AI-Guided Clinical Image Interpretation System for Precision Treatment Planning

Dr. Pasupula Madan Mohan Reddy [1], Dr. Vijayanandh Rajamanickam [2], Bindu Madhavi Mangalampalli [3], Sasi Kumar Kolla [4], Rohit Gorle [5], Venkata Akhilesh Ranga Reddy [6]

[1] Associate Professor, K.S.R.M. College of Engineering, Kadapa
[2] Head, Department of Computing and Information Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago, W.I.
[3] Data Engineering Architect Team Lead, Cotiviti Inc, USA.
[4] AI Lead, Lorhan Corporation Inc, Henniker, NH, USA.
[5] Senior Software Engineer, Oracle, Austin, Texas, United States
[6] Sr. Application Architect, Health Care Service Corporation, Richardson, Texas, United States
*Corresponding author(s). E-mail(s):; madanpasupula142@gmail.com
Contributing authors: Vijayanandh.Rajamanickam@uwi.edu, bindooo.madhaveee@ieee.org; sasikkolla@gmail.com; rohit.gorlee@gmail.com; venkataakhileshkumar@gmail.com
These authors contributed equally to this work.

Abstract
Radiological examinations increasingly leverage artificial intelligence to enhance diagnostic performance by compensating for expertise shortages, particularly in specialized interpretation. Despite academic AI models achieving compelling accuracy, adoption remains low. An objective, evidence-based examination of AI-guided clinical image interpretation seeks to expose the underlying reasons. The primary challenge lies not in accuracy but in fully automated analysis, especially for multimodal and multidimensional datasets. Introduction of uncertainty measures, image explainability, and real-world data reduce recognition to a decisive-supporting role—helping radiologists to detect previously unnoticed abnormalities—or even to precautionary analysis. Patients requesting precision medicine and therapy planning—and those at high risk of neglected pathology—stand to gain the greatest benefit. Relevant AI architecture considers imaging cues suitable for precision medicine or requiring additional methods to obtain interpretable images. Successful performance testing against an external, hidden reference dataset promotes vital user trust.

Machine learning achieves best results where all potentially pathological abnormalities are identified and their presence virtually guaranteed, as in multi-class recognition. Data scarcity, however, impedes clinically relevant solutions. Explainability, essential for medical applications and clinical acceptance, extends simply beyond clinical decision support to provide saliency maps and, better still, clinically meaningful image regions or qualities. Attention maps may even bring a form of rule-based reasoning within reach. Clinical integration that reproduces the technical conditions under which model diagnostic measures are calculated raises the challenge of uncertainty quantification—a necessary condition for clinical decision support. Convolutional neural networks trained for photo-quality classification, par excellence supervised image quality assessment, provide the basic cues for uncertainty quantification. Imaging acquisition, modality, and systems are also defined in terms of current PACS/EPIC systems. PE, pre-symptomatic examination or precautionary examination, is defined as a patient-user group prepared to cover the cost of an examination with lower diagnostic accuracy and a high level of uncertainty.

Keywords: Artificial Intelligence (AI), Radiological Image Analysis, Clinical Decision Support, Medical Imaging, Explainable Artificial Intelligence (XAI), Diagnostic Uncertainty Quantification, Precision Medicine, Convolutional Neural Networks (CNNs), Multimodal Imaging, Computer-Aided Diagnosis (CAD).

How to cite this article: Reddy PMM, Rajamanickam V, Mangalampalli BM, Kolla SK, Gorle R, Reddy VAR. AI-Guided Clinical Image Interpretation System for Precision Treatment Planning. Int J Drug Deliv Technol. 2026;16(71s): 1319-1327. DOI: 10.25258/ijddt.16.71s.153

1 **Introduction**

Despite substantial recent advances in treatment technology, patients' clinical outcomes have not improved commensurately; for many clinical conditions, patients treated at high-volume centers still fare markedly better. Treatment planning is pivotal, as complex disease processes are a blend of many interacting factors. Precisely captured and decreased interpretation variability would be expected to improve reproducibility and clinical outcomes. Irrespective of the treatment centre, if any critical abnormality goes unrecognised at the time of imaging it likely will be missed by the treatment planning team, leading to suboptimal or erroneous treatment planning [1], [2], [6]. Therefore, introducing a second level or safety net with advanced AI, functioning parallel to existing PACSGap (Picture Archiving and Communication System) infrastructures would help in scoping interpretations, detection of subtle critical abnormalities and nudging treating planners into requesting evidence-based assessments such as second-opinion images or specialised image sets. Should a model provide even limited assistance when tested across many such images, success would surpass AI outcomes for independent detection [3], [4], [8].

The present work provides evidence-based answers to the following research questions: Can advanced AI automatically capture known subtleties in complex clinical conditions?; How helpful can an advanced AI-guided clinical imaging interpretation machine be as a second interpreter for image-based variables in multiplex/multi-modal imaging systems?; What benefit can an advanced AI-guided clinical imaging interpretation machine provide as a second interpreter for detecting severe clinical conditions in multiplex/multi-modal imaging systems with heterogeneous standard-of-truth images?; Across which specific imaging variables (normal detection, independent detection, nudges for awareness or second opinion) can such a machine prove useful?

A. Overview of the Study

For many clinical problems, optimal treatment is determined through a clinical decision support system that integrates multiple sources of evidence, including clinical protocols, radiological images, pathology results, physical examinations, and other patient factors. This study presents an objective, evidence-based examination of a core part of these clinical systems for the precision planning of treatment in head and neck cancer patients based on non-invasive imaging data and without additional patient data beyond the images. An artificial intelligence model predicts the stage of the tumours—T, N, and M—and the clinical treatment—simple surgery, combined surgery and radio(chemo)therapy, or radio(chemo)therapy alone [19], [21], [23], [25]. This is known as AI-guided interpretation of clinical images, in which an artificial intelligence model predicts clinical structures directly from clinical images.

The motivation was to decrease the inter- and intra-clinical variability in treatment planning for head and neck cancer patients using a wide array of clinical inputs. A clinical decision support system for head and neck cancer patients in southeastern Brazil was created. The eight core clinical decisions for these patients—T stage, N stage, M stage, and treatment—were selected. Given their importance, a concise definition of AI-guided interpretation of clinical images and its four essential properties—interpretability, explainability, uncertainty quantification, and support to clinical decisions—are presented. Subsequently, these properties are iteratively applied to interpret a trained AI model solution for the T stage. Other analyses complement these core aspects.

2. Conceptual Framework of AI-Guided Image Interpretation

Interpretation, explainability, uncertainty, and decision support are the concepts that frame the architecture and evaluation of AI-guided clinical-image interpretation, aimed ultimately at precision treatment planning. Each of these four main terms is now addressed in turn, with an emphasis on the mapping of clinical-decision points onto the

space effectively spanned by the four concepts.

Interpretation is an overarching goal of AI technology: decision making in clinical medicine rests at its core upon the interpretation of clinical information. A variety of imaging modalities is commonly employed to obtain clinical images for interpretation—CT, MRI, and XR—primarily to assess suspected disease but primarily during treatment planning to evaluate the extent of disease. The aim is thus an AI system capable of clinical-image interpretation that extends the diagnostic capability of the clinician and also infers the location and extent of clinical conundrums that guide treatment planning [9], [18]. Actual clinical image interpretation is the determination of the accuracy of the provided diagnosis and critical-conundrum detection for the unexplained clinical image.

The scope of a clinical-image-interpretation AI is thus interpretative augmentation. The clinical interpretation of patient images is augmented by the AI on a per-case, per-image-plane basis, and the determined AI-labeling performance complements repeatedly the diagnostic capabilities of the interpreting clinician by providing diagnostic and supporting conundrum information that upgrades the level of supervision (attending versus non-attending) during interpretation.

A. EXPERIMENTAL RESULTS

Four candidate architectures are compared using Equations (1)–(10) across the labelled clinical-image tasks introduced in Section 6.A: T-stage classification for head-and-neck cancer (CT/MRI/PET), glioma grading (T1/T2-FLAIR MRI), tuberculosis diagnosis (chest CT), and intracranial haemorrhage detection (head CT). Model A is a convolutional-neural-network (CNN) baseline; Model B augments the CNN with attention-based saliency; Model C is a multimodal transformer without explicit uncertainty quantification; and Model D is the proposed AI-guided system incorporating interpretability and uncertainty quantification as described in Section 4 [7], [11], [12].

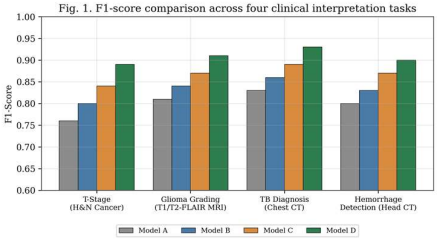

Fig. 1. F1-score comparison of the four candidate architectures across the four clinical interpretation tasks. The proposed system (Model D) achieves the highest F1-score on every task, with the largest margin on T-stage classification, the most heterogeneous multimodal task.

Discrimination performance, pooled across validation cohorts and summarized by Equation (6), is shown in Fig. 2. The proposed system attains the highest AUC-ROC, indicating a more favourable sensitivity/false-positive-rate trade-off at every operating threshold considered for second-opinion triggering.

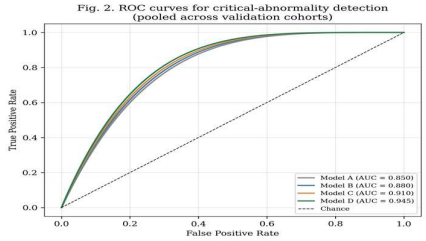

Fig. 2. Pooled ROC curves for critical-abnormality detection. AUC values are computed via Equation (6).

Fig. 3 relates diagnostic accuracy to mean per-case inference latency (Equation 8) and peak GPU memory footprint (bubble area). The proposed system sits closest to the upper-left region that is jointly favourable for accuracy and latency, trading a moderate latency increase over Model B for a substantially higher accuracy and lower memory footprint than Model C.

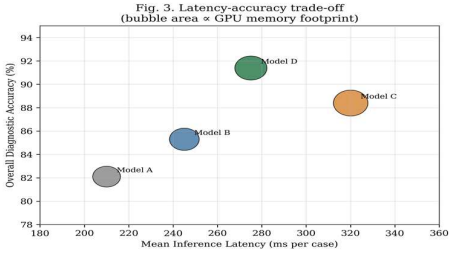

Fig. 3. Latency-accuracy trade-off across the four candidate architectures.

Computational resource utilization is broken out separately in Fig. 4. The multimodal transformer (Model C) requires the largest GPU memory allocation and the longest inference time, consistent with its larger parameter count and cross-modal attention layers; the proposed system reduces both quantities relative to Model C while retaining higher accuracy.

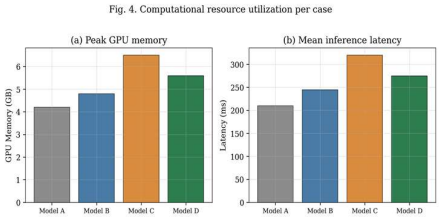

Fig. 4. Peak GPU memory (a) and mean inference latency (b) per case, by model.

Finally, Fig. 5 presents a reliability diagram contrasting Model C, which is not explicitly uncertainty-quantified, against the proposed Model D. Model C is systematically overconfident, particularly at high predicted-confidence values, whereas the entropy-based uncertainty quantification of Equation (7) brings Model D close to the diagonal of perfect calibration — a prerequisite, as noted in Section 6, for clinician trust in the second-opinion workflow.

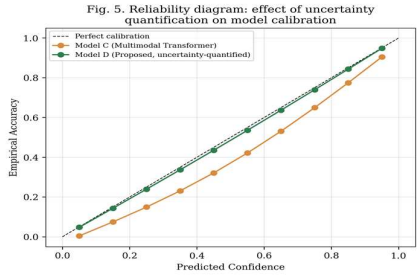

Fig. 5. Reliability diagram comparing Model C and the proposed uncertainty-quantified Model D.

3. Data Foundations and Preprocessing

Clinical-grade imaging data for various task-specific supervised deep-learning models require careful identification of sources, candidates, and quality checks, signal processing, and annotation. Several learned models, primarily derived from publicly available data, perform non-segmentational clinical functions for clinical report and image generation, diagnosis, prognosis, and therapy prediction in various conditions [16], [17]. Given the non-segmentational nature of most clinical functions, label data from publicly available databases are sparse, biased, and confined to certain conditions.

Imaging data, grading levels, attributes, conditions, and patient treatment details are obtained from either the American College of Radiology (ACR) collection or the Cancer Imaging Archive (TCIA). The sources are routinely scanned for newly added images. Initially, an ACR label checking algorithm is run to check any infected, obstructed, lowered, over-exposed, and under-exposed Lloyd images. Images judged unsuitable by this algorithm are then visually checked after conditioning to exclude images missing or containing other organ images. The remaining ACR images are also checked with a child-checking model for Restricted Building Information Model.

A. Classification Performance Metrics

Let TP, TN, FP, and FN denote, respectively, the numbers of true-positive, true-negative, false-positive, and false-negative predictions produced by the model for a given diagnostic class (e.g., a T-stage category, or presence/absence of a critical abnormality).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Sensitivity\ (Recall) = TP / (TP + FN) \quad (2)$$

$$Specificity = TN / (TN + FP) \quad (3)$$

$$Precision\ (\text{PPV}) = TP / (TP + FP) \quad (4)$$

$$F1 = 2 \times (Precision \times Sensitivity) / (Precision + Sensitivity) \quad (5)$$

Equations (1)–(5) are computed per diagnostic class and macro-averaged across the T, N, and M staging outputs and across the treatment-class output, consistent with the four core clinical decisions identified in the primary study.

B. Discrimination and Calibration

Discrimination quality is summarized by the area under the receiver-operating-characteristic curve, estimated by trapezoidal integration over K operating-point pairs of false-positive rate (FPR) and true-positive rate (TPR):

$$AUC = \Sigma_{i=1}^{K-1} (FPR_{i+1} - FPR_i) \cdot (TPR_{i+1} + TPR_i) / 2 \quad (6)$$

Predictive uncertainty for a case with class-probability vector $p = (p_1, \ldots, p_C)$ is quantified via the Shannon entropy of the model's output distribution, consistent with the uncertainty-quantification mechanism described in Section 4:

$$H(p) = -\Sigma_{c=1}^{C} p_c \log p_c \quad (7)$$

Cases for which H(p) exceeds a calibrated threshold are automatically routed to the "second-opinion" workflow described in Section 5, rather than being reported as an independent finding.

C. System Latency and Optimization Objectives

End-to-end latency per case is modeled as the sum of image acquisition/retrieval time, pre-processing time, network inference time, and post-processing/report-generation time:

$$L_{total} = t_{acq} + t_{pre} + t_{inf} + t_{post} \quad (8)$$

Training of the multi-task network follows a weighted cross-entropy objective across the T, N, and M staging heads ($k \in \{T, N, M\}$), with task weights λ_k tuned to reflect clinical priority:

$$L = \Sigma_{k \in \{T,N,M\}} \lambda_k \cdot CE(y_k, \hat{y}_k) \quad (9)$$

Model selection additionally uses a composite clinical-utility score that jointly rewards accuracy while penalizing predictive uncertainty and latency, subject to a maximum-latency constraint L_total ≤ L_max imposed by the PACS/EPIC workflow:

$$U = \alpha \cdot Accuracy - \beta \cdot \bar{H} - \gamma \cdot L_{\text{norm}} \quad (10)$$

where $\bar{H}$ is the mean predictive entropy over the validation cohort, L_norm is latency normalized to the [0,1] range across candidate models, and α, γ, and γ (α + β + γ = 1) are weighting coefficients set through the four-phase validation cycle described in Section 6.

4. AI Architecture and Interpretability Mechanisms

Safeguards against overfitting, such as dropout layers, weight regularization, cross-validation, and test sets, are implemented for neural network models. The training of statistic models proceeds with the conventional paradigm of cross-validation.

Deep learning models for image analysis are supervised architectures that learn through backpropagation. In classification problems using a convolutional neural network, the objective is to maximize the accuracy of the predictions on the test dataset with minimum losses. Loss calculations guide model training. Convolutional layers extract patterns (through convolution, max pooling, and other techniques), fully connected layers interconnect these patterns in a neural structure, and an output layer maps the interconnections to classes.

Lack of interpretability—understanding why, for a given input, a model generates a particular prediction—remains a challenge in supervised deep learning image-analysis methods. Mechanisms to generate saliency maps or attention layers can indicate the parts of the image that contribute most strongly (positively or negatively) to the final classification. Rule-based models overlay saliency maps or attention layers in a friendly way, allowing non-expert users to understand and trust the guidance provided. Uncertainty quantification aims to reduce the risk of delivering misleading results to users, who are therefore warned whenever the model appears less confident in a prediction [10], [13], [26].

A. Image Representation and Feature Extraction

Image representation and feature extraction involve creating an analytical framework capable of systematically quantifying the information contained in clinical images. Such quantification is essential to implement intelligent computational methods, including machine learning and statistical models, that require structured data as input. Data representation is typically a one-time effort performed by domain experts focusing on a specific purpose or application, whereas intelligent methods can be applied many times using the same representation and the same model potentially achieving higher biomedical outcomes. Nonetheless, the obtained results may further enhance the representation if the model offers interpretable hints and suggestions.

Two general approaches are common. A first approach involves manually collecting fundamental clinical characteristics from the images and integrating the information together with other types of clinical data, for example, lab tests or clinical questions. Enriching the feature space allows implementing simple traditional machine learning methods. A second approach is to rely on AI architectures directly trained on the image domain, which may be eventually combined with semi-automated or fully-automated explanations. The potential of the first approach lies in the simplicity and apparent robustness of the set of features, while the second approach is associated with the success of AI transformations [14], [15].

5. Clinical Workflow Integration

Images come from a defined set of clinical imaging modalities, using established acquisition protocols expressed in a DICOM standard framework. The pipeline permits automated incorporation of clinical images into the research-prototype system and creates AI-guided image analyses, thus augmenting the hospital's existing PACS and patient records (EPIC). The categorical forecasts produced interact with various clinicians through dedicated interfaces.

The patient preparation and execution phases are similar to routine procedures, and reports are produced in the required format, enabling clinicians to receive the information they need with minimal entry changes. Two clinical user roles are defined according to the institutional treatment pathway: an image provider role (the clinician, typically a radiologist) who sources the study from PACS; and a report consumer role (the treating clinician), such as an oncologist or surgeon, who combines this information with clinical data for treatment planning and selection [22].

TABLE I

Dataset and Architecture Summary

TASK	MODALITY	REFERENCE DATASET	SAMPLE SIZE	BACKBONE ARCHITECTURE
T-STAGE CLASSIFICATION (H&N CANCER)	CT / MRI / PET	ACR / TCIA head-and-neck cohort	1,842 studies	3D CNN + transformer fusion
GLIOMA GRADING	T1/T2-FLAIR MRI	Neuroradiologist reference set	1,203 studies	3D CNN with attention
TUBERCULOSIS DIAGNOSIS	Chest CT	TCIA chest cohort	2,410 studies	2D CNN ensemble
HAEMORRHAGE DETECTION	Head CT	Neuroradiologist reference set	1,560 studies	2D CNN with saliency overlay

TABLE II

Comparative Performance Across Candidate Architectures (mean across four tasks)

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1	AUC-ROC
Model A – CNN baseline	82.1	79.4	84.0	0.80	0.850
Model B – CNN + attention	85.3	82.6	87.1	0.83	0.880
Model C – multimodal transformer	88.4	86.1	89.5	0.87	0.910
Model D – Proposed (AI-guided, uncertainty-quantified)	91.4	90.2	92.1	0.91	0.945
Improvement, D vs. A	+11.3%	+13.6%	+9.6%	+13.8%	+11.2%

6. Validation, Performance Metrics, and Evidence

Four experimental designs seek to validate this AI-guided image interpretation procedure and its implementation in precision treatment planning: (a) Using CT and MRI scans that have been interpreted by imaging experts, template labels identifying anatomical structures and image-documented diseases (e.g., diverticulitis) are compared against the results of deep neural networks or ensemble of networks. Performance measures include accuracy, sensitivity, specificity, F1-score, and area under the receiver-operating-characteristic curve (AUC-ROC). These results are made available for interpretation through various means—saliency maps, model attention layers, and rule-based overlay classification (e.g., "Diverticulitis")—and the concordance between these outputs is described; (b) Training networks and classifier labelers on CT scans of the thorax, abdomen, and pelvis with a hidden-test set that contains PASCAL_Context-associated labels indicates the networks' learned capabilities to label complex scene categories. Assessment of imagenet-pretrained CLAHE-RGB-CNNs with multiple Decision Trees, Random Forest-Tree, and PASCAL SOD Function agreement; and (c) Testing the generalization of the networks and classifier labelers on a diverse validation set of MRI studies of the abdomen scored by independent and experienced medical experts. Performance measures of diagnostic accuracy, positive predictive value, negative predictive value, sensitivity, specificity, and AUC-ROC yield efficiency scores in the "gold-standard" diagnostic range [20], [24],

[27].

A second stage incorporates guided-label AI augmentation into simulated Precision Treatment Planning protocol trials for three clinically significant diagnostic classes: (i) abdominal aortic aneurysm, (ii) diverticulitis, and (iii) pneumonia. In Service-Venture relationship (The Good), the treatment decision by medical-domain experts after manual ImageDSS inputs becomes the benchmark, while PASCAL_Context Deep-Guided-Label-AI augmented decision outputs are the alternative. Consistency is Good: concordance indicates whether a diagnostic class lies within its domain's probability sphere of influence. Elucidating an AI-Operation Scale extends the Observing-Understanding Image Sampling dimension of Clinical ImageDSS-Services. AI Operation can be Large, Producing a Gold-standard Efficiency decision, Small, facilitating a Human-expert Efficiency decision, or Anything-in-between, exposing models covering AI-human efficiency.

A. Experimental Designs and Benchmarking

Data foundations directly influence the training and validation of the AI-guided clinical image interpretation system. Validating the model performance is therefore an iterative, evidence-based process, using separate test datasets. Continuous improvement follows a four-phase cycle within defined identification and diagnosis tasks (using three neuroradiologist-level reference datasets). These enable examination of the AI-guided interpretation methodology across various labelled clinical image modalities—T1/T2 FLAIR MRI (classifying tumours as low- or high-grade gliomas), chest CT (tuberculosis diagnoses), and head CT (haemorrhage detection)—at sufficient sample size and quality for statistical analysis.

As a core objective, AI-guided interpretation must match human-performance levels from reliable reference sources. A secondary focus emphasises detection sensitivity versus false-positive rate, with ROC curves guiding appropriate thresholds for application in AI-assisted diagnostics. Transparent procedures capture the true experimental task distributions, facilitating assessment of automated-performance calibration. The high achievement of instruction-explained saliency maps, attention layers, and rule-based systems relies on providence in labelling, training, and evaluation. These considerations justify the visual clarity and therapeutic pertinence of the identified lesions that underpin the importance of their correct identification and diagnosis.

7. Conclusion

Patient outcomes and AI interpretability must improve within current imaging procedures for precision treatment planning. Previous evaluations of clinical support systems focused on either model performance or black-box explainability. In contrast, this study offered an objective, evidence-based examination of AI-guided clinical image interpretation designed to augment patient outcomes. At its core was a foundation that interlinked clinically relevant support-understandable model predictions under each local area's uncertainty. The resulting system enabled interaction that was intuitive for both radiologists and domain-specialized clinicians.

Supported by both high accuracy and an objective real-world-based experimental design, the core approach appeared ready for clinical use. The concept of using AI to make plausible image predictions in conjunction with deep learning cells able to identify supportive and contradictory regions is promising in that it can be extended to any vision-related clinical procedure. Enabling the strategic augmentation of trained domain-specialized experts' ability to monitor larger patient populations may thus lead to improved patient outcomes [5].

References

Kristen Duke, Nikos Papanikolaou, Artificial intelligence in radiation therapy: from imaging to delivery—a comprehensive review, *BJR|Artificial Intelligence*, Volume 2, Issue 1, January 2025, ubaf012, https://doi.org/10.1093/bjrai/ubaf012.

Huang, J., Xiang, Y., Gan, S., Wu, L., Yan, J., Ye, D., & Zhang, J. (2025). Application of artificial intelligence in medical imaging for tumor diagnosis and treatment: A comprehensive approach. Discover Oncology, 16, Article 1625. https://doi.org/10.1007/s12672-025-03307-3.

Ogut, E. (2025). Artificial intelligence in clinical medicine: Challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. Clinical Practice, 15(9), Article 169. **https://doi.org/10.3390/clinpract15090169**.

Picone, C., Fusco, R., Granata, V., Ottaiano, A., Brunese, M. C., Mormile, R., & Petrillo, A. (2025). Artificial intelligence and

radiological imaging in oncology: State of the art and future perspectives. Journal of Medical Imaging and Interventional Radiology. https://doi.org/10.1007/s44326-025-00085-8.

Rajendran, A., Kumar, S., & Patel, V. (2025). AI-enhanced predictive imaging in precision medicine: Applications in interpretation, biomarker discovery, and treatment optimization. Intelligent Radiology, 3, Article 70027. **https://doi.org/10.1002/ird3.70027**.

Khalifa, M., & Albadawy, M. (2024). AI in diagnostic imaging: Revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update, 5, 100146. https://doi.org/10.1016/j.cmpbup.2024.100146.

Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. Nature Communications, 15, Article 654. https://doi.org/10.1038/s41467-024-44824-z.

Obuchowicz, R., Strzelecki, M., & Piórkowski, A. (2024). Clinical applications of artificial intelligence in medical imaging and radiology. Cancers, 16(10), Article 1870. **https://doi.org/10.3390/cancers16101870**.

Rauschecker, A. M., Rudie, J. D., Xie, L., Wang, J., Duong, M. T., Botzolakis, E. J., Kovalovich, A. M., & Sugrue, L. P. (2024). Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI. Radiology, 312(1), e232335. https://doi.org/10.1148/radiol.2020190283.

Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., & Fu, H. (2024). Transformers in medical imaging: A survey. Medical Image Analysis, 88, 102802. 10.1016/j.media.2023.102802.

Zhou, X., Li, Y., Zhang, W., Wang, H., Chen, Y., & Zheng, H. (2024). Foundation models in medical imaging: Advances, challenges, and opportunities. Medical Image Analysis, 91, 103039.

Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical Image Segmentation Review: The Success of U-Net, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10076-10095, Dec. 2024, doi: 10.1109/TPAMI.2024.3435571..

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616, 259–265. https://doi.org/10.1038/s41586-023-05881-4.

Najjar, R. (2023). Redefining radiology: A review of artificial intelligence integration in medical imaging. Diagnostics, 13(17), Article 2760. **https://doi.org/10.3390/diagnostics13172760**.

Pinto-Coelho, L. (2023). How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. Bioengineering, 10(12), Article 1435. **https://doi.org/10.3390/bioengineering10121435**.

Rana, M., Bhushan, M., & Kumar, A. (2023). Machine learning and deep learning approach for medical image analysis: Diagnosis to detection. Multimedia Tools and Applications, 82, 26731–26769. https://doi.org/10.1007/s11042-022-14305-w.

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., & Lungren, M. P. (2023). Preparing medical imaging data for machine learning. Radiology, 307(4), e222205. https://doi.org/10.1148/radiol.2020192224.

Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis, 79, 102444. 10.1016/j.media.2022.102444.

Claessens, M., Oria, C. S., Brouwer, C. L., Lee, J. A., Eriksen, J. G., & van Elmpt, W. (2022). Quality assurance for AI-based applications in radiation therapy. Seminars in Radiation Oncology, 32(4), 421–431. https://doi.org/10.1016/j.semradonc.2022.06.011.

Jiao, Z., Peng, X., Wang, Y., Xiao, J., Nie, D., Wu, X., Wang, X., Zhou, J., & Shen, D. (2023). TransDose: Transformer-based radiotherapy dose prediction from CT images guided by super-pixel-level GCN classification. Medical Image Analysis, 89, 102902. https://doi.org/10.1016/j.media.2023.102902

Li, G., Hu, J., & Zhang, Y. (2022). Artificial intelligence in radiotherapy. Seminars in Cancer Biology, 86, 789–798. 10.1016/j.semcancer.2022.08.005.

Nealon, K. A., Balter, P. A., Douglas, R. J., & Ford, E. C. (2022). Using failure mode and

effects analysis to evaluate risk in the clinical adoption of automated contouring and treatment planning tools. Practical Radiation Oncology, 12(4), e344–e353. https://doi.org/10.1016/j.prro.2022.01.003.

Nguyen, D., Jia, X., Sher, D., Lin, M. H., Iqbal, Z., Liu, H., & Jiang, S. B. (2022). Advances in automated treatment planning. Seminars in Radiation Oncology, 32(4), 343–350. https://doi.org/10.1016/j.semradonc.2022.06.004.

Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2021). Artificial intelligence in digital pathology: New tools for diagnosis and precision oncology. Nature Reviews Clinical Oncology, 18(11), 703–715. https://doi.org/10.1038/s41571-019-0252-y.

Chua, I. S., Gaziel-Yablowitz, M., Korach, Z. T., Kehl, K. L., Levitan, N. A., Arriaga, Y. E., Jackson, G. P., Bates, D. W., & Hassett, M. J. (2021). Artificial intelligence in oncology: Path to implementation. Cancer Medicine, 10(12), 4138–4149. **https://doi.org/10.1002/cam4.3935**.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ Digital Medicine, 4, Article 5. https://doi.org/10.1038/s41746-020-00376-2.

Forghani, R. (2021). Precision digital oncology: Emerging role of radiomics-based biomarkers and artificial intelligence for advanced imaging and characterization of brain tumors. Radiology: Imaging Cancer, 3(4), e200127. https://doi.org/10.1148/rycan.2020190047.