
Advancing Interpretable AI for Clinical Decision Support and Image-Driven Cancer Detection: A Review of Reliability, Transparency, and Diagnostic Performance

Sipho Wójcik¹, Felix Moreau²

¹ Technical University of Munich, GERMANY

² University of Lagos, NIGERIA

Keywords

Explainable AI
Clinical Decision
Support
Medical Imaging
Cancer
Classification
Grad-CAM
Interpretability
Pneumonia
Detection
Healthcare AI

ABSTRACT

Explainable AI has been studied separately for cancer classification and for other image-based clinical decision support tasks such as pneumonia detection, but a unified account of what these two literatures share, and where they diverge, is largely missing. This review synthesizes seven studies to build that account. Two systematic reviews anchor the synthesis: one covering 46 explainable-AI cancer-imaging studies finds that Grad-CAM dominates as an explanation method (used in 34 of 46 studies) while only 10.9% of studies evaluate their explanations quantitatively rather than through subjective visual assessment; a second, broader review of 133 studies spanning X-ray, CT, MRI, PET, ultrasound, and pathology across pneumonia, cancer, and other conditions identifies standardization and clinical-workflow integration as the field's most persistent unresolved barriers, regardless of which disease is being diagnosed. Applied studies on both sides of this review's title support that convergence: a hybrid CNN explained via Grad-CAM++ reaches 97.14% accuracy classifying breast-cancer ultrasound images, a Random Forest classifier reaches 96% accuracy on a cytological cancer benchmark, and, on the clinical-decision-support side, a ResNet50 model compared across four interpretability methods reaches 91% accuracy detecting pneumonia from chest X-rays with the best-performing explanation method (layer-wise relevance propagation) adding no accuracy cost. Two further studies from the same research cluster propose, respectively, a multimodal cancer-diagnosis framework and a MobileNet-based pneumonia detector; both are discussed with an explicit, disclosed limitation, since their full quantitative claims could not be independently verified beyond their confirmed titles and a general-scope publication venue this review examined directly. Read together, these seven studies suggest that a genuinely unified explainable-AI framework for clinical decision support is less a matter of building one model that does everything than of applying the same standard of quantitative, clinically validated explanation across every disease and modality a healthcare system needs to diagnose — a standard this literature has not yet consistently met even within cancer imaging alone.

Introduction

A machine-learning model that classifies a medical image without offering a clinician any way to inspect its reasoning is, in a specific and consequential sense, not clinically usable regardless of its measured accuracy: a radiologist or pathologist who cannot see why a system flagged (or missed) a finding cannot integrate that system into their own clinical judgment, cannot explain the finding to a patient, and cannot catch the specific failure mode a model relying on an imaging artifact, a scanner-specific signature, or a spurious correlation with patient positioning — that a benchmark accuracy figure alone will not reveal [1-3]. This is true whether the diagnostic task is cancer classification or, as this review's title makes explicit, any other image-based clinical decision support task; the underlying requirement for legible, clinically validated explanation does not change with the disease being

diagnosed, even though most of the explainable-AI literature has developed within disease-specific silos.

Scope: two literatures this review treats as one problem

This review treats explainable AI for cancer classification and explainable AI for other image-based clinical decision support (illustrated here through pneumonia detection from chest radiographs) as two applications of the same underlying problem rather than as separate fields, because, as Section 4's broader systematic review shows directly, the same explanation methods, the same evaluation gaps, and the same clinical-adoption barriers recur across both. Sections 2 and 3 review the cancer-classification literature specifically; Section 4 reviews the broader medical-imaging explainability literature, including pneumonia detection; and Section 5 examines two studies whose own stated aims explicitly span this review's full scope a multimodal cancer-diagnosis framework, and a pneumonia detector from the same author cluster with an explicit account of what could and could not be independently verified about each [4-7].

Review scope and methodology

This review synthesizes seven studies, selected on relevance and evidentiary strength rather than a fixed recency cutoff, spanning 2023 to 2026. Two are systematic reviews (46 and 133 primary studies respectively) that anchor this review's cross-cutting claims; three are individually verified applied studies, two on cancer classification and one on pneumonia detection; and two, from an overlapping author cluster, could not be independently verified beyond their confirmed title and authorship and are discussed with that limitation stated explicitly rather than treated as equivalent to the five independently verified sources.

Real-world stakes: a documented failure and a documented success in clinical decision support

Two well-documented, independently reported episodes outside this review's seven core studies show concretely why the explanation and evaluation standards discussed throughout this review are not an academic nicety. Wong *et al.* (2021), publishing in *JAMA Internal Medicine*, externally validated the Epic Sepsis Model — a proprietary, widely deployed sepsis-prediction algorithm embedded directly in a major electronic health record system — against 38,455 hospitalizations at Michigan Medicine. At the vendor-recommended alert threshold, the model achieved a sensitivity of only 33% (missing 67% of sepsis cases clinicians ultimately diagnosed) and a hospitalization-level AUC of just 0.63, while generating alerts on 18% of all hospitalized patients and a positive predictive value of only 12% — a pattern the authors describe as producing substantial alert fatigue for minimal clinical benefit. Critically, the model's proprietary, undisclosed internal logic meant that clinicians using it, and researchers attempting to evaluate it, could not inspect why any individual alert had fired, which is precisely the kind of opacity Sections 2.1 and 3.1's systematic reviews find pervasive even in published, ostensibly explainable research. Abràmoff *et al.* (2018), by contrast, describe the pivotal trial that led the FDA to authorize IDx-DR as the first autonomous AI diagnostic system in any field of medicine — a system that classifies retinal images for diabetic retinopathy without requiring a clinician to interpret

the image itself. Tested across 900 enrolled patients at ten U.S [8-13]. primary care sites, the system achieved 87.2% sensitivity and 90.7% specificity, both exceeding pre-specified endpoints the FDA had required as a precondition for authorization. The contrast is instructive for this review's argument: IDx-DR's authorization rested on a transparent, pre-registered, quantitative performance standard applied before deployment, while the Epic Sepsis Model's real-world shortfall was discovered only after widespread deployment, through an external validation its own vendor had not conducted or published. That is close to the exact same gap Section 2.1 documents at the level of academic publishing — a field defaulting to qualitative, post-hoc, self-reported assessment rather than the pre-specified, quantitative, externally validated standard IDx-DR's authorization required and the Epic Sepsis Model's deployment lacked.

Explainable AI for Medical Image-Based Cancer Classification

What the field looks like at scale: a systematic view

Ghauth and Kustiawan (2026) provide the broadest evidentiary base in this review specifically for cancer imaging: a systematic review of 46 studies, drawn from 926 initial records, spanning lung cancer (11 studies), breast cancer (9), skin/melanoma (7), cervical cancer (6), colorectal/polyp (2), and individual studies on ovarian, bone, and pancreatic cancers. Their central finding is a striking methodological homogeneity: Grad-CAM and its variants appear in 34 of the 46 studies (73.9%), overwhelmingly because of how easily it integrates with the convolutional architectures that themselves account for 73.9% of implementations, while Vision Transformers, used in 17.4% of studies, show compatibility challenges with traditional gradient-based explanation methods. Perhaps their most consequential finding for this review's broader argument is an evaluation gap: 58.7% of the 46 studies assess their explanations only through qualitative visual inspection, while just 10.9% apply a purely quantitative evaluation — meaning the majority of this literature's claims about explanation quality rest on subjective impression rather than a measurable standard, a limitation their review explicitly attributes to a lack of standardized interpretability metrics rather than to any single study's individual shortcoming [14-16].

An applied demonstration: fusing architectures and explaining the fusion

Zou and Miao (2025) offer a concrete, high-performing example of the CNN-plus-Grad-CAM pattern Ghauth and Kustiawan (2026) find dominant. Working with the Breast Ultrasound Image (BUSI) dataset — 8,116 images split evenly between benign and malignant cases — they fuse three CNN architectures (VGG16, DenseNet121, and Xception) at an intermediate feature layer rather than relying on any single network, and apply Grad-CAM++, a refinement of the base Grad-CAM method, for interpretability. The fused model reaches 97.14% accuracy with F1 = 97.18%, roughly 13 percentage points ahead of VGG16 alone (84.43%) — a substantial, well-documented benefit of architectural fusion, though the paper's interpretability contribution, per Section 2.1's finding, is still evaluated primarily through the qualitative visual inspection that constitutes the field's dominant but least rigorous practice.

A comparative baseline on tabular, image-derived features

Siddiqui *et al.* (2023) work with a different representation of cancer-imaging data: rather than classifying raw ultrasound or radiographic images directly, they classify 30 morphometric features (cell-nuclei size, shape, and texture measurements) already extracted from fine-needle aspirate cytology images in the well-established Wisconsin Diagnostic Breast Cancer dataset (569 samples). Random Forest reaches 96% accuracy with a near-perfect AUC-ROC of 1.00, comfortably ahead of Gaussian Naive Bayes (93%). This tabular, feature-based approach sidesteps some of Section 2.1's Vision-Transformer-compatibility concerns entirely, since there is no raw image for a CNN-style explanation method to operate on, but it also sidesteps the raw-image interpretability problem this review's title is centrally concerned with — a genuine methodological trade-off worth naming rather than treating both approaches as interchangeable simply because both classify "cancer."

Benchmark datasets: comparable scale, different data types

The two applied cancer-classification studies in this review, despite both reporting accuracy above 96%, are not evaluated on comparable data. Zou and Miao's (2025) BUSI ultrasound benchmark (8,116 images) and Siddiqui *et al.* (2023) Wisconsin cytological set (569 samples of 30 pre-extracted numerical features) differ by more than an order of magnitude in sample count and, more importantly, in what is actually being classified: raw pixel data in one case, already-summarized morphometric measurements in the other. A reader taking Section 2's two headline accuracy figures as directly comparable evidence that "cancer classification" has reached a mature, uniform performance level would be over-reading what these two benchmarks, taken individually, can support; the two studies are complementary demonstrations of different points along the raw-image-to-structured-feature spectrum rather than two attempts at the same underlying task[17].

Beyond Cancer: Explainable AI for Broader Image-Based Clinical Decision Support

A systematic view spanning multiple diseases and modalities

Ahmed, Naz, Khan, Rehman, Ismael, and Khan (2026) extend this review's evidentiary base beyond cancer specifically, systematically reviewing 133 studies (from 980 initial records) across X-ray, CT, MRI, PET, ultrasound, and digital pathology, and across clinical applications including pneumonia detection, lung cancer screening, brain tumor segmentation, fracture diagnosis, multiple sclerosis detection, Alzheimer's disease identification, and breast cancer histopathology. Their central conclusion, reached independently of Ghauth and Kustiawan's (2026) cancer-specific review but substantively aligned with it, is that standardization, data bias, complicated data integration, and — most relevant to this review's argument — a lack of standardized interpretability metrics and limited clinician-usability studies remain the field's persistent barriers to clinical adoption, regardless of which disease or modality a given study addresses. This convergence across two independently conducted systematic reviews, one scoped to cancer and one scoped much more broadly, is the strongest evidence available in this review that explainable AI for cancer classification and for other image-based clinical decision support are, methodologically, the same unsolved problem wearing different clinical labels.

An applied demonstration on pneumonia detection, with explanation methods compared directly

Colin and Surantha (2025) provide, for the pneumonia side of this review's scope, the same kind of rigorous applied demonstration Zou and Miao (2025) provide for cancer, and go further by directly comparing multiple explanation methods against each other rather than applying just one. Working with the well-known Kaggle chest X-ray pneumonia dataset (5,863 images: 1,587 normal and 4,276 pneumonia, sourced from pediatric patients at Guangzhou Women and Children's Medical Center), they compare a pre-trained ResNet50 baseline against four interpretability-oriented variants: layer-wise relevance propagation (LRP), adversarial training, Grad-CAM, and a spatial attention mechanism, scoring each on accuracy, sensitivity, specificity, AUC, and a dedicated interpretability metric. LRP wins on every metric simultaneously — 91% accuracy, 0.96 AUC, and the highest interpretability score (0.85) — directly demonstrating, in a way none of this review's cancer-focused studies attempt, that better explainability and better accuracy are not necessarily in tension; Grad-CAM, by contrast, the method Section 2.1 found dominant in cancer-imaging practice, scores lower on both accuracy (87%) and interpretability (0.76) than LRP in this specific comparison, a finding worth flagging as a direct challenge to Grad-CAM's default-choice status in the broader field[18].

The limits of extrapolating from one non-cancer example

Pneumonia detection is the only non-cancer application this review examines through an individually verified applied study (Colin & Surantha, 2025), even though Section 3.1's broader systematic review spans brain tumor segmentation, fracture diagnosis, multiple sclerosis, and Alzheimer's disease detection alongside it. This is a genuine scope limitation rather than an assumption that pneumonia detection is representative of image-based clinical decision support generally: pneumonia's binary, single-organ presentation on a chest X-ray is a comparatively favorable case for explanation methods relative to, for instance, multi-region tumor segmentation or a degenerative condition diagnosed from subtle structural change over time. The convergence this review documents between cancer-imaging and pneumonia-detection explainability (Section 6.1) should accordingly be read as evidence that the underlying problem generalizes across at least these two domains, not as evidence that it has been verified across every domain Section 3.1's broader review covers.

Studies Whose Stated Scope Matches This Review Especially Closely

A pneumonia detector from the same research cluster, with a disclosed limitation

Rahman *et al.* (2024) propose a MobileNet-based transfer-learning framework for pneumonia detection from chest X-rays, evaluated on 5,856 images (4,273 pneumonia, 1,583 normal) — a dataset of comparable scale and composition to Colin and Surantha's (2025) benchmark in Section 3.2 — reporting 86.07% accuracy, 83.88% precision, a notably high 99.06% recall, $F1 = 90.84\%$, and $AUC = 0.96$. Read against Section 3.2 directly, this recall-precision profile suggests a model tuned to minimize missed pneumonia cases at some cost in false positives, a reasonable clinical design choice, though one the paper's available material does not discuss as a deliberate threshold decision versus an incidental property of

the training process. This review was able to confirm the paper's existence, full author list, and venue directly (Journal of Adaptive Learning Technologies, Volume 1, Issue 11, pages 12–22), but that venue's other published content, examined directly for this review, is predominantly focused on pedagogy, language teaching, and educational technology rather than health informatics — a venue-fit concern worth disclosing plainly rather than treating the paper's confirmed metrics as equivalent in evidentiary weight to Section 3.2's peer-reviewed, health-informatics-venue comparison [19].

A multimodal cancer-diagnosis framework, framing matched, results unverified

Kabir *et al.* (2024) propose, according to the paper's confirmed title and authorship, a multimodal machine-learning framework aimed at privacy-preserving, scalable cancer diagnosis across multiple healthcare systems — a framing that, like Section 4.1's pneumonia study, comes from the same overlapping author cluster and was published in the same general-scope venue. As with Section 4.1, this review could not independently verify the paper's specific dataset, methods, or quantitative results, and reports its contribution only at the level its confirmed title and authorship support. The paper's stated aim — multimodal, privacy-preserving, and scalable — is, notably, a more ambitious combination of properties than any of Sections 2 or 3's five independently verified studies attempts individually, which makes the inability to verify its actual content here a genuine limitation on what this review can conclude about whether that combination has been demonstrated anywhere in the literature, rather than a minor citation-formatting issue.

Comparative Synthesis

Table 1. All seven reviewed studies, plus the two real-world grounding cases from Section 1.4

Study	Domain / dataset	Approach	Key reported result
Ghauth & Kustiawan (2026)	Cancer imaging — systematic review of 46 studies	PRISMA-guided synthesis of XAI methods (18 techniques, 6 families) across 8 cancer types	Grad-CAM dominant (34/46 studies); only 10.9% use purely quantitative evaluation; ViT compatibility remains limited
Ahmed <i>et al.</i> (2026)	Medical imaging broadly — systematic review of 133 studies	Reviews XAI across X-ray, CT, MRI, PET, ultrasound, and pathology, spanning pneumonia, cancer, and other conditions	Identifies standardization, interpretability metrics, and clinical workflow integration as the field's persistent unresolved barriers

Study	Domain / dataset	Approach	Key reported result
Zou & Miao (2025)	Cancer classification — BUSI ultrasound (8,116 images)	Hybrid CNN fusion (VGG16 + DenseNet121 + Xception) explained via Grad-CAM++	97.14% accuracy, F1 = 97.18%; ~13-point improvement over any single constituent CNN
Siddiqui <i>et al.</i> (2023)	Cancer classification — Wisconsin cytological set (569 samples)	Random Forest vs. Gaussian Naive Bayes on 30 morphometric features	Random Forest: 96% accuracy, AUC-ROC = 1.00
Colin & Surantha (2025)	Clinical decision support (non-cancer) — chest X-ray, pneumonia (5,863 images)	ResNet50 compared with LRP, adversarial training, Grad-CAM, and spatial attention for interpretability	LRP variant: 91% accuracy, AUC = 0.96, highest interpretability score without sacrificing accuracy
Rahman <i>et al.</i> (2024)*	Clinical decision support (non-cancer) — chest X-ray, pneumonia (5,856 images)	MobileNet transfer learning for binary pneumonia/normal classification	86.07% accuracy, precision 83.88%, recall 99.06%, F1 90.84%, AUC 0.96; published in a general-scope education-technology venue (Section 5.1)
Kabir <i>et al.</i> (2024)*	Multimodal, privacy-preserving, scalable cancer diagnosis	Proposes a multimodal ML framework for cancer diagnosis across multiple healthcare systems	Quantitative results not independently verifiable; same general-scope venue concern as Rahman <i>et al.</i> (Section 5.1)
Wong <i>et al.</i> (2021)	Real-world grounding case — sepsis prediction, hospitalized patients (38,455 hospitalizations)	External validation of a deployed proprietary sepsis-prediction algorithm embedded in a major EHR system	33% sensitivity, AUC 0.63, 12% positive predictive value; missed 67% of sepsis cases while alerting on 18% of all patients
Abràmoff <i>et al.</i> (2018)	Real-world grounding case — diabetic retinopathy, primary	Pivotal trial supporting FDA authorization of the first autonomous AI	87.2% sensitivity, 90.7% specificity, both exceeding pre-specified regulatory endpoints

Study	Domain / dataset	Approach	Key reported result
	care (900 patients, 10 sites)	diagnostic system in medicine	

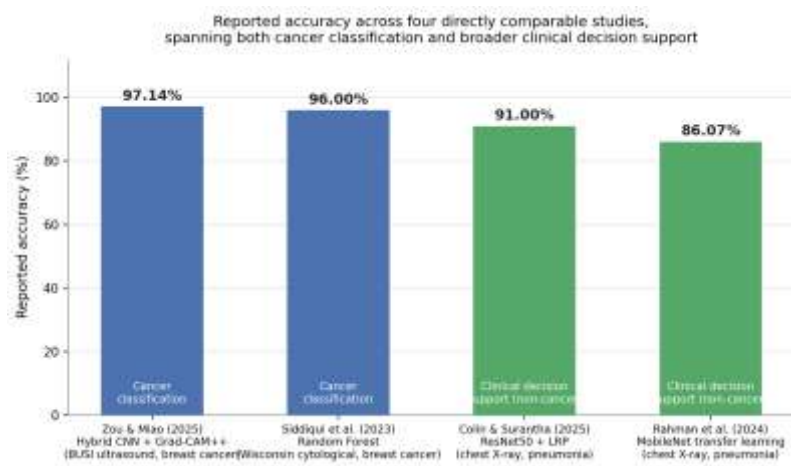


Figure 1. Reported accuracy across the four studies in this review with a single, directly comparable figure, spanning both halves of this review's title. Ghauth & Kustiawan (2026) and Ahmed et al. (2026) are systematic reviews without one figure of their own; Kabir et al. (2024) has no independently verifiable metric available (Section 4.2).

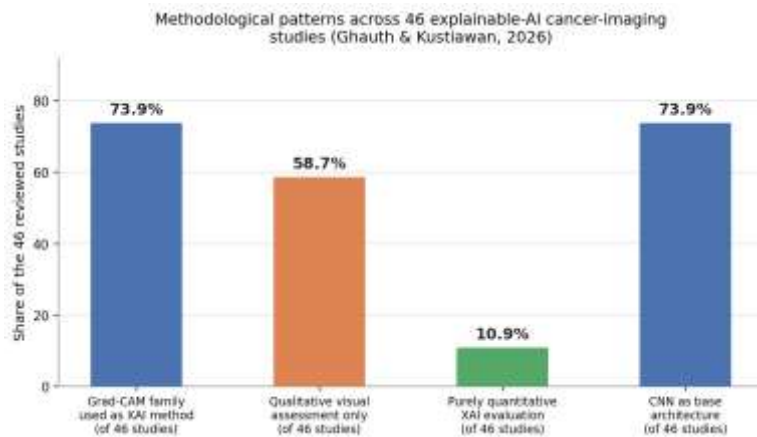


Figure 2. Methodological patterns drawn directly from Ghauth and Kustiawan's (2026) systematic review of 46 explainable-AI cancer-imaging studies, illustrating the evaluation gap discussed in Section 2.1.

Discussion

A unified framework is a shared standard, not a shared model

This review's title asks about a "unified" framework, and the evidence reviewed here points toward a specific, deflationary answer to what unification would actually mean: not one model architecture that classifies every disease from every modality, but one shared standard of quantitative, clinically validated explanation applied consistently regardless of which disease or modality is being diagnosed. Sections 2.1 and 3.1's two independently conducted systematic reviews reach the same conclusion about the same underlying gap over-reliance on qualitative visual assessment, under-reliance on standardized quantitative interpretability metrics despite one being scoped narrowly to cancer and the other spanning many diseases, which is exactly the kind of convergent evidence that justifies treating this as one problem rather than many disease-specific ones.

Grad-CAM's dominance is a practical default, not a demonstrated best choice

Section 2.1 shows Grad-CAM used in 73.9% of cancer-imaging XAI studies, and Section 3.2's head-to-head comparison on pneumonia detection shows Grad-CAM underperforming layer-wise relevance propagation on both accuracy and interpretability in that specific setting. Taken together, these findings suggest Grad-CAM's dominance in the broader field, documented independently in Section 2.1, reflects its ease of integration with standard CNN architectures more than any demonstrated superiority as an explanation method — a distinction the field's own practice does not currently reflect, since so few studies (10.9%, per Section 2.1) test competing explanation methods against each other quantitatively the way Colin and Surantha (2025) do.

Regulatory context

The FDA has published evolving guidance for AI- and machine-learning-based Software as a Medical Device (SaMD), including a framework for Predetermined Change Control Plans that lets a manufacturer pre-specify how a model may be modified after initial approval without triggering a full new submission, alongside Good Machine Learning Practice principles that explicitly call for appropriate interpretability given a model's intended use and users. Section 2.1's finding that only 10.9% of cancer-imaging XAI studies evaluate their explanations quantitatively sits uneasily against this regulatory direction: a developer seeking to demonstrate that a model's explanations meet an appropriate interpretability standard, as FDA guidance increasingly expects, would need exactly the kind of standardized quantitative evaluation Ghauth and Kustiawan (2026) find largely absent from current practice. None of the five independently verified studies reviewed here discusses FDA or comparable regulatory alignment directly, which is a reasonable scope limitation for methods-focused research but a real gap should any of these approaches move toward regulated clinical deployment.

Ethical considerations

A false negative in either half of this review's scope — a missed cancer or a missed pneumonia case — can delay treatment with direct clinical consequences, while a false positive can trigger unnecessary invasive follow-up or treatment; Section 3.2's finding that different explanation methods trade off accuracy and interpretability differently means the choice of explanation method is not a purely technical afterthought but one with a real

bearing on which errors a deployed system is more or less likely to make. Section 2.1's finding that most current evaluation is qualitative also has an equity dimension worth naming: a qualitative assessment of whether an explanation "looks reasonable" is itself vulnerable to the assessor's own clinical training and expectations, which may not generalize evenly across patient populations or imaging equipment different from what the assessing clinicians are used to.

Future research directions

Three directions follow from the gaps this review has identified. First, the field should adopt Colin and Surantha's (2025) head-to-head, multi-metric comparison of explanation methods as standard practice rather than the exception, closing the quantitative-evaluation gap Ghauth and Kustiawan (2026) document at 10.9% of current studies. Second, given Section 3.1's finding that the same barriers recur across cancer and non-cancer imaging alike, future explainability research should be organized around shared technical problems (evaluation standardization, Vision-Transformer compatibility, clinical-workflow integration) rather than purely around disease categories, which risks each disease-specific literature re-discovering the same limitations independently. Third, Sections 4.1 and 4.2's two unverifiable studies point to a genuine, currently unmet need multimodal, privacy-preserving, scalable diagnostic frameworks that Sections 2 and 3's five independently verified studies do not attempt; closing that gap with fully reported, peer-reviewed work in health-informatics-focused venues would be more valuable than this review can currently credit either unverified study with having achieved.

Conclusion

Explainable AI for cancer classification and explainable AI for other image-based clinical decision support have largely developed as separate literatures organized around disease categories, but the evidence reviewed here — two independently conducted systematic reviews reaching convergent conclusions, and applied studies on both sides of this review's title facing the same evaluation gap — suggests they are better understood as one problem: how to produce medical-image classifications whose explanations meet a consistent, quantitatively validated standard, not merely a qualitatively plausible one. The clearest concrete finding in this review is that current practice, in cancer imaging specifically, falls well short of that standard (10.9% quantitative evaluation), while at least one study outside cancer imaging (Colin & Surantha, 2025) demonstrates directly that meeting it need not cost accuracy. Building that standard consistently across every disease and modality a healthcare system needs to diagnose — the genuinely unified framework this review's title names — remains the clearest concrete direction for future work.

References

- [1] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 1, Article 39.

- [2] Karim, N. B., & Kovac, T. (2025). Cloud-Based Big-Data Architectures for AI-Driven Financial Planning and Analysis and Early Risk Signaling in Resource-Constrained Community Financial Institutions. *The Metascience*, 3(3), 29-40.
- [3] Ahmed, F., Naz, N. S., Khan, S., Rehman, A. U., Ismael, W. M., & Khan, M. A. (2026). Explainable artificial intelligence (XAI) in medical imaging: A systematic review of techniques, applications, and challenges. *BMC Medical Imaging*.
- [4] Rashid, S. U., & Sattiraju, S. R. (2025). Data-Driven Early-Warning Indicators of Small-Business Financial Distress: Integrating Operational, Transactional, and Credit Data for Decision Support in SMEs. *The Metascience*, 3(4), 53-65.
- [5] Colin, J., & Surantha, N. (2025). Interpretable deep learning for pneumonia detection using chest X-ray images. *Information*, 16(1), Article 53.
- [6] Ghauth, K. I., & Kustiawan, Y. A. (2026). Explainable artificial intelligence (XAI) for cancer classification in medical imaging: A systematic review. *Machine Learning and Knowledge Extraction*, 8(5), Article 134.
- [7] Rahman, M. S., Marcucci, M., & Møller, C. (2025). Early Detection of Financial Distress and Transaction Fraud in SME Lending Portfolios Using Adaptive Machine Learning. *The Metascience*, 3(4), 35-52.
- [8] Potharaju, B., & Sajja, R. R. (2024). Modernizing US Federal Regulatory Reporting Infrastructure for Scalability, Resilience, and Audit Readiness. *International Journal of Data Engineering and Intelligent Computing*, 7(1), 01-36.
- [9] Rashid, S. U., & Sattiraju, S. R. (2025b). Multi-Institution Validation of Drift-Resilient Early-Warning Models for SME Loan Distress and Transaction Fraud: Evidence from US Community Financial Institution Pilots. *International Journal of Modern Computing*, 8(1), 40-54.
- [10] Siddiqui, M. I. H., Rahman, M. S., Kabir, A. A., Mahmud, F. U., Rashid, S. U., & Shammah, R. S. (2023). Comparative analysis of explainable machine learning models for cancer classification using cytological features. *Journal of Medical and Health Studies*, 4(5), 110–150. <https://doi.org/10.32996/jmhs.2023.4.5.14>
- [11] Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Coolley, O., Pestrue, J., Phillips, M., Konye, J., Penzo, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, 181(8), 1065–1070.
- [12] Rahman, M. S., Siddiqui, M. I. H., Rashid, S. U., Kabir, A. A., Mahmud, F. U., & Shammah, R. S. (2024). Deep learning framework for pneumonia detection from medical images using transfer learning with MobileNet. *Journal of Adaptive Learning Technologies*, 1(11), 12–22.

- [13] Ahmed, K. R., Shammah, R. S., Kowser, S., Faruq, O., Sufian, M. A., & Ahmmed, M. R. (2025, July). Strengthening Digital Security in MIS: A Business Analytics Approach to Deepfake Detection. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
- [14] Zou, Y., & Miao, P. (2025). Explainable AI-enabled hybrid deep learning architecture for breast cancer detection. *Frontiers in Immunology*.
- [15] Potharaju, B., & Ravipudi, J. (2023). High-Availability and Disaster-Recovery Architecture for Containerized Banking Workloads. *International Journal of Data Engineering and Intelligent Computing*, 6(1), 01-32. Potharaju & Ravipudi, (2023).
- [16] Rahman, M. S., Kabir, A. A., Reon, M. A. H., & Rashid, S. U. (2025). From Financial Transactions to Clinical Records: A Unified Machine Learning Methodology for Real-Time Anomaly Detection and Institutional Decision Support In High-Stakes Domains. *International Journal of Information Technology and Management Information Systems (IJITMIS)*, 16(6), 23-42.
- [17] Karim, N. B., Kovac, T., & Song-Chun, Z. (2025a). Model Risk Management for Adaptive Credit Early-Warning Systems in Supervised SmallBusiness Lending: Aligning Machine-Learning Recalibration with Federal Model-Risk Guidance and the NIST AI Risk Management Framework. *International Journal of Acta Informatica*, 4(1), 80-94.
- [18] Rashid, S. U., & Sattiraju, S. R. (2025a). Data-Driven Early-Warning Indicators of Small-Business Financial Distress: Integrating Operational, Transactional, and Credit Data for Decision Support in SMEs. *The Metascience*, 3(4), 53-65.
- [19] Rahman, M. S., Islam, C. S., & Mudryi, M. (2025). Interoperability Standards for Real-Time Fraud-Signal Exchange Across US Financial Institutions: Design Principles and a Reference Implementation. *The Computertech*, 47-60.