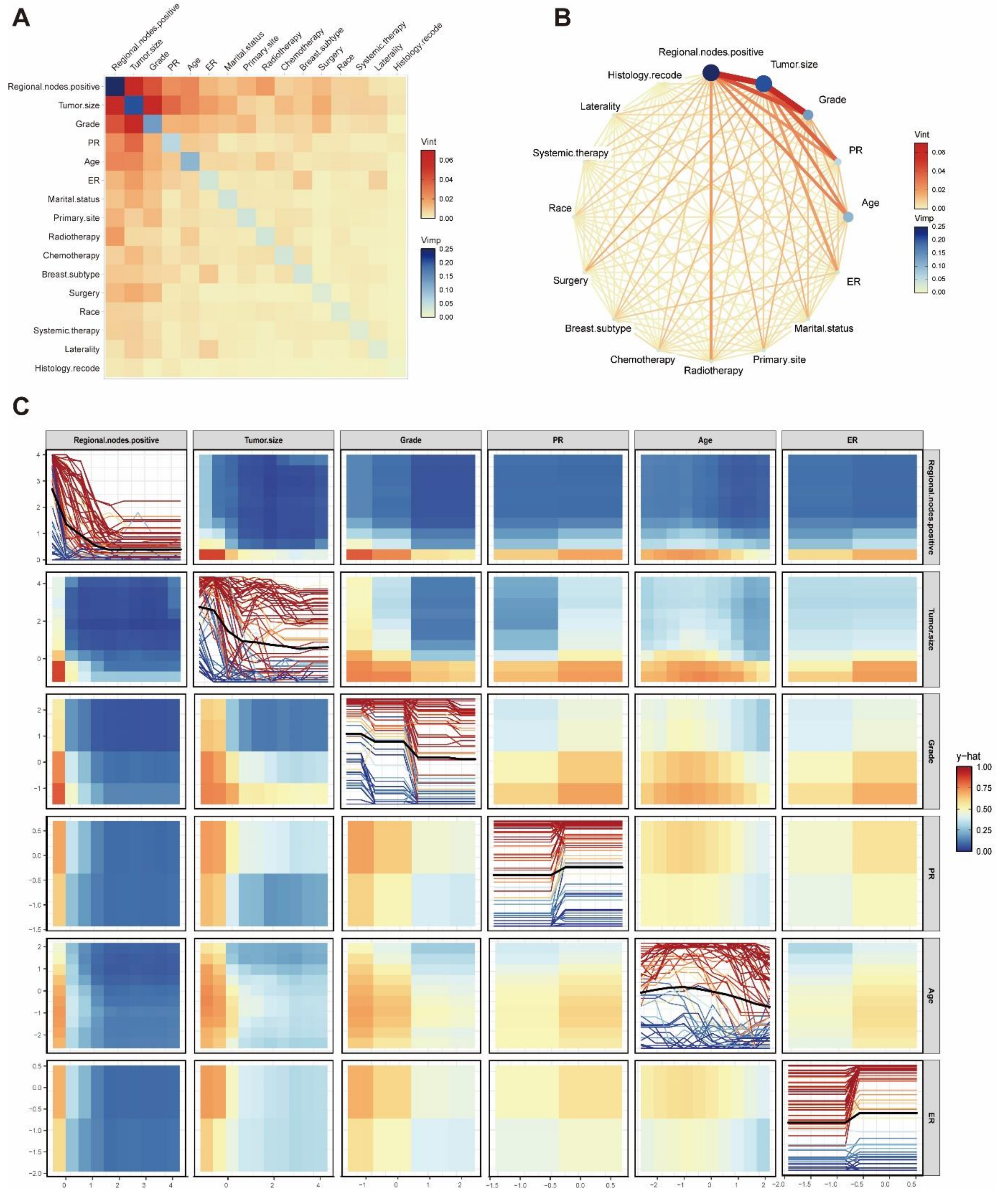

Figure 5. Random Forest Model Vivid analysis results, **(A)** the random forest-fitted heatmap shows the strength of the diagonal’s two-way interactions and the importance of individual variables on the diagonal; **(B)** the random forest-fitted network plot shows the strength of two-way interactions and the importance of individual variables; **(C)** the pairs partial dependence plot for the random forest model.

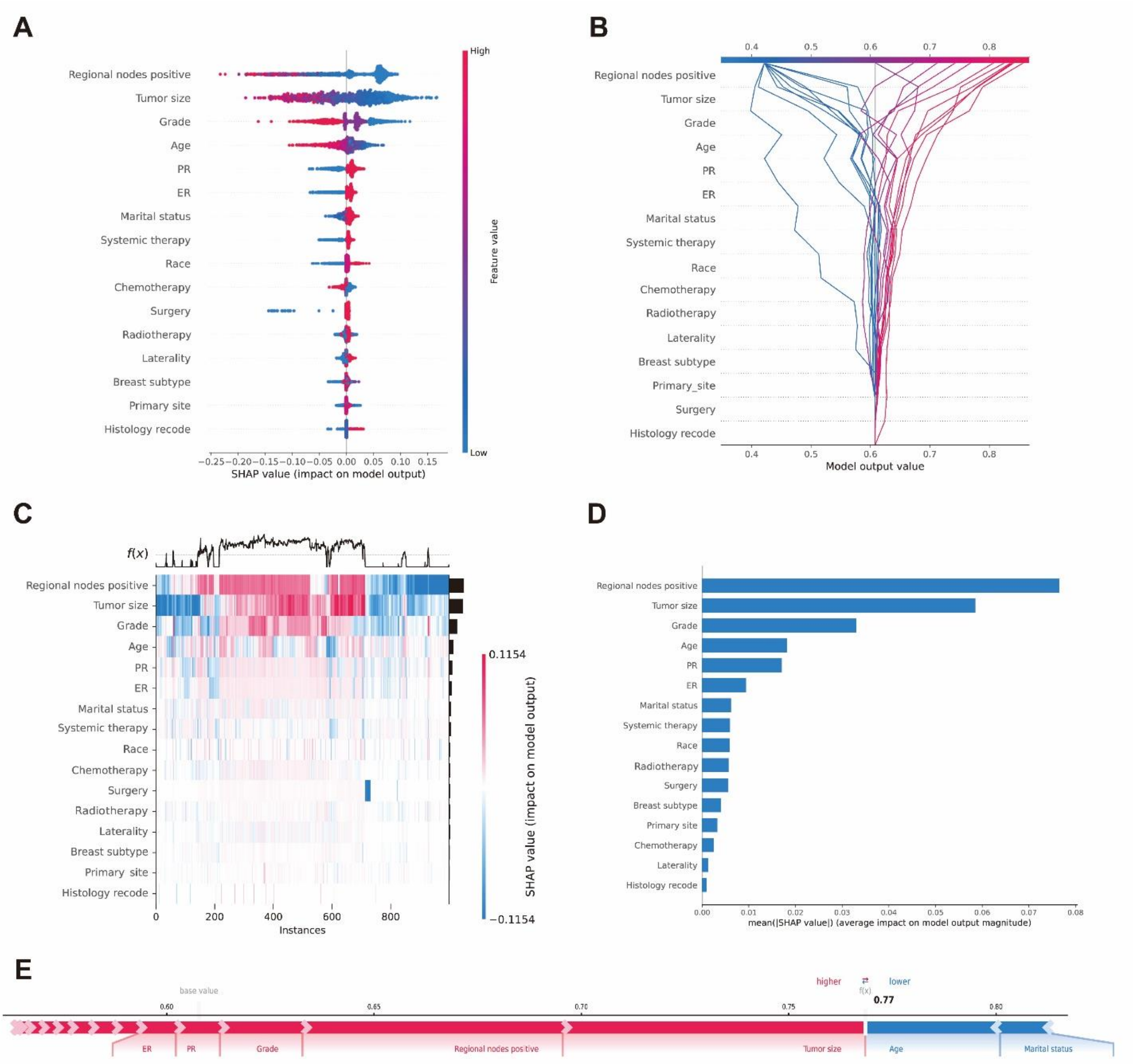

Figure 6. DNN model SHAP analysis results, **(A)** bee swarm plot ranks the features based on the sum of the size of the SHAP values for all samples. The color represents the feature values (high in red, low in blue). The *X*-axis indicates the effect on the model output (positive on the right, negative on the left); **(B)** the decision plot shows the 20 test observations. The *X*-axis represents the model output (in log odds), and the *Y*-axis lists the features in order of importance. Each colored line represents an observation, and the line color corresponds to the predicted value of the observation; **(C)** heatmap of SHAP values with instances on the *X*-axis and model inputs on the *Y*-axis; colors represent SHAP values. The samples are sorted by similarity hierarchy clustering, and the model outputs are displayed above the heatmap, with the importance bars of the input features on the right; **(D)** SHAP feature importance as measured by the average absolute shapley value of the DNN model; **(E)** a force plot is used to interpret predictions for individual samples. Each attribute value acts as a force that increases (red) or decreases (blue) the prediction. Predictions start at the baseline, which is a constant for the model, and each attributed value is represented by an arrow showing a positive or negative contribution to the prediction.

Figure 6 illustrates the results of the SHAP analysis of the DNN model. Each point in the bee swarm plot represents a sample, the sample size is stacked vertically, and the colors indicate the eigenvalues (red corresponds to high values, and blue corresponds to low values). Tumor size, for example, suggests that a larger tumor size (red) negatively affects prediction and a smaller tumor size (blue) positively impacts prediction. It is worth noting that when the marital status is unmarried or domestic partner and married (including common law), there is a positive effect on prediction, and when the marital status is separated, widowed, and single. Race has a minor impact on the predictions, but interestingly, for the predicted outcomes, other (American Indian/AK Native, Asian/Pacific Islander), white, and black have a negative effect. Race's risk of death is progressively decreasing (**Figure 6A**). Besides, the decision plot shows the model output analyzed using SHAP values, as different features affect the predicted outcomes. The contribution of each feature on the model output value is represented by a line, with values ranging from 0.4 to 0.8. **Figure 6A** highlights that the feature regional nodes positive has the most significant impact on the model output, exhibiting a markedly higher output value than the other features. This suggests that this feature plays a pivotal role in the prediction, which is consistent with the clinical practice of associating a positive status of lymph nodes with a worse prognosis.

Additionally, the SHAP analysis reveals that tumor size and grade significantly influence the model output. The tumor size and grade output values range from 0.5 to 0.7, indicating that these variables substantially influence prognosis. In comparison, features such as age, PR, and ER exhibited a relatively lesser impact, with output values between 0.4 and 0.6 (**Figure 6B**). The heatmap illustrates the influence of SHAP values for each feature on the model output. The color shade of each line represents the positive (red) or negative (blue) impact of the corresponding feature on the model output, with the uppermost black line representing the overall output value of the model. The regional node's positive feature exerts a discernible positive influence in most instances, thereby underscoring its pivotal role in model determinations. Additional characteristics, such as tumor size and grade, also demonstrate a notable impact, albeit to a lesser extent than that observed for regional nodes positive.

As illustrated in the heatmap, the SHAP values of distinct instances exhibit disparate patterns, suggesting that the impact of specific characteristics on an individual may deviate from the prevailing trend. This variability underscores the necessity of incorporating patient characteristics into individualized treatment plans (**Figure 6C**). The influence of each feature on the mean SHAP value of the model output is subsequently quantified via a bar chart, demonstrating that regional nodes positive tumor size, and grade are the three most influential features on the model output. These features significantly influence the decision-making process after the bee swarm plot, decision plot, and heatmap results. Features such as age, PR, and ER are ranked lower (**Figure 6D**).

Moreover, the force plot illustrates the distribution of the SHAP values of the features concerning the baseline values. The black line in the center represents the baseline value of the model (0.77), which is shifted upward or downward with the influence of different features. The red and blue arrows in the figure indicate an increase and decrease in the output, respectively. Features such as regional nodes positively and significantly boost the model output, while age and PR have a relatively small effect (**Figure 6E**).

4. Discussion

This study introduced a method for predicting patient survival outcomes using breast cancer patient data from the SEER database. This was achieved using ML and DNN models, with the corresponding results demonstrating that the DNN model, when combined with the attention mechanism, significantly outperforms the traditional ML method regarding classification performance. Furthermore, it has high interpretability.

During the model construction process, it was determined that the correlations between the factors were minimal, indicating that each feature provided independent information that could be effectively utilized for model training. The correlation between most factors and the target is less than 0.5, which is a challenge for model training, as the model cannot effectively capture the relationship between the input features and the output results, reducing prediction performance. Accordingly, this study proposed using DNN models and attention mechanisms to mitigate the effects of small correlations between variables and targets. Indeed, DNNs can learn complex hierarchical feature representations and extract meaningful patterns from data even when initial correlations are weak. At the same time, the attention mechanism in the embedded model dynamically assigns weights to the input features, adjusting them according to their importance to the prediction task [34–36]. This combination emphasizes the key features while reducing the impact of irrelevant features. The proposed adaptive weighting strategy can also identify critical feature interactions for prediction performance. Furthermore, incorporating a dropout layer allows regularizing the model, mitigating overfitting, and enhancing the model's generalization ability [37].

A comparison of multiple machine learning algorithms revealed that the DNN model demonstrated superior accuracy, precision, recall, and F1 score on the validation and test sets, particularly in AUC metrics, which reached 0.96. This suggests that the DNN model is more robust and reliable in predicting survival outcomes for breast cancer patients. Besides, we identified several variables that contributed most to the model prediction through Vivid and SHAP analyses, including positive regional nodes, tumor size, grade, PR, age, and ER. It is worth noting the significant positive effect observed for regional nodes in several models, which aligns with the established knowledge of lymph node status as a prognostic indicator for breast cancer in clinical practice. In this study, Vivid and SHAP analyses identified important features and revealed their interactions. For instance, there was a moderate level of interaction between regional nodes positive and PR, age, ER, and radiotherapy, which provides further justification for clinical decision-making. Indeed, the probability of patient survival decreased significantly with increasing tumor size. Subsequent analysis revealed a significant correlation between tumor size, age, and ER status, which collectively influenced patient prognosis. PR and ER status demonstrated significant effects in multiple models. Precisely, the results demonstrated that PR and ER were positively associated with a superior survival prognosis. This finding is consistent with existing clinical studies and further confirms the importance of hormone receptor status as a prognostic indicator. Noteworthy, a significant interaction was observed between PR and ER status with radiotherapy and tumor grade, which provides additional considerations for treatment decisions.

Our study has a significant advantage over previous studies in this field. The intricacy of ML presents a considerable challenge for clinicians seeking transparent and interpretable models, as the decision paths within these systems are often difficult to comprehend [38]. In lieu of this, we employed interpretable analysis to elucidate the decision-making process of DNN, thereby effectively addressing the issue mentioned above. In this study, data from 37,917 female breast cancer patients in the SEER database were utilized, which, due to its large sample size, enhances the reliability and statistical significance of the results. The SEER database encompasses patients from diverse ethnicities and geographic regions, strengthening the study's representativeness and improving the model's applicability to different populations. Combining DNN with the attention mechanism can effectively learn complex nonlinear relationships, enhancing the model's classification performance and interpretability. The study encompasses various variables influencing patient survival, improving the predictive model's comprehensiveness and accuracy. This approach avoids the limitations of relying on a single factor.

Despite the notable advancements in model performance and interpretability, this study has inherent limitations. While the SEER database encompasses a significant proportion of the US population, its data may not be wholly representative of breast cancer patients globally, as patients from different regions and ethnicities may have different disease characteristics and treatment responses. Therefore, accessing more international datasets to validate the model's generalizability would be beneficial [39]. The data utilized in this study was from the SEER database, which may have resulted in incomplete or inaccurate data. For instance, some crucial biological characteristics (such as gene mutations and immune status) are not meticulously documented in the SEER database, potentially compromising the precision of the model's predictions. Despite the efficacy of DNN models, their intrinsic complexity and demand for substantial computational resources during training and deployment present significant challenges. This may prove challenging in some resource-limited healthcare environments, necessitating the development of more efficient algorithms and optimization techniques.

In this study, tumor size was found to be a central variable in survival prediction, a result that may correlate with reprogramming of the tumor mechanistic microenvironment. Larger tumors are often accompanied by increased ECM cross-linking and up-regulation of transforming growth factor-β (TGF-β) secretion, which promotes collagen deposition through activation of stromal fibroblasts to create a pro-metastatic rigid microenvironment. This biomechanical remodeling may explain why increased tumor volume is significantly associated with poor prognosis. Regional lymph node positivity, as the strongest predictor, may reflect the metastatic ability of cancer cells through mechanosensitive migration pathways. It has been shown that elevated ECM stiffness enhances cytoskeletal contractility and promotes invasive pseudopod formation through activation of RhoA/ROCK signaling. This hypothesis could be tested in the future by quantifying in situ stiffness through elastography of isolated tumor tissues (e.g., ultrasound shear wave elastography).

Risk stratification based on model predictions allows for differentiated treatment regimens:

- High-risk patients (>80% predicted mortality): intensive treatment regimens (e.g., neoadjuvant chemotherapy + radical surgery + postoperative radiotherapy) are recommended, with close monitoring of ECM stiffness (via ultrasound elastography) to assess metastatic risk.
- Intermediate-risk patients (30%–80% predicted mortality): targeted therapy (e.g., CDK4/6 inhibitors) combined with biomechanical interventions (e.g., LOX inhibitors to reduce tumor stress) is recommended.
- Low-risk patients (<30% predicted mortality): breast-conserving surgery + endocrine therapy with regular follow-up on quality of life (via FACT-B scale) to optimize rehabilitation plan.

The above strategies refer to the NCCN guidelines (2023) and the weighting of tumor size, lymph node status and ethnicity in treatment selection was validated by SHAP analysis.

The generalizability of the present model needs further validation. For example, the estrogen receptor positivity rate of Asian breast cancer patients (85%) was significantly higher than that of the SEER database (78.76%), whereas the validation of other races in the dataset showed a decrease in the AUC of the present model from 0.96 to 0.92. This discrepancy may be related to the tumor biology (e.g., a higher HER2 positivity rate) and the treatment pattern of Asian patients. Future multicenter data need to be included to optimize the cross-ethnic applicability of the model.

The findings of this study offer compelling evidence in favor of making treatment decisions for breast cancer patients on an individual basis. By identifying key prognostic factors, clinicians can more accurately assess a patient’s prognosis and develop individualized treatment plans. For instance, more aggressive treatment strategies may be indicated for patients with positive regional lymph nodes and a high tumor grade. Furthermore, the high interpretability of the model in this study renders it more actionable in clinical applications. By interpreting the model prediction results, physicians can communicate treatment plans and expectations more effectively with patients. Future research could consider the following avenues to enhance the model’s applicability. The first is extending the data set, combining more regions and larger datasets, thereby strengthening the model’s generalizability. Furthermore, integrating additional modal data, including genomics, imaging, and other multidimensional data, can enhance the model’s prediction performance and interpretability. In addition to real-time model updating, the model is updated regularly by continuously collecting new patient data, thus improving its predictive performance. It is anticipated that, through further optimization and extension, this method will play an essential role in clinical practice and contribute to the development of individualized medicine.

5. Conclusions

A DNN-based method was successfully constructed to predict survival outcomes of female breast cancer patients and was analyzed using a large-scale dataset from the

SEER database. The results highlighted that the DNN model incorporating the attention mechanism significantly outperformed traditional machine learning models in survival prediction and presented good interpretability. Regarding the interpretability analysis, we identified key prognostic factors such as regional lymph node positivity, tumor size, and tumor grade using the Vivid method and SHAP analysis. These factors were critical in model prediction and revealed the complex interrelationships between variables, providing valuable clinical decision-making references. In addition, SHAP analysis demonstrated the specific impact of each feature on the model output, further enhancing the transparency and credibility of the model in clinical applications.

The DNN model constructed in this study reveals the prognostic value of tumor size and lymph node status as biomechanical microenvironmental markers. Increased tumor size may drive malignant progression through elevated internal solid stress and pro-fibrotic signaling, whereas lymph node metastasis suggests activation of mechanosensitive migratory pathways. Follow-up studies may further quantify these mechanisms through multimodal biomechanical assays (e.g., tumor elastography, single-cell tensiometry) and facilitate the precise implementation of mechanically targeted therapies.

The methodology of this study has the potential for broad clinical application to support personalized treatment decisions and lays the foundation for further research. Through continuous expansion of the dataset and multimodal data integration, future studies are expected to enhance the model's accuracy and applicability and contribute to individualized medicine for breast cancer.

Future research can be expanded in the following directions: first, integrating single-cell sequencing technology to resolve tumor cell heterogeneity and identify mechanosensitive genes (e.g., RhoA/ROCK pathway-related genes) to quantify biomechanical drivers; second, combining spatial transcriptomics technology to map the spatial correlation between the distribution of collagen fibrils and gene expression in the tumor microenvironment, revealing the mechanism of mechanical stress regulation on metastasis; and third, combining spatial transcriptomics technology to map the spatial correlation between collagen fiber distribution and gene expression in the tumor microenvironment., develop lightweight DNN models (e.g., knowledge distillation techniques) to adapt to resource-limited medical scenarios.

Supplementary materials: Figure S1: Loss curves and Accuracy curves of DNN models on training and validation sets; **Figure S2**: The following section outlines the specific model details for TP, TN, FP, and FN on both the validation and test sets; **Figure S3**: Taylor diagram: standard deviation and correlation coefficient analysis of different models; **Figure S4**: Univariate partial dependence plot: characterizing impact analysis with ICE curves.

Acknowledgments: Thanks to the authors in the references, and to Junyi Geng and Hongyu Chen for helpful discussions.

Conflict of interest: The author declares no conflict of interest.

Abbreviations

AUC	Area under curve
CNN	Convolutional neural networks
DNN	Deep neural networks
FN	False negative
FP	False positive
ICE	Individual conditional expectation
KNN	K-nearest neighbor
ML	Machine learning
MLP	Multilayer perceptron
PDPs	Partial dependence plots
ROC	Receiver operating characteristic
SVM	Support vector machine
TN	True negative
TP	True positive
VImp	Variable importance
VInt	Variable interaction measures

References

1. Wu Y, Zhang Y, Duan S, et al. Survival prediction in second primary breast cancer patients with machine learning: An analysis of SEER database. Computer Methods and Programs in Biomedicine. 2024; 254: 108310. doi: 10.1016/j.cmpb.2024.108310
2. Dell'Aquila K, Vadlamani A, Maldjian T, et al. Machine learning prediction of pathological complete response and overall survival of breast cancer patients in an underserved inner-city population. Breast Cancer Research. 2024; 26(1). doi: 10.1186/s13058-023-01762-w
3. Yu Y, Ren W, He Z, et al. Machine learning radiomics of magnetic resonance imaging predicts recurrence-free survival after surgery and correlation of LncRNAs in patients with breast cancer: a multicenter cohort study. Breast Cancer Research. 2023; 25(1). doi: 10.1186/s13058-023-01688-3
4. Aldrighetti CM, Niemierko A, Van Allen E, et al. Racial and Ethnic Disparities Among Participants in Precision Oncology Clinical Studies. JAMA Network Open. 2021; 4(11): e2133205. doi: 10.1001/jamanetworkopen.2021.33205
5. Naik K, Goyal RK, Foschini L, et al. Current Status and Future Directions: The Application of Artificial Intelligence/Machine Learning for Precision Medicine. Clinical Pharmacology & Therapeutics. 2024; 115(4): 673-686. doi: 10.1002/cpt.3152
6. Deo RC, Nallamothu BK. Learning About Machine Learning: The Promise and Pitfalls of Big Data and the Electronic Health Record. Circulation: Cardiovascular Quality and Outcomes. 2016; 9(6): 618-620. doi: 10.1161/circoutcomes.116.003308
7. Mahoro E, Akhloufi MA. Applying Deep Learning for Breast Cancer Detection in Radiology. Current Oncology. 2022; 29(11): 8767-8793. doi: 10.3390/curroncol29110690
8. Ming C, Viassolo V, Probst-Hensch N, et al. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Research. 2019; 21(1). doi: 10.1186/s13058-019-1158-4
9. Zhou BY, Wang LF, Yin HH, et al. Decoding the molecular subtypes of breast cancer seen on multimodal ultrasound images using an assembled convolutional neural network model: A prospective and multicentre study. eBioMedicine. 2021; 74: 103684. doi: 10.1016/j.ebiom.2021.103684
10. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nature Communications. 2020; 11(1). doi: 10.1038/s41467-020-15027-z

11. Wang Y, Acs B, Robertson S, et al. Improved breast cancer histological grading using deep learning. Annals of Oncology. 2022; 33(1): 89-98. doi: 10.1016/j.annonc.2021.09.007
12. Stashko C, Hayward MK, Northey JJ, et al. A convolutional neural network STIFMap reveals associations between stromal stiffness and EMT in breast cancer. Nature Communications. 2023; 14(1). doi: 10.1038/s41467-023-39085-1
13. Jiang M, Li CL, Luo XM, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. European Journal of Cancer. 2021; 147: 95-105. doi: 10.1016/j.ejca.2021.01.028
14. Poirion OB, Jing Z, Chaudhary K, et al. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. Genome Medicine. 2021; 13(1). doi: 10.1186/s13073-021-00930-x
15. Hussain H, Tamizharasan PS, Rahul CS. Design possibilities and challenges of DNN models: a review on the perspective of end devices. Artificial Intelligence Review. 2022; 55(7): 5109-5167. doi: 10.1007/s10462-022-10138-z
16. Du M, Liu N, Hu X. Techniques for interpretable machine learning. Communications of the ACM. 2019; 63(1): 68-77. doi: 10.1145/3359786
17. Bifarin OO. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. PLOS ONE. 2023; 18(5): e0284315. doi: 10.1371/journal.pone.0284315
18. Farzipour A, Elmi R, Nasiri H. Detection of Monkeypox Cases Based on Symptoms Using XGBoost and Shapley Additive Explanations Methods. Diagnostics. 2023; 13(14): 2391. doi: 10.3390/diagnostics13142391
19. Ren J, Li Y, Zhou J, et al. Developing machine learning models for personalized treatment strategies in early breast cancer patients undergoing neoadjuvant systemic therapy based on SEER database. Scientific Reports. 2024; 14(1). doi: 10.1038/s41598-024-72385-0
20. Rochlin DH, Barrio AV, McLaughlin S, et al. Feasibility and Clinical Utility of Prediction Models for Breast Cancer–Related Lymphedema Incorporating Racial Differences in Disease Incidence. JAMA Surgery. 2023; 158(9): 954. doi: 10.1001/jamasurg.2023.2414
21. Huang S, Cai N, Pacheco PP, et al. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genom. Cancer Genomics & Proteomics. 2018; 15(1). doi: 10.21873/cgp.20063
22. Langarizadeh M, Moghbeli F. Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. Acta Informatica Medica. 2016; 24(5): 364. doi: 10.5455/aim.2016.24.364-369
23. Boateng EY, Abaye DA. A Review of the Logistic Regression Model with Emphasis on Medical Research. Journal of Data Analysis and Information Processing. 2019; 07(04): 190-207. doi: 10.4236/jdaip.2019.74012
24. de Ville B. Decision trees. WIREs Computational Statistics. 2013; 5(6): 448-455. doi: 10.1002/wics.1278
25. Parmar A, Katariya R, Patel V. A review on random forest: An ensemble classifier. In: Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. 7–8 August 2019; Coimbatore, India. pp. 758-763.
26. Cunningham P, Delany SJ. k-Nearest Neighbour Classifiers - A Tutorial. ACM Computing Surveys. 2021; 54(6): 1-25. doi: 10.1145/3459665
27. Rana A, Singh Rawat A, Bijalwan A, et al. Application of multi layer (perceptron) artificial neural network in the diagnosis system: A systematic review. In: Proceedings of the 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE). 22–24 August 2018; San Salvador, El Salvador. pp. 1-6.
28. Clift AK, Dodwell D, Lord S, et al. Development and internal-external validation of statistical and machine learning models for breast cancer prognostication: cohort study. BMJ. Published online May 10, 2023: e073800. doi: 10.1136/bmj-2022-073800
29. Inglis A, Parnell A, Hurley C. vivid: An R package for variable importance and variable interactions displays for machine learning models. arXiv. 2022; arXiv:2210.11391. doi: 10.48550/arXiv.2210.11391
30. Friedman JH, Popescu BE. Predictive learning via rule ensembles. The Annals of Applied Statistics. 2008; 2(3). doi: 10.1214/07-aoas148
31. Friedman JH. Greedy function approximation: A gradient boosting machine. The Annals of Statistics. 2001; 29(5). doi: 10.1214/aos/1013203451
32. Goldstein A, Kapelner A, Bleich J, et al. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. Journal of Computational and Graphical Statistics. 2015; 24(1): 44-65. doi: 10.1080/10618600.2014.907095

33. Vimbi V, Shaffi N, Mahmud M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Informatics. 2024; 11(1). doi: 10.1186/s40708-024-00222-1
34. Guang Y, Wang W, Song H, et al. Prediction of external corrosion rate for buried oil and gas pipelines: A novel deep learning method with DNN and attention mechanism. International Journal of Pressure Vessels and Piping. 2024; 209: 105218. doi: 10.1016/j.ijpvp.2024.105218
35. Hacene GB, Mauch L, Uhlich S, et al. DNN quantization with attention. arXiv. 2021; arXiv:2103.13322v1. doi: 10.48550/arXiv.2103.13322
36. Senda J, Tanaka M, Iijima K, et al. Auditory stimulus reconstruction from ECoG with DNN and self-attention modules. Biomedical Signal Processing and Control. 2024; 89: 105761. doi: 10.1016/j.bspc.2023.105761
37. Kukačka J, Golkov V, Cremers D. Regularization for deep learning: A taxonomy. arXiv. 2017; arXiv:1710.10686v1. doi: 10.48550/arXiv.1710.10686
38. Gao J, Lu Y, Ashrafi N, et al. Prediction of sepsis mortality in ICU patients using machine learning methods. BMC Medical Informatics and Decision Making. 2024; 24(1). doi: 10.1186/s12911-024-02630-z
39. Sammut SJ, Crispin-Ortuzar M, Chin SF, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2021; 601(7894): 623-629. doi: 10.1038/s41586-021-04278-5
40. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO reports. 2014; 15(12): 1243-1253. doi: 10.15252/embr.201439246
41. Stylianopoulos T, Martin JD, Chauhan VP, et al. Causes, consequences, and remedies for growth-induced solid stress in murine and human tumors. Proceedings of the National Academy of Sciences. 2012; 109(38): 15101-15108. doi: 10.1073/pnas.1213353109
42. Xiao W, Pahlavanneshan M, Eun CY, et al. Matrix stiffness mediates pancreatic cancer chemoresistance through induction of exosome hypersecretion in a cancer associated fibroblasts-tumor organoid biomimetic model. Matrix Biology Plus. 2022; 14: 100111. doi: 10.1016/j.mbplus.2022.100111
43. Zhou BY, Wang LF, Yin HH, et al. Decoding the molecular subtypes of breast cancer seen on multimodal ultrasound images using an assembled convolutional neural network model: A prospective and multicentre study. eBioMedicine. 2021; 74: 103684. doi: 10.1016/j.ebiom.2021.103684
44. Wu Y, Zhang Y, Duan S, et al. Survival prediction in second primary breast cancer patients with machine learning: An analysis of SEER database. Computer Methods and Programs in Biomedicine. 2024; 254: 108310. doi: 10.1016/j.cmpb.2024.108310