Supplementary

Machine learning algorithm

Logistic regression models the conditional probability using a logistic function, which depends on the odds, based on the values of the independent variables. Logistic regression tries to minimize the negative log-likelihood of conditional probability via optimization algorithms. It is designed to find cumulative logistic distribution by measuring the relationship between one dependent and one or more independent variables (34).

Boosting algorithms change the training data distribution iteratively using a base classifier to predict the hard-to-classify exemplars. Boosting assigns a weight to each training exemplar, and they can be changed at the end of each round of boosting adaptively (35).

SGD is a method for unconstrained optimization problems that try to learn linear scoring by assigning weights and intercept parameters to minimize the loss function (36).

LDA finds a linear combination of attributes, which separates two or more classes by determining a subspace of lower dimension of the original data. Statistical measures, such as variance and mean, are used to determine separability. LDA maximizes the projected class means by minimizing the class variance in that direction by fitting a Gaussian density to each class with the assumption that all classes share the same covariance matrix (37).

RF are based on the concept of decision trees. A decision tree represents patterns and structures in the input data with hierarchical and sequential nodes that form tree-like structures. A decision tree comprises (I) internal nodes, (II) branches, and (III) terminal nodes. RF constructs many decision trees from a dataset, combines the results from all the trees, and makes a prediction with a majority vote (bootstrap aggregation) to make predictions on classification or regression (19).

Linear SVM is a classifier that finds a hyperplane based on the maximal margin rule to separate the data into two classes. It can be applied to linearly separable data sets and nonlinearly separable data sets using nonlinear kernels. The nonlinear data have to be transformed into a new linear space from its original coordinate space. Thus, in a new coordinate space, the linear decision boundaries could separate the sample data (38). The specific parameter setting of machine learning is shown in Figure S1.

In this study, the optimum ranking of the features based on their importance in the models was reported employing RFE. This method can select features by recursively considering smaller and smaller sets of features by training a classifier on the initial set of features and weights. Then, features with the absolute minimum weights are pruned from the current set features. By repeating this procedure, the desired optimum number of features with the maximum accuracy is identified.

References

34. Falconieri N, Van Calster B, Timmerman D, Wynants L. Developing risk models for multicenter data using standard logistic regression produced suboptimal predictions: a simulation study. Biom J. 2020 Jul;62:932-44.
35. Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. From machine learning to statistical modelling. Methods Inf Med 2014;53:419-27.
36. Ilboudo WEL, Kobayashi T, Sugimoto K. Robust Stochastic Gradient Descent With Student-t Distribution Based First-Order Momentum. IEEE Trans Neural Netw Learn Syst 2020. [Epub ahead of print]. doi: 10.1109/TNNLS.2020.3041755.
37. Dornaika F, Khoder A. Linear embedding by joint Robust Discriminant Analysis and Inter-class Sparsity. Neural Netw 2020;127:141-59.
38. Citak-Er F, Firat Z, Kovanlikaya I, Ture U, Ozturk-Isik E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. Comput Biol Med 2018;99:154-60.


```
LogisticRegression(penalty='l2',C=0.04,class_weight="balanced",solver='saga', dual=False,
                   tol=0.1,fit_intercept=True, intercept_scaling=1,random_state=None,
                   max_iter=1000, multi_class='auto', verbose=0, warm_start=False),
XGBClassifier(base_score=0.5, booster=None, colsample_bylevel=1,
          colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
          importance_type='gain', interaction_constraints=None,
          learning_rate=0.300000012, max_delta_step=0, max_depth=6,
          min_child_weight=1, monotone_constraints=None,
          objective='binary:logistic', missing=None,
          n_estimators=100, n_jobs=0, num_parallel_tree=1,
          random_state=0, reg_alpha=0,
          reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method=None,
          validate_parameters=False, verbosity=None),
SGDClassifier(alpha= 0.0001, average= False, class_weight= None, early_stopping= False,
              epsilon= 0.1, eta0= 0.0, fit_intercept= True, l1_ratio= 0.15, learning_rate= 'optimal',
              loss= 'hinge', max_iter= 1000, n_iter_no_change= 5, n_jobs= None, penalty= 'l2',
              power_t= 0.5, random_state= None, shuffle= True,
              tol= 0.001, validation_fraction= 0.1, verbose= 0, warm_start= False),
LinearDiscriminantAnalysis(covariance_estimator= None, n_components= None, priors= None,
                        shrinkage= None, solver= 'svd', store_covariance= False, tol= 0.0001),
AdaBoostClassifier(base_estimator=None, n_estimators=50, learning_rate=1.0, algorithm='SAMME.R',
                   random_state=None),
RandomForestClassifier(n_estimators=100, criterion='gini', max_depth=10, min_samples_split=2,
                       min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto',
                       max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None,
                       bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0,
                       warm_start=False, class_weight=None, ccp_alpha=0.0, max_samples=None),
DecisionTreeClassifier(criterion='gini', splitter='best', max_depth=5, min_samples_split=2, min_samples_leaf=1,
                       min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None, class_weight=None,
                       presort='deprecated', ccp_alpha=0.0),
SVC(C=1.0, kernel='linear', degree=3, gamma='scale', coef0=0.0, shrinking=True, probability=False,
    tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=-1, decision_function_shape='ovr',
    break_ties=False, random_state=None)
```

Figure S1 Specific parameters of machine learning

 https://dx.doi.org/10.21037/qims-21-39

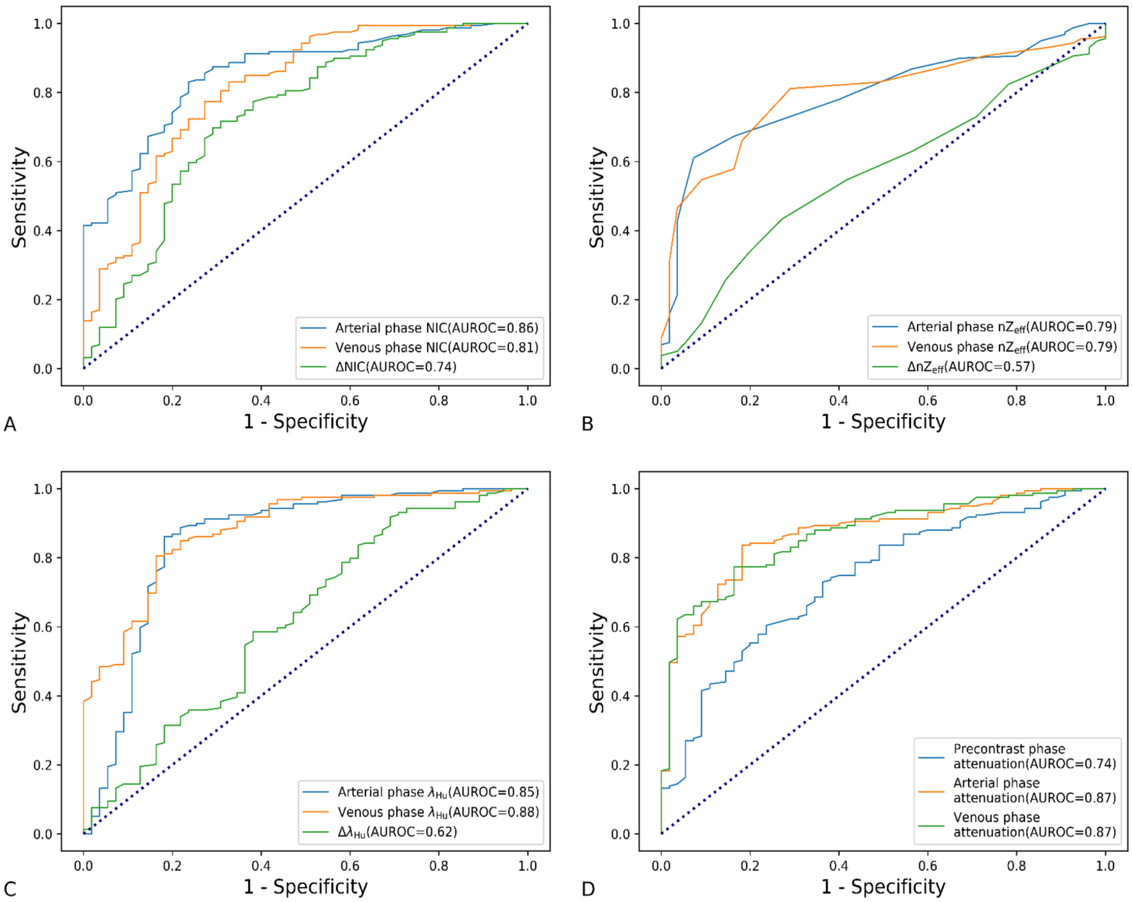
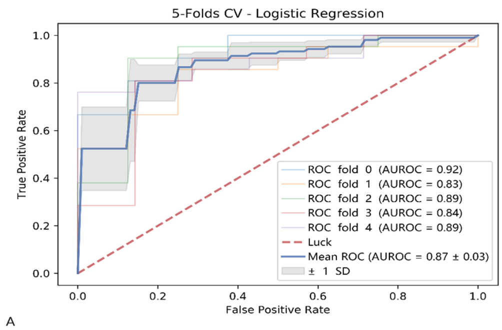
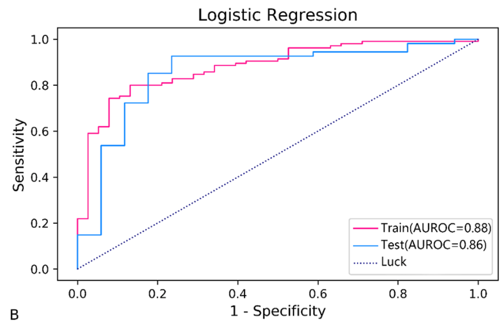
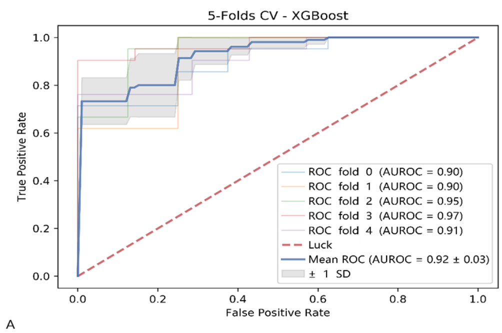
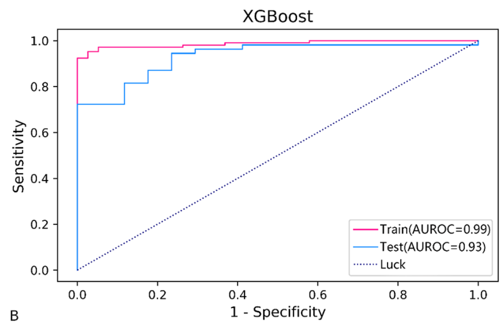
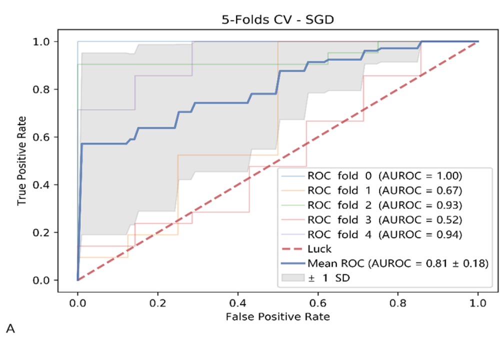
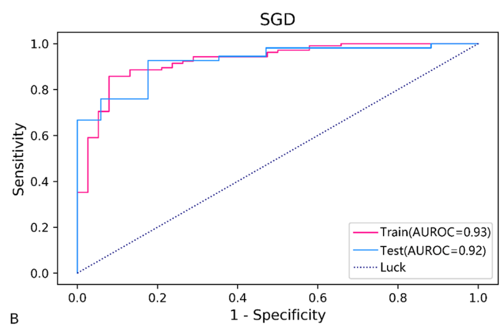
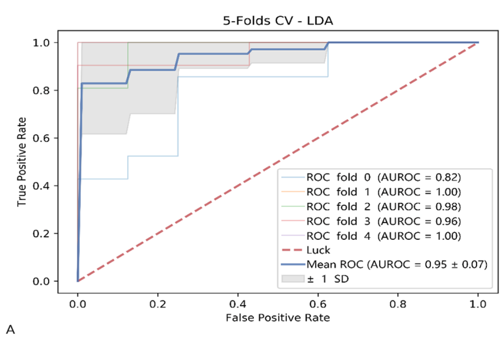
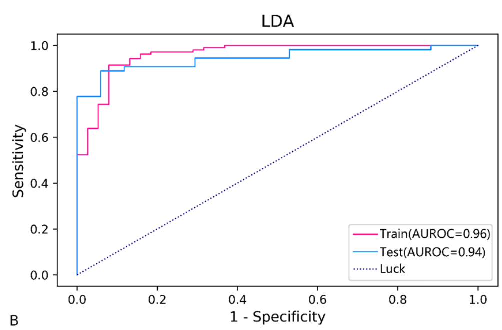


Figure S2 ROC curves of univariate logistic regression model using mpDECT to predict malignant and benign breast lesions. (A) The NIC; (B) the nZ_{eff}; (C) the λ_{Hu}; and (D) attenuation. ROC, receiver operating characteristic; mpDECT, multiparametric dual-energy computed tomography; NIC, normalized iodine concentration; nZeff, normalized effective atomic number; λ_{Hu}, slope of the spectral Hounsfield unit curve; AUROC, area under the receiver operating characteristic curve.

https://dx.doi.org/10.21037/qims-21-39

<table>
<tr><th></th><th>Five-fold CV</th><th>Training and testing datasets</th></tr>
<tr><td>Logistic regression</td><td>
</td><td>
</td></tr>
<tr><td>XGBoost</td><td>
</td><td>
</td></tr>
<tr><td>SGD</td><td>
</td><td>
</td></tr>
<tr><td>LDA</td><td>
</td><td>
</td></tr>
</table>

https://dx.doi.org/10.21037/qims-21-39

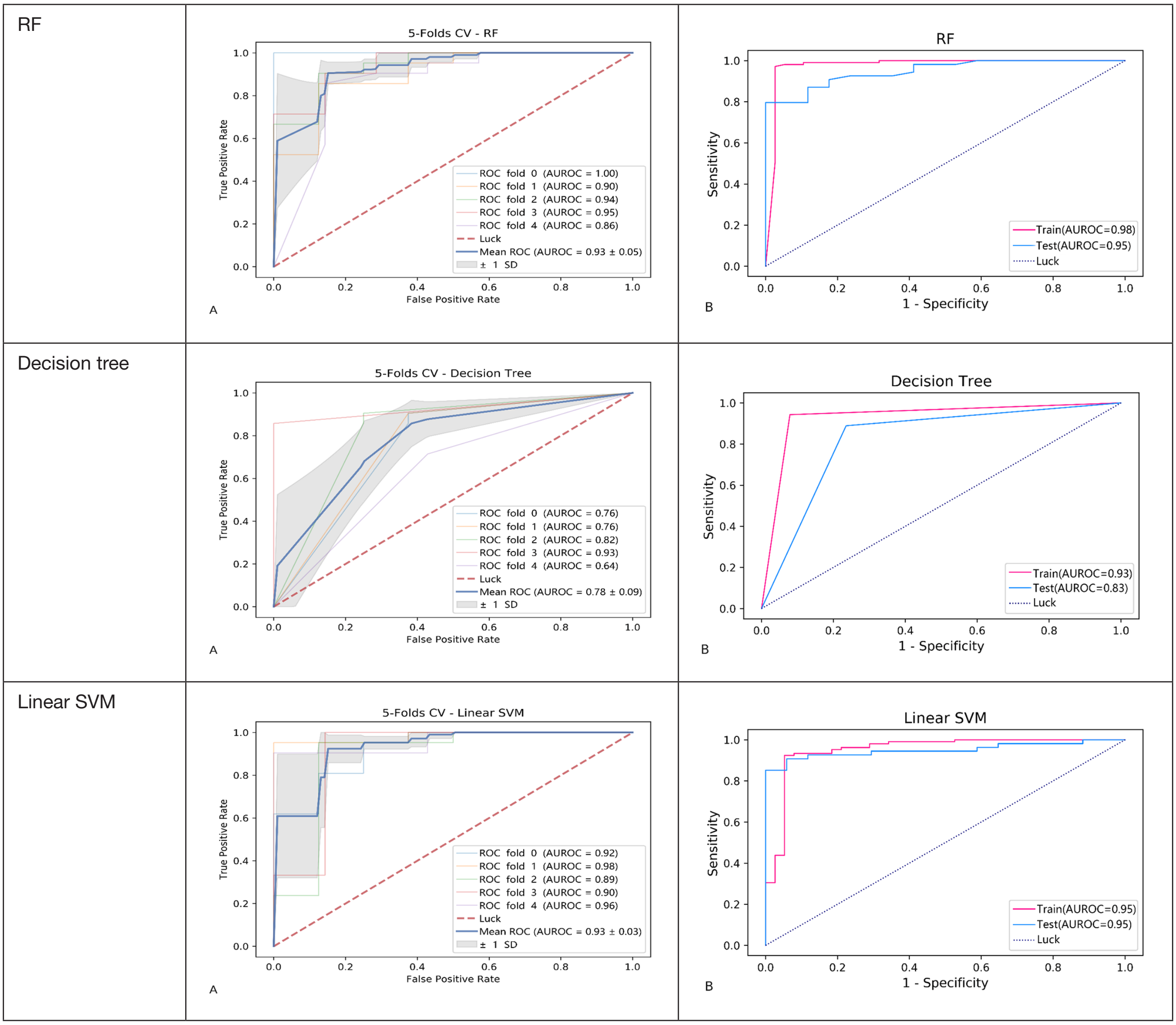


Figure S3 ROC curves of mpDECT model using logistic regression, XGBoost, SGD, LDA, RF, decision tree, and linear SVM classifier using five-fold CV in the prediction of malignant and benign breast lesions. (A) Five-fold CV; (B) training and testing dataset. ROC, receiver operating characteristic; mpDECT, multiparametric dual-energy computed tomography; XGBoost, extreme gradient boosting; SGD, stochastic gradient descent; LDA, linear discriminant analysis; AdaBoost, adaptive boosting; RF, random forest; SVM, support vector machine; CV, cross-validation; AUROC, area under the receiver operating characteristic curve.

 https://dx.doi.org/10.21037/qims-21-39