Multimodal Prediction of Breast Cancer Relapse Prior to Neoadjuvant Chemotherapy Treatment

Simona Rabinovici-Cohen[1], Ami Abutbul[1], Xosé Fernandez[2], Oliver Hijano Cubelos[2], Shaked Perek[1], Tal Tlusty[1]
[1]IBM Research – Haifa, Mount Carmel, Haifa 3498825, Israel
simona@il.ibm.com
[2]Institut Curie, 26 Rue d'Ulm, 75005 Paris, France

Abstract. Neoadjuvant chemotherapy (NAC) is one of the treatment options for women diagnosed with breast cancer, in which chemotherapy is administered prior to surgery. In current clinical practice, it is not possible to predict whether the patient is likely to encounter a relapse after treatment and have the breast cancer reoccur in the same place. If this outcome could be predicted prior to the start of NAC, it could inform therapeutic options. We explore the use of multimodal imaging and clinical features to predict the risk of relapse following NAC treatment. We performed a retrospective study on a cohort of 1738 patients who were administered with NAC. Of these patients, 567 patients also had magnetic resonance imaging (MRI) taken before the treatment started. We analyzed the data using deep learning and traditional machine learning algorithms to increase the set of discriminating features and create effective models. Our results demonstrate the ability to predict relapse prior to NAC treatment initiation, using each modality alone. We then show the possible improvement achieved by combining MRI and clinical data, as measured by the AUC, sensitivity, and specificity. When evaluated on holdout data, the overall combined model achieved 0.735 AUC and 0.438 specificity at a sensitivity operation point of 0.95. This means that almost every patient encountering relapse will also be correctly classified by our model, enabling the reassessment of this treatment prior to its start. Additionally, the same model was able to correctly predict in advance 44% of the patients that would not encounter relapse.

Keywords: Breast MRI, Convolutional Neural Networks, Neoadjuvant Chemotherapy Treatment.

1 Introduction

Neoadjuvant chemotherapy (NAC), in which chemotherapy treatment is administered prior to surgical therapy, is one of the approaches available in treating locally advanced breast cancer. The potential clinical advantages of NAC have been largely studied and include improved success of breast-conserving therapy, minimized nodal surgery, and more accurate *in-vivo* observation of tumor sensitivity [1]. Today, the clinical parameters used to select the NAC option are based on breast cancer subtype, tumor size, disease grade, number of affected nodes, age, and tumor growth amongst others. Imaging

is used to evaluate the position of the tumor and its size, but not to predict the outcome of the treatment. Moreover, quantitative models based on clinical and imaging features are not considered when evaluating the potential success of the treatment.

Predicting a relapse after NAC, essentially determining whether the breast cancer is likely to recur in the same location, is an important clinical question. If this future outcome could be predicted based on data available prior to the initiation of NAC treatment, it could impact the treatment selection. Because administering chemotherapy may weaken or even prevent other treatments, it is vital to correctly assess the contribution of NAC treatment in advance.

About 10% of NAC treated patients will suffer a relapse, but clinicians have difficulty estimating who is at risk for this outcome. Artificial intelligence models that predict relapse can empower clinicians in their treatment selection and decision-making. Specifically, predicting treatment outcome using medical imaging is an emerging area of interest in the medical community. It aims to extract large numbers of quantitative features from the patient's own medical images, and thus is an important enabler of precision medicine.

In this paper, we describe methods to improve relapse prediction using multi-modal data of different types. We describe a deep learning (DL) model for magnetic resonance imaging (MRI) data, a traditional machine learning (ML) model for clinical data, and an ensemble model of the individual clinical and MRI models. For the imaging modality, we use Dynamic Contrast Enhanced MRI (DCE-MRI) of the breast. DCE-MRI imaging acquires T1 changes in tissues before and after injection of gadolinium-based contrast agents at several points in time. We trained and evaluated our models on a cohort of 1738 patients, out of which 567 have MRI data. The results show that our approach is able to identify those patients likely to encounter relapse.

The paper is organized as follows. In Section 2, we describe the work related to this topic. We present the methods used to develop our multimodal predictor in Section 3 and the evaluation of our models in Section 4. We then discuss our results and conclusions in Section 5.

2 Related Work

Recent studies have explored various methods to predict cancer repetition after chemotherapy treatment. There are three types of cancer repetition: (i) relapse, the tumor redevelops at the same location where it was diagnosed before treatment; (ii) metastasis, the spread of cancer from the original tumor where it first formed to other areas in the body; (iii) recurrence, the cancer may come back to the same place as the original tumor or to another place in the body. Previous work can be divided into two major methods: predicting repetition using clinical features [2-6] and prediction using features that are extracted from imaging modalities [7-11].

Abreuet et al. [2] present a comprehensive review of 17 published studies done between 1997 and 2014, which predict the recurrence of breast cancer from clinical data using different machine learning techniques. The work shows the gaps in current studies, such as the lack of data. Most of the works use very small datasets for training and

evaluation, imbalanced cohorts, and problematic feature selection. Recent work by Chen et al. [3] present a comparison study for the evaluation of a single classifier to predict recurrence from clinical data features. Different ML methods are used to evaluate a public set of 286 patients, and the results are compared using different metrics. Other works fuse the predictions of more than one classifier to achieve more accurate results [4,5]. Tseng et al. [6] use ML techniques on a cohort of 148 patients to predict metastasis from clinical features, such as demographic data, tumor information, pathology data, and laboratory data.

The use of MRI data to predict the repetition of cancer has not been widely explored. Hylton at el. [7] and Drukker at el. [8] use tumor volume approximations, as extracted from DCE-MRI modality, to predict recurrence. A significant correlation between approximated volumes and recurrence is presented in both works for tumor volumes extracted at different times during neoadjuvant chemotherapy treatment. In an extension to [8], Drukker at el. [9] add 7 kinetic curve features extracted from within the tumor area of MRI, in addition to the approximated tumor volume. The extracted features were used to train their Long Short Term Memory (LSTM) model. This model was then evaluated on the ISPY1 publicly published dataset, which contains 222 patients [10], of which 157 patients were selected for the analysis. Another approach for using the MRI modality extracts texture features, which indicate the tumor's heterogeneity in addition to its size feature [11]. This work shows a significant correlation between these features and the ability to predict recurrence.

There are other works related to our problem that predict NAC therapy response using imaging modality. Hyunjong et al. [12] use radiomic features extracted from ROI PET/CT volumes as well as clinical and pathological features to train a logistic regression model to predict NAC treatment response. Rabinovici-Cohen at el. [13] explore features extracted from mammography and clinical data for response prediction, and present possible improvements by fusing mammograms and clinical features. Most of the works for response prediction use features extracted from the MRI modality [14-18]. Eben et al. [14] created a multi-representation-based prediction of response to NAC therapy in breast cancer. They use both CNN and radiomics on DCE-MRI volumes to extract features from within the tumor area and from the peritumoral region outside the tumor. Work by Haarburger et al. [15] uses radiomic texture features extracted from the DCE-MRI tumor area of 38 patients. A retrospective study by Ravichandran et al. [16] on 42 patients, examined the ability to predict response using a two-branch CNN with cropped DCE-MRI images around a lesion, from before and after chemotherapy is administered. Ha et al. [17] used the open dataset of ISPY1 [10] for the task. Here, a CNN was applied to MRI-DCE tumor slices with consideration of pre and post contrast via the input channel of the images. He et al. [18] applied a VGG-like network to predict the response on lesion patches extracted from a 3D voxel segmentation.

Our work differs from previous methods in four main aspects: (i) We focus on the prediction of relapse that has its own clinical importance, rather than the prediction of recurrence or response to treatment. (ii) We work specifically on patient data before NAC treatment, allowing us to focus on a more defined task. (iii) Compared to previous works, we use a relatively large set of imaging data and train our end-to-end neural network using full images extracted from the MRI volume. In addition, we evaluate the

results on a large holdout set of 100 patients (iv) We perform multimodal analysis that includes both imaging and clinical features.

3 Methods

Our dataset was collected from patients prior to NAC treatment, and contains data from several modalities, including MRI, mammography, ultrasound, and clinical data. Because our work is focused on MRI and clinical data modalities, our model consists of two branches. Each branch was trained using one of the modalities. We then combined the two branches into one final ensemble model. In this section, we elaborate on each of these components. We present our dataset, describe the MRI model branch and the clinical model branch, and then detail the final ensemble model that combines the two branches.

3.1 Dataset and Annotations

Our dataset is from Institut Curie in France and includes a cohort of 1738 breast cancer patients that received NAC treatment between 2012 and 2018. For each patient, the data includes a label marking whether this patient encountered a relapse since her treatment ended. From this cohort, we excluded 100 patients that had clinical and MRI data, for holdout evaluation. The remaining 1638 patients were considered for our cross-validation experiments.

We used two data subsets in our experiments because some patients had only clinical data while other patients had clinical and MRI data. The first data subset is a large cohort of 1638 patients for clinical data evaluation. The clinical data included demographics such as age, weight, height, and tumor properties such as breast cancer histology, grade of the tumor, Ki67, and molecular subtypes based on estrogen, progesterone, and HER2. The second data subset was a small cohort of 467 patients who, in addition to the clinical data, also had MRI scans taken prior to NAC treatment. The small cohort is a subset of the larger cohort.

A DCE-MRI scan of a patient with breast cancer includes multiple volumes. The volumes are taken before a contrast agent is injected, and at several intervals after the injection. For our analysis, we used a digital subtraction of the volume acquired after injection of the contrast agent and the baseline volume acquired before the injection. We chose to use the subtraction volumes because this type of imaging is used by radiologists for medical diagnosis and was likely to contain the information relevant for our analysis.

As mentioned, we held out the data for a cohort of 100 patients that included clinical and MRI data. The inherent distribution between negative and positive samples for the holdout set was similar to the cross-validation set. This holdout data was annotated by expert radiologists. They annotated the most important subtraction volume in which the tumor appeared to be the brightest in terms of relative illumination. In the selected volume, they also annotated the significant slice in which the tumor was the largest.

The rest of the MRI volumes were used for cross-validation. The most important subtraction as well as the significant slice were annotated by non-expert researchers.

Data distribution in the cross-validation cohorts and the holdout cohort is presented in Table 1.

Table 1. Number of patients for relapse predication

	Total number of patients	Relapse	Relapse-free
Large cohort	1638	187	1451
Small cohort	467	46	421
Holdout cohort	100	10	90

3.2 MRI Model

Our MRI model branch consists of two components. The first component is a CNN model that produces embedding features containing the CNN output score, together with 32 features taken from the previous CNN layer. These features are then used as input to the second component, which includes a logistic regression classifier. In this section, we begin by describing the pre-processing steps we applied to the raw MRI subtractions and continue with a description of each model.

Pre-processing: The input to the CNN is the significant slice and the two pre and post adjacent slices (i.e., three slices in total) that are extracted from the selected MRI subtraction volume. The significant slice is the slice in which the tumor is most visible and appears to be largest. This slice is not well-defined and different clinicians may select different slices as the significant slice.

The selected slices undergo a cropping and resizing process. Our data consisted of axial MRI volumes, which contain both sides of the breast. Hence, we cropped the image vertically and continued processing only the relevant side in which the tumor was located. Then, we cropped the image horizontally to exclude non-breast parts that appeared in the image. This process was done automatically using a sliding window, where we searched the most enhanced organs within the first slice in the MRI volume, and found a cut line above them that was used for our three selected slices. Each of the vertically and horizontally cropped slices was then resized to 512 x 256 pixels to bring them all to the same size.

The last two steps of the pre-processing included rotating the slices, so the breast was facing in the same direction for all slices. We also under sampled the slices where there was overlap between slices in the volume.

CNN model: Our CNN model is a modification of ResNet [18] as a classifier. We specifically used ResNet18 formulation, but reduced the number of filters per layer to speed up training and avoid over-fitting. The original Resnet18 consists of blocks of convolutions, with residual connections between the blocks. Each convolution layer is followed by a batch normalization layer and ReLU activation. For our network, we used 7 residual blocks with [32, 64, 128, 128, 256, 256] filters per convolutional layer. This 2D-CNN model was applied simultaneously to the 3 slices, i.e., the same 2D-CNN

model with the same weights was applied to each slice. Next, a 4D-tensor was used to aggregate features produced from the 3 input slices. Finally, a 3D convolution layer was applied, followed by a 3D average global pooling layer. The output of the pooling layer was treated as an embedding vector $v \in \mathbb{R}^{32}$. On top of this embedding layer, we added a simple sigmoid-activated linear layer as an output layer. A detailed diagram of the CNN model is depicted in Figure 1.

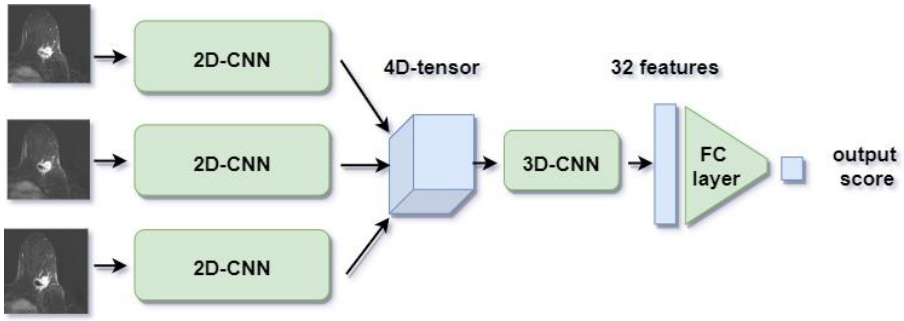

Fig. 1. **MRI CNN architecture.** Three adjacent MRI slices (pre-significant, significant, post-significant) form the input to three 2D-CNN that have the same weights. The features are aggregated into a 3D-CNN followed by an average global pooling layer and a fully connected layer that outputs the probability of a patient having a relapse.

CNN Training: we split the small cohort of 467 patients with MRI scans into 5 folds with equally distributed positive and negative samples among folds. We implemented our ResNet model in Keras with TensorFlow as backend [19]. We trained from scratch using the ADAM optimizer with binary cross-entropy as a loss function. We used a batch size of 24 samples and an initial learning rate of 10^{-4} with 'reduce on plateau' scheduler. With the training limited to 150 epochs using an early stopping protocol, we then selected the best epoch weights based on the AUC on validation set. For each volume (3 selected slices), we applied a standardization of zero mean and unit variance independently from other volumes. Also, as part of the practice for small datasets, we applied several regularization methods to prevent over-fitting. These included L_2 regularization and augmentation, which included shifts, zoom, rotation, and horizontal flip.

We refer to the model that outputs the CNN scores as the "MRI-scores" model. We also used the CNN as a feature extractor. For each subtraction volume, it produced a feature vector of size 33, which was a concatenation of the output score and the 32 features from the output of the pooling layer.

Embeddings model: We used 33 embedding features coming from the CNN and applied a scaler that scales all features to the [0,1] range. We then trained the embeddings with logistic regression and created a model that we refer to as "MRI-embeddings".

In summary, we produced two models: MRI-scores and MRI-embeddings. MRI-scores is based on the CNN scores without the logistic regression step and MRI-

embeddings is the output after the logistic regression step. We performed cross-validation and computed the receiver operating characteristic (ROC) area under the curve (AUC) with a confidence interval, specificity at sensitivity for each fold, as well as the mean values. We then selected the best model and used it to evaluate the holdout data AUC and specificity at several sensitivity operation points.

3.3 Clinical Model

We split the large cohort of 1638 patients with clinical information into 5 folds with equally distributed positive and negative samples among folds; this covered approximately 90% negative relapse and 10% positive relapse patients. The folds were created in correlation with the folds of the small cohort that includes the patients with MRI; namely, a patient remains in the same fold in both datasets.

We created our model using the 26 pre-treatment features per patient, described in Section 3.1. The features have values in different ranges and some values are missing; thus, we preprocessed the data by applying a scaler that scales all features to the [0,1] range. An imputation process replaced missing values with the mean value. One feature that suffers from a lot of missing values is lymph node involvement. This weakened our model as this feature was found to be a strong predictor of relapse [20].

To select the best classifier for our task, we trained the data with three known ML algorithms: Random Forest, Logistic Regression, and XGBoost. We performed cross-validation and computed the ROC, AUC with confidence interval, specificity at sensitivity for each fold, and the mean values across folds. We then selected the best model, found to be Random Forest, and used it to evaluate the holdout data AUC and specificity at several sensitivity operation points. We also examined the features of importance produced by our models.

3.4 Ensemble Model

The ensemble model depicted in Figure 2 receives six scores per patient: three scores based on clinical data and three scores based on the MRI data. To improve generalization, we created multiple variations of each model where each different variation started its train from a different initialization. Thus, the three scores for clinical data are produced from three variations of the clinical model that differ in their training initialization. Likewise, the three scores for MRI data are produced from three variations of the MRI model that differ in their training initialization.

We tried both options of the MRI model, namely MRI-scores and MRI-embeddings. We created an ensemble of the MRI-scores model with the clinical model, and then compared it to an ensemble of MRI-embeddings model with the clinical model.

We examined several strategies for combining the models and evaluated the cross-validation AUC and specificity at sensitivity for each option. We first tried the stacking classifier, in which we trained a meta model on top of the six models' scores using the small cohort folds. We also tried several voting strategies. However, we found that the most effective strategy used the average value of all available scores per patient.

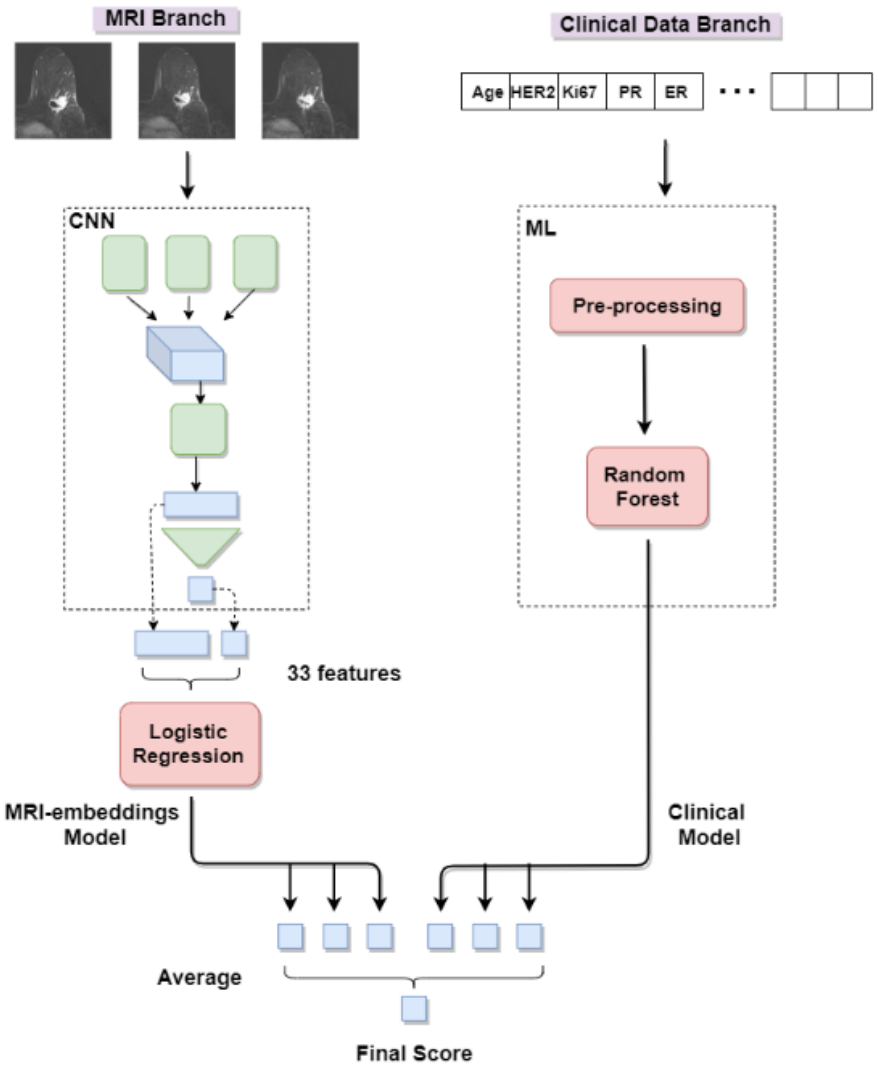

Fig. 2. Ensemble model architecture. (left) MRI model branch with three slices as input, (right) clinical model branch with 26 clinical features as input, and (bottom) merge of the two branches into one ensemble. The final score is the average of three scores from MRI-embeddings model variations and three scores from clinical model variations.

4 Results

For each modality, we evaluated the individual models, as well as the final ensemble model. Each model was trained on the largest dataset it could use. For example, the MRI was trained on the small cohort while the clinical model was trained on the large

cohort. We evaluated the final ensemble model on the small cohort since it included both MRI and clinical data, and we could compare the contribution of each modality.

As part of our evaluation, we performed a 5-fold cross-validation as well as test on a holdout dataset of 100 patients. For both, cross-validation and holdout test, we report AUC with 95% confidence interval and specificity at a sensitivity operation point of 0.95. For the holdout evaluation, the patient score was the average of the scores of the 5 models selected from the 5-fold cross-validation.

Table 2 below summarizes the results for the cross-validation and holdout test. In the MRI-only models branch, we evaluated two models: the MRI-scores model whose score is the output of the CNN (row 1) and the MRI-embeddings model, in which the score is the output of the logistic regression applied on the CNN embedding features (row 2). In the cross-validation, the MRI-scores model achieved 0.716 AUC with 95% confidence interval [0.672, 0.756] and specificity 0.363 at sensitivity operation point of 0.95. In the holdout test, the MRI-scores model achieved 0.682 [0.639, 0.722] AUC and 0.409 specificity. For the second MRI-embeddings model, the cross-validation achieved a slightly worse result with AUC of 0.708 [0.665, 0.749] and the same specificity. In the holdout data, the MRI-embeddings model achieved 0.704 [0.661, 0.743] AUC and specificity 0.424.

In the clinical-only model branch, similar results were achieved using either XGBoost or Random Forest classifiers, but Random Forest was slightly better and thus selected. It is possible that further hyperparameters tuning is needed when using XGBoost. In cross-validation, we obtained 0.687 [0.642, 0.728] AUC and 0.321 specificity. In the holdout test, we obtained 0.671 [0.627, 0.711] AUC and 0.2 specificity (row 3). The important features found by the model were age, BMI, Ki67, tumor grade, and molecular subtypes HER2, estrogen, and progesterone.

In the MRI with clinical ensemble model, we evaluated both MRI-scores with clinical (row 4) and MRI-embeddings with clinical (row 5). The second option achieved better AUC and specificity in the cross-validation evaluation and thus was selected as our final model. In the cross-validation, the final model achieved 0.745 [0.702, 0.784] and 0.442 specificity, while the holdout test achieved 0.735 [0.694, 0.773] and 0.438 specificity.

Table 2. Evaluation of the models on cross-validation and holdout test

	Cross-validation		Holdout test	
	AUC	Spec at Sens=0.95	AUC	Spec at Sens=0.95
MRI-scores	0.716	0.363	0.682	0.409
MRI-embeddings	0.709	0.363	0.704	0.424
Clinical	0.687	0.321	0.671	0.2
MRI-scores with Clinical	0.737	0.385	0.716	0.233
MRI-embeddings with Clinical (final model)	**0.745**	**0.442**	**0.735**	**0.438**

Figure 3 below shows the cross-validation and the holdout test ROC curves. They exhibit similar trends. In both, the MRI model shows promise in predicting relapse after NAC treatment with good specificity for above 0.95 sensitivity. The clinical model shows the ability to predict relapse with higher specificity around the 0.5 sensitivity but lower specificity around the 0.95 sensitivity. The ensemble of MRI and clinical leveraged both modalities and improved the AUC and specificity at various operation points.

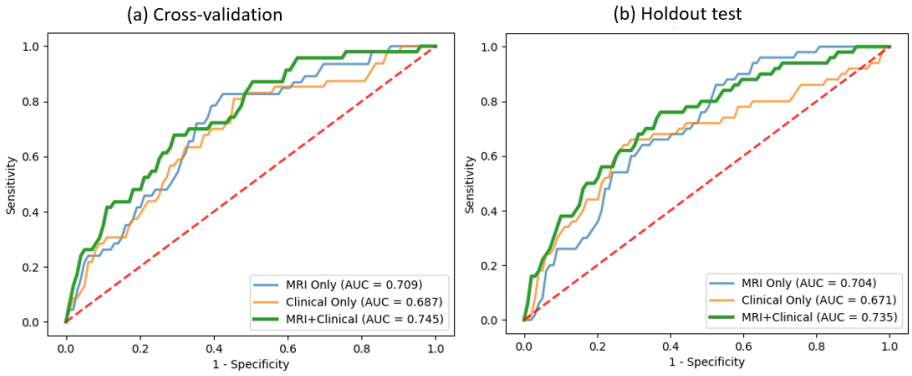

Fig. 3. Cross-validation and holdout ROC curves. (a) Cross-validation evaluation with MRI+Clinical ensemble mean AUC of 0.745 (b) Holdout evaluation with MRI+Clinical ensemble mean AUC of of 0.735.

5 Discussion and Conclusion

We demonstrated our ability to predict relapse using multimodal algorithms that include features extracted from MRI images and clinical data prior to neoadjuvant chemotherapy treatment. Each modality alone shows the ability to offer predictions in this problem setting, but the multimodal model offers better results.

We used deep learning algorithms to analyze our MRI models and traditional machine learning algorithms to analyze the clinical data. Then, we combined the two branches to create an ensemble model that produced the final prediction. Using two branches enabled us to use the best method per modality and utilize the maximum available data for each data type. After excluding the data of 100 patients for holdout, we had a cohort of 1638 patients with clinical information, out of which 467 patients also had MRI data. We were able to use the large cohort to train the clinical model and the small subset cohort to train the MRI model. Moreover, experimental training of a clinical model on the small cohort obtained an almost random model, so using the large cohort for the clinical model training was essential.

In the MRI branch, we examined two models: MRI-scores and MRI-embeddings. The cross-validation MRI-scores produced results that were slightly better in AUC than the MRI-embeddings but slightly worse in specificity. When we examined the cross-validation of these models ensembled with the clinical model, the MRI-embeddings seemed to be more calibrated and it outperformed in both AUC and specificity. Thus, the MRI-embeddings was selected for the final model.

While our MRI data for predicting relapse after NAC treatment is one of the largest compared to those reported in prior art, it is relatively small for deep learning networks. Moreover, MRI has no standardized protocol for scan acquisition and high variance of image resolution, voxel size, and image contrast dynamics. We selected special MRI preprocessing and neural network to adjust for these limitations, and the major contribution of this modality to our prediction is clear. Yet, to get robust models that are not sensitive to fold partitions and generalize better, we need to retrain our models on much larger datasets.

We used a holdout of 100 patients, which is relatively large when compared to our train set of 467 patients with MRI. However, comparing the results on the holdout with the results on the cross-validation shows that there isn't a significant decrease in performance, and we still have good generalization in the holdout cohort. This holds the promise that our models may be able to generalize to unseen but similar datasets.

High sensitivity was important in our problem setting since we wanted almost all patients that encountered relapse to be correctly classified by our model and enable NAC treatment options to be reassessed in advance. It is also important to know the specificity in these high sensitivity operation points. A false negative could potentially turn into a life-threatening situation if the patient thinks she may be cured, while in fact the cancer will come back. Adding the MRI modality enabled us to improve the specificity at high sensitivity operation points.

A future direction we intend to follow lies in improving our imaging-based model by using additional imaging data, professional annotations, and different advanced methods. In addition, using a larger cohort from additional sites is expected to help produce more generalized models.

Acknowledgements

We thank Prof. Fabien Reyal and Dr. Beatriz Grandal Rejo of Institut Curie for defining the clinical use case. We thank Chani Sacharen from IBM Research - Haifa for her help in editing the manuscript.

Research reported in this publication was partially supported by European Union's Horizon 2020 research and innovation program under grant agreement No 780495. The authors are solely responsible for the content of this paper. It does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of data appearing therein.

References

1. Teshome, M. and Hunt, K. K.: Neoadjuvant therapy in the treatment of breast cancer. In: Surgical oncology clinics of North America, 23(3), pp. 505–523 (2014).
2. Abreu, Pedro Henriques, et al.: Predicting breast cancer recurrence using machine learning techniques: a systematic review. In: ACM Computing Surveys (CSUR) 49.3 pp. 1-40 (2016).
3. Goyal, Kashish, Preeti Aggarwal, and Mukesh Kumar: Prediction of Breast Cancer Recurrence: A Machine Learning Approach. In: Computational Intelligence in Data Mining. Springer, Singapore, pp. 101-113 (2020).

4. Chen, Xi, et al.: A Reliable Multi-classifier Multi-objective Model for Predicting Recurrence in Triple Negative Breast Cancer. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, (2019).
5. Al-Quraishi, Tahsien, et al.: Breast cancer recurrence prediction using random forest model. In: International Conference on Soft Computing and Data Mining. Springer, Cham, (2018).
6. Tseng, Yi-Ju, et al.: Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies. In: International journal of medical informatics 128 pp. 79-86 (2019).
7. Hylton, Nola M., et al. "Neoadjuvant chemotherapy for breast cancer: functional tumor volume by MR imaging predicts recurrence-free survival—results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL." *Radiology* 279.1 (2016): 44-55
8. Drukker, Karen, et al.: Most-enhancing tumor volume by MRI radiomics predicts recurrence-free survival "early on" in neoadjuvant treatment of breast cancer. In: Cancer imaging 18.1 12 (2018).
9. Drukker, Karen, et al.: Deep learning predicts breast cancer recurrence in analysis of consecutive MRIs acquired during the course of neoadjuvant chemotherapy. In: Medical Imaging 2020: Computer-Aided Diagnosis. Vol. 11314. International Society for Optics and Photonics, (2020).
10. ISPY1 homepage, https://wiki.cancerimagingarchive.net/display/Public/ISPY1.
11. Li, Hui, 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. In: Radiology 281.2 pp. 382-391 (2016).
12. Lee, Hyunjong, et al.: Predicting response to neoadjuvant chemotherapy in patients with breast cancer: combined statistical modeling using clinicopathological factors and FDG PET/CT texture parameters. In: Clinical Nuclear Medicine 44.1 pp. 21-29 (2019)
13. Rabinovici-Cohen, Simona, et al.: Radiomics for predicting response to neoadjuvant chemotherapy treatment in breast cancer. In: Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications. Vol. 11318. International Society for Optics and Photonics, (2020).
14. Eben, Jeffrey E., Nathaniel Braman, and Anant Madabhushi.: Response Estimation Through Spatially Oriented Neural Network and Texture Ensemble (RESONATE). In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, (2019).
15. Haarburger, Christoph, et al.: Multi scale curriculum CNN for context-aware breast MRI malignancy classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, (2019).
16. Ravichandran, Kavya, et al.: A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast DCE-MRI. In: Medical Imaging 2018: Computer-Aided Diagnosis. Vol. 10575. International Society for Optics and Photonics, (2018).
17. Ha, Richard, et al.: Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset. In: Journal of digital imaging 32.5 pp. 693-701 (2019).
18. He, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016).
19. TensorFlow Homepage, https://www.tensorflow.org. Last accessed 5 July 2020.
20. Klein, Jonathan, et al.: Locally advanced breast cancer treated with neoadjuvant chemotherapy and adjuvant radiotherapy: a retrospective cohort analysis. In: BMC Cancer 19, 306, (2019).