Decoding breast cancer Histopathology: Machine Learning-Enhanced Advanced Mathematical Models in Multi-B-Value MR Diffusion Imaging

*Behnam Amini[+1], Moein Ghasemi[+1], Dorreh Farazandeh[1], Hana Farzaneh[2], Sarah Torabi[1], Mona Sedaghat[3], Niloofar Jafarimehrabady[4], Mobasher Hajiabbasi[5], Ashkan Azizi[1], OmidReza Gorjestani[1], Anahita Naviafar[6], Maryam Alaei[7], Mohammad M. Hosseini[8], Nastaran Karimi[9], Amirhossein Parsaei[1], Maryam Vajihinejad[*10], Reza Doshmanziari[2,11], Ali Z. Dehnavi[12], Masih Rikhtehgar[2,11], Alireza Almasi Nokiani[2,11]*

[1]School of Medicine, Tehran University of Medical Science, Tehran, Iran
[2]Department of Radiology, Iran University of Medical Sciences, Tehran, Iran
[3]School of Medicine, University of Social Welfare and Rehabilitation, Tehran, Iran
[4]Department of Clinical-Surgical Diagnostic, University of Pavia, Pavia, Italy
[5]School of Medicine, Islamic Azad University of Tonekabon, Tonekabon, Iran
[6]School of Medicine, Isfahan University of medical sciences, Isfahan, Iran
[7]School of Medicine, Shahid Beheshti University of Medical Science, Tehran, Iran
[8]School of Medicine, Azad Ardabil University of Medical Sciences, Ardabil, Iran.
[9]School of Medicine, Azad Sari University of Medical Sciences, Sari, Iran
[10]Department of Pathology, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.
[11]Firoozabadi Clinical Research Development Unit (FCRDU), Iran University of Medical Sciences, Tehran, Iran
[12]Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science, Rochester, USA

[+]B.A. and M.G. contributed equally to this work
**Corresponding author: Email: Maryam.vajihinejadd@gmail.com*
Associate professor, Department of Pathology, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.
Address: Shahid Sadoughi University of Medical Sciences, Daneshjoo Blvd., Yazd, Iran.
Phone: +98-35-36285406 PO: 8916188635 Fax: +98-35-36238561

Co-authors' emails:

Behnam Amini, Email: Behnamamini717@gmail.com, Orcid ID: 0000-0002-6295-0058
Moein Ghasemi, Email: Moein.sobhan@gmail.com
Dorreh Farazandeh, Email: doriya95@gmail.com
Hana Farzaneh, Email: Farzanehhana@gmail.com
Sarah Torabi, Email: Torabisara396@gmail.com
Mona Sedaghat, Email: dr.monasedaghat@gmail.com
Niloofar Jafarimehrabady, Email: niloofar.jafari.m@gmail.com
Mobasher Hajiabbasi, Email: mobasherhajiabbasi93@gmail.com, ORCID ID: 0000-0003-2050-4065
Ashkan Aziz, Email: Ashkanazizi.md@gmail.com
Omidreza Gorjestani, Email: Omid.gorjy@gmail.com
Anahita Naviafar, Email: anahitanr1375@gmail.com
Maryam Alaei, Email: Alaaei.maryam@gmail.com
Mohammad Mehdi Hosseini, Email: mehdi1374598@gmail.com
Nastaran Karimi, Email: nastarankarimi202020@gmail.com
Amirhossein Parsaei, Email: ah-parsaei@alumnus.tums.ac.ir
Reza Doshmanziari, Email: Doshmanziarireza@gmail.com
Maryam Vajihinejad, Email: Maryam.vajihinejadd@gmail.com, ORCID ID: 0000-0003-2322-5731
Ali Zare Dehnavi, Email: alizared9494@gmail.com
Masih Rikhtehgar, Email: masih.rikhtehgar83.mr@gmail.com
Alireza Almasi Nokiani, Emails: almasinokiani.ar@iums.ac.ir, a_almasi_n@hotmail.com

Appendix 1

S1. Histopathological Analysis

The study employed a comprehensive histopathological analysis, adhering to standardized protocols in line with the College of American Pathologists (CAP) guidelines. In addition to the primary markers of ER, PR, and c-erb-B2 (HER2), the study also assessed Ki67 protein expression. A cutoff value of 14% was employed for Ki67, based on its established clinical relevance as delineated in the St Gallen International Expert Consensus [1]. This cutoff value was further cross-validated with patient outcomes in our dataset. Ki67 positivity was defined as staining in more than 14% of cancer cell nuclei. Molecular markers were assessed based on the percentage of tumor cells exhibiting nuclear staining for ER and PR, and membrane staining for HER2. ER or PR positivity was ascertained if stained nuclei were present in over 10% of cancer cells. HER2 expression was graded on a scale from 0 to 3+, with scores of 0 and 1+ indicating HER2 negativity and a score of 3+ indicating HER2 positivity. For cases with ambiguous HER2 status, defined as an immunohistochemistry score of 2+, Fluorescence In Situ Hybridization (FISH) was utilized for further clarification. All tissue samples were preserved using formalin-fixed, paraffin-embedded (FFPE) methods, considered the gold standard in histopathological analysis. The integrity of these preservation methods was verified through periodic quality checks, which included both visual inspection for tissue morphology and molecular assays for cellular integrity.

S2. MRI Acquisition

Quality Assurance and Calibration

To maintain optimal performance, the MRI system is subjected to rigorous monthly Quality Assurance (QA) assessments. These assessments employ a specialized fat-water phantom to evaluate the efficacy of fat suppression. Real-time monitoring of key performance indicators supplements these monthly checks. Any deviations from the established benchmarks trigger immediate recalibration, ensuring that the system consistently operates within optimal parameters. The initial calibration and standardization of the MRI system were conducted using a standardized phantom to validate the system's suitability for the study.

Contrast Agent and Safety Protocols

During the MRI procedure, gadopentetate dimeglumine was selected based on its demonstrated efficacy and safety in prior breast imaging studies. A comprehensive pre-screening protocol, encompassing renal function tests and allergy history, is rigorously implemented for all patients. Alternative contrast agents are available for those who do not meet the criteria for gadopentetate dimeglumine, thereby ensuring both safety and efficacy.

Image Optimization and Artifacts

The Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) technique is employed to enhance the signal-to-noise ratio and mitigate aliasing artifacts. The GRAPPA acceleration factor has been fine-tuned to 2, a value determined through rigorous phantom and volunteer studies. Post-acquisition, artifact correction algorithms and a multi-step, iterative reconstruction algorithm are applied to further minimize GRAPPA-induced artifacts and geometric distortions.

Acquisition Parameters

The parameters were validated through pilot studies that encompassed a diverse array of breast lesions. The chosen b-values were rigorously optimized to capture both rapid and slow diffusion components, thereby facilitating a nuanced characterization of tissue microstructure. T2-weighted fast spin-echo imaging was acquired with a repetition time (TR) and echo time (TE) of 5100/70 ms, a field of view measuring 350 × 350 mm^2, and a slice thickness of 3.8 mm. The acquisition time for this sequence was 2 minutes and 15 seconds. For diffusion-weighted imaging (DWI), readout-

segmented echo-planar imaging (rsEPI) was employed with a TR/TE of 5600/85 ms, a slice count of 20, and a flip angle of 180 degrees. The readout segments were set to 4, and the slice thickness was adjusted to 4.8 mm. The field of view was 330 × 330 mm^2, and the parallel acquisition technique (PAT) mode used was GRAPPA with a matrix size of 168 × 168. Fat saturation was achieved through spectral adiabatic inversion recovery (SPAIR). The b-values selected for this sequence were 0, 10, 20, 30, 40, 50, 100, 200, 400, 800, 1000, 1500, 2000, and 2500, with an acquisition time of 10 minutes and 35 seconds. The signal-to-noise ratio was maintained at 1.0 for this sequence. Dynamic contrast-enhanced T1-weighted imaging (DCE T1WI) was performed using a fat-suppressed gradient-echo sequence before, and four times continuously after, the intravenous administration of a gadolinium contrast medium (Magnevist, Bayer HealthCare Pharmaceuticals INC.) at a dose of 0.1 mmol/kg of body weight and an injection rate of 2.5 mL/s. This was followed by a 20-mL saline flush at the same injection rate. The scan parameters for DCE T1WI were a TR/TE of 4.6/1.7 ms, a field of view of 335 × 335 mm^2, a matrix of 380 × 380, a flip angle of 10 degrees, and a slice thickness of 1.0 mm. The total number of slices was 106, and the total duration of T1-weighted imaging was 7 minutes and 40 seconds.

S3: Detailed Methodology for Machine Learning Algorithms

Data Preprocessing

Data Cleaning

The dataset, which includes MRI-derived parameters and cancer subtype classifications for 148 BC patients, underwent a meticulous examination to identify inconsistencies or missing values. To create a more robust dataset, we employed multiple imputation techniques using chained equations, replacing the previously used median imputation method.

Scaling

Recognizing that not all machine learning algorithms necessitate feature scaling, we implemented parallel pipelines. One pipeline includes scaled features, standardized to have a mean of 0 and a standard deviation of 1 using the StandardScaler function from the scikit-learn library. The other pipeline retains the original feature magnitudes. This dual approach allows us to determine the most suitable preprocessing strategy for each algorithm.

Model Selection

Algorithms

In this study, a diverse array of machine learning algorithms was selected based on their proven applicability in classification tasks, adaptability to the dataset's unique characteristics, and performance in preliminary testing. Each algorithm was chosen with a comprehensive rationale, focusing on its proven efficacy in medical diagnostic tasks and its suitability for the dataset's characteristics.

- **Linear Support Vector Machine (SVM)**: Employed for its effectiveness in high-dimensional spaces, this algorithm uses a linear kernel. The regularization parameter C is finely tuned to balance margin maximization and classification error minimization.
- **SVM with Radial Basis Function (RBF) Kernels**: This variant captures more complex relationships in the data and employs two primary parameters: C for regularization and gamma for the Gaussian radial basis function.
- **Linear Discriminant Analysis (LDA)**: Selected for its ability to optimally separate different classes, LDA uses various solver parameters such as 'svd', 'lsqr', and 'eigen'.

- **Logistic Regression (LR)**: Known for its simplicity and effectiveness in binary classification, LR employs the C parameter for regularization strength and allows for both l1 and l2 regularization techniques.
- **Random Forests (RF)**: This ensemble method uses multiple decision trees and averages their outputs. Parameters like max_depth, min_samples_split, and min_samples_leaf are finely tuned.
- **Naïve Bayes**: Chosen for its computational efficiency and suitability for high-dimensional datasets, the var_smoothing parameter is optimized for predictive accuracy.
- **Stochastic Gradient Descent (SGD)**: Utilized for its efficiency in large-scale problems, SGD employs hyperparameters like alpha for learning rate and l1_ratio for regularization.
- **Decision Tree**: Known for its interpretability, this algorithm uses parameters like criterion ('gini' or 'entropy'), max_depth, and min_samples_split.
- **AdaBoost**: This ensemble method boosts the performance of weak learners, using parameters like n_estimators and learning_rate.
- **Extreme Gradient Boosting (XGBoost)**: Employed for its high performance and speed, parameters like learning_rate, max_depth, n_estimators, and min_child_weight are optimized.
- **Feed Forward Neural Networks (FFNN)**: This deep learning algorithm models complex nonlinear relationships, with architectural parameters like the number of hidden layers and neurons, as well as activation functions ('relu', 'sigmoid'), being finely tuned.

Hyperparameter Tuning

To optimize the predictive power of each algorithm, we employed Bayesian optimization as an alternative to the previously used grid search method. This change enhances computational efficiency. Furthermore, we extended the use of regularization techniques to other algorithms where applicable, providing a detailed justification for the choice of each regularization parameter. For each algorithm, the hyperparameter space was exhaustively explored, aided by 4-fold cross-validation to ensure robustness. The optimal set of hyperparameters for each model was determined based on the highest average cross-validation accuracy. Below are details for each algorithm:

- **Linear SVM**: The key hyperparameters, including the regularization parameter **C**, were tuned. The range for **C** was set between 0.1 and 10.
- **SVM with RBF Kernels**: Apart from the regularization parameter **C**, the kernel coefficient **gamma** was also optimized. Values of **C** ranged from 0.1 to 10, and **gamma** was set between 0.01 and 1.
- **Linear Discriminant Analysis (LDA)**: The solver parameter was tested with values such as 'svd', 'lsqr', and 'eigen'.
- **Logistic Regression**: The regularization type (l1 or l2) and strength (**C**) were tuned. **C** values ranged from 0.001 to 10.
- **Random Forests**: Multiple parameters like **n_estimators** (number of trees), **max_depth** (maximum depth of the tree), and **min_samples_split** were adjusted.
- **Naïve Bayes**: The var_smoothing parameter was fine-tuned, ranging from 1e-9 to 1e-1.
- **Stochastic Gradient Descent (SGD)**: Learning rate (**alpha**) and regularization (**l1_ratio**) were among the hyperparameters tuned.
- **Decision Tree**: Parameters like **criterion** ('gini' or 'entropy'), **max_depth**, and **min_samples_split** were optimized.
- **AdaBoost**: Number of estimators and learning rate were the primary focus. **n_estimators** ranged from 50 to 200.
- **XGBoost**: Comprehensive tuning was done for **learning_rate**, **max_depth**, **n_estimators**, and **min_child_weight**. Each of these parameters was tested across multiple values to find the optimal settings.
- **Feed Forward Neural Networks (FFNN)**: Several architecture-related parameters, including the number of hidden layers, number of neurons in each layer, and activation functions, were optimized.

For each algorithm, the hyperparameter space was exhaustively searched using grid search, aided by 4-fold cross-validation to ensure robustness. The optimal set of hyperparameters for each model was determined based on the highest average cross-validation accuracy.

Model Evaluation

Cross-Validation

Cross-validation remains a cornerstone for validating the robustness of our machine learning models and mitigating overfitting. In this refined approach, we justify the use of 4-fold cross-validation by comparing its performance with other k-values and leave-one-out cross-validation. This comparison aims to elucidate the trade-offs between computational efficiency and model validation rigor.

Stratification

Stratified k-fold cross-validation is employed, with an additional layer of complexity: the Synthetic Minority Over-sampling Technique (SMOTE) is integrated during the cross-validation process to ensure balanced class representation within each fold, particularly when the minority class is inadequately represented.

Data Splitting

The dataset is partitioned into four equally sized folds. In each cross-validation iteration, three folds serve as the training set, and the remaining fold is used for testing. This procedure is iteratively executed four times, each with a different test fold.

Hyperparameter Validation

Hyperparameter tuning is conducted concurrently with cross-validation. Bayesian optimization replaces the previously used grid search method, enhancing computational efficiency while searching the hyperparameter space. The optimal set of hyperparameters is determined based on the highest average cross-validation accuracy.

Model Robustness

We assess the stability and reliability of each model by analyzing the variability in performance metrics, such as accuracy and AUC-ROC, across the four folds. A low standard deviation in these metrics indicates a more robust model.

Handling of Imbalanced Classes

Identification of Imbalance

Prior to model training, exploratory data analysis is conducted to ascertain any class imbalances in the target variables. This involves calculating the ratio of the minority to the majority class for each target variable.

Resampling Techniques

To address the limitations of SMOTE, Adaptive Synthetic Sampling (ADASYN) is also employed as an alternative. The performance of models trained on these differently balanced datasets is then compared.

Cost-sensitive Learning

A detailed methodology is provided for determining the misclassification costs, including sensitivity analyses to assess the impact of different cost structures. This approach makes it more 'expensive' for the model to misclassify the minority class.

Balanced Metrics

Evaluation metrics are adapted to account for class imbalance. Balanced accuracy, which calculates the arithmetic mean of sensitivity and specificity, is employed to give equal weight to both classes.

Evaluation with Imbalance Handling

Models are rigorously evaluated both before and after the application of these imbalance-handling techniques. Metrics such as the F1 Score, which are more sensitive to class imbalance, are emphasized in the evaluation process.

Algorithmic Approach

Intrinsic parameters for handling class imbalance are utilized in algorithms like Random Forests and XGBoost. For instance, in Random Forests, the **class_weight** parameter is set to 'balanced', and in XGBoost, the **scale_pos_weight** parameter is finely tuned.

Feature Importance

The methodology for assessing feature importance was meticulously designed to provide a robust measure of each variable's contribution to the predictive models. Below are the specific methods and parameters used for each type of algorithm:

Linear SVM:

- Kernel: Linear
- C-value: 1
- **Feature Importance Method Conducted in the Study**: We used the coefficients of the hyperplane, available via the **coef_** attribute, to assess the importance of each feature. These coefficients were normalized and ranked to present a clearer picture of feature significance.

Logistic Regression (LR):

- Regularization: L1
- C-value: 1
- Solver: SAGA
- **Feature Importance Method Conducted in the Study**: We extracted the coefficients via the **coef_** attribute and normalized them to a [0, 1] scale for easier interpretation and ranking.

Random Forests (RF):

- Number of Trees: 100
- Criterion: Gini Impurity
- **Feature Importance Method Conducted in the Study**: Feature importance was directly accessible through the **feature_importances_** attribute, and these were ranked and plotted for interpretation.

Decision Tree:

- Criterion: Gini Impurity
- Max Depth: None (full expansion)
- **Feature Importance Method Conducted in the Study**: The **feature_importances_** attribute was used to rank the features based on how often they appear in the decision nodes.

AdaBoost:

- Base Estimator: DecisionTreeClassifier with max_depth=1
- Number of Estimators: 50
- **Feature Importance Method Conducted in the Study**: Weighted average of feature importances from base estimators was computed and ranked.

XGBoost:

- Learning Rate: 0.1
- Max Depth: 3
- n_estimators: 100
- **Feature Importance Method Conducted in the Study**: The **plot_importance** function was used to directly obtain and visualize the Gini Importance of each feature.

Feed Forward Neural Networks (FFNN):

- Architecture: Input layer (12 neurons), Hidden Layer 1 (50 neurons, ReLU activation), Hidden Layer 2 (20 neurons, ReLU activation), Output Layer (1 neuron, Sigmoid activation).
- **Feature Importance Method Conducted in the Study**: We employed connection weights from the input layer to the first hidden layer to approximate feature importance. These weights were normalized and ranked for interpretation.

SVM with RBF Kernels:

- Kernel: Radial Basis Function (RBF)
- C-value: 1
- Gamma: 0.1

- **Feature Importance Method Conducted in the Study**: Due to the non-linear nature of the RBF kernel, obtaining direct feature importance is not straightforward. Therefore, we employed Recursive Feature Elimination (RFE) over 10 iterations to approximate the significance of each feature.

Linear Discriminant Analysis (LDA):

- Solver: Singular Value Decomposition (SVD)
- Priors: Uniform
- **Feature Importance Method Conducted in the Study**: Linear discriminant coefficients were accessed via the **coef_** attribute of the fitted model, normalized to a [0, 1] scale, and used to determine the features' importance.

Naïve Bayes:

- Type: Gaussian Naïve Bayes
- Priors: None (empirical probabilities from data)
- **Feature Importance Method Conducted in the Study**: Although Naïve Bayes doesn't provide feature importance directly, we approximated it by computing and sorting the variance of each feature in descending order.

Stochastic Gradient Descent (SGD):

- Loss Function: Modified Huber
- Regularization: L2
- Learning Rate: 0.01
- Max Iterations: 1000
- **Feature Importance Method Conducted in the Study**: We accessed the **coef_** attribute after model fitting, sorted the coefficients by their absolute values, and ranked the features based on this sorting to identify their importanc

Table S4.

Group Classification of MRI Variables through Principal Component Analysis and Hierarchical Clustering

MRI Variable	Contribution to Variance (%)	Hierarchical Clustering Analysis			
		Diffusion Cluster category	Euclidean Distance to cancer Cluster 1 *	Euclidean Distance to cancer Cluster 2	Euclidean Distance to cancer Clusters 3
Mean ADC	2.9	1	10.23	10.95	4.59
Median ADC	8.2				
First ADC Quartile	1.9				
Last ADC Quartile	6.2				
ADC Kurtosis	23.3				
ADC Skewness	17.6				
ADC Entropy	39.9	2	83.03	88.20	80.42
Dt	29.7				
fp	56.8				
Dp	13.5				
MD	76.1				
MK	23.9				

The table delineates the outcomes of Principal Component Analysis (PCA) and hierarchical clustering on MRI variables, specifically apparent diffusion coefficient (ADC), intravoxel incoherent motion (IVIM), and diffusion kurtosis imaging (DKI). It enumerates the contribution ratios of dominant variables in each methodology and the Euclidean distances between MRI and target clusters. Cancer clusters are defined as follows: Cluster 1 corresponds to luminal A and HR profile; Cluster 2 to Luminal B and Luminal HER2; Cluster 3 to HER2, HER2 profile, ki67 index, and Triple-Negative (TN). Abbreviations: ADC, apparent diffusion coefficient; IVIM, intravoxel incoherent motion; DKI, diffusion kurtosis imaging.

Table S5.

The Euclidean Distances of diffusion variables based on the unsupervised learning approach (hierarchical clustering).

MRI variables	First ADC Quartile	Mean ADC	Median ADC	Last ADC Quartile	ADC Kurtosis	ADC Skewness	ADC Entropy	Dt	fp	Dp	MD	MK
First ADC Quartile	0.0	8.1	7.6	7.3	6.6	9.2	75.2	8.4	230.5	239.2	10.2	7.9
Mean ADC	8.1	0.0	8.7	8.4	7.1	9.4	75.3	8.8	230.4	239.0	10.5	8.7
Median ADC	7.6	8.7	0.0	8.1	7.2	9.5	74.9	9.0	231.0	239.3	10.4	8.8
Last ADC Quartile	7.3	8.4	8.1	0.0	7.5	9.9	74.0	8.5	229.6	238.4	10.1	8.6
ADC Kurtosis	6.6	7.1	7.2	7.5	0.0	5.7	76.7	8.1	232.1	240.9	10.2	6.8
ADC Skewness	9.2	9.4	9.5	9.9	5.7	0.0	80.2	11.0	235.5	244.7	12.9	8.2
ADC Entropy	75.2	75.3	74.9	74.0	76.7	80.2	0.0	73.0	183.0	184.0	72.7	75.6
Dt	8.4	8.8	9.0	8.5	8.1	11.0	73.0	0.0	228.9	237.8	9.2	9.9
fp	230.5	230.4	231.0	229.6	232.1	235.5	183.0	228.9	0.0	140.1	228.0	230.7
Dp	239.2	239.0	239.3	238.4	240.9	244.7	184.0	237.8	140.1	0.0	236.6	239.8
MD	10.2	10.5	10.4	10.1	10.2	12.9	72.7	9.2	228.0	236.6	0.0	11.0
MK	7.9	8.7	8.8	8.6	6.8	8.2	75.6	9.9	230.7	239.8	11.0	0.0

This table delineates the Euclidean distances among various MRI variables, as determined through hierarchical clustering analysis. The variables include those from apparent diffusion coefficient (ADC), intravoxel incoherent motion (IVIM), and diffusion kurtosis imaging (DKI) methodologies.

Table S6.

The Euclidean Distances of cancer variables based on the unsupervised learning approach (hierarchical clustering).

Tumor varables	Luminal A	Luminal B	Luminal HER2	HER2	TN	HR status	HER2 status	Ki67 status
Luminal A	0.0	9.4	8.8	9.1	7.3	10.6	7.3	11.5
Luminal B	9.4	0.0	5.9	6.4	9.7	8.4	9.7	6.7
Luminal HER2	8.8	5.9	0.0	5.5	10.3	7.7	11.4	8.0
HER2	9.1	6.4	5.5	0.0	10.0	5.5	11.7	7.9
TN	7.3	9.7	10.3	10.0	0.0	11.4	7.7	10.0
HR status	10.6	8.4	7.7	5.5	11.4	0.0	10.3	6.5
HER2 status	7.3	9.7	11.4	11.7	7.7	10.3	0.0	9.5
Ki67 status	11.5	6.7	8.0	7.9	10.0	6.5	9.5	0.0

Table S7.

MANOVA of Diffusion Variables and OLS Regression of Interaction Effects among PCA-Derived Variables across Different Cancer Categories

categories	MANOVA			OLS Regression of Interaction Effects (p-values)						
	Pr(>F)	Pillai	approx F	ADC	IVIM	DKI	ADC:IVIM	ADC:DKI	IVIM:DKI	ADC:IVIM:DKI
HR status	<0.001	0.473	10.103	0.055	<0.001	0.009	0.633	0.600	0.212	0.905
HER2 status	0.008	0.174	2.372	0.123	0.016	0.302	0.017	0.978	0.824	0.110
Ki67 status	<0.001	0.325	5.404	0.934	<0.001	0.834	0.591	0.505	0.978	0.151
Luminal A	<0.001	0.490	10.816	0.678	<0.001	0.116	0.683	0.724	0.515	0.164
Luminal B	0.639	0.067	0.810	0.493	0.389	0.269	0.854	0.128	0.646	0.275
Luminal HER2	0.320	0.093	1.158	0.254	0.464	0.810	0.286	0.467	0.903	0.356
HER2	0.114	0.121	1.549	0.375	0.022	0.298	0.047	0.554	0.870	0.253
TN	<0.001	0.409	7.793	0.003	<0.001	0.043	0.015	0.246	0.114	0.222

This table presents the outcomes of the MANOVA and OLS regression analyses on diffusion variables and interaction effects among PCA-derived variables across various BC categories. The MANOVA section highlights significant multivariate effects in categories like Luminal A, TN, HR, and Ki67, supported by notable Pillai's trace values. The OLS section details the interaction effects between different PCA approaches and BC categories, providing a nuanced understanding of these complex relationships. Ridge and Lasso regression were employed to mitigate multicollinearity. Abbreviations: MANOVA (Multivariate Analysis of Variance), ANOVA (Analysis of Variance), PCA (Principal Component Analysis), ADC (Apparent Diffusion Coefficient), IVIM (Intravoxel Incoherent Motion), DKI (Diffusion Kurtosis Imaging), TN (Triple-Negative), HR (Hormone Receptor), HER2 (Human Epidermal Growth Factor Receptor 2), Ki67 (Proliferation Marker). Additional terms: Pr(>F) represents the p-value, Pillai refers to Pillai's trace which is a measure of effect size, and approx F refers to the approximate F-statistic used in the MANOVA analysis.

Table S8.

Comparative Analysis of Supervised and Deep Learning Approaches in Cancer Feature Classification Leveraging Original MRI Variables

Categories	Linear SVM	RBF-SVM	LDA	LR	RF	Naïve Bayes	SGD	Decision Tree	AdaBoost	XGBoost	FFNN
HR status	0.904	0.785	0.910	0.911	0.903	0.920	0.861	0.790	0.934	0.885	0.889
HER2 status	0.854	0.687	0.763	0.763	0.694	0.813	0.764	0.618	0.599	0.681	0.556
Ki67 status	0.638	0.791	0.747	0.787	0.587	0.760	0.760	0.555	0.611	0.679	0.754
Luminal A	0.825	0.774	0.859	0.841	0.778	0.811	0.788	0.569	0.839	0.708	0.809
Luminal B	0.456	0.647	0.683	0.681	0.496	0.788	0.586	0.545	0.635	0.576	0.576
Luminal HER2	0.629	0.607	0.714	0.507	0.716	0.764	0.717	0.511	0.654	0.779	0.728
HER2	0.559	0.610	0.776	0.767	0.807	0.712	0.769	0.568	0.563	0.550	0.714
TN	0.775	0.815	0.799	0.787	0.830	0.820	0.861	0.748	0.816	0.694	0.876

The table delineates the performance metrics of various machine learning algorithms, including Linear SVM, RBF-SVM, LDA, LR, RF, Naïve Bayes, SGD, Decision Tree, AdaBoost, XGBoost, and FFNN. The highest AUC values are highlighted for each BC category, offering a nuanced understanding of algorithmic efficacy. Abbreviations: SVM (Support Vector Machines), RBF (Radial Basis Function), LDA (Linear Discriminant Analysis), LR (Logistic Regression), RF (Random Forest), SGD (Stochastic Gradient Descent), FFNN (Feed-forward Neural Networks), AUC (Area Under the Curve), MRI (Magnetic Resonance Imaging), HR (Hormone Receptor), HER2 (Human Epidermal Growth Factor Receptor 2), TN (Triple-Negative), Ki67 (Proliferation Marker).

1 Goldhirsch, A. *et al.* Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of oncology* **24**, 2206-2223 (2013).