Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement

Illan, Ignacio Alvarez, Tahmassebi, Amirhessam, Ramirez, Javier, Gorriz, Juan , Foo, Simon, et al.

Ignacio Alvarez Illan, Amirhessam Tahmassebi, Javier Ramirez, Juan M. Gorriz, Simon Y. Foo, Katja Pinker-Domenig, Anke Meyer-Baese, "Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement," Proc. SPIE 10669, Computational Imaging III, 106690W (14 May 2018); doi: 10.1117/12.2304588

SPIE.

Event: SPIE Commercial + Scientific Sensing and Imaging, 2018, Orlando, Florida, United States

Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement

Ignacio Alvarez Illan[a,b,*], Amirhessam Tahmassebi[a], Javier Ramirez[b], Juan M Gorriz[b], Simon Y. Foo[c], Katja Pinker-Domenig[d,e], and Anke Meyer-Baese[a]

[a]Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA
[b]Department of Signal Theory, Networking and Communications, University of Granada, Spain
[c]Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering, Tallahassee, Florida 32310-6046, USA
[d]Department of Radiology, Breast Imaging Service, Memorial Sloan Kettering Cancer Center, New York, USA
[e]Department of Biomedical Imaging and Image-Guided Therapy, Division of Molecular and Gender Imaging, Medical University of Vienna, Vienna, Austria

ABSTRACT

Accurate methods for breast cancer diagnosis are of capital importance for selection and guidance of treatment and optimal patient outcomes. In dynamic contrast enhancing magnetic resonance imaging (DCE-MRI), the accurate differentiation of benign and malignant breast tumors that present as non-mass enhancing (NME) lesions is challenging, often resulting in unnecessary biopsies. Here we propose a new approach for the accurate diagnosis of such lesions with high resolution DCE-MRI by taking advantage of seven robust classification methods to discriminate between malignant and benign NME lesions using their dynamic curves at the voxel level, and test it in a manually delineated dataset. The tested approaches achieve a diagnostic accuracy up to 94% accuracy, sensitivity of 99 % and specificity of 90% respectively, with superiority of high temporal compared to high spatial resolution sequences.

Keywords: Breast Cancer, Computer Aided Diagnosis (CAD), Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI), Machine learning

1. INTRODUCTION

Since the introduction of dynamic-contrast-enhanced magnetic resonance imaging (DCE-MRI) for the diagnosis of breast cancer, a set of computer aided diagnosis (CAD) systems have been developed with satisfactory performance results. CAD systems are usually designed for the problem of lesion detection and segmentation, such as modifications on the fuzzy c-means algorithm,[1,2] combining kinetic and morphological features,[3–6] like shape, margins, and internal enhancement distribution,[7] textural kinetic,[8] or more recently using deep neural networks,[9,10] among others. The performance of the CAD is then measured by comparing the automatic detection and segmentation solutions with manually delineated regions of interest, performed by experts. Although the performance of existing CAD systems is promising, the high false positive rate is still a problem.[11–13]

While a big effort has been made to overcome the limitation of a high false positive rate with a significant number of important contributions, the problem of distinguishing between malign and benign lesion has received less attention. However, accurate means for the objective and reliable differentiation of malignant and benign are warrant to reduce unnecessary breast biopsies that cause patient anxiety and costs. Machine learning approaches can provide with a solution to this problem, by 'learning' to differentiate between different patterns. The machine learning solution is both accurate and objective, and has been shown to outperform human experts in medical imaging, or at least have comparable performance.[14]

* Corresponding Author: Ignacio Alvarez Illan
E-mail: illan at ugr.es

Computational Imaging III, edited by Abhijit Mahalanobis, Amit Ashok, Lei Tian, Jonathan C. Petruccelli, Proc. of SPIE Vol. 10669, 106690W · © 2018 SPIE
CCC code: 0277-786X/18/$18 · doi: 10.1117/12.2304588

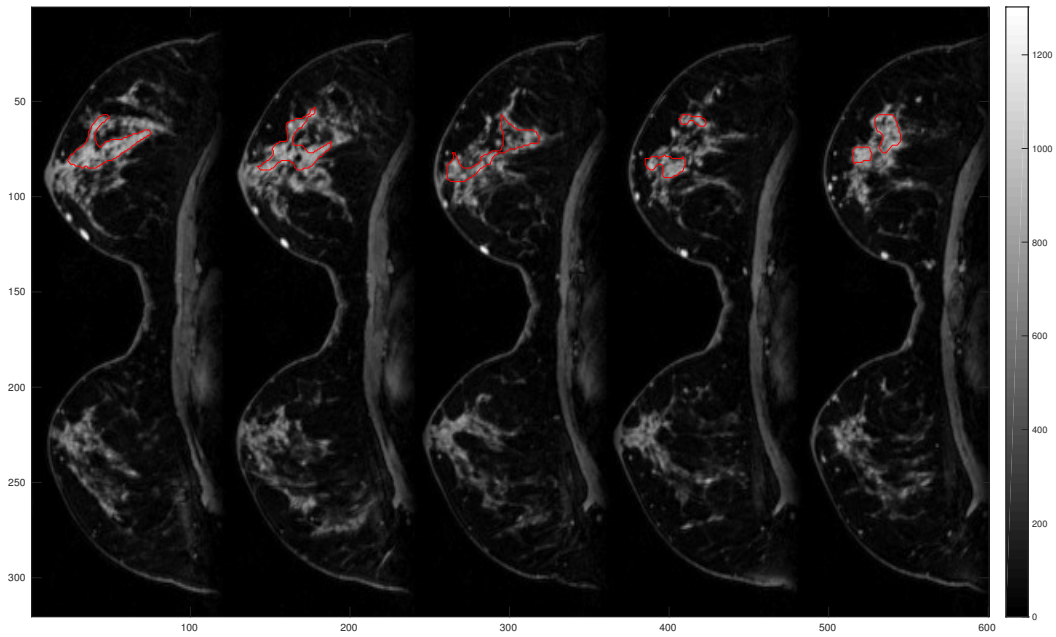

Figure 1. Representative axial slices of a malignant NME lesion delineated in red.

Table 1. Acquisition parameters

	TR/TE (millisec)	FOVr (mm)	SI (mm)	Axial slices	flip angle	matrix
3T FLASH sequence	877/3.82	320	1 isotropic	96	9°	320 / 134
1.5T VIBE sequence	3.61/1.4	320	1.7 isotropic	72	6°	192 / 192

In the case of non-mass enhancing (NME) lesions, the accurate differentiation of benign from malignant lesions is especially challengings.[15] In this work, we study the suitability of machine learning techniques for benign-malignant differentiation in NME lesions, using the kinetic curve dynamics of the tissues at the voxel level. We analyze the importance of having a high spatial resolution versus a high temporal resolution, using 7 well-known machine learning paradigms.

2. METHODS

2.1 Dataset

We collected a set of data from 13 patients with NME. The data represented a patient subset from a larger cohort undergoing MRI following patient inclusion criteria described in detail in.[16] The imaging protocol comprised of both high spatial and temporal resolution MRI. Low-temporal high-spatial resolution (LTHSR) imaging was acquired with a 3T MRI scanner (Tim Trio, Siemens, Erlangen, Germany) with a dedicated, bilateral, 4-channel breast coil in vivo (Orlando, FL). Three high spatial resolution images were taken, pre-contrast, peak, and post-contrast as coronal T1-weighted (3D) FLASH sequence, with water excitation and fat suppression, with 2 minutes average acquisition time 2 minutes. Low-spatial high-temporal resolution (LSHTR), contrast-enhanced, coronal T1-weighted (VIBE) sequence was obtained with the following sequence parameters with 13.2 seconds average acquisition time per volume leading to 3.45 minutes for 17 measurements. The parameters used in the high and low resolution series are summarized in table 1. A second set of high spatial resolution T1-weighted imaging (repeated 3D-FLASH) was acquired after these 17 low spatial VIBE resolution images, as the peak enhancement of the lesion could be expected at the end of this time span([16] and references therein). Finally,

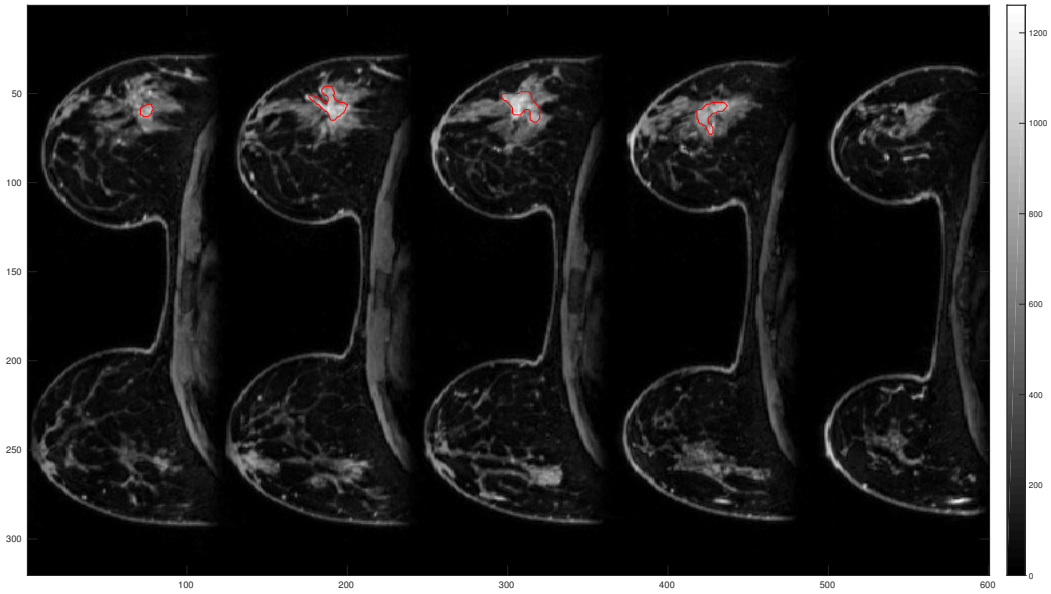

Figure 2. Representative axial slices of a benign tumor delineated in red.

high temporal resolution (repeated VIBE with 25 measurements, leading to an acquisition time of 5 minutes 35 seconds and repeated 3D-FLASH for dynamic assessment of lesion wash-out) was performed, and then high spatial resolution T1-weighted images were recorded. The contrast agent used was Gd-DOTA (generic name: Gadoterate meglumine; Dotarem, Guerbet, France), injected intravenously as a bolus (0.1 mmol per kilogram body weight) and administered with a power injector (Spectris Solaris EP, Medrad, Pittsburgh, PA) at 4 mL/s followed by a 20 mL saline flush. The contrast agent was injected 75 seconds after starting the first coronal T1-weighted VIBE.

The NME were visually assessed by expert radiologists following the American College of Radiology BI-RADS atlas,[17] and delineated using the Osirix software on the 3T high spatial resolution volumes. All NME were classified as BI-RADS 4: suspicious or BI-RADS 5: highly suspicious of malignancy. The lesions found were histopathologically confirmed, from which 5 where invasive ductal carcinoma (IDC), 5 ductal carcinoma in-situ (DCIS) and 3 invasive lobular carcinoma (ILC).

2.2 Preprocessing

All dynamic sequences were registered to the pre-contrast volume. This pre-processing step was applied to remove all possible artifacts produced by patient movements. The algorithm employed to perform this task was the SPM12[18] registration algorithm, which performs affine and non-affine transformations on the data by minimizing the mutual information metric.

The manual delineations of the lesions were performed on 3T high-spatial resolution images using the Osirix software, recorded as a set of axial point coordinates. The Bresenham algorithm[19] was used to transform the coordinate points into 3D binary masks, and a decimation was employed to down-sample the masks to the size of the low spatial high-temporal resolution images. Thus, the down-sampled masks were used to define the class labels of each voxel, 1 if the voxel was in the mask, and 0 otherwise.

The pre-contrast sequences were averaged and subtracted from the post-contrast ones in order to normalize

intensities in an interpatient level. Concretely, the final intensity was obtained as:

$$I(x,y,z,t) = I(x,y,z,t) - \frac{1}{N}\sum_{t=1}^{t=N} I(x,y,z,t) \tag{1}$$

where N is the number of pre-contrast images, 7 in the LSHTR case, and 1 in the LTHSR case.

2.3 Evaluation

A selection of images was used to evaluate the CAD: 5 DCE-MRI images containing 8 delineated lesions, five DCIS, one IDC, and 2 benign tumors. The data were considered at the voxel level, being for the LSHTR images $\approx 2 \cdot 10^3$ samples, and $\approx 3 \cdot 10^5$ samples for the LTHSR ones. The dataset was split following a stratified 10-fold cross validation scheme, and relevant parameters were obtained for each iteration: accuracy, sensitivity, specificity, and area under ROC curve. The confidence intervals were obtained from the set of 10 fold iterations as obtaining the standard deviation of the desired value.

We employed seven different machine learning algorithms as an explanatory tool to detect tumors in breast and distinguish between two classes of tumors, benign and malignant. The classification task was done by learning seven different target functions such as, Support Vector Machine (SVM) with linear and kernel, Nearest Neighbors (NN), Quadratic Discriminant Analysis (QDA), Decision Trees (DT), Random Forests (RF), AdaBoost, and Artificial Neural Network (ANN) which map each feature set x to one of the class labels y.

- **Support Vector Machine (SVM)** The essential idea behind a support vector machine is to find a hyperplane that separates the classes with the largest margin and with the minimal number of errors, in the case the data classes are not separable. In our setting, we make use of a logistic function to estimate the probabilities of the SVM classifier, and vary the a posterior value to obtain a receiver operation characteristic (ROC) curve.
- **Nearest Neighbors (NN)** According to a defined metric, usually euclidean, the class of the test example in the nearest neighbor classifier is assigned to be the class of the closest n-samples computed by the defined metric. We use n=3.
- **Decision Tree (DT)** Decision trees are non-parametric models for supervised learning in which the features are learned by simple decision rules inferred from the data. The input data is split into different branches so that it matches the target values. We use here the gini impurity measure to guarantee the quality of the splitting.
- **Random Forests (RF)** Random forest as an ensemble method employs multiple decision tree classifiers to predict the target values, and average their results after resampling. The used parameters are: number of trees is 10 and the subsamples are drawn with replacement using bootstrap.
- **AdaBoost** Although Adaboost can be used with any weak learner, we use here decision trees as weak learners. Doing so, Adaboost is closely related to random forest, but instead of simple resampling, it applies a weight to each input instance depending on how hard to classify they are such that subsequent classifiers focus more on difficult cases. We use 50 estimators before the boosting is stopped.
- **Artificial Neural Network (ANN)** An artificial neural network find a series of outputs from some inputs by adjusting the weights of a network that may have one or several layers. ANN might use different types of activation functions such as logistic, hyperbolic tangent, sigmoid, and linear. This makes possible to solve non-linear problems. ANN assigns the learning weights in the network that minimize the sum of squared errors between inputs and expectations, and actualize them iteratively by backpropagating the errors. The parameters used here are: 2-layers of size 100 and 50, $\alpha = 0.01$, with rectifier linear unit (relu) as activation function.

Table 2. Performance Parameters for the LSHTR.

	Accuracy	Sensitivity	Specificity	AUC
Linear SVM	**0.94 ± 0.09**	0.98 ± 0.03	0.90 ± 0.18	0.94 ±0.11
Nearest Neighbors	0.92 ± 0.08	0.94 ± 0.04	0.90 ± 0.20	0.94 ± 0.10
Decision Tree	0.91 ± 0.07	0.94 ± 0.03	0.82 ± 0.17	0.90 ± 0.08
Random Forest	0.85 ± 0.03	0.85 ± 0.02	0.84 ± 0.21	0.89 ± 0.07
ANN	0.91 ± 0.12	0.93 ± 0.05	0.91 ± 0.23	0.92 ± 0.15
Adaboost	0.93 ± 0.09	0.95 ± 0.03	**0.90 ± 0.18**	0.95 ± 0.08
QDA	0.93 ± 0.14	**0.99 ± 0.02**	0.89 ± 0.22	**0.96 ± 0.08**

- **Quadratic Discriminant Analysis (QDA)** Likewise linear discriminant analysis, QDA can be derived from Bayes rule. However, in QDA model, there is no assumption on covariance matrices, and the class is assigned to the one with maximum a posteriori probability. If there is a high correlation on the data, QDA will fail to find the target classes, as it depends on the computation of the covariance and its inverse. Also, Naive Bayes can be seen as an special case of QDA if the features are independent.

3. RESULTS

Figure 3 shows the cross-validated ROC results for the defined classifiers. The results are obtained for the LSHTR images, so the dimension of the feature space is 42. The mean value and its standard deviation is showed for the area under curve (AUC), obtained form the 10 iterations. The performance results for this experiment are summarized in table 2. Although some of the reported values can never be greater than 1, the confidence intervals sometimes cover values higher than 1. We interpret the confidence intervals as a measure of robustness of the tested classifier, corresponding small confidence intervals with more robust classifiers, and not strictly to valid range of values. In that sense, the best performance is achieved by the QDA classifier, although the performance results of most of the tested classifiers are in general optimal. Therefore, a supervised or semi-manual segmentation of lesions could benefit from a CAD system trained on benign-malign discrimination to decrease the number of biopsies.

Figure 4 shows the cross-validated ROC results for LTHSR images. In this other case the dimension of the feature space is 3. Followed by equation 1, the number of independent features is reduced to 2 after intensity normalization, and so is the feature space dimension. The main observation from the ROC analysis, compared to the LSHTR one, is that the classification performance decreases drastically with a small feature space dimension, and a number of samples two orders of magnitude bigger. Although Decision Tree (DT) and Random Forest (RF) classifiers perform slightly worse than than the rest in the LSHTR case, they are the classifiers reaching the higher AUC values in LTHSR case. Comparing also the robustness of these classifiers from table 2, DT and RF have the narrower confidence intervals in most of the parameters evaluated. Therefore, we can attribute the fact that DT and RF have better performance in the LTHSR case to its robustness, which is important when the number of samples is increased by two orders of magnitude.

Other effect measured in the LTHSR case is that the QDA is ill defined, since the variables appear to be highly correlated when increasing the number of samples (the variability reduces). Since discriminant analysis performs a matrix inversion, it gives inaccurate results when it is close to 0, which corresponds to collinear variables. Therefore, the results of QDA are not reported for the LTHSR case in figure 4.

4. CONCLUSIONS

CAD systems for differentiation of benign and malignant breast tumors presenting as NME lesions in DCE-MRI provide tools that aid the radiologist in accurate diagnosis. Here, we employed seven different machine learning approaches to the problem of differentiation between benign and malign NME lesions, by using LSHTR DCE-MRI images, and LTHSR DCE-MRI images. In general, machine learning techniques are successful in differentiating benign from a malignant tumors with a superiority of high temporal compared to a high spatial resolution sequences. Taking into account that the BI-RADS criteria for the lesions studied in this work recommended biopsy for all the lesions, and the malignancy was established histopathologically, the results presented here

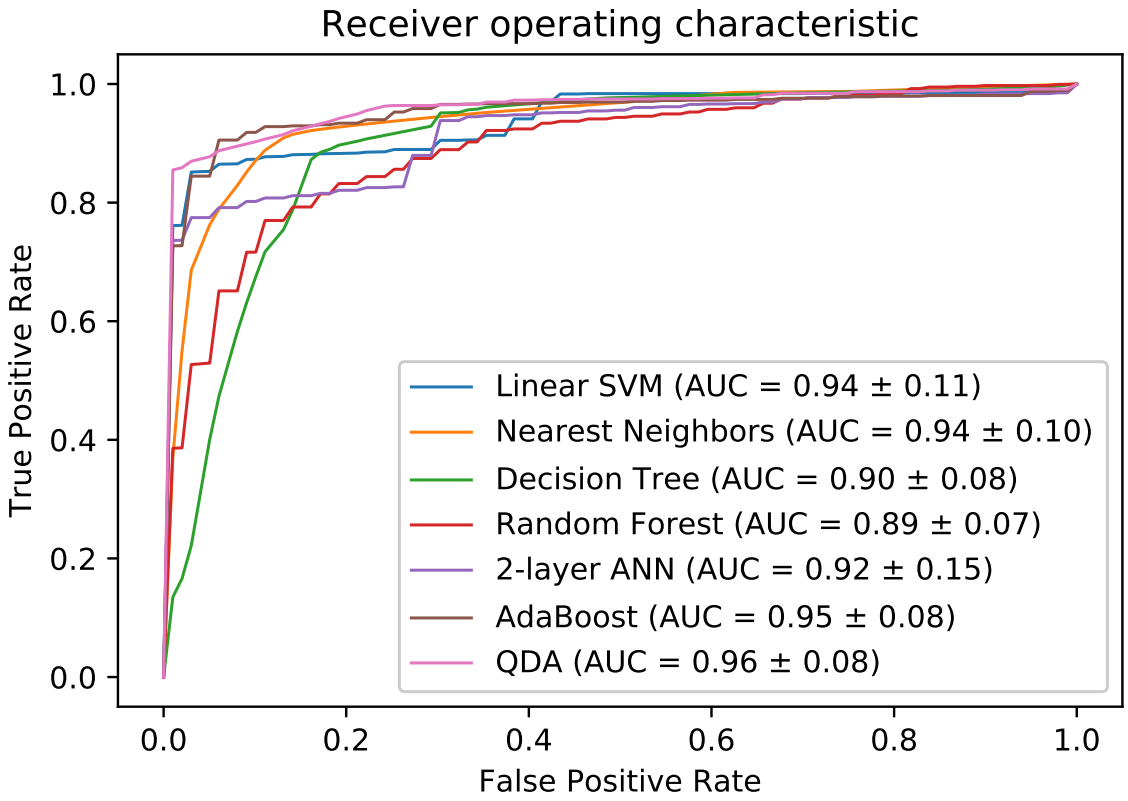

Figure 3. Receiver operator curve for detection of benign-malignant tissues for LSHTR.

indicate that using machine learning could facilitate the accurate differentiation of malignant and benign breast tumors, and reduce the number of unnecessary biopsies in breast cancer.

5. ACKNOWLEDGEMENTS

This work is partially supported by ONR Grant "Gulf of Mexico Spring School (GMSS) Deep Learning Workshop". This work has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 656886.

REFERENCES

[1] Chen, W., Giger, M. L., and Bick, U., "A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images," *Academic Radiology* **13**, 63–72 (Jan. 2006).

[2] Chang, Y.-C., Huang, Y.-H., Huang, C.-S., Chang, P.-K., Chen, J.-H., and Chang, R.-F., "Classification of breast mass lesions using model-based analysis of the characteristic kinetic curve derived from fuzzy c-means clustering," *Magnetic Resonance Imaging* **30**, 312–322 (Apr. 2012).

[3] Hoffmann, S., Shutler, J. D., Lobbes, M., Burgeth, B., and Meyer-Bse, A., "Automated analysis of non-mass-enhancing lesions in breast MRI based on morphological, kinetic, and spatio-temporal moments and joint segmentation-motion compensation technique," *EURASIP Journal on Advances in Signal Processing* **2013**(1), 172 (2013).

[4] Gubern-Merida, A., Marti, R., Melendez, J., Hauth, J. L., Mann, R. M., Karssemeijer, N., and Platel, B., "Automated localization of breast cancer in DCE-MRI," *Medical Image Analysis* **20**(1), 265–274 (2015).

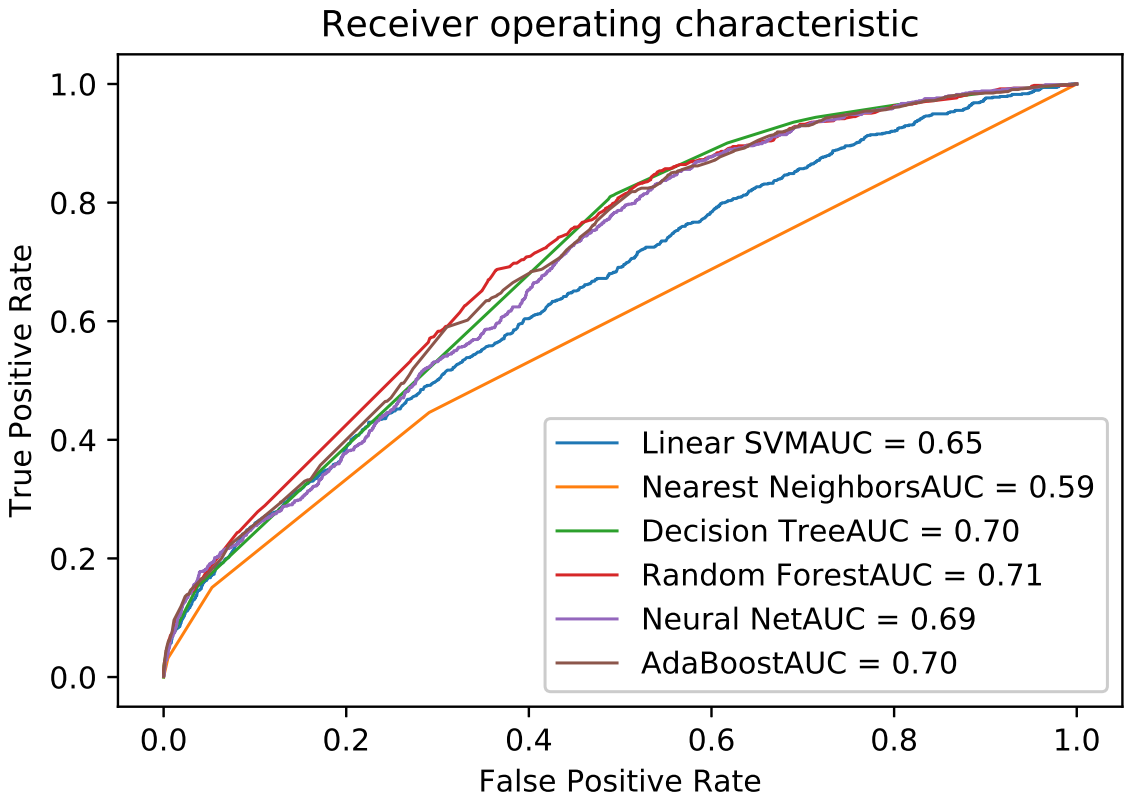

Figure 4. Receiver operator curve for detection of benign-malignant tissues for LTHSR.

[5] Chang, Y.-C., Huang, Y.-H., Huang, C.-S., Chen, J.-H., and Chang, R.-F., "Computerized breast lesions detection using kinetic and morphologic analysis for dynamic contrast-enhanced MRI," *Magnetic Resonance Imaging* **32**(5), 514–522 (2014).

[6] Wang, T.-C., Huang, Y.-H., Huang, C.-S., Chen, J.-H., Huang, G.-Y., Chang, Y.-C., and Chang, R.-F., "Computer-aided diagnosis of breast DCE-MRI using pharmacokinetic model and 3-D morphology analysis," *Magnetic Resonance Imaging* **32**, 197–205 (Apr. 2014).

[7] Agliozzo, S., De Luca, M., Bracco, C., Vignati, A., Giannini, V., Martincich, L., Carbonaro, L. A., Bert, A., Sardanelli, F., and Regge, D., "Computer-aided diagnosis for dynamic contrast-enhanced breast MRI of mass-like lesions using a multiparametric model combining a selection of morphological, kinetic, and spatiotemporal features," *Medical Physics* **39**(4), 1704–1715 (2012).

[8] Agner, S. C., Soman, S., Libfeld, E., McDonald, M., Thomas, K., Englander, S., Rosen, M. A., Chin, D., Nosher, J., and Madabhushi, A., "Textural kinetics: a novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification," *Journal of Digital Imaging* **24**(3), 446–463 (2011).

[9] Rasti, R., Teshnehlab, M., and Phung, S. L., "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognition* **72**, 381–390 (Dec. 2017).

[10] Antropova, N., Huynh, B., and Giger, M., "SU-D-207b-06: Predicting Breast Cancer Malignancy On DCE-MRI Data Using Pre-Trained Convolutional Neural Networks," *Medical Physics* **43**, 3349–3350 (June 2016).

[11] Hu, L., Cheng, Z., Wang, M., and Song, Z., "Image manifold revealing for breast lesion segmentation in DCE-MRI," *Bio-Medical Materials and Engineering* **26**, S1353–S1360 (Jan. 2015).

[12] McClymont, D., Mehnert, A., Trakic, A., Kennedy, D., and Crozier, S., "Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph," *Journal of Magnetic Resonance Imaging* **39**, 795–804 (Apr. 2014).

[13] Jayender, J., Chikarmane, S., Jolesz, F. A., and Gombos, E., "Automatic Segmentation of Invasive Breast Carcinomas from DCE-MRI using Time Series Analysis," *Journal of magnetic resonance imaging : JMRI* **40**, 467–475 (Aug. 2014).

[14] Illan, I. A., Gorriz, J. M., Ramirez, J., Salas-Gonzalez, D., Lopez, M. M., Segovia, F., Chaves, R., Gomez-Rio, M., and Puntonet, C. G., "18f-FDG PET imaging analysis for computer aided alzheimers diagnosis," *Information Sciences* **181**(4), 903–916 (2011).

[15] Sakamoto, N., Tozaki, M., Higa, K., Tsunoda, Y., Ogawa, T., Abe, S., Ozaki, S., Sakamoto, M., Tsuruhara, T., Kawano, N., Suzuki, T., Yamashiro, N., and Fukuma, E., "Categorization of non-mass-like breast lesions detected by MRI," *Breast Cancer* **15**, 241–246 (July 2008).

[16] Pinker, K., Grabner, G., Bogner, W., Gruber, S., Szomolanyi, P., Trattnig, S., Heinz-Peer, G., Weber, M., Fitzal, F., Pluschnig, U., Rudas, M., and Helbich, T., "A combined high temporal and high spatial resolution 3 tesla MR imaging protocol for the assessment of breast lesions: Initial results," *Investigative Radiology* **44**(9), 553–558 (2009).

[17] Reston, V. A. C. o. R., [*Breast Imaging Reporting and Data System Atlas BI-RADS-MRI*], American College of Radiology (2003).

[18] Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W., [*Statistical Parametric Mapping: The Analysis of Functional Brain Images*], Academic Press (2007).

[19] Bresenham, J. E., "Algorithm for computer control of a digital plotter," *IBM Systems Journal* **4**(1), 25–30 (1965).