

The Importance of Temporal Integration in DCE-MRI for Improved Breast Cancer Diagnosis

Francesco Prinzi^{1,2†*}, Francesco Ceccarelli^{2†}, Sean Holden², Pietro Liò², and Salvatore Vitabile¹

¹ Department of Biomedicine, Neuroscience and Advanced Diagnostics (BiND), University of Palermo, Palermo, Italy

² Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

Abstract. Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is a sequence of MRI scans acquired after the administration of a contrast agent. In Breast cancer, it is used to capture dynamic changes in tissue enhancement over time, which is shown to be different between benign and malignant lesions. In fact, the DCE-MRI is a temporal sequence of MRI composed of one pre-contrast and six post-contrast instant. This study leverages machine learning techniques to enhance breast cancer classification using the full sequence of DCE-MRI data, addressing the common oversight of underutilizing its temporal dimension. We analyze an in-house dataset, integrating radiomic features from each time instant through *i*) feature concatenated from all instants of the sequence and a random forest model (multi-instant Random Forest), and *ii*) a graph neural network (GNN) to extract informative embeddings in which nodes correspond to the seven time instants of the DCE-MRI sequence. Our findings indicate that incorporating temporal information significantly improves classification performance, both in terms of accuracy and particularly in terms of Positive Predictive Value (PPV), which is crucial for reducing false positives in clinical decisions. Despite the complexity of GNNs, their performance gains are marginal compared to the simpler multi-instant RF, suggesting that shallower models may be equally effective with smaller datasets. Explainable AI methods further reveal that the pre-contrast and third post-contrast instants are most informative for classification, offering new insights for radiology physicians.

Keywords: Breast Cancer · Radiomics · Graph Neural Networks · Explainable AI.

1 Introduction

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) is an advanced imaging technique for medical diagnostics. Unlike conventional MRI,

* Corresponding author. Email: francesco.prinzi@unipa.it

† These authors contributed equally to this work

DCE-MRI involves the intravenous injection of a contrast agent to enhance the visibility of vascular perfusion and permeability within the target tissue. As a result, DCE-MRI captures dynamic changes in tissue enhancement over time, providing invaluable insights into tissue microvasculature and permeability characteristics [3]. In the case of breast cancer, the DCE-MRI sequence is composed of 1 pre-contrast MRI followed by 6 post-contrast MRI, and its signal intensity time curves have shown that malignant tissues generally enhance early compared to benign tissues [9]. Considering this marked difference between benign and malignant lesions, the entire sequence rather than individual instants has to be evaluated for an enhanced diagnosis.

Machine learning methodologies were employed for breast cancer classification across diverse contexts [7], and notable advancements have also emerged within the domain of DCE-MRI. Despite the physician diagnostic process typically relying on classification based on contrast medium absorption trends, many studies fail to fully leverage the temporal dimension of a DCE-MRI sequence [6, 2, 12], undermining the effectiveness of the classification approaches. For this reason, new recent methods were proposed to exploit all the DCE-MRI instants, proposing time series classifier methods for classification [13] and a local-global cross-attention fusion network for classification and segmentation [17]. Although these approaches have focused on improving the models’ accuracy, they have overlooked the study of how each instant affects the final prediction and whether there is an instant that contributes most to the prediction. This step is critical to validate the models, analyze the clinical implications, and provide new insights to radiology physicians. This aspect necessitates methodologies that incorporate data from all time instants and employ *post-hoc* model explanations to assess the significance of these instants.

Our work aims to analyze an in-house dataset for breast cancer classification combining all the sequence instants within the DCE-MRI sequence. Each DCE-MRI is composed of one pre-contrast image and six post-contrast. Radiomic features [5, 12] are extracted at each time instants of the sequence and are integrated through *i*) a features concatenation; and *ii*) a graph neural network, implemented to extract informative embeddings by assimilating information from all nodes, here represented by the time instants of the DCE-MRI sequence. To highlight the importance of temporal integration in DCE-MRI, these models are compared with a Random Forest trained on the pre-contrast instant and third post-contrast instant because proven the most informative [12]. In addition, Explainable AI algorithms are developed to evaluate the most informative instants for prediction when the instants are integrated, providing new insights to radiology physicians.

2 Materials and Methods

This work aims to evaluate the importance of aggregating all instants of the DCE-MRI sequence for classification. Radiomic feature extraction is implemented to convert the lesions of each instant into informative signatures. The pre-

contrast and third post-contrast instants, found to be the most predictive in Prinzi et al. [12], are compared with two implemented approaches that integrate information from multiple instants:

- A Random Forest model trained by concatenating the features extracted from all instants of the sequence.
- A Fully-Connected Graph Neural Network in which a node is represented by each instant of the DCE-MRI sequence and the features vectors are represented by the radiomic features of each instant.

The following sections outline all the steps in the workflow, including the employed dataset and the feature extraction process, feature preprocessing and selection, and model training.

2.1 Dataset Description and Radiomic Feature extraction

Our in-house dataset was collected using a 1.5T MR scanner and consists of T1-weighted sequences. The process of a DCE-MRI involves first obtaining a pre-contrast MRI (one without the contrast agent) and then a set of MRI (six in our case) obtained subsequent to the endovenous injection of the contrast agent. Each instant is acquired every 70/90 seconds. A total of 166 breast lesions were collected and manually segmented by expert radiologists. The dataset is balanced, with 84 malignant lesions and 82 benign. In addition, our dataset is multi-protocol with 81 and 85 lesions for protocols 1 and 2, respectively.

For each lesion, radiomic features were extracted using the pyradiomics [18], compliant with the Imaging Biomarker Standardization Initiative (IBSI) [4]. A total of 93 features were extracted from un-processed images, considering statistical and textural features. Using the Haar kernel, the same features were extracted from the wavelet transform and from Laplacian of Gaussian (LoG) filtered images ($\sigma = [2, 3]$), collecting 744 wavelet-derived and 186 LoG-derived features, respectively. Figure 1 mentions the extracted categories.

2.2 Feature preprocessing and Selection

Considering the two acquisition protocols employed for image acquisition, ComBat was used for feature harmonization [8]. Features with variance $\sigma^2 < 0.01$ were removed, as well as highly correlated features using a Spearman correlation coefficient ($|\rho| > 0.9$ as the threshold). These two steps were repeated for each instant within the DCE-MRI sequence. Finally, only the features statistically significant simultaneously in at least 5 instants were selected. The Mann-Whitney U test was employed as a statistical test, considering $p < 0.05$ as significance level.

2.3 Classification using Shallow methods

Each instant of the DCE-MRI sequence was previously evaluated showing that each instant, analyzed separately, can provide high diagnostic performance [12].

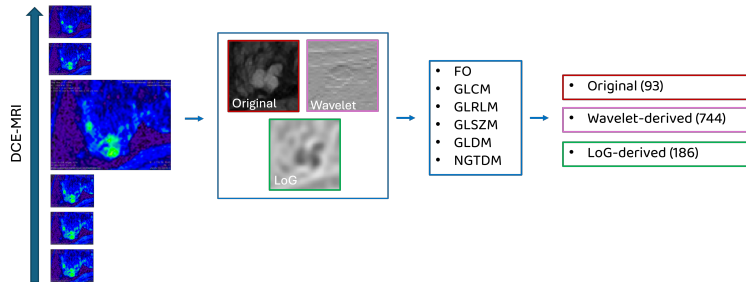


Fig. 1. Workflow for radiomic feature extraction. From each instant original, wavelet-derived and log-derived features were extracted. (GLCM: Gray Level Co-occurrence Matrix; FO: first-order; GLRLM: Gray Level Run Length Matrix; GLSZM: Gray Level Size Zone Matrix; GLSM: Gray Level Dependence Matrix; NGTDM: Neighbouring Gray Tone Difference Matrix;)

Random Forest emerged as the most performing model across all instants. In addition, the pre-contrast instant and the third post-contrast instant showed higher predictive capability. In this work, to integrate information from all instances, a new Random Forest model was trained concatenating all selected features coming from each instance. We refer to this model as the multi-instant Random Forest. Before the concatenation process, correlated features were again discarded, considering Spearman’s correlation index < 0.8 . In addition, to maintain a proper proportion between input samples and dataset dimensionality [10], the sequential feature selector (SFS) [14] was used to select an informative signature. SFS was set with Random Forest as the classifier, and accuracy as the metric to maximize.

2.4 Classification using a Fully-Connected Graph Neural Network

A graph neural network was implemented to integrate all instances of the DCE-MRI sequence. Specifically, each instant represents the node of the graph, and the selected radiomic features represent the features of each node. Formally, a graph $G = (V, E)$ is a structure consisting of a set V of n nodes and a set of edges E . Each node $v \in V$ is equipped with a d -dimensional feature vector \mathbf{x}_v , here v are the seven DCE-MRI instants, \mathbf{x}_v are the radiomic features associated with each instant. These can be grouped into a feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ by stacking all the $n = |V|$ feature vectors vertically. Its adjacency matrix \mathbf{A} , whose entry i, j is equal to 1 if node i is connected to node j and 0 otherwise, fully captures the connectivity structure of G . Each layer of a GNN is made up of a nonlinear function that maps a feature matrix into a new, hidden feature matrix. This allows the connectivity of the underlying network to account for the pairwise relationships that are present in it. Formally,

$$\mathbf{H}^{(l)} = f(\mathbf{H}^{(l-1)}; \mathbf{A}) \quad (1)$$

where $\mathbf{H}^{(l)}$ is the hidden feature matrix at layer l and $\mathbf{H}^{(0)} = \mathbf{X}$. A single graph represents a breast tumor consisting of seven nodes. A fully connected connectivity was implemented hypothesizing that important relations exist between non-consecutive time steps. This construction strategy resulted in a dataset composed of 166 graphs with seven nodes each. After applying GNN layers, the representations learned for each node were then aggregated using mean pooling as readout function to obtain graph-level embeddings. Finally, these embeddings were used for binary classification. Graph Attention Networks (GATs) [15] was employed as architecture.

3 Experimental Results

The trained models were evaluated using a stratified 10-fold cross-validation. Accuracy, AUROC, specificity, sensitivity, PPV, and NPV were considered as metrics. Random forest was trained using 100 estimators. Experiments provided in [12] for the pre-contrast and the third post-contrast instants were performed again using the same pipeline and evaluation protocol.

3.1 Performance

Forty-four features were selected after the feature selection described in 2.2. In particular, our strategy guarantees that in at least 5 out of 7 DCE-MRI instants features are informative, uncorrelated, and statistically significant. In addition, 6 and 8 features were selected via SFS for the training signature on the pre-contrast and third post-contrast instants, respectively. Seven features were selected via SFS when the features of all instants have been concatenated (multi-instant Random Forest). All 44 features were employed to train the GNN. Table 1 presents the performance metrics of different models, including individual instants as well as combined approaches (Multi-instant Random Forest and GNN).

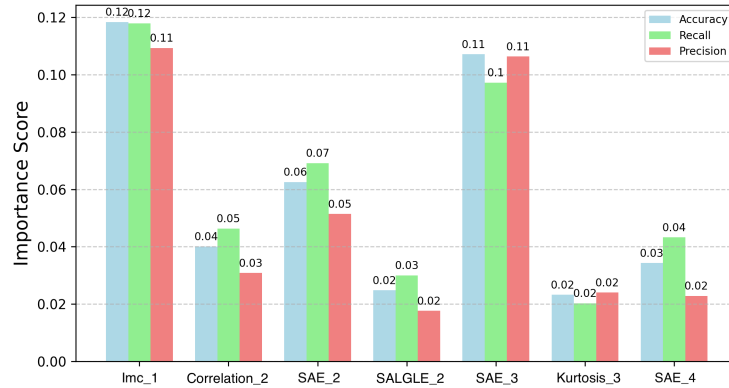
When considering only the features from instant 1 (pre-contrast) and instant 4 (third post-contrast) independently, both achieved comparable performance, with RF trained on instant 1 slightly outperforming RF on instant 4. This observation suggests that these individual time instants contain valuable information for classification. However, the real breakthrough emerges when the information across all time instants are combined. The multi-instant RF model trained using all the instant employing a feature concatenation exhibits a notable boost in performance across all metrics compared to single-instant models. This proves the significance of leveraging temporal context and sequential dependencies within the data to improve classification accuracy. The performance of the GNN model, which incorporates a fully connected graph neural network architecture, closely parallels that of the Multi-instant RF model. In terms of performance, a gap between single instant RF and models that incorporate all the instants is highlighted. This is particularly evident when analyzing PPV metrics, in which the multi-instant RF model exhibited a PPV of 0.775 ± 0.125 , indicating its ability to accurately identify malignant lesions while the GNN model shows a PPV of 0.797 ± 0.146 .

Table 1. Calculated performance. Inst 1 and 4 are RF trained on pre-contrast and post-contrast instants. Multi is the multi-instant RF. GAT is the fully connected GNN.

Model	Accuracy	AUROC	Sensitivity	Specificity	PPV	NPV
Inst 1	0.710 ± 0.107	0.774 ± 0.109	0.720 ± 0.157	0.699 ± 0.168	0.721 ± 0.129	0.721 ± 0.123
Inst 4	0.700 ± 0.108	0.747 ± 0.108	0.694 ± 0.148	0.706 ± 0.173	0.725 ± 0.130	0.699 ± 0.126
Multi	0.742 ± 0.095	0.782 ± 0.113	0.722 ± 0.134	0.763 ± 0.159	0.775 ± 0.125	0.734 ± 0.107
GAT	0.742 ± 0.132	0.788 ± 0.089	0.686 ± 0.178	0.797 ± 0.181	0.797 ± 0.146	0.723 ± 0.147

3.2 Explaining predictivity of DCE-MRI sequence

Explaining the Multi-Instant RF using Permutation Importance Permutation importance was implemented to explain the multi-instant Random Forest [1]. This method consists of a baseline model evaluation, then a feature column is permuted and the metrics are computed again to evaluate the impact of that feature. Accuracy, precision, and recall were considered. Figure 2 shows the importance of each feature contributing to the radiomic signature. The plot shows the main importance of wavelet-HLH_glm_Imc2 coming from the first instant, but overall, all features contribute to the model.

**Fig. 2.** Feature importance computed using Permutation Importance for the multi-instant RF. All the features are wavelet-derived and the last number represents the time-instant from which the feature comes from. Imc: Informational Measure of Correlation; SAE: Small Area Emphasis; SALGLE: Small Area Low Gray Level Emphasis;)

Explaining the FC-GNN using Occlusion Sensitivity Occlusion sensitivity was implemented to evaluate the contribution of each node within the graph. This technique, which has been widely used in image analysis [16], is based on assessing how much a model’s output changes when specific parts of an image

are occluded. The fundamental assumption is that the greater the change in the output caused by occluding a particular area, the more influential and important that area is for the task. Similarly, in the context of GNNs, the objective is to determine the importance of each node and, by extension, each time instant in the DCE-MRI sequence. Figure 3 illustrates the variation in error corresponding to the exclusion of each node. A higher error relative to the model using all nodes (indicated by the red dashed line) signifies a greater influence of that node on classification accuracy. The pre-contrast and third post-contrast instants emerge as the most important. This finding aligns with the instant-wise analysis presented in [12].

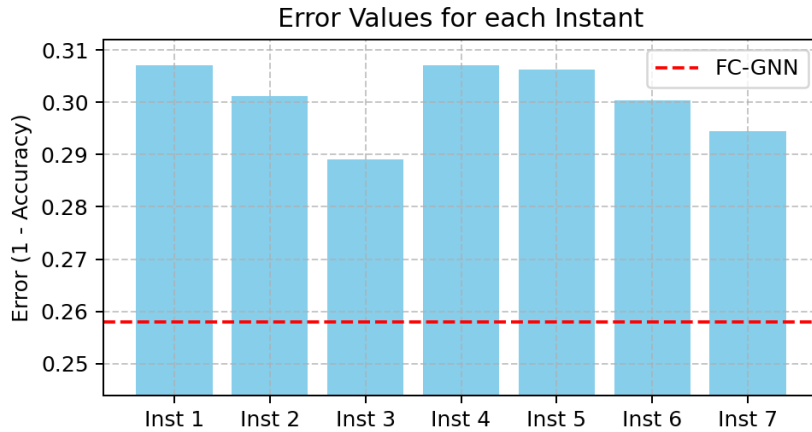


Fig. 3. Instant importance computed via occlusion sensitivity.

4 Discussion and Conclusions

While classification using individual instants yields respectable results, the integration of temporal information through concatenation or employing deeper architectures such as GNNs leads to substantial performance improvements. This result is particularly evident when the PPV metric is analyzed. The high PPV achieved by the multi-instant RF and GNN models in our study underscores the importance of reliable malignancy predictions in breast cancer classification. A high PPV signifies a low rate of false positives, which is critical in clinical decision-making processes, particularly when determining whether further invasive procedures such as biopsies are warranted. Our findings suggest that these models hold promise in assisting healthcare professionals in accurately identifying malignant lesions, thereby optimizing patient care and resource allocation. Despite the added complexity of GNNs, their performance gains are not significantly superior to the simpler multi-instant RF in this scenario. This observation

suggests that while deep learning models like GNNs offer powerful capabilities to capture intricate patterns within the data, their advantage over shallower models may not be pronounced when dealing with relatively smaller datasets [11].

The explanation of GNN and multi-instant RF models underscored the significance of integrating all temporal instances for enhanced breast cancer classification. Moreover, it was supported that the pre-contrast instant, along with the third post-contrast instant, are the most informative for classification [12]. The third post-contrast instant result is particularly intuitive, as by this instant, both benign and malignant lesions typically exhibit full absorption of the contrast agent. Regarding the pre-contrast instance, the lesion's contours and the lesion itself have not yet been highlighted by the contrast agent, thus preserving all original information, with unaltered and undistorted graylevels.

In addition to presenting an enhanced model for breast cancer classification in DCE-MRI, the results derived from explainable AI methods have yielded valuable insights for radiology clinicians.

Acknowledgement

Funding This research was co-funded by the Italian Complementary National Plan PNC-I.1 "Research initiatives for innovative technologies and pathways in the health and welfare sector" D.D. 931 of 06/06/2022, "DARE - Digital life-long pRevEntion" initiative, code PNC0000002, CUP: B53C22006460001; and Finanziato dall'Unione europea – Next Generation EU - Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) 2022, Prot. 2022ENK9LS. Project: "EX-EGETE: Explainable Generative Deep Learning Methods for Medical Image and Signal Processing", CUP: B53D23013040006.

References

1. Altmann, A., Tološi, L., Sander, O., Lengauer, T.: Permutation importance: a corrected feature importance measure. *Bioinformatics* **26**(10), 1340–1347 (2010)
2. Dong, H., Kang, L., Cheng, S., Zhang, R.: Diagnostic performance of dynamic contrast-enhanced magnetic resonance imaging for breast cancer detection: an update meta-analysis. *Thoracic cancer* **12**(23), 3201–3207 (2021)
3. Gordon, Y., Partovi, S., Müller-Eschner, M., Amarteifio, E., Bäuerle, T., Weber, M.A., Kauczor, H.U., Rengier, F.: Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. *Cardiovascular diagnosis and therapy* **4**(2), 147 (2014)
4. van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S., Aerts, H.J.: Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* **77**(21), e104–e107 (10 2017). <https://doi.org/10.1158/0008-5472.CAN-17-0339>
5. Lambin, P., Leijenaar, R.T., Deist, T.M., Peerlings, J., De Jong, E.E., Van Timmeren, J., Sanduleanu, S., Larue, R.T., Even, A.J., Jochems, A., et al.: Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology* **14**(12), 749–762 (2017)

6. Militello, C., Rundo, L., Dimarco, M., Orlando, A., Woitek, R., D'Angelo, I., Russo, G., Bartolotta, T.V.: 3d dce-mri radiomic analysis for malignant lesion prediction in breast cancer patients. *Academic Radiology* **29**(6), 830–840 (2022)
7. Murtaza, G., Shuib, L., Abdul Wahab, A.W., Mujtaba, G., Mujtaba, G., Nweke, H.F., Al-garadi, M.A., Zulficar, F., Raza, G., Azmi, N.A.: Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review* **53**, 1655–1720 (2020)
8. Orlhac, F., Eertink, J.J., Cottureau, A.S., Zijlstra, J.M., Thieblemont, C., Meignan, M., Boellaard, R., Buvat, I.: A guide to combat harmonization of imaging biomarkers in multicenter studies. *Journal of Nuclear Medicine* **63**(2), 172–179 (2022). <https://doi.org/10.2967/jnumed.121.262464>
9. Padhani, A.R.: Dynamic contrast-enhanced mri in clinical oncology: current status and future directions. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **16**(4), 407–422 (2002)
10. Papanikolaou, N., Matos, C., Koh, D.M.: How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* **20**(1), 1–10 (2020). <https://doi.org/10.1186/s40644-020-00311-4>
11. Prinzi, F., Currier, T., Gaglio, S., Vitabile, S.: Shallow and deep learning classifiers in medical image analysis. *European Radiology Experimental* **8**(1), 26 (2024)
12. Prinzi, F., Orlando, A., Gaglio, S., Midiri, M., Vitabile, S.: ML-based radiomics analysis for breast cancer classification in dce-mri. In: *International Conference on Applied Intelligence and Informatics*. pp. 144–158. Springer (2022)
13. Prinzi, F., Orlando, A., Gaglio, S., Vitabile, S.: Breast cancer classification through multivariate radiomic time series analysis in dce-mri sequences. *Expert Systems with Applications* **249**, 123557 (2024)
14. Raschka, S.: Mlxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack. *Journal of Open Source Software* **3**(24), 638 (2018). <https://doi.org/10.21105/joss.00638>
15. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. *stat* **1050**(20), 10–48550 (2017)
16. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I* 13. pp. 818–833. Springer (2014)
17. Zhao, X., Liao, Y., Xie, J., He, X., Zhang, S., Wang, G., Fang, J., Lu, H., Yu, J.: Breastdm: A dce-mri dataset for breast tumor image segmentation and classification. *Computers in Biology and Medicine* **164**, 107255 (2023)
18. Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R.J., Boellaard, R., et al.: The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**(2), 328 (2020). <https://doi.org/10.1148/radiol.2020191145>