Title

Convolutional Neural Network Using a Breast MRI Tumor Dataset Can Predict Oncotype Dx Recurrence Score.

Journal

Journal of Magnetic Resonance Imaging, 49(2)

Authors

Ha, Richard

Chang, Peter

Mutasa, Simukayi

et al.

Publication Date

2019-02-01

DOI

10.1002/jmri.26244

Peer reviewed

Published in final edited form as:

J Magn Reson Imaging. 2019 February ; 49(2): 518–524. doi:10.1002/jmri.26244.

Convolutional Neural Network Using a Breast MRI Tumor Dataset Can Predict Oncotype Dx Recurrence Score

Richard Ha, MD, MS[1,*], **Peter Chang, MD**[2], **Simukayi Mutasa, MD**[2], **Jenika Karcich, MD**[2], **Sarah Goodman, MD**[2], **Elyse Blum, MD**[2], **Kevin Kalinsky, MD, MS**[3], **Michael Z. Liu, MS**[4], **Sachin Jambawalikar, PhD**[4]

[1]Breast Imaging Section, Department of Radiology, Columbia University Medical Center, New York, New York, USA;

[2]Department of Radiology, Columbia University Medical Center, New York, New York, USA;

[3]Division of Hematology/Oncology in the Department of Medicine at Columbia University Medical Center, New York, New York, USA;

[4]Department of Medical Physics, Columbia University Medical Center, New York, New York, USA

Abstract

Background: Oncotype Dx is a validated genetic analysis that provides a recurrence score (RS) to quantitatively predict outcomes in patients who meet the criteria of estrogen receptor positive / human epidermal growth factor receptor-2 negative (ER+/HER2−)/node negative invasive breast carcinoma. Although effective, the test is invasive and expensive, which has motivated this investigation to determine the potential role of radiomics.

Hypothesis: We hypothesized that convolutional neural network (CNN) can be used to predict Oncotype Dx RS using an MRI dataset.

Study Type: Institutional Review Board (IRB)-approved retrospective study from January 2010 to June 2016.

Population: In all, 134 patients with ER+/HER2− invasive ductal carcinoma who underwent both breast MRI and Oncotype Dx RS evaluation. Patients were classified into three groups: low risk (group 1, RS <18), intermediate risk (group 2, RS 18–30), and high risk (group 3, RS >30).

Field Strength/Sequence: 1.5T and 3.0T. Breast MRI, T_1 postcontrast.

Assessment: Each breast tumor underwent 3D segmentation. In all, 1649 volumetric slices in 134 tumors (mean 12.3 slices/tumor) were evaluated. A CNN consisted of four convolutional layers and max-pooling layers. Dropout at 50% was applied to the second to last fully connected layer to prevent overfitting. Three-class prediction (group 1 vs. group 2 vs. group 3) and two-class prediction (group 1 vs. group 2/3) models were performed.

[*]Address reprint requests to: R.H., Breast Imaging Section, Department of Radiology, 622 West 168th St., PB-1-301, Columbia University Medical Center, New York, NY 10032. rh2616@columbia.edu.

Conflict of Interest

The authors report no conflicts of interest.

Statistical Tests: A 5-fold crossvalidation test was performed using 80% training and 20% testing. Diagnostic accuracy, sensitivity, specificity, and receiver operating characteristic (ROC) area under the curve (AUC) were evaluated.

Results: The CNN achieved an overall accuracy of 81% (95% confidence interval [CI] ± 4%) in three-class prediction with specificity 90% (95% CI ± 5%), sensitivity 60% (95% CI ± 6%), and the area under the ROC curve was 0.92 (SD, 0.01). The CNN achieved an overall accuracy of 84% (95% CI ± 5%) in two-class prediction with specificity 81% (95% CI ± 4%), sensitivity 87% (95% CI ± 5%), and the area under the ROC curve was 0.92 (SD, 0.01).

Data Conclusion: It is feasible for current deep CNN architecture to be trained to predict Oncotype DX RS.

Breast cancer is the most ubiquitous malignancy afflicting women worldwide and is the second most common cause of cancer deaths among women in the United States.[1] Not all breast cancers are the same, with a wide spectrum of intrinsic biologic diversity seen across multiple subtypes indicating variable biologic behavior and treatment options.[2]

If a patient meets the criteria of estrogen receptor positive (ER+), human epidermal growth factor receptor-2 negative (HER2−), and node-negative, adjuvant chemotherapy may not be indicated, as the risk of recurrence is comparable to the harm from toxicity.[3,4] These patients may receive surgery, endocrine therapy, or radiation.[4]

Oncotype Dx (Genomic Health, Redwood City, CA) is a validated 21-gene reverse transcriptase polymerase chain reaction (RT-PCR) assay involved in tumor cell proliferation and hormonal response, which provides a recurrence score (RS) to quantitatively predict outcomes in patients who meet the criteria of ER+/HER2−/node negative invasive breast carcinoma.[5–7] In 2016 the updated guidelines of the American Society of Clinical Oncology (ASCO) recommended use of this RS in ER+/HER2−/node negative breast cancer to help determine the utility of adjuvant systemic chemotherapy.[3]

Although effective, genetic analysis such as Oncotype Dx is invasive and expensive, which has motivated the investigation of imaging analysis to determine tumor heterogeneity. Magnetic resonance imaging (MRI) is a common modality used in the diagnosis of breast cancer given its high soft-tissue contrast and sensitivity.[8] In recent years there have been investigations into quantitative analysis of specific extracted imaging features termed “radiomics.” Further correlation of these quantitative imaging features to molecular gene expression defines “radiogenomics.”[9]

The field of radiomics and radiogenomics has developed largely due to the contribution of machine-learning techniques utilizing the extraction of pertinent imaging features and correlating with clinical data. More recently, due to advances in computer hardware technology, a subset of machine learning utilizing a type of artificial neural network called convolutional neural networks (CNNs) has begun to proliferate for medical imaging analysis. In contrast to traditional algorithms that utilize hand-crafted features based on human extracted patterns, neural networks allow the computer to automatically construct predictive statistical models, tailored to solve a specific problem subset.[10] The laborious task of human engineers inputting specific patterns to be recognized could be replaced by

inputting curated data and allowing the technology to self-optimize and discriminate through increasingly complex layers.

The purpose of this study was to develop a novel CNN algorithm to predict Oncotype DX RS using a breast MRI tumor dataset.

Materials and Methods

Patient Population

A Health Insurance Portability and Accountability Act (HIPAA)-compliant, Institutional Review Board-approved retrospective review from January 1, 2010 to June 30, 2016 identified 134 patients with ER+/HER2− invasive ductal carcinoma who underwent Oncotype Dx RS evaluation and had preoperative bilateral breast MRI performed prior to definitive breast surgery. The need for informed consent was waived. These patients also had no known hormonal therapy at the time of diagnosis. The mean age of patients was 55.9 years (standard deviation [SD], 11 years). Menopausal status was 53% premenopausal (71/134) and 47% postmenopausal (63/134). The average tumor diameter was 1.44 cm (SD, 0.63 cm).

MRI Acquisition and Analysis

MRI was performed on a 1.5T or 3.0T commercially available system (Signa Excite, GE Healthcare, Milwaukee, WI) using an eight-channel breast array coil. The imaging sequences included a triplane localizing sequence followed by a sagittal fat-suppressed T_2-weighted sequence (repetition time / echo time [TR/TE], 4000–7000/85; section thickness, 3 mm; matrix, 256 × 192; field of view [FOV], 18–22 cm; no gap). A bilateral sagittal T_1-weighted fat-suppressed fast spoiled gradient-echo sequence (17/2.4; flip angle, 35°; bandwidth, 31–25 Hz) was then performed before and three times after a rapid bolus injection (gadobenate dimeglumine/Multihance; Bracco Imaging, Princeton, NJ; 0.1 mmol/kg) delivered through an IV catheter. Image acquisition started after contrast material injection and was obtained consecutively with each acquisition time of 120 seconds. Section thickness was 2–3 mm using a matrix of 256 × 192 and an FOV of 18–22 cm. Frequency was in the anteroposterior direction. After the examination, postprocessing was performed including subtraction of the unenhanced images from the first contrast-enhanced images on a pixel-by-pixel basis and reformation of sagittal images to axial images.

Oncotype Dx RS

Each tumor specimen was sent to Genomic Health as standard of care and the Oncotype Dx RS was determined ranging from 0–100. Patients were classified into three groups based on the risk of recurrence 10 years after treatment: low risk (group 1, RS <18), intermediate risk (group 2, RS 18–30), and high risk (group 3, RS >30).

Computer-Based Image Analysis

IMAGE PREPROCESSING.—For all patients, breast tumor regions were manually annotated by a board-certified radiologist using a region-of-interest (ROI) drawn in 3DSlicer[11] based on first postcontrast dynamic contrast-enhanced (DCE)-MRI images. For

134 tumors, 1649 volumetric slices (mean 12.3 slices per tumor) in 32 × 32 voxel resolution were evaluated from the segmented tumor data. The intensity values at each pixel of the image were normalized by subtracting the mean intensity value of the image and dividing by the SD for each image. Representative preprocessed single slice image of DCE-MRI breast tumors are shown in Fig. 1. Images in each row correspond to low (A), moderate (B), and high (C) Oncotype Dx groups.

NEURAL NETWORK ARCHITECTURE.—CNN is structured as a sequential set of convolution filters applied to the original image followed by activations functions. The filters apply learnable function that is trained with each new batch of input images. The filter weights are updated by minimizing the cost function, which compares the predicted output with ground truth training labels (here, Oncotype Dx group). The L2 regularization, which adds “squared magnitude” of coefficient as a penalty term to the loss function, was used to discourage parameters of this learnable filter from becoming too large and to prevent overfitting of the model to the training data. In our network, we used L2-norm on the fully connected layer. L2-norm loss function is also known as least squares error (LSE). It minimizes the sum of the square of the differences (S) between the target value (Yi) and the estimated values (f(xi):

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

The activation function following convolutional filtering introduces nonlinearities that create a hierarchy of layers. This layered hierarchy is fundamentally what allows depth in a network. Hierarchical depth in the network is what allows filters to represent more complex features. The optimization of the network mostly involves proper scaling of the input data and the learning rate step size. A proper preprocessing normalization of the data is essential to a network’s convergence.

The overall network architecture is shown in Fig. 2. The CNN is implemented completely by series of 3 × 3 convolutional kernels to prevent overfitting, as described by Simonyan et al.[12] Max-pooling with kernel of 2 × 2 is used as shown. All nonlinear functions are modeled by the rectified linear unit (ReLU).[13] In deeper layers the number of feature channels was increased from 32 to 64, reflecting increasing representational complexity. Dropout at 50% was applied to the second to last fully connected layer to prevent overfitting by limiting coadaptation of parameters.[14] Training was done over 200 epochs using the Adam optimizer with a base and a learning rate of 0.001. For better generalization and to prevent overfitting of the model, a L2-regularization penalty of 0.01 was used.

For each breast tumor, a final softmax score threshold of 0.5 was used for classification. The softmax score, also known as softmax function, is a normalized exponential function. It is a generalization of the logistic function that “squashes” a K-dimensional vector of arbitrary real values to a K-dimensional vector of real values, where each entry is in the range (0, 1), and all the entries add up to 1. The softmax score provides the probability for each class label. The probability of each class will sum to 1 as dictated by the normalization constraint.

Two sets of experiments were run, one three-class model to train the CNN model to predict low, moderate, or high Oncotype Dx RS and the second to predict two-class low vs. (moderate+high) Oncotype Dx RS. Five-fold crossvalidation was performed with 80% of the data used as training and 20% used for testing purposes. In the three-class model, three different sensitivity and specificity metrics are given, one for each class. The performance metrics are calculated from the test dataset reserved for performance characterization upon which the training model was never exposed to. Training was implemented using the Adam optimizer, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments.[15,16] Parameters are initialized using the heuristic described by He et al.[17] To account for training dynamics, the learning rate is annealed whenever training loss plateaus.

Software code for this study was written in Python using Keras software (https://github.com/fchollet/keras) with TensorFlow[18] backend. Experiments and CNN training were done on a Linux workstation. On a single consumer NVIDIA Geforce GTX 1080 ti card with 11 GB video RAM with 3584 CUDA cores running at the stock boost clock of 1582 MHz, training the dataset with 5-fold crossvalidation takes 90 minutes over 1000 epochs. Two sets of experiments were run, one to train the CNN model to predict three-class low, moderate, or high Oncotype-Dx RS and the second to predict two-class low vs. (moderate+high) Oncotype-Dx RS.

Statistical Analysis

Statistical analysis was performed using the IBM SPSS software (Armonk, NY, v. 24). Age was calculated at the time of diagnosis. Descriptive statistics were used to summarize clinical, imaging, and pathologic parameters. Classification performance was evaluated using a multiclass receiver operating characteristics (ROC) analysis. This involved generating ROC plots for each group versus the other two combined. For each of these two-class classifications the sensitivity and specificity is reported.

Results

The tumor grade was 17.9% low grade (24/134), 65.7% intermediate grade (88/134), and 16.4% high grade (22–134). Axillary lymph node status was 92.5% negative (124/134) and 7.5% positive (10/134). Based on the American Joint Committee on Cancer, TNM classifications were as follows: T1 (73.8%, 99/134), T2 (25.4%, 34/134), T3 (0.7%, 1/134), T4 (0%); N0 (92.5%, 124/134), N1 (7.5%, 10/134), N2 (0%), N3 (0%); M0 (100%, 134/134), M1 (0%). Most (97%, 130/134) of the patients had unifocal disease. Four patients had multifocal disease. Three our of four patients had one additional tumor. One out of four patients had two additional tumors. No contralateral tumors were present. Only the primary tumor that underwent Oncotype Dx evaluation was matched for MRI image analysis. The additional tumors did not undergo Oncotype Dx evaluation. Breast MRI was performed on 1.5T in 61.2% (82/134) of the patients and on 3.0T in 38.8% (52/134) of the patients. 11.2% (15/134) of the tumors demonstrated nonmass enhancement. 88.8% (119/134) of the tumors demonstrated mass enhancement.

The median Oncotype Dx score was 16 (range, 1–75). Patients were classified into three groups based on the risk of recurrence 10 years after treatment: low risk (group 1, RS <18), intermediate risk (group 2, RS of 18–30), and high risk (group 3, RS >30). The low-risk group consisted of 77 patients. The intermediate-risk group consisted of 40 patients. The high-risk group consisted of 17 patients.

A total of 134 breast cancer cases with Oncotype Dx recurrence scores were included in this study. For each breast tumor, a final softmax score threshold of 0.5 was used for classification. The CNN was trained for a total of 200 epochs (batch size of 32) before convergence. Based on this, mean 5-fold validation accuracy was calculated. Initially, a three-class prediction model was utilized, classifying results into a low-risk group, intermediate-risk group, and high-risk group. The CNN achieved an overall accuracy of 81% (95% confidence interval [CI] ± 4%). Subsequently, a two-class Oncotype Dx prediction model was evaluated in two groups consisting of 77 and 57 patients (group 1 vs. groups 2 and 3). The CNN achieved an overall accuracy of 84% (95% CI ± 5%) in two-class prediction.

The ROC plot is shown in Figs. 3 and 4. For the two-class prediction model, the area under the ROC curve was 0.92 (SD, 0.01) with specificity 81% (95% CI ± 4%) and sensitivity 87% (95% CI ± 5%). For the three-class prediction model, the area under the ROC curve was 0.92 (SD, 0.01) with specificity 90% (95% CI ± 5%) and sensitivity 60% (95% CI ± 6%).

Discussion

The CNN algorithm used in our study achieved an overall accuracy of 84% in predicting patents with low Oncotype Dx RS compared to patients with intermediate/high Oncotype Dx RS. Our results indicate the feasibility of utilizing the CNN algorithm to predict Oncotype Dx RS. Potential improvement of our model with a larger dataset may result in a useful predictive tool for determining patients' likelihood of breast cancer distant recurrence.

It has been well established that the 21-gene Oncotype Dx RS is an accurate prognostic and predictive determinant of breast cancer recurrence and treatment response in ER+/HER2−/node negative breast cancer.[5,6] The most updated guidelines of the ASCO strongly recommend the use of the Oncotype Dx RS to decide on adjuvant systemic chemotherapy in ER+/HER−/node negative breast cancer, based on high-quality evidence.[3]

While Oncotype Dx RS has been validated for clinical use, there is a need for the development of additional methods. First, the Oncotype Dx test is expensive and invasive, requiring utilization of valuable tumor tissue. Second, it is limited in availability. According to the National Cancer Database (NCDB), ~20% of ER + breast cancer patients acquire an RS in the United States, with skewed prevalence among Caucasians, higher levels of education, higher incomes, and Middle Atlantic residents.[19]

The prevalence of preoperative breast MRI continues to rise as its role in the diagnosis and treatment of breast cancer evolves.[20] Additionally, utilizing these already acquired MR images to determine an Oncotype Dx RS has the potential to more efficiently utilize resources to personalize treatment decisions. In an effort to improve the clinical application

of these MRI images, various quantitative computer algorithms have been developed and used to further delineate gross tumor features into additional computer-generated characteristics. Over the past decade there has been a surge of investigation into quantitative analysis of specific extracted imaging features, known as radiomics. Recently, radiogenomics, defined as correlation of these quantitative imaging features to genetic information, has shown further potential to utilize images in guiding clinical management.

Machine-learning technology continues to evolve, as it plays a powerful role in modern-day society. Traditional algorithms are based on human designed pattern-recognition utilizing computer-generated imaging features such as morphologic histogram and Gray-Level Co-Occurrence Matrix characteristics. In 2014, Ashraf et al[21] were one of the first groups to apply a traditional machine-learning algorithm to investigate intrinsic imaging phenotypes in breast tumors and their association with prognostic gene expression profiles. In that study, which included 56 patients, a multiparametric imaging phenotype vector was extracted for each tumor by using quantitative morphologic, kinetic, and spatial heterogeneity features yielding a moderate correlation ($r = 0.71$, R (2) = 0.50, $P < 0.001$) between DCE MRI features and the recurrence score. MRI features were predictive of recurrence risk as determined by the surrogate assay, with area under the receiver operating characteristic curve of 0.82.

Subsequently, in 2015, Sutton et al[22] investigated the association between Oncotype Dx RS and morphological and texture-based image features extracted from MRI. The study included 95 patients. Two MRI-derived image features, kurtosis in the first and third postcontrast images and histologic nuclear grade, were found to be significantly correlated with the Oncotype Dx RS. The overall model resulted in a statistically significant correlation with Oncotype Dx RS with an R-squared value of 0.23 (adjusted R-squared = 0.20; $P = 0.0002$) and a Spearman's rank correlation coefficient of 0.49 ($P < 0.0001$). Overall, this study demonstrated that MRI image-based features could be used to predict the likelihood of recurrence and magnitude of chemotherapy benefit.

Recently, in 2016, Li et al[23] investigated relationships between computer-extracted breast MRI phenotypes and multigene assays of MammaPrint, Oncotype DX, and PAM50 to assess the role of radiomics in evaluating the risk of breast cancer recurrence in 84 patients. For each case, computer-extracted tumor phenotypes of size, shape, margin morphology, enhancement texture, and kinetic assessment were evaluated, yielding significant associations between radiomics signatures and multigene assay recurrence scores (R2 = 0.25–0.32, $r = 0.5–0.56$, $P < 0.0001$). Important radiomics features included tumor size and enhancement texture, which reflected tumor heterogeneity. Use of radiomics in the task of distinguishing between good and poor prognosis yielded area under the receiver operating characteristic curve values ranging from 0.55–0.88, depending on the multigene assay. The authors concluded that quantitative breast MRI radiomics shows promise for image-based phenotyping in assessing the risk of breast cancer recurrence.

While the studies discussed above show promising results, they are dependent on feature engineering methods. Feature engineering involves the process of putting domain knowledge into the creation of feature extractors to simplify the complexity of the data in order to make

patterns more visible and enable algorithms to work. These methods have limitations, including the dependence on accurate extraction of important features.

In contrast, neural network algorithms are trained to automatically extract the features of an input important to the defined problem domain. This process involves creation of an algorithm that iteratively improves its ability to evaluate features of an input in an end-to-end manner, utilizing stacked and increasingly complex layers, in order to improve its ability to predict a desired output. In addition, CNNs consist of structured layers of learnable filters. The architecture itself is a feature extractor; applications in future unrelated fields may borrow from previous architecture. Training data and subsequent hyperparameter optimization will need to be specific to the application, but network architecture does not usually have to be changed significantly. Filters and weights from superficial layers are generalizable, as they represent nonspecific features such as edges, corners, blobs, etc. Filters and weights from deeper layers will necessarily be retrained on application-specific data.

We applied a convolutional neural network to this problem domain, and the largest study (n = 136) applying radiogenomics to predict Oncotype Dx RS using 3D volumetric MRI data. Using a CNN, our algorithm predicted Oncotype Dx RS with high accuracy using an MRI tumor dataset. Our study demonstrates significant potential for further advancement in the field of radiogenomics with the utilization of CNNs.

Our study has a few limitations. It was a small, retrospective study in a single institution. The performance of CNN has been shown to increase logarithmically with larger datasets.[25] Larger MRI datasets are likely to significantly improve an Oncotype Dx RS prediction model. In addition, patients in this study underwent MRI imaging at different magnetic field strengths (1.5 or 3.0T), potentially affecting the image quality. However, selection bias is likely negligible, given that the choice of patients undergoing MRI on a 1.5 or 3.0T magnet were randomly determined purely based on availability of the scanner. Other limitations include potentially long training times when utilizing convolutional neural networks. This is related to many factors, including the size of the dataset and number of network parameters. Traditional algorithms comparatively takes much less time to train. This effect, however, is reversed during testing time, where CNNs take much less time to run. Lastly, because training a CNN is an end-to-end process, it does not clearly reveal the reasoning behind the final result in a deterministic manner. Many methods have been developed to improve human understanding and intuition behind the predictions of a neural network; however, this is an ongoing area of research.

In conclusion, in a relatively small sample size we were able to predict Oncotype Dx RS based on an MRI tumor dataset with an accuracy of up to 84%. Future research with a prospective randomized study is needed to validate the potential of predicting Oncotype Dx RS, as well as directly correlating MRI with clinical outcome.

Acknowledgment

This work was presented in an oral presentation format at the 2017 RSNA meeting.

References

1. Siegel RL. Cancer statistics, 2017. CA Cancer J Clin 2017;67:7–30. [PubMed: 28055103]
2. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumors. Nature 2000;406:747–752. [PubMed: 10963602]
3. Harris LN, Ismaila N, McShane LM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol 2016;34:1134–1150. [PubMed: 26858339]
4. Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26(Suppl 5):v8–v30. [PubMed: 26314782]
5. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node negative breast cancer. N Engl J Med 2004; 351:2817–2826. [PubMed: 15591335]
6. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J Clin Oncol 2006;24:3726–3734. [PubMed: 16720680]
7. Gluz O, Nitz UA, Christgen M, et al. West German Study Group Phase III PlanB Trial: First prospective outcome data for the 21-gene recurrence score assay and concordance of prognostic markers by central and local pathology assessment. J Clin Oncol 2016;34:2341–2349. [PubMed: 26926676]
8. Song JL, Chen C, Yuan JP, et al. Progress in the clinical detection of heterogeneity in breast cancer. Cancer Med 2016;5:3475–3488. [PubMed: 27774765]
9. Fan M, Li H, Wang S, et al. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. PloS one 2017;12: e0171683. [PubMed: 28166261]
10. LeChun Y, Bengio T, Hinton G. Deep learning. Nature 2015;521: 436–444. [PubMed: 26017442]
11. Pieper S, Halle M, Kikinis R. 3D Slicer. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No 04EX821). IEEE 2004;632–635.
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations 2015; p 1–14.
13. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) 2010; pp. 807–814.
14. Srivastava N, Hinton GE, Krizhevsky A, et al. Dropout?: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15: 1929–1958.
15. Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014; arXiv preprint arXiv:1412.6980.
16. Mandic DP. A generalized normalized gradient descent algorithm. IEEE Signal Process Lett 2004;11:115–118.
17. He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. arXiv:1502.01852 (2015).
18. Abadi M, Barham P, Chen J, et al. TensorFlow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI ‘16) [Internet] 2016; p 265–84.
19. Orucevic A, Heidel RE, Bell JL. Utilization and impact of 21-gene recurrence score assay for breast cancer in clinical practice across the Unites States: Lessons learned from the 2010 to 2012 National Cancer Data Base analysis. Breast Cancer Res Treat 2016;157:427–435. [PubMed: 27206678]
20. Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. Lancet 2011;378:1804–1811. [PubMed: 22098853]
21. Ashraf AB, Daye D, Gavenonis S, et al. Identification of intrinsic imaging phenotypes of breast cancer tumors: Preliminary associations with gene expression profiles. Radiology 2014;272:374–384 [PubMed: 24702725]
22. Sutton EJ, Oh JH, Dashevsky BZ, et al. Breast cancer subtype intertumor heterogeneity: MRI-based features predict results of a genomic assay. J Magn Reson Imaging 2015;42:1398–1406. [PubMed: 25850931]

23. Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. Radiology 2016;281:382–391. [PubMed: 27144536]

24. Sun C, Shrivastava1 A, Singh S, et al. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. arXiv preprint arXiv:1707.02968 (2017).

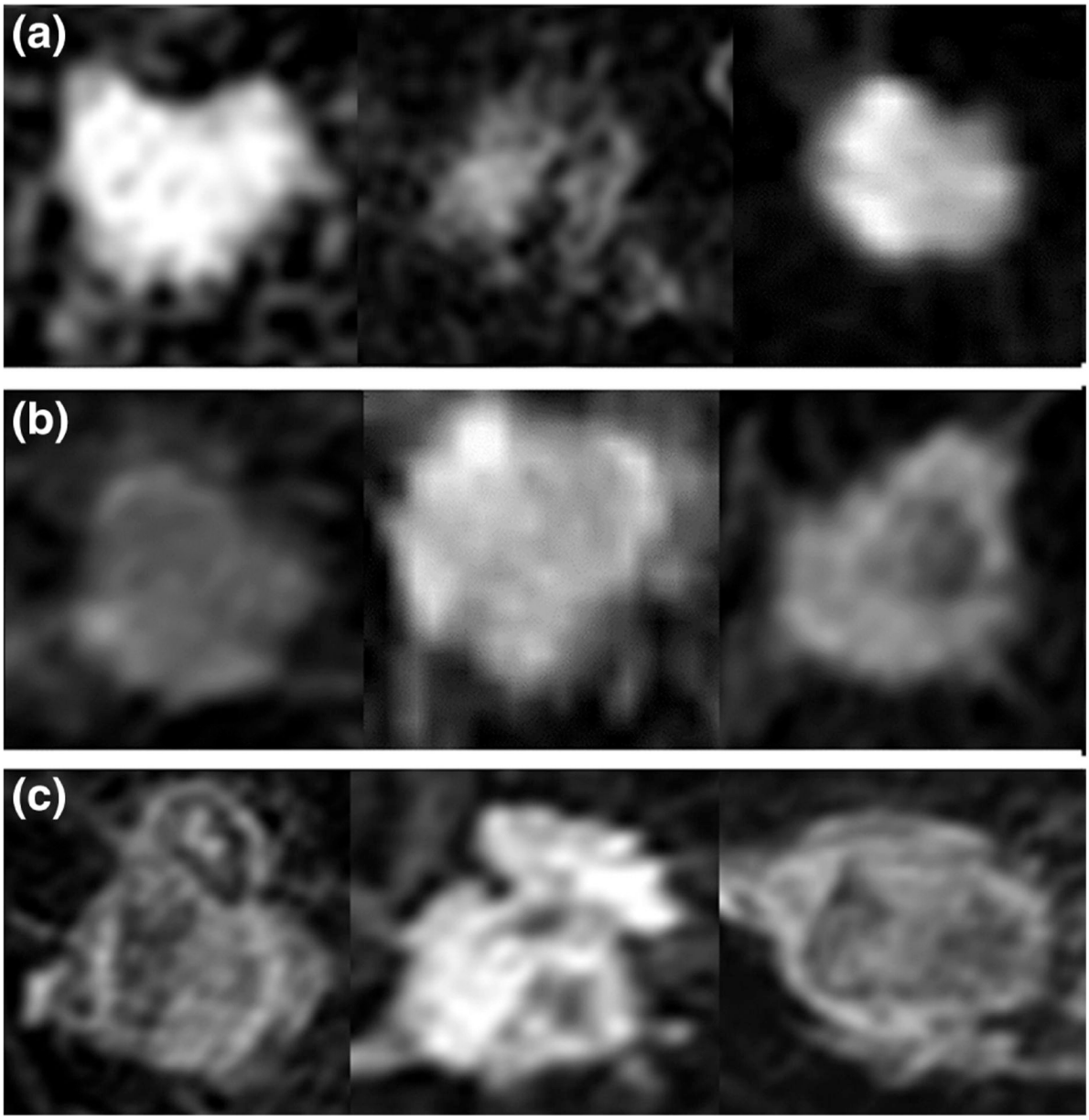

FIGURE 1:
Preprocessed DCE-MRI breast tumors. Example cases of DCE tumor images corresponding to low (top row, a), intermediate middle row, (b) and high (bottom row, c) Oncotype Dx RS.

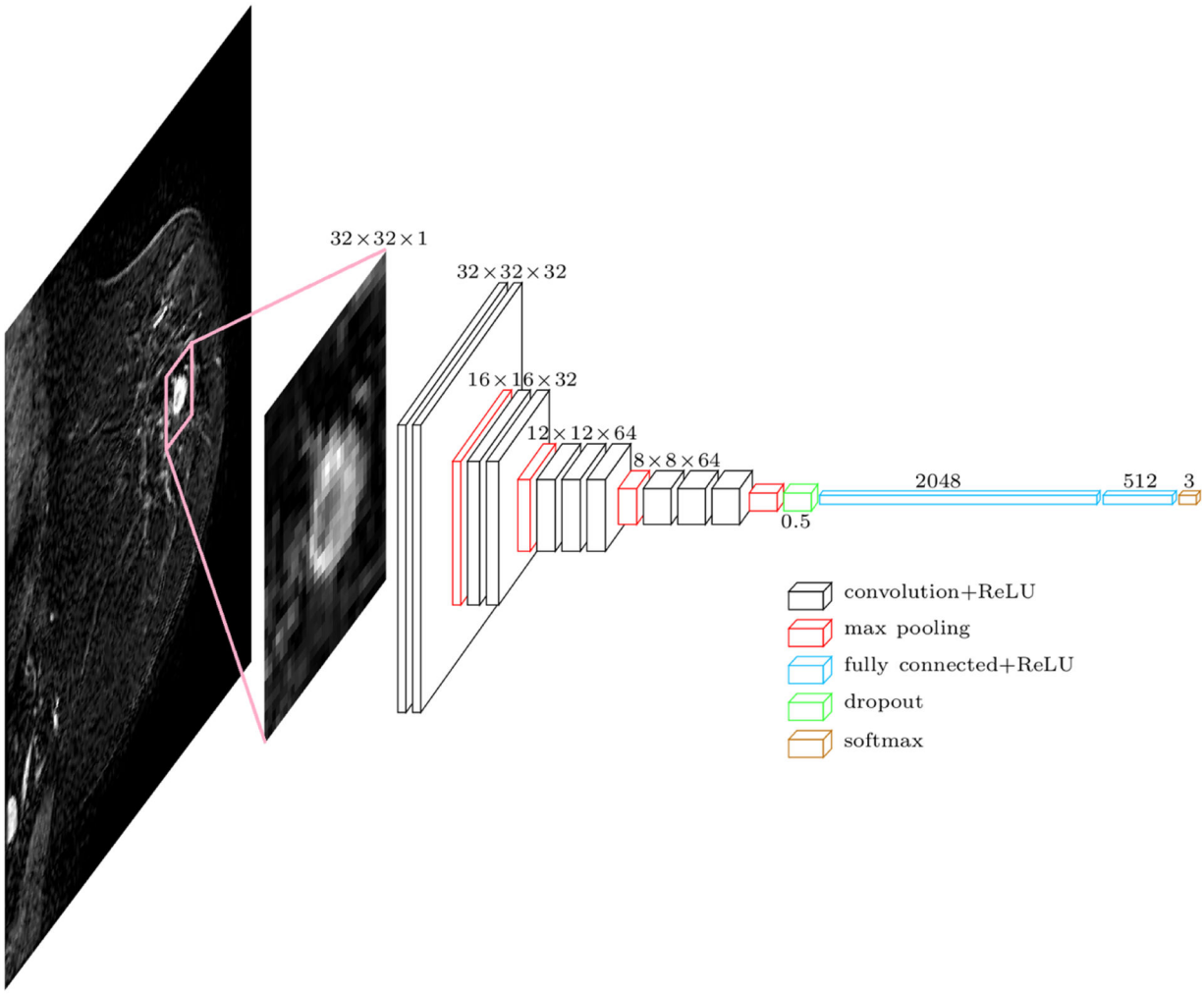

FIGURE 2:
CNN Architecture for two- and three-class classification models.

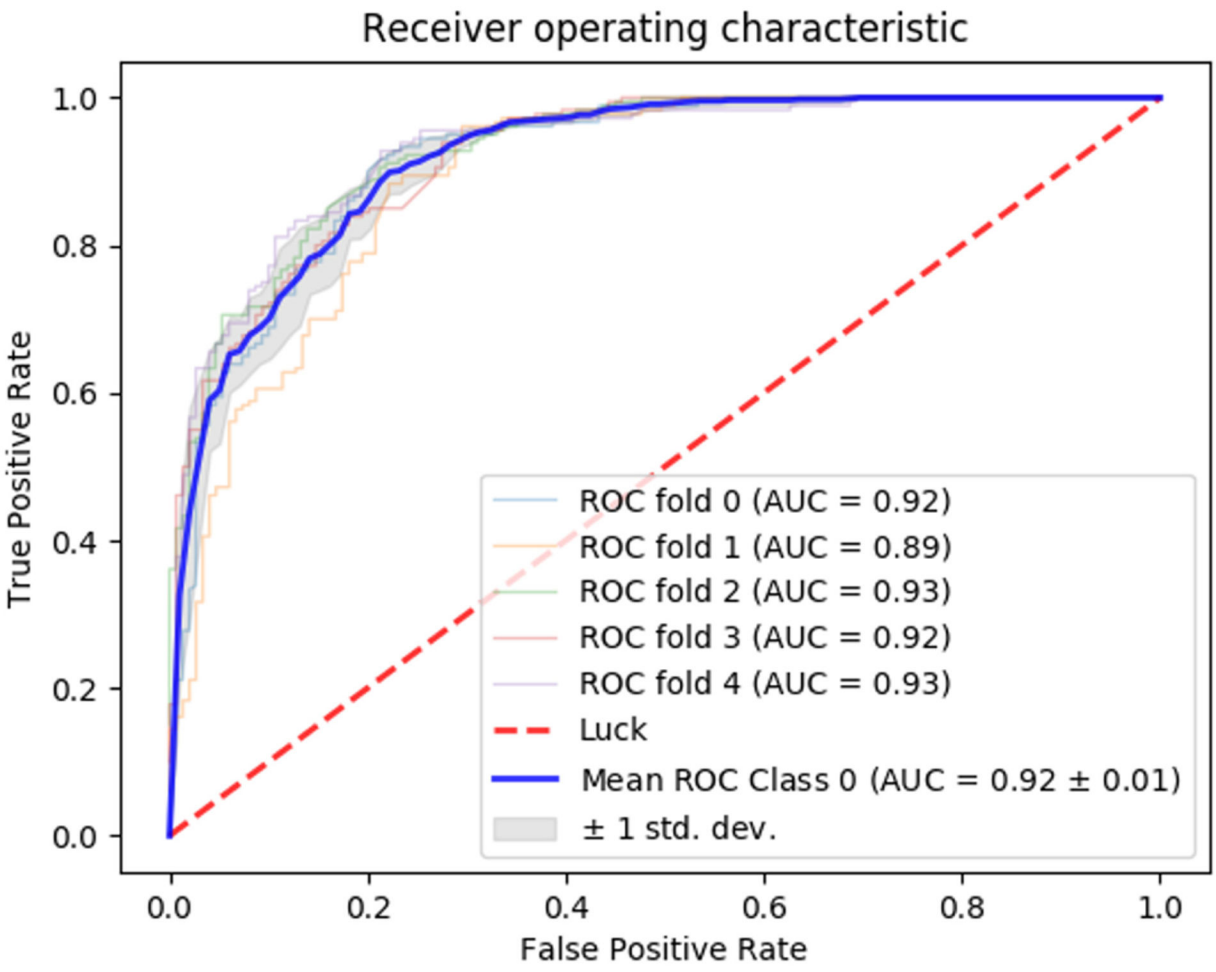

FIGURE 3:
ROC analysis for three-class CNN Oncotype DX RS prediction model.

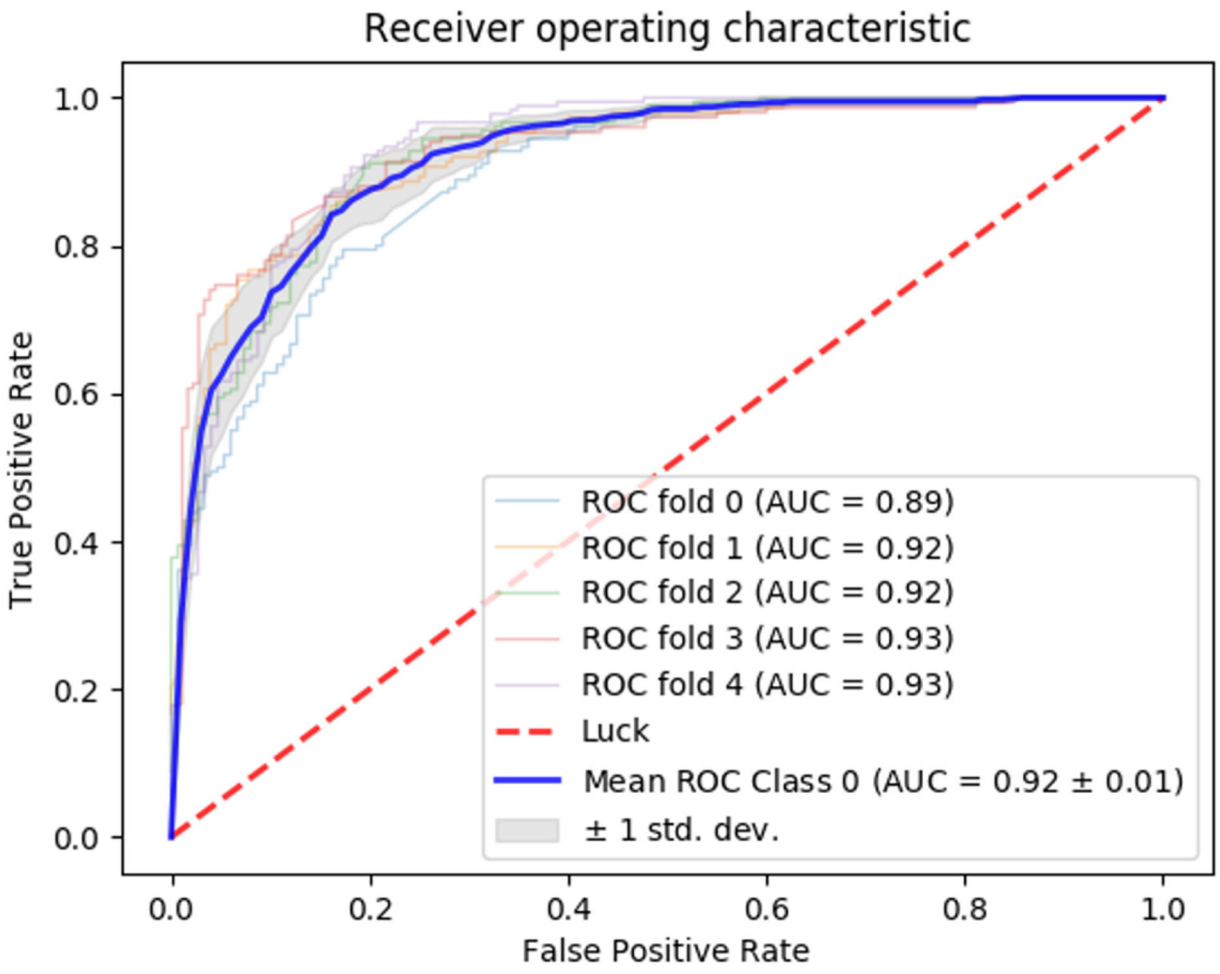

FIGURE 4:
ROC analysis for two-class CNN Oncotype DX RS prediction model.