OPEN ACCESS

Original research

Next-generation AI for visually occult pancreatic cancer detection in a low-prevalence setting with longitudinal stability and multi-institutional generalisability

Sovanlal Mukherjee,[1] Ajith Antony,[1,2] Nandakumar G Patnam,[1,3] Kamaxi H Trivedi,[1,3] Aashna Karbhari,[1,3] Khurram Khaliq Bhinder,[1] Armin Zarrintan,[1] Joel G Fletcher,[1] Mark Truty,[4] Matthew P Johnson,[5] Suresh T Chari ,[6] Ajit Harishkumar Goenka [1]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/gutjnl-2025-337266).

[1]Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA
[2]Radiology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA
[3]Radiology, University of Washington, Seattle, Washington, USA
[4]Surgery, Mayo Clinic, Rochester, Minnesota, USA
[5]Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA
[6]Gastroenterology, Hepatology and Nutrition, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Correspondence to
Dr Ajit Harishkumar Goenka; goenka.ajit@mayo.edu

Received 12 October 2025
Accepted 5 March 2026

To cite: Mukherjee S, Antony A, Patnam NG, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2025-337266

ABSTRACT

Background Failure of conventional imaging to detect pancreatic ductal adenocarcinoma (PDA) at its visually occult pre-diagnostic stage is a primary barrier to improving its otherwise poor rate of survival.

Objective To develop and validate the Radiomics-based Early Detection MODel (REDMOD), an AI framework to identify subvisual radiomic signatures of pre-diagnostic PDA on standard-of-care CT.

Designs REDMOD was trained on a multi-institutional cohort (n=969; 156 pre-diagnostic, 813 control) and tested on an independent set (n=493; 63 pre-diagnostic, 430 control), simulating a low prevalence (~1:6) early detection paradigm. The fully automated framework couples AI-driven segmentation with a heterogeneous ensemble architecture trained on a 40-feature radiomic signature derived from Synthetic Minority Over-sampling Technique (SMOTE)-balanced data. A tunable Youden Index-optimised classification threshold enables performance calibration without retraining. Validation included direct comparison with radiologists, longitudinal test–retest analysis and external specificity validation across two independent cohorts (n=539 and n=80).

Results On an independent test set (n=493), REDMOD identified occult PDA (AUC 0.82; 73.0% sensitivity) at a median 475-day lead time. This represented nearly twofold higher sensitivity than radiologists (38.9%; p<0.001), which grew to nearly threefold (68.0% vs 23.0%) at >24 months lead time. REDMOD showed strong longitudinal stability (90–92% concordance) and generalisable specificity across multi-institutional (81.3%; n=539) and public (87.5%; n=80) datasets. Mechanistic analyses confirmed predictive power derived principally from multi-scale wavelet-filtered textural features (90% of selected signature), which outperformed unfiltered features (AUC 0.82 vs 0.74; p=0.007) in capturing subvisual architectural disruptions.

Conclusions REDMOD is an automated, mechanistically grounded, longitudinally stable, externally validated AI that surpasses radiologists for PDA detection at its visually occult pre-diagnostic stage. These attributes position it for prospective validation in high-risk cohorts, a necessary step towards shifting the paradigm from late-stage symptomatic diagnosis to proactive pre-clinical interception.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Early detection of pancreatic ductal adenocarcinoma (PDA) is the most powerful approach to improve survival; however, this objective is fundamentally impeded by the morphologically normal appearance of the pancreas on conventional imaging during its curable pre-clinical phase.

⇒ This diagnostic deficit limits the clinical utility of NICE recommendations for urgent imaging in high-risk cohorts (new onset diabetes and weight loss), thereby necessitating technologies to augment standard CT interpretation.

⇒ Initial AI investigations showed potential but were constrained by non-scalable manual methodologies, unrealistic case–control ratios that do not accurately reflect clinical prevalence and a lack of rigorous validation regarding longitudinal stability and multi-institutional generalisability.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDA) is among the deadliest malignancies.[1] Its poor prognosis is attributable to the prevailing clinical paradigm of symptom-driven detection, leading to the diagnosis of over 85% of cases at an unresectable stage where therapeutic interventions are solely palliative. Yet, PDA stands as an anomaly among prevalent malignancies, lacking a recognised early detection strategy for sporadic cases, which constitute the predominant proportion (85–90%) of all PDAs.[2] Consequently, the early detection of sporadic PDA represents the most impactful pathway to enhance overall survival, as even modest advancements in early diagnosis would yield substantial reductions in national mortality rates.[3 4]

Recently, glycaemically-defined new onset diabetes (gNOD) with an Enriching NOD for Pancreatic Cancer (ENDPAC) score of ≥3 has emerged as a validated high-risk cohort for sporadic PDA with a 3-year risk of ~3.6%, which is nearly 20-fold higher than the general population.[5 6] In fact, the National Institute for Health and Care Excellence (NICE)

WHAT THIS STUDY ADDS

⇒ This study introduces Radiomics-based Early Detection MODel (REDMOD), a next-generation AI framework featuring a robust ensemble architecture (logistic regression, random forest, XGBoost) and fully automated volumetric pancreas segmentations.

⇒ Validated in a large low-prevalence cohort (~6:1 control–case ratio) and incorporating a clinically adaptable design with a modifiable diagnostic threshold, REDMOD detects the subvisual signature of pre-clinical PDA a median of 475 days before clinical diagnosis (AUC 0.82).

⇒ The performance of REDMOD surpasses that of radiologists, demonstrating nearly double the sensitivity (73.0% vs 38.9%; p<0.001) for detecting visually occult PDA, with an advantage that increases to nearly threefold for cases detected more than 24 months prior to diagnosis.

⇒ For the first time, this work establishes the longitudinal stability of a pre-clinical radiomic signal (90–92% test–retest concordance) and confirms the robust specificity of the model across independent multi-institutional (81.3%) and public (87.5%) datasets.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ REDMOD offers an automated and longitudinally stable anchor for multicentre prospective studies in risk-enriched early detection programmes, with pre-specified thresholds and an observe–retest strategy for intermediate results.

⇒ REDMOD could operate as a triage and longitudinal monitoring tool for CT requests triggered by NICE guidelines, aligning with NHS pathways while awaiting prospective clinical evidence and cost-effectiveness.

⇒ The release of a fully automated and externally benchmarked pipeline offers a reproducible platform for method comparisons, bias audits and cost-effectiveness modelling to guide policy.

now recommends urgent abdominal imaging for individuals aged ≥60 with gNOD and weight loss,[7] which underscores the growing recognition of gNOD as an early clinical manifestation of PDA. However, even in such high-risk cohorts, conventional imaging fails to reliably detect PDA at a curable stage due to the absence of overt imaging abnormalities.[8] Furthermore, it is critical to distinguish between 'missed' and 'imaging-occult' PDA. 'Missed' cancers are those where a retrospective expert review can identify a discrete lesion that was overlooked due to perceptual error or technical factors such as suboptimal contrast timing or artefacts. In contrast, 'imaging-occult' cancers show no discernible mass even on expert re-review. Thus, they represent a distinct biological state where disease presence is encoded in subvisual parenchymal alterations rather than macroscopic morphological changes.[4] Our group achieved a breakthrough by demonstrating that machine learning (ML) models, using radiomic features derived from pre-diagnostic CT scans—incidental CTs acquired months to years prior to clinical diagnosis—exhibit high accuracy in identifying such subclinical alterations indicative of carcinogenesis.[9 10] These models substantially outperformed radiologists in detecting visually occult PDA (stage 0 disease) with a considerable lead time before clinical diagnosis. However, consistent with other investigations in the field at that time,[11–13] our study relied on manual pancreas segmentations, which are labour-intensive and introduce interobserver variability.[14] Furthermore, the dataset possessed a control–case ratio of ~1, which did not fully reflect low prevalence early detection scenarios. Finally, the model used a standard classifier without specific mechanisms to manage the profound class imbalance inherent in such settings. These factors underscored a distinct trajectory for methodological refinement towards enhanced clinical applicability.

To overcome current limitations, several critical unmet needs must be addressed. First, fully automated pancreas segmentation is essential to remove the labour and variability of manual segmentation.[15] Second, because PDA is rare, models must be trained and tested on larger more heterogeneous datasets that approximate the high control–case ratios in early detection cohorts to the extent feasible.[16] Third, more advanced radiomic feature engineering may reveal subtle reproducible pre-clinical disease patterns. Fourth, there is a need for modelling strategies that explicitly address extreme class imbalance in early detection cohorts. Beyond accuracy, two issues are central for clinical adoption: understanding how different groups of radiomic features contribute to the model's predictions and linking these imaging signatures to known biological processes. Equally critical—and largely unaddressed—is the longitudinal performance and test–retest reliability of these AI models. For an AI tool to be trusted in early detection contexts, its ability to provide consistent and reliable predictions over time, despite minor variations in patient status or image acquisition, must be rigorously established.[17]

The current study introduces Radiomics-based Early Detection MODel (REDMOD), an AI framework developed to systematically address these challenges. REDMOD integrates our validated, fully automated volumetric pancreas segmentation tool.[18] It was developed using a large cohort with a clinically oriented control–case ratio and employs an expanded radiomic feature set with advanced selection methods. Its classification engine is a heterogeneous ensemble model, explicitly trained using the Synthetic Minority Over-sampling Technique (SMOTE)[19] to manage class imbalance. This study also incorporates novel analytical components: evaluating differential feature-type contributions and assessing longitudinal model stability. We also present its head-to-head comparison against expert radiologists, analyse its detection capabilities across stratified lead time intervals preceding clinical diagnosis, and explore the characteristics of radiomic features associated with early disease. By systematically addressing previous limitations through comprehensive automation, a more challenging and representative dataset, advanced feature engineering, a robust ensemble classification strategy and by including novel analyses of feature contributions and longitudinal model stability, this study seeks to develop a more robust and clinically relevant tool for early PDA detection.

MATERIALS AND METHODS

Study participants

Pre-diagnostic PDA cohort

The pre-diagnostic cohort was assembled through a stringent process to capture only truly subclinical PDA cases. Patients with pathologically confirmed PDA (~3500) were first identified from our electronic medical record, and all contrast-enhanced pre-diagnostic abdominal CT scans obtained 3–36 months prior to tissue diagnosis were retrieved. From this initial pool, eligibility required several sequential safeguards:

- Prospective clinical interpretation: only CT scans prospectively reported by the reading radiologist during routine care as negative for any focal pancreatic mass or other imaging features suggestive of PDA were considered.

- Independent verification: each candidate CT scan then underwent re-review by a board-certified abdominal radiologist with inclusion contingent on confirmation of no discernible focal mass.

Exclusion criteria were predefined as any CT scan with image quality insufficient for detailed ductal and parenchymal assessment, non-standard contrast phase (eg, non-contrast, arterial or delayed only) or substantial artefact (such as motion or streak artefact). This process yielded 219 pre-diagnostic CT scans with a median lead time to diagnosis of 427 days (range 90–1092). The cohort was intentionally diverse, comprising scans from multiple institutions (70.3% external) and four major CT vendors with varied acquisition parameters, as described in online supplemental methods section S1.

Control cohort

A large control cohort was assembled to provide a rigorous comparator to the pre-diagnostic cases. Abdominal CT scans were selected if their contemporaneous clinical reports indicated no evidence of PDA, specifically:

- Radiology report criteria: negative for focal pancreatic masses or suspicious features, although scans with incidental but common benign findings (eg, small cysts, calcifications, lipomatosis) were permitted to reflect routine practice.
- Independent radiologist review: each control scan underwent re-review by a board-certified abdominal radiologist to confirm the absence of suspicious features.
- Longitudinal verification: included subjects were required to remain free of a PDA diagnosis for at least 3 years of subsequent clinical follow-up.

From this process we assembled 1243 control CT scans, yielding a control–pre-diagnostic case ratio of ~6:1 to simulate a low-prevalence early detection context. This cohort mirrored the multi-institutional (83% external) and technical diversity of the pre-diagnostic set, as shown in online supplemental methods section S2. Specifically, to strengthen internal validity and limit selection bias, controls were selected using a systematic matching process rather than convenience sampling. They were drawn from the same institutional CT archives and calendar period as the pre-diagnostic cases matched on age, sex, scan acquisition date and similar abdominal CT indications (eg, abdominal pain or surveillance), so that the control cohort was demographically, clinically and technically comparable to the case cohort. Although this design is not a formal incidence-density sample, it closely approximates a nested case–control structure within the source population of patients undergoing abdominal imaging.

Dataset allocation

The combined dataset included 219 pre-diagnostic PDA CT scans and 1243 control CT scans (total n=1462). To enable both model development and independent evaluation, the dataset was randomly divided into two parts:

- Training set: 969 scans (156 pre-diagnostic, 813 control)
- Independent test set: 493 scans (63 pre-diagnostic, 430 control)

This allocation provided a sufficiently powered training cohort while preserving a separate test cohort for unbiased evaluation of the REDMOD framework. Figure 1 shows the workflow for constructing the pre-diagnostic and control cohorts, stratifying the pre-diagnostic cohort by lead time, and the train–test split.

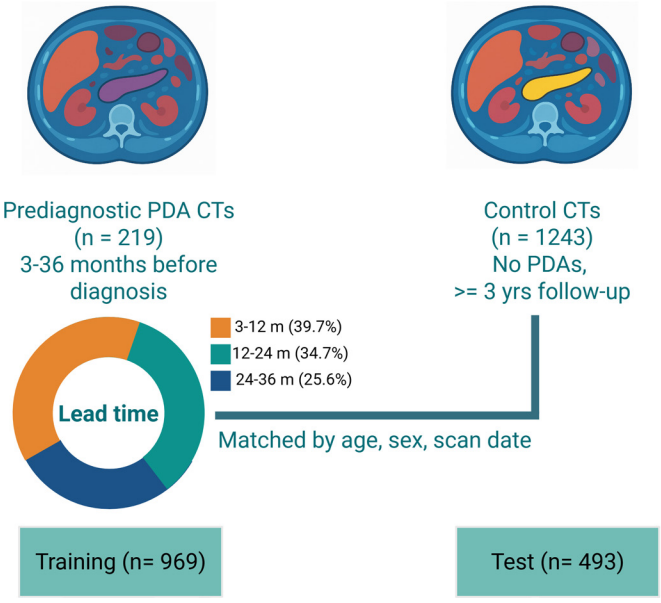

Figure 1 Cohort construction and dataset split. Pre-diagnostic CT scans from patients who later developed pancreatic ductal adenocarcinoma (PDA; n=219) were obtained 3–36 months before diagnosis and grouped by lead time. Control CT scans (n=1243) were from patients without PDA and with ≥3 years of follow-up. The final dataset was split into training (n=969) and test (n=493) cohorts.

Assessment of temporal detection window and external generalisability within the test subset

To assess the practical utility of REDMOD, two complementary analyses were performed on the pre-diagnostic CT scans within the test subset:

- Temporal detection window: for each subject the interval between the pre-diagnostic CT and histopathological diagnosis of PDA was calculated. Lead times were stratified into three clinically relevant bins: 3–12 months (n=20), 12–24 months (n=24) and >24 months (n=19). Model sensitivity was then evaluated across these intervals to identify the earliest stage at which REDMOD could reliably detect subtle radiomic signatures of stage 0 PDA.
- External generalisability: to test robustness across acquisition environments, scans were stratified by source institution. Of the 63 pre-diagnostic CT scans, 18 (28.6%) were obtained at our hospital and 45 (71.4%) were acquired externally. This allowed assessment of the stability of REDMOD across diverse scanners, acquisition protocols and patient populations beyond the primary training environment.

Fully automated AI-driven volumetric pancreas segmentation

A foundational component of the REDMOD framework is the precise and reproducible delineation of the pancreas, which is essential for subsequent radiomic feature extraction. To overcome the inherent limitations and variability of manual contouring and to ensure the scalability required for large-scale quantitative analyses, we implemented a fully automated volumetric pancreas segmentation algorithm. This segmentation was performed using a deep learning model based on the 3D nnU-Net architecture.[20] Its development, rigorous validation and detailed performance characteristics are reported in our recent publication.[18]

Radiomic feature extraction and reduction

A comprehensive panel of 968 quantitative radiomic features was extracted from each automatically segmented pancreas following a standardised preprocessing pipeline that included multiscale image filtering (wavelet and Laplacian-of-Gaussian (LoG)). To reduce dimensionality and select for the most informative biomarkers, the Minimum Redundancy Maximum Relevance (mRMR) algorithm[21] was employed, yielding a final set of 40 features for model input. A detailed description of the image preprocessing, feature computation and selection methodology is shown in online supplemental methods section S3.

Ensemble classifier construction, balanced training and feature importance analyses

To develop a classification model capable of distinguishing subtle pre-diagnostic PDA signatures from normal pancreatic tissue, several advanced ML strategies were employed, focusing on addressing class imbalance, leveraging diverse algorithmic strengths and ensuring model interpretability.

Addressing class imbalance in training

A common challenge in developing models for rare diseases like PDA is the extreme class imbalance in representative datasets, where pre-diagnostic scans are significantly outnumbered by controls. To mitigate the risk of the model becoming biased towards the majority class and to enhance its sensitivity to the minority (pre-diagnostic) class, the SMOTE was applied to the training subset. SMOTE generates synthetic data points for the minority class by interpolating between existing instances in the feature space, thereby creating a more balanced 1:1 class distribution for model training without simply duplicating existing data. This approach allows the learning algorithms to better discern the defining characteristics of the pre-diagnostic state.

Heterogeneous ensemble classifier development

Recognising that different ML algorithms capture distinct aspects of complex data, we constructed a heterogeneous ensemble classifier to synergise the predictive capabilities of three complementary algorithms:

- Logistic regression (LR): included for its ability to provide well-calibrated probabilistic outputs and establish a robust linear decision boundary.
- Random forest (RF): an ensemble of decision trees used for its strength in modelling non-linear relationships within the radiomic feature space and its inherent variance reduction through bootstrap aggregation.
- Extreme gradient boosting (XGBoost): a gradient boosting framework that sequentially builds decision trees, with each new tree correcting the errors of its predecessors, known for its high predictive accuracy and regularisation capabilities.

The hyperparameters for each individual algorithm within the ensemble were optimised using a randomised search strategy evaluated through stratified fivefold cross-validation on the SMOTE-augmented training data to prevent overfitting. The final prediction from the ensemble was derived using a soft-voting mechanism, where the class probability outputs from each of the three optimised models were averaged. The class with the highest mean probability was then assigned as the final classification. The complete REDMOD pipeline—from fully automated volumetric pancreas segmentation through radiomic feature extraction and selection to the construction of an ensemble model for final prediction—is shown in figure 2.

Determination of optimal classification threshold

To convert the probabilistic outputs of the ensemble into a binary classification (pre-diagnostic vs control), an optimal decision threshold was determined using the Youden Index. This index was calculated from the receiver operating characteristic (ROC) curve generated on the training subset, identifying the threshold that maximised the difference between the true positive rate (sensitivity) and the false positive rate (1 – specificity). This objectively derived threshold was then fixed and applied to the independent test subset for all subsequent performance evaluations.

Feature importance analysis and model interpretability

To gain insights into the radiomic biomarkers most critical for the detection of stage 0 PDA, a feature ablation study was conducted using the final selected feature set (n=40). Features were systematically removed one at a time and the resultant impact on the sensitivity of the ensemble model on the training set was recorded. This process helped identify features whose absence most significantly degraded detection performance, thereby highlighting the key imaging signatures associated with the earliest detectable stages of disease and offering a degree of model interpretability.

Error analysis

Finally, all CT scans misclassified by REDMOD on the test subset were independently reviewed by two board-certified abdominal radiologists (who were not involved in the reader study) to identify any potential morphological findings, image quality issues or other technical factors that might have contributed to the classification errors, providing qualitative insights for future model refinement.

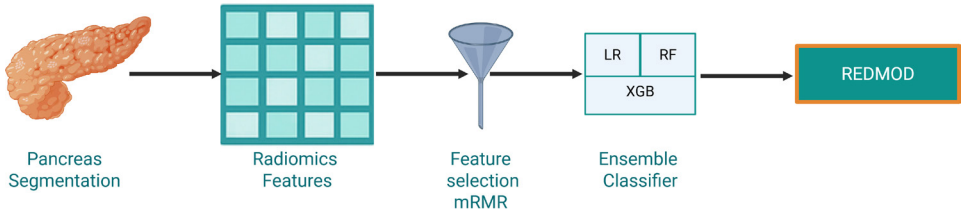

Figure 2 Model pipeline for REDMOD. The pipeline includes fully automated AI-based pancreas segmentation followed by extraction of radiomic features (first order and grey level: original and wavelet/Laplacian-of-Gaussian filters). Feature selection was performed using the minimum redundancy maximum relevance (mRMR) method. Selected features were used to train an ensemble classifier comprising logistic regression (LR), random forest (RF) and extreme gradient boosting (XGB). The final ensemble model, REDMOD, combines predictions from all three classifiers using a 'soft voting' mechanism where the class probability outputs from each of the three optimised models were averaged and the class with the highest mean probability was assigned as the final classification.

Methodological validation and component analyses

To evaluate the impact of key architectural choices and assess methodological robustness, we performed targeted component analyses which included the independent contributions of original versus multiscale filtered radiomic features, the performance of the ensemble classifier against its individual components and the stability of the model's predictions against expert refinement of the automated pancreas segmentations. The detailed methodology for these validation studies is provided in online supplemental methods section S4.

Benchmarking against clinical practice and evaluating model reliability

To evaluate the clinical translation potential of REDMOD we conducted a series of analyses to benchmark its performance against current practice and to assess its reliability and generalisability.

Comparing REDMOD against expert radiologists

A head-to-head multireader study was conducted to directly compare REDMOD with human interpretation.

- Readers: two board-certified abdominal radiologists (3 years post-residency experience).
- Dataset: all 493 CT scans in the independent test subset (63 pre-diagnostic, 430 control).
- Design: readers were blinded to all clinical and group labels. Even the case–control ratio was not disclosed. Using a standardised 5-point ordinal scale (1=definitely normal; 5=definitely abnormal), they rated the likelihood of stage 0 PDA and annotated any relevant findings.
- Purpose: this design enabled (1) direct comparison of the incremental diagnostic value of REDMOD and (2) assessment of interobserver variability.
- Analysis: radiologist performance metrics were calculated using the same ground-truth labels as REDMOD for consistency.

Longitudinal stability and test–retest reliability

To be viable for screening, REDMOD must produce stable predictions across serial imaging. A test–retest analysis was therefore performed on subjects with temporally distinct CT scans.

- Control cohort: when multiple eligible scans were available, a secondary scan within ±12 months of the index was selected. To maximise robustness, the scan with the greatest temporal separation was prioritised.
- Pre-diagnostic cohort: for available cases, the CT scan immediately prior to the index scan was selected to minimise interval length. This approach ensured that concordance testing reflected the stability of radiomic signatures already present in early carcinogenesis rather than being confounded by disease evolution or by intervals too remote to reliably capture a signal.

All secondary scans were processed through the same pre-trained REDMOD framework without retraining. Predictions for each secondary scan were compared with its corresponding index scan to quantify concordance.

Validation of specificity on multi-institutional and public datasets

Given the absence of publicly available pre-diagnostic PDA datasets for external validation of sensitivity, we assessed the specificity of REDMOD—its ability to correctly identify truly normal pancreas—on two independent cohorts across diverse settings and institutions.

- Multi-institutional normal CT validation set: this independent dataset included 539 CT scans from patients referred to our quaternary care centre who showed no evidence of PDA during at least 3 years of follow-up. It presented a robust challenge due to varied scanner vendors (Siemens: 58%, Toshiba: 33%, GE: 9%) and slice thicknesses (0.6–5 mm).
- Public NIH-PCT dataset: the publicly available National Institutes of Health–Pancreas Computed Tomography (NIH-PCT) dataset,[22] comprising 80 CT scans from healthy volunteers with no pancreatic pathology, was also acquired on different scanners (Siemens and Philips).

This multi-pronged validation strategy, shown in figure 3, was designed to build confidence in the negative predictive performance of REDMOD across a wide range of clinical scenarios and institutional contexts.

Statistical analyses

Statistical analyses were performed using R and Python. The primary endpoint was the area under the curve (AUC), with 95% CIs calculated via bootstrapping. The optimal classification threshold was determined using the Youden Index on the training set. Model AUCs were compared using DeLong's test. A rigorous Multi-Reader Multi-Case (MRMC) ROC analysis was used for the comparison between REDMOD and radiologists. Comprehensive statistical methodologies are shown in online supplemental methods section S5.

RESULTS

Pre-diagnostic cohort is characterised by smaller pancreatic volume

The study included 219 patients with pre-diagnostic PDA (median age 69 years, range 34–88; male to female ratio 1.3) and 1243 control subjects (median age 64 years, range 34–88; ratio 1.4), with comparable demographics across training and test subsets. The median interval from the pre-diagnostic CT scan to histopathological diagnosis of PDA was 427 days (range

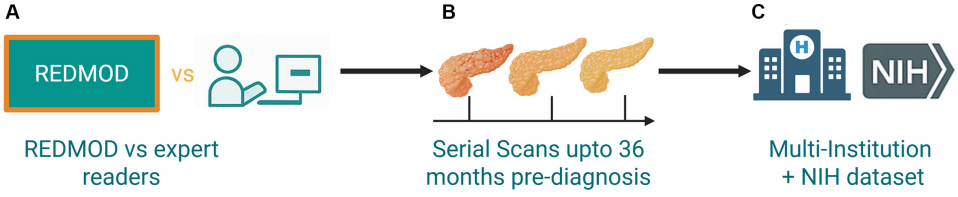

Figure 3 Validation strategy for REDMOD. (A) Reader study: a blinded comparison of REDMOD and two board-certified abdominal radiologists was performed on the independent test set. (B) Longitudinal robustness: test–retest analysis assessed the stability of REDMOD predictions across serial scans. (C) External validation: the specificity of the model was evaluated in a multi-institutional cohort and a public National Institutes of Health (NIH) dataset.

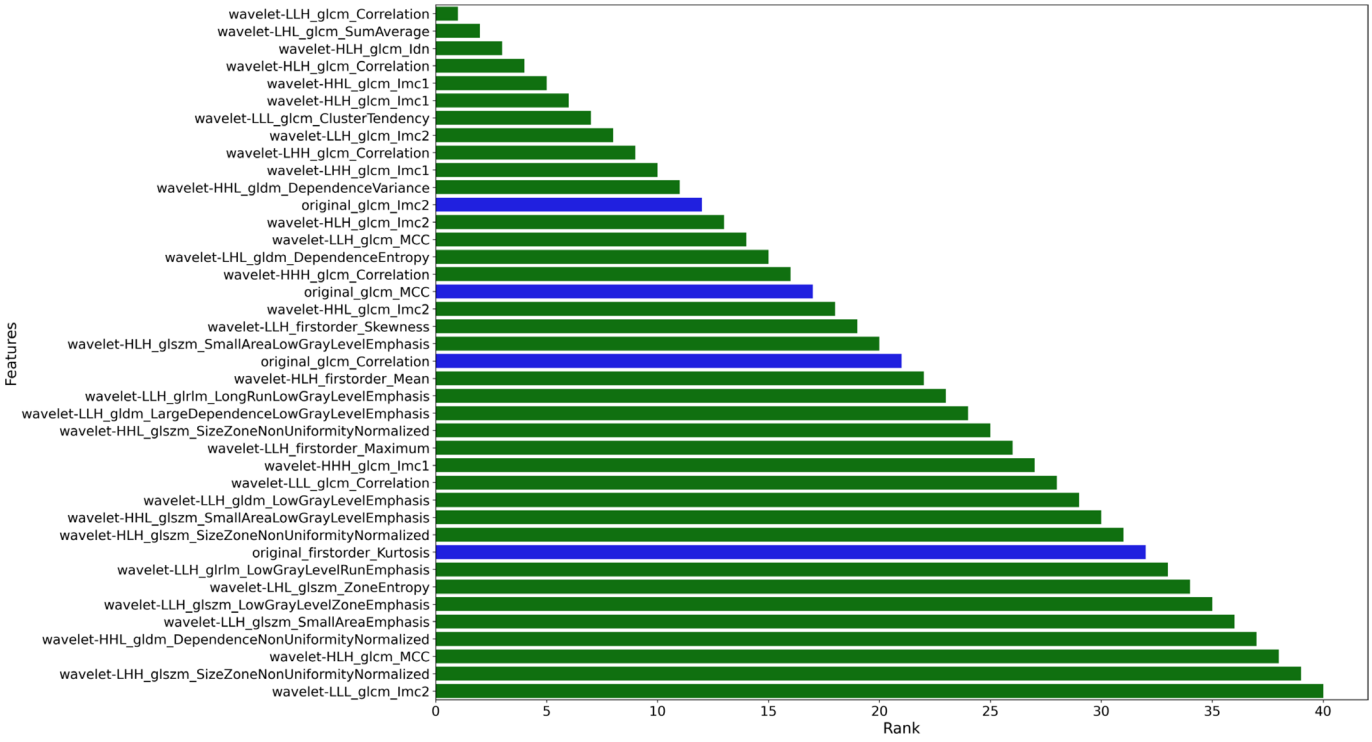

Figure 4 Forty radiomic features selected by the minimum redundancy maximum relevance (mRMR) method for the ensemble REDMOD model. The selected features included four first-order features and 36 texture features. Among these, 36 were filtered features (green) and four were original features (blue). GLCM, grey-level co-occurrence matrix; GLDM, grey-level dependence matrix; GLSZM, grey-level size zone matrix; GLRLM, grey-level run-length matrix.

90–1092 days) for the full cohort and 475 days (range 97–1076 days) for the 63 cases in the test subset. Lead time distribution was 87 patients (39.7%) at 3–12 months, 76 (34.7%) at 12–24 months and 56 (25.6%) at 24–36 months. Subsequent diagnostic imaging localised PDA primarily to the pancreatic head (64%), with fewer cases in the body (18%) and tail (18%).

AI-driven automated segmentation showed significantly smaller pancreatic volumes in the pre-diagnostic group than in the controls (68.9 (23.8) mL vs 74.7 (23.6) mL; p=0.005). A similar trend was observed in the test subset (67.1 (22.7) mL vs 72.9 (24.1) mL), but this did not reach statistical significance (p=0.09). Pancreatic volumes from independent validation cohorts were consistent, measuring 71.2 (24.8) mL in the multi-institutional set (n=539) and 68.7 (17.5) mL in the NIH-PCT dataset (n=80).

Robust predictive performance driven primarily by filtered radiomic features

From 968 initial radiomic features per pancreatic volume, the mRMR algorithm selected 40 for the REDMOD ensemble classifier. These included four first-order statistical features and 36 higher-order grey-level textural features, of which 36 (90%) originated from wavelet-filtered images, highlighting the central role of multi-scale filtered information in capturing subtle pre-clinical pancreatic alterations. Figure 4 shows the 40 features selected by mRMR.

On the test subset (n=493; 63 pre-diagnostic, 430 control), REDMOD achieved an AUC of 0.82 (95% CI 0.81 to 0.83) for detecting stage 0 PDA. At the optimal classification threshold (0.41, derived from the Youden Index), the model correctly identified 46 of 63 pre-diagnostic CT scans, yielding a sensitivity of 73.0% (95% CI 60.0% to 78.7%), and 349 of 430 control CT scans, yielding a specificity of 81.1% (95% CI 75.2% to 93.1%). The overall performance was strong with an accuracy of 80.1% (95% CI 75.3% to 88.6%), precision of 36.2% (95% CI 31.5% to 60.9%) and F1 score of 48.4% (95% CI 44.3% to 58.1%). Based on the observed AUC of 0.82, the control–case ratio of ~6.8:1 and a significance level (α of 0.05), our test subset size of 493 subjects provided >80% statistical power to detect the observed effect size. This post-hoc analysis confirms that, despite the inherent constraints in accruing pre-diagnostic CT scans, the study was sufficiently powered to validate the performance of REDMOD. Furthermore, given that REDMOD was trained on SMOTE-balanced data but applied to an imbalanced test subset (~14% prevalence), we evaluated the potential for calibration drift using post-hoc calibration (see online supplemental results section S6) to ensure the reliability of the probabilistic outputs in a clinically representative setting. The findings (Expected

Table 1 Comparison of model performance with original, filtered and combined radiomic features

	Model with original features	Model with filtered features	REDMOD (original + filtered)
Metric	Value (95% CI)	Value (95% CI)	Value (95% CI)
Sensitivity	66.6 (53.1 to 76.2)	73.0 (61.9 to 80.2)	73.0 (60.0 to 78.7)
Specificity	73.2 (66.0 to 85.7)	79.7 (73.8 to 91.2)	81.1 (75.2 to 93.1)
Accuracy	72.4 (66.7 to 81.2)	78.9 (74.3 to 87.7)	80.1 (75.3 to 88.6)
Precision	26.7 (23.3 to 39.1)	34.6 (33.1 to 56.9)	36.2 (31.5 to 60.9)
F1 score	38.2 (34.6 to 46.1)	46.9 (45.9 to 58.1)	48.4 (44.3 to 58.1)
AUC	0.74 (0.71 to 0.76)	0.82 (0.81 to 0.84)	0.82 (0.81 to 0.83)

All values are percentages (%) except AUCs.
AUC, area under the curve.

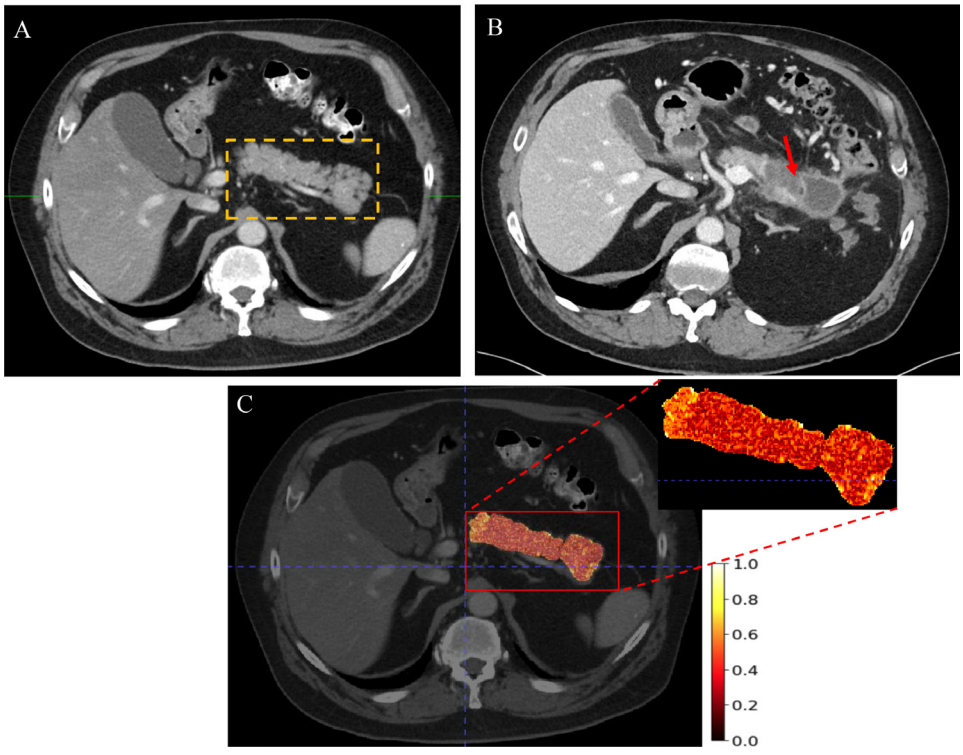

Figure 5 Identification of occult radiomic signatures by REDMOD in a pre-diagnostic CT scan 2.4 years prior to clinical diagnosis. (A) Axial contrast-enhanced CT scan of a 63-year-old man, prospectively interpreted as normal, with the pancreas outlined (yellow dashes). (B) A diagnostic CT scan from the same patient 2.4 years later shows a large pancreatic ductal adenocarcinoma (red arrow). (C) Radiomic texture maps generated by the REDMOD framework, overlaid on the original pre-diagnostic CT scan, visualise the spatial distribution of a key predictive feature (wavelet-filtered GLCM-IMC2). The colour map indicates that areas of high feature expression (red/yellow) are concentrated in the region of the pancreas where the tumour subsequently developed.

Calibration Error 0.04, Brier score 0.189) (see online supplemental figure 1) support that REDMOD provides both high discrimination and reliable risk estimation for an early detection paradigm.

Radiologist review of misclassified scans did not reveal systemic anatomical abnormalities or technical artefacts to explain the errors. Two radiologists, blinded to model predictions, also stratified the test subset into a 'normal pancreas' subgroup (n=341) and a 'confounding imaging features' subgroup (n=152) with the latter comprising cases with cystic lesions, fatty replacement, ductal dilation or irregularity, calcifications or distinct fibrosis/scarring and focal atrophy. We found no statistically significant difference in performance between these groups (AUC 0.79 vs 0.82; DeLong test p=0.90).

Finally, the REDMOD performance remained stable across technical variations: accuracy was comparable by CT vendor (Siemens 78.8% vs non-Siemens 81.8%; p=0.48) and by slice thickness (≤3 mm 80.7% vs >3 mm 76.6%; p=0.49), indicating resilience to common variations in image acquisition.

High sensitivity for stage 0 PDA detection >2 years prior to diagnosis

A critical measure of early detection is the lead time achievable prior to clinical diagnosis. REDMOD maintained high sensitivity across extended pre-diagnostic intervals:

- 3–12 months before diagnosis: 75.0% (15/20)
- 12–24 months before diagnosis: 75.0% (18/24)
- >24 months before diagnosis: 68.4% (13/19)

Importantly, performance was consistent across sources with comparable sensitivity for scans from our institution (72.2% (13/18)) and from external hospitals (73.3% (33/45)), suggesting

Table 2 Comparison of individual machine learning models against the ensemble REDMOD model

	Logistic regression	Random forest	XGBoost	Ensemble (REDMOD)
Metric	Value (95% CI)	Value (95% CI)	Value (95% CI)	Value (95% CI)
Sensitivity	63.5 (49.2 to 73.1)	61.9 (56.3 to 75.4)	63.5 (57.1 to 79.4)	73.0 (60.0 to 78.7)
Specificity	81.1 (70.5 to 93.6)	92.3 (80.1 to 94.6)	90.0 (70.5 to 92.4)	81.1 (75.2 to 93.1)
Accuracy	78.9 (70.7 to 87.9)	88.4 (79.0 to 89.7)	86.6 (71.7 to 88.4)	80.1 (75.3 to 88.6)
Precision	33.0 (25.8 to 53.3)	54.2 (34.2 to 60.7)	48.2 (27.9 to 54.2)	36.2 (31.5 to 60.9)
F1 score	43.5 (37.3 to 51.5)	57.7 (45.8 to 59.3)	54.8 (41.4 to 57.7)	48.4 (44.3 to 58.1)
AUC	0.78 (0.74 to 0.80)	0.82 (0.80 to 0.84)	0.81 (0.78 to 0.83)	0.82 (0.81 to 0.83)

All values are percentages (%) except AUCs.
AUC, area under the curve.

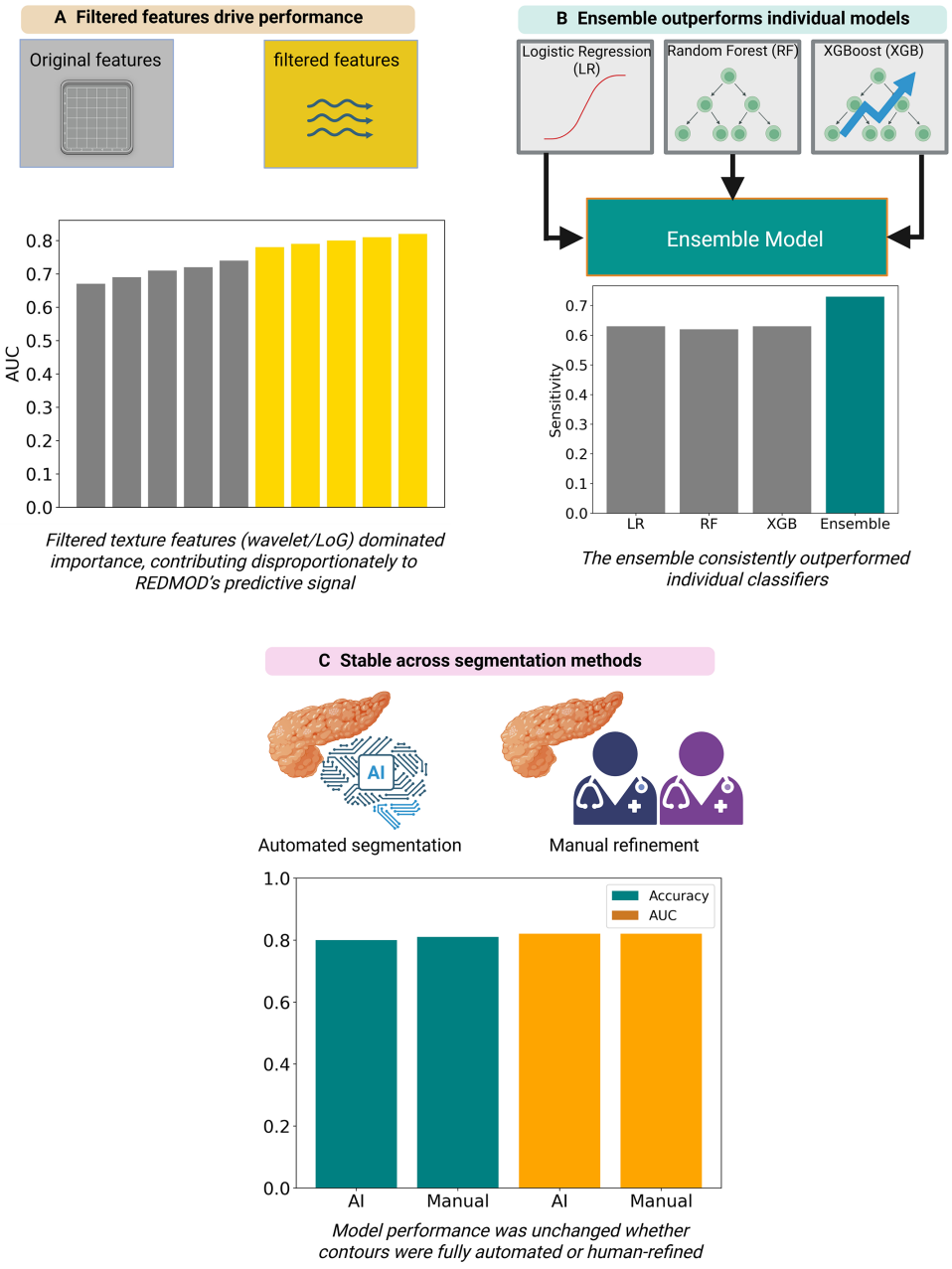

Figure 6 Methodological validation of REDMOD. (A) Filtered texture features provided superior predictive performance compared with original unfiltered features. (B) An ensemble learning approach consistently outperformed individual models in sensitivity. (C) Model performance remained stable whether using fully automated or manually refined pancreas segmentations.

generalisability across diverse imaging environments. Likewise, the sensitivity to predict pre-diagnostic PDA was spatially invariant (p=0.36), with sensitivities being 79% (30/38) for head lesions, 58% (7/12) for body lesions, 73% (8/11) for tail lesions and 50% (1/2) for multifocal lesions.

High specificity across multi-institutional and public control cohorts

In the large external validation set of 539 control CT scans, REDMOD achieved a specificity of 81.3% (438/539), mirroring its performance on the test subset. On the public NIH-PCT dataset (n=80), specificity was even higher at 87.5% (70/80), further supporting robust generalisability in correctly classifying normal pancreatic tissue.

Ablation studies demonstrate the critical role of filtered features and ensemble architecture

Contribution of filtered radiomic features

Models built exclusively with filtered radiomic features significantly outperformed those using only unfiltered features (AUC 0.82 (95% CI 0.81 to 0.84) vs AUC 0.74 (95% CI 0.71 to 0.76); p=0.007). Performance with filtered-only features was comparable to the full REDMOD model (AUC 0.82; p=0.54), indicating that wavelet and LoG-derived textural information were the primary drivers of predictive capability (table 1).

This was consistent with the finding that 90% of features selected for REDMOD were filter-derived. The most influential features identified by ablation included wavelet-based GLCM (Informational Measure of Correlation (IMC) 1 and 2), GLSZM

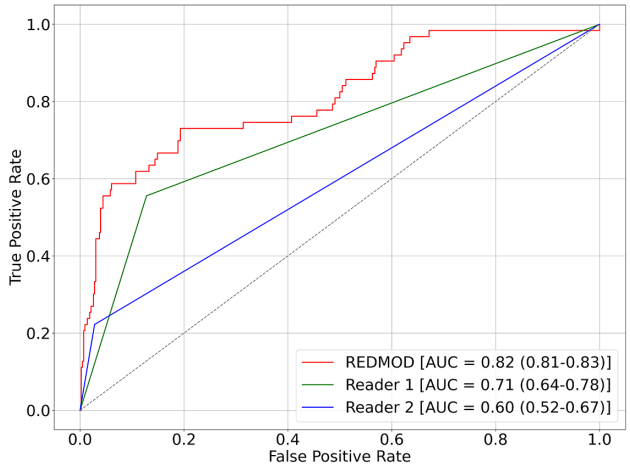

Figure 7 ROCs of the ensemble REDMOD and two radiologist readers on the test subset (n=493; 430 control, 63 pre-diagnostic). False positive rate: 1 – specificity; true positive rate: sensitivity. The model significantly outperformed two radiologist readers for the pancreas assessment (control or pre-diagnostic) of the test subset.

(Size Zone Non-Uniformity Normalised) and GLDM (Large Dependence Low Grey Level Emphasis). Removal of these top features reduced sensitivity from 73.0% to 66.6%, underscoring their central role in capturing subtle multiscale textural biomarkers. In an illustrative case (figure 5), the radiomic textural pattern of wavelet-filtered GLCM-IMC2 on a pre-diagnostic CT scan showed focal high-expression regions within the pancreas that corresponded to the site where PDA later developed.

Benefit of ensemble classification

While individual classifiers (logistic regression, random forest, XGBoost) showed varied performance (AUCs 0.78, 0.82 and 0.81, respectively), the REDMOD ensemble (AUC 0.82) provided a modest but consistent advantage. Notably, it achieved the highest sensitivity (73.0%) across configurations. This demonstrates that the soft-voting ensemble strategy contributed to a more balanced and robust overall classification (table 2).

Performance is robust to variations in pancreas segmentation

To test whether minor segmentation inaccuracies drove classification errors we compared the predictions of REDMOD using fully automated contours versus human-refined contours. Of the 98 cases misclassified in the test subset (17 pre-diagnostic, 81 control), radiologists recommended minor segmentation corrections in 50 instances (9 pre-diagnostic, 41 control). Substituting radiomic features from these refined segmentations produced nearly identical results to the originals (accuracy 81.1% vs 80.1%; specificity 82.3% vs 81.1%; AUC 0.82; p=0.62–1.0). These findings confirm that the performance of REDMOD is not materially affected by small segmentation variations, underscoring the reliability of the fully automated pipeline and its ability to extract stable radiomic signatures in a high-performance real-world setting. Figure 6 shows the methodological validation strategies of REDMOD.

REDMOD outperforms expert radiologists

In the head-to-head evaluation, REDMOD showed significantly superior accuracy for detecting pre-diagnostic PDA compared with two independent radiologists (AUC 0.82 vs 0.69; p<0.001; figure 7). REDMOD achieved a sensitivity of 73.0% (95% CI 61.0% to 82.4%), nearly double that of the pooled radiologists (38.9% (95% CI 30.2% to 47.5%); p<0.001).

Given the limited number of pre-diagnostic CT scans within each lead time bin, formal statistical comparison was not feasible but qualitative evaluation indicated a progressive widening of the sensitivity advantage of REDMOD with increasing diagnostic interval:

- 3–12 months: 75.0% vs 50.0% (1.5-fold higher)
- 12–24 months: 75.0% vs 41.0% (1.8-fold higher)
- >24 months: 68.0% vs 23.0% (nearly 3-fold higher)

Performance of the individual radiologists further highlighted the trade-off between sensitivity and specificity. Reader 1 had sensitivity of 55.5% with specificity of 87.2%, while Reader 2 showed lower sensitivity (22.2%) but higher specificity (97.2%). In contrast, REDMOD achieved both balanced and superior performance, with sensitivity of 73.0% and specificity of 81.1%. Notably, inter-reader agreement between the radiologists was modest (weighted Cohen's κ=0.22), underscoring the variability and subjectivity inherent to human interpretation.

High concordance and reliability on longitudinal scans

The REDMOD framework showed stable predictive performance in a longitudinal test–retest analysis limited to subjects with serial CT scans available.

- Control cohort (n=61): among subjects with serial CT scans available within ±12 months of the index test CT scan, the predictions of REDMOD were concordant on retest. Consistency reached 92% (47/51) for cases initially classified as true negatives and 70% (7/10) for cases initially classified as false positives.
- Pre-diagnostic cohort (n=10): using earlier pre-index CT scans performed a median of 210 days before the index CT scan (range 6–761 days), the model preserved its true positive classification in 90% (9/10). For these correctly classified serial CT scans the median lead time to PDA diagnosis was 675 days (range 505–993).

Together, these results show that outputs of REDMOD remained reproducible across temporally distinct studies, which is further illustrated in figure 8.

DISCUSSION

This study validates REDMOD as a fully automated AI framework that can identify the imaging signature of stage 0 PDA from routine CT scans with substantial lead times and performance superior to expert radiologists. By leveraging a large clinically representative dataset and advanced AI methodologies, this work establishes the clinical validity of the REDMOD framework for pre-clinical PDA detection, consistent with a phase 3 biomarker validation as defined by the NCI's Early Detection Research Network.[23]

AI-driven radiomics as a solution to the challenge of stage 0 PDA detection

Early detection of PDA at a curable stage is a formidable challenge due to its imperceptibility on conventional imaging during the crucial pre-clinical window (stage 0).[4 24 25] This 'invisibility' of early-stage disease on standard imaging, combined with the aggressive and rapid progression characteristic of PDA,[26] mandates the development of innovative solutions. While studies on smaller datasets have shown initial promise in identifying pre-diagnostic signals,[9–13] REDMOD was evaluated on a substantially larger (n=1462) multi-institutional cohort to closely reflect the real-world early detection prevalence (~6.8:1

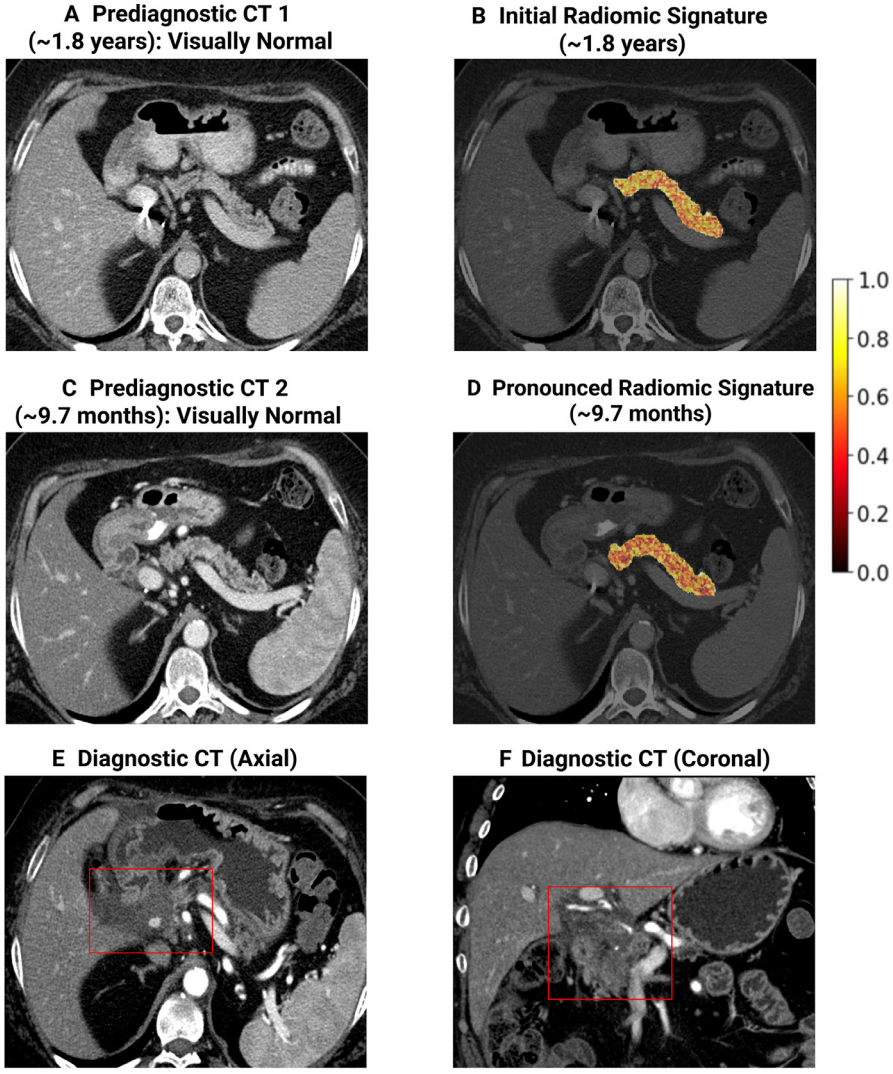

Figure 8 Longitudinal tracking of pre-clinical pancreatic ductal adenocarcinoma (PDA). A patient correctly identified by REDMOD at ~1.8 years (A) and ~9.7 months (C) prior to clinical diagnosis. The corresponding radiomic feature maps (B, D) show an evolving intensifying signature in a normal pancreas. Diagnostic CT scans (E, F) later confirmed tumour development in the highlighted region.

control–pre-diagnostic case ratio). Under these stringent conditions, REDMOD detected the subtle textural and architectural signatures of stage 0 PDA with high sensitivity (73%) and accuracy (AUC 0.82) at a median lead time of 475 days. This temporal window holds profound significance, as attaining such early detection would substantially augment the probability of cure and improved survival. In fact, modelling studies indicate that increasing the proportion of localised PDAs from 10% to 50% would more than double survival rates,[2 27 28] thereby underscoring that the timing of diagnosis is the single most critical determinant of survival outcomes.

Furthermore, REDMOD demonstrated superior accuracy to radiologists (AUC 0.82 vs 0.69; $p<0.001$). It is imperative to contextualise these metrics within the operational realities of an early detection paradigm. In cohorts characterised by a high control–case ratio, the AUC metric can be misleading because increased specificity across a substantial proportion of normal cases can disproportionately inflate the AUC, thereby obscuring a clinically significant deficiency in sensitivity. This statistical fallacy is evident in the performance metrics of the radiologists. Their high specificities (87–97%) yielded a respectable AUC despite a pooled sensitivity of merely 38.9% and substantial interobserver variability ($\kappa=0.22$). The paramount objective of an early detection test is to achieve a clinically efficacious equilibrium: to accurately identify rare instances of disease (sensitivity) while concurrently maintaining an acceptably low false-positive rate (specificity). In this regard, REDMOD showed a superior equilibrium, with a profoundly greater sensitivity (73.0% vs 38.9%) while simultaneously preserving a robust specificity (81.1%). This balanced performance, particularly a nearly threefold higher sensitivity (68.0% vs 23.0%) at the longest lead times (>24 months), suggests that the primary strength of REDMOD lies in its capacity to identify the earliest imaging signatures of visually occult disease. Finally, this complementary performance profile—higher sensitivity from REDMOD and high specificity from radiologists—suggests that an AI-augmented workflow could lead to a synergistic improvement in overall accuracy for early PDA detection. In the absence of visible tumour cues, however, automation bias—where readers uncritically accept AI predictions—remains a risk. To address this, our upcoming prospective AI-PACED (Artificial Intelligence for Pancreatic Cancer Early Detection) trial[29] will quantify automation bias and establish optimal override criteria so that algorithmic alerts

prompt appropriate risk-stratified evaluation rather than premature intervention.

At the prespecified threshold (0.41; Youden Index), REDMOD delivered 36.2% precision while maintaining 73.0% sensitivity, notably at a visually occult stage where all CT scans had been prospectively reported as negative for PDA. Its precision already exceeds the 3% precision recommended by NICE at the first step of UK cancer referral, where avoiding missed cancers takes precedence over false positives.[7] Established early detection programmes reinforce this principle: low-dose CT and mammography accept lower precision initially, relying on follow-up imaging to adjudicate borderline findings[30 31] to boost precision. For instance, nodule volume and doubling time stratify lung cancer risk after low-dose CT,[32] while mammographic comparison reduces recall rates while balancing initial sensitivity.[33] In this context, the 90–92% test–retest concordance of REDMOD and signal persistence of at least 12–24 months support an 'observe-and-retest' strategy for intermediate results. Moreover, unlike low-dose CT and mammography, which are predicated on radiologically visible abnormalities, REDMOD identifies PDA signatures before any lesion is apparent. Thus, it extends the window of detection into a pre-diagnostic phase.

Conversely, there is a need for higher precision when invasive procedures are considered. Guided by these principles, the Youden Index-derived operative threshold of REDMOD was intentionally designed to be adjustable to pretest risk. Crucially, this design allows recalibration without retraining, enabling thresholds to be shifted to emphasise sensitivity or precision in accordance with clinical context. Weighted or asymmetrical indices can further penalise false positives where procedural risk is high. This flexibility permits alignment with clinical objectives across the pathway—non-invasive triage early, confirmatory imaging next and stringent precision before surgery.

Methodological advances and their implications for robust early detection

REDMOD integrates automated pancreas segmentation[18] for scalability and reproducibility. Our analysis confirmed the robustness of the downstream classification to minor segmentation inaccuracies (AUC unchanged at 0.82; p=0.62–1.0). Furthermore, unlike approaches that focus on a limited number of pre-selected features, REDMOD used an extensive feature engineering pipeline. This pipeline commenced with 968 radiomic features per patient and employed the mRMR algorithm to derive a parsimonious and highly predictive set of 40 features. A key insight is the predominance (90% of selected features) and superior contribution of filtered (wavelet and LoG) features vis-à-vis original unfiltered features (AUC 0.82 vs 0.74, p=0.007) to the predictive power of REDMOD. This suggests that the earliest manifestations of PDA are subtle multiscale textural disruptions and alterations in local intensity gradients, which are more effectively captured by these advanced feature engineering techniques. To elucidate the mechanistic basis of REDMOD, an ablation study systematically removed feature classes to quantify performance degradation. Textural heterogeneity features (eg, specific wavelet-transformed GLCM, GLSZM and GLDM metrics) emerged as the most predictive, which aligns with prior evidence that early carcinogenesis disrupts pancreatic parenchymal architecture at a microscopic level before mass formation.[34 35] These findings suggest that REDMOD detects biologically relevant remodelling associated with early malignancy. Furthermore, the use of a heterogeneous ensemble classifier, which averaged the probabilistic outputs of three distinct algorithms, had the highest sensitivity (73.0%) among all configurations with comparable AUC (0.82), suggesting the soft-voting mechanism provides a more balanced classification.

Limitations and future directions

We acknowledge the limitations of this study. First, the study was not designed to evaluate performance across different racial and ethnic groups, a critical consideration for future validation given known disparities in PDA risk among individuals with gNOD.[36] Second, while specificity was validated on external cohorts, sensitivity requires validation on public pre-diagnostic datasets which are not yet available. Third, the REDMOD framework was intentionally designed to assess the entire pancreatic gland, reflecting the hypothesis that stage 0 disease manifests as a diffuse field effect rather than a discrete focal lesion. At this early visually occult stage, there is no discernible lesion to serve as ground truth for training a localisation algorithm. Consequently, REDMOD could serve a distinct and critical purpose as an advanced risk-enrichment tool. A 'high-risk' flag from the model could prompt further investigation with molecular imaging (eg, using fibroblast activation protein (FAP)-targeted positron emission tomography radiotracers) capable of pinpointing areas of cellular activity to direct endoscopic ultrasound-guided sampling. Moreover, the therapeutic landscape for early stage PDA is undergoing rapid evolution. The emergence of systemic interception strategies, such as novel multi-antigen nanovaccines engineered to elicit a curative antitumour immune response throughout the organ,[37] may obviate the necessity for lesion localisation. Within such a paradigm, the primary clinical imperative transitions from identifying a focal lesion for resection to identification of the high-risk individual (HRI) whose immune system warrants activation, a role for which a robust whole-gland risk assessment tool like REDMOD is eminently suitable. Finally, this study validates the diagnostic performance of REDMOD rather than PDA survival outcomes, so classical screening biases such as lead time and length time bias do not directly affect our primary endpoints. Nonetheless, future prospective evaluations must apply methods to account for these biases and competing risks of death, as in hepatocellular carcinoma surveillance,[38 39] to define the true net benefit of early interception.

CONCLUSION

This study validates REDMOD as a fully automated AI framework capable of identifying the imaging signatures of stage 0 PDA in normal pancreas, achieving this with substantial lead times and performance superior to expert radiologists. The demonstrated ability of the framework to consistently detect these occult signals on a large clinically-oriented dataset, combined with its high longitudinal stability and validated specificity, establishes a robust foundation for AI-augmented early detection. This work overcomes key barriers in the field by providing a scalable objective tool that addresses a critical diagnostic gap. While prospective validation is paramount to confirm clinical utility, the REDMOD framework represents a significant advance towards shifting the paradigm for sporadic PDA from a late-stage symptomatic diagnosis to proactive pre-clinical interception, offering tangible hope for improving outcomes in this challenging disease.

Contributors SM: Study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis. AA: Acquisition of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. NGP: Acquisition of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. KHT: Acquisition of

data; analysis and interpretation of data; critical revision of the manuscript for important intellectual content. AK: Acquisition of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. KKB: Acquisition of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. AZ: Analysis and interpretation of data; critical revision of the manuscript for important intellectual content. JGF: Study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content. MT: Study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content. MPJ: Analysis and interpretation of data; critical revision of the manuscript for important intellectual content; statistical analysis. STC: Study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content. AHG: Study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; obtained funding; study supervision. AHG is the guarantor. All the authors have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analyses. All authors were responsible for the critical revision of the manuscript for important intellectual content.

Funding Funding was provided by National Institutes of Health (Grant numbers: R01CA272628 and R01CA256969), Mayo Clinic Comprehensive Cancer Center (Grant number: not applicable), Mayo Clinic Department of Radiology (Grant number: not applicable), Champions for Hope Pancreatic Cancer Research Program of the Funk Zitiello Foundation (Grant number: not applicable), Centene Charitable Foundation (Grant number: not applicable) and Hoveida Family Foundation (Grant number: not applicable).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This retrospective minimal risk case-control study received Mayo Clinic Institutional Review Board approval (ID: 19-003018), complied with the HIPAA and was granted a waiver of informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Access to datasets from the Mayo Clinic Foundation should be requested directly via their data access request forms. Subject to the institutional review boards' ethical approval, de-identified data could be made available. All experiments and implementation details are described thoroughly in the Methods section so they can be independently replicated with non-proprietary libraries.

ORCID iDs
Suresh T Chari https://orcid.org/0000-0002-3924-0971
Ajit Harishkumar Goenka https://orcid.org/0000-0002-7804-2695

REFERENCES

1 Siegel RL, Kratzer TB, Giaquinto AN, *et al*. Cancer statistics, 2025. *CA Cancer J Clin* 2025;75:10–45.
2 Kenner B, Chari ST, Kelsen D, *et al*. Artificial Intelligence and Early Detection of Pancreatic Cancer: 2020 Summative Review. *Pancreas* 2021;50:251–79.
3 Chari ST, Kelly K, Hollingsworth MA, *et al*. Early detection of sporadic pancreatic cancer: summative review. *Pancreas* 2015;44:693–712.
4 Antony A, Mukherjee S, Bhinder K, *et al*. Artificial Intelligence-Augmented Imaging for Early Pancreatic Cancer Detection. *Visc Med* 2025;2025:1–9.
5 Chari S, Leibson C, Rabe K, *et al*. Probability of Pancreatic Cancer Following Diabetes: A Population-Based Study. *Gastroenterology* 2005;129:504–11.
6 Sharma A, Kandlakunta H, Nagpal SJS, *et al*. Model to Determine Risk of Pancreatic Cancer in Patients With New-Onset Diabetes. *Gastroenterology* 2018;155:730–9.
7 National Institute for Health and Care Excellence (NICE). Suspected cancer: recognition and referral. NICE Guideline NG12. 2025. Available: https://www.nice.org.uk/guidance/ng12/chapter/Recommendations-organised-by-site-of-cancer [Accessed 26 Sep 2025].
8 Antony A, Mukherjee S, Bi Y, *et al*. AI-driven insights in pancreatic cancer imaging: from pre-diagnostic detection to prognostication. *Abdom Radiol (NY)* 2025;50:3214–24.
9 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* 2022;163:1435–46.
10 Mukherjee S, Korfiatis P, Patnam NG, *et al*. Assessing the robustness of a machine-learning model for early detection of pancreatic adenocarcinoma (PDA): evaluating resilience to variations in image acquisition and radiomics workflow using image perturbation methods. *Abdom Radiol (NY)* 2024;49:964–74.
11 Qureshi TA, Gaddam S, Wachsman AM, *et al*. Predicting pancreatic ductal adenocarcinoma using artificial intelligence analysis of pre-diagnostic computed tomography images. *Cancer Biomark* 2022;33:211–7.
12 Javed S, Qureshi TA, Gaddam S, *et al*. Risk prediction of pancreatic cancer using AI analysis of pancreatic subregions in computed tomography images. *Front Oncol* 2022;12:1007990.
13 Chen W, Zhou Y, Asadpour V, *et al*. Quantitative Radiomic Features From Computed Tomography Can Predict Pancreatic Cancer up to 36 Months Before Diagnosis. *Clin Transl Gastroenterol* 2023;14:e00548.
14 Khasawneh H, Patra A, Rajamohan N, *et al*. Volumetric Pancreas Segmentation on Computed Tomography: Accuracy and Efficiency of a Convolutional Neural Network Versus Manual Segmentation in 3D Slicer in the Context of Interreader Variability of Expert Radiologists. *J Comput Assist Tomogr* 2022;46:841–7.
15 Panda A, Korfiatis P, Suman G, *et al*. Two-stage deep learning model for fully automated pancreas segmentation on computed tomography: Comparison with intra-reader and inter-reader reliability at full and reduced radiation dose on an external dataset. *Med Phys* 2021;48:2468–81.
16 Korfiatis P, Suman G, Patnam NG, *et al*. Automated Artificial Intelligence Model Trained on a Large Data Set Can Detect Pancreas Cancer on Diagnostic Computed Tomography Scans As Well As Visually Occult Preinvasive Cancer on Prediagnostic Computed Tomography Scans. *Gastroenterology* 2023;165:1533–46.
17 Huang C, Hecht EM, Soloff EV, *et al*. Imaging for Early Detection of Pancreatic Ductal Adenocarcinoma: Updates and Challenges in the Implementation of Screening and Surveillance Programs. *AJR Am J Roentgenol* 2024;223:e2431151.
18 Mukherjee S, Antony A, Patnam NG, *et al*. Pancreas segmentation using AI developed on the largest CT dataset with multi-institutional validation and implications for early cancer detection. *Sci Rep* 2025;15:17096.
19 Ni Z, Zhu Y, Qian Y, *et al*. Synthetic minority over-sampling technique-enhanced machine learning models for predicting recurrence of postoperative chronic subdural hematoma. *Front Neurol* 2024;15:1305543.
20 Isensee F, Jaeger PF, Kohl SAA, *et al*. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–11.
21 Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005;27:1226–38.
22 Suman G, Patra A, Korfiatis P, *et al*. Quality gaps in public pancreas imaging datasets: Implications & challenges for AI applications. *Pancreatology* 2021;21:1001–8.
23 Zheng Y, Wagner PD, Singal AG, *et al*. Designing Rigorous and Efficient Clinical Utility Studies for Early Detection Biomarkers. *Cancer Epidemiol Biomarkers Prev* 2024;33:1150–7.
24 Overbeek KA, Goggins MG, Dbouk M, *et al*. Timeline of Development of Pancreatic Cancer and Implications for Successful Early Detection in High-Risk Individuals. *Gastroenterology* 2022;162:772–85.
25 Chhoda A, Vodusek Z, Wattamwar K, *et al*. Late-Stage Pancreatic Cancer Detected During High-Risk Individual Surveillance: A Systematic Review and Meta-Analysis. *Gastroenterology* 2022;162:786–98.
26 Yu J, Blackford AL, Dal Molin M, *et al*. Time to progression of pancreatic ductal adenocarcinoma from low-to-high tumour stages. *Gut* 2015;64:1783–9.
27 Conlon KC, Klimstra DS, Brennan MF. Long-term survival after curative resection for pancreatic ductal adenocarcinoma. Clinicopathologic analysis of 5-year survivors. *Ann Surg* 1996;223:273–9.
28 Singhi AD, Koay EJ, Chari ST, *et al*. Early Detection of Pancreatic Cancer: Opportunities and Challenges. *Gastroenterology* 2019;156:2024–40.
29 PYMNTS. Mayo Clinic Uses AI to Detect Pancreatic Cancer Earlier. 2025. Available: https://www.pymnts.com/artificial-intelligence-2/2025/mayo-clinic-uses-ai-to-detect-pancreatic-cancer-earlier/ [Accessed 17 Dec 2025].
30 de Koning HJ, van der Aalst CM, de Jong PA, *et al*. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med* 2020;382:503–13.
31 Welch HG, Prorok PC, O'Malley AJ, *et al*. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *N Engl J Med* 2016;375:1438–47.
32 Walter JE, Heuvelmans MA, Ten Haaf K, *et al*. Persisting new nodules in incidence rounds of the NELSON CT lung cancer screening study. *Thorax* 2019;74:247–53.

33 Akwo JD, Trieu P, Lewis S. Does the availability of prior mammograms improve radiologists' observer performance? A scoping review. *BJR Open* 2023;5:20230038.
34 Torphy RJ, Wang Z, True-Yasaki A, *et al*. Stromal Content Is Correlated With Tissue Site, Contrast Retention, and Survival in Pancreatic Adenocarcinoma. *JCO Precis Oncol* 2018;2018:PO.17.00121.
35 Liudahl SM, Betts CB, Sivagnanam S, *et al*. Leukocyte Heterogeneity in Pancreatic Ductal Adenocarcinoma: Phenotypic and Spatial Features Associated with Clinical Outcome. *Cancer Discov* 2021;11:2014–31.
36 Chari ST, Wu B, Lopez C, *et al*. Risk of Pancreatic Cancer in Glycemically Defined New-Onset Diabetes: A Prospective Cohort Study. *Gastroenterology* 2026;170:106–17.
37 NIH RePORT. Anticancer ELNP nanovaccines for curative treatment of pancreatic cancer. 2025. Available: https://reporter.nih.gov/search/Oz5oAFm3kUqjvhzx1Kz7gQ/project-details/11040015#details [Accessed 26 Sep 2025].
38 Daher D, Seif El Dahan K, Rich NE, *et al*. Hepatocellular Carcinoma Screening in a Contemporary Cohort of At-Risk Patients. *JAMA Netw Open* 2024;7:e248755.
39 Rich NE, Singal AG. Overdiagnosis of hepatocellular carcinoma: Prevented by guidelines? *Hepatology* 2022;75:740–53.