PROCEEDINGS OF SPIE

Determining the importance of parameters extracted from multi-parametric MRI in the early prediction of the response to neo-adjuvant chemotherapy in breast cancer

Tahmassebi, Amirhessam, Pinker-Domenig, Katja, Wengert, Georg, Helbich, Thomas, Bago-Horvath, Zsuzsanna, et al.

SPIE.

Determining the importance of parameters extracted from multi-parametric MRI in the early prediction of the response to neo-adjuvant chemotherapy in breast cancer

Amirhessam Tahmassebi[a,*], Katja Pinker-Domenig[a,b,c], Georg Wengert[c], Thomas Helbich[c], Zsuzsanna Bago-Horvath[d], and Anke Meyer-Baese[a]

[a]Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA
[b]Department of Radiology, Breast Imaging Service, Memorial Sloan Kettering Cancer Center, New York, USA
[c]Department of Biomedical Imaging and Image-Guided Therapy, Division of Molecular and Gender Imaging, Medical University of Vienna, Vienna, Austria
[d]Department of Pathology, Medical University Vienna, Vienna, Austria

ABSTRACT

Neo-adjuvant chemotherapy (NAC) is the treatment of choice in patients with locally advanced breast cancer to reduce tumor burden, and potentially enable breast conservation. Response to treatment is assessed by histopathology from surgical specimen, a pathological complete response (pCR), or a minimal residual disease are associated with an improved disease-free, and overall survival. Early identification of non-responders is crucial as these patients might require different, or more aggressive treatment. Multi-parametric magnetic resonance imaging (mpMRI) using different morphological and functional MRI parameters such as T2-weighted, dynamic contrast-enhanced (DCE) MRI, and diffusion weighted imaging (DWI) has emerged as the method of choice for the early response assessments to NAC. Although, mpMRI is superior to conventional mammography for predicting treatment response, and evaluating residual disease, yet there is still room for improvement. In the past decade, the field of medical imaging analysis has grown exponentially, with an increased numbers of pattern recognition tools, and an increase in data sizes. These advances have heralded the field of radiomics. Radiomics allows the high-throughput extraction of the quantitative features that result in the conversion of images into mineable data, and the subsequent analysis of the data for an improved decision support with response monitoring during NAC being no exception. In this paper, we determine the importance and ranking of the extracted parameters from mpMRI using T2-weighted, DCE, and DWI for prediction of pCR and patient outcomes with respect to metastases and disease-specific death.

Keywords: Breast cancer, Radiomics, Neo-adjuvant chemotherapy, Multi-parametric MRI, Machine Learning.

1. INTRODUCTION

In recent years, neo-adjuvant chemotherapy (NAC) is widely used in patients with locally advanced breast cancer (LABC) offering several advantages such as reduction of tumor, and enabling breast-conservation surgery instead of mastectomy as well as response-guided NAC approaches.[1–3] In patients undergoing neo-adjuvant chemotherapy for breast cancer, the achievement of a pathological complete response (pCR) is associated with a significantly improved disease-free, and overall survival.[4–7] However, a pCR is achieved in only 30% of the patients after the completion of NAC, and clinical studies have shown that the therapeutic outcome can be improved after the treatment modifications during the NAC. Therefore, accurate means to predict treatment response as early as possible are desirable to identify women who do not benefit from cytotoxic therapy. Several studies have

---

*Corresponding Author: Amirhessam Tahmassebi
E-mail: atahmassebi@fsu.edu
URL: www.amirhessam.com

demonstrated that dynamic contrast-enhanced (DCE) MRI is the most sensitive method for the assessment, and prediction of treatment response.[3,8,9] In addition, it has been demonstrated that multi-parametric MRI (mpMRI) using morphological as well as additional functional parameters such as diffusion-weighted imaging (DWI) has potential for an improved prediction of treatment response.

In the past decade, the field of medical image analysis has grown exponentially, with an increased numbers of pattern recognition[10] tools, and an increase in data set sizes. These advances have facilitated the development of processes for high-throughput extraction of quantitative features that result in the conversion of images into mineable data, and the subsequent analysis of these data for decision support. This emerging field in medical research is termed radiomics.[11] O'Flynn et al.[12] have recently shown that machine learning algorithms such as linear discriminant analysis along with statistical methods based on multi-parametric MRI features such as enhancement fraction, tumor volume, initial area under the gadolinium curve, and also maco-kinetic parameters such as K_{trans} and K_{ep} can be employed to predict patients in terms of responders and non-responders to NAC. However, it should be noted that they have employed only seven features along with one machine learning algorithm based on thirty-two patients.

The aim of this study was to assess radiomics with multi-parametric MRI using, T2-weighted, DCE with pharmako-kinetic modeling MRI, apparent diffusion coefficient (ADC) with DWI for the early prediction of pCR in breast cancer patients undergoing neo-adjuvant chemotherapy.[6]

2. MR IMAGING ACQUISITION & FEATURE EXTRACTION

In this IRB approved prospective study, forty-one women (median age 50 years; range 25-80 years) with biopsy-proven breast cancer scheduled for neo-adjuvant chemotherapy were included and underwent multi-parametric MRI with T2-weighted, DCE-MRI, and DWI study. Breast MRI was performed prior to and after two cycles of NAC, and pCR was assessed after surgery.
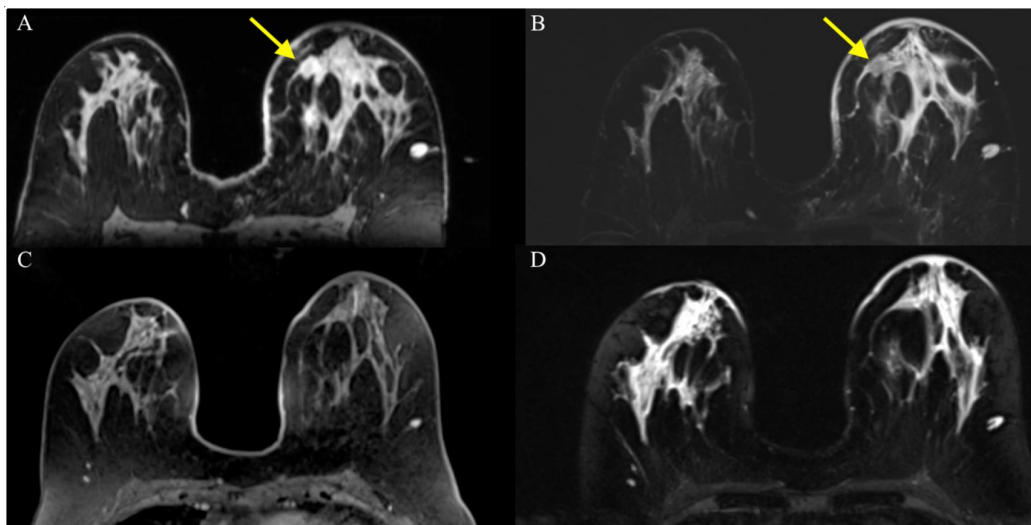

Figure 1. An illustration of complete imaging and pathological response after two cycles of neo-adjuvant chemotherapy.

Figure 1 illustrates a breast scan of a 65 years old patient with multi-centric breast cancer and index lesion left breast retroareolar medial (invasive ductal carcinoma grade 3, triple negative molecular subtype: estrogen receptor/ progesterone receptor/ human epidermal growth factor receptor 2 negative, ki67 90%): complete imaging and pathological response after two cycles of neo-adjuvant chemotherapy.

For each lesion a total number of 14 features were extracted ranging from morphological, quantitative kinetic, and ADC parameters. Table 1 presents the list of the features extracted in this study from the mpMRI including: T2-weighted: hyper-, hypo-, and isointense, presence of peritumoral edema, size: diameters in all three planes in millimeters, mass shape in terms of round, oval, and irregular, mass margins: circumscribed, irregular, and

Table 1. Features extracted from mpMRI using morphological, and functional imaging.

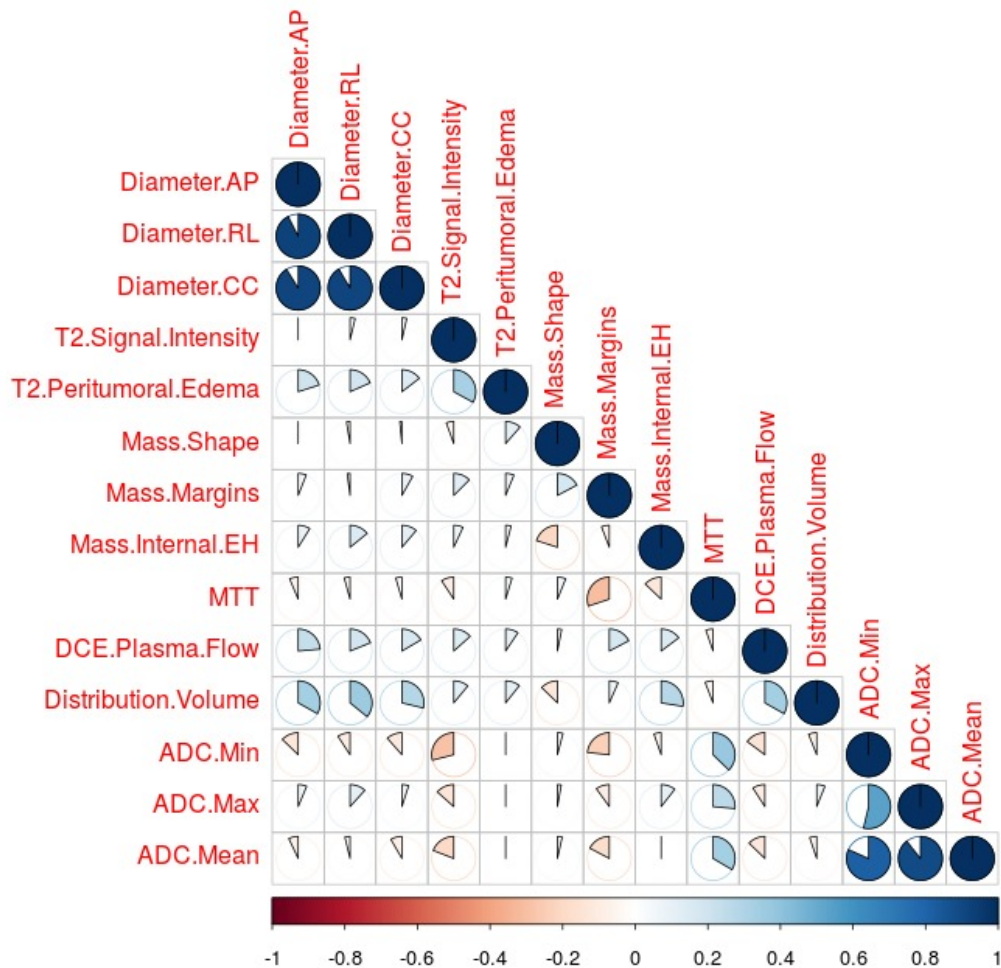|	
1. AP Diameter (mm)	8. Mass Internal EH
2. RL Diameter (mm)	9. MTT (Sec)
3. CC Diameter (mm)	10. DCE Plasma Flow (ml/min)
4. T2 Signal Intensity	11. Distribution Volume (ml/100ml)
5. T2 Peritumoral Edema	12. ADC (Min)
6. Mass Shape	13. ADC (Max)
7. Mass Margins	14. ADC (Mean)

Figure 2. Correlation matrix plot of the extracted features.

spiculated: mass internal enhancement characteristics: homogeneous, heterogeneous, rim enhancement, dark internal spatations: quantitative enhancement kinetics: mean transit time (MTT) in seconds, plasma flow (ml/min), distribution volume (ml/100 mls), and ADC min, max, and mean. Figure 2 depicts the correlation matrix of the extracted features with color bar which highlights the probability of the correlation of each of the extracted features with each other. It is clear that the diagonal has the probability of one. In this plot, to

illustrate the correlation matrix, the pie method was employed. Positive correlations are displayed in blue and negative correlations in red color. Color intensity and the size of the pie slices are proportional to the correlation coefficients. As shown, as the pie slice is getting bigger, it means the features are more correlated which can be both positive or negative. For example, the mean value of ADC is correlated with the minimum and maximum values of ADC which was expected. The other point which was expected is the correlation of the distribution volume with the diameters AP, RL, and CC.

3. MACHINE LEARNING METHODS

A machine learning method is aimed to construct a classification model that maps each feature set x to one of the class label y. In this paper, four linear classification algorithms including: 1- Linear Support Vector Machine (SVM),[13] 2- Linear Discriminant Analysis (LDA),[14] 3- Logistic Regression (LR),[15] 4- Stochastic Gradient Descent (SGD),[16] and an ensemble algorithm 5- Random Forests (RF)[17] were employed. The classifiers were applied to the extracted features set to predict three principal class labels. Figure 3 presents a schematic illustration of a linear, and an ensemble classifier.
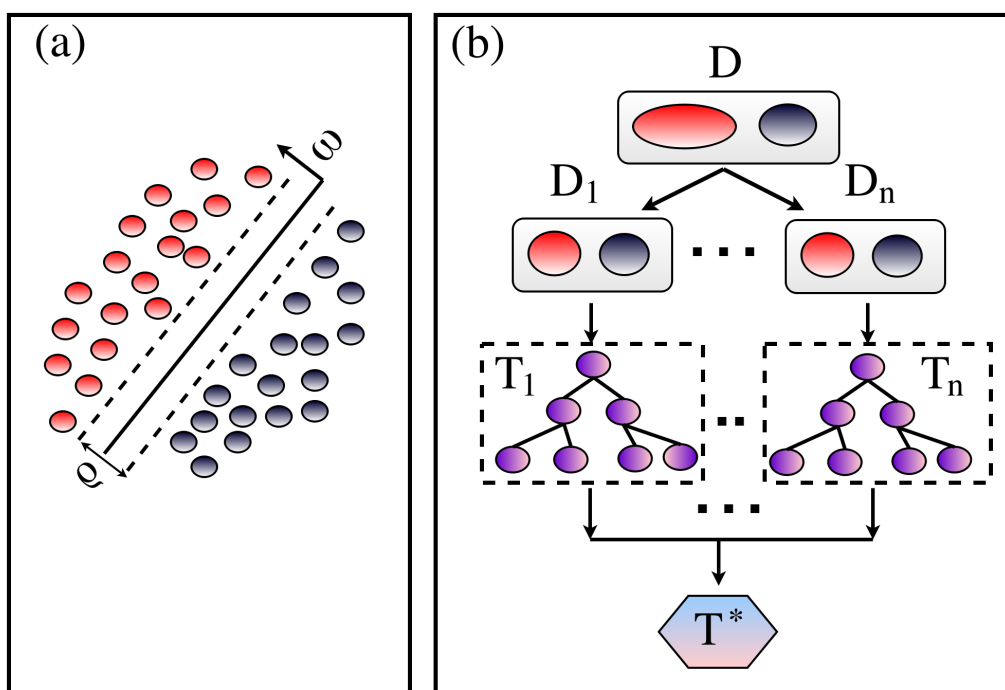

Figure 3. Schematic illustration of (a) a linear classifier, and (b) an ensemble classifier. (a) presents the classification of a binary data using linear support vector machine with maximum margin ρ via assigning a weight vector w to the data. However, (b) depicts random forest classifier which subsets D_i were randomly sampled from the data D and trees T_i were constructed, respectively. The final prediction would be based on the tree T^*.

The first class label in prediction was the residual cancer burden (RCB) score which describes the presence, and amount of residual cancer after NAC. In terms of RCB score, a patient could be classified as a complete response after treatment without no cancer cells, or be classified into different categories based on the number of cancer cells found after the treatment.[18]

In this study for the RCB score prediction, the patients were classified as complete or incomplete responses. The second class label was the recurrence free survival (RFS) classified by the development of the distant metastases. In this study, the patients also were classified into metastatic and non-metastatic categories. Finally, the third class label was disease-specific death (DSS). In this regard, the patients were classified into two classes based on the death and non-death outcomes after the chemotherapy.

Algorithm 1: Recursive Feature Elimination incorporating k-Folds Cross-Validation
___

Input: Data D, Number of folds k

Output: Optimum Number of Features ONF, Variable Ranking R, Cross-Validation Accuracy $\bar{Acc}$

1 Partition the data D into k subsets $\{D_1, \ldots, D_k\}$;

2 **for** $i \in \{1, \ldots, k\}$ **do**

3 $S_i \leftarrow D - D_i$;

4   Train the model with S_i data;

5   Test the model on D_i data;

6   Calculate the classification accuracy Acc_i for subset D_i;

7   Calculate the variable importance R_i;

8   Determine the optimum number of features ONF;

9 **end**

10 $\bar{Acc} \leftarrow \frac{\sum_{i=1}^{k} Acc_i}{k}$ Cross-Validation Accuracy;

11 Determine the optimum list of features incorporating their ranking;
___

Each method employs a specific learning algorithm to build the best model which fits the input data, and predicts the class labels correctly. SVM is a classifier that finds a hyperplane based on maximal margin rule to separate the data into two classes as shown in Figure 3(a). It could be possible to apply a linear decision boundary to nonlinear data conditions. The nonlinear data has to be transformed into a new linear space from its original coordinate space. Thus, in the new coordinate space, the linear decision boundaries could separate the sample data. LDA finds linear combination of features which separates two or more classes by determination a subspace of lower dimension of the original data. Statistical measures, such as variance and mean, are used to determine separability. LDA can also be derived from Bayesian rule by assigning a pattern with the maximal probability by comparing the posterior probability of all classes. In other words, it is designed to maximize the projected class means with minimization of the classes variance in that direction by fitting a Gaussian density to each class having the assumption that all classes share the same covariance matrix.[19]

LR models the conditional probability using logistic function based on the odds, and the values of the independent variables. LR tries to minimize the negative log-likelihood of conditional probability via optimization algorithms. It is designed to find cumulative logistic distribution by measuring the relationship between one dependent, and one or more independent variables. RF method is based on the concept of decision trees as shown in Figure 3(b). A decision tree represents patterns, and structures in the input data with hierarchical, and sequential nodes that form tree-like structures.[17] RF constructs many decision trees from a dataset, and combines the results from all the trees to make predictions on classification or regression. SGD as a method for unconstrained optimization problems, tries to learn a linear scoring by assigning weights, and intercept parameter. The parameters would be found using minimizing the loss function. Employing SGD gives us more degrees of freedom to choose different loss functions. Different loss function would lead us to different classifiers.[16]

In this study, recursive feature elimination method as presented in Algorithm 1 along with five different classifiers was employed to find the optimum ranking of the features extracted from mpMRI. By employing recursive feature elimination, we can select features by recursively considering smaller and smaller sets of features by training a classifier on the initial set of features, and weights.[20] Then, features whose absolute weights are the minimum are pruned from the current set features. By repeating this procedure the desired optimum number of features with the maximum accuracy would be found. It should be noted that, the number of individuals in each class was not equal. In better words, we were dealing with biased classification tasks. In this regard, 10-folds cross-validations (CV) was employed to decrease the possibility of over-fitting the models. For each classifier, the optimum number of features (ONF) is also reported. It should be noted that all the computing codes, and analysis were written in `R`, and `Python` programming languages.[20]

4. RESULTS & DISCUSSION

As discussed, the recursive feature elimination was applied for three different classes including: RCB score, RFS, and DSS. In addition, the optimum number of features (ONF) for each class along with each algorithm was also reported.
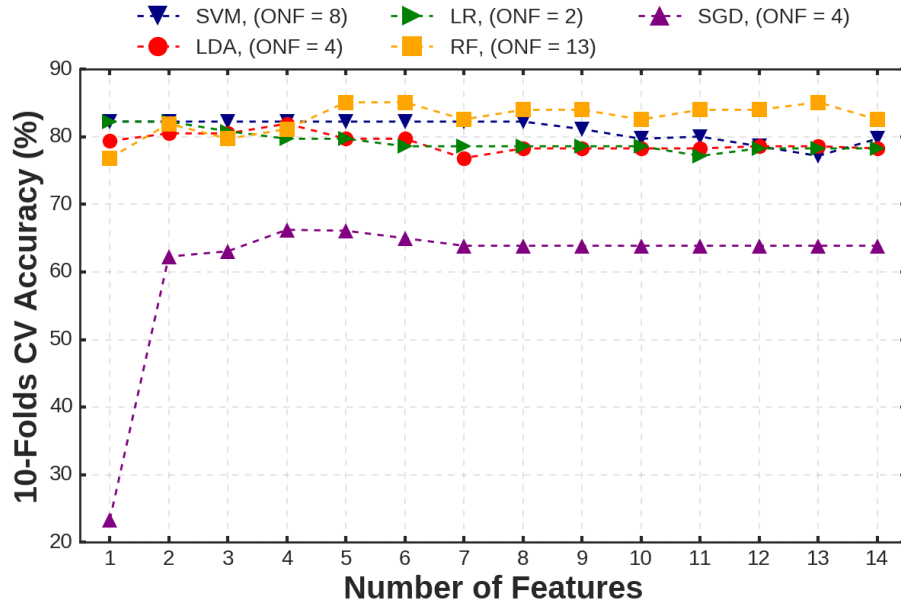
Figure 4. Recursive feature elimination along with 10-folds cross-validation incorporating different classifiers to predict RCB score.
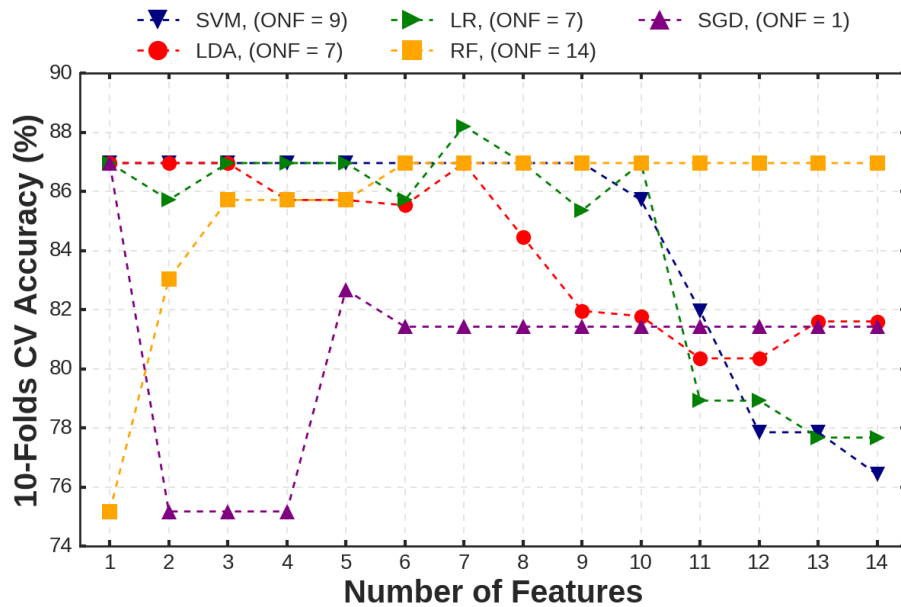

Figure 5. Recursive feature elimination along with 10-folds cross-validation incorporating different classifiers to predict RFS.

Figure 4 presents the performance of the employed classifiers using recursive feature elimination algorithm in terms of RCB score prediction. It should be noted that due to the unbalanced number of individuals in each class, random predictions might happen. As seen in Figure 4 LR showed a reasonable prediction accuracy by employing only two features in the training model. RF due to its ensemble nature has employed 13 features, and
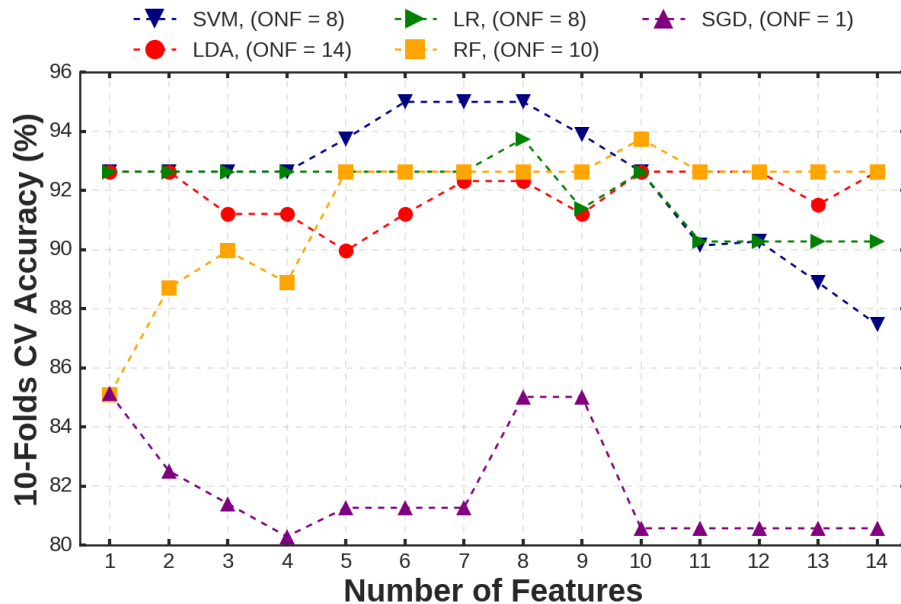
Figure 6. Recursive feature elimination along with 10-folds cross-validation incorporating different classifiers to predict DSS.

showed a similar performance in terms of RCB score like SVM with 8 and LDA with 4 features. The performance of SGD is quite interesting where the number of feature increased, the classification reached a stable point after incorporating 4 features. The optimization process of SGD is clearly presented in Figure 4.

Figure 5 depicts the RFS prediction using different classifiers along with recursive feature elimination incorporating 10-folds cross-validation. As shown, the performance of SVM, LDA, and LR decreased through the number of features. On the other hand, RF and SGD showed better performance as the number of features increased. The interesting point is the turning point, where SGD, and RF started predicting with higher accuracy, while SVM, LDA, and LR started predicting with lower accuracy. This might be due to the over-fitting of the models. These features (the best five) can be employed as the selected features group to be employed in prediction of all the three classes RCB, RFS, and DSS to help radiologists to consider these features at the beginning.

In addition to this, the prediction of DSS is shown in Figure 6. Based on the recursive feature elimination along with 10-folds cross-validation accuracy, LR and SVM showed the highest classification accuracy. Similar to RFS prediction, all the classification algorithms could predict DSS with a reasonable accuracy via employing the best five features.

As discussed, we took a deeper look into the details of the features ranking for each of the prediction classes. Then, five features which are most common in all three classes were chosen. These features are including 1-mass internal enhancement characteristics, 2- mass shape, 3- mass margins, 4- tissue edema, and 5- T2 signal intensity. The other question should be addressed here is that what classifier can be employed along with the new features set? In this regard, we have employed five classifiers using 10-folds cross-validation along with the new features set to predict RCB score, RFS, and DSS in all patients. Figure 7 illustrates the performance of the classifiers as explained. As shown, based on the new features set, we have also reached a ranking based on the classifiers performance. As expected, SGD has the lowest performance with the largest deviation in prediction of all three classes, and Linear-SVM showed the best performance in all three classification tasks. By referring to Figures 4, 5, and 6, it can be concluded that Linear-SVM has always employed about 8 or 9 features in the classification tasks which is about 60% of the information of the features set. In this task, we have decreased this share to even 35% of the information of the original feature set with 14 mpMRI features by including only the best 5 features, and Linear-SVM outperformed the other classifiers. In this study, as we are dealing with
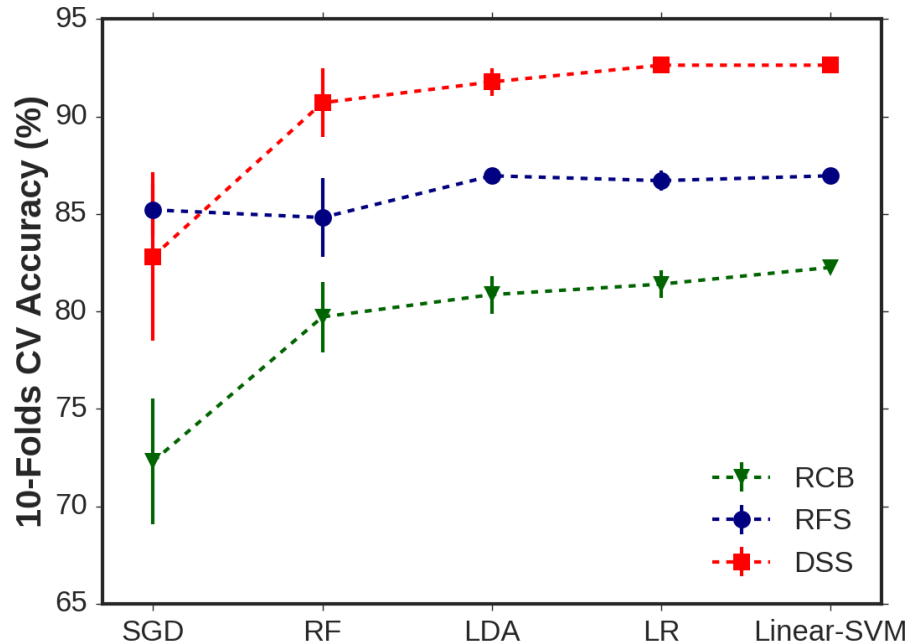
Figure 7. The performance of the different classifiers in terms of 10-folds cross-validation accuracy with error-bars for predicting RCB score, RFS, and DSS.

unbalanced classes in terms of the number of individuals in each class, employing the most stable classifier in performance is vital. This choice along with incorporating 10-folds cross-validation would decrease the chance of over-fitting drastically.

5. CONCLUSIONS

Early identification of non-responders is crucial for therapeutic decision making since these patients require an aggressive treatment and ineffective and toxic chemotherapy agents should be avoided. Predicting the pathological response after NAC in breast cancer patients is crucial and quantitative computerized methods represent an important step towards an accurate and effective breast cancer treatment. In our study, based on the prediction results from a total of fourteen extracted mpMRI parameters the following conclusions are drawn:

1. It is shown with a very high accuracy that mass internal enhancement, mass shape, mass margins, tissue edema, and T2 signal intensity are the most important parameters for all the three RCB score, RFS, and DSS predictions.

2. Linear-SVM has shown the best and most stable performance and outperformed the other linear classifiers including LR, LDA, SGD, and even RF with an ensemble-nature in prediction of all three classes.

3. It is shown that with even choosing the best five features based on the recursive feature elimination we can still predict all three classes with a very reasonable accuracy and also low deviation from the mean values.

4. It is shown that the recursive feature elimination along with k-folds cross-validation can be a useful method to be employed in radiomics with mpMRI using T2-weighted, DCE-MRI, and DWI for the early prediction of pCR in breast cancer patients undergoing NAC.

REFERENCES

[1] Kaufmann, M., Von Minckwitz, G., Bear, H., Buzdar, A., McGale, P., Bonnefoi, H., Colleoni, M., Denkert, C., Eiermann, W., Jackesz, R., et al., "Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006," *Annals of Oncology* **18**(12), 1927–1934 (2007).

[2] Abramson, R. G., Li, X., Hoyt, T. L., Su, P.-F., Arlinghaus, L. R., Wilson, K. J., Abramson, V. G., Chakravarthy, A. B., and Yankeelov, T. E., "Early assessment of breast cancer response to neoadjuvant chemotherapy by semi-quantitative analysis of high-temporal resolution dce-mri: preliminary results," *Magnetic resonance imaging* **31**(9), 1457–1464 (2013).

[3] Arlinghaus, L. R., Li, X., Levy, M., Smith, D., Welch, E. B., Gore, J. C., and Yankeelov, T. E., "Current and future trends in magnetic resonance imaging assessments of the response of breast tumors to neoadjuvant chemotherapy," *Journal of oncology* **2010** (2010).

[4] Lindenberg, M. A., Miquel-Cases, A., Retèl, V. P., Sonke, G. S., Wesseling, J., Stokkel, M. P., and van Harten, W. H., "Imaging performance in guiding response to neoadjuvant therapy according to breast cancer subtypes: A systematic literature review," *Critical Reviews in Oncology/Hematology* **112**, 198–207 (2017).

[5] Liu, Q., Wang, C., Li, P., Liu, J., Huang, G., and Song, S., "The role of 18f-fdg pet/ct and mri in assessing pathological complete response to neoadjuvant chemotherapy in patients with breast cancer: a systematic review and meta-analysis," *BioMed research international* **2016** (2016).

[6] Zöllner, F. G., Weisser, G., Reich, M., Kaiser, S., Schoenberg, S. O., Sourbron, S. P., and Schad, L. R., "Ummperfusion: an open source software tool towards quantitative mri perfusion analysis in clinical routine," *Journal of digital imaging* **26**(2), 344–352 (2013).

[7] Rauch, G. M., Adrada, B. E., Kuerer, H. M., van la Parra, R. F., Leung, J. W., and Yang, W. T., "Multimodality imaging for evaluating response to neoadjuvant chemotherapy in breast cancer," *American Journal of Roentgenology* **208**(2), 290–299 (2017).

[8] Wu, L.-A., Chang, R.-F., Huang, C.-S., Lu, Y.-S., Chen, H.-H., Chen, J.-Y., and Chang, Y.-C., "Evaluation of the treatment response to neoadjuvant chemotherapy in locally advanced breast cancer using combined magnetic resonance vascular maps and apparent diffusion coefficient," *Journal of Magnetic Resonance Imaging* **42**(5), 1407–1420 (2015).

[9] Minarikova, L., Bogner, W., Pinker, K., Valkovič, L., Zaric, O., Bago-Horvath, Z., Bartsch, R., Helbich, T. H., Trattnig, S., and Gruber, S., "Investigating the prediction value of multiparametric magnetic resonance imaging at 3 t in response to neoadjuvant chemotherapy in breast cancer," *European radiology* **27**(5), 1901–1911 (2017).

[10] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern classification*], John Wiley & Sons (2012).

[11] Gillies, R. J., Kinahan, P. E., and Hricak, H., "Radiomics: images are more than pictures, they are data," *Radiology* **278**(2), 563–577 (2015).

[12] AM OFlynn, E., Collins, D., DArcy, J., Schmidt, M., and de Souza, N. M., "Multi-parametric mri in the early prediction of response to neo-adjuvant chemotherapy in breast cancer: Value of non-modelled parameters," *European journal of radiology* **85**(4), 837–842 (2016).

[13] Cortes, C. and Vapnik, V., "Support-vector networks," *Machine learning* **20**(3), 273–297 (1995).

[14] Fisher, R. A., "The use of multiple measurements in taxonomic problems," *Annals of human genetics* **7**(2), 179–188 (1936).

[15] Cox, D. R., "The regression analysis of binary sequences," *Journal of the Royal Statistical Society. Series B (Methodological)* , 215–242 (1958).

[16] Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al., "Learning representations by back-propagating errors," *Cognitive modeling* **5**(3), 1 (1988).

[17] Breiman, L., "Random forests," *Machine learning* **45**(1), 5–32 (2001).

[18] Symmans, W. F., Wei, C., Gould, R., Yu, X., Zhang, Y., Liu, M., Walls, A., Bousamra, A., Ramineni, M., Sinn, B., et al., "Long-term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype," *Journal of Clinical Oncology* **35**(10), 1049–1060 (2017).

[19] Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K.-R., "Fisher discriminant analysis with kernels," in [*Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*], 41–48, IEEE (1999).

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research* **12**(Oct), 2825–2830 (2011).