npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

Article in Press

https://doi.org/10.1038/s41698-026-01649-x

Graph-radiomic learning (GrRAiL) descriptor to characterize imaging heterogeneity in confounding tumor pathologies

Dheerendranath Battalapalli[1,2], Apoorva Safai[2], María Jaramillo[2], Hyemin Um[2], Gustavo Adalfo Pineda Ortiz[2], Ulas Bagci[3], Manmeet Singh Ahluwalia[4], Marwa Ismail[2] & Pallavi Tiwari[1,4,5] ✉

Distinguishing benign confounding pathologies from malignant neoplasms on routine imaging remains a significant challenge in solid tumors. Radiomics approaches attempt to capture lesion heterogeneity on clinical imaging (CT,MRI), but typically aggregate feature values across the region of interest (ROI), losing the complex spatial relationships among intensity compositions within the lesion. We present Graph-Radiomic Learning (GrRAiL), a new descriptor for characterizing intralesional heterogeneity (ILH) on clinical MRI. GrRAiL (1) identifies clusters of sub-regions from per-voxel radiomic measurements and (2) computes graph-theoretic measurements that quantify the spatial associations of these radiomic clusters. The resulting weighted graphs capture higher-order spatial relationships within the ROI to reliably distinguish confounding pathologies from malignancies. GrRAiL was evaluated on $n = 947$ subjects across three multi-institutional use-cases: distinguishing radiation effects from tumor recurrence in glioblastoma ($n = 106$) and brain metastasis ($n = 233$), and differentiating no+low- versus high-risk pancreatic intraductal papillary mucinous neoplasms (IPMNs) ($n = 608$). GrRAiL consistently outperformed Graph Neural Networks, textural radiomics, and intensity graph analysis, achieving cross-validation and test accuracies of 89% and 78% (glioblastoma), 84% and 74% (metastasis), and 85% and 79% (IPMN), representing >10%, >13%, and >10% test-accuracy improvements over comparative approaches, respectively.

A critical challenge in oncology is the frequent diagnostic ambiguity in differentiating between benign confounding pathologies and malignant neoplasms on clinical MRI scans[1]. For instance, in neuro-oncology, the distinction between true tumor recurrence (TuR) and benign treatment-related effects (i.e. psuedoprogression (PsP), radiation necrosis (RN)) on post-treatment MRI scans, presents a well-known diagnostic dilemma[2–4]. TuR and benign treatment effects exhibit similar MRI visual characteristics, which often require invasive surgical interventions for a definitive diagnosis[5,6]. Similarly, in the screening for pancreatic cancer, the early identification of premalignant lesions represents a crucial strategy to improve patient prognosis. Despite an increased detection rate for incidental pancreatic cysts, their accurate classification into "high-risk" and "low-to-no-risk" categories remains a significant challenge[7].

Fundamentally, malignant lesions are characterized by a complex and distinct microenvironment organization with different tissue types disbursed through the lesion, in contrast to benign lesions[8,9]. Thus, the ability to analyze the subtle organizational structure of tumors and their underlying heterogeneity could enable reliable discrimination of benign confounding pathologies from malignant lesions. This noninvasive characterization could also significantly impact clinical decision-making, thus mitigating the need for unnecessary biopsies and treatments in patients with benign conditions.

[1]Department of Biomedical Engineering, University of Wisconsin–Madison, Madison, WI, USA. [2]Department of Radiology, University of Wisconsin–Madison, Madison, WI, USA. [3]Department of Radiology, Northwestern University, Evanston, IL, USA. [4]Miami Cancer Institute, Baptist Health South Florida, Miami, FL, USA. [5]William S. Middleton Memorial Veterans Affairs Healthcare System, Madison, WI, USA. ✉e-mail: ptiwari9@wisc.edu

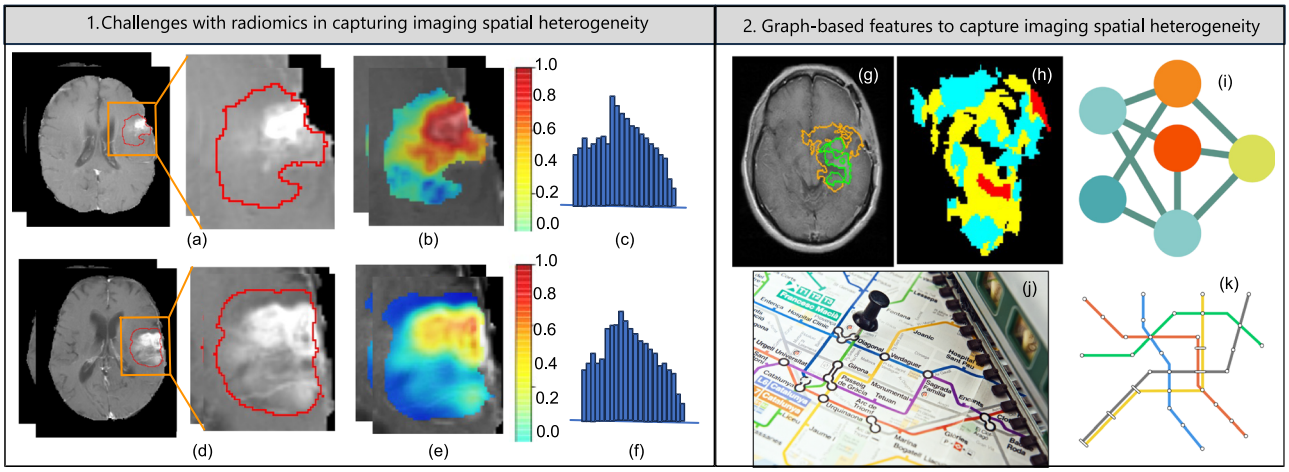

Fig. 1 | Radiomic feature expression maps and graph representations reveal distinct spatial heterogeneity patterns that distinguish tumor recurrence from pseudoprogression in glioblastoma. Panels **a**, **d** demonstrate Gd-T1w MRI scans for TuR and PsP in glioblastoma studies, respectively. **b**, **e** Highlight the corresponding GLCM entropy expression maps, while (**c**, **f**) are the histogram plots from subjects with confirmed TuR and PsP, respectively. It is worth noting that while the spatial distribution of the radiomic feature expressions in (**b**) and (**e**) is distinct, the histogram distribution of the features in (**c**) and (**f**) appears similar, which could potentially contribute to misclassification of confounding pathologies (i.e., TuR versus PsP). **h** Represents a clustered GLCM radiomics map (with 3 clusters) corresponding to a (**g**) glioblastoma Gd-T1w MRI scan, and the analogy of capturing spatial interactions across radiomic clusters via a graphical representation (**i**) with that of a subway map shown in (**j**) and (**k**).

Several approaches have attempted to capture surrogate markers of lesion heterogeneity on imaging (e.g. MRI, CT) to distinguish between confounding pathologies, including (1) radiomics that encompasses intensity-based, textural, and topological characteristics of the lesion[10–17], and (2) deep learning models that extract complex features directly from the images without the need for semantic segmentation, including convolutional neural networks (CNN) and graph neural networks (GNN). While deep learning approaches may demonstrate advantages over radiomics with larger training cohorts, DL approaches tend to be 'opaque' with regard to feature interpretability.

Among radiomic techniques, the Gray Level co-occurrence Matrix (GLCM) quantifies textural variations within the region of interest (ROI) at a voxel-level, correlating imaging heterogeneity with tumor aggressiveness across cancers[18–28]. However, a significant methodological limitation with radiomic features, including GLCM, lies in the subsequent consolidation of the texture maps into aggregated feature metrics (e.g., mean, mode, kurtosis)[29]. Although these aggregated values seek to capture the 'image-based heterogeneity' associated with disease aggressiveness, they inherently lose the spatial information regarding the organization of different tissue types in a lesion and may not represent the distinct pathological variabilities within the same lesion[29].

For example, consider the radiomic expression maps in Fig. 1, which represent textural patterns derived from GLCM feature maps for two confounding pathologies (TuR, TrE). Visual inspection of these expression maps reveals discernible spatial differences in lesion organization, reflected by variations in the color distribution on the GLCM entropy expression heatmaps from the segmented ROI. In a way, these radiomic feature expression maps (Fig. 1h) are analogous to a subway map (Fig. 1j, k). The radiomic maps often represent overlapping regions where different cell populations interact, influencing heterogeneity, while subway maps demonstrate spatial intersections across transfer stations with multiple lines converging. Unfortunately, both machine learning (ML) and DL classifiers may not explicitly encode the spatial interactions of the 'sub-populations' (i.e., clusters) within the ROI for reliable distinction of confounding pathologies from malignant lesions. Consequently, there exists an opportunity to advance beyond aggregated feature values, by investigating intensity composition and complex interrelationships within the tumor microenvironment, at the spatial level. Such an approach could enable a comprehensive characterization of the lesion's inherent heterogeneity and improve the distinction of confounding pathologies from malignant tumors/cysts.

Recently, some studies have attempted to go beyond conventional radiomic analysis to explore structural connectivity within the lesion and how it impacts heterogeneity patterns and tumor behavior. For instance[30,31], explored the utility of graph theory approaches in characterizing primary and metastatic brain tumor heterogeneity by quantifying spatial interactions within intratumoral sub-regions. In addition[32], presented an automated pipeline that leverages structural connectivity graphs derived from T1 and DTI data, along with global graph metrics, to classify multiple sclerosis (MS) patients into four clinical subtypes. Similarly, our group previously developed a method called RADIomic Spatial TexturAl descripTor (RADISTAT) to analyze spatial interactions within tumor regions[29]. RADISTAT categorized voxels into predefined sub-regions of low, intermediate, and high radiomic expression and quantified pairwise adjacencies among these regions to assess tumor spatial organization in the context of brain and rectal cancer post-treatment response assessment. While RADISTAT provided an important step toward capturing spatial heterogeneity, it quantified organization using three predefined heterogeneity levels and pairwise adjacency counts, thereby constraining the analysis to local, fixed-resolution interactions.

In this work, we present a new Graph Radiomic Learning approach (GrRAiL) to comprehensively characterize imaging heterogeneity by (1) identifying clusters of sub-regions based on per-voxel radiomic measurements, followed by (2) computing graph-theoretic measurements that analyze the spatial associations of the radiomic clusters. The weighted graphs constructed via the GrRAiL descriptor capture the complexity of the higher-order spatial relationships within the lesion microenvironment to differentiate between confounding pathologies. A preliminary implementation of 3D GrRAiL was previously presented in ref. 33. Our current work builds on the initial work to include a detailed analysis of the GrRAiL algorithm with extended parameter sensitivity analysis. Additionally, we perform extensive comparisons of GrRAiL with state-of-the-art approaches in the context of clinically relevant problems in brain and pancreatic lesions. We summarize our key contributions as follows:

- Capturing the spatial organization of radiomic patterns within the lesion environment on structural MRI scans through clustering radiomic expression maps and generating graph theoretic features that characterize the spatial interactions of the radiomic patterns within the

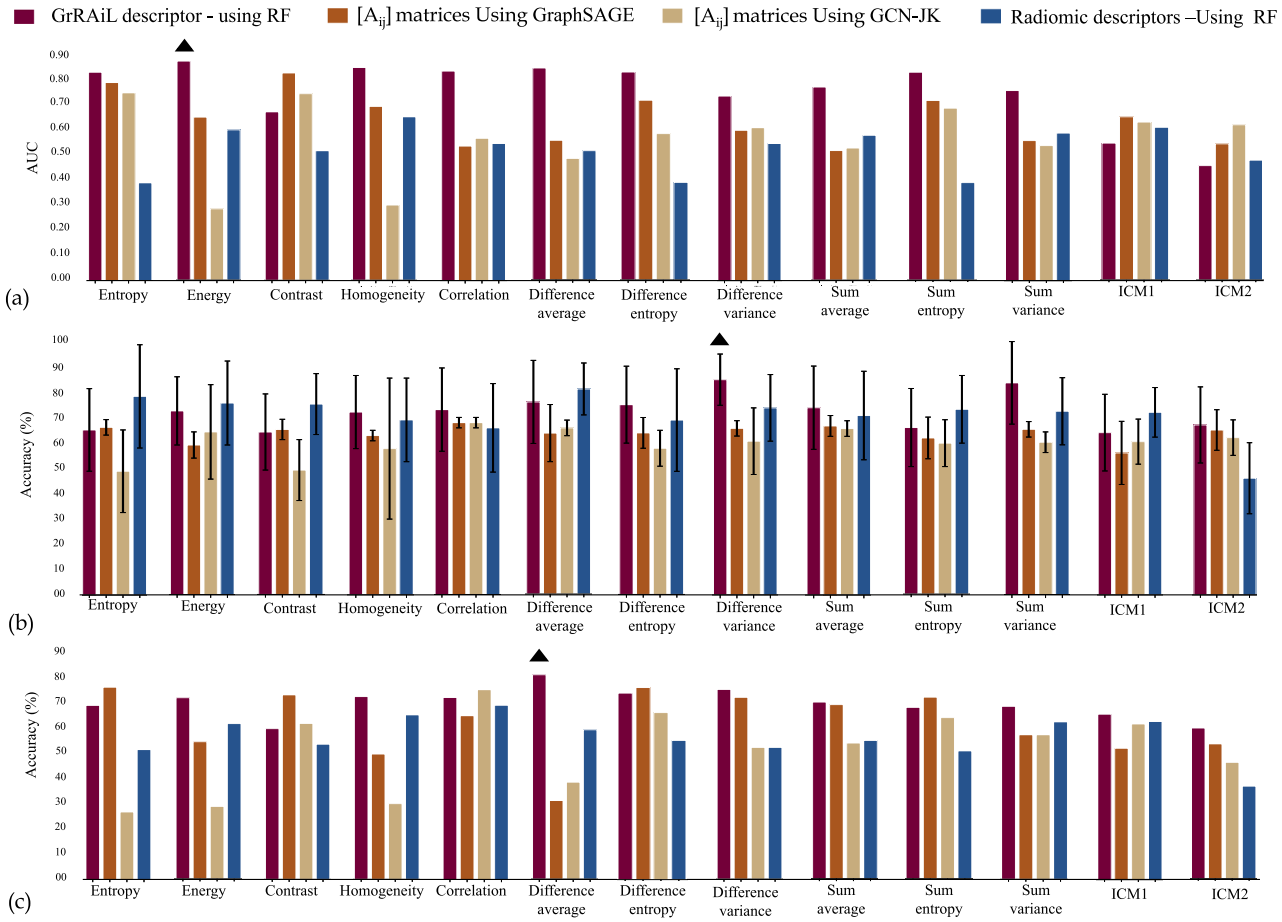

Fig. 2 | Classification performance metrics for GrRAiL, GraphSAGE, GCN-JK, and radiomic methods across 13 GLCM features in the context of glioblastoma tumors for the classification of TuR from PsP. The bar plots correspond to (**a**) AUC values, **b** cross-validation accuracy values on the training data, as well as (**c**) test accuracy values on unseen data from a different institution. Black upward-pointing triangles indicate the best-performing feature in each panel.

lesion. GrRAiL descriptor thus quantifies the spatial connectivity of "sub-populations" within the lesion (analogous to a subway system), and their influence on disease malignancy.

- Evaluating GrRAiL's efficacy across clinical challenges in: (a) differentiating tumor recurrence (TuR) from radiation necrosis (RN) in metastatic brain tumors after radiotherapy on a multi-institutional cohort (n = 233 studies), (b) distinguishing TuR from pseudo-progression (PsP) in glioblastoma on a multi-institutional cohort (n = 106 studies), and (c) Classifying Intraductal Papillary Mucinous Neoplasms (IPMNs), the most common pancreatic cysts, into no-risk + low-risk versus high-risk on a multi-institutional study comprising T2w MRI scans from n = 608 studies.

Results

Experiment 1: Distinguishing tumor recurrence from pseudo-progression in glioblastoma

GrRAiL analysis. Figure 2 illustrates qualitative 3D GLCM entropy feature maps alongside their corresponding GrRAiL-generated graphs for glioblastoma patients with lesions exhibiting TuR (a, d) and those exhibiting PsP (g, j). The GrRAiL features appear to be capturing the distinctive spatial patterns from the corresponding radiomic textural maps across PsP (i, l) and TuR (c, f). For instance, lesions characterized as PsP largely exhibited a markedly lower number of nodes and edges in the graph, suggesting greater homogeneity within the lesion microenvironment. Conversely, the larger n of nodes found in the graphs corresponding to patients with TuR reflected higher heterogeneity within the lesion microenvironment. These trends were consistently observed across both training and test cohorts of patients with PsP and TuR analyzed using GrRAiL.

Bar plots corresponding to the performance metrics (AUC, CV accuracy, and test accuracy) for each radiomic feature map are shown in Fig. 3. Among the 13 radiomic maps, the GrRAiL feature from the energy map achieved the highest AUC value (0.86), the difference variance map yielded the highest CV accuracy (83% ± 10%), and the difference average map yielded the highest test accuracy (78%), on the test set. The top-performing metrics in Fig. 3 are highlighted with a delta above the corresponding bar plot. The performance metrics of GrRAiL are shown in Fig. 4a, b. The AUC value was 0.85, the CV accuracy was 89% ± 7%, and the accuracy on the test set was 78%. The SHAP summary plot in Fig. 4c highlights the top 10 GrRAiL features that significantly contributed to the classification accuracies for distinguishing PsP from TuR. In Fig. 4d–h, we highlighted the features with p value ≤0.01, indicated with a single asterisk above the box plot. These features include the assortativity graph feature from the ICM2 expression map, modularity from the energy expression map, network entropy from the ICM2 expression map, average path length from the contrast expression map, clustering coefficient from the homogeneity expression map, and small-worldness from the homogeneity expression map. Comparison of GrRAiL test-set accuracy metric with other comparative approaches using z-statistics, and the corresponding p values are detailed in Supplementary Table 7.

Ablation studies

Assessment of graph-derived adjacency matrices. The adjacency matrices obtained for each graph from individual radiomic feature maps through

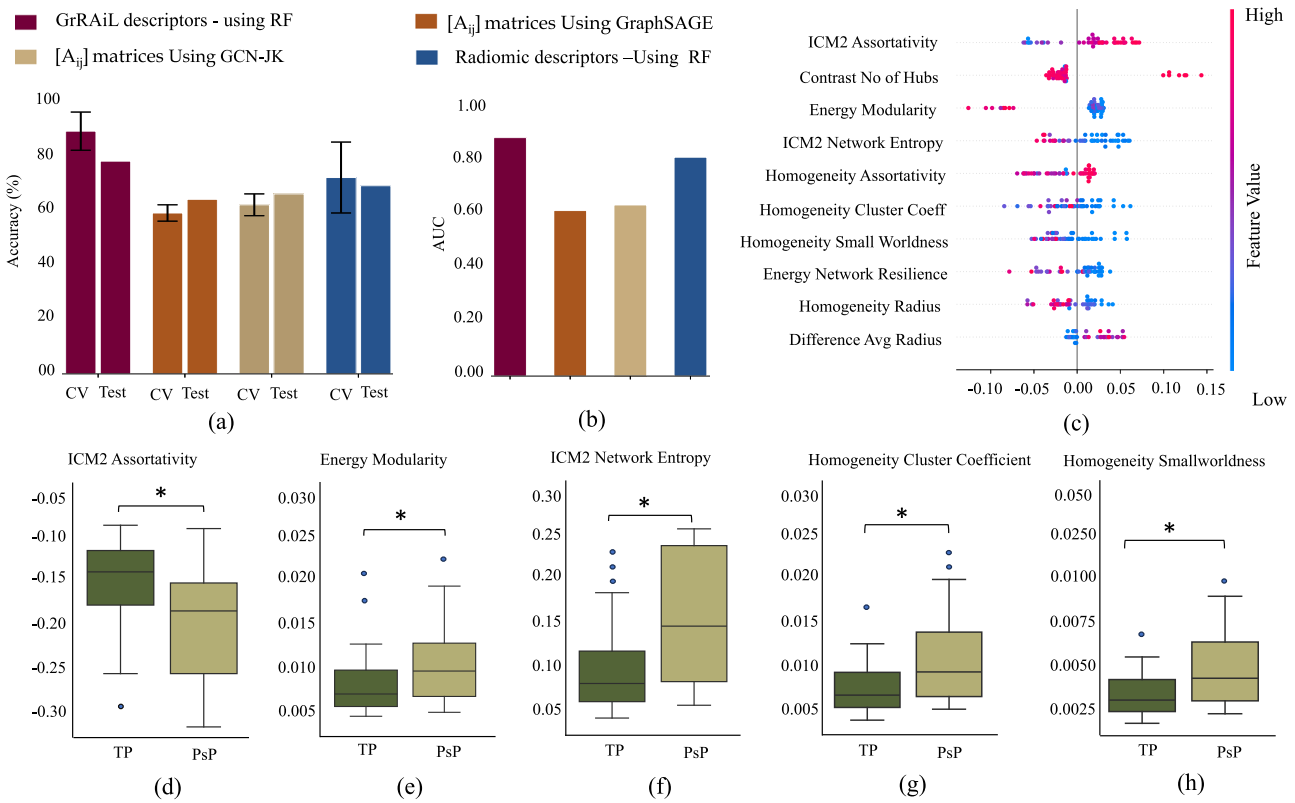

Fig. 3 | GrRAiL outperforms comparative graph- and radiomics-based approaches and identifies graph features that significantly discriminate tumor recurrence from pseudoprogression in glioblastoma. Bar plots with (**a**) CV and test accuracy, and **b** AUC values for graph features, GrRAiL, GraphSAGE, GCN-JK, and Radiomics features, **c** SHAP summary plot for the top 10 features contributing to the GrRAiL classification accuracy in distinguishing TuR from PsP in glioblastoma tumors. The remaining box plots are based on the statistical significance with $p \leq 0.01$ and include (**d**) ICM2 Assortativity, **e** Energy Modularity, **f** ICM2 Network Entropy, **g** homogeneity Cluster coefficient, and **h** homogeneity small worldness. A single asterisk indicates $p \leq 0.01$. Each feature demonstrated significant differences across TuR and PsP, highlighting the efficacy of GrRAiL features in differentiating true tumor progression from PsP.

GraphSAGE and GCN-JK neural networks using more than 2 convolutional layers (e.g., 5, 7) did not yield significant improvement; instead, the accuracy values decreased. Thus, these two convolutional layers were used for further analysis (Fig. 3). GraphSAGE achieved an AUC of 0.80 using the contrast map. The entropy map yielded a test accuracy of 74%, and the contrast map yielded the highest CV accuracy of 69.3% ± 11%. GCN-JK achieved an AUC of 0.73 using the entropy map. The correlation map yielded a CV accuracy of 66% ± 2%, and a test accuracy of 73%, as shown in Fig. 3. Furthermore, when the adjacency matrices from all 13 radiomic feature maps were combined and fed into GraphSAGE and GCN-JK, GraphSAGE achieved a CV accuracy of 59% ± 3%, its performance on the test set was 64% and an AUC of 0.59. The GCN-JK model yielded a CV accuracy of 62% ± 0.4%, a test accuracy of 66%, and an AUC of 0.61, suggesting that the model may be limited in its ability to capture discriminatory features across PsP and TuR (Fig. 4a, b). Results, when applying the GAT and GIN schemes, are provided in Supplementary Table 2. Both GAT and GIN yielded poor performance in distinguishing PsP from TuR, compared to GrRAiL.

Assessment of radiomic features. Performance metrics (AUC, CV accuracy, and test accuracy) for each radiomic feature map are depicted using bar plots in Fig. 3. Among the 13 radiomic maps, the radiomic descriptors from the homogeneity map achieved the highest AUC value (0.63), while the contrast map yielded the highest CV accuracy (79% ± 10%) and correlation map achieved a test accuracy of (68%). Additionally, when the radiomic descriptors from all 13 radiomic feature maps were combined into the classification model, the performance metrics showed an increase in the AUC and test accuracy values, as shown in Fig. 4a, b. The AUC value was 0.79, CV accuracy was 72% ± 13%, and the test accuracy was 69%. Overall, combining features from all 13 radiomic maps significantly enhanced the radiomic analysis performance but led to decreased performance for the adjacency matrices. When comparing individual analyses, GrRAiL outperformed both adjacency matrices and radiomic descriptors in the classification task.

Assessment of intensity graph analysis. The intensity graph analysis yielded a cross-validation accuracy of 69% ± 11%, a test accuracy of 59%, and an AUC value of 0.56, respectively. Both GrRAiL and the radiomic analysis was found to outperform the intensity graph analysis, with statistically significant improvements in accuracy (p value = 0.04, z-score = 2.019) (Supplementary Table 7). The performance metric values are provided in supplementary Table 1.

Experiment 2: Distinguishing tumor recurrence from radiation necrosis in metastatic brain tumors

GrRAiL analysis. Figure 5 illustrates qualitative 3D GLCM entropy feature maps alongside their corresponding GrRAiL-generated graphs for patients with brain metastases. Similar to our findings in glioblastoma tumors, GrRAiL effectively captured unique spatial patterns from the radiomic texture maps across RN and TuR in metastatic brain tumors. Lesions classified as RN typically exhibited fewer nodes and edges in their graphs, indicating a higher degree of homogeneity within the lesion microenvironment. In contrast, TuR graphs displayed a greater number of nodes and edges, reflecting increased heterogeneity in lesions that corresponded to tumor progression. These observations were consistent across both training and test cohorts analyzed using GrRAiL. Performance metrics (AUC, CV accuracy, and test accuracy) for each radiomic feature map are depicted using bar plots in Fig. 6. Among the 13 radiomic

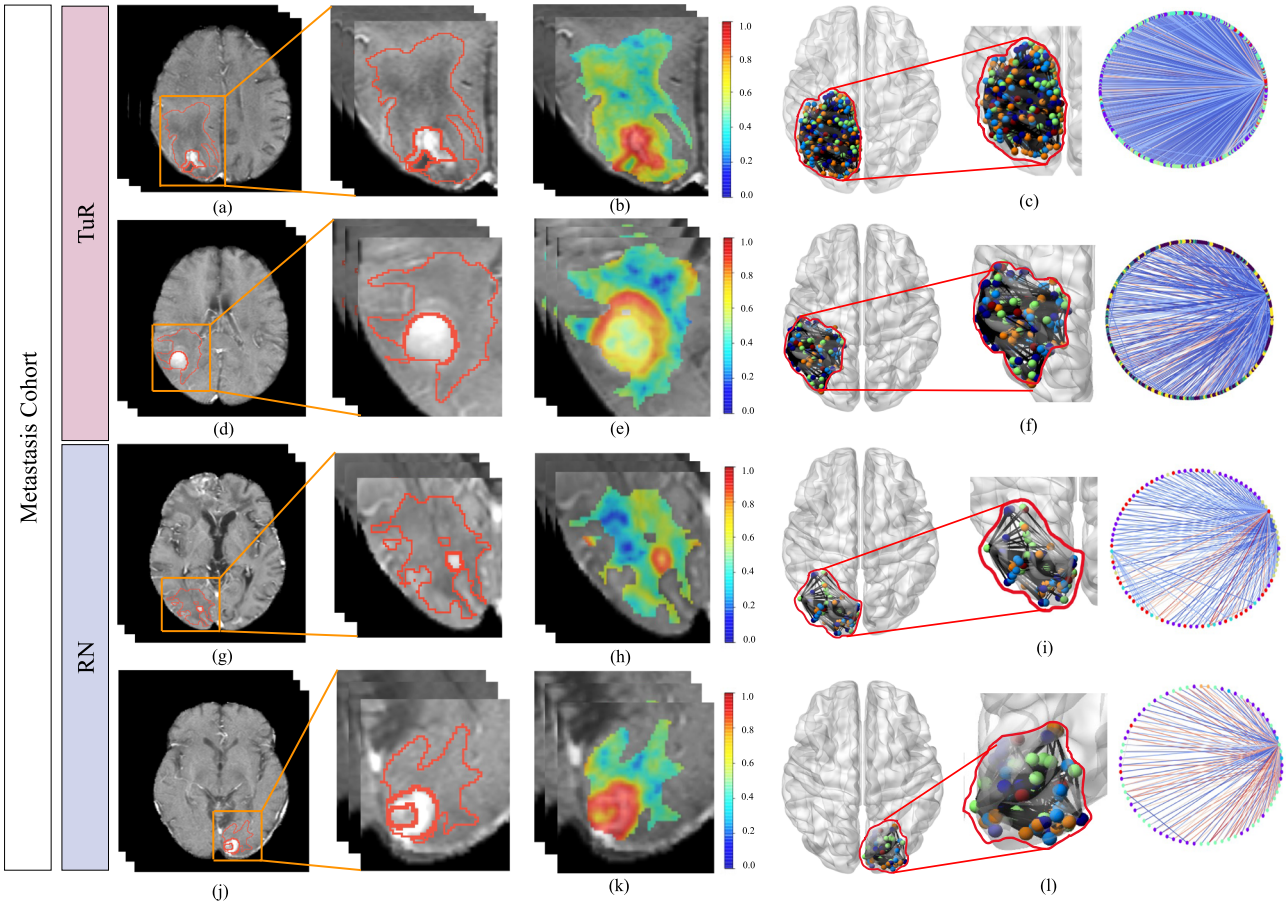

Fig. 4 | Representative metastatic brain tumor cases show distinct GLCM entropy expression maps and GrRAiL-generated graph structures for tumor recurrence versus radiation necrosis. Illustrates metastatic brain tumor cases comparing tumor recurrence (TuR; top two rows) and radiation necrosis (RN; bottom two rows). Panels **a**, **d** show Gd-T1w MR images and radiologist-provided lesion ROIs for TuR cases, while **g**, **j** show the corresponding images and ROIs for RN cases. Panels **b**, **e** and **h**, **k** show GLCM entropy maps within the lesion ROI, where blue color indicate lower radiomic feature expression and red color indicate higher radiomic feature expression. Panels **c**, **f** and **i**, **l** show the corresponding GrRAiL-generated graph visualizations. In these graphs, nodes represent GMM-derived radiomic subregions, node colors indicate distinct subregion labels, and edges represent EMD dissimilarity weights between adjacent subregions. The circular graphs shown at the right are topology visualizations to highlight differences in graph connectivity and complexity between TuR and RN cases.

maps, the GrRAiL features from the entropy map achieved the highest AUC (0.81), and the highest CV accuracy (75% ± 11%), and the contrast map yielded the highest test accuracy (78%). The top-performing metrics in Fig. 6 are highlighted with an delta above the corresponding bar plot. The performance metrics of GrRAiL descriptor are shown in Fig. 7a, b. We obtained an AUC of 0.86, CV accuracy of 84% ± 6%, and test accuracy of 74% in distinguishing RN from TuR in the metastatic brain tumor cohort. The SHAP summary plot in Fig. 7c highlights the top 10 GrRAiL features that significantly contributed to classification accuracy with a focus on features that yielded a p value of ≤0.001, indicated by double asterisks on the box plots. Statistical comparisons across GrRAiL and other comparative approaches using z-statistics and the corresponding p values, are provided in Supplementary Table 7.

Ablation studies

Assessment of graph-derived adjacency matrices. GraphSAGE achieved an AUC of 0.71 using the correlation map. The difference variance map resulted in a test accuracy of 62%, and the contrast map achieved the highest CV accuracy at 84% ± 16%. GCN-JK attained an AUC of 0.73 with the correlation map, a CV accuracy of 77% ± 8% using the sum variance map, and a test accuracy of 73% with the correlation map. However, when adjacency matrices from all 13 radiomic maps were combined and input into GraphSAGE and GCN-JK, the performance metrics declined. GraphSAGE's CV accuracy was 47% ± 5, its test accuracy dropped to 44% with an AUC of 0.55. GCN-JK gave a CV accuracy of 55% ± 7% and a test accuracy of 58%, and an AUC of 0.65, indicating relatively poor performance in distinguishing TuR from RN in metastatic brain tumors (Fig. 7). Results using the GAT and GIN schemes are provided in supplementary Table 4. The statistical comparison of test accuracies revealed that GrRAiL significantly outperformed GraphSAGE (p value = 0.0002, z-score = 3.786) and GCN-JK (p value = 0.04, z-score = 2.05) (Supplementary Table 7).

Assessment of radiomic features. Performance metrics for each radiomic feature map are shown in Fig. 6. Among the 13 radiomic maps, the contrast map's radiomic descriptors achieved the highest AUC of 0.69, the homogeneity map yielded the highest CV accuracy at 72% ± 7%, and the contrast map achieved a test accuracy of 66%. When we combined the radiomic descriptors from all 13 maps into the classification model, there was a slight increase in AUC and test accuracy, as shown in Fig. 7. The AUC improved to 0.79, CV accuracy was 78% ± 12%, and test accuracy was 61%. However, the overall performance remained lower compared to the GrRAiL features.

Assessment of intensity graph analysis. The intensity graph analysis yielded a cross-validation accuracy of 73% ± 11%, a test accuracy of 55%, and AUC value of 0.55. These performance metrics indicate that both GrRAiL as well as the radiomic analysis outperformed the intensity graph analysis. Besides, there was a statistically significant difference between the test accuracies of the GrRAiL approach and the intensity graph approach (p value = 0.017, z-score = 2.37) (Supplementary Table 7). The performance metric values are provided in Supplementary Table 3.

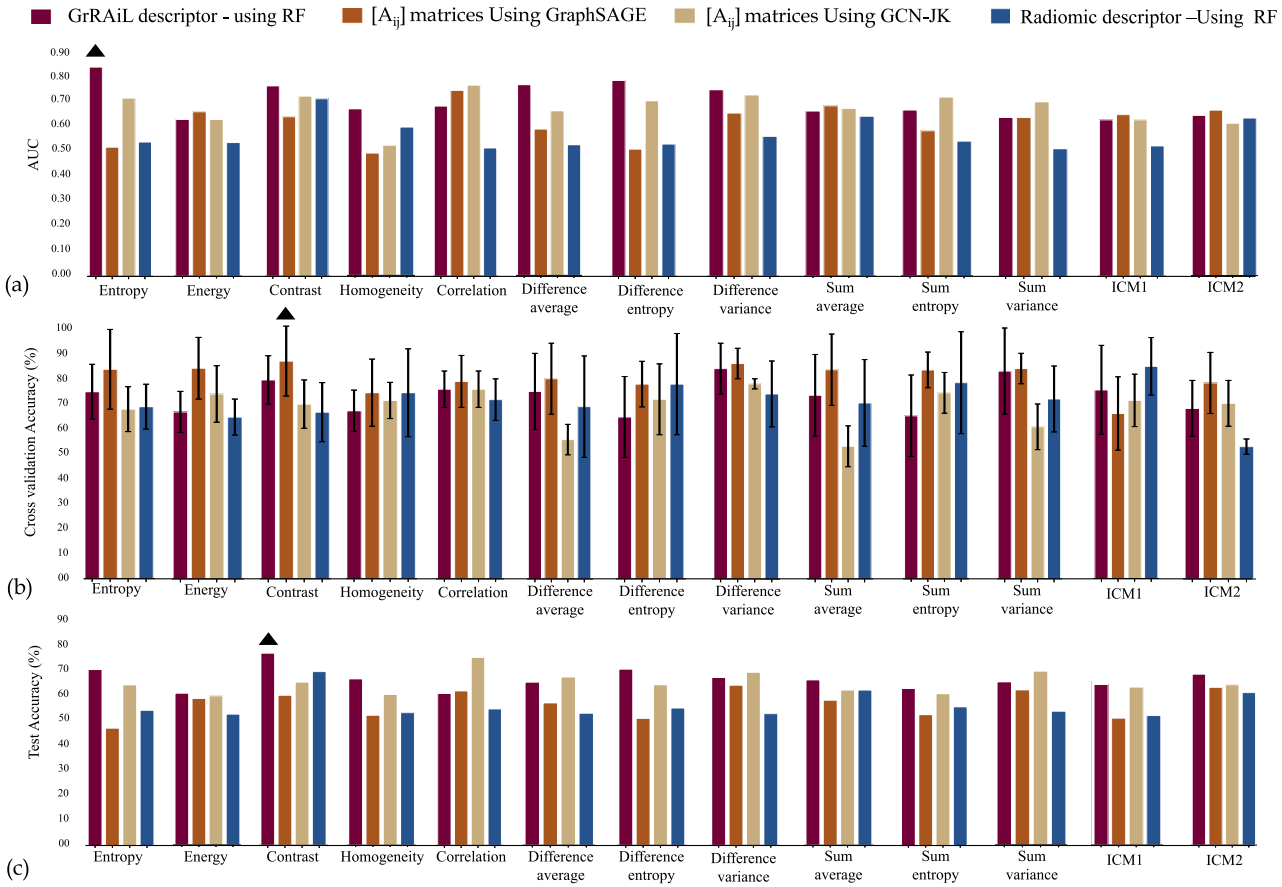

Fig. 5 | Classification performance metrics for GrRAiL, GraphSAGE, GCN-JK, and radiomic methods across 13 GLCM features in the context of metastasis brain tumors for the classification of TuR from RN. Bar plots of (**a**) AUC values, **b** cross-validation accuracy values on the training data, and **c** test accuracy values reflecting the model's ability to generalize to unseen data. Black upward-pointing triangles indicate the best-performing feature in each panel.

Experiment 3: Distinguishing no-risk + low-risk vs. high-risk pancreatic IPMN lesions

GrRAiL analysis. Figure 8 presents qualitative 3D GLCM entropy feature maps and their corresponding GrRAiL-generated graphs for patients categorized as having no-risk + low-risk versus high-risk IPMN lesions. Lesions in the no-risk + low-risk IPMNs category exhibited a lower number of nodes and edges in their graph representations, suggesting a more homogeneous tissue microenvironment. In contrast, patients classified as high-risk displayed graphs with a significantly larger number of nodes and edges, reflecting increased lesion heterogeneity.

Performance metrics (AUC, CV accuracy, and test accuracy) for each radiomic feature map are shown in Fig. 9. Among the 13 radiomic maps, the difference average map yielded the highest CV accuracy of 84% ± 4.3%, the homogeneity map achieved the highest test accuracy of 75%, and the difference variance maps showed an AUC of 0.79. The top-performing metrics in Fig. 9 are highlighted with an delta. The performance of GrRAiL is illustrated in Fig. 10, where we obtained an AUC value of 0.83, CV accuracy of 85% ± 4.5%, and test accuracy of 79%. Figure 10 presents the SHAP summary plot highlighting the top 10 GrRAiL features that contributed significantly to classification accuracy, with features demonstrating p values ≤0.001 (indicated by double asterisks above the box plots) in distinguishing no-risk + low-risk versus high-risk IPMN lesions. The z-statistics and the corresponding p-values towards statistical comparisons across GrRAiL test accuracy metric and other comparative approaches are detailed in Supplementary Table 7.

Ablation studies

Assessment of graph-derived adjacency matrices. As presented in Fig. 10, GraphSAGE achieved an AUC of 0.73 using the contrast map, while the difference average map resulted in the highest CV accuracy of 80% ± 4% and contrast map achieved a test accuracy of 73%. Similarly, GCN-JK achieved an AUC of 0.73 using the contrast map, while the difference variance map resulted in the highest CV accuracy of 80% ± 4% and the contrast map achieved a test accuracy of 72.2%. Interestingly, when adjacency matrices from all 13 radiomic maps were combined and input into GraphSAGE and GCN-JK, the performance declined, where GraphSAGE's CV accuracy dropped to 75.2% ± 5% and its test accuracy dropped to 65% with an AUC of 0.61. GCN-JK showed similar trends, with a CV accuracy of 73.3% ± 7.4%, a test accuracy of 68%, and an AUC of 0.57, demonstrating limited effectiveness when adjacency matrices from multiple radiomic maps were combined. Results using the GAT and GIN schemes are provided in Supplementary Table 6. GrRAiL was found to statistically outperform GraphSAGE for identifying high-risk from no-risk+low-risk IPMNs on the test-set (p-value = 0.02, z-score = 2.209) (Supplementary Table 7).

Assessment of radiomic features. Performance metrics for individual radiomic feature maps are shown in Fig. 9. Among the 13 radiomic maps, the energy map radiomic feature achieved the highest AUC of 0.75, while the difference average map yielded the highest CV accuracy of 85% ± 8%, and the entropy map achieved a test accuracy of 72.14%. When radiomic features from all 13 maps were combined, classification performances improved slightly. As shown in Fig. 10, the AUC increased to 0.76, the CV accuracy to 86% ± 5%, and the test accuracy to 71%. However, GrRAiL still outperformed radiomic feature-based classification in every metric.

Assessment of intensity graph analysis. The intensity graph analysis achieved a cross-validation accuracy of 83% ± 3.2%, a test accuracy of 61%, and an AUC of 0.52. These performance metrics indicate that both GrRAiL

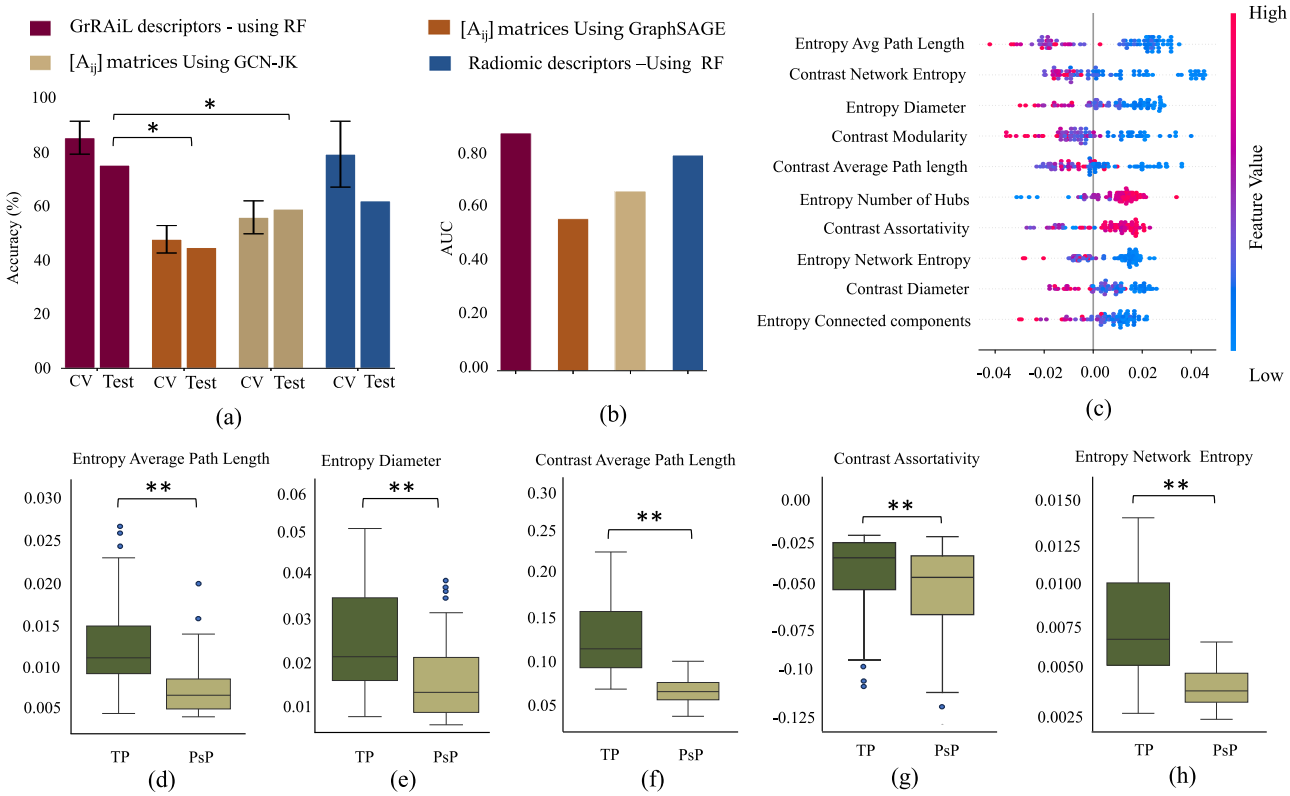

Fig. 6 | GrRAiL outperforms comparative graph- and radiomics-based approaches and identifies graph features that significantly discriminate tumor recurrence from radiation necrosis in metastatic brain tumors. a Bar plots with cross validation accuracy and test accuracy for GrRAiL, GraphSAGE, GCN-JK and Radiomics features from metastasis cohort. **b** Bar plots with AUC values for graph features, GraphSAGE, GCN-JK and Radiomics features from metastasis cohort. **c** SHAP summary plot for the top 10 features contributing to the GrRAiL classification accuracy in distinguishing TuR from RN in metastasis brain tumors. The remaining box plots are based on the statistical significance with $p \leq 0.0001$ and include (**d**) Entropy average path length, **e** Entropy diameter, **f** Contrast average path length, **g** Contrast assortativity and **h** Entropy network entropy. A double asterisk indicates $p \leq 0.0001$. Each feature demonstrates significant differences between TuR and RN, highlighting the efficacy of graph-based features in differentiating true tumor progression from RN.

and radiomic feature analysis significantly outperformed the intensity graph approach. Besides, there was a statistically significant difference between the test accuracies of the GrRAiL approach and the intensity graph approach (p value = 0.005, z-score = 2.78) (Supplementary Table 7). The performance metric values are provided in Supplementary Table 5.

Discussion

In this work, we presented a new descriptor called Graph-Radiomic Learning (GrRAiL), which leverages a combination of radiomic and graph-based attributes to quantify the intratumoral heterogeneity, towards distinguishing confounding pathologies from malignant lesions on MRI scans.

GrRAiL was employed to differentiate between tumor recurrence and treatment-induced radiation effects in glioblastoma (GB) and metastatic brain tumors, as well as to classify Intraductal papillary mucinous neoplasms, the most common cystic tumors in pancreatic cancer, according to their risk level (no+low versus high-risk). This was achieved by using radiomic expression maps that capture local heterogeneity within the tumor region, and then connecting these heterogeneity patterns through a graph, followed by extracting graph-theoretic measurements that construct GrRAiL, to classify the different pathologies addressed in this work. Our results showed that GrRAiL outperformed the state-of-the-art approaches, e.g., the graph-derived adjacency matrices, in classifying the different pathologies across brain and pancreatic lesions. Notably, our results support the hypothesis that the global graph features that GrRAiL comprises may capture the spatial interactions of lesion 'sub-populations', thus allowing for effectively representing heterogeneity within the lesion. Our findings demonstrated superior diagnostic performance compared to previous work in differentiating the different pathologies evaluated in this study, as detailed below.

It is worth contextualizing GrRAiL against our group's earlier RADISTAT descriptor, which also sought to capture spatial heterogeneity from radiomic expression maps. RADISTAT[29] quantified spatial organization by binning voxels into three predefined heterogeneity levels and counting pairwise adjacencies among these binned regions. GrRAiL differs from RADISTAT in three substantive ways (i) the number of radiomic clusters is selected per scan by BIC rather than consolidated to three clusters (ii) the graph is weighted by the 1-D Wasserstein distance between cluster-level radiomic value distributions rather than by binary pairwise adjacency counts and (iii) the resulting descriptor consists of global graph-topological features (density, modularity, small-worldness, network entropy, network resilience, and others) that characterize the entire tumor heterogeneity network rather than local pairwise interactions.

In distinguishing tumor recurrence versus pseudoprogression in glioblastoma, our results showed that Assortativity (Fig. 4d; p = 0.0049)-which measures the tendency of graph nodes to connect to other nodes with similar attribute values[34]-was higher in PsP than in TuR. In other words, clusters of voxels with similar GLCM ICM2 values in pseudo-progression lesions tended to form more tightly interconnected neighborhoods, which may reflect localized, treatment-related homogeneity. By contrast, the lower assortativity in true-progression graphs suggests that cluster centroids with differing texture characteristics are more likely to connect, consistent with a more heterogeneous lesion architecture. Similarly, Energy Modularity (Fig. 4e; p = 0.012)-a metric of how cleanly a network can be partitioned into distinct modules[35]-was elevated in PsP and reduced in TP. Higher modularity in pseudo-progression indicates stronger subnetwork organization,

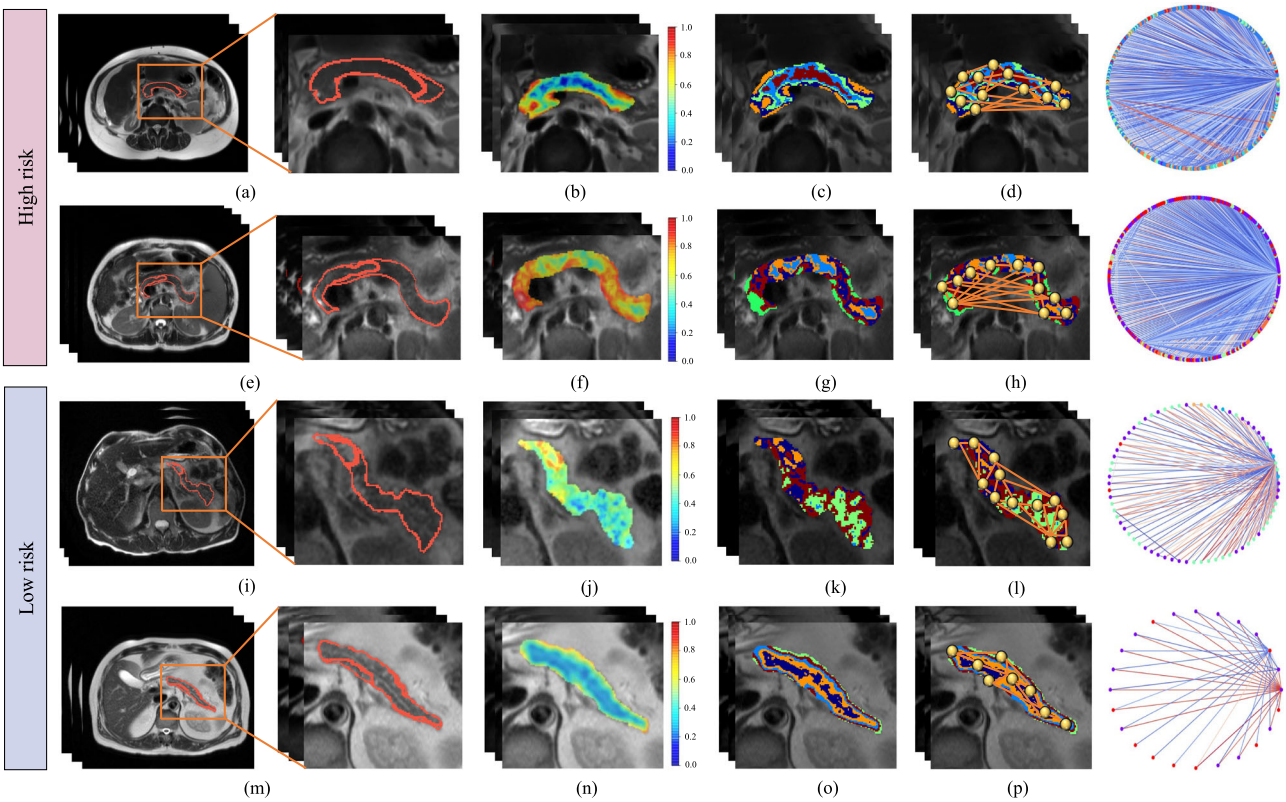

Fig. 7 | Representative pancreatic IPMN cases show distinct GLCM entropy expression maps, GMM-derived cluster maps, and GrRAiL-generated graph structures for high-risk versus no-risk/low-risk lesions. Illustrates pancreatic IPMN cases comparing high-risk IPMNs (top two rows) and low-risk/no-risk + low-risk IPMNs (bottom two rows). Panels **a**, **e** show T2w MR images and radiologist-provided ROIs for high-risk cases, while (**i**, **m**) show the corresponding images and ROIs for low-risk cases. Panels **b**, **f** and **j**, **n**) show GLCM entropy maps within the lesion ROI, where blue color indicate lower radiomic feature expression and red color indicate higher radiomic feature expression. Panels **c**, **g** and **k**, **o** show the corresponding GMM-derived cluster maps, where different colors represent distinct radiomic subregions. Panels **d**, **h** and **l**, **p** show the corresponding GrRAiL-generated graph visualizations. In these graphs, nodes represent GMM-derived radiomic subregions, node colors indicate distinct subregion labels, and edges represent EMD dissimilarity weights between adjacent subregions. The circular graphs shown at the right are topology visualizations to highlight differences in graph connectivity and complexity between high-risk and low-risk IPMN cases.

whereas the lower values in true progression could imply that the infiltrative tumor growth may disrupt clear graph communities, making module delineation more difficult. Finally, the Homogeneity Clustering Coefficient (Fig. 4g; $p = 0.008$)-which quantifies the degree of local interconnectedness among nodes sharing similar attributes[36]-was greater in PsP cases. This may suggest that treatment-affected regions form compact clusters of similar voxel intensities, while the more irregular and invasive growth patterns in TP lesions may lead to reduced local clustering.

Our results demonstrated that GrRAiL achieved a CV accuracy of 89%, a test accuracy of 78%, and an AUC value of 0.85 in classifying true progression from pseudo-progression in glioblastoma studies. While Patel et al.[37] focused on radiomic features from the multiparametric MRI scans and reported an AUC of 0.80, our approach combined graph features from all 13 radiomic maps, which significantly enhanced performance metrics. Specifically, combining these features increased the AUC to 0.85 in glioblastoma studies and an AUC of 0.86 in metastatic brain tumors. This indicates that integrating multiple radiomic-derived graph features may provide a more comprehensive characterization of tumor heterogeneity and improve the distinction of TuR from PsP. Other studies employing graph-theory approaches, such as Cao et al.[31], developed a multidimensional framework combining clinical factors, connectomics, and radiomics features to distinguish tumor recurrence from treatment effects, achieving promising results with an AUC of 0.89. In their approach, the tumor network was modeled by treating each voxel as a node and applying an empirical threshold of 0.1 to reduce unnecessary connections in the graph. However, this thresholding may have led to the loss of information by omitting weaker but potentially significant connections within the tumor network. Additionally, their model was trained and evaluated using leave-one-out cross-validation without testing on cross-institutional data, which may limit the generalizability of their findings. Lee et al.[30] explored the association of graph-based spatial features with the overall survival status of glioblastoma patients. They grouped the tumor region into high and low-intensity habitats based on MRI scans and constructed graphs by connecting these regions, focusing primarily on the center slice of the tumor for graph construction. By extracting graph features such as the minimum spanning tree (MST) and gray-level run-length matrix (GRLM), they achieved AUC values of 0.83 for MST and 0.77 for GRLM in predicting overall survival. While their study demonstrated the utility of graph-based features in a prognostic application, it was limited to a two-dimensional analysis of a single slice, which may not capture the full complexity of tumor heterogeneity.

In distinguishing tumor recurrence versus radiation necrosis in Metastatic Brain Tumors, our metastatic brain tumor graphs showed that the Entropy Average Path Length (Fig. 7d, $p < 0.0001$) and the Entropy Diameter (Fig. 7e, $p < 0.0001$) were both greater in the TP cohort than in the RN cohort. The average path length-calculated as the mean of all shortest path distances between cluster centroid nodes-was higher in TuR lesions, which may indicate that, on average, more graph steps are required to traverse between typical texture clusters. The diameter, defined as the single longest shortest path in the graph, was also larger in TuR, suggesting that the most distant pair of clusters in TuR lesions could be separated by a wider network span than in RN lesions. Together, these two measures-mean versus maximum shortest path distance-may reflect a more dispersed

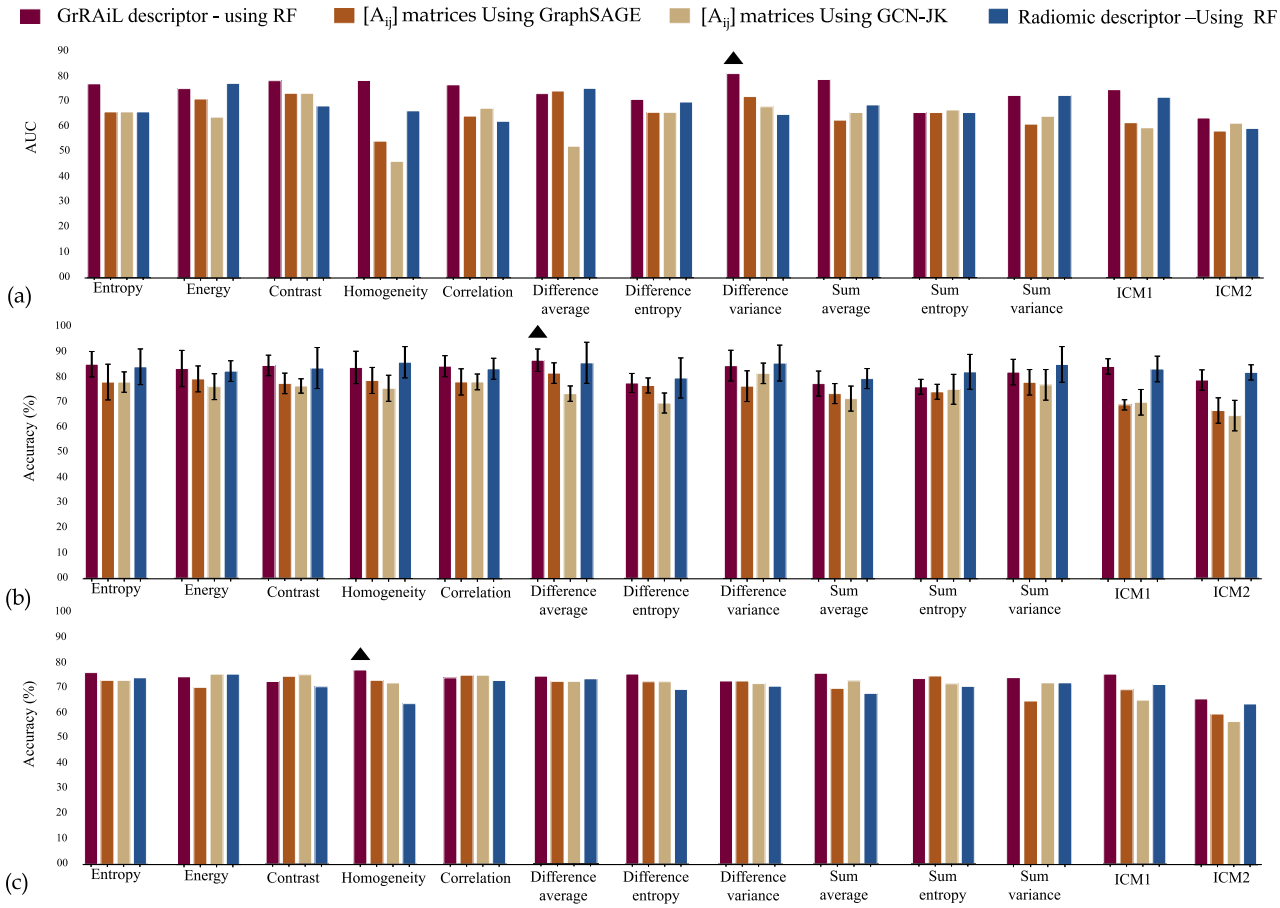

Fig. 8 | Classification performance metrics for GrRAiL, GraphSAGE, GCN-JK, and radiomic methods across 13 GLCM features in the context of IPMN for the classification of high risk from no + low risk lesions. Bar plots of (**a**) AUC values, **b** cross-validation accuracy values on the training data, and **c** test accuracy values reflecting the model's ability to generalize to unseen data. Black upward-pointing triangles indicate the best-performing feature in each panel.

arrangement of texture clusters in TP compared to the relatively compact network topology observed in RN[38].

Overall, GrRAiL demonstrated a descent performance in distinguishing tumor recurrence (TuR) from radiation necrosis (RN) in metastatic brain tumor cases, achieving a cross-validation accuracy of 84%, an external test accuracy of 74%, and an AUC of 0.86. Importantly, GrRAiL was validated using a multi-institutional dataset consisting of 332 lesions from 233 patients, underscoring its clinical relevance and potential generalizability. By contrast, many prior studies on this clinical issue have been single-institutional with limited external validation[39]. For instance, Basree et al.[40] reported an AUC of 0.76 using conventional radiomic features from a single-center cohort. Additionally[41], employed support vector machine (SVM) classifiers based on radiomic features from pre- and post-contrast T1-weighted and T2-FLAIR subtraction images in a two-center study involving 56 glioma patients, achieving a test accuracy of 93.3%, and an AUC of 0.94. While these results appear impressive[41], study faced several limitations, including a relatively small sample size, and the absence of external validation cohorts. In contrast, our approach not only leveraged advanced graph-based metrics to capture spatial heterogeneity but also benefited from a substantially larger dataset and external validation across multiple institutions.

In distinguishing no-risk + low-risk versus high-risk pancreatic intraductal papillary mucinous neoplasm lesions, contrast diameter, sum entropy number of hubs, and energy network resilience emerged as the top performing GrRAiL features to discriminate no + low versus high risk IPMNs ($p = 0.00022$, 0.0024, and 0.008, respectively). The box plot for contrast diameter shows higher values in the high risk group and lower values in the low risk group, which may reflect greater heterogeneity and spatial dispersion of texture clusters in high risk lesions. Similarly, a markedly larger sum entropy number of hubs was observed in high risk cases, suggesting a more heterogeneous and interconnected network structure, whereas low risk lesions tended to exhibit fewer hubs and potentially simpler topology. In contrast, energy network resilience was higher in the low risk class (Fig. 10e), which could indicate more uniform or robust connectivity patterns compared with the reduced resilience seen in high risk lesions. Notably, contrast derived graph features consistently ranked among the strongest discriminators across all three use cases in our study, suggesting they may effectively capture key aspects of tumor heterogeneity and its spatial organization.

Our GrRAiL framework achieved robust performance, demonstrating a cross-validation accuracy of 85%, external test accuracy of 79%, and an AUC of 0.83. Notably, GrRAiL was validated using a significantly larger and heterogeneous multi-institutional dataset (n=609), underlining its strong generalizability and applicability to real-world clinical scenarios. In contrast[42], utilized conventional radiomic features extracted from CT scans of a comparatively smaller, single-institution cohort ($n = 103$), obtaining an internal validation mean AUC of 0.77, improving slightly to 0.81 with the addition of clinical variables. Although[43] reported higher external validation AUCs (0.88 and 0.87) for predicting pathological grades of IPMN lesions using MRI radiomics, their analysis was constrained by a smaller dataset ($n = 202$), manual segmentation of lesions, exclusion of peritumoral regions, and fewer clinical parameters, which potentially limit scalability and broader generalizability. Thus, while[43] reported slightly higher AUC values, GrRAiL may offer stronger overall performance in terms of dataset size, generalizability, external validation rigor, and the integration of advanced graph-

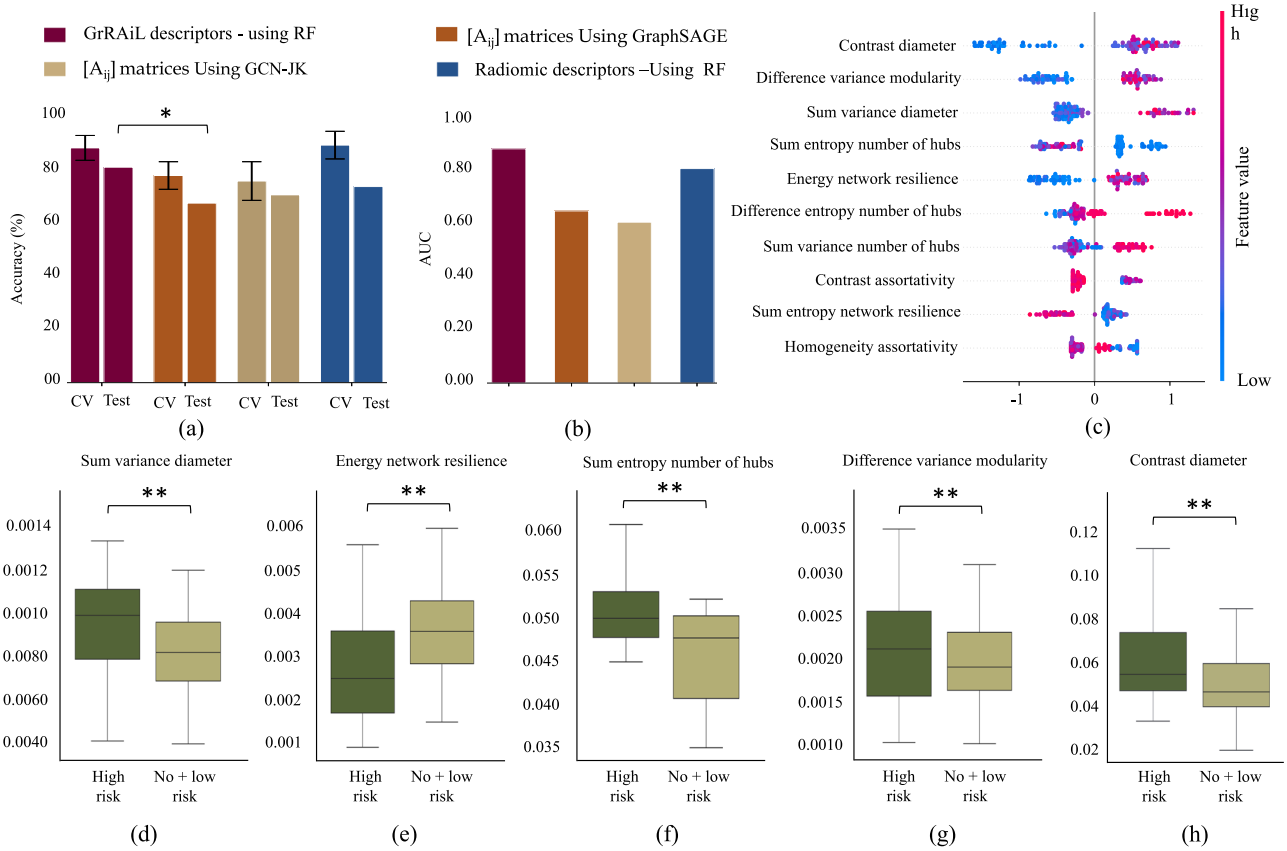

Fig. 9 | GrRAiL outperforms comparative graph- and radiomics-based approaches and identifies graph features that significantly discriminate high-risk from no-risk/low-risk IPMN lesions. a Bar plots with cross validation accuracy and test accuracy for GrRAiL, GraphSAGE, GCN-JK and Radiomics features from IPMN cohort. **b** Bar plots with AUC values for graph features, GraphSAGE, GCN-JK and Radiomics features from IPMN cohort. **c** SHAP summary plot for the top 10 features contributing to the GrRAiL classification accuracy in distinguishing high risk from No + low risk lesions in IPMN patients. The remaining box plots are based on the statistical significance with $p \leq 0.001$ and include (**d**) sum variance diameter, **e** energy network resilience, **f** sum entropy number of hubs, **g** difference variance modularity, and **h** contrast diameter. A double asterisk indicates $p \leq 0.0001$. Each feature demonstrates significant differences between high-risk and no + low-risk, highlighting the efficacy of GrRAiL features.

based metrics, potentially enhancing its clinical reliability and practical utility in predicting high-risk lesions.

Although GrRAiL consistently demonstrated improved diagnostic performance across different use-cases in brain and pancreatic neoplasms, there remains room for further improvement. In the current study, GrRAiL was developed using post-contrast T1w MRI for the brain tumor cohorts and T2w MRI for the IPMN cohort. Future work will explore the integration of multimodal and multiparametric imaging data, including T1w, T2w, FLAIR, diffusion, and perfusion imaging, which may further improve the diagnostic performance of the trained models. We also aim to expand the application of GrRAiL to additional cancer types and investigate advanced graph neural network architectures to improve the performance of models that directly use graph-generated adjacency matrices.

Beyond these clinical and modality-level extensions, the GrRAiL framework is not limited to GLCM feature maps. Since GrRAiL relies on clustering spatially resolved voxel-wise values and constructing weighted graphs over the resulting sub-regions, it may be extended to voxel-wise maps derived from other radiomic feature families, including first-order, GLRLM, GLSZM, GLDM, and NGTDM features, as well as quantitative imaging maps such as ADC, perfusion, T1/T2 relaxation, or PET-derived maps.

While we employed multi-institutional studies across different experimental validations of GrRAiL, future studies would involve the study pool to demonstrate consistent performance across institutions and disease domains. GrRAiL may also be less suitable for very small lesions, highly fragmented ROIs, poor-quality scans, or highly anisotropic acquisitions, where stable clustering and graph construction may not be supported and the resulting graphs may contain too few nodes or edges for reliable topology-based measurements. In particular, the IPMN cohort had thicker native through-plane resolution than the target isotropic grid used for 3D GLCM feature-map extraction. Although isotropic resampling enabled a unified 3D GrRAiL workflow across cohorts, it may introduce interpolated through-plane information in highly anisotropic abdominal MRI acquisitions and could influence voxel-wise 3D GLCM maps. Future work will therefore evaluate acquisition-specific preprocessing strategies, including coarser isotropic resampling, anisotropic feature extraction that preserves native slice spacing, and 2D in-plane GLCM extraction for cohorts with thick-slice imaging.

In addition, although the graph-based representation in GrRAiL may reduce sensitivity to certain global intensity shifts by summarizing the spatial organization and connectivity of clustered radiomic expression patterns, this does not fully eliminate scanner, acquisition, or preprocessing-related effects. ComBat harmonization was considered but was not applied in the primary analysis. Future work will evaluate ComBat and related harmonization strategies, including longitudinal ComBat and GAM-based ComBat, to further quantify the reproducibility of GrRAiL features across institutions[44].

In this study, we presented and evaluated the Graph-Radiomic Learning (GrRAiL) descriptor for distinguishing benign confounding pathologies from malignant neoplasms on clinical MRI scans. By leveraging graph representations of radiomic feature maps, GrRAiL seeks to capture

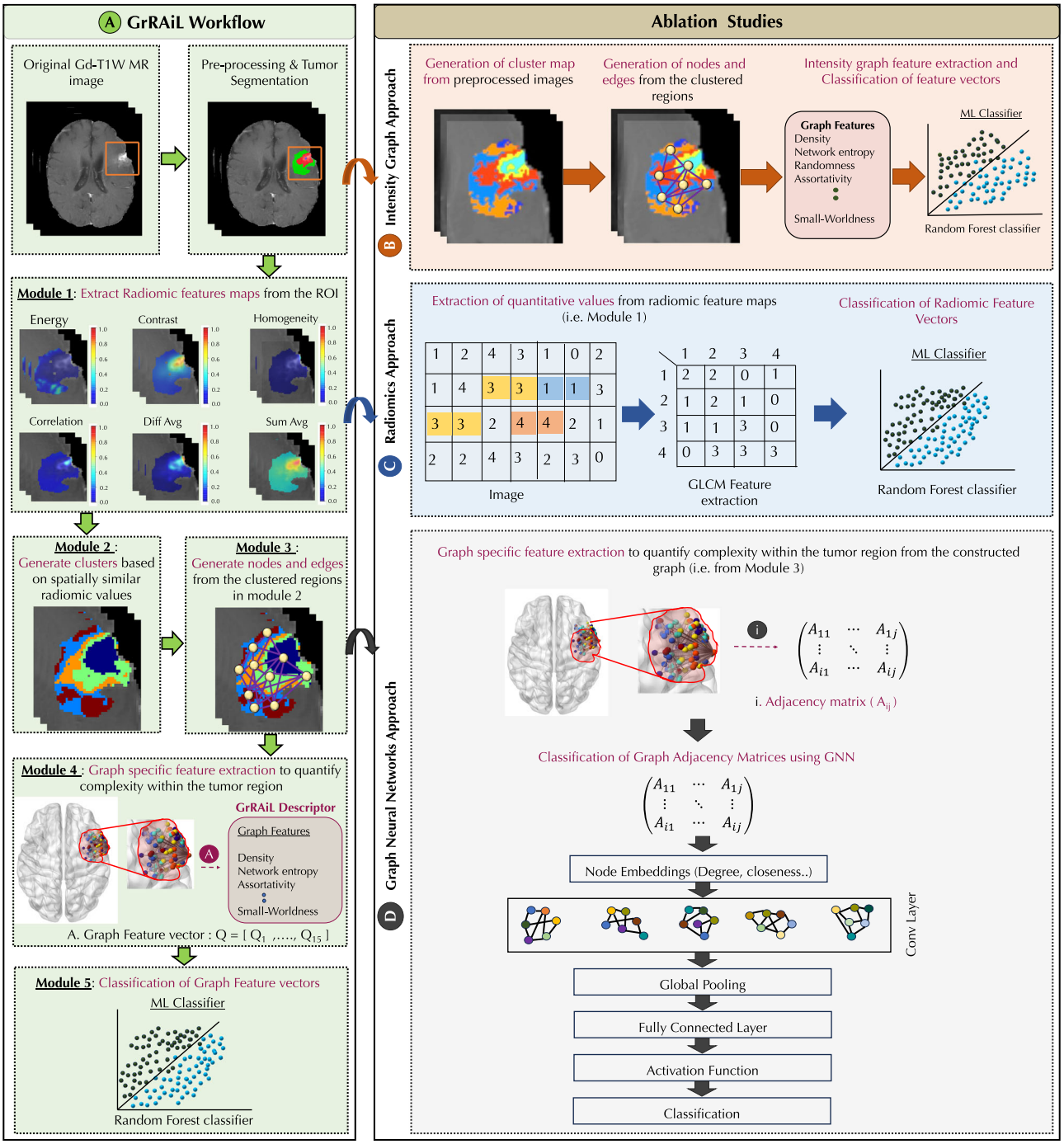

Fig. 10 | Workflow illustrating the GrRAiL framework and accompanying ablation studies. a GrRAiL Framework: (1) Radiomic feature extraction: Compute voxel wise gray level co-occurrence matrix (GLCM) feature maps within the segmented region of interest. Heatmap overlay (red = high expression, blue = low). (2) Subregion clustering: Apply Gaussian Mixture Modeling (GMM) to each radiomic map; select optimal cluster count via Bayesian Information Criterion (BIC). (3) Graph construction: Extract spatially connected radiomic sub-regions from each clustered map using 3D 26-neighborhood connectivity, represent each sub-region as a node, define edge weights as the 1-D EMD between the radiomic-value distributions of adjacent sub-regions (4) Graph theory feature extraction: Derive weighted graph metrics (e.g., node degree, clustering coefficient, path length) to quantitatively characterize spatial heterogeneity. (5) Classification: Perform feature selection on the graph theory metrics and train a classifier to distinguish confounding pathologies. Ablation Studies:**b**intensity graph approach: Apply GMM clustering directly on the raw ROI intensities (bypassing GLCM maps), construct graphs from those clusters, extract graph theory features as in Module 4, and classify via Random Forest. **c** Radiomics approach: Quantitative values were extracted from Module 1 GLCM feature maps were fed directly into a Random Forest classifier for classification. **d** Graph neural network approach: Follow the same clustering and graph construction pipeline (as in GrRAiL pipeline), but instead of extracting hand crafted graph metrics, input the full graph adjacency and node attribute matrices into a Graph Neural Network for direct classification.

the intricate spatial relationships within the lesion microenvironment, to be able to reliably distinguish confounding pathologies from malignant neoplasms. Our multi-institutional analysis across different use cases encompassing glioblastoma, brain metastasis, and IPMN lesions cohorts demonstrated that GrRAiL may serve as a surrogate image-based biomarker for distinguishing between recurrent brain tumors and radiation-induced treatment effects, as well as malignancies with no+low versus high-risk,

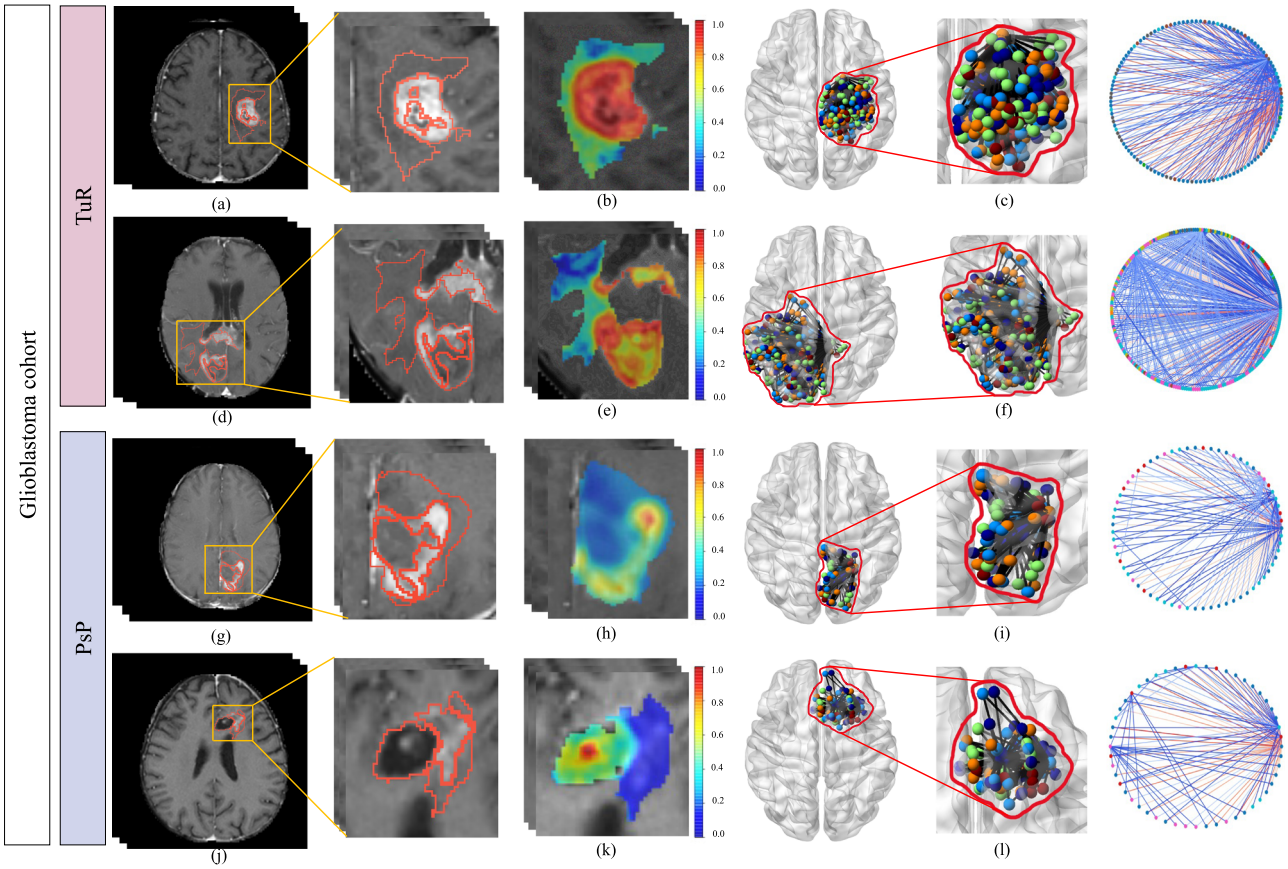

Fig. 11 | Representative glioblastoma cases show distinct GLCM entropy expression maps and GrRAiL-generated graph structures for tumor recurrence versus pseudoprogression. Illustration of glioblastoma cases from the primary brain tumor cohort comparing tumor recurrence (TuR; top two rows) and pseudoprogression (PsP; bottom two rows). Panels (**a**, **d**) show Gd-T1w MR images and radiologist-provided lesion ROIs for TuR cases, while (**g**, **j**) show the corresponding images and ROIs for PsP cases. Panels **b**, **e** and **h**, **k** show GLCM entropy maps within the lesion ROI, where blue colors indicate lower radiomic feature expression and red colors indicate higher radiomic feature expression. Panels **c**, **f** and **i**, **l** show the corresponding GrRAiL-generated graph visualizations. In these graphs, nodes represent GMM-derived radiomic subregions, node colors indicate distinct subregion labels, and edges represent EMD dissimilarity weights between adjacent subregions. The circular graphs shown at the right are topology-only visualizations that highlight differences in graph connectivity and complexity between TuR and PsP cases.

outperforming conventional radiomic and SOTA graph-neural network-based approaches, on clinical MRI scans.

Methods

Image notation

We define a GLCM feature expression map as $I = (K, f)$, where K denotes a spatial grid composed of voxels $k \in \mathbb{R}^3$. Each voxel $k \in K$ is associated with a specific value of the radiomic feature $f(k)$. The GLCM extraction parameters, including gray-level discretization, number of gray levels (N_g), window size (W), stride (s), voxel distance (δ), direction set, boundary handling, and directional aggregation strategy, are detailed in Section IIB. Figure 11 illustrates an overview of the GrRAiL framework. The notations and acronyms used throughout this paper are summarized in Table 1.

GLCM feature map extraction

Voxel-wise GLCM feature maps were computed within the lesion ROI using a custom Python implementation based on NumPy and SciPy, following Image Biomarker Standardisation Initiative (IBSI) definitions[45]. Image intensities were normalized and discretized using fixed-bin-number discretization into N_g gray levels. For each tumor voxel, a local $3 \times 3 \times 3$ cubic window was centered on the voxel and moved across the ROI with a stride of $s = 1$ voxel. Co-occurrence pairs were computed at a voxel distance of $\delta = 1$ across 13 non-redundant 3D directions corresponding to the 26-neighborhood. Counts from all directions were merged into a single symmetric local GLCM, normalized, and then used for feature calculation. The resulting voxel-wise GLCM maps were used for clustering and graph-construction. A total of 13 GLCM feature maps $I_1, \ldots, I_{13}$ were derived: energy, entropy, contrast, correlation, homogeneity, sum average, sum variance, sum entropy, difference entropy, difference average, difference variance, information measure of correlation 1 (ICM1), and information measure of correlation 2 (ICM2)[46].

Clustering of radiomic feature expression maps

The clustering of I is performed using a modified Gaussian Mixture Model (GMM) algorithm, to derive u clusters[47]. The objective is to robustly delineate spatially contiguous regions in the radiomic expression map based on similar-valued voxels in $\mathbb{R}^3$, in an unsupervised manner. Our modified GMM approach assumes that the voxel intensities are generated from a finite mixture of Gaussian distributions, each symbolizing a unique cluster. The probability that a voxel k with intensity vector x_k belongs to cluster u is defined by the following equation:

$$P(u|x_k) = \frac{\pi_u \mathcal{N}(x_k|\mu_u, \Sigma_u)}{\sum_{i=1}^{M} \pi_i \mathcal{N}(x_k|\mu_i, \Sigma_i)}, \quad (1)$$

where π_u denotes the prior probability of cluster u, and $\mathcal{N}(x_k|\mu_u, \Sigma_u)$ is the Gaussian probability density function with mean μ_u and covariance Σ_u for the u^{th} component. To determine the most suitable number of clusters for each individual radiomic expression map, we employ the Bayesian

Table 1 | List of the notations and acronyms used in this paper

Notation	Description	Notation	Description
I	Radiomic feature expression map (K, f)	$\hat{I}$	Clustered radiomic feature expression map (K, g)
N_g	Number of gray levels used for GLCM discretization	W	Local window size for voxel-wise GLCM computation
s	Stride used to move the local GLCM window	δ	Voxel distance used for GLCM co-occurrence computation
K	Spatial grid in $\mathbb{R}^3$	k	Voxel in K
$f(k)$	Radiomic feature value at voxel k	$g(k)$	Average radiomic feature value within the cluster of k
$\hat{K}_u$	Set of voxels in the u-th cluster	u	Number of clusters (via BIC)
M	Total number of GMM components	π_u	Prior probability of component u
μ_u	Mean vector of Gaussian component u	Σ_u	Covariance matrix of component u
$\mathcal{N}(x_k \mid \mu_u, \Sigma_u)$	Gaussian PDF for x_k under component u	$P(u \mid x_k)$	Posterior probability that voxel k belongs to u
D_{v_i}	Distribution of radiomic feature values within sub-region v_i	w_{ij}	Edge weight; 1-D EMD between distributions D_{v_i} and D_{v_j} of adjacent sub-regions.
$G = (V, E)$	Undirected graph of nodes and edges	A_{ij}	Adjacency matrix entry (1 if edge (i, j) exists)
$Q = [Q_1, \ldots, Q_{15}]$	Global graph-theoretic feature vector	$\mathbf{x}_k$	Intensity-feature vector at voxel k
I_i, $i = 1, \ldots, 13$	The i-th GLCM feature map (energy, entropy, …, ICM2)		
Acronyms			
Acronym	**Full Form**	**Acronym**	**Full Form**
GLCM	Gray-Level Co-occurrence Matrix	GNN	Graph Neural Network
GMM	Gaussian Mixture Model	GIN	Graph Isomorphism Network
BIC	Bayesian Information Criterion	GAT	Graph Attention Network
RAG	Region Adjacency Graph	GCN-JK	GCN with Jumping Knowledge
EMD	Earth Mover's Distance	SOTA	State-Of-The-Art
HIPAA	Health Ins. Portability & Accountability Act	T1w	T1-weighted (MRI)
IRB	Institutional Review Board	PsP	Pseudo-progression
TuR	Tumor Recurrence	RN	Radiation Necrosis
IPMN	Intraductal Papillary Mucinous Neoplasm	T2w	T2-weighted (MRI sequence)
MRI	Magnetic Resonance Imaging	PanSegNet	Pancreas Segmentation Network
ITK-SNAP	Insight Segmentation and Registration Toolkit - SNAP	CystX	Cystic Neoplasm eXchange dataset
NYU	New York University Langone Health	MCF	Mayo Clinic Florida
NWU	Northwestern University	AHN	Allegheny Health Network
MCA	Mayo Clinic Arizona	IU	Istanbul University Faculty of Medicine
EMC	Erasmus Medical Center	ROI	Region of Interest

Information Criterion (BIC). It is worth noting that the number of clusters (u) resulting from this step can vary as a function of the unique expression values in each I. Based on the GMM clustering, I is quantized to obtain a cluster map $\hat{I} = (K, g)$, where every $k \in \hat{K}_u \subset K$, $g(k)$ is the average radiomic feature value within cluster $\hat{K}_u$.

Graph construction

Our framework quantifies the clustered regions of radiomic expression levels in $\hat{I}$ to capture the spatial interactions among radiomic sub-regions. After Gaussian mixture model (GMM)-based clustering, each voxel is assigned to a radiomic cluster label. We then identify spatially connected radiomic sub-regions within each clustered map using standard 3D 26-neighborhood connectivity, in which two voxels are considered connected if they share a face, edge, or corner.

Using these sub-regions, we construct an undirected weighted region-adjacency graph (RAG) $G = (V, E)$, where V is the set of nodes, with each node representing one connected radiomic sub-region $\hat{K}_u \subset K$, and E is the set of edges. An edge $(v_i, v_j) \in E$ is created only when two sub-regions are directly adjacent in the 3D tumor volume. For each node v_i, we store the empirical distribution of normalized radiomic feature values within that sub-region, denoted as D_{v_i}. The weight of each edge is defined as the 1-D Earth mover's distance (EMD), between the radiomic-value distributions of the two adjacent sub-regions:

$$w_{ij} = \mathrm{EMD}\left(D_{v_i}, D_{v_j}\right). \tag{2}$$

The EMD cost is therefore defined in radiomic feature space rather than in Euclidean space. Spatial information is encoded implicitly through the RAG topology, since edges exist only between physically adjacent 3D sub-regions. The adjacency matrix of G is denoted by $A = [A_{ij}]$, where

$$A_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E, \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

Construction of GrRAiL descriptor

From the resulting G, the GrRAiL descriptor is obtained for every subject by computing the global features $\mathbf{Q} = [Q_1, \ldots, Q_N]$ for each radiomic map $I_1, \ldots, I_{13}$, where $N = 15$ is the total number of graph-theory based features extracted from the constructed graphs to quantify the structural properties of G. All global graph features are then concatenated across the extracted radiomic expression maps per subject. The corresponding features and their description are provided in Table 2.

Table 2 | Graph-based features and their descriptions

Feature	Description
Size	Number of nodes in the graph.
Density	Proportion of possible edges that are actual edges.
Diameter	Maximal distance between any pair of nodes.
Average Shortest Path Length	Average distance across all pairs of nodes.
Clustering Coefficient	$C = \frac{3 \times \text{numberoftrianglesinthegraph}}{\text{numberofconnectedtripletsofnodes}}$ Measures the degree of clustering among the nodes.
Modularity	Strength of division of a network into modules or clusters.
Small-worldness	Compares the clustering coefficient to the characteristic path length relative to a random graph.
Connected Components	Number of connected subgraphs in the graph.
Assortativity	Correlation between node degrees at either end of an edge.
Radius	The minimum eccentricity of any node, where eccentricity is the greatest distance between a node and any other node.
Global Efficiency	The average inverse shortest path length, measuring how efficiently information is exchanged over the network.
Network Entropy	Measures the randomness or disorder within the network structure.
Number of Hubs	Number of nodes with significantly higher degrees compared to other nodes, acting as central points in the network.
Randomness Value	Quantifies the unpredictability of the network's structure.
Network Resilience	The ability of the network to maintain connectivity after the removal of nodes or edges.

Algorithm 1. Computation of GrRAiL Descriptor

Input: Radiomic feature expression map $I = (K, f)$, where K is the spatial grid and f is the voxel-wise radiomic feature.

Output: Global graph-theory feature vector **Q**

Initialization :

1: Choose radiomic feature map I from $I_1, \ldots, I_{13}$ for graph extraction.

Clustering Process :

2: Generate u clusters in K using GMM, where $\hat{K}_u \subset K$, $u \in \{1, 2, 3, 4, 5\}$, and apply BIC to select optimal u.

Sub-region Identification :

3: For each cluster $\hat{K}_u$, extract spatially connected radiomic sub-regions using 3D 26-neighborhood connectivity. Let $\{v_1, v_2, \ldots, v_n\}$ denote the resulting sub-regions.

Distribution Estimation :

4: **for** each sub-region v_i **do**

5: Compute D_{v_i}, the empirical distribution of normalized radiomic feature values within v_i.

6: **end for**

Graph Construction :

7: Construct the weighted Region-Adjacency Graph (RAG) $G = (V, E)$:

- Create an edge $(v_i, v_j) \in E$ if sub-regions v_i and v_j are physically adjacent in the 3D tumor volume.
- Assign edge weight $w_{ij} = \text{EMD}(D_{v_i}, D_{v_j})$, the 1-D Wasserstein distance between the radiomic-value distributions of the adjacent sub-regions.

8: Store the adjacency matrix $A_{i,j}$ of G.

Graph Metrics Extraction :

9: Compute graph metrics $Q_1, \ldots, Q_{15}$ (e.g., modularity, average path length, diameter, network entropy, small-worldness).

10: Obtain graph features as a $1 \times N$ vector **Q** as described.

Save Output : 11: Save the metrics to a file.

12: **reture Q**

Data description

Our retrospective data curation across different use-cases was granted a waiver of informed consent by the institutional review board and was performed in compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations and in accordance with the Declaration of Helsinki. This multi-institutional retrospective study was approved by the University of Wisconsin-Madison Institutional Review Board: Glioblastoma and Metastasis cohorts (Approval No. 2023-0057) and IPMN cohort (Approval No. 2025-1191). The requirement for informed consent was waived for this retrospective analysis of de-identified data. A summary of patient characteristics, imaging parameters, and associated patient outcome groups across the 3 cohorts, is summarized in Table 3.

1) *Glioblastoma Cohort:* Our retrospective study consisted of MRI scans of 106 patients: 46 studies from Dana-Farber Cancer Center and 60 studies from the Cleveland Clinic. University of Wisconsin-Madison Institutional Review Board (2023-0057); informed consent was waived for this retrospective analysis. The studies from Cleveland Clinic (TuR: $n = 38$, PsP: $n = 22$) were used as the training set, while those from the Dana-Farber Cancer Center (TuR: $n = 33$, PsP: $n = 13$) served as the test set. Glioblastoma patients who underwent chemoradiation treatment following the Stupp protocol and presented with a suspicious enhancing lesion within three months of treatment completion were included in the study. Each patient had a post-treatment Gadolinium-enhanced T1w MRI (Gd-T1w MRI), along with histological reports available for disease confirmation[48]. To address known intensity discrepancies, all brain T1w MRI scans were processed for bias field correction, skull-stripping, and intensity normalization. The manual segmentations for enhancing lesions were obtained on every MRI slice that displayed more than 5 mm of rim enhancement, via consensus across two collaborating radiologists to minimize inter-reader variability. Ground truth definition for PsP or TuR was based on histologic confirmation via surgical resectionor multiple biospies.
2) *Metastatic Brain Tumor Cohort:* We collected $n = 233$ patient MRI studies of metastatic brain tumor patients: $n = 109$ from Cleveland Clinic, $n = 48$ from University Hospitals (Cleveland), and $n = 76$ from the University of Wisconsin-Madison. Of the $n = 233$ studies, some metastasis patients had more than one tumor, and thus we evaluated a total of 332 lesions across $n = 233$ studies. For the training set, 157 studies from Cleveland Clinic and the University Hospitals at Cleveland were used, comprising 246 lesions: Cleveland Clinic (TuR: $n = 74$, RN: $n = 86$) and University Hospitals at Cleveland (TuR: $n = 38$, RN: $n = 48$). For the test set, 76 studies from the University of Wisconsin-Madison Hospital were used, comprising 86 lesions (TuR: $n = 38$, RN: $n = 48$). The inclusion criteria were: (1) patients treated with surgery and radiation therapy, (2) suspicious lesions on follow-up MRI before second surgical resection, (3) availability of surgically resected histopathological sections used for confirmation of disease or

Table 3 | Summary of patient characteristics, imaging parameters, and outcome groups in discovery and validation cohorts

	Metastatic brain tumor		Glioblastoma		IPMN	
	Train (*n*=157)	Test (*n*=76)	Train (*n*=60)	Test (*n*=46)	Train (*n*=509)	Test (*n*=99)
Patient Characteristics						
Sex, men:women:n/a	29:38:90	42:34	39:21	30:16	232:277	48:51
Age, mean (range), years	56.85 (44–67)	64.5 (55–74)	60.6 (26–74)	55.6 (25–76)	65 (52–78)	60 (46–74)
Imaging Parameters						
MRI Protocol	Gd-T1w	Gd-T1w	Gd-T1w	Gd-T1w	T2w	T2w
In-plane resolution (mm)	1.00	1.00	1.00	1.00	1.00	1.00
Slice thickness (mm)	0.8–10	1–2.40	1.00	1.00	4.5–4.6	5.6–6.1
Repetition time (msec)	7.6–2200	7.7–2100	263–764	263–850	NA	NA
Echo time (msec)	2.45–23	2.84–6	5–20	5–20	—	—
3T	54	33	0	0	NA	NA
1.5T	95	43	60	46	NA	NA
1T	8	0	0	0	NA	NA
Lesion Pathology						
Tumor Recurrence	112	38	37	33	–	–
Treatment Effects	134	48	23	13	–	–
No risk + Low risk	–	–	–	–	394	72
High risk	–	–	–	–	115	27

* In metastatic brain tumor studies, a single patient can have more than one lesion. Lesion-pathology counts are per scan. See data description for details.

treatment effects, and (4) availability of post-contrast T1w, T2w, and FLAIR MRI scans. Patients were excluded if (1) there were any artifacts in the MRI scans, and (2) lesions were less than 5 mm in the largest axis. Patients receiving planned concurrent stereotactic radiosurgery (SRS) or whole-brain radiotherapy (WBRT) with immune checkpoint inhibitors (ICI) were included[49].

3) *IPMN Cohort:* Our retrospective multicenter IPMN study included n = 608 patients with MRI scans (with both T1w and T2w) from seven centers. Participating centers in this study were the following: New York University (NYU) Langone Health (n = 147), Mayo Clinic Florida (MCF) (n = 125), Northwestern University (NWU) (n = 175), Allegheny Health Network (AHN) (n = 12), Mayo Clinic Arizona (MCA) (n = 15), Istanbul University Faculty of Medicine (IU) (n = 62), and Erasmus Medical Center (EMC) (n = 72). Following pathological confirmation, cases were split across three categories: no-risk (n = 147 cases), low-risk IPMN (n = 319 cases), and high-risk IPMN (n = 142 cases). Clinical information and MRI specifications of the scans are summarized in Table 3. Of the 608 total patients, 509 from IU (n = 62), MCF (n = 125), NWU (n = 175), and NYU (n = 147) were used for training, whereas the remaining 99 from AHN (n = 12), MCA (n = 15), and EMC (n = 72) served as the test set. This training-testing procedure was performed for no-risk+ low-risk vs high-risk classification task. To segment the pancreas, we developed a semiautomatic method in two stages. In the first round, we implemented a state-of-the-art segmentation algorithm named 'PanSegNet'[50]. For this training, five radiologists, each representing one of the five centers, manually segmented the pancreas on axial 385 T1w and 382 T2w scans from 499 adult patients across the centers. Before the annotation process, the radiologists established a protocol to standardize their approach, and all used the same software, ITK-SNAP[51] for these annotations[52]. In the second round, a total of 341 T1W and 364 T2-weighted (T2W) scans from NU and two additional centers, IU and EMC, were automatically segmented using PanSegNet. Two abdominal radiologists then reviewed and corrected the segmentations to establish the ground-truth segmentations for the second round segmentations.

Pre-processing and radiomic feature map extraction

All imaging data and corresponding annotations were resampled to a uniform isotropic resolution of $1 \times 1 \times 1$ mm^3 in the X, Y, and Z axes before voxel-wise 3D GLCM feature-map extraction. Image volumes were resampled using linear interpolation, whereas segmentation annotations were resampled using nearest-neighbor interpolation to preserve discrete label values. This common spatial grid was used to standardize voxel spacing across subjects and institutions because the downstream GrRAiL pipeline relies on local 3D texture neighborhoods, GMM-derived cluster centroids, inter-cluster spatial relationships, and graph-theoretic measurements. For the IPMN cohort, the native slice thickness was larger than the target isotropic spacing, ranging from 4.5–4.6 mm in the training cohort and 5.6–6.1 mm in the test cohort. Therefore, resampling was used for spatial standardization and was not assumed to introduce new anatomical information in the through-plane direction. For each region of interest (ROI), 13 GLCM (Gray Level Co-occurrence Matrix) features[46] were derived on a voxel-wise basis, employing a local $3 \times 3 \times 3$ sliding window to capture intensity co-occurrences across all spatial dimensions. These GLCM features, frequently utilized in radiomic studies[53], produced 13 distinct voxel-wise GLCM radiomic feature maps for each ROI, labeled I_1 through I_{13}.

GrRAiL development

The algorithm was implemented in Python to compute $\mathbf{Q}$ for each $I_1, \ldots, I_{13}$. The features were specifically selected to capture the 15 global attributes of the graph, per radiomic expression map, with the number and type of features kept consistent across all experiments. The extracted graph-theoretic measurements ($15 \times 13 = 195$) are concatenated to form the GrRAiL descriptor, then evaluated for different diagnostic tasks using machine learning classifiers. Finally, to assess the impact of gray-level discretization on feature extraction and model performance, we performed a parameter sensitivity analysis by varying the number of discrete intensity levels, $N_g \in \{4, 16, 64\}$, used to bin the continuous voxel intensities..

To identify the optimal number of clusters to spatially segment radiomic expression maps using a GMM, we used the BIC in different

cluster combinations, ranging from 1 to 12. Empirical analysis consistently identified that $u = 5$ clusters yielded the highest predictive performance across different use-cases, while maintaining reduced computational load and time requirements.

Experimental setup

For all analyses, the classification models were trained and optimized using fivefold cross-validation on the training cohort, and their performance was subsequently evaluated on a hold-out test cohort. Specifically, we performed feature selection followed by random forest (RF) classification to distinguish the different pathologies and obtain performance metrics. Recursive feature elimination[54] was applied to the GrRAiL descriptor to obtain uncorrelated features. These features were then employed in an RF classifier on the training set to distinguish the different pathologies. The top features resulting from the training model were then used on the test set for the classification task. Performance metrics, including the area under the receiver operating characteristic curve (AUC), cross-validation (CV) accuracy, and test accuracy, were used to assess the diagnostic performance of GrRAiL. To evaluate the contributions of different features that comprise GrRAiL, we performed SHapley Additive exPlanations (SHAP) analysis[55] on the graph features to identify the top 10 descriptors that contributed most significantly to the model's performance. We conducted statistical analysis using the Mann-Whitney U test on the SHAP-identified features to provide insights into the discriminatory ability of different features within the GrRAiL descriptor, across different use-cases. Additionally, statistical differences in model performance metrics (test accuracy) between GrRAiL and comparative methods were evaluated using a two-tailed z-test, to determine statistical significance of observed performance differences.

Ablation studies

Graph neural networks for adjacency matrices analysis. A weighted region adjacency graph (RAG) was created from K clusters for each radiomic feature map, with edge weights computed using the EMD. From the constructed graph $G = (V, E)$, the adjacency matrix $A = [A_{ij}]$ was computed, resulting in a total of 13 adjacency matrices (A_1, …, A_{13}) derived from I_1, …, I_{13}, respectively. The resulting adjacency matrices were evaluated using Graph Neural Network algorithms.

We conducted multiple experiments to identify the optimal approach for processing the generated adjacency matrices from the constructed graphs using several state-of-the-art (SOTA) Graph Neural Network (GNN) algorithms. Specifically, we selected four GNNs for graph-level classification: GraphSAGE[56], Graph Isomorphism Network (GIN)[57], Graph Attention Network (GAT)[58], and Graph Convolutional Network with Jumping Knowledge (GCN-JK)[59]. To ensure a fair comparison, we maintained consistent training hyperparameters and loss functions across all models. GraphSAGE (Graph Sample and Aggregation) generates node embeddings by sampling and aggregating features from a node's local neighborhood, enabling generalization to unseen nodes and dynamic graphs. In our study, we used initial node features derived from several graph properties, such as node degrees, closeness centrality, clustering coefficients, and average neighbor degree. The GraphSAGE model then used these embeddings for classification. The GIN, known for its strong discriminative power, closely mimics the Weisfeiler-Lehman graph isomorphism test. The GAT employs attention mechanisms to assign varying importance to neighboring nodes, enabling it to focus on the most relevant parts of the graph. GCN-JK incorporates a Jumping Knowledge mechanism, which aggregates information from various layers of the network. For GraphSAGE and GCN-JK, the 13 adjacency matrices were aggregated and directly input into the neural network models to derive performance metrics that were similarly computed for GrRAiL classification.

GLCM radiomics analysis. To assess GrRAiL's performance, a comparative assessment that employs GLCM-based features in a voxel-wise manner was conducted. Similar to the GrRAiL approach, recursive feature elimination was applied to the voxel-wise radiomic descriptors extracted from individual radiomic feature maps (I_1, …, I_{13}) to obtain uncorrelated features. Then, for each of the 13 voxel-wise GLCM features, we calculated five statistical descriptors: mean, median, standard deviation, kurtosis, and skewness[60,61]. These descriptors capture the distribution of radiomic features across the ROI, resulting in a total of 65 voxel-wise descriptors. These features were then employed in an RF classifier on the training set to distinguish the different pathologies. The top features resulting from the training model were then used on the test set for the classification task.

Intensity graph analysis. For this experiment, we performed clustering directly on the tumor region using a GMM rather than generating radiomic expression maps first, then performing clustering. BIC was used to determine the optimal number of clusters, ranging between 1 and 5. Nodes in the graph represent the centroids of each cluster, while edges were computed using the EMD, resulting in a graph that encapsulates diverse patterns of heterogeneity from the intensity clusters[62]. From the constructed weighted graph, we extracted 15 graph theory-based features to evaluate the graph architecture and quantify lesion complexity.

Data availability

The MRI datasets for the three study cohorts—glioblastoma, metastatic brain tumors, and intraductal papillary mucinousneoplasms (IPMN)—were acquired under institutional approvals at multiple participating sites. Because of institutional and patient-privacy protections, raw imaging and associated clinical data cannot be publicly released. Upon reasonable request, a data sharing agreement can be initiated between the interested parties and the clinical institution following institution-specific guidelines.

Code availability

The GrRAiL implementation will be made publicly available upon publication to support reproducibility and facilitate application of the proposed descriptor to independent datasets. The public repository will include the core code for voxel-wise radiomic feature map processing, graph construction and GrRAiL descriptor extraction. Until publication, the code may be made available to qualified researchers upon reasonable request from the corresponding author.

Received: 24 October 2025; Accepted: 2 August 2026;
Published online: 11 September 2026

References

1. Verhaak, R. G. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, AND NF1. *Cancer cell* **17**, 98–110 (2010).
2. Zhou, Q., Xue, C., Ke, X. & Zhou, J. Treatment response and prognosis evaluation in high-grade glioma: An imaging review based on mri. *J. Magn. Reson. Imaging* **56**, 325–340 (2022).
3. Pope, W. B. Brain metastases: Neuroimaging. *Handb. Clin. Neurol.* **149**, 89–112 (2018).
4. Derks, S. H., van der Veldt, A. A. & Smits, M. Brain metastases: The role of clinical imaging. *Br. J. Radiol.* **95**, 20210944 (2022).
5. Mitsuya, K. et al. Perfusion weighted magnetic resonance imaging to distinguish the recurrence of metastatic brain tumors from radiation necrosis after stereotactic radiosurgery. *J. Neuro-Oncol.* **99**, 81–88 (2010).
6. Mayo, Z. S. et al. Radiation necrosis or tumor progression? A review of the radiographic modalities used in the diagnosis of cerebral radiation necrosis. *J. Neuro-Oncol.* **161**, 23–31 (2023).
7. Chakraborty, J. et al. Preliminary study of tumor heterogeneity in imaging predicts two year survival in pancreatic cancer patients. *PloS one* **12**, e0188022 (2017).
8. Gerlinger, M. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).

9. Sottoriva, A. et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* **110**, 4009–4014 (2013).
10. Zhou, M. et al. Radiologically defined ecological dynamics and clinical outcomes in glioblastoma multiforme: Preliminary results. *Transl. Oncol.* **7**, 5–13 (2014).
11. Zhou, M. et al. Identifying spatial imaging biomarkers of glioblastoma multiforme for survival group prediction. *J. Magn. Reson. Imaging* **46**, 115–123 (2017).
12. Wu, J. et al. Robust intratumor partitioning to identify high-risk subregions in lung cancer: A pilot study. *Int. J. Radiat. Oncol. * Biol. * Phys.* **95**, 1504–1512 (2016).
13. Sala, E. et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin. Radiol.* **72**, 3–10 (2017).
14. Zhou, M. et al. Radiomics in brain tumor: Image assessment, quantitative feature descriptors, and machine-learning approaches. *Am. J. Neuroradiol.* **39**, 208–216 (2018).
15. Corral, J. E. et al. Deep learning to classify intraductal papillary mucinous neoplasms using magnetic resonance imaging. *Pancreas* **48**, 805–810 (2019).
16. LaLonde, R. et al. Inn: inflated neural networks for ipmn diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 101–109 (Springer, 2019).
17. Hussein, S., Kandel, P., Bolan, C. W., Wallace, M. B. & Bagci, U. Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. *IEEE Trans. Med. imaging* **38**, 1777–1787 (2019).
18. Van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H. & Baessler, B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging* **11**, 91 (2020).
19. Horvat, N. et al. MR imaging of rectal cancer: Radiomics analysis to assess treatment response after neoadjuvant therapy. *Radiology* **287**, 833–843 (2018).
20. McCague, C. et al. Introduction to radiomics for a clinical audience. *Clin. Radiol.* **78**, 83–98 (2023).
21. Ismail, M. et al. Radiomic deformation and textural heterogeneity (r-depth) descriptor to characterize tumor field effect: Application to survival prediction in glioblastoma. *IEEE Trans. Med. imaging* **41**, 1764–1777 (2022).
22. Ismail, M. et al. Shape features of the lesion habitat to differentiate brain tumor progression from pseudoprogression on routine multiparametric mri: a multisite study. *Am. J. Neuroradiol.* **39**, 2187–2193 (2018).
23. Battalapalli, D. et al. Fractal dimension: analyzing its potential as a neuroimaging biomarker for brain tumor diagnosis using machine learning. *Front. Physiol.* **14**, 1201617 (2023).
24. Yang, G., Nawaz, T., Barrick, T. R., Howe, F. A. & Slabaugh, G. Discrete wavelet transform-based whole-spectral and subspectral analysis for improved brain tumor clustering using single voxel mr spectroscopy. *IEEE Trans. Biomed. Eng.* **62**, 2860–2866 (2015).
25. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. cancer* **48**, 441–446 (2012).
26. A Antolin, A., Workman, P., Mestres, J. & Al-Lazikani, B. Polypharmacology in precision oncology: current applications and future prospects. *Curr. Pharm. Des.* **22**, 6935–6945 (2016).
27. Aerts, H. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014; 5: 4006 (2014).
28. Ibrahim, A. et al. Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework. *Methods* **188**, 20–29 (2021).
29. Antunes, J. T. et al. Radiomic spatial textural descriptor (radistat): Quantifying spatial organization of imaging heterogeneity associated with tumor response to treatment. *IEEE J. Biomed. health Inform.* **26**, 2627–2636 (2022).
30. Lee, J., Narang, S., Martinez, J., Rao, G. & Rao, A. Association of graph-based spatial features with overall survival status of glioblastoma patients. *Sci. Rep.* **13**, 17046 (2023).
31. Cao, Y. et al. A multidimensional connectomics-and radiomics-based advanced machine-learning framework to distinguish radiation necrosis from true progression in brain metastases. *Cancers* **15**, 4113 (2023).
32. Kocevar, G. et al. Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Front. Neurosci.* **10**, 478 (2016).
33. Battalapalli, D. et al. Graph-radiomics learning (grrail): Application to distinguishing recurrence from pseudo-progression on structural mri. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)* (2024).
34. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
35. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
36. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
37. Patel, M. et al. Machine learning-based radiomic evaluation of treatment response prediction in glioblastoma. *Clin. Radiol.* **76**, 628–e17 (2021).
38. Bartolomei, F. et al. Disturbed functional connectivity in brain tumour patients: Evaluation by graph analysis of synchronization matrices. *Clin. Neurophysiol.* **117**, 2039–2049 (2006).
39. Bhandari, A. et al. Machine learning imaging applications in the differentiation of true tumour progression from treatment-related effects in brain tumours: A systematic review and meta-analysis. *J. Med. Imaging Radiat. Oncol.* **66**, 781–797 (2022).
40. Basree, M. M. et al. Leveraging radiomics and machine learning to differentiate radiation necrosis from recurrence in patients with brain metastases. *J. Neuro-Oncol.* 1–10 (2024).
41. Gao, X.-Y. et al. Differentiation of treatment-related effects from glioma recurrence using machine learning classifiers based upon pre- and post-contrast t1wi and t2 flair subtraction features: a two-center study. *Cancer management and research* 3191–3201 (2020).
42. Chakraborty, J. et al. Ct radiomics to predict high-risk intraductal papillary mucinous neoplasms of the pancreas. *Med. Phys.* **45**, 5019–5029 (2018).
43. Cui, S. et al. Radiomic nomogram based on mri to predict grade of branching type intraductal papillary mucinous neoplasms of the pancreas: A multicenter study. *Cancer Imaging* **21**, 1–13 (2021).
44. Jodoin, P.-M. et al. Challenges and best practices when using combat to harmonize diffusion mri data. *Sci. Rep.* **15**, 41508 (2025).
45. Zwanenburg, A. et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).
46. Haralick, R. M., Shanmugam, K. & Dinstein, I. H. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* 610–621 (1973).
47. Zhao, H. et al. Anomaly detection for medical images using self-supervised and translation-consistent features. *IEEE Trans. Med. Imaging* **40**, 3641–3651 (2021).
48. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
49. Gondi, V. et al. Radiation therapy for brain metastases: an astro clinical practice guideline. *Pract. Radiat. Oncol.* **12**, 265–282 (2022).
50. Zhang, Z. et al. Large-scale multi-center ct and mri segmentation of pancreas with deep learning. *Med. image Anal.* **99**, 103382 (2025).

51. Yushkevich, P. A., Gao, Y. & Gerig, G. Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 3342–3345 (IEEE, 2016).
52. Tustison, N. J. et al. N4itk: Improved n3 bias correction. *IEEE Trans. Med. imaging* **29**, 1310–1320 (2010).
53. Prasanna, P., Patel, J., Partovi, S., Madabhushi, A. & Tiwari, P. Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric mr imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur. Radiol.* **27**, 4188–4197 (2017).
54. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**, 389–422 (2002).
55. Lundberg, S. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874* (2017).
56. Hamilton, W. L., Ying, R. & Leskovec, J. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 1024–1034 (2017).
57. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)* (2019).
58. Veličković, P. et al. Graph attention networks. In *International Conference on Learning Representations (ICLR)* (2018).
59. Xu, K. et al. Representation learning on graphs with jumping knowledge networks. In *International Conference on Machine Learning (ICML)* (2018).
60. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
61. Van Griethuysen, J. et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
62. Battalapalli, D. et al. Spatial interactions via graph-based learning (sigl) to distinguish glioblastoma recurrence from pseudo-progression on clinical mri. In *Medical Imaging 2024: Computer-Aided Diagnosis*, vol. 12927 of *Proceedings of SPIE*, 256–261 (SPIE, 2024).

Acknowledgements

The authors would like to thank Virginia Hill, Volodymyr Statsevych, Raymond Huang, Andrew Baschnagel, Alan McMillan, Michael Veronesi, and Ankush Bhatia for providing access to anonymized brain tumor MRI scans used in this study. This research was supported by NIH/NCI/ITCR (1U01CA248226-01), NIH/NCI (R01CA277728-01A1), NIH/NCI (1R01CA264017-01A1), NIH/NCI (3U01CA248226-03S1), the DOD/PRCRP Career Development Award (W81XWH-18-1-0404), the Dana Foundation David Mahoney Neuroimaging Program, the V Foundation Translational Research Award, the Johnson & Johnson WiSTEM2D Award, the Musella Foundation Grant, the R & D Pilot Award, the Departments of Radiology and Medical Physics at the University of Wisconsin--Madison, and the WARF Accelerator Oncology Diagnostics Award.

Author contributions

D.B.: Data curation, formal analysis, investigation, visualization, methodology, algorithm development, validation, and writing-original draft and editing of the manuscript. M.I.: Formal analysis, project administration, review, and editing of the manuscript. A.S.: Review and editing of the manuscript. H.U., M.J., G.A.P.O.: Data preprocessing. U.B., M.S.A.: Data provision and review of the manuscript. P.T.: Conceptualization, funding acquisition, project administration, review, and editing of the manuscript.

Competing interests

P.T. is an equity holder in LivAI Inc. The authors declare no Competing Non-Financial Interests. All other authors declare no Competing Financial Interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-026-01649-x.

Correspondence and requests for materials should be addressed to Pallavi Tiwari.