MUGI-MRI: Enhancing Breast Cancer Classification through Multiplex Graph Neural Networks in DCE-MRI

Anonymous Authors

***Abstract*—Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) involves acquiring a sequence of MRIs during the administration of a contrast agent. Radiologists then aim to discern the contrast uptake differences between malignant and benign lesions for tumor classification. Regrettably, existing literature underutilizes the temporal structure inherent to DCE-MRI time series, leading to tumor classifications based on individual instants rather than entire sequences. This research introduces two Graph Neural Network (GNN)-based methods designed to aggregate information from multiple instants within the DCE-MRI sequence. Each lesion undergoes manual segmentation, and radiomic features are individually extracted from each time instant of the DCE-MRI sequence. Two graph construction methodologies are proposed: *(i)* a fully connected graph topology, aiming to represent each temporal instant as a node in a graph; *(ii)* a multiplex network, named MUGI-MRI (MUltiplex Graph neural network for Integration of MRI), where each layer identifies an instant of the DCE-MRI sequence. MUGI-MRI achieves an AUROC of 0.8017 $\pm$ 0.1146, showcasing promising performance in lesion classification. In addition to improving upon current state-of-the-art, the integration capability of MUGI-MRI addresses the problem of imbalance between sensitivity and specificity, which affects numerous studies in the realm of DCE-MRI. Our findings strongly indicate that the aggregation of information across all time instants is pivotal for enhancing the diagnostic process, and vastly superior to a simplistic instant-wise analysis. While applied to MRI sequences, our approach can be extended to general problems of multimodal data integration. We make our code publicly available on GitHub (link omitted to keep the manuscript anonymous).**

***Index Terms*—Graph Neural Network, Multiplex Network, Radiomics, Breast Cancer, MRI.**

I. Introduction

Breast cancer represents the main disease that threatens womens' health (1). Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) has been widely used for breast cancer classification. The DCE-MRI measures changes in tissues over time after administration of a contrast agent, resulting in a sequence of MRI scans. Radiologists then analyze the resulting sequence to assess the differences in contrast agent uptake between malignant and benign lesions (2). From a radiological point of view, malignant lesions exhibit an immediate increase in signal intensity, while benign lesions show a slower increase (3). Unfortunately, despite DCE-MRI being widely used to improve MRI in characterizing breast lesions (4), its specificity remains low (around 72%) (5).

Several works have recently attempted to analyze DCE-MRI sequences for breast cancer classification using Machine Learning methods (6; 7; 8; 9; 10). Specifically, through Radiomics (11), imaging features are extracted to provide a quantitative viewpoint complementary to the physician's visual/qualitative one. However, despite the demonstrated predictive capabilities of radiomic features, current works do not fully exploit the DCE-MRI potential (6; 7; 8; 9; 10). Most work focuses on extracting features at single instants, failing to exploit the temporal information of the sequence. The temporal dimension embedded in DCE-MRI incorporates the assumption that a model should possess the ability to extract and integrate information from all time points to take full advantage of its capabilities. From a clinical point of view, this reflects the physician's diagnostic process of examining the entire sequence to assess contrast agent uptake. From a computer science perspective, it involves creating a model capable of aggregating information across all time points.

Graphs are ubiquitous data structures, widely used in many fields of research. They are particularly pervasive in biology and medicine where they can be used to represent a multitude of biological entities and their associations. Graph Neural Networks (GNNs) are a class of neural networks designed to operate on data defined over graphs (12). Since their introduction (13), GNNs have shown outstanding results in a plethora of fields, such as computational chemistry (14), bioinformatics (15), social networks (16), physics simulations (17) and many more. The key idea of GNNs is to exploit the inductive bias induced by the topology of graph-structured data, represented by the connectivity structure, to perform graph representation learning tasks.

In this work, we exploit the power and flexibility of Graph Neural Networks to develop a learning framework that achieves accurate breast cancer classification from DCE-MRI sequences. The radiomic workflow was utilized to extract MRI features for each instant of the DCE-MRI sequence. We developed two alternative graph construction methodologies, a simple one based on a fully connected topology, and a more sophisticated one, named MUGI-MRI (MUltiplex Graph neural network for Integration of MRI), based on a multiplex network, to incorporate the dynamic information from the DCE-MRI scans and exploit their temporal dimension to achieve more accurate classification performance. This paper introduces new perspectives for the analysis of the DCE-MRI, addressing a gap in the existing literature, which overlooks the importance of leveraging all temporal instants offered by the DCE-MRI sequences. In particular, new GNN-based methods for aggregating information from short time series

are proposed. Our results, in addition to surpassing previous work in terms of accuracy, demonstrated a remarkable balance between model sensitivity and specificity, improving this balance significantly compared to the literature. The proposed methods, while developed and tested for DCE-MRI analysis, offer interesting insights for further study and investigation in numerous other contexts.

II. Background and Related Work

In recent times, Radiomics has become a prevalent approach for extracting features from radiological images. The Radiomics workflow is designed to develop data-driven models, involving various image analysis steps such as data acquisition, identification, and Region of Interest (ROI) segmentation, as well as feature extraction and selection (11). Specifically, ROIs are transformed into highly informative quantitative radiomic features, which have been shown to be highly correlated to clinical outcomes (18). These radiomic features quantify the shape of ROIs, as well as the texture and intensity of gray levels. Further, more sophisticated features can be incorporated using techniques like wavelet transforms, Laplacian-of-Gaussian filtering, and other preprocessing methods on the images (19). The process of radiomic feature extraction can offer two primary advantages over deep extraction methods. Firstly, achieving precise feature extraction does not demand an extensive dataset, in contrast to deep learning where large amounts of data are indispensable for training convolutional-based or transformer-based architectures (20). Secondly, Radiomics enables efficient feature extraction at the original spatial resolution, preserving the essential information even in very small lesions.

Guided by these advantages, several works have implemented the radiomic workflow for DCE-MRI analysis. These works mainly involve the use of very small datasets and shallow learning methods. For example, a Support Vector Machine (SVM) model was trained using radiomic features to classify sub 1-cm breast lesions in a 165-sample dataset (6). A logistic regression model was employed to classify 133 lesions in (7). In (8), radiomic features from several MRI modalities were aggregated and classified using the isomap method for dimensionality reduction (21) and SVM as classifier, trained on a dataset composed of 124 multiparametric breast MR images. Five imaging modalities for feature extraction, including DCE-MRI, and 207 histopathology-confirmed breast lesions were used to train an SVM model in (9). In (10), several feature selection and shallow learning methods were explored in a 111-patient cohort and a SVM was eventually trained for classification. All time instants of the DCE-MRI sequences were evaluated separately to determine the best instant for classification in (22), using SVM, Random Forest and XGBoost. Aside from shallow methods, several deep learning architectures have been proposed to analyze DCE-MRI sequences. A local–global cross attention fusion network (LG-CAFN) was implemented for DCE-MRI breast segmentation and classification in (23). The authors explored several 2D and 3D deep architectures trained considering and aggregating pre-contrast, post-contrast, and subtraction of sequences. Finally, object detection architectures such as Faster-RCNN-based models were employed for lesion detection and classification in (24).

Despite demonstrating promising performance in both radiomic and deep workflow, prior research has not fully harnessed the potential of DCE-MRI. The temporal relationships between the different moments in the sequences have not been fully exploited. Instead, individual instants are often treated separately or modeled through sequence subtraction, leading to an approach that can be *(i)* suboptimal for effectively aggregating information across multiple time points, and *(ii)* inconsistent with the radiological process of DCE-MRI evaluation, which focuses on the evolution of the contrast medium rather than individual instants.

III. Materials and Methods

A. Dataset Description

In our work, a collection of 166 breast mass enhancements was included. The dataset was acquired at the XXX section of Hospital XXX. A consensus classification was carried out by two experienced breast radiologists for the 166 breast mass enhancements. The breast MRI protocols performed consisted of a dynamic contrast-enhanced (DCE) sequence; in detail, a 3D GRE T1w with fat saturation sequence (named Vibrant), with one acquisition before and six acquisitions after contrast media administration was employed. A temporal resolution of ≈ 80 seconds was used. Figure 1 shows an example of a DCE-MRI sequence.

As shown in Figure 2, the main components of the proposed workflow prior to the deep model training are: ROI segmentation, feature extraction, preprocessing and selection. Details for each are as follows.

1) ROI Segmentation: The segmentation process was performed using the 3D-Slicer software (25) by three distinct operators along with three senior radiologists. Each operator selected the post-contrast phase that best depicted the lesion contours. To validate the segmentation accuracy and consistency, an independent breast radiologist with a decade of expertise in breast MR imaging assessed and confirmed the segmentation outcomes for all 166 breast masses.

2) Feature Extraction: A set of 1,023 radiomic features was extracted following the Imaging Biomarker Standardization Initiative (IBSI) (26) and using pyradiomics (Version 3.0.1) (27). Table I summarizes the extracted features. Ninety-three features were extracted from the original (un-processed) images, consisting of First Order (FO) intensity histogram statistics, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence Matrix (GLDM) and Neighboring Gray Tone Difference Matrix (NGTDM). The above-mentioned 93 features were also extracted from Laplacian of Gaussian (LoG) filtered images ($\sigma = [2, 3]$), collecting 186 LoG-derived features. In the end, wavelet-derived features were also extracted using the Haar kernel,

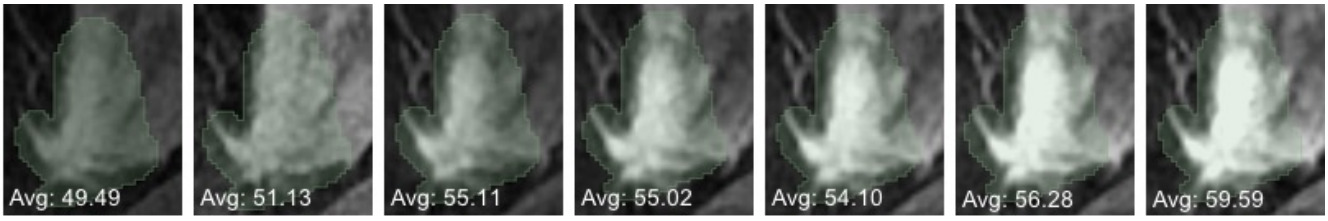

Fig. 1. An example of a DCE-MRI sequence, depicting the manually segmented ROI in green. The average gray level intensities for each ROI are also shown.

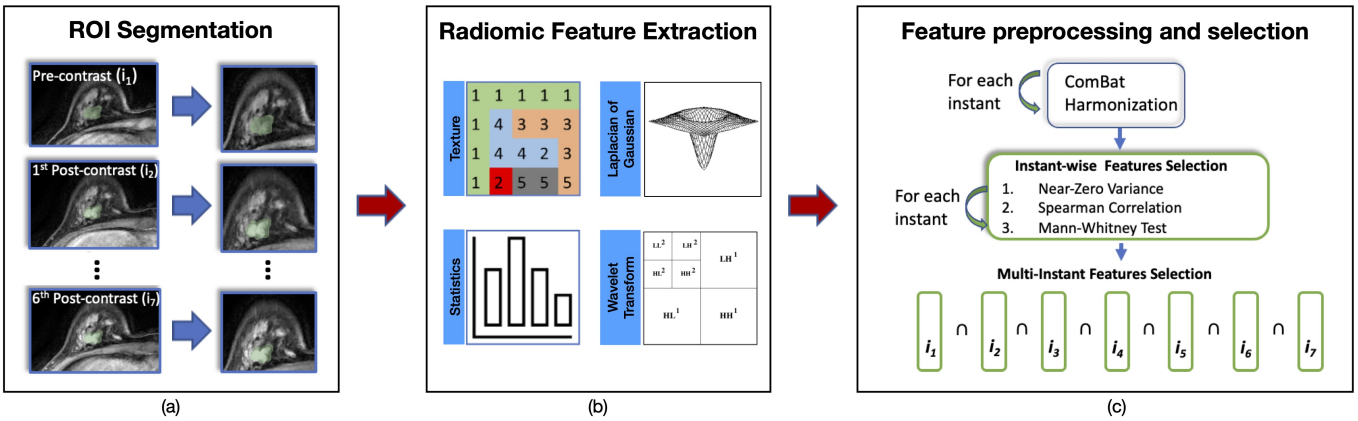

Fig. 2. Workflow for feature extraction, preprocessing, and selection: (a) DCE-MRI sequences underwent manual segmentation and validation by expert radiologists. (b) Textural, statistical, LoG-derived, and Wavelet-derived radiomic features were extracted from each instant. (c) Feature harmonization was performed to align feature distributions in our multi-protocol dataset. Instant-wise feature selection was then applied to eliminate redundant features and retain only statistically significant ones for each instant. Finally, multi-instant feature selection was employed to select features that were statistically significant in at least five out of the seven time instants of the DCE-MRI sequences.

generating a total of 744 features, considering 93 features for each of the 8 wavelet decompositions.

Features were extracted independently from each of the seven time instants in the DCE-MRI sequence. Overall, our dataset consisted of 166 samples, and 1,023 distinct features across the seven (one pre-contrast and six post-contrast) time instants.

TABLE I
SUMMARY OF ALL EXTRACTED FEATURES. L AND H REPRESENT LOW-PASS FILTER AND HIGH-PASS FILTER, AND σ IDENTIFIES THE LOG KERNEL SIZE.

Features	Dimension	Description
Original	93	Features extracted from the un-processed images
Wavelet	744	93 x 8, where 93 are the Original features and 8 are the LLH, LHL, LHH, HLL, HLH, HHL, HHH, and LLL wavelet 3D decompositions.
LoG	186	93 x 2, where 93 are the Original features and 2 are $\sigma = [2, 3]$.
Total	**1023**	Extracted features for each time instant

3) Feature preprocessing and selection: Since the DCE-MRI sequences were acquired using two different imaging protocols, ComBat was employed to harmonize this multi-protocol dataset (28). ComBat was directly applied to the radiomic features of each instant within the sequence. For each sequence instant, low variance features ($\sigma^2 < 0.01$), and highly correlated features (Spearman's rank correlation $|\rho| > 0.9$) were discarded (29). The Mann-Whitney U test was employed to compare the malignant and benign distributions. A significance level of $p < 0.05$ was considered. Finally, the selected features from each time instant were incorporated into the multi-instant feature selection process, where only the features found statistically significant simultaneously in at least 5 (out of the 7) instants were selected.

B. Graph Neural Networks

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 $\mathbf{x}_v$, and 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. The connectivity structure of G is fully captured by its adjacency matrix $\mathbf{A}$, in which the entry i, j of $\mathbf{A}$ is equal to 1 if node i is connected to node j and 0 otherwise. Here, $\mathbf{A}$ is *symmetric* (that is, $\mathbf{A} = \mathbf{A}^T$). In GNNs, each layer consists of a nonlinear function that maps a feature matrix into a new (hidden) feature matrix, accounting for the pairwise relationships in the underlying graph captured by its connectivity. Formally,

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

where $\mathbf{H}^{(l)}$ is the hidden feature matrix at layer l and $\mathbf{H}^{(0)} = \mathbf{X}$. By adapting different implementations of Equation 1, a plethora of Graph Neural Network architectures have been developed over the years (30; 31; 32; 33; 34).

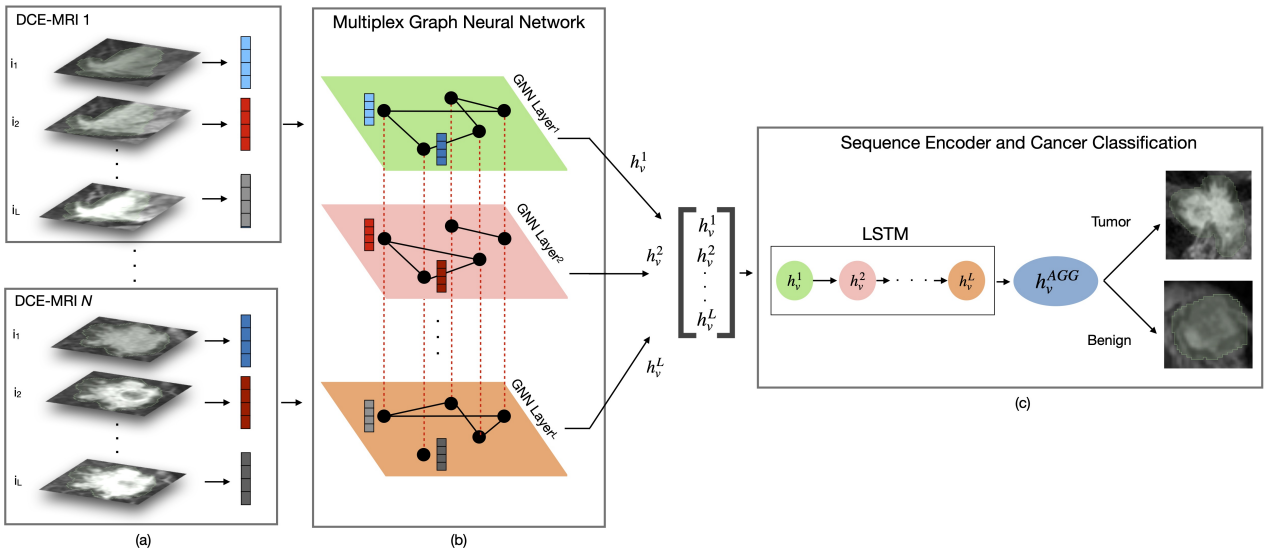

Fig. 3. Proposed workflow for breast cancer classification using MUGI-MRI. (a) For every DCE-MRI sequence composed of time instants from i_1 to i_L, where L is the length of the sequence, radiomic features are extracted from the ROIs. (b) We construct a multiplex network where each node is replicated over L layers. For a single DCE-MRI sequence, the initial features of a node v in a layer l correspond to the features extracted in step (a) from i_l. We then parameterize each layer with GNNs to learn a hidden representation h_v^l for each node v in each layer l. (c) We finally generate a temporally-aware representation for each sample through an LSTM sequence encoder. The final aggregated representation is used for cancer classification.

C. Multiplex Graph Neural Networks

A multiplex network is a multi-layer representation of a graph defined by the quadruple $M = (V, R, B, \mathbb{M})$. V corresponds to the set of nodes that are separately instantiated in each layer of the multiplex system. R represents the set of relations which form the basis for each layer graph of the multiplex network. The separate instantiations of each node $v \in V$ across each relation $r \in R$ are given by a binary relation on V and R, such that $B = (V, R, \beta)$, $\beta \subseteq V \times R$. The set $\mathbb{M} = \{G^r\}_{r \in R}$ defines the relation-specific graphs. Finally, we can define a coupling graph G^C on β in which there is an edge between two node-layer pairs (v, i) and (u, j) only if $v = u$. Essentially, G^C defines the coupling between the same node $v \in V$ across different layers of the multiplex graph. As suggested in (35), training GNNs on multiplex networks involves parameterizing relation-specific representations of nodes, and then aggregating these representations to model the multi-relational context depending on the ad hoc task.

IV. Representing Sequences as Graphs

In order to represent a sequence of MRI scans as a graph, we evaluated two graph construction strategies: *(i)* a simple one which leverages a fully-connected topology among all the time instants of a DCE-MRI sequence, and *(ii)* a more sophisticated one in which each time instant is modeled as a layer of a multiplex graph. Details for each are given as follows.

A. Nodes as time steps: Fully Connected Graphs

Each sample in the dataset was represented as a single graph, consisting of seven nodes (each identifying one time instant of the contrast agent administration). We defined the connectivity of these graphs as fully connected to allow for information flow between all time instants. By allowing a fully-connected topology, we hypothesized that important relations exist between non-consecutive time steps, and exploiting a complete graph structure could be useful to accurately perform cancer classification. 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.

B. Nodes as patients: Temporal Multiplex Graph

For the second graph construction strategy, each sample was represented as a node in a multiplex graph, consisting of seven relational layers. Each layer identified one of the seven time instants, and the connectivity at each layer was defined by intra-layer nearest-neighbor similarity. This resulted in a dataset composed of a single multiplex network, with seven graph layers of 166 nodes each. In order to aggregate the information across the different layers and to capture their temporal dependency, the individual graph-generated node representations for the same node across all layers were fed into a Long Short-Term Memory (LSTM) sequence encoder (36) to generate a final, temporally-aware representation for each sample. This aggregated representation was then used for the final classification. The proposed workflow, that we named MUGI-MRI (MUltiplex Graph neural network for Integration of MRI), is depicted in Figure 3.

V. Experimental Results

In this section, we present the results of the proposed framework, distinguishing between the two above-mentioned

TABLE II
CLASSIFICATION RESULTS FOR THE PROPOSED FULLY-CONNECTED GRAPH ARCHITECTURES.

MODEL	Accuracy	AUC-ROC	Specificity	Sensitivity	PPV	NPV
GCN	0.7357 ± 0.1146	0.7731 ± 0.1079	0.7708 ± 0.1475	0.7000 ± 0.1807	0.7730 ± 0.1357	0.7344 ± 0.1409
GAT	0.7415 ± 0.1323	0.7882 ± 0.0891	0.7972 ± 0.1808	0.6861 ± 0.1782	0.7966 ± 0.1457	0.7232 ± 0.1472
SAGE	0.7415 ± 0.1185	0.7715 ± 0.0980	0.7833 ± 0.1505	0.7000 ± 0.1684	0.7853 ± 0.1383	0.7289 ± 0.1303

TABLE III
ABLATION STUDY: DEEPSET

MODEL	Accuracy	AUC-ROC	Specificity	Sensitivity	PPV	NPV
GCN	0.7059 ± 0.1486	0.7628 ± 0.1305	0.6986 ± 0.1354	0.7139 ± 0.1887	0.6998 ± 0.1447	0.7187 ± 0.1572
GAT	0.6934 ± 0.1443	0.7760 ± 0.1024	0.6972 ± 0.1492	0.6889 ± 0.1986	0.6967 ± 0.1390	0.7031 ± 0.1570
SAGE	0.7114 ± 0.0944	0.7729 ± 0.1078	0.7097 ± 0.1478	0.7097 ± 0.1307	0.7323 ± 0.1297	0.7135 ± 0.1007

TABLE IV
ABLATION STUDY: SEQUENTIAL GRAPHS – CHAINS

MODEL	Accuracy	AUC-ROC	Specificity	Sensitivity	PPV	NPV
GCN	0.7000 ± 0.1542	0.7646 ± 0.1270	0.7097 ± 0.1369	0.6917 ± 0.2053	0.6965 ± 0.1581	0.7075 ± 0.1571
GAT	0.6934 ± 0.1443	0.7736 ± 0.1011	0.7097 ± 0.1478	0.6778 ± 0.2201	0.6967 ± 0.1390	0.7038 ± 0.1565
SAGE	0.7055 ± 0.0999	0.7717 ± 0.1101	0.7097 ± 0.1478	0.6986 ± 0.1383	0.7281 ± 0.1323	0.7065 ± 0.1087

graph construction strategies. To evaluate model performances, Accuracy, Area Under the Receiver Operating Characteristic (AUC-ROC), Specificity, Sensitivity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) were considered. We report mean and standard deviation scores evaluated with stratified 10-fold cross-validation. In this work, we evaluate the proposed framework using Graph Convolutional Networks (GCNs) (30), Graph Attention Networks (GATs) (31), and GraphSAGE (32) trained to minimize the binary cross-entropy loss between tumor and normal samples. Adam optimizer (37) with a learning rate of 0.001 was used to train the models for 100 epochs with a patience of 10 epochs. The batch size was set to 32, and we employed weight decay (L2) regularization to reduce overfitting. For the fully connected graph construction strategy, all the models were implemented with a single graph layer. For the MUGI-MRI workflow, we selected the 30 nearest neighbors for each sample in each layer to define its connectivity and implemented all models with three graph layers.

A. Features Selected

After undergoing the preprocessing and multi-instant selection steps, a total of 44 features were selected. These 44 features exhibited no near-zero variance, were uncorrelated, and demonstrated statistical significance in at least 5 out of 7 DCE-MRI instants. Among these features, 11 were extracted from the original images; 7 features were LoG-filtered images, and 26 were wavelet-derived.

B. Fully Connected GNNs

Table II reports the scores obtained using the fully connected graph construction strategy discussed above. Overall, GAT and GraphSAGE exhibited better performances compared to GCN. To analyze the benefits of exploiting the topology induced by the fully connected graph structures, we performed two ablation studies that disregarded such information. Introduced in (38), DeepSet considers objective functions defined on sets, that are invariant to permutations. Using a DeepSet formulation, we constructed graphs with features where each node is only connected to itself. As for the graph models, we trained DeepSet to minimize the binary cross-entropy between tumor and normal samples and report the results in Table III. Furthermore, we also constructed chain-like graphs where each time step (node) i was only connected to itself and time step (node) $i+1$. Results of this evaluation are reported in Table IV. All the scores in Table III and Table IV, computed by ignoring the fully connected information, are lower than those reported in Table II. While directly exploiting the temporal sequence of MRI scans to perform cancer classification mimics the radiologist's decision process, these results demonstrated that complex, non-linear interactions exists between non-consecutive time steps, and by leveraging a more complex, complete topology, graph neural networks are able to produce more accurate predictions.

C. Temporal Multiplex GNNs

For the MUGI-MRI workflow (Figure 3), we separately applied a graph neural network to each layer of the network to learn a fixed-size node representation of 32. For each sample, we then constructed a sequence by stacking the learned representations corresponding to the same node instantiation across all layers, before feeding it to an LSTM sequence encoder to learn an aggregated node representation. The results of this evaluation, for the three graph architectures explored, are reported in Table V. Similarly to Table II, GAT reached

TABLE V
CLASSIFICATION RESULTS FOR THE PROPOSED MUGI-MRI ARCHITECTURES WITH LSTM AGGREGATION.

MODEL	Accuracy	AUC-ROC	Specificity	Sensitivity	PPV	NPV
GCN	0.7276 ± 0.1422	0.7884 ± 0.1353	0.7819 ± 0.2165	0.6778 ± 0.1226	0.7841 ± 0.1841	0.6953 ± 0.1141
GAT	0.7518 ± 0.0902	0.8017 ± 0.1146	0.7431 ± 0.1796	0.7597 ± 0.1217	0.7724 ± 0.1299	0.7628 ± 0.1042
SAGE	0.7114 ± 0.1014	0.7703 ± 0.1286	0.6944 ± 0.1506	0.7222 ± 0.1542	0.7151 ± 0.1088	0.7306 ± 0.1327

TABLE VI
CLASSIFICATION RESULTS FOR THE PROPOSED MUGI-MRI ARCHITECTURES WITH MLP AGGREGATION.

MODEL	Accuracy	AUC-ROC	Specificity	Sensitivity	PPV	NPV
GCN	0.6971 ± 0.1111	0.7717 ± 0.1230	0.6694 ± 0.2248	0.7236 ± 0.1315	0.7139 ± 0.1238	0.6986 ± 0.1617
GAT	0.7048 ± 0.1005	0.7493 ± 0.1146	0.6931 ± 0.2202	0.7111 ± 0.1032	0.7321 ± 0.1342	0.6948 ± 0.0868
SAGE	0.6882 ± 0.1118	0.7649 ± 0.1014	0.6708 ± 0.1777	0.7000 ± 0.1604	0.6995 ± 0.1276	0.7100 ± 0.1433

the highest classification performance. Compared to the fully connected topology, the proposed multiplex networks were able to distinguish between benign and malignant lesions more accurately. The use of the LSTM sequence encoder allows us to aggregate the graph representations learned from each layer and to create a temporally-aware final sample representation. As an ablation study, we replaced the LSTM sequence encoder in the MUGI-MRI workflow with a simple 2-layer Multilayer perceptron (MLP). Results of this evaluation are reported in Table VI. Compared to Table V, disregarding the sequential ordering of the MRI scans and their temporal dimensions, resulted in lower performances across all metrics.

D. Comparison with SOTA

A valid comparison of our framework can be drawn with Prinzi et al. (22), wherein the same dataset is employed. In (22), the authors made use of a Random Forest (RF) model trained using instant-wise information. Table VII reports the cross-validated mean and standard deviation values calculated for the instant-wise RF model reported in (22) and for the proposed fully-connected GNNs (FC-GNN) and multiplex GNNs (MUGI-MRI). For the FC-GNN and MUGI-MRI setting, we selected the GAT models as they exhibited on average a higher AUROC in all tested configurations. Compared to the instant-wise analysis, both graph construction strategies showed better performances across all metrics. By exploiting the inductive bias induced by the topology of the graph structure, we were able to explicitly incorporate and model the sequential nature of the DCE-MRI scans. In addition to exhibiting significantly superior performance in all metrics, MUGI-MRI is distinguished by a higher balance achieved between specificity and sensitivity. This indicates an equal ability to classify both benign and malignant lesions. In contrast, the instant-wise model and FC-GNN exhibit a pronounced imbalance. This substantial imbalance is a prevalent problem affecting numerous studies in the realm of DCE-MRI, spanning conventional radiological diagnostics and machine learning works. In particular, DCE-MRI is well-known to provide high sensitivity but low specificity for radiologist evaluation (39). While the specificity-sensitivity imbalance has been reported in numerous radiomic works (6; 7; 9; 40; 41; 42), the proposed multiplex approach demonstrates a remarkable ability to accurately distinguish between benign and malignant lesions.

TABLE VII
COMPARISON WITH THE CURRENT STATE-OF-THE-ART ON THE SAME DATASET. THE BEST PERFORMANCES ARE IN **BOLD**.

Metrics	Instant-wise (22)	FC-GNN	MUGI-MRI
Accuracy	0.7100 ± 0.1300	0.7415 ± 0.1323	$\mathbf{0.7518 \pm 0.0902}$
AUROC	0.7410 ± 0.1350	0.7882 ± 0.0891	$\mathbf{0.8017 \pm 0.1146}$
Specificity	0.7380 ± 0.1770	$\mathbf{0.7972 \pm 0.1808}$	0.7431 ± 0.1796
Sensitivity	0.6830 ± 0.1780	0.6861 ± 0.1782	$\mathbf{0.7597 \pm 0.1217}$
PPV	0.7430 ± 0.1530	$\mathbf{0.7966 \pm 0.1457}$	0.7724 ± 0.1299
NPV	0.7030 ± 0.1450	0.7232 ± 0.1472	$\mathbf{0.7628 \pm 0.1042}$

VI. DISCUSSION AND CONCLUSION

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) plays a pivotal role in the diagnosis of breast lesions. In this work, we presented a novel framework to perform breast cancer classification from DCE-MRI through Graph Neural Networks. Specifically, we developed two graph construction strategies which allowed us to model and integrate information from all time instants of a DCE-MRI sequence: (i) a fully connected topology, where each node represented a time point of the DCE-MRI sequence, and (ii) a more complex representation based on a flexible multiplex network, named MUGI-MRI (MUltiplex Graph neural network for Integration of MRI), where each node represented an individual sample (patient). Our findings revealed the effectiveness of the proposed framework to enhance the temporal analysis of DCE-MRI sequences, by surpassing the state-of-the-art on the same dataset in distinguishing between benign and malignant lesions. The developed multiplex framework allowed us to achieve a remarkable specificity-sensitivity balance, which has been reported to be a problem affecting numerous DCE-MRI studies. We performed several ablation studies which, aiming to disrupt the connectivity structure of the proposed graph architectures, served as proof of the importance of integrating the temporal dimension in DCE-MRI sequences to obtain accurate classification results. Indeed, optimal outcomes are

achieved through a gradual elevation of integration levels between instants. The least favorable performance is observed when treating each instant separately (22). Subsequently, enhanced performance is achieved when instants are aggregated into a sequence (Table IV) and through DeepSet (Table III). The highest performances are achieved when a fully connected graph topology (Table II) and a multiplex network (Table V) are implemented to aggregate all time instants. These results underscore the significance of quantitative analysis in DCE-MRI, as it enables the exploration of previously unidentified relationships among multiple temporal instants.

Finally, the proposed multiplex approach offers interesting insights for further study and numerous potential applications in various other fields, such as multi-omics analysis and integration, pharmacology, and neuroscience.

VII. Acknowledgments

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] J. Xiao, H. Rahbar, D. S. Hippe, M. H. Rendi, E. U. Parker, N. Shekar, M. Hirano, K. J. Cheung, and S. C. Partridge, "Dynamic contrast-enhanced breast mri features correlate with invasive breast cancer angiogenesis," *NPJ breast cancer*, vol. 7, no. 1, p. 42, 2021.

[3] A. R. Padhani, "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*, vol. 16, no. 4, pp. 407–422, 2002.

[4] R. H. El Khouli, K. J. Macura, M. A. Jacobs, T. H. Khalil, I. R. Kamel, A. Dwyer, and D. A. Bluemke, "Dynamic contrast-enhanced mri of the breast: quantitative method for kinetic curve type assessment," *AJR. American journal of roentgenology*, vol. 193, no. 4, p. W295, 2009.

[5] A. Orlando, M. Dimarco, R. Cannella, and T. V. Bartolotta, "Breast dynamic contrast-enhanced-magnetic resonance imaging and radiomics: state of art," *Artificial Intelligence in Medical Imaging*, vol. 1, no. 1, pp. 6–18, 2020.

[6] P. Gibbs, N. Onishi, M. Sadinski, K. M. Gallagher, M. Hughes, D. F. Martinez, E. A. Morris, and E. J. Sutton, "Characterization of sub-1 cm breast lesions using radiomics analysis," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 5, pp. 1468–1477, 2019.

[7] J. Zhou, Y. Zhang, K.-T. Chang, K. E. Lee, O. Wang, J. Li, Y. Lin, Z. Pan, P. Chang, D. Chow *et al.*, "Diagnosis of benign and malignant breast lesions on dce-mri by using radiomics and deep learning with consideration of peritumor tissue," *Journal of Magnetic Resonance Imaging*, vol. 51, no. 3, pp. 798–809, 2020.

[8] V. S. Parekh and M. A. Jacobs, "Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric mri," *NPJ breast cancer*, vol. 3, no. 1, p. 43, 2017.

[9] Q. Zhang, Y. Peng, W. Liu, J. Bai, J. Zheng, X. Yang, and L. Zhou, "Radiomics based on multimodal mri for the differential diagnosis of benign and malignant breast lesions," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 2, pp. 596–607, 2020.

[10] C. Militello, L. Rundo, M. Dimarco, A. Orlando, R. Woitek, I. D'Angelo, G. Russo, and T. V. Bartolotta, "3d dce-mri radiomic analysis for malignant lesion prediction in breast cancer patients," *Academic Radiology*, vol. 29, no. 6, pp. 830–840, 2022.

[11] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[13] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

[14] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[15] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[16] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The world wide web conference*, 2019, pp. 417–426.

[17] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, "Learning to simulate complex physics with graph networks," in *International conference on machine learning*. PMLR, 2020, pp. 8459–8468.

[18] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. De Jong, J. Van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," *Nature reviews Clinical oncology*, vol. 14, no. 12, pp. 749–762, 2017.

[19] A. Demircioğlu, "The effect of preprocessing filters on predictive performance in radiomics," *European Radiology Experimental*, vol. 6, no. 1, p. 40, 2022.

[20] P. Wei, "Radiomics, deep learning and early diagnosis in oncology," *Emerging topics in life sciences*, vol. 5, no. 6, pp. 829–835, 2021.

[21] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[22] F. Prinzi, A. Orlando, S. Gaglio, M. Midiri, and S. Vitabile, "Ml-based radiomics analysis for breast cancer classification in dce-mri," in *International Conference on Applied Intelligence and Informatics*. Springer, 2022, pp. 144–158.

[23] X. Zhao, Y. Liao, J. Xie, X. He, S. Zhang, G. Wang, J. Fang, H. Lu, and J. Yu, "Breastdm: A dce-mri dataset for breast tumor image segmentation and classification," *Computers in Biology and Medicine*, vol. 164, p. 107255, 2023.

[24] Y. Chen, L. Wang, R. Luo, S. Wang, H. Wang, F. Gao, and D. Wang, "A deep learning model based on dynamic contrast-enhanced magnetic resonance imaging enables accurate prediction of benign and malignant breast lessons," *Frontiers in oncology*, vol. 12, p. 943415, 2022.

[25] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, "3d slicer: a platform for subject-specific image analysis, visualization, and clinical support," in *Intraoperative imaging and image-guided therapy*. Springer, 2013, pp. 277–289.

[26] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard *et al.*, "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

[27] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[28] J.-P. Fortin, N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. McGrath *et al.*, "Harmonization of cortical thickness measurements across scanners and sites," *Neuroimage*, vol. 167, pp. 104–120, 2018.

[29] C. Militello, F. Prinzi, G. Sollami, L. Rundo, L. La Grutta, and S. Vitabile, "Ct radiomic features and clinical biomarkers for predicting coronary artery disease," *Cognitive Computation*, vol. 15, no. 1, pp. 238–253, 2023.

[30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[31] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio *et al.*, "Graph attention networks," *stat*, vol. 1050, no. 20, pp. 10–48 550, 2017.

[32] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[33] C. Bodnar, F. Frasca, N. Otter, Y. Wang, P. Lio, G. F. Montufar, and M. Bronstein, "Weisfeiler and lehman go cellular: Cw networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 2625–2640, 2021.

[34] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[35] J. Melton and S. Krishnan, "muxgnn: Multiplex graph neural network for heterogeneous graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[38] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. R. Salakhutdinov, and A. J. Smola, "Deep sets," *Advances in neural information processing systems*, vol. 30, 2017.

[39] A. Orlando, M. Dimarco, R. Cannella, and T. V. Bartolotta, "Breast dynamic contrast-enhanced-magnetic resonance imaging and radiomics: state of art," *Artificial Intelligence in Medical Imaging*, vol. 1, no. 1, pp. 6–18, 2020.

[40] N. Aristokli, I. Polycarpou, S. Themistocleous, D. Sophocleous, and I. Mamais, "Comparison of the diagnostic performance of magnetic resonance imaging (mri), ultrasound and mammography for detection of breast cancer based on tumor type, breast density and patient's history: A review," *Radiography*, 2022.

[41] H. Dong, L. Kang, S. Cheng, and R. Zhang, "Diagnostic performance of dynamic contrast-enhanced magnetic resonance imaging for breast cancer detection: an update meta-analysis," *Thoracic Cancer*, vol. 12, no. 23, pp. 3201–3207, 2021.

[42] L. Zhang, M. Tang, Z. Min, J. Lu, X. Lei, and X. Zhang, "Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis," *Acta radiologica*, vol. 57, no. 6, pp. 651–660, 2016.