Leakage-Free Multimodal Fusion of Radiomics and 3D Deep Learning for CT-Based PDAC Classification

Anoushka [1*] and Saurabh Singhal [2]

[1,2] Greater Noida Institute of Technology, Engineering Institute, Greater Noida, UP, India
[1*]anoushkatomar30@gmail.com
[2]me.ssaurabh@gmail.com

Abstract. PDAC is inherently lethal, which can be mostly explained by late detection and low sensitivity of visual inspection during contrast-enhanced CT. Minor changes in the textures, isoattenuation lesions, and large inter-observers' variability still remain obstacles to consistent detection. AI-based methods have shown real promise here, but single-modality approaches, radiomics alone, or deep learning alone, struggle when scanner protocols vary across sites and patient populations don't cooperate. Here we propose a no-leakage multi-approach framework for CT-based PDAC classification that takes a combination of handcrafted radiomics, 3D deep learning and decision-based fusion. Radiomics features are extracted from anatomically aligned regions of interest and then harmonized across scanners to account for variability introduced by differences in imaging equipment and acquisition settings. In parallel, a 3D early fusion ResNet uses whole-body CT scans and body part masks to get to know the spatial context of the cancer. To avoid data leakage, deep features are extracted using an out-of-fold strategy before being combined with radiomics features. Two fusion approaches are evaluated: average fusion and stacked fusion. The performance of our framework is evaluated using a very strict protocol of nested cross-validation where area under the ROC curve was used as the primary metric. Across all three pipelines i.e. radiomics, deep features, and fusion it was observed that non-linear models consistently outperformed their linear counterparts. The stacked fusion model reached an AUC of 0.955, the volumetric deep learning model hit 0.96. Both discriminate well between PDAC and non-PDAC cases. The results suggest that combining modalities with strict leakage controls produces a more stable classifier than either approach alone, which matters if the goal is something a radiologist can actually trust.

Keywords: Pancreatic ductal adenocarcinoma, computed tomography, radiomics, 3D deep learning, multimodal fusion, nested cross-validation, explainable AI.

1 Introduction

The poor prognosis of pancreatic ductal adenocarcinoma, with survival rates that have made little progress, can be explained by the fact that this cancer is diagnosed at a late stage in most patients. Contrast-enhanced CT scans are used in the detection, staging, and treatment planning of this cancer. In the early stages, the problem with PDAC is

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_4

detectability. PDACs are detected as areas of low attenuation with poor contrast to surrounding tissues and ill-defined margins. In clinical practice, radiologists do not detect PDACs. Even if two radiologists do, they do not agree.

Artificial intelligence has been actively explored in the analysis of CT scans of PDAC. Radiomics aims to quantify tumor heterogeneity using handcrafted intensity and texture descriptors, while in deep learning, models learn hierarchical representation from volumetric data. However, unimodal approaches face inherent weaknesses: radiomic descriptors are bounded by mathematical definitions and often confounded by acquisitions variation, while deep models are prone to overfitting, lack interpretability and may overestimate results without rigorous validation by the authors [1]. Multimodal fusion has emerged as a solution for these pitfalls by integrating complementary representations. By combining radiomics with deep learning, descriptors that are biologically interpretable can be combined with high-level spatial abstractions, and non-linear learning can capture complex feature interactions relevant to PDAC with greater precision. However, most exciting studies use naïve fusion schemes which, do not sufficiently control data leakage, or inadequately validate models for reliable conclusions to be drawn.

In this work, we propose a rigorously validated multimodal fusion framework for CT-based PDAC classification. The main contributions are:

- A harmonized radiomics and volumetric deep learning pipeline for PDAC classification.
- Leakage-free deep feature extraction via out-of-fold training within nested cross-validation.
- Direct comparison of linear and non-linear classifiers across unimodal and multimodal settings.
- A finding that stacked, non-linear fusion outperforms single-modality models across evaluation runs.

2 Related Work

CT-based PDAC analysis has been approached through radiomics, deep learning, and multimodal fusion, each targeting limitations that visual assessment leaves unaddressed by the authors [19]. Radiomics captures tumor heterogeneity through handcrafted descriptors, volumetric deep learning improves sensitivity by incorporating spatial context across the full scan volume by the authors [9]. Most existing CT-based PDAC methods fall into the same modality and have been validated using small and restricted datasets. This makes it difficult to generalize, and in a real-world application, this is significant as said by the authors [23]. The aforementioned limitations have been addressed in this work by using a leakage-free feature extraction, non-linear fusion, and cross-validation approach.

Table 1. Comparative Summary of Representative Studies on CT-Based PDAC Analysis

Modality	Method	Validation Strategy	Key Limitation	Study
CT	Radiomics	Multi-study methodological appraisal	Low IBSI compliance, reproducibility across studies not assured	Malcom. [8]
CT	Radiomics	Retrospective case-control with train-test split	Handcrafted texture features only, no volumetric context or learned representations	Chu. [21]
CT	Deep learning (3D CNN)	Nationwide train-test split	Single centred data, geographic generalizability untested	Malcom. [8]
CT	Deep learning (3D CNN)	Large-scale external validation	East Asian cohort only, no integration with handcrafted radiomic biomarkers	Cao. [22]
CT	Radiomics + deep learning (fusion)	Two centres train-test validation	Early fusion without leakage control, no cross-site harmonization	Zhang. [6]
CT	Radiomics + 3D deep learning (fusion)	Cross-validation	Prognostic scope, generalizability across institutions not validated	Dou. [7]
CT	Radiomics + 3D deep learning + stacked fusion	Nested cross-validation with OOF strategy	Small dataset, multi-institutional prospective validation still needed	Proposed work

3 Materials and Methods

The methodology pipeline (Fig.1) runs through five stages: standardized preprocessing, radiomics modeling, volumetric deep learning with early fusion, leakage-free deep feature extraction, and decision-level fusion, each designed to prevent data leakage and support reproducible validation.

3.1. Dataset Description

This study drew from two publicly available TCIA collections: CPTAC-PDA and Pancreas-CT. CPTAC-PDA contains approximately 103 patients with confirmed PDAC.

80 control scans from the Pancreas-CT dataset, acquired at the NIH Clinical Center, were included to represent healthy pancreatic anatomy.

During preprocessing, subjects were excluded where segmentation masks were unavailable or CT scans were corrupted, either condition made feature extraction unreliable. After quality filtering, the final dataset contained 149 subjects out of which 69 were PDAC cases and 80 were controls.

Data integrity checks were run before model training. The processed dataset contained no duplicate subjects and no feature-label inconsistencies, confirming that the data used for analysis was internally consistent and suitable for subsequent experiments.

3.2. Common Preprocessing

As depicted in Fig 1, all the CT volumes and segmentation masks underwent a shared preprocessing pipeline prior to branching. The raw DICOM series underwent reconstruction into three-dimensional volumes, followed by conversion to Hounsfield Units and resampling to a uniform isotropic resolution, thereby mitigating inter-scanner variability. To enhance the soft-tissue contrast pertinent to pancreatic imaging, intensity windowing and normalization were implemented. Segmentation masks, including tumor masks for CPTAC-PDA and Pancreas-CT, were rasterized, aligned, and subsequently verified to guarantee voxel-wise correspondence with the CT volumes. This standardized preprocessing protocol ensures the provision of consistent inputs for both radiomics and deep learning methodologies by the authors [2].

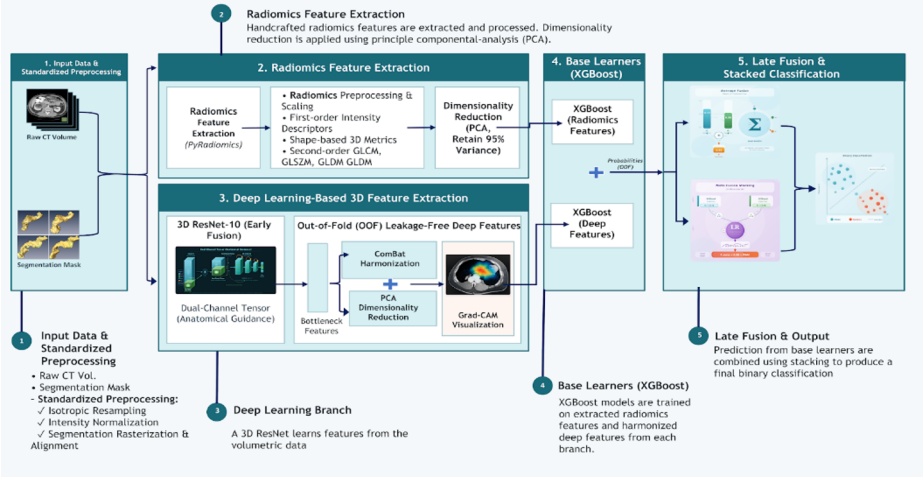

Fig. 1. Integrated multimodal framework for CT-based PDAC classification

3.3. Radiomics Branch

The radiomics branch (Fig 1A) extracted the manually designed quantitative features from the co-registered CT mask volumes. One region of interest per subject was chosen to maintain consistency at the subject level. A total of 107 radiomics features were

extracted using PyRadiomics. Of these, 15 shape-based features were removed early on since shape features are known to be heavily correlated with segmentation geometry and tumor size, and hence not amenable to inter-subject comparisons. The PyRadiomics diagnostic columns and dataset metadata variables were removed since they do not carry any predictive information. The texture and intensity features that survived the cleaning process were used as input to all the models. These features capture biologically meaningful characteristics such as heterogeneity, hypodensity, and structural irregularity said by the authors [3].

Given the multi-cohort nature of the data, feature harmonization was performed using ComBat, treating dataset origin as a batch variable while protecting diagnostic labels by the authors [4]. This step mitigates scanner and protocol-induced variability without suppressing disease-related signals. Following harmonization, dimensionality reduction was carried out using PCA, applied after Z-score standardization. Rather than fixing the number of components in advance, the analysis retained however many components were needed to capture 95% of the variance in the training data, so the number varied depending on the data rather than being set arbitrarily, as summarized in Fig 1. This reduced redundancy in the feature space without discarding the patterns most relevant to classification.

3.4. Deep Learning Branch with Early Fusion

The deep learning branch (Fig.1B) uses a volumetric 3D ResNet-10 with an early-fusion input. Each sample is represented as a dual-channel volume, the preprocessed CT scan and its segmentation mask concatenated along the channel dimension. The CT channel carries structural and density information; the mask channel supplies anatomical localization by the authors [5].

All volumes were normalized and resized 64 ×128 ×128 voxels prior to training. A stratified 70/15/15 train-validation-test split preserved class distribution across subsets. Optimization used AdamW with a learning rate of 1 ×10-4 and weight decay of 1 ×10-5. Binary cross-entropy loss with logits was applied with class-weighted adjustments to address the PDAC/control imbalance. Training ran for up to 50 epochs, with early stopping (patience = 10) monitored on validation AUC to prevent overfitting.

By fusion CT and mask at the input, the network learns spatially aligned features from the first convolutional layers, anchoring representation learning to the pancreas region rather than surrounding anatomy. The network outputs a PDAC probability score and produces intermediate feature representations used in the downstream fusion stage by the authors [6].

Overview of the end-to-end pipeline illustrating standardized CT preprocessing, parallel radiomics and anatomically guided 3D deep learning feature extraction, leakage-free deep feature generation, XGBoost-based base learners, and late fusion via averaging and stacked meta-learning for final PDAC versus control classification.

3.5. Out-of-Fold Deep Feature Extraction

To enable robust multimodal fusion, intermediate deep features were extracted from the trained 3D ResNet using a strict out-of-fold (OOF) strategy Fig.1. Within each

cross-validation fold, the network was trained only on the training subset, after which bottleneck features were extracted for the held-out data. Deep features were extracted from an intermediate layer of the 3D ResNet, after which adaptive global average pooling collapsed the spatial dimensions of each feature map down to a single vector. The result was a 256-dimensional representation per subject, encoding the high-level volumetric patterns the network had learned from the CT data. This ensures that every deep feature vector is generated by a model that has not seen the corresponding subject, preventing information leakage.

The same dimensionality reduction approach used for the radiomics branch was applied here: PCA with a 95% variance retention cutoff. This kept the deep feature space compact while holding onto the representation that actually mattered for classification by the authors [7]. These OOF deep features constitute a complementary modality for fusion-based classification.

3.6. Fusion Strategies and Validation

Decision-level fusion is performed using two strategies shown in Fig 1. Average fusion computes the mean of probabilistic outputs from base learners trained on radiomics and deep features, serving as a simple baseline. In contrast, in the Stacked Fusion technique, a meta-learner is employed that learns from the OOF probabilistic predictions obtained from the base models. This technique allows for data-driven weighting and non-linear interaction modeling between the data modalities. A nested cross-validation procedure is used for all the experiments. This is done throughout the entire pipeline. The outer loop is used for obtaining unbiased estimates of the performance of the model, and the inner loop is used for preprocessing, harmonization, model selection, and fusion learning. All data-dependent operations are restricted to the training data. AUC is used as the primary evaluation criterion since it is robust to class imbalance and relevant for decision support systems by the authors [8]. This leakage-aware strategy is used for fair comparisons between unimodal and multimodal models. It also ensures that focus is kept on testing for general ability, not just fit according to the authors [9].

4. Results

This section compares four types of models for the classification of PDAC from CT images: radiomics models, volumetric deep learning models, deep feature classifiers, and multimodal fusion models given by the authors [10]. The comparison criterion is AUC, which is threshold-independent and robust for class distributions as observed in common clinical image datasets. The numerical comparison is presented in Table 2, and the ROC curves and confusion matrices are presented in Fig.2 (a)-(d).

4.1. Radiomics Results: Linear vs. Non-Linear Models

The performance of radiomics was highly dependent on the classifier used. The logistic regression classifier gave an AUC of 0.549, which is nearly random, indicating that

linear decision boundaries are not effective enough to describe the relationships between texture, intensity, and shape features that describe the heterogeneity of PDAC.

XGBoost gave an AUC of 0.808 (Fig.2 (a-b)). This indicates that a particular pattern is followed, as the features have some diagnostic capability, but it is not linear. The linear classifier fails to describe it, and the nonlinear classifier fails to describe it as well, which is consistent with the findings of the study that ensemble-based classifiers are more suitable for utilizing the high-dimensional feature space made available by radiomics.

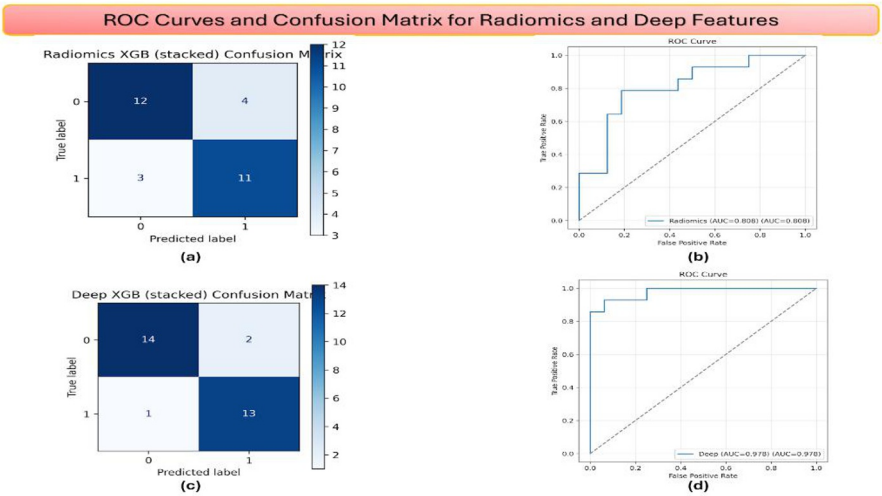

Fig. 2. Classification performance of radiomics and deep feature models. (a) Confusion matrix for the Radiomics XGBoost model with (b) its ROC curve (AUC = 0.808). (c) Confusion matrix for the Deep Feature XGBoost model with (d) its ROC curve (AUC = 0.978), demonstrating superior predictive performance of deep features.

4.2. Deep 3D ResNet Results

The volumetric deep learning model had an AUC value of 0.96 on the held-out test set. Out of 34 total cases, 19 PDAC and 14 controls had been correctly classified, with one false positive and no false negatives (Fig 3). This suggests a high sensitivity and specificity in classifying PDAC from control subjects. The training behavior also suggests that there is stable learning in the model. From the training and validation loss curves, we notice that there is gradual convergence in the training process. There is no significant divergence in the training and validation loss curves. We also notice that the validation AUC is consistently high throughout the training process. However, we notice that the training stops at 24 epochs due to early stopping, as validation performance does not improve beyond this point.

We also notice from the ROC curve that there is good class separation across a range of decision thresholds. This is consistent with the reported AUC value. We notice that

the volumetric DL model has performed very well in classifying PDAC from control subjects. This suggests that there is good generalization in this data set given by authors [12].

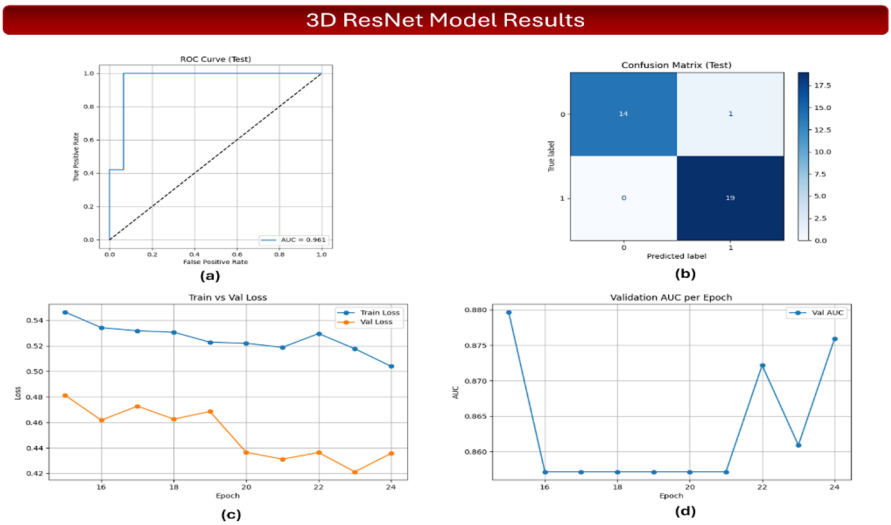

Fig. 3. Performance evaluation of the proposed 3D ResNet model on the test dataset: (a) ROC curve with an AUC of 0.961, demonstrating strong discriminative capability; (b) confusion matrix showing classification outcomes; (c) training and validation loss trends across epochs; and (d) validation AUC progression during training

4.3. Deep Feature Results: LR vs. XGBoost

Experiments performed on the deep feature representation obtained from the 3D ResNet model showed a marked difference in performance between linear and non-linear classifiers. When logistic regression was applied to the extracted embeddings, the performance of classification was poor, resulting in an AUC of 0.357. This indicates that the learned feature space is not literally separable and that a linear decision boundary is not sufficient to model the underlying diagnostic patterns.

However, when XGBoost was applied to the same deep feature space, the performance was substantially better, achieving an AUC of 0.978 on the test set Fig. 3 (c-d)). Further statistical analysis of the results confirmed the robustness of the performance. The 95% confidence interval obtained from the boot-strap analysis was 0.92-1.00, and multiple experimental runs showed a mean AUC of 0.961 ± 0.034, indicating robust predictive performance across varying splits of the data.

These results show that the deep feature representation obtained from the volumetric network captures complex, non-linear relationships that can be harnessed effectively by flexible classifiers such as XGBoost. These results suggest that deep feature representations contain more diagnostic information than a linear classifier can model and

that non-linear classifiers are required to tap into this information fully by the authors [13].

4.4. Average Fusion Results: Linear vs non-linear

Average fusion was first experimented as a baseline for the fusion of the radiomics features and the deep feature predictions. It was observed that when logistic regression was used as the classifier, the average fusion resulted in a low AUC value of 0.406, which clearly showed that the information was not complementary across the modalities for simple linear combinations.

But when the average feature representation was experimented using XGBoost as the classifier, it resulted in an AUC of 0.942 for the test set Fig.4 (b). Repeated experiments further confirmed the stability of this result, yielding a mean AUC of 0.983 with a standard deviation of 0.017, indicating that non-linear decision boundaries are better suited to model the interactions between radiomics descriptors and deep imaging features. The averaging strategy has a structural limitation, though, it assigns fixed weights to both modalities across all samples, with no mechanism to adjust per-sample contributions by the authors[14].

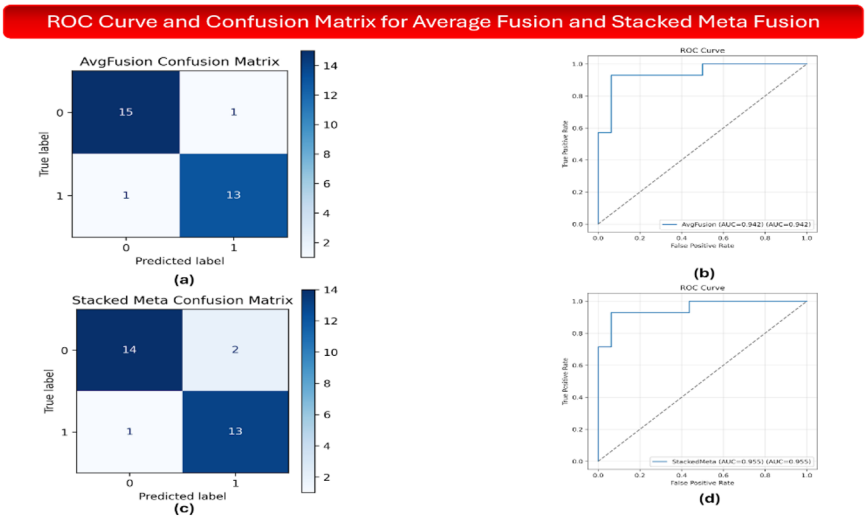

Fig. 4. Performance comparison of two fusion strategies. (a) Confusion matrix for the Average Fusion approach and (b) its corresponding ROC curve (AUC = 0.942). (c) Confusion matrix for the Stacked Meta Fusion model and (d) its ROC curve (AUC = 0.955), indicating improved classification performance over average fusion

4.5. Stacked Fusion Results: LR vs. XGBoost

Stacked meta-fusion was then evaluated to improve adaptive modality-specific prediction fusion. In this case, the logistic regression model was used, and the AUC was found

to be 0.683. Although this indicates a reasonable level of improvement compared to the linear aver-age fusion, the linear assumption between the modalities acted as a constraint.

The best multimodal performance was achieved when the XGBoost model was used as the stacked meta-learner. In this case, the AUC was found to be 0.955 for the test data Fig.4 (d). On average, the experimental runs gave an AUC of 0.983, with a standard deviation of 0.017. This indicates that the stacked fusion method yields stable multimodal predictive performance, provided that the predictions of the two modalities are fused using a non-linear framework [15].

4.6. Explainability Analysis

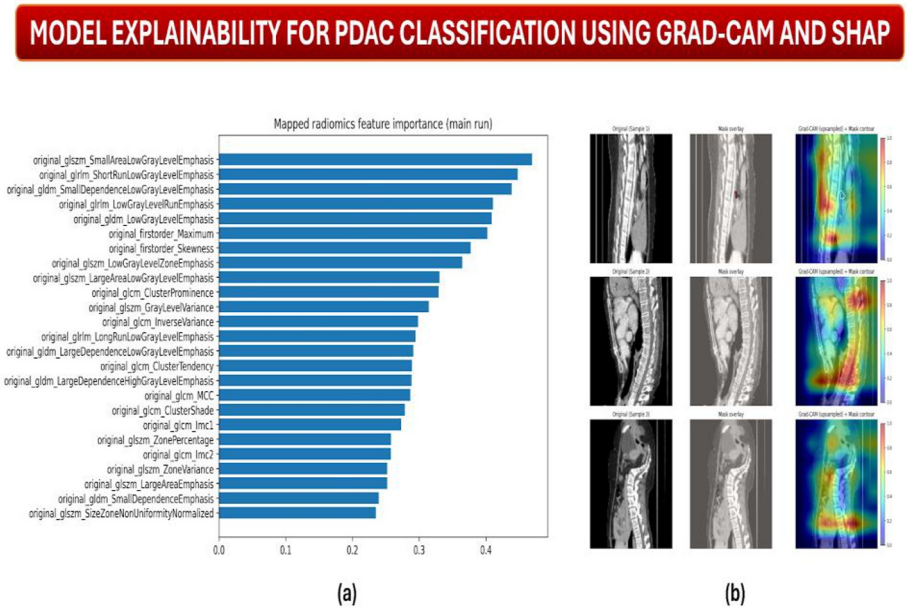

Fig. 5. Explainability analysis for PDAC classification using SHAP and Grad-CAM. (a) SHAP feature importance for the XGBoost radiomics model emphasizing the importance of prominent low gray-level and texture features. (b) Grad CAM analyses from the 3D ResNet illustrating the correspondence between network focus and regions of pancreatic tumors in exemplary cases.

Explainability analysis was conducted to evaluate whether the learned representations capture meaningful imaging patterns. As shown in Fig. 5 (b), the Grad-CAM visualization results of the 3D Res-Net are concentrated on the regions that overlap with the pancreas and tumor masks, and a large amount of attention is paid to the hypodense and morphologically irregular regions. This spatial alignment indicates that the network's predictions are driven by anatomically relevant structures rather than background artefacts by the authors [16].

Feature-level explainability was examined using SHAP analysis for both XGBoost and logistic regression models. For XGBoost, the most influential radiomics features include SmallAreaLowGrayLevelEmphasis, LowGrayLevelRunEmphasis, and related heterogeneity measures, which are associated with hypo vascularity and necrotic components characteristic of PDAC (Fig.5 (a)) by the authors [17]. In contrast, logistic regression assigns importance primarily to global intensity statistics, reflecting its limited capacity to model complex feature interactions. Together, these findings demonstrate biologically plausible decision-making and help explain the superior performance of non-linear and fusion-based models.

Table 2. Comparative performance of radiomics, deep learning, and fusion-based models for PDAC classification

Model Category	Input Representation	Classifier	AUC
Radiomics	Handcrafted radiomics features	Logistic Regression	0.549
Radiomics	Handcrafted radiomics features	XGBoost	0.808
Deep Learning	Raw volumetric CT (early-fusion)	3D ResNet	0.96
Deep Features	ResNet bottleneck embeddings	Logistic Regression	0.357
Deep Features	ResNet bottleneck embeddings	XGBoost	0.978
Average Fusion	Radiomics + deep features	Logistic Regression	0.406
Average Fusion	Radiomics + deep features	XGBoost	0.942
Stacked Fusion	Radiomics + deep features	Logistic Regression	0.683
Stacked Fusion	Radiomics + deep features	XGBoost	0.955

5. Discussion

This study shows that the combination of radiomics and volumetric deep learning through multimodal fusion generates a better result in classifying pancreatic ductal adenocarcinoma (PDAC) using CT, compared with unimodal techniques. Radiomics and deep learning capture complementary information: handcrafted features encode interpretable descriptors of tissue heterogeneity and intensity distribution, while deep networks learn high-level spatial and contextual representations directly from volumetric data. By integrating these complementary representations through fusion, the framework combines the strengths of each modality and enables improved discrimination between PDAC and control cases by the authors [18].

Across all of the experiments, non-linear modeling revealed itself to be a key performance determinant. Linear classifiers always performed poorly when applied to radiomics, deep features and fusion outputs, thus suggesting that imaging biomarkers of pancreatic ductal adenocarcinoma (PDAC) are not linearly separable. In contrast, gradient-boosted ensembles were able to witness non-linear relationships in both the within and across feature spaces, which provide significant improvements in discrimination. These results highlight the years of importance of model expressiveness as being equally critical as feature design, especially, within heterogeneous clinical imaging datasets.

The proposed framework contributes to ongoing research on multimodal PDAC classification by combining volumetric deep learning features with radiomics representations within leakage-aware validation pipeline. Many previous investigations are based on single modality pipelines, naïve fusion methods, or lack cross-validation, which may lead to estimating optimistic performance. Through an emphasis on robust validation and harmonization, this work provides a more reliable assessment of multimodal integration.

In the translational view, the proposed approach paves the way for the development of reliable decision support systems for PDAC screening/diagnosis. Even if the approach is not meant to replace the expertise of radiologists, the presence of such models could help the clinician identify patients with high risk and contribute to the confidence of uncertain CT scans, especially in the early phases of the disease [19].

6. Conclusion

This paper proposes a leakage-aware multimodal framework for CT-based pancreatic ductal adenocarcinoma (PDAC) classification that combines radiomics, volumetric deep learning, and decision-level fusion. The results demonstrate that non-linear models substantially outperform linear baselines across all representations and that multimodal integration improves robustness over unimodal approaches. Among the evaluated strategies, stacked fusion with a non-linear meta-learner achieved the most consistent and reliable discrimination, highlighting the contemporary strengths of handcrafted and learned features by the authors [20].

The primary contribution of this study lies in the systematic combination of early-fusion 3D deep learning, out-of-fold feature extraction, and rigorous nested cross-validation within a unified pipeline. Future work will focus on external multi-institutional validation and extension of the framework to longitudinal and multimodal clinical data to support earlier and more reliable PDAC detection.

Acknowledgment. The authors would like to thank the faculty and research staff of the Department of Computer Science for their guidance and support throughout this work. We also thank the institute's High Performance Computing Unit, whose infrastructure enabled it to manage the computing requirements of the clinical and imaging datasets. Furthermore, we appreciate the TCIA community for upholding open-access imaging

libraries that continue to facilitate significant scholarly investigation and repeatable studies.
Disclosure of Interests The authors declare that they have no competing interests that could have influenced the work reported in this paper.

Acknowledgments. The authors would like to thank the faculty and research staff of the Department of Computer Science for their guidance and support throughout this work. We also thank the institute's High Performance Computing Unit, whose infrastructure enabled it to manage the computing requirements of the clinical and imaging datasets. Furthermore, we appreciate the TCIA community for upholding open-access imaging libraries that continue to facilitate significant scholarly investigations and repeatable studies.

Disclosure of Interests. The authors declare that they have no competing interests that could have influenced the work reported in this paper.

References

1. Varoquaux, G., Cheplygina, V.: Machine learning for medical imaging: Methodological failures and recommendations for the future. NPJ Digit. Med. 5, 48 (2022). https://doi.org/10.1038/s41746-022-00592-y
2. Clark, K., Vendt, B., Smith, K., et al.: The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. J. Digit. Imaging 26, 1045–1057 (2013). https://doi.org/10.1007/s10278-013-9622-7
3. Mukherjee, S., Patra, A., Khasawneh, H., et al.: Radiomics-based machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis. Gastroenterology 163(5), 1435–1446.e3 (2022). https://doi.org/10.1053/j.gastro.2022.06.066
4. Da-Ano, R., Masson, I., Lucia, F., Doré, M., Robin, P., et al.: Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. Sci. Rep. 10, 10248 (2020). https://doi.org/10.1038/s41598-020-66110-w
5. MONAI Consortium: MONAI: Medical Open Network for AI. Zenodo (2022). https://doi.org/10.5281/zenodo.6639453
6. Zhang, Y., Lobo-Mueller, E.M., Karanicolas, P., Gallinger, S., Haider, M.A., Khalvati, F.: Improving prognostic performance in resectable pancreatic ductal adenocarcinoma using radiomics and deep learning feature fusion in CT images. Sci. Rep. 11, 15459 (2021). https://doi.org/10.1038/s41598-021-80998-y
7. Dou, Z., Lu, C., Shen, X., Gu, C., Shen, Y., Xu, W., Qin, S., Zhu, J.: Development of a radiomics–3D deep learning fusion model for prognostic prediction in pancreatic cancer. BMC Cancer 25, 14889 (2025). https://doi.org/10.1186/s12885-025-14889-0
8. Malcolm, J.A., Tacey, M., Gibbs, P., et al.: Current state of radiomic research in pancreatic cancer: Focusing on study design and reproducibility of findings. Eur. Radiol. 33, 6659–6669 (2023). https://doi.org/10.1007/s00330-023-09653-6
9. Martí-Bonmatí, L., Koh, D.M., Rojo-Carrión, R., et al.: Pancreatic cancer, radiomics and artificial intelligence. Br. J. Radiol. 95(1137), 20220072 (2022). https://doi.org/10.1259/bjr.20220072

10. Huang, B., Huang, H., Zhang, S., Zhang, D., Shi, Q., Liu, J., Guo, J.: Artificial intelligence in pancreatic cancer. Theranostics 12(16), 6931–6954 (2022). https://doi.org/10.7150/thno.77949
11. Khalvati, F., Zhang, Y., Baig, S., Lobo-Mueller, E.M., et al.: Prognostic value of CT radiomic features in resectable pancreatic ductal adenocarcinoma. Sci. Rep. 9, 5449 (2019). https://doi.org/10.1038/s41598-019-41728-7
12. Chen, P.T., Wu, T., Wang, P., Chang, D., Liu, K.L., Wu, M.S., et al.: Pancreatic cancer detection on CT scans with deep learning: A nationwide population-based study. Radiology 305(1), 134–143 (2022). https://doi.org/10.1148/radiol.220152
13. Bakasa, W., Viriri, S.: VGG16 feature extractor with extreme gradient boost classifier for pancreas cancer prediction. J. Imaging 9(7), 138 (2023). https://doi.org/10.3390/jimaging9070138
14. Jia, W., Li, H., Ali, R., Shanbhogue, K.P., Masch, W.R., et al.: Investigation of ComBat harmonization on radiomic and deep features from multi-center abdominal MRI data. J. Digit. Imaging 38, 1016–1027 (2025). https://doi.org/10.1007/s10278-024-01253-0
15. Yao, S., Yao, D., Huang, Y., Qin, S., Chen, Q.: A machine learning model based on clinical features and ultrasound radiomics features for pancreatic tumor classification. Front. Endocrinol. 15, 1381822 (2024). https://doi.org/10.3389/fendo.2024.1381822
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int. J. Comput. Vis. 128, 336–359 (2020). https://doi.org/10.1007/s11263-019-01228-7
17. Park, Y.J., Park, Y.S., Kim, S.T., Hyun, S.H.: A machine learning approach using [18F]FDG PET-based radiomics for prediction of tumor grade and prognosis in pancreatic neuroendocrine tumor. Mol. Imaging Biol. 25, 897–910 (2023). https://doi.org/10.1007/s11307-023-01832-7
18. Lyu, G., Tong, T., Yang, G., Zhao, J., Xu, Z.: Bibliometric and visual analysis of radiomics for evaluating lymph node status in oncology. Front. Med. 11, 1501652 (2024). https://doi.org/10.3389/fmed.2024.1501652
19. Alves, N., Schuurmans, M., Viergever, M.A., et al.: Artificial intelligence and radiologists in pancreatic ductal adenocarcinoma detection on CT (PANORAMA): A nationwide population-based study. Lancet Oncol. 26(5), 601–611 (2025)
20. Grossberg, A.J., Chu, L.C., Deig, C.R., et al.: Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. CA Cancer J. Clin. 70, 375–403 (2020). https://doi.org/10.3322/caac.21626
21. Chu, L.C., Park, S., Kawamoto, S., Fouladi, D.F., et al.: Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue. AJR Am. J. Roentgenol. 213(4), 889–896 (2019). https://doi.org/10.2214/AJR.18.20901
22. Cao, K., Xia, Y., Yao, J., Han, X., Lambert, L., Zhang, T., et al.: Large-scale pancreatic cancer detection via non-contrast CT and deep learning. Nat. Med. 29, 3033–3043 (2023). https://doi.org/10.1038/s41591-023-02640-w
23. Patnam, N.G., Huang, C., Miao, S., et al.: Pancreas segmentation using AI developed on the largest CT dataset with multi-institutional validation. Sci. Rep. 15, 8542 (2025). https://doi.org/10.1038/s41598-025-01802-9