

Original Research

Received 29 May 2026

Revised 28 July 2026

Accepted 06 August 2026

Natural Resources for Human Health
2026, Vol. 6 (14s): 1437–1447

KEYWORDS: Artificial Intelligence; Machine Learning; Radiation Therapy; Quality Assurance; Quality Management; Patient Safety; Medical Physics; Vision 2030

Integrating Artificial Intelligence into Radiation Therapy Quality Management: Opportunities and Challenges

A Structured Narrative Review with a Proposed Governance Framework (PRISM) for the Saudi Vision 2030 Context

¹Sultan Al Nujaiman,²Amer AlAmeer,³Areej alfayez,⁴Mohammed Almutairi,⁵Abdulrahman Alotaibi

Workplaces/Radiation Therapist, King Abdulaziz Medical City- Ministry Of National Guard Health Affairs

ABSTRACT:

Background: Radiation therapy (RT) is among the most technically complex and safety-critical services in modern medicine. Its quality management (QM) programme spans machine performance, treatment planning, patient-specific verification, and delivery. Conventional QM relies heavily on labour-intensive, measurement-based checks that are increasingly strained by rising patient volumes and advanced techniques such as intensity-modulated, stereotactic, and adaptive radiotherapy.

Objective: To synthesise current evidence on the application of artificial intelligence (AI) and machine learning (ML) to RT quality management, to critically examine the associated technical, clinical, ethical, and regulatory challenges, and to propose an implementation framework relevant to the Kingdom of Saudi Arabia.

Methods: A structured narrative review of peer-reviewed literature, professional task group reports (AAPM, ESTRO), and regulatory guidance published between 2009 and 2026, interpreted through Donabedian's structure–process–outcome model, Reason's Swiss cheese model, and the risk-based philosophy of AAPM Task Group 100.

Findings: AI demonstrates substantial promise in predictive machine QA, virtual patient-specific QA, automated plan and chart review, contour error detection, real-time delivery monitoring, and natural-language analysis of incident reports. However, adoption is constrained by limited and heterogeneous data, poor generalisability across institutions, model drift, limited explainability, automation bias, the unresolved need to quality-assure the AI tools themselves, and evolving regulatory and liability frameworks.

Conclusion: AI should be positioned as an additional, independent layer of defence rather than a replacement for expert oversight. The proposed PRISM framework (Proactive risk intelligence, Robust validation, Integrated human–AI oversight, Standards-aligned governance, and Monitoring for lifecycle learning) offers a pathway aligned with SFDA, NRRC, SDAIA, SCFHS, and CBAHI requirements and with the Health Sector Transformation Program.

1. INTRODUCTION

Radiation therapy is a cornerstone of cancer care, indicated for roughly half of all patients with cancer at some point in their disease course. It is also one of the most technologically intensive disciplines in healthcare. A single course of treatment involves a chain of interdependent steps: simulation imaging, target and organ-at-risk delineation, inverse treatment planning, independent dose verification, patient positioning with image guidance, and the delivery of highly conformal dose distributions over days or weeks. Each link in that chain offers an opportunity for error, and because radiation effects are cumulative and frequently irreversible, a small systematic mistake can harm a patient in ways that may not become apparent for months (Ford

& Evans, 2018). For this reason, radiation oncology has long maintained one of the most rigorous quality management cultures in medicine. Periodic machine checks are codified in documents such as AAPM Task Group 142 (Klein et al., 2009); patient-specific verification of modulated plans is guided by Task Group 218 (Miften et al., 2018); and physics plan and chart review is structured by Task Group 275 (Ford et al., 2020). Beyond prescriptive checklists, Task Group 100 moved the field toward a risk-based philosophy, encouraging departments to analyse their own processes with failure modes and effects analysis (FMEA) and to direct effort where risk is greatest (Huq et al., 2016). Despite this maturity, the conventional QM model faces growing strain. Measurement-based patient-specific QA consumes considerable physicist and linac time, yet its ability to detect clinically relevant errors has been questioned. Manual chart reviews depend on individual vigilance and are vulnerable to fatigue and inattention blindness. Meanwhile, techniques such as stereotactic body radiotherapy and online adaptive radiotherapy compress the time available for verification from days to minutes. The volume of data generated by modern linear accelerators, imaging systems, and oncology information systems now far exceeds what human reviewers can meaningfully interrogate. Artificial intelligence, and in particular machine learning, has emerged as a potential answer to these pressures. ML models can learn patterns from historical QA measurements, treatment plans, and delivery logs; they can predict outcomes of verification measurements before they are performed, flag anomalous plans, and forecast machine faults before they breach tolerance (Kalet et al., 2020; Chan et al., 2020). At the same time, AI introduces new failure modes of its own: opaque reasoning, sensitivity to data shifts, and the risk that clinicians defer to automated outputs uncritically (Parasuraman & Manzey, 2010). This review addresses three questions. First, where along the RT process has AI demonstrated credible value for quality management? Second, what technical, human, ethical, and regulatory barriers stand between promising research and safe routine use? Third, how might a health system undergoing rapid transformation, such as that of the Kingdom of Saudi Arabia under Vision 2030, integrate AI into RT quality management responsibly? To answer the third question, the paper proposes an original governance and implementation framework, **PRISM**.

2. REVIEW METHODOLOGY

This article adopts a structured narrative review design, which is appropriate for a heterogeneous and fast-moving field where applications, study designs, and outcome measures vary too widely for meta-analysis. Literature was identified through searches of PubMed/MEDLINE, Scopus, IEEE Xplore, and Google Scholar using combinations of the terms “artificial intelligence”, “machine learning”, “deep learning”, “radiotherapy”, “radiation oncology”, “quality assurance”, “quality control”, “patient-specific QA”, “plan review”, “auto-segmentation”, and “incident learning”. Reference lists of key reviews were hand-searched to identify additional sources. Sources were eligible if they (a) described the development, validation, or clinical implementation of an AI or ML method for a quality-related task in external beam radiotherapy or brachytherapy, (b) provided professional guidance on QA of AI tools in RT, or (c) set out regulatory or ethical requirements for AI-based medical software. Priority was given to peer-reviewed studies, AAPM and ESTRO reports, and publications from regulators including the Saudi Food and Drug Authority (SFDA), the Saudi Data and AI Authority (SDAIA), and the World Health Organization (WHO). The evidence was synthesised thematically and interpreted through the theoretical lens described in Section 3. As a narrative rather than systematic review, this work does not claim exhaustive coverage, and selection bias cannot be excluded. The intention is to provide a critical, practice-oriented synthesis that supports decision-makers, medical physicists, radiation oncologists, and quality leaders.

3. THEORETICAL FOUNDATIONS

3.1 Donabedian’s Structure–Process–Outcome Model

Donabedian’s (1988) model holds that quality of care can be evaluated through structure (resources, equipment, staffing, and organisation), process (what is done in giving and receiving care), and outcome (the effect of care on health status). In RT quality management, structure includes linacs, imaging and dosimetry equipment, and qualified medical physicists; process covers commissioning, periodic QA, planning, and verification; outcome encompasses dosimetric accuracy, tumour control, toxicity, and freedom from incidents. AI touches all three levels. It becomes part of the structure as a new software asset requiring commissioning, it reshapes the process by automating or augmenting checks, and it can influence outcomes by catching errors earlier or, if poorly implemented, by introducing new ones.

3.2 Reason’s Swiss Cheese Model

Reason’s (2000) Swiss cheese model conceptualises safety as a series of imperfect barriers, each with latent holes. Harm occurs when the holes momentarily align. RT QM is, in effect, a deliberate stacking of such barriers: physician peer review, physics plan checks, independent monitor-unit calculation, pre-treatment measurement, therapist time-outs, and image guidance. The model offers a useful framing for AI: an ML-based checker is best understood as an additional slice whose holes are positioned

differently from those of human reviewers. Its value lies precisely in that difference. However, if AI replaces a human slice rather than adding to it, or if humans relax their vigilance because an AI slice is present, the overall number of effective barriers may fall.

3.3 Risk-Based Quality Management (AAPM TG-100)

Task Group 100 recommended process mapping, FMEA, and fault-tree analysis so that departments can rank failure modes by occurrence, severity, and lack of detectability, and allocate QA resources accordingly (Huq et al., 2016). This philosophy is highly compatible with AI. Predictive models can make risk estimates dynamic rather than static, updating the probability that a specific plan or machine state will fail. Conversely, TG-100 methods should be applied to AI tools themselves, identifying how a model might fail silently and what detection controls are needed.

3.4 Human–Automation Interaction

Research in human factors shows that people working with automated decision aids are susceptible to automation complacency and automation bias: they monitor less carefully and accept automated recommendations even when contradictory evidence is available (Parasuraman & Manzey, 2010). These effects are strongest when automation is highly reliable, which is paradoxically the condition that successful AI tools aim to achieve. Any framework for AI in RT QM must therefore design for sustained human engagement, not merely for algorithmic accuracy.

4. THE CURRENT LANDSCAPE OF RADIATION THERAPY QUALITY MANAGEMENT

Before examining AI applications, it is useful to outline the principal domains of RT QM and the limitations that motivate interest in automation. Table 1 summarises these domains.

Table 1. Principal domains of radiation therapy quality management and limitations of conventional approaches

QM domain	Conventional approach	Key limitations
Machine QA	Daily, monthly, and annual checks against tolerances (TG-142); output, symmetry, MLC positioning, imaging alignment	Periodic sampling misses drift between checks; reactive rather than predictive; time-consuming
Patient-specific QA	Phantom or detector-array measurement with gamma analysis (TG-218)	Consumes linac and physicist time; limited sensitivity to clinically relevant errors; delays start of treatment
Treatment plan review	Physician and physicist review of dose distribution, dose-volume metrics, and plan parameters	Inter-reviewer variability; dependent on experience; plan quality benchmarks rarely patient-specific
Chart and data transfer checks	Manual checklists across planning and record-and-verify systems (TG-275)	Fatigue, inattention blindness, and checklist fatigue; many items low-yield
Contour review	Peer review of target and organ-at-risk delineation	High inter-observer variability; time pressure; limited audit capacity
Delivery monitoring	Image guidance, in-vivo dosimetry, log-file review	Large data volumes rarely analysed systematically
Incident learning	Voluntary reporting (e.g., RO-ILS, SAFRON) and root-cause analysis	Under-reporting; free-text reports difficult to analyse at scale

Two themes cut across these domains. First, much of current QA is **measurement-heavy and sample-based**, meaning it verifies a subset of parameters at discrete time points. Second, the **detection power** of many checks is modest. Studies of

incident learning systems suggest that a significant fraction of errors reaching patients passed through several existing checks undetected, and that physics plan review and therapist checks are among the most effective barriers when performed well (Ford et al., 2020; Hoopes et al., 2015). These observations suggest that the greatest gains may come not from doing more of the same checks but from making checks smarter and more targeted.

5. OPPORTUNITIES: AI APPLICATIONS ACROSS THE QUALITY CHAIN

5.1 Predictive Machine QA and Maintenance

Linear accelerators generate continuous streams of performance data through daily QA devices, machine log files, and electronic portal imaging devices (EPIDs). ML approaches, including time-series models and artificial neural networks, have been used to forecast trends in beam output and symmetry, allowing physicists to intervene before a parameter drifts beyond tolerance (Li & Chan, 2017). El Naqa et al. (2019) demonstrated that unsupervised learning on EPID images could automatically identify anomalous machine behaviour, offering a proof of principle for automated machine QA. By shifting from fixed-interval checks to condition-based monitoring, predictive models can reduce unplanned downtime, a particularly important benefit in settings where a single linac outage displaces dozens of patients per day.

5.2 Virtual Patient-Specific QA

Virtual QA is perhaps the most extensively studied application. Valdes et al. (2016) developed a Poisson regression model using plan complexity features to predict gamma passing rates of IMRT plans, and subsequent work applied deep convolutional neural networks directly to fluence maps, achieving comparable or better accuracy without hand-crafted features (Interian et al., 2018). The clinical vision is a risk-stratified workflow: plans predicted to pass comfortably proceed with reduced measurement, while plans predicted to be marginal receive full measurement or are re-optimised before QA. Two caveats temper this promise. Gamma passing rate is itself an imperfect surrogate for clinical dose accuracy, so predicting it reproduces the surrogate's weaknesses. Furthermore, models trained on plans from one planning system, treatment machine, or detector may not transfer to another. Virtual QA is therefore best framed as a **triage tool** complementing, rather than eliminating, measurement.

"The goal of AI in radiotherapy QA is not to measure less for its own sake, but to measure wisely — concentrating expert attention where the risk truly lies."

5.3 Automated Plan and Chart Review

Physics plan review is among the most effective error-detection steps in RT, yet it is cognitively demanding. Rule-based automation already verifies prescriptions, energies, and data consistency between systems; ML extends this by learning what a "normal" plan looks like. Luk et al. (2019) characterised a Bayesian network that estimates the probability that combinations of plan parameters are consistent with historical practice, flagging outliers for human review. Knowledge-based planning models, originally developed to improve plan quality, also function as QA tools: by predicting achievable organ-at-risk doses from patient anatomy, they identify plans that are suboptimal even when they meet generic protocol constraints (Moore et al., 2011). Deep learning dose-prediction models extend this to full three-dimensional dose distributions (Nguyen et al., 2019).

5.4 Auto-Segmentation and Contour Quality Control

Deep learning auto-segmentation is now widely deployed commercially and has reduced contouring time considerably (Cardenas et al., 2019). From a QM perspective it plays two roles. It is a source of consistency, reducing inter-observer variability, but it is also a new source of systematic error that must itself be checked. Complementary AI tools have been developed to detect contouring errors automatically; Rhee et al. (2019) used convolutional neural networks to flag erroneous contours by comparing clinical structures against independently generated predictions. The question of what constitutes a "clinically acceptable" auto-contour remains an active area of debate, and dosimetric impact rather than geometric overlap alone is increasingly recommended as the benchmark (Baroudi et al., 2023).

5.5 Treatment Delivery Monitoring and In-Vivo Verification

EPID-based transit dosimetry, surface-guided radiotherapy, and machine log files provide rich records of what was actually delivered. ML models can compare expected and measured signals in real time, distinguishing clinically meaningful deviations (for example, a wrong patient, gross anatomical change, or MLC malfunction) from benign noise. In stereotactic treatments, where margins are small and doses per fraction high, such monitoring is particularly valuable (Mancosu et al., 2022). In adaptive workflows, AI may assist with rapid verification of re-optimised plans when time for conventional QA is measured in minutes.

5.6 Natural Language Processing of Incident Reports

Incident learning systems such as RO-ILS collect thousands of narrative reports, but their free-text format limits systematic analysis (Hoopes et al., 2015). Natural language processing (NLP) can classify reports by process step, severity, and contributing factors, cluster similar events across institutions, and surface emerging risks that individual safety committees might miss. Large language models, which have advanced rapidly since 2023, offer further potential to summarise root-cause analyses and suggest corrective actions, although their outputs require careful validation given their tendency to produce plausible but incorrect statements.

Table 2. Summary of AI applications for RT quality management

Application	Typical AI methods	Representative evidence	Maturity
Predictive machine QA	Time-series models, ANN, anomaly detection	Li & Chan (2017); El Naqa et al. (2019)	Research / early clinical
Virtual patient-specific QA	Regression, CNN on fluence maps	Valdes et al. (2016); Interian et al. (2018)	Early clinical
Automated plan review	Bayesian networks, outlier detection	Luk et al. (2019)	Early clinical
Plan quality benchmarking	Knowledge-based planning, dose prediction	Moore et al. (2011); Nguyen et al. (2019)	Clinical (commercial)
Contour QA	CNN-based error detection	Rhee et al. (2019); Baroudi et al. (2023)	Research / early clinical
Delivery monitoring	Signal comparison, classification of EPID/log data	Mancosu et al. (2022)	Research / early clinical
Incident learning analytics	NLP, topic modelling, LLMs	Hoopes et al. (2015); Ford & Evans (2018)	Research

6. CHALLENGES AND RISKS

The literature is consistent that the principal obstacles to AI in RT QM are less about algorithmic capability and more about data, validation, human factors, and governance (Kelly et al., 2019; Vandewinckele et al., 2020). The key challenges are discussed below and summarised in Table 3.

6.1 Data Availability, Quality, and Class Imbalance

Supervised learning requires labelled examples, yet in a well-functioning RT department genuine errors are rare. QA datasets are therefore severely imbalanced: models see thousands of passing plans and very few failures. This makes it difficult to train and, crucially, to validate sensitivity to the very events QA exists to catch. Many studies compensate by simulating errors, which may not represent the diversity of real-world failures. Data are also fragmented across treatment planning, record-and-verify, and QA systems that use different formats and identifiers.

6.2 Generalisability and Model Drift

Models trained at one institution often degrade when applied elsewhere because of differences in equipment, planning practice, dose calculation algorithms, and patient populations. Even within a single department, performance may decline over time as clinical protocols, software versions, or machines change, a phenomenon known as model or data drift. Unlike a linac, whose drift is measured by routine QA, an AI model can degrade silently unless monitoring is designed deliberately (Claessens et al., 2022).

6.3 Explainability and Trust

Deep learning models are often opaque. A physicist told that a plan is “high risk” without an explanation cannot easily decide whether to act, and may either ignore a valid warning or accept a spurious one. Explainability techniques such as feature attribution and saliency maps can help but are themselves imperfect. For QM purposes, explanations should be actionable, pointing to the plan parameter, structure, or machine component driving the concern.

6.4 Automation Bias, Complacency, and Deskilling

As discussed in Section 3.4, the more reliable an AI tool becomes, the greater the risk that human reviewers disengage (Parasuraman & Manzey, 2010). Over time, reduced hands-on practice may erode the expertise needed to recognise when the AI is wrong, which is a form of deskilling with implications for training future medical physicists and dosimetrists. In Reason’s terms, the human slice becomes thinner precisely because the AI slice is present.

6.5 Quality Assurance of the AI Itself

A distinctive challenge is that AI tools used for QA must themselves be quality-assured. Professional recommendations advocate a lifecycle approach: acceptance testing and commissioning using local data, definition of clinically meaningful performance metrics, end-to-end testing within the clinical workflow, and ongoing periodic QA analogous to that applied to any other clinical system (Vandewinckele et al., 2020; Claessens et al., 2022). Few departments yet have standardised procedures, test datasets, or staff with the combined physics and data-science skills required.

6.6 Regulatory, Legal, and Ethical Considerations

AI software intended for diagnostic or therapeutic decision support is generally regulated as a medical device. Regulators have issued guidance on good machine learning practice, including requirements for representative data, transparency, and management of post-market changes (U.S. Food and Drug Administration et al., 2021). Questions of liability remain unsettled: if an AI tool fails to flag an error that a human then misses, responsibility may be shared ambiguously among vendor, institution, and clinician. Ethical principles emphasised by the WHO (2021) include human autonomy, safety, transparency, accountability, inclusiveness, and sustainability.

6.7 Cybersecurity, Privacy, and Infrastructure

AI tools frequently require data integration across networked systems and, in some cases, cloud computing. This expands the attack surface of radiation oncology departments, which have already been targets of ransomware incidents internationally. Privacy obligations apply to the patient data used for training and inference, and local legislation governs cross-border data transfer.

6.8 Workforce, Cost, and Equity

Implementation requires investment in software, computing infrastructure, and personnel with hybrid expertise. Smaller centres may lack these resources, raising the risk that AI widens rather than narrows quality gaps between institutions. At the same time, well-designed shared models could offer smaller centres access to expertise they could not otherwise afford.

Table 3. Key challenges and mitigation strategies

Challenge	Risk to quality and safety	Mitigation strategies
Data scarcity and imbalance	Unknown sensitivity to rare but serious errors	Multi-institutional datasets; realistic error simulation; federated learning
Poor generalisability / drift	Silent performance degradation	Local commissioning; drift monitoring dashboards; revalidation after system changes
Limited explainability	Warnings ignored or blindly accepted	Actionable explanations; uncertainty estimates; user training
Automation bias and deskilling	Thinner human defence layer	Keep AI as an additional layer; periodic blinded audits; competency maintenance

Challenge	Risk to quality and safety	Mitigation strategies
QA of AI tools	Unverified tool used in safety-critical checks	Lifecycle QA programme with acceptance, commissioning, and periodic tests
Regulatory and liability uncertainty	Unclear accountability after incidents	SFDA-registered tools; documented roles; governance committee sign-off
Cybersecurity and privacy	Data breaches; service disruption	PDPL compliance; network segmentation; vendor security assessment
Workforce and cost	Unequal access; failed implementations	Hybrid physics–data science training; national shared services

7. THE SAUDI ARABIAN CONTEXT

The Kingdom of Saudi Arabia offers a distinctive setting for AI integration in RT quality management. Under Vision 2030 and its Health Sector Transformation Program, the health system is undergoing restructuring toward integrated care, value-based purchasing, and digital health, while cancer services are expanding to meet the needs of a growing and ageing population. Tertiary centres such as King Faisal Specialist Hospital and Research Centre, King Abdulaziz Medical City, and other Ministry of Health and university hospitals operate advanced RT programmes, and new radiotherapy centres are being commissioned in several regions.

The national regulatory ecosystem is relatively well positioned for AI governance:

- ◆ **Saudi Food and Drug Authority (SFDA):** regulates medical devices, including software, and has issued dedicated guidance on AI- and ML-based medical devices (MDS-G010), covering clinical evaluation, risk classification, and management of changes after market authorisation (SFDA, 2022).
- ◆ **Nuclear and Radiological Regulatory Commission (NRRC):** oversees radiation protection and the licensing of facilities using ionising radiation; AI tools that influence dose verification fall within the broader radiation safety programme under its oversight.
- ◆ **Saudi Data and AI Authority (SDAIA):** has published AI Ethics Principles emphasising fairness, privacy, transparency, accountability, and human-centricity, and oversees implementation of the Personal Data Protection Law (PDPL) (SDAIA, 2023).
- ◆ **Saudi Commission for Health Specialties (SCFHS):** licenses medical physicists, radiation therapists, and radiation oncologists and can embed AI literacy and data science competencies within training and continuing professional development.
- ◆ **Saudi Central Board for Accreditation of Healthcare Institutions (CBAHI):** sets hospital accreditation standards, including those for quality, patient safety, and medical equipment management, into which AI governance requirements can be integrated.
- ◆ **King Abdullah International Medical Research Center (KAIMRC)** and university research centres: provide research capacity for local validation studies and multi-centre data collaborations.

Several opportunities follow. The concentration of RT services in a relatively small number of large public institutions favours the creation of **national reference datasets** and shared knowledge-based models, possibly using federated learning so that data remain within each institution in compliance with the PDPL. A **national RT incident learning system**, analysed with NLP, could accelerate system-wide learning. Local validation is especially important because patient anatomy, disease presentation, and practice patterns may differ from the Western populations on which most commercial models were trained. Finally, AI-enabled efficiency gains can support Vision 2030 goals of improving access and reducing waiting times without compromising safety.

Challenges also exist. The pool of medical physicists with data-science expertise is limited, reliance on international vendors raises questions about data sovereignty, and interoperability across hospital information systems remains uneven. Addressing these requires coordinated action across regulators, providers, and academic institutions.

8. THE PRISM FRAMEWORK FOR AI-ENABLED RT QUALITY MANAGEMENT

Drawing on the evidence and theory reviewed above, this paper proposes the **PRISM framework**, a five-pillar model to guide the safe and effective integration of AI into RT quality management. Like a prism that separates light into distinct components, the framework separates the complex task of AI adoption into distinct but complementary domains. It maps onto Donabedian's structure, process, and outcome dimensions and treats AI as an additional barrier within Reason's layered model of defence.

Table 4. The PRISM framework

Pillar	Core principle	Key actions	Donabedian link
P — Proactive risk intelligence	Use AI to anticipate failures rather than react to them	Deploy predictive machine QA and virtual QA for risk-stratified checks; integrate with TG-100 FMEA	Process
R — Robust validation	No AI tool enters clinical use without local evidence	Acceptance testing, commissioning on local data, error-injection testing, dosimetric impact assessment	Structure
I — Integrated human–AI oversight	AI adds a defensive layer; it does not replace one	Define human review points; design interfaces showing uncertainty; train against automation bias; blinded audits	Process
S — Standards-aligned governance	Align with national and international requirements	SFDA registration; NRRC and CBAHI alignment; SDAIA ethics and PDPL compliance; AI governance committee	Structure
M — Monitoring for lifecycle learning	Performance is tracked continuously and fed back	Drift dashboards; periodic QA of AI; link to incident learning; outcome and efficiency metrics	Outcome

8.1 Proactive Risk Intelligence

The first pillar reorients QA from uniform checking toward risk-informed allocation. AI-derived risk scores for plans and machines should be embedded within the department's TG-100 analysis so that high-risk cases receive intensified scrutiny, while routine low-risk cases follow streamlined pathways. The objective is not merely efficiency but improved detection where it matters most.

8.2 Robust Validation

Every AI tool should undergo a structured commissioning process analogous to that required for a new dose calculation algorithm. This includes defining intended use, assembling a representative local test set, deliberately injecting known errors to establish sensitivity, and assessing performance in terms of clinical consequences rather than abstract accuracy metrics alone (Claessens et al., 2022). Commissioning results should be documented and approved by a qualified medical physicist.

8.3 Integrated Human–AI Oversight

The third pillar operationalises the principle that AI is an additional slice of cheese. Workflows should specify which decisions remain human, how AI outputs are displayed, and how disagreements are resolved. Periodic blinded audits, in which reviewers assess plans without AI assistance, help maintain competence and reveal complacency. Training programmes should explicitly address automation bias.

“An AI tool that is never questioned has quietly stopped being a safety barrier and become a single point of failure.”

8.4 Standards-Aligned Governance

Institutions should establish a multidisciplinary AI governance committee including radiation oncologists, medical physicists, radiation therapists, IT and cybersecurity specialists, quality and patient safety officers, and legal or ethics representatives. The committee approves tools for clinical use, verifies regulatory status with the SFDA, ensures compliance with the PDPL and SDAIA principles, and aligns documentation with CBAHI standards and NRRC radiation safety requirements.

8.5 Monitoring for Lifecycle Learning

Finally, AI tools require continuous monitoring. Key performance indicators might include agreement between AI predictions and measurements, false-positive and false-negative rates, time saved, and the proportion of flagged items that led to plan changes. Any software upgrade, new machine, or protocol change should trigger revalidation. Monitoring data should feed into the incident learning system and inform periodic reviews by the governance committee.

9. PHASED IMPLEMENTATION ROADMAP

Translating PRISM into practice benefits from a staged approach that builds capability and evidence progressively. Table 5 outlines a suggested roadmap for RT departments and health system leaders.

Table 5. Suggested phased roadmap for AI integration in RT quality management

Phase	Focus	Example activities	Success indicators
1. Readiness (0–6 months)	Governance and data foundations	Form AI committee; audit data infrastructure; map QM process with FMEA; staff AI literacy training	Committee charter approved; data inventory complete
2. Pilot (6–18 months)	Low-risk, high-value use cases in shadow mode	Run virtual QA and automated plan checks in parallel with existing QA without changing decisions	Validated sensitivity and specificity on local data
3. Integration (18–36 months)	Risk-stratified clinical use	Adopt reduced-measurement pathways for low-risk plans; implement drift monitoring	No increase in incidents; reduced QA time; user acceptance
4. Scale and network (36+ months)	System-wide learning	Multi-centre federated models; national incident analytics; shared benchmarks	Cross-institution benchmarking; published outcomes

A key feature of this roadmap is the **shadow mode** phase. Running AI alongside existing QA, without allowing it to influence decisions, generates local evidence of performance and builds staff familiarity before any reduction in conventional checks. This mirrors the conservative, evidence-driven culture of medical physics and reduces the risk of premature reliance.

10. FUTURE DIRECTIONS AND RESEARCH AGENDA

Several developments are likely to shape the next phase of AI in RT quality management.

- ◆ **Outcome-linked QA metrics:** moving beyond gamma passing rates toward models that predict clinically meaningful dose errors and, ultimately, patient outcomes.
- ◆ **Federated and privacy-preserving learning:** enabling multi-institutional model development without centralising patient data, which is particularly relevant under the PDPL.
- ◆ **QA for online adaptive radiotherapy:** developing fast, reliable AI-based verification suitable for workflows in which plans are re-optimised while the patient is on the couch.

- ◆ **Uncertainty quantification:** equipping models to report their own confidence so that low-confidence cases are routed automatically to human experts.
- ◆ **Large language models and agents:** exploring careful use of generative AI for protocol checking, incident analysis, and documentation, with rigorous safeguards against fabricated outputs.
- ◆ **Human factors research:** conducting prospective studies of how AI changes reviewer behaviour, error detection, and workload in real clinical settings.
- ◆ **Health economic evaluation:** quantifying the costs and benefits of AI-enabled QM to support investment decisions within value-based healthcare models.

For Saudi Arabia specifically, priorities include local validation studies of commercial AI tools, development of national reference datasets, and integration of AI competencies into SCFHS training programmes for medical physics and radiation therapy. Such efforts would position the Kingdom not only as an adopter but as a contributor to international evidence.

11. CONCLUSION

Artificial intelligence offers radiation therapy quality management a genuine opportunity to move from periodic, uniform, and largely reactive checking toward continuous, risk-informed, and predictive oversight. The evidence reviewed here shows credible progress in predictive machine QA, virtual patient-specific QA, automated plan review, contour quality control, delivery monitoring, and incident analytics. Yet the same evidence highlights serious challenges: scarce and imbalanced data, limited generalisability, opaque models, automation bias, and the pressing need to quality-assure AI tools with the same rigour applied to any other element of the treatment chain.

The central message of this review is that AI should strengthen, not replace, the layered defences that make radiotherapy safe. The PRISM framework, grounded in Donabedian's model, Reason's Swiss cheese model, and TG-100 risk-based methods, provides a structured approach to achieving this. In the Saudi context, strong regulatory foundations through the SFDA, NRRC, SDAIA, SCFHS, and CBAHI, combined with the transformation momentum of Vision 2030, create favourable conditions for responsible adoption. With deliberate governance, local validation, and sustained human engagement, AI can help ensure that every patient receives radiation therapy that is not only more efficient but demonstrably safer.

REFERENCES

- [1] Baroudi, H., Brock, K. K., Cao, W., et al. (2023). Automated contouring and planning in radiation therapy: What is 'clinically acceptable'? *Diagnostics*, 13(4), 667.
- [2] Cardenas, C. E., Yang, J., Anderson, B. M., Court, L. E., & Brock, K. B. (2019). Advances in auto-segmentation. *Seminars in Radiation Oncology*, 29(3), 185–197.
- [3] Chan, M. F., Witzum, A., & Valdes, G. (2020). Integration of AI and machine learning in radiotherapy QA. *Frontiers in Artificial Intelligence*, 3, 577620.
- [4] Claessens, M., Oria, C. S., Brouwer, C. L., et al. (2022). Quality assurance for AI-based applications in radiation therapy. *Seminars in Radiation Oncology*, 32(4), 421–431.
- [5] Donabedian, A. (1988). The quality of care: How can it be assessed? *JAMA*, 260(12), 1743–1748.
- [6] El Naqa, I., Irrer, J., Ritter, T. A., et al. (2019). Machine learning for automated quality assurance in radiotherapy: A proof of principle using EPID data description. *Medical Physics*, 46(4), 1914–1921.
- [7] Ford, E. C., & Evans, S. B. (2018). Incident learning in radiation oncology: A review. *Medical Physics*, 45(5), e100–e119.
- [8] Ford, E., Conroy, L., Dong, L., et al. (2020). Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275. *Medical Physics*, 47(6), e236–e272.
- [9] Hoopes, D. J., Dicker, A. P., Eads, N. L., et al. (2015). RO-ILS: Radiation Oncology Incident Learning System: A report from the first year of experience. *Practical Radiation Oncology*, 5(5), 312–318.
- [10] Huq, M. S., Fraass, B. A., Dunscombe, P. B., et al. (2016). The report of Task Group 100 of the AAPM: Application of risk analysis methods to radiation therapy quality management. *Medical Physics*, 43(7), 4209–4262.

- [11] Interian, Y., Rideout, V., Kearney, V. P., et al. (2018). Deep nets vs expert designed features in medical physics: An IMRT QA case study. *Medical Physics*, 45(6), 2672–2680.
- [12] Kalet, A. M., Luk, S. M. H., & Phillips, M. H. (2020). Radiation therapy quality assurance tasks and tools: The many roles of machine learning. *Medical Physics*, 47(5), e168–e177.
- [13] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 195.
- [14] Klein, E. E., Hanley, J., Bayouth, J., et al. (2009). Task Group 142 report: Quality assurance of medical accelerators. *Medical Physics*, 36(9), 4197–4212.
- [15] Li, Q., & Chan, M. F. (2017). Predictive time-series modeling using artificial neural networks for Linac beam symmetry: An empirical study. *Annals of the New York Academy of Sciences*, 1387(1), 84–94.
- [16] Luk, S. M. H., Meyer, J., Young, L. A., Cao, N., Ford, E. C., Phillips, M. H., & Kalet, A. M. (2019). Characterization of a Bayesian network-based radiotherapy plan verification model. *Medical Physics*, 46(5), 2006–2014.
- [17] Mancosu, P., Lambri, N., Castiglioni, I., et al. (2022). Applications of artificial intelligence in stereotactic body radiation therapy. *Physics in Medicine & Biology*, 67(16), 16TR01.
- [18] Miften, M., Olch, A., Mihailidis, D., et al. (2018). Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Medical Physics*, 45(4), e53–e83.
- [19] Moore, K. L., Brame, R. S., Low, D. A., & Mutic, S. (2011). Experience-based quality control of clinical intensity-modulated radiotherapy planning. *International Journal of Radiation Oncology, Biology, Physics*, 81(2), 545–551.
- [20] Nguyen, D., Long, T., Jia, X., Lu, W., Gu, X., Iqbal, Z., & Jiang, S. (2019). A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. *Scientific Reports*, 9, 1076.
- [21] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.
- [22] Reason, J. (2000). Human error: Models and management. *BMJ*, 320(7237), 768–770.
- [23] Rhee, D. J., Cardenas, C. E., Elhalawani, H., et al. (2019). Automatic detection of contouring errors using convolutional neural networks. *Medical Physics*, 46(11), 5086–5097.
- [24] Saudi Data and AI Authority. (2023). *AI ethics principles*. SDAIA.
- [25] Saudi Food and Drug Authority. (2022). *Guidance on artificial intelligence (AI) and machine learning (ML) technologies based medical devices (MDS-G010)*. SFDA.
- [26] S. Food and Drug Administration, Health Canada, & Medicines and Healthcare products Regulatory Agency. (2021). *Good machine learning practice for medical device development: Guiding principles*.
- [27] Valdes, G., Scheuermann, R., Hung, C. Y., Olszanski, A., Bellerive, M., & Solberg, T. D. (2016). A mathematical framework for virtual IMRT QA using machine learning. *Medical Physics*, 43(7), 4323–4334.
- [28] Vandewinckele, L., Claessens, M., Dinkla, A., Brouwer, C., Crijns, W., Verellen, D., & van Elmpt, W. (2020). Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. *Radiotherapy and Oncology*, 153, 55–66.
- [29] World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. WHO.