CEM and MR Radiomics-based Biomarkers to Predict Immunohistochemistry Breast Cancer Subtypes: A comparative study

Ashif Iquebal  
Arizona State University  https://orcid.org/0000-0001-5922-3311

Siqiong Zhou (siqiong.zhou@asu.edu)  
Arizona State University  https://orcid.org/0000-0001-9131-5267

Nicholas Pfeiffer  
Mayo Clinic Arizona

Sara Ranjbar  
Mayo Clinic Arizona

Imon Banerjee  
Mayo Clinic Arizona

Kristin Swanson  
Mayo Clinic Arizona

Felipe Batalini  
Mayo Clinic Arizona

Karen S. Anderson  
Mayo Clinic Arizona

Muhammad Murtaza  
University of Wisconsin-Madison

Barbara A. Pockaj  
Mayo Clinic Arizona

Bhavika K. Patel  
Mayo Clinic Arizona

Research Article

**Keywords:** Breast cancer, Machine learning, Radiomics, Breast cancer subtype, Contrast-enhanced mammography, Breast MR imaging

**Posted Date:** November 7th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-2232518/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Purpose Accurately predicting the clinical breast cancer subtypes could be extremely helpful for radiologists, pathologists, surgeons, and clinicians and inform future treatment prediction algorithms. Therefore, we evaluate and compare the accuracy of radiomic features extracted from contrast enhanced mammography (CEM) and magnetic resonance imaging (MRI) scans to make predictions to subtypes of breast cancer.

Methods This HIPAA-compliant prospective single institution study was approved by the local institutional review board with written informed consent. Women with breast tumors 2 cm or larger underwent dynamic contrast-enhanced MRI and/or CEM for surgical staging. Semi-manual regions of interest were drawn by radiologist using Cancer Imaging Phenomics Toolkit (CaPTk). Radiomic features were obtained using PyRadiomics and MR- and CEM- based classification models were built on a low-dimensional representation of the features obtained via kernel principal component analysis. We subscribed to an ensemble tree-based learning approach called extremely randomized trees (ERT) to predict cancer subtypes captured via immunohistochemistry markers.

Results For MRI analysis, 124 women with newly diagnosed breast cancer were included in the analysis comprising 49 HR+HER2-, 37 HR+HER2+, 11 HR-HER2+, and 27 HR-HER2- cases. For CEM analysis, models were built using data from 170 female patients including 74 HR+HER2-, 41 HR+HER2+, 14 HR-HER2+, and 43 HR-HER2-. CEM based model resulted in accuracies of 55%, 72%, 88%, and 71% respectively for HR+HER2-, HR+HER2+, HR-HER2+, and HR-HER2- whereas MRI based model alone led to accuracies of 54%, 62%, 89%, and 76% respectively for HR+HER2-, HR+HER2+, HR-HER2+, and HR-HER2-.

Conclusions Radiomic features extracted from CEM and MR were strong predictors of breast cancer subtypes with CEM-based radiomic features performing slightly better, though not statistically significantly better (p = 0.82), than its MRI counterpart.

Background

Breast cancer is the most common female cancer globally, accounting for more than 12% of all new annual cancer cases [1]. It is a heterogeneous disease harboring varying biology, treatment responses, and immunohistochemistry (IHC) markers. Among these, the hormone receptor subtype of breast cancer has particularly received significant attention in clinical practice since the early 2000s and is routinely adopted into clinical guidelines for breast cancer management. The National Comprehensive Cancer Network guideline classifies breast cancer subtypes into the following four categories according to the status of hormone receptor or HR and human epidermal growth factor receptor 2 (HER2): HR+HER2-, HR+HER2+, HR-HER2+, and HR-HER2- (referred to as triple negative breast, TNBC) [1]. Treatment and surveillance plans are entirely dependent on these immunohistochemistry markers.

Breast MRI primarily detects breast cancer by highlighting areas of tumor neoangiogenesis, which demonstrates the underlying tumor vascular physiology. Breast MRI has been the gold standard imaging tool to characterize and locally stage newly diagnosed breast cancer [2–15]. An emerging imaging modality, contrast enhanced mammography (CEM), has also recently attracted broad clinical research interest with the potential to be a low-cost alternative to MR imaging. As a low-cost tool, CEM could become accessible to patients in developing countries where access to MRI tends to be limited. However, unlike MRI, CEM only provides a two-dimensional view of the breast. Therefore, establishing the efficacy of CEM biomarkers compared to MRI and its predictive capability will be crucial for its widespread adoption.

Breast cancer subtype is a significant independent prognostic factor [16]. Accurately predicting the tumor subtype of breast cancer by using MRI and CEM images could be extremely helpful for radiologists, pathologists, surgeons, and clinicians. Prediction of tumor subtype would allow personalized treatment such that aggressive surgical procedures
could be avoided. In addition, it would allow treatment costs and side effects reduction, and avoidance of patient inconvenience [17].

This study aims to examine new prognostic, radiomics-based biomarkers (used in combination with current clinical markers) to accurately predict the tumor subtype of breast cancer using MRI and CEM images and compare their predictive performance. Given the importance of breast cancer subtypes as prognostic factors in women with operable breast cancer, we aim to analyze imaging biomarkers embedded within the standard of care imaging studies performed for staging to predict immunohistochemistry-based subtypes. The successful development of such a predictive model will potentially help radiologists, pathologists, surgeons, and clinicians understand features driving breast cancer phenotypes and, more importantly, prognoses using CEM and MRI.

**Methods**

**Study population**

Data used in this study was acquired from patients enrolled in a large clinical trial at our institution assessing circulating breast tumor DNA. This is a retrospective analysis of prospectively collected patients with T2 or larger tumors who intended to undergo NST. Our patient cohort included women with invasive breast cancer of 2 centimeters or greater at clinical examination or imaging and who were undergoing neoadjuvant systemic therapy (NST) and underwent MR or CEM imaging were between 01/2015 to 01/2021, where CEM and MRI are standard clinical care at our site. Each patient was offered both MRI and CEM, however, some patients opted for one study and declined the other based on claustrophobia, allergy, or reimbursements. The images are collected prior to NST. Pregnant patients and those with ferromagnetic prosthesis were excluded from the study. Patients signed a single consent form for blood and tissue samples and imaging. The Health Insurance Portability and Accountability Act-compliant protocol and the informed consent process were approved by the local site Institutional Review Board.

Patient data were obtained from an IRB approved institutional database containing breast cancer patients. Analysis was conducted on de-identified data. At the time of our study, 124 breast cancer patients had pre-treatment MR imaging, and 170 patients had CEM imaging available for review. Pre-treatment imaging was performed before the start of NST. The breast MRI cohort comprised 49 HR+HER2-, 37 HR+HER2+, 11 HR-HER2+, and 27 TNBC cases. The CEM cohort included 74 HR+HER2-, 41 HR+HER2+, 14 HR-HER2+, and 43 HR-HER2-.

Clinical tumor size in centimeters, as well as tumor location with clock position, was recorded. The institution's breast specialized pathologist performed the histopathologic analysis of surgical specimens. HR positivity (estrogen receptor (ER) positivity or progesterone (PR) positivity or both) and HER2 expression were determined from pre-treatment core biopsy by immunohistochemistry according to ASCO/CAP guidelines and Allred score.

**Breast MR Imaging Technique**

MRI was performed using a 3.0-T imaging system (GE, Discovery 750W) in the prone position with a dedicated 16-channel breast coil (Sentinelle, In vivo Corp.). Each study included a non–fat-saturated T1-weighted sequence (FSE) in the axial orientation (TR/TE = 700/8.3 ms; matrix = 256 x 192) and a fat-suppressed T2- weighted FSE ASPIR sequence in the sagittal orientation (TR/TE = 4800/79.5 ms; matrix = 256 x 224).

A dynamic contrast-enhanced image set was also acquired, with the first series being an unenhanced fat-saturated gradient-recalled echo T1-weighted sequence (VIBRANT) followed by three dynamic contrast-enhanced fat-saturated T1-weighted gradient-recalled echo series (VIBRANT) performed after IV administration of Gadobutrol (Gadavist, Bayer).
at 30 sec, 3 min and 6 min with the use of a weight-based dosing protocol. The dynamic contrast images were acquired in the Sagittal orientation (TR/TE = 5.2/2.3 ms; matrix = 256 x 256) in 2.6 mm slices. Automatic post-processing included the generation of subtraction images between pre, and post-contrast images were produced after each phase. Late gadolinium fat-suppressed T1-weighted fast-spoiled-gradient-echo (FSPGR) sequences were also acquired for both right and left sides separately in the axial orientation (TR/TE = 115/3.15 ms; matrix = 256 x192).

For each examination, T1 weighted dynamic contrast material enhanced MRI images were analyzed for the study. Early phase dynamics subtraction images obtained at 2 mins were used for analysis. Images from the MRI were reviewed, and the lesion was located independently by a radiologist with consensus review by a second radiologist with at least five years of experience (NP, BKP). The tumor location was confirmed on the associated radiology report. MRI imaging assessment included measurement of tumor in the longest diameter. The clinical size was also recorded prior to treatment. The location of the primary tumor on the MRI examination was annotated using CaPTk [18]. Representative MRI images along with regions of interest (ROI) for each of the subtypes are shown in Fig. 1. Here, ROIs for MRI images are in 3D and we only show a representative image here for clarity.

Breast Dual Energy (DE) CEM Technique

All of our DE-CEM examinations were performed on a single commercial DE-CEM model (Hologic). Contraindications to performing CEM at our institution include renal insufficiency (glomerular filtration rate < 30 mL/min/1.73 m²), prior iodinated contrast allergy, and pregnancy. Details regarding the technique have been previously published [19; 20]. For each CEM examination, the recombined subtracted images were analyzed for the study. Images from the CEM were reviewed, and the lesion was located independently by a radiologist with consensus review by a second radiologist (NP, BKP). The location of the primary tumor on the CEM examination was annotated using Horos. Representative images of the CEM and annotations for each of the IHC subtype are shown in Fig. 2.

Radiomic Analysis and feature extraction of CEM and MRI

Before extracting radiomic features from MRI and CEM images, the intensity profile of each of the images was normalized to a scale of 100 and resampled to a pixel spacing of 3, 3, 3 in each X, Y, and Z dimension to standardize all the images. We also discretize the MRI images with a bin width of 5 pixels so that the effect of noisy pixels may be reduced. Radiomic features were extracted using the PyRadiomics v3.0.1 package in Python [21] for MRI and CEM images based on the ROIs delineated by two radiologists with 3 and 14 years of experience through a consensus review. The extracted features were a combination of first-order statistics (19 features), shape-based features that included 16 3D (only extracted for MRI) and ten 2D features, and 75 higher-order statistical features. These higher-order statistical features included 24 gray level co-occurrence matrix (glcm), 16 gray level run length matrix (glrlm), 16 gray level size zone matrix (glszm), five neighboring gray tone difference matrix (ngtdm), and 14 gray level dependence matrix (gldm). A set of 120 radiomic features for each of the MRI and CEM images were extracted from the corresponding raw images. A graphical illustration of these features is presented in Fig. 3. In addition to extracting features from the raw images, we also extract features after processing the images through certain filters. Specifically, we employ two filters, namely, Laplacian of Gaussian (LoG) and wavelet with either a high-pass filter or a low-pass filter in X, Y, and Z axis (when applicable). In total, we extracted 960 features for MRI images and 688 features for CEM images.

Dimensionality reduction and data imbalance

We employed nonlinear principal component analysis (PCA), known as kernel PCA to reduce the dimensionality of radiomic features using the sigmoid kernel [22]. After performing kernel PCA, we retained only the top 50 principal
components that explained more than 95% of the variability in the original radiomics feature space.

**Classification model**

We compare four different machine learning algorithms, including support vector machine (SVM) [23], random forests [24], adaptive boosting (AdaBoost) [25], and extremely randomized trees (ERT) [26]. Table 1 shows the performance comparison of these four algorithms on the original dataset for MRI-based phenotypes for prediction to one of the subtypes. Based on their performances over all metrics, we finally use the ensemble tree-based supervised classification methodology ERT. To avoid overfitting and obtain a consistent estimate of the performance of ERT for classifying IHC subtypes in the absence of an unseen test set, we performed 10-fold cross-validation. Thus, the ratio of training set and testing set is 90:10. It is also important to note that the IHC subtype prediction considered in this work is a four-class problem. However, due to limited sample size, we performed four binary classifications, namely HR + HER2 + vs. others, HR-HER2- vs. others, HR-HER2+ vs. others, and HR + HER2- vs. others. Similar strategies have been adopted in the literature, for instance, [27; 28].

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Positive Predictive Value</th>
<th>Recall</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.54</td>
<td>0.15</td>
<td>0.042</td>
<td>0.88</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>[0.48, 0.6]*</td>
<td>[0.0, 0.5]</td>
<td>[0.0, 0.111]</td>
<td>[0.8, 0.933]</td>
<td>[0.36, 0.493]</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.58</td>
<td>0.38</td>
<td>0.12</td>
<td>0.89</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>[0.542, 0.64]</td>
<td>[0.0, 0.667]</td>
<td>[0.0, 0.2]</td>
<td>[0.8, 0.933]</td>
<td>[0.473, 0.617]</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.53</td>
<td>0.39</td>
<td>0.31</td>
<td>0.68</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>[0.4, 0.64]</td>
<td>[0.222, 0.571]</td>
<td>[0.2, 0.444]</td>
<td>[0.533, 0.8]</td>
<td>[0.38, 0.6]</td>
</tr>
<tr>
<td>ERT</td>
<td>0.54</td>
<td>0.42</td>
<td>0.37</td>
<td>0.63</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>[0.36, 0.6]</td>
<td>[0.2, 0.5]</td>
<td>[0.2, 0.556]</td>
<td>[0.4, 0.667]</td>
<td>[0.413, 0.689]</td>
</tr>
</tbody>
</table>

*Range for each of the metric is provided within the square brackets.

**Results**

**Patient Characteristics**

Table 2 shows the distribution and characteristics of all the patients analyzed, after exclusion of patients with incomplete annotation. Based on two sample t-test and a chi-square test for categorical variables, no significant difference is observed between the patient population in CEM and MRI datasets for each IHC subtype.

**Predictive Ability of MRI-based Phenotypes for IHC subtype prediction**

Using the top 50 principal components obtained from kernel PCA, we performed the classification of the IHC subtypes of tumors. The performance metrics reported here were accuracy, positive predictive value, recall, and specificity defined as follows:

\[
\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}
\]
Positive Predictive Value = \frac{True Positive}{True Positive + False Positive}

Recall = \frac{True Positive}{True Positive + False Negative}

Specificity = \frac{True Negative}{True Negative + False Positive}

While accuracy remains the primary indicator for the performance of the classification, we included recall and specificity to establish the performance of the proposed approach considering data imbalance. Table 3 shows the results obtained via the MRI-based phenotypes obtained after performing 10-fold cross-validation. The ROC curves corresponding to the HR-HER2- and HR + HER2 + are shown in Fig. 4.

**Predictive Ability of CEM-based Phenotypes for IHC subtype prediction**

Data for CEM is available for both the low energy (LE) and DES. High energy images are not used for clinical analysis as low energy images are equivalent to standard full-field mammogram using both low energy and high energy [29]. However, no specific difference was noted in the predictive performance between the LE and DES images. Therefore, to assess the performance of CEM images, we considered the LE CEM images with mediolateral oblique (MLO) view. The same set of features were extracted as the MR images with the exception of 3D shape features as CEM images are two-dimensional. Predictive results are presented in Tables 4, respectively. The ROC curve for HR-HER2- and HR + HER2 + cases are shown in Fig. 5.
### Table 2

Characteristic and distribution of patients considered in this study.

<table>
<thead>
<tr>
<th></th>
<th>HR + HER2-</th>
<th></th>
<th>HR + HER2+</th>
<th></th>
<th>HR-HER2+</th>
<th></th>
<th>TNBC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CEM</td>
<td>MRI</td>
<td>CEM</td>
<td>MRI</td>
<td>CEM</td>
<td>MRI</td>
<td>CEM</td>
<td>MRI</td>
</tr>
<tr>
<td><strong>Number of patients</strong></td>
<td>74 (43.0%)</td>
<td>49 (39.5%)</td>
<td>41 (23.8%)</td>
<td>37 (29.8%)</td>
<td>14</td>
<td>(8.10%)</td>
<td>11</td>
<td>(8.9%)</td>
</tr>
<tr>
<td><strong>ER+</strong></td>
<td>46 (62.20%)</td>
<td>31 (63.30%)</td>
<td>25 (60.9%)</td>
<td>23 (62.2%)</td>
<td>11</td>
<td>(78.60%)</td>
<td>9</td>
<td>(81.20%)</td>
</tr>
<tr>
<td><strong>PR+</strong></td>
<td>41 (55.40%)</td>
<td>28 (57.10%)</td>
<td>24 (58.5%)</td>
<td>23 (62.20%)</td>
<td>9</td>
<td>(64.30%)</td>
<td>9</td>
<td>(81.20%)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>54</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>53</td>
<td>58</td>
<td>64</td>
</tr>
<tr>
<td>p-value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.734</td>
<td>0.975</td>
<td>0.995</td>
<td>0.233</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>27.7</td>
<td>27.35</td>
<td>25.4</td>
<td>25.4</td>
<td>24.79</td>
<td>26.7</td>
<td>27.9</td>
<td>27.7</td>
</tr>
<tr>
<td>p-value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.746</td>
<td>0.857</td>
<td>0.965</td>
<td>0.721</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>18.4–44.0</td>
<td>18.4–41.1</td>
<td>18.9–39.2</td>
<td>17.7–38.5</td>
<td>18.9–45.6</td>
<td>19.2–39.0</td>
<td>19.7–43.9</td>
<td>19.7–43.9</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>4.1</td>
<td>3.7</td>
<td>3.1</td>
<td>2.55</td>
<td>4.95</td>
<td>4.95</td>
<td>2.8</td>
<td>3.7</td>
</tr>
<tr>
<td>p-value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.403</td>
<td>0.338</td>
<td>0.978</td>
<td>0.391</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td>0.7–14.0</td>
<td>1.7–12.2</td>
<td>0.6–12.5</td>
<td>0.9–7.0</td>
<td>1.7–8.5</td>
<td>4–10.5</td>
<td>1.1–10.5</td>
<td>1.1–9.6</td>
</tr>
<tr>
<td><strong>Postmenopausal status</strong></td>
<td>63.93%</td>
<td>59.20%</td>
<td>56.40%</td>
<td>52.78%</td>
<td>50%</td>
<td>54.50%</td>
<td>71.10%</td>
<td>80%</td>
</tr>
<tr>
<td>p-value&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.69</td>
<td>0.31</td>
<td>0.49</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dense Breast</strong></td>
<td>57.63%</td>
<td>55%</td>
<td>65.50%</td>
<td>68.57%</td>
<td>66.67%</td>
<td>63.64%</td>
<td>65.80%</td>
<td>64%</td>
</tr>
<tr>
<td>p-value&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.29</td>
<td>0.46</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>49 (66.20%)</td>
<td>43 (87.80%)</td>
<td>31 (75.60%)</td>
<td>27 (72.9%)</td>
<td>11</td>
<td>(78.60%)</td>
<td>8</td>
<td>(72.70%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8 (10.80%)</td>
<td>5 (10.20%)</td>
<td>4 (9.80%)</td>
<td>2 (5.40%)</td>
<td>0</td>
<td>(9.10%)</td>
<td>1</td>
<td>(9.30%)</td>
</tr>
</tbody>
</table>

The table shows the characteristic and distribution of patients considered in this study, including the number of patients, ER+, PR+, age, BMI, size, postmenopausal status, dense breast, and race.
<table>
<thead>
<tr>
<th></th>
<th>HR + HER2-</th>
<th>HR + HER2+</th>
<th>HR-HER2+</th>
<th>TNBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian/ Pacific Islander</td>
<td>3 0</td>
<td>1 3</td>
<td>1 1</td>
<td>3 0</td>
</tr>
<tr>
<td></td>
<td>(4.50%)</td>
<td>(2.40%)</td>
<td>(8.10%)</td>
<td>(7.10%)</td>
</tr>
<tr>
<td>Native American</td>
<td>0 0</td>
<td>1 0</td>
<td>0 0</td>
<td>1 1</td>
</tr>
<tr>
<td></td>
<td>(2.40%)</td>
<td></td>
<td>(2.40%)</td>
<td>(3.70%)</td>
</tr>
<tr>
<td>African American</td>
<td>1 1</td>
<td>0 2</td>
<td>0 0</td>
<td>0 0</td>
</tr>
<tr>
<td></td>
<td>(1.4%)</td>
<td>(2.00%)</td>
<td>(5.40%)</td>
<td></td>
</tr>
<tr>
<td>Other/NA</td>
<td>13 0</td>
<td>4 3</td>
<td>2 1</td>
<td>6 1</td>
</tr>
<tr>
<td></td>
<td>(17.60%)</td>
<td>(9.80%)</td>
<td>(8.10%)</td>
<td>(14.30%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.10%)</td>
<td>(9.10%)</td>
<td>(13.96%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.70%)</td>
<td></td>
</tr>
</tbody>
</table>

*a two-sample t-test

*b Chi-square test for categorical variables based on Postmenopausal status and Dense Breast

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Positive Predictive Value</th>
<th>Recall</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR-HER2+</td>
<td>0.89</td>
<td>0.08</td>
<td>0.13</td>
<td>0.84</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>[0.84, 0.92]*</td>
<td>[0.0, 0.333]</td>
<td>[0.0, 0.5]</td>
<td>[0.818, 0.955]</td>
<td>[0.341, 0.609]</td>
</tr>
<tr>
<td>HR + HER2+</td>
<td>0.62</td>
<td>0.35</td>
<td>0.31</td>
<td>0.74</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>[0.5, 0.72]</td>
<td>[0.222, 0.5]</td>
<td>[0.143, 0.375]</td>
<td>[0.301, 0.675]</td>
<td>[0.353, 0.643]</td>
</tr>
<tr>
<td>HR + HER2-</td>
<td>0.54</td>
<td>0.42</td>
<td>0.37</td>
<td>0.63</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>[0.36, 0.6]</td>
<td>[0.2, 0.5]</td>
<td>[0.2, 0.556]</td>
<td>[0.4, 0.667]</td>
<td>[0.413, 0.689]</td>
</tr>
<tr>
<td>HR-HER2-</td>
<td>0.76</td>
<td>0.25</td>
<td>0.2</td>
<td>0.89</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>[0.667, 0.846]</td>
<td>[0.0, 1.0]</td>
<td>[0.0, 0.667]</td>
<td>[0.7, 1.0]</td>
<td>[0.421, 0.65]</td>
</tr>
</tbody>
</table>

*Range for each of the metric is provided within the square brackets.
Table 4
Performance of ERT on the original dataset for CEM imaging-based phenotypes for IHC subtype prediction.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Positive Predictive Value</th>
<th>Recall</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR-HER2+</td>
<td>0.88</td>
<td>0.08</td>
<td>0.13</td>
<td>0.98</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>[0.871, 0.933]*</td>
<td>[0.0, 0.25]</td>
<td>[0.0, 0.333]</td>
<td>[0.821, 1.0]</td>
<td>[0.339, 0.714]</td>
</tr>
<tr>
<td>HR + HER2+</td>
<td>0.72</td>
<td>0.43</td>
<td>0.15</td>
<td>0.91</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>[0.677, 0.742]</td>
<td>[0.167, 0.5]</td>
<td>[0.125, 0.25]</td>
<td>[0.818, 1.0]</td>
<td>[0.449, 0.614]</td>
</tr>
<tr>
<td>HR + HER2-</td>
<td>0.55</td>
<td>0.41</td>
<td>0.34</td>
<td>0.74</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>[0.533, 0.6]</td>
<td>[0.375, 0.5]</td>
<td>[0.25, 0.5]</td>
<td>[0.556, 0.778]</td>
<td>[0.421, 0.57]</td>
</tr>
<tr>
<td>HR-HER2-</td>
<td>0.71</td>
<td>0.35</td>
<td>0.22</td>
<td>0.91</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>[0.677, 0.767]</td>
<td>[0.0, 0.5]</td>
<td>[0.0, 0.286]</td>
<td>[0.826, 1.0]</td>
<td>[0.523, 0.671]</td>
</tr>
</tbody>
</table>

*Range for each of the metric is provided within the square brackets.

Discussion

Both MRI- and CEM-based models are able to distinguish the breast cancers with different IHC-based subtypes using a radiomics-based machine learning approach. CEM-based model performs numerically better than the MRI-based model (71.5% vs. 70.2%), although this is not statistically significant (p-value = 0.82). The similar performance between CEM- and MR-based models is somewhat surprising given that CEM only provides a two-dimensional view of the tumor, whereas MRI provides a complete three-dimensional view. We also notice that the overall predictive performance of the HR-HER2- in terms of accuracy, PPV, recall, sensitivity, and AUC is superior to other classes both across CEM and MRI test cases. These results provide evidence that CEM imaging could be as informative as MRI from a machine learning perspective. The unexpected performance of CEM-based model could be potentially attributed to its higher resolution (nearly ten times) in comparison to MR images. High-resolution images preserve the details of the tumor, especially the geometric (or shape-based) features that consistently remain the prominent features in all the classifications, and resolves the presence of microcalcifications. Alternatively, this could be explained by the larger size of the CEM cohort, which could lead to the slightly better model training and prediction.

While the problem of IHC subtype classification has been recently studied in the literature [27; 28], radiomics-based predictive models are still emerging. Because radiomic features can be automatically extracted from segmented images, they allow fast, quantitative, and reproducible features. This is varied from the current state of BI-RADS classification, which requires trained experts, and has been shown to demonstrate both inter- and intra-reader expert variability [30; 31]. Similar studies have emerged in recent years that focus on the classification of tumor subtypes using radiomics, clinical features, BI-RADS, or a combination. For instance, Wu et al. [32] employed BI-RADS features to classify four different IHC subtypes: Luminal A, Luminal B, HER2, and basal-like breast cancer achieving an accuracy of 74.1% on a cohort of 363 patients. Leithner et al. [28] employed radiomic signatures extracted from CEM images to develop a predictive model using 91 patients from one institution and validated on another institution consisting of 52 patients with an accuracy of 79.4% for Luminal A vs. Luminal B and 77.1% for Luminal B vs. TNBC. However, the authors did not report the recall and specificity of their performance. More recently, Son et al. [27] performed the prediction of IHC subtypes using radiomics signatures of synthetic mammography constructed from the digital breast tomosynthesis (DBT) for a cohort of 365 patients with an accuracy of 81.7%, 76.1%, and 56.3% for TNBC, HER2, and luminal A and B, respectively in an one class vs. others framework using the craniocaudal (CC) view. There was no
improvement in the performance when the features from the CC and MLO views were combined. Similarly, our work also demonstrated no significant difference between performance of the model when using CC versus MLO views.

We also studied the importance of radiomic features (without performing kernel PCA) using a game-theoretic approach known as Shapley values [33]. In agreement with existing studies, we note that several shape-based features emerged as prominent features in both the MRI- and CEM-based models. Some of the features that were consistently prominent across all the predictive models include: shape sphericity, axis length, shape flatness, and shape surface area. Specifically, we noted that for HR-HER2- patients, the tumors were consistently round and spherical in shape, whereas for HR + HER2- patients, the tumors were irregular. These observations are aligned with the findings reported in the literature and observed in clinical practice. For instance, Son et al. [27] reported that triple-negative tumors tend to be round or oval. In addition to the shape-based features, we also noted several intensities and correlation-based features to be significant in model prediction, particularly, first-order features such as correlation and entropy extracted from gray level co-occurrence matrix and gray level dependence matrix.

From the present study, as well as other recent reports in the literature, it is evident that radiomic features are effective in distinguishing IHC subtypes. Limitations include classic large $p$ small $n$ problem in machine learning, [34] caused by limited number of patients ($n = 170$ for CEM and $n = 124$ for MRI) studied in relationship to the high-dimensionality of features in the dataset ($p = 960$ for MRI features). This also limits the development of multi-class predictive models [35]. Class imbalance leads to poor precision and recall performance of the predictive models, and while synthetic resampling strategies could help augment the existing datasets, they seldom improve the predictive performance. The issue of class imbalance could partially be addressed by analyzing larger cohorts. To avoid a problem of data harmonization from inter-scanner and/or inter-radiologist variability, we use data collected from a single scanner annotated by the same set of radiologists, which leads to limited generalizability. The high model complexity and black-box nature of machine learning models employed limits their interpretability. The authors’ ongoing work is focused on making these machine learning models more interpretable so that the inference generated from these models may not only lead to more understanding of the biology but also to informing practitioners in the decision-making process. Interpretability and model fairness also allow for monitoring against potential biases associated with the underrepresentation of racial minorities in most datasets. In general, these limitations are being addressed via multi-institutional collaboration for the generation of much larger and diverse datasets for generation, and comparison of different models.

For the breast MRI image analyses in this study, we utilized only dynamic contrast-enhanced images and did not incorporate the associated T2 weighted imaging in our analysis [36; 37]. The expectation would be that these sequences would provide additional surrogates for biological data of the tumors and will be included in future studies. The patients included in the study had biopsy-proven invasive breast cancer prior to undergoing MRI and CEM contrast-enhanced imaging, where by post biopsy change may confound our results. The heterogeneous enhancement that can occur in a post biopsy bed can alter the appearance of the native tumor. However, this is standard clinical care and beneficial to build models as imaging in true clinical practice is available.

**Conclusion**

MRI- and CEM-based machine learning models demonstrate a comparable performance based on radiomic features to classify breast cancer according to known IHC subtypes. Via an ensemble machine learning algorithm known as extremely randomized trees, we show that MRI and CEM based radiomic features can predict IHC subtypes with 70.2% and 71.5% accuracy, respectively. Using feature importance, we note that shape-based features such as shape sphericity, elongation, and shape surface area are consistently the most prominent features across all the predictive models developed in this study. Our current and future works focus on validating our machine learning model on a large
cohort of patients, generating interpretability for black-box machine learning methods. Doing so, will help improve clinician uptake of these models in clinical practice. Quantitative prognostic and predictive models to predict complete pathological response to neoadjuvant chemotherapy will help strengthen precision medicine and lead to improved patient survival, without the otherwise additional unnecessary treatment side effects from therapies that would have been futile.

**Abbreviations**

DE-CEM: Dual Energy Contrast Enhanced Mammography

MRI: Magnetic Resonance Imaging

HIPAA: Health Insurance Portability and Accountability Act

CaPTk: Cancer Imaging Phenomics Toolkit

ERT: Extremely Randomized Trees

IHC: Immunohistochemistry

HR: Hormone Receptor

HER2: Human Epidermal Growth Factor Receptor 2

TNBC: Triple Negative Breast Cancer

NST: Neoadjuvant Systemic Therapy

FSE: Fast Spin Echo

TR/TE: Repetition Time and Echo Time

FSPGR: Fast Spoiled Gradient Echo

ASPIR: Adiabatic Spectral Inversion Recovery

DNA: Deoxyribonucleic Acid

VIBRANT: T1-weighted Volume Imaging for Breast Assessment

ASCO/CAP: American Society of Clinical Oncology/College of American Pathologists

ROI: Region of Interest

LoG: Laplacian of Gaussian

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

LE: Low-Energy

DES: Dual-Energy Subtraction
Declarations

Acknowledgments

The authors are grateful for the kind support provided by the Gerstner Family and the Brandt Young Scholarship from the Centers of Individualized Medicine and ASU-Mayo Clinic Summer residency program. The data employed in this study was collected at the Mayo Clinic Arizona.

Funding. This study has received funding from Gerstner Family and the Brandt Young Scholarship from the Centers of Individualized Medicine and ASU-Mayo Clinic Summer residency program.

Competing interests. The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Authors’ contributions. Ashif Iquebal, Siqiong Zhou, Nicholaus Pfeiffer, Sara Ranjbar, Imon Banerjee, and Bhavika K. Patel analyzed and interpreted patient data. Ashif Iquebal, Siqiong Zhou, Nicholaus Pfeiffer, and Bhavika K. Patel developed the statistical models. Ashif Iquebal, Sara Ranjbar, Imon Banerjee, Kristin Swanson, Felipe Batalini, Karen S. Anderson, Muhammad Murtaza, and Barbara A. Pockaj prepared and revised the manuscript.

Data availability. The data in the study was obtained by Bhavika K. Patel at Mayo Clinic of Arizona, under the IRB approved institutional database containing breast cancer patients. The codes developed in this study will be available from the authors upon request.

Ethics approval. Institutional Review Board approval was obtained.

Consent to participate. Only if the study is on human subjects: Written informed consent was obtained from all subjects (patients) in this study.

Consent to publish. The authors affirm that human research participants provided informed consent for publication of the images in Fig. 1, Fig. 2, and Fig.3.

References


