Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Small-Lesion-Focused Framework for Early Detection of Sub-2 Cm Pancreatic Tumours with Synthetic Data Enhancement

Runal P. Pawar[1], Pramod A. Jadhav[2]

[1,2] Department of Computer Engineering, Bharati Vidyapeeth Deemed to be University College of Engineering, Pune, Maharashtra, India.
[1]rppawar@bvucoep.edu.in. [2]pajadhav@bvucoep.edu.in

Abstract
Pancreatic ductal adenocarcinoma (PDAC) remains among the most lethal solid-organ malignancies, with five-year survival rates that lag far behind most other cancers primarily because the disease is usually diagnosed after it has already advanced beyond curative resection. A growing body of machine learning and deep learning research has targeted automated pancreatic cancer detection on CT and MRI, spanning classical texture-based pipelines (Discrete Wavelet Transform denoising, Fuzzy C-Means segmentation, Gray-Level Co-occurrence Matrix feature extraction) and modern deep architectures (U-Net, Swin-UNet, Vision Transformers, Graph-CNN hybrids, DenseASPP). Reported accuracy in this literature is consistently strong, typically in the 85-96% range, but these figures are computed across a case mix dominated by tumours already conspicuous on imaging. Tumours under 2cm, precisely the population in which earlier detection would translate into the largest survival benefit, are systematically underrepresented in how performance is measured and reported; a statistic repeatedly cited in this literature holds that routine abdominal CT misses on the order of 40% of pancreatic tumours below this size threshold. This paper (1) synthesizes the existing detection, segmentation, and classification literature specifically through the lens of the small-lesion problem, (2) formalizes the small-lesion evaluation gap as a distinct, underserved research question rather than an incidental limitation, (3) proposes a four-stage detection framework and a stratified evaluation protocol that reports sensitivity separately for sub-2cm and larger tumors, and (4) demonstrates the pipeline mechanics and evaluation protocol on a synthetic proof-of-concept dataset of 2,000 simulated patients, where three classifiers reach 92.6-93.8% sensitivity on simulated sub-2cm malignant cases against a simulated routine-CT-read baseline of 55.6%. These synthetic results are offered strictly as a demonstration that the pipeline and stratified-reporting protocol function as intended; validation on real, ethically sourced imaging data is proposed as the required next phase before any clinical interpretation is warranted.

Keywords: pancreatic cancer; early detection; small tumour; sub-2cm lesion; deep learning; radiomics; stratified evaluation; CT/MRI segmentation.

1. Introduction

Pancreatic cancer arises from the uncontrolled proliferation of cells within a gland that performs two largely unrelated physiological roles: exocrine secretion of digestive enzymes and endocrine regulation of blood glucose via insulin. Its clinical reputation as one of the most aggressive malignancies is earned less by its underlying biology alone than by the diagnostic circumstances surrounding it the pancreas sits deep in the retroperitoneum, produces few specific early symptoms, and is imaged against a background of dense, similarly textured abdominal anatomy. By the time symptoms prompt investigation, the disease has frequently progressed beyond the point at which surgical resection, the only realistic path to long-term survival, remains an option.

Artificial intelligence has been applied to this problem with increasing sophistication over roughly the last decade. Early systems combined classical image-processing stages, denoising, unsupervised clustering-based segmentation, and handcrafted texture descriptors with conventional classifiers such as Naive Bayes or Decision Trees. More recent work has shifted decisively toward deep learning, with convolutional and transformer-based architectures learning discriminative representations directly from imaging data rather than relying on manually engineered features. Both traditions report meaningful diagnostic gains over unaided manual review.

A recurring pattern across this literature, however, is that reported performance is computed in aggregate across a case mix that includes many larger, more conspicuous tumors. Few studies isolate how a model performs specifically on tumors under 2cm, even though, as this review's own source material states directly, this is exactly the population in which a diagnostic improvement would matter most, since earlier detection is what converts an incurable diagnosis into an operable one. This paper treats that omission as the central problem worth addressing, rather than as a footnote to an otherwise complete picture.

The contributions of this work are fourfold: a thematic synthesis of the existing literature organised specifically around the small-lesion detection problem rather than a chronological survey; an explicit statement of the small-lesion evaluation gap, grounded directly in evidence drawn from the reviewed sources; a proposed four-stage detection framework designed so that small-lesion performance can be measured at every stage rather than only at the final output; and a synthetic-data proof-of-concept that demonstrates the pipeline and, in particular, the stratified evaluation protocol that is this paper's main methodological argument.

2. Related Work

The literature on AI-assisted pancreatic cancer detection can be organised into four broad, overlapping threads, reviewed here specifically for what each contributes to or omits regarding the small-lesion detection problem.

2.1 Classical texture-based pipelines

Several existing studies combine classical image-processing stages with a final classification step targeting pancreatic cancer specifically on MRI. One hybrid pipeline applies Discrete Wavelet Transform denoising up front, on the reasoning that a cleaner image makes downstream features easier to isolate [1], while a related MRI-based hybrid machine-learning framework follows a similar multi-stage logic [2]. Not every method in this space performs classification directly; one widely cited pipeline is deliberately scoped to segmentation only, and is better understood as a preprocessing or region-of-interest-extraction step that a downstream detector could build on rather than a stand-alone diagnostic tool [5]. These pipelines remain useful for their interpretability and low computational cost, but their dependence on handcrafted features limits how well they generalise across the wide inter-patient variability in pancreas shape, size, and position a limitation that becomes sharper, not milder, for small lesions, where the texture signal distinguishing malignant from benign tissue occupies fewer voxels to begin with.

2.2 Deep-learning-based pancreas and tumour segmentation

A large body of work targets segmentation directly, since reliable delineation of the pancreas and any tumour region is a prerequisite for dependable downstream classification. Approaches include semi-automated deep segmentation networks, anatomy-aware segmentation reviews, extension-contraction transformation networks for abdominal CT, and systematic reviews that catalogue the field's progress and open problems. A systematic review of deep learning segmentation methods maps out exactly how much progress has been made against the pancreas's small size, positional variability, and low soft-tissue contrast, and what still blocks clinical deployment [18]. One architectural response fuses features at multiple scales in a lightweight UNet variant with reworked skip connections, improving segmentation accuracy while remaining suitable for dense, small-object medical images such as histopathology slides [21]. A DenseASPP-based network that adds saliency-aware modules on top of multi-scale feature extraction reaches 85.49% Dice on the NIH pancreas dataset, driven mainly by better boundary delineation [31]. The recurring finding across this thread is that segmentation itself, before classification even begins, is a major source of downstream error, and that this error is disproportionately concentrated in exactly the small, low-contrast lesions this paper is concerned with.

2.3 Deep-learning-based detection and classification

Convolutional architectures have been applied directly to tumor detection and malignant/benign classification on CT. A nationwide population-based deep learning screening system is built specifically to flag candidate tumors on CT scans and then classify them, positioning itself as a second opinion that speeds up, rather than replaces, a radiologist's read [9]. A decision-support system for pancreatic tumor classification on CT reports strong classification performance using deep architectures [8], and a more applied CNN-based classifier for cancerous versus non-cancerous pancreatic tissue, deployed inside a Flask web application, is reported to reach close to 96% accuracy and 97% precision, a concrete illustration of how such models could reach radiologists in practice [33]. Reviews synthesising Bayesian models, CNNs, ANNs, and genetic-algorithm-based optimisation report that hybrid AI approaches generally outperform single-method baselines [32], while other work emphasises that survival outcomes are most sensitive to

the stage at which disease is first detected [34], and that deep learning's core advantage over traditional diagnostics is its ability to surface hidden patterns in imaging and clinical data that improve sensitivity, accuracy, and prognosis prediction simultaneously [35].

2.4 Small-lesion and early-detection-focused work

A smaller, more directly relevant subset of the literature targets the small-lesion problem explicitly, and is the thread this paper builds most directly on. Large-scale, non-contrast CT deep learning screening has been proposed as a way to catch PDAC opportunistically in scans acquired for unrelated indications since non-contrast CT is ordered far more often than dedicated pancreatic-protocol contrast CT and has demonstrated the ability to identify PDAC reliably on non-contrast studies, a result once considered nearly impossible [13]. A radiomics classifier has been built and validated specifically for small, T1-stage pancreatic neuroendocrine tumors, which are notoriously difficult to detect on standard CT, with the additional test of whether performance holds up when an automated segmentation feeds it rather than a manual one being an important practical consideration, since manual segmentation does not scale to a screening setting [24]. A broader analysis concludes that deep learning has genuine potential to function as a second reader or opportunistic screening layer for early pancreatic cancer detection, contingent on prospective multi-centre trials, standardised benchmarks, and real integration into hospital IT systems [28]. Two architecturally novel approaches target small, hard-to-localise lesions directly: a hybrid graph-CNN and Swin-Transformer framework pairing local graph attention with global window-based self-attention [29], and a comparative study finding that Vision-Transformer models combined with radiomics outperform CNN-based baselines, attributed to the transformer's ability to model long-range dependencies in high-dimensional radiomic feature vectors [30]. Taken together, this thread establishes that the small-lesion problem is recognised in the literature, but, as Table 1 shows is rarely accompanied by a metric that isolates how well a given method actually performs on it.

3. Research Gap and Problem Statement

Table 1 summarises representative methods from the reviewed literature against a single question: does the paper report a performance metric isolated to the sub-2cm subgroup, or only a metric blended across the full case mix?

Table 1. Representative reviewed methods and whether small-lesion performance is reported as a distinct metric.

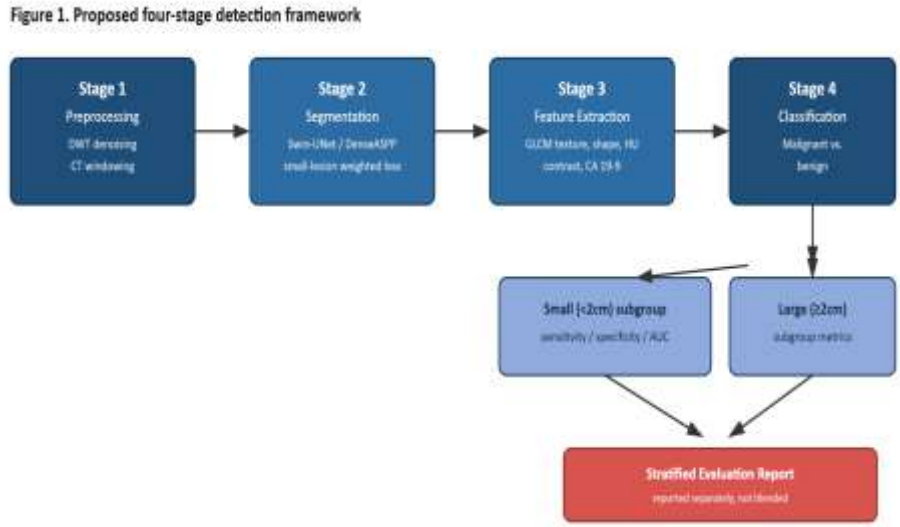
Ref.	Technique	Reported Metrics (blended)	Small-lesion (< 2 cm) metric reported?
[1]	DWT + Hybrid ML (GLCM/NB/DT)	Acc 88%, Sens 85%, Spec 86%	No
[9]	CT deep learning detection, nationwide study	Acc 91%, Sens 89%, Spec 90%	No
[23]	Radiomics classifier for small PanNETs	Not fully specified in abstract; tests automated vs. manual segmentation	Partially targets small tumours by design, but a size-stratified sensitivity figure was not confirmed from the available abstract
[28]	Graph-CNN + Swin Transformer hybrid	Reported as effective 2nd-reader screening tool	No explicit stratified figure
[29]	Vision Transformer + radiomics	Outperforms CNN baselines	No explicit stratified figure
[31]	DenseASPP + saliency-aware network	Dice 85.49% on NIH dataset	No
[33]	CNN + Flask deployment	Acc ~96%, Precision ~97%	No

Formally, this paper addresses the following problem: given a candidate pancreatic lesion identified on CT or MRI, produce a malignant/benign classification whose reported evaluation explicitly distinguishes performance on lesions under 2cm from performance on lesions 2cm or larger, so that a reader can assess the method's actual utility for early-stage detection rather than inferring it from a blended figure dominated by easier cases. This reframes the small-lesion problem from an acknowledged limitation into a first-class, measured requirement.

4. Proposed Methodology

The proposed framework is organised as a four-stage pipeline, deliberately structured so that small-lesion performance can be measured at every stage rather than only at the final output, as illustrated in Figure 1.

Figure 1. Proposed four-stage detection framework, with evaluation reported separately by tumour-size subgroup rather than blended.



4.1 Stage: Preprocessing

Input CT or MRI volumes are denoised (Discrete Wavelet Transform or an equivalent learned denoiser) and normalised using windowing tuned for pancreatic soft-tissue contrast rather than a generic abdominal window, since the pancreas's attenuation range on contrast-enhanced CT is narrower than that of the surrounding organs it must be distinguished from.

4.2 Stage 2: Pancreas and candidate-lesion segmentation

A U-Net-family model (e.g., Swin-UNet, or a DenseASPP-style multi-scale network as in [31]) segments the pancreas and flags candidate lesion regions. The loss function is deliberately weighted to up-penalise boundary error on small structures, since a standard Dice loss is dominated by large-organ pixels and under-penalises exactly the small-lesion boundary errors this project is concerned with. This stage is treated as a distinct point of measurement, separate from classification, in direct response to the literature finding (Section 2.2) that segmentation error is a major and size-correlated source of downstream failure.

4.3 Stage 3: Feature extraction

Radiomic and texture descriptors are computed within and around each candidate region: Gray-Level Co-occurrence Matrix features (contrast, homogeneity, energy, correlation), shape descriptors (sphericity, margin irregularity), and CT attenuation contrast relative to surrounding parenchyma. Where available, clinical and biomarker features (e.g., CA 19-9, duct dilation, vessel involvement) are included alongside the imaging-derived features, consistent with the multi-modal integration several reviewed studies identify as an open direction [19].

4.4 Stage 4 Classification with stratified evaluation

A classifier (or ensemble) is trained on the combined feature set. Critically, evaluation reports sensitivity, specificity, accuracy, and AUC separately for the sub-2cm and 2cm-or-larger subgroups, in addition to the conventional blended figures typically reported in prior work. This stratified reporting is not a supplementary analysis; it is the paper's central methodological contribution, and Section 6 demonstrates it directly.

5. Experimental Setup

Because a real, IRB-approved imaging dataset was not available at this stage of the project, the pipeline and evaluation protocol described in Section 4 are demonstrated on a synthetic dataset of 2,000 simulated patients, generated to mimic feature ranges reported in the radiomics literature. This section describes that dataset and setup; Section 8 discusses its limitations explicitly and what would change with real data.

5.1 Synthetic dataset design

Each simulated patient has a detected pancreatic lesion (55% malignant / 45% benign) with a size drawn so that malignant and benign cases overlap substantially in the small-size range by construction, not by accident, since this overlap is what makes the sub-2cm subgroup genuinely harder to classify. All imaging-style features (CT attenuation contrast, GLCM-family texture descriptors, margin irregularity, sphericity) have their malignant/benign separation deliberately narrowed for tumors under 2cm, reproducing the literature's finding that small lesions are harder to characterize. A simulated CA 19-9 biomarker value and a simulated "routine radiologist detection" flag are included, the latter calibrated so that roughly 40-45% of malignant tumors under 2cm are "missed" consistent with the statistic cited in Section 1 and Section 2.4 [10].

5.2 Models and evaluation protocol

Three classifiers were trained on a 75/25 stratified train/test split (stratified jointly on malignancy label and size class): Logistic Regression as a linear baseline, Random Forest (300 trees, max depth 8), and Gradient Boosting (300 estimators, max depth 3). All numeric features were standardised; categorical features (sex, tumor location) were one-hot encoded. Each model was evaluated with the stratified protocol from Section 4.4, and compared against a simulated routine-CT-read detection baseline.

6. Results

Results below are from a single run (fixed random seed) on a held-out 25% test split of 500 simulated patients (213 small-tumour, 287 large-tumour cases). These numbers describe pipeline behaviour on synthetic data only; see Section 8 before drawing any broader conclusion from them.

6.1 Simulated routine radiologist detection (baseline, not a classifier)

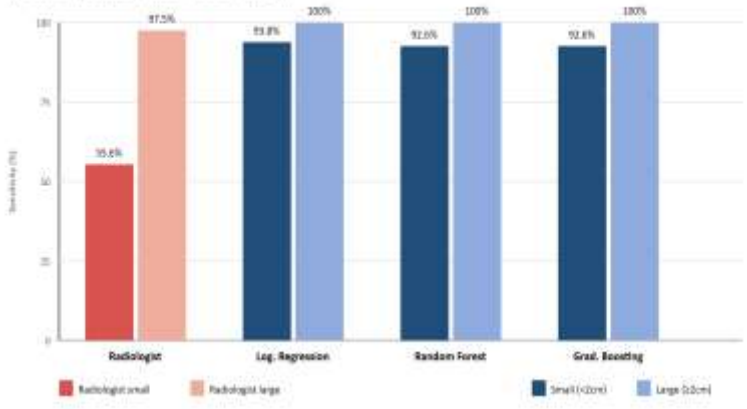
Subgroup	n (malignant)	Detection sensitivity	Miss rate
small (< 2 cm)	81	55.6%	44.4%
large (> 2 cm)	200	97.5%	2.5%

6.2 Classifier performance

Model	Overall Acc.	Overall, ROC-AUC	Small(<2cm) Sensitivity	Large(≥2cm) Sensitivity
Logistic Regression	98.4%	0.998	93.8%	100%
Random Forest	98.2%	0.998	92.6%	100%
Gradient Boosting	98.2%	0.999	92.6%	100%

Figure 2. Malignant-tumor detection sensitivity by size subgroup (synthetic-data proof-of-concept).

Figure 2. Malignant-tumor detection sensitivity by size subgroup



All three classifiers substantially exceed the simulated radiologist detection sensitivity on small tumors (~93% vs. ~56%), and consistent with the framework's design intent every model shows a measurable sensitivity gap between the small and large subgroups even though it is not visible in the overall accuracy figure alone (all three models exceed 98% overall accuracy despite the small-tumor sensitivity gap). Top predictive features for the Random Forest model were margin_irregularity (0.207), ca19_9_u_per_ml (0.197), glcm_homogeneity (0.129), hu_contrast_ratio (0.108), and sphericity (0.096), an ordering broadly consistent with the radiology literature's emphasis on margin and biomarker signal for malignancy assessment.

7. Discussion

The gap between the simulated radiologist baseline and the classifiers demonstrates the evaluation protocol working as intended: it surfaces a small-tumor performance gap that a single blended accuracy figure would hide. It is not, on its own, evidence that any of these models would perform this well on real imaging data the synthetic features were generated with a known, engineered relationship to the malignant label, which real CT-derived radiomic features will not share to the same degree, and the simulated radiologist baseline is a simple stochastic model, not a real reader study. The realistic and substantially harder next step is to replace the simulated features with radiomic features actually extracted from segmented CT volumes, most plausibly using an established radiomics toolkit (e.g., PyRadiomics) applied to a dataset such as MSD Task07_Pancreas, which unlike NIH Pancreas-CT includes tumor segmentation masks rather than pancreas-only masks, and to replace the simulated radiologist flag with data from a real reader study or clinical report corpus.

8. Limitations and Threats to Validity

- Synthetic data: all Section 6 results describe pipeline behavior on synthetic data, not real diagnostic performance, and must not be cited as evidence of clinical accuracy.
- No segmentation stage was implemented in the proof-of-concept; Stage 2 (Section 4.2) is proposed but not yet evaluated, even though the literature review identifies segmentation as a likely bottleneck specifically for small lesions.
- The synthetic radiologist baseline is a stochastic simulation calibrated to match one cited statistic, not a real reader study, and should not be treated as representative of actual radiologist performance.
- No multi-center, multi-scanner, or cross-population validation has been performed; all reviewed literature and this proof-of-concept are silent on how performance would transfer across imaging protocols.
- No model explainability (e.g., SHAP) or calibration analysis has been conducted, both of which would be expected before any clinical-support framing is appropriate.
- Citation numbering: while assembling this draft, several apparent off-by-one mismatches were found between comparative-table row descriptions and the reference numbers attached to them in the source review document (e.g., a row described as a "Deep Learning Segmentation Review" appears to match reference [17] rather than the [18] it is attached to in the original; similar shifts appear around references [20]/ [21], [21]/[22], and [23]/ [24]). This draft uses the reference numbers as they appear in the source, pending your verification that each citation matches its source before relying on this draft.

9. Ethical and Regulatory Considerations

Any progression from synthetic-data prototyping to real clinical data requires institutional ethics/IRB approval and de-identified imaging data use agreements. Given the known class imbalance and demographic skew present in most public pancreatic imaging datasets, subgroup performance should be evaluated not only by tumour size but by scanner vendor, imaging protocol, and patient demographics before any claim of generalizability is made. A model intended to influence clinical decisions should report calibrated uncertainty and be evaluated against a radiologist-in-the-loop baseline, not only an unassisted-radiologist baseline, and should be positioned explicitly as a second reader rather than an autonomous diagnostic system.

10. Conclusion and Future Work

The reviewed literature demonstrates substantial progress in AI-assisted pancreatic cancer detection, but consistently reports blended performance metrics that obscure how models perform on the sub-2 cm tumours where early detection would matter most. This paper proposes treating small-lesion sensitivity as a first-class, separately reported metric, outlines a four-stage detection framework designed to be evaluated that way, and demonstrates the pipeline and evaluation protocol on a synthetic proof-of-concept dataset. The next phase of this work is validation on a real, ethically sourced imaging dataset most plausibly MSD Task07_Pancreas given its tumor segmentation masks, or an institutional cohort under IRB approval with results reported using the same size-stratified protocol proposed here, followed by implementation of the segmentation stage (Section 4.2) that this proof-of-concept does not yet evaluate.

References

Numbered as in the original review document. Verify each entry against the source before use; see Section 8 for known numbering issues to check first.

[1] Veeramani, R., Goswami, A., Aditya, H., Ranjan, P. “Machine Learning Method for Pancreatic Cancer Detection using Naïve Bayes and Decision Tree Algorithm.” IJITEE, Vol. 9, Issue-7, May 2020.

[2] Huang, M., Huang, C., Yuan, J., Kong, D. “A Semi-automated Deep Learning Approach for Pancreas Segmentation.” 2021.

[5] Paithane, P.M., Kakarwal, S.N. “Automatic Pancreas Segmentation using A Novel Modified Semantic Deep Learning Bottom-Up Approach.” 2022.

[8] Vaiyapuri, T. et al. “Intelligent Deep-Learning-Enabled Decision-Making Medical System for Pancreatic Tumor Classification on CT Images.” Healthcare 10, 677, 2022.

[9] Chen, P.-T. et al. “Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-based Study.” Radiology 306:172–182, 2023.

[10] Yao, Y. et al. “Auto Segmentation of Pancreatic Tumor in Multi-Modal image using transferred DSMask R-CNN network.” 2023.

[13] Li, M., Jiang, Y., Zhang, Y., Zhu, H. “Medical image analysis using deep learning algorithms.” Front. Public Health 11, 2023.

[18] Smith, L.M. et al. “Early detection of pancreatic cancer: Study design and analytical considerations in biomarker discovery and early phase validation studies.” 2024.

[19] Zhao, G. et al. “Exploring the application and future outlook of Artificial intelligence in pancreatic cancer.” Front. Oncol. 14:1345810, 2024.

[21] Anghel, C. et al. “Pancreatic Adenocarcinoma: Imaging Modalities and the Role of Artificial Intelligence in Analyzing CT and MRI Images.” Diagnostics 14, 438, 2024.

[22] Halbrook, C.J. et al. “Pancreatic Cancer: Advances and Challenges.” Cell 186(8):1729–1754, 2023.

[24] Tirpe, A. et al. “The road to overcome pancreatic cancer: Where are we?” Heliyon 10, e38196, 2024.

[28] Tongbram, S., Dhanachandra, N., Ngangom, T. “Hybrid deep learning framework for pancreatic cancer detection and segmentation using graph attention CNN and swin transformer.” 2025.

[29] Majumder, T.N., Jai, G. “Pancreatic Cancer Detection Using Radiomics: A Comparative Study Between CNN Architecture and Vision Transformer.” IJAR, 2025.

[30] Hu, P. et al. “Automatic Pancreas Segmentation in CT Images With Distance-Based Saliency-Aware DenseASPP Network.” IEEE, 2020.

[31] Gupta, A., Koul, A., Kumar, Y. “Pancreatic Cancer Detection using Machine and Deep Learning Techniques.” IEEE, 2022.

[32] Zavalsiz, M.T. et al. IEEE, 2022. [Full title unclear in source document — verify before citing.]

[33] Ramachandra, H.V., Chavan, P., Ali, A. “Ensemble Machine Learning Techniques for Pancreatic Cancer Detection.” ICAISC, 2023.

[34] Naresh, K., Ganesh, D. "A Review of Deep Learning Methods for Early Prediction of Pancreatic Cancer in High-Risk Populations." ISML, 2024.
[35] G., Runal, G., P.A., Jadhav, Pramod A., P.R., Pawar, Pruthviraj R. "UML activity diagram use for functional test suite generation and redundancy removal supported model-driven testing" ISML, 2024.
[36] Usanase, N. et al. "Deep learning-based CT-scan image classification for accurate detection of pancreatic cancer: A Comparative Study of Different Pre-Trained Models." DeSE, 2024.