

Review

Fluoroscopy-Guided Motion Management in Particle Therapy: Evolution, Challenges, and AI-Enabled Opportunities

Feifei Li *, Keith M. Furutani  and Chris J. Beltran

Department of Radiation Oncology, Mayo Clinic, Jacksonville, FL 32224, USA; furutani.keith@mayo.edu (K.M.F.); beltran.chris@mayo.edu (C.J.B.)

* Correspondence: li.feifei2@mayo.edu

Simple Summary

Particle therapy delivers radiation that stops sharply at a chosen depth, sparing healthy tissue near the tumor. This precision can be undermined when the target moves during respiration, as in many lung, liver, and pancreatic tumors, where small displacements can cause underdosing of the tumor or unintended dose to adjacent organs. Fluoroscopy enables real-time imaging of the target during treatment and is therefore a promising imaging modality for motion-managed particle therapy. This review traces the evolution of fluoroscopy hardware from image intensifiers to modern flat-panel detectors integrated with proton therapy units, summarizes vendor-supported fluoroscopy-guided systems, and examines why reliable tracking still relies on implanted fiducial markers. We then survey emerging AI-based methods that could lead to marker-less tumor tracking using on-treatment fluoroscopy. Technical and clinical challenges are discussed.

Abstract

The sharp dose gradients that underpin the dosimetric advantage of particle therapy over photon therapy can be undermined by the interplay effects due to intra-fraction motion in modern pencil beam scanning systems. Fluoroscopy-Guided Particle Therapy (FGPT) offers a promising path to improved motion management through real-time tracking of tumors or surrogate signals. The advent of flat-panel detector (FPD)-based technology has enabled tighter integration of fluoroscopy/fluorography into treatment units and accelerated clinical adoption and research, with commercial systems such as Hitachi's Real-time Gated Particle Therapy (RGPT) now available. However, the need for implanted fiducial markers, with the associated invasiveness and risk of complications, limits the utility of RGPT to a few anatomic sites in selected patients. The full potential of FGPT, therefore, depends on reliable marker-less tumor tracking, which remains challenging because soft-tissue targets are obscured by overlapping anatomy along the X-ray path, leading to reduced reliability of traditional image-registration algorithms in the projection domain. Recent advances in deep learning and AI-driven image registration have renewed hope for overcoming these barriers, enabling real-time marker-less tracking for particle therapy. This review outlines the evolution of fluoroscopy technology from image intensifier (II) to FPD-based systems, summarizes historical and recent vendor-supported FGPT strategies, and surveys emerging AI-based algorithms in the literature. A general review of machine learning-based image registration is provided, challenges in generalizability and interpretability are highlighted, and potential paths toward reliable, clinically deployable FGPT are discussed.

Keywords: fluoroscopy; particle therapy; machine learning



Academic Editor: Emilio Quaia

Received: 13 March 2026

Revised: 30 April 2026

Accepted: 1 May 2026

Published: 9 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The clinical application of particle therapy traces its origins back to Robert Wilson's seminal 1946 paper [1]. In the aftermath of World War II, physicist Robert R. Wilson left the Manhattan Project to join Harvard University, determined to seek peaceful applications of atomic energy. Wilson proposed a therapeutic usage of protons based on their unique physical property known as the Bragg peak—a phenomenon where charged particles deposit the majority of their energy at a precise, controllable depth before stopping completely. He argued that by manipulating this peak, clinicians could deliver high doses of radiation to deep-seated tumors while sparing the healthy tissue surrounding them—a revolutionary concept that effectively birthed the field of particle therapy. Wilson's vision found its first clinical realization at his own institution: in 1975, Gragoudas and colleagues at Massachusetts Eye and Ear used the Harvard Cyclotron Laboratory to deliver the first proton-beam treatment for choroidal melanoma, an eye-preserving alternative to enucleation that remains a standard of care today [2]. The technique exploits the precision of the Bragg peak in its most demanding setting: curative doses must be deposited millimeters from the optic nerve, retina, and lens. The same approach now extends to ocular surface tumors at the dedicated 60–70 MeV proton beamlines that have continued the Harvard tradition, including the Centre Antoine-Lacassagne in Nice and the CATANA Centre at INFN-LNS in Catania, where 48–60 Gy RBE is delivered to conjunctival melanoma and conjunctival squamous cell carcinoma in only four daily fractions, with a 90-to-10% dose fall-off of approximately 1 mm at 3 cm depth [3,4]. Ocular proton therapy thus represents both the historical proof and the geometric extreme of the precision agenda that motivates particle therapy more broadly—an agenda that, as the next sections detail, becomes substantially harder to realize when the target lies deep within a moving thoracic or abdominal anatomy. As of May 2026, there are 128 particle therapy facilities in clinical operations worldwide [5].

The steep dose gradient of particle therapy is a double-edged sword. It enables superior normal tissue sparing compared to conventional photon radiotherapy, yet it also makes particle therapy inherently more sensitive to geometric uncertainties, especially those arising from motion. Unlike photon therapy, where radiation passes completely through the patient, protons and heavy ions stop at a specific depth that is highly sensitive to the density and composition of the materials traversed. For a typical photon treatment beam, a 1.0 cm depth variation beyond the depth of maximum dose leads to approximately 3% dose difference. In contrast, for a proton beam, the same depth variation near the distal edge can result in up to 90% dose error [6]. Any motion that alters the radiological path length can shift the stopping point, potentially causing the beam to underdose the tumor or overdose healthy tissue immediately beyond it. This is especially critical in the lung, where solid tumor motion leads to significant alteration of water equivalent path length (WEPL). Lung tumors can move up to 30 mm [7], and respiratory motion can lead to WEPL change of 20.4 mm near the heart [8]. The benefit of particle therapy diminishes without precise patient and tumor positioning.

The impact of organ motion on radiation therapy has been the subject of numerous reviews [9–14]. In the context of particle therapy, dosimetric uncertainties arise from the combination of three primary sources of motion. Inter-fraction motion: discrepancies in patient positioning and anatomy relative to the baseline planning CT that occur between treatment fractions; typical examples include patient weight fluctuations and tumor progression or regression. Intra-fraction motion: physical movement of the patient's body or internal organs during the delivery of a single treatment fraction; typical examples are respiration, digestion, and organ filling. Beam scanning motion: the temporal movement of the particle beam relative to the tumor volume; this is intrinsic to delivery mechanisms

such as modern pencil beam scanning (PBS) [15], which requires a finite amount of time to scan through the entire target, both laterally and depth-wise. Inter-fraction motion can be mitigated through improved immobilization, precise positioning systems, and adaptive re-planning. However, intra-fraction motion remains a significant challenge for radiation therapy. Respiratory motion is a major concern for therapy delivered to the thoracic and abdominal regions [11,16]. Because particle therapy dosimetry is uniquely sensitive to positioning and range uncertainties, robust motion management is even more critical to particle therapy as compared to conventional X-ray therapy [12,16].

Motion management for particle therapy depends heavily on the mechanism of beam delivery. In early proton delivery systems, beams were delivered via passive scattering (PS) [17]. In this method, the narrow proton beam from the accelerator is broadened laterally by a scattering system and shaped in depth using ridge filters or modulator wheels. The lateral spread occurs on a timescale that is effectively instantaneous, while spreading along the distal direction occurs within one modulator revolution (typically < 0.1 s) [16]. Because the beam is broad and continuous, motion management was historically handled through robust planning margins. Clinicians utilized the internal target volume (ITV) [18], expanding the beam aperture to encompass the tumor's entire trajectory. However, unlike photons, protons are highly sensitive to variations in WEPL. To mitigate motion-induced range uncertainty, custom-milled plastic compensators were designed using a technique called "smearing". Smearing involves physically modifying the compensator's topography to ensure distal coverage remains adequate even as the tumor and heterogeneities shift. While this method is robust, it adheres to a "static dose cloud" approximation that is only partially valid, and it often leads to the irradiation of significant healthy tissue to ensure coverage.

The inability of passive delivery to shape the dose proximal to the target led to the development of pencil beam scanning (PBS) [15,17]. Pioneered at the Paul Scherrer Institute (PSI), PBS replaces broad scattering foils with dipole magnets that steer a narrow "pencil" beam to paint the dose spot-by-spot and layer-by-layer. PBS systems operate in three primary modes: discrete spot scanning [15], raster scanning [19], and line scanning [20]. In discrete scanning mode, the lateral beam scans a grid of discrete spots and the beam is turned off between spots, resulting a few milliseconds of dead time; in raster scanning modes, the lateral beam scans through the same grid but the beam remains on during transitions, leaving small transient dose in between spots; in line scanning mode, the lateral beam scans continuously in space and time, not bounding to any grid. In PBS, lateral scanning is rapid (5–20 m/s), but changing the beam's penetration depth (energy switching) is significantly slower. For Hitachi ProBeatV's synchrotron system, energy switching can take 1 to 2 s for single energy extraction (SEE) and 0.2 s for multiple energy extraction (MEE) [21]. The time scale involved in energy switching is of the same order of magnitude as the respiration period. Without synchronization between beam delivery and respiration, the beam may be delivered preferentially to certain respiratory phases—a phenomenon known as the interplay effect [22,23]. This "dynamic-on-dynamic" problem cannot be solved by static margins or compensator smearing; it requires real-time synchronization between the beam delivery and the patient's anatomy changes to ensure dosimetric robustness [24].

To mitigate the interplay effect on a PBS delivery system, one can adopt a passive management scheme by adapting the treatment plan or beam delivery methods in a way not regulated by the state of the patient's breathing. A widely adopted approach is to improve the treatment plan robustness through 4D planning [25–31]. In terms of PBS delivery, rescanning [12,32–40] is the dominant practice with variants such as layered rescanning, volumetric rescanning, or phase-controlled rescanning (PCR) [20,29,41–43]. The drawbacks of rescanning include increased treatment time, some degree of sensitivity to the timing with respect to tumor motion, and the inability to deliver very small doses accurately.

Motion management can also be accomplished with active management: either the patient breathing or beam delivery, or both are modified to synchronize with each other so that the uncertainty of the delivered dose can be reduced. Breath hold (BH) [44] and abdominal compression (AC) [45] are two main approaches to reduce breathing motion in particle therapy. BH and AC are not always tolerable by patients, in which case gating is an alternative [46]. Respiratory gating is defined as the synchronization of radiation delivery with the respiratory cycle, correlating the beam-on and beam-off states with the physical location of the tumor to ensure accurate targeting. While external respiratory gating systems such as the real-time position management (RPM) system, Respiratory Gating for Scanners (RGSC), and Surface Guided Radiation Therapy (SGRT) have been widely adopted to approximate this strategy, they fundamentally rely on the assumption that external surface motion correlates predictably with internal tumor position. This assumption, however, is frequently challenged by physiological discrepancies. Hanley et al. [47] observed instances where the diaphragm moved 38 mm while the chest wall moved only 2.5 mm, highlighting a potential disconnect in motion magnitude between external surrogates and internal anatomy. It was also demonstrated that 3D tumor motion often differs significantly from surface motion due to hysteresis and phase shifts [48]. While some studies suggest using the diaphragm as a superior surrogate, Cerviño et al. emphasized that its correlation with tumor position is not universal and must be verified in a strictly patient-by-patient fashion [49]. Therefore, an ideal gating approach should generate the gating signal directly from internal tumor motion to ensure that beam delivery is synchronized to the actual position of the target volume [50,51].

The benefit of active motion management is reported in a review by Riboldi et al., who summarized 12 studies involving 445 patients with lung, liver, and pancreatic cancers and concluded that real-time respiratory tumor tracking yields superior outcomes compared to those without [52]. Similarly, Zhang et al. found that integrated mega-voltage (MV) and kilo-voltage (kV) on-treatment tracking leads to a clinically significant reduction in late urinary toxicity [53]. While these successes in photon therapy underscore the potential of motion compensation, there remains an unmet clinical need in particle therapy, where the implementation of real-time tracking remains strikingly limited. Despite the heightened sensitivity of particle beams to anatomical motion, a 2023 worldwide survey revealed that only a small minority of particle therapy centers have successfully adopted either marker-based or marker-less real-time tracking [54]. While tracking based on internal signals is theoretically the “gold standard” for motion management, its scarcity in clinical practice highlights significant technical barriers and a clear opportunity for innovation. One promising avenue for addressing this gap is fluoroscopy-based tumor tracking, which can be implemented via both marker-based and marker-less approaches, and is the primary focus of this review. Figure 1 situates FGPT within the broader landscape of imaging modalities used across the radiation therapy patient pathway. The colored path traces the convergence that motivates this review—from intra-fraction motion management scheme at the treatment-delivery stage, through kV fluoroscopy, to fluoroscopy-guided particle therapy which Sections 2–4 develop in detail.

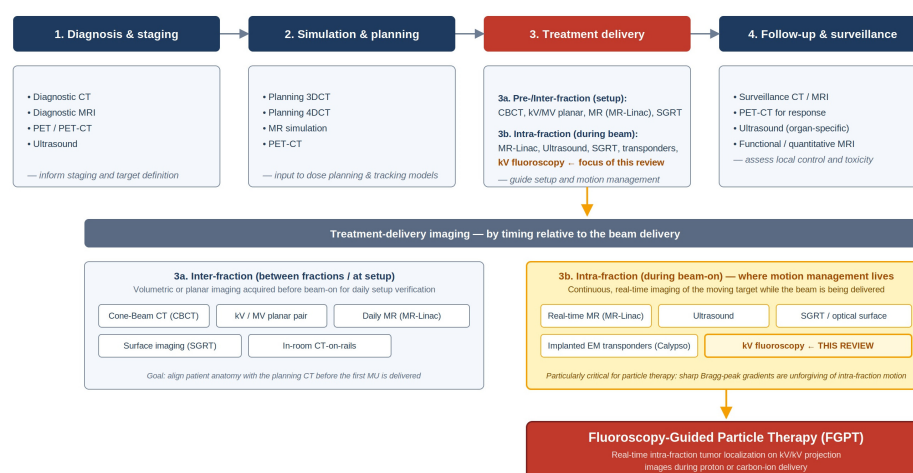


Figure 1. Imaging modalities commonly used across radiation therapy and the position of fluoroscopy-guided particle therapy (FGPT) as an intra-fraction motion management technique.

2. The Evolution of Fluoroscopy Imaging System

Fluoroscopy and radiography are both X-ray-based imaging techniques with an intricate, intertwined history, though they differ in their traditional output and application. Fluoroscopy produces real-time, continuous moving images on a screen, allowing clinicians to observe dynamic processes. Radiography, on the other hand, captures a single static photograph of the fluoroscopic image onto a medium. However, technological advances such as pulsed X-ray tubes and digital FPDs have brought these two modalities converging—modern systems can now seamlessly switch between real-time dynamic imaging and high-quality static image capture within a single digital platform, blurring the traditional distinctions between them [55].

The history of fluoroscopy and radiography is inseparable from the discovery of X-rays itself. In 1895, Wilhelm Röntgen became the first to observe real-time X-ray fluoroscopy when he watched the bones of his own hand moving between an X-ray source and a phosphor-coated screen. When he captured the now-famous radiograph of his wife's hand and wedding ring, he produced the first radiograph. Thomas Edison subsequently refined the technology, discovering that calcium tungstate (CaW), as a phosphor material, produced brighter images than Röntgen's barium platinocyanide. Edison developed a dedicated viewing device and named it the "fluoroscope". This innovation enabled non-invasive visualization of internal anatomy and was rapidly adopted in medicine. However, the early era was marked by a dangerous lack of safety awareness. Unchecked enthusiasm led to inappropriate commercial applications, such as shoe-fitting machines in retail stores, and widespread unregulated exposure to ionizing radiation resulted in significant harm—including radiation burns and deaths among early operators and the public [56,57].

In the early generations of X-ray imaging, phosphor screens lacked the efficiency to produce adequate exposure rapidly; capturing an image on film required tens of minutes, rendering radiography impractical due to motion-induced blurring. Consequently, real-time fluoroscopy became the preferred diagnostic modality rather than radiography. However, despite the remarkable sensitivity of human vision, the luminance of early fluorescent screens was insufficient for daylight viewing. Clinicians were required to undergo a 30 min period of "dark adaptation" prior to procedures. This physiological necessity ensured the sensitization of retinal rod cells, which was essential for visualizing the low-luminance images produced by early fluoroscopic screens [56]. While rod-mediated (scotopic) vision is effective in low-light conditions, it suffers from significantly reduced visual acuity—approximately ten times lower than that of cone-mediated (photopic) vision.

As a result, early fluoroscopy was characterized by poor image detail and often necessitated high radiation doses as physicians attempted to compensate for the dimness of the anatomical images.

A critical advance in fluoroscopy technology came in the 1950s with the introduction of the image intensifier (II)—a vacuum tube device that enabled fluoroscopy viewing in ambient light. The II is a chain of imaging components that converts X-ray photons to significantly intensified visible light photons [55], with one X-ray photon converted into several thousand visible light photons [58]. The process is done in stages along the imaging chain. First, X-ray photons strike an input scintillator material and are turned into visible light photons. The visible light photons then hit a photoelectric cathode and knock out photoelectrons that carry the visual signals. Major amplification is accomplished when photoelectrons are accelerated and focused by a high voltage of about 25,000 V [58], subsequently impinging onto an output phosphor material, generating a very bright visible light image. The first generation of II uses silver-activated zinc-cadmium-sulfide (ZnCdS:Ag) for the input phosphor material, but it was replaced by sodium-activated cesium-iodide (CsI:Na) in the mid-1970s. The introduction of CsI:Na doubled the scintillating efficiency of ZnCdS:Ag, owing to its higher X-ray stopping power; the required imaging dose is halved [56]. CsI remains the most common input scintillating layer on today's flat-panel detector (FPD) based system. Prior to the emergence of FPD, the output of the II required optical coupling to viewing or recording devices. This was achieved through an optical distributor—a complex assembly of lenses and beam-splitting mirrors designed to direct the intensified light to various video components. The video capture technology itself underwent significant evolution, shifting from bulky analog vacuum tubes like the Vidicon and Plumbicon in the 1960s—which were prone to image lag and spatial distortion—to Charge-Coupled Devices (CCDs) in the 1980s [59].

By the 1990s, X-ray image intensifier technology had reached a high state of refinement with not much room for significant improvement in terms of Detective Quantum Efficiency (DQE) [56]. However, the device's vacuum tube architecture remained subject to inherent physical constraints and artifacts [60]. Clinical image quality was compromised by “vignetting” (peripheral light loss) and “veiling glare” (a reduction in object contrast at the output phosphor caused by the internal scattering of light and electrons). Geometric fidelity was degraded by “pincushion distortion” which occurs because the X-ray beam is projected onto a curved input window, resulting in nonlinear magnification at the image periphery. Additionally, the electron optics were sensitive to external magnetic fields, altering electron paths to produce “S-distortion”. It was during this period that a new generation of detectors emerged, poised to replace the bulky image intensifier and its optical coupling with an integrated, compact digital system—the flat-panel detector (FPD). This technological shift was fueled by the development of large-area active-matrix liquid crystal displays. While initially intended for laptop computers, these arrays were validated as adequate for radiological applications as well [61–64]. FPDs became increasingly prevalent at the turn of the century. Their adoption enabled the standardization of digital radiography and eventually cone beam CT (CBCT), revolutionizing standard radiotherapy by transforming previously ‘blind’ treatments into high-precision, image-guided procedures [62,65–68].

The design of flat-panel detectors (FPD) is based on large-area active matrix arrays [61,69], typically constructed using hydrogenated amorphous silicon (a-Si:H) thin-film transistors (TFTs). Widely utilized in solar cell technology, a-Si is well-suited for use as a photodiode in large-area detectors; it demonstrates minimal radiation-induced degradation, a stability evidenced by its ability to withstand light intensities in solar applications that are orders of magnitude higher than those generated by scintillator screens in a fluoroscopy system [70]. The signal chain of FPD begins at the conversion layer, which follows either

an indirect or direct design. In indirect-conversion detectors, a scintillator screen composed of Thallium-doped cesium iodide (CsI:Tl) absorbs incident X-rays and converts them into visible light photons. This structured CsI configuration utilizes needle-like crystals that act as light guides, directing photons toward an a-Si photodiode integrated into the individual detector element (dixel), where the optical signal is converted into proportional electrical charges. Conversely, direct-conversion detectors utilize a semiconductor layer, such as amorphous selenium (a-Se), to convert X-ray energy directly into electron-hole pairs under a high voltage bias, eliminating the stage involving intermediate light production [55]. Regardless of the conversion method, the generated charge is collected and stored in a local capacitor located within each individual dixel, which is paired with a TFT switch. To read the detector, gate lines trigger the TFTs one row at a time. This allows the accumulated charge to flow down parallel drain (or data) lines, where it is captured and processed by column-based charge amplifiers. These amplifiers boost the analog signal before it is passed to Analog-to-Digital Converters (ADCs) for digitization and final image formation. It is worth noting that while this architecture preserves spatial resolution, temporal artifacts can arise. The intrinsic lag of a-Si photodiodes [63] and the persistence of scintillation luminescence—where visible photon emission continues after X-ray excitation ceases—can result in image ghosting (afterglow) at high frame rates. The physical origin of this lag is attributed to the deep trapping and subsequent emission of electrons within the a-Si diodes, as well as the inherent afterglow of the CsI scintillator [71]. Consequently, significant residual signals can persist; Hoheisel et al. demonstrated that it can take up to 5 s for the residual signal to decay below the signal level of the image's darkest regions [70]. This phenomenon may necessitate algorithmic correction to subtract the residual signal from the prior frame [72].

The transition to flat-panel detectors (FPD) offered distinct advantages over traditional image intensifier (II) systems, particularly regarding image fidelity and physical design. The inherent digital nature of FPDs allows for advanced post-processing—a significant upgrade over fluorescence screen-based systems—while providing the geometric accuracy required for precise image guidance. Unlike II systems, FPDs provide a completely distortion-free readout [68], a much wider dynamic range, no geometric distortion, and a wider field of view (FOV). They exhibit excellent image uniformity across the field of view by eliminating the “vignetting” and “veiling glare” inherent to II/TV chains. Indeed, FPDs have been shown to surpass II-based detectors for cone beam CT (CBCT) in terms of FOV, contrast, and resolution [73]. Physically, the FPD's compact, thin-profile design significantly improves ergonomics and patient access compared to the bulky vacuum tube architecture of the II. In terms of performance, these detectors demonstrate high DQE, high frame rates, high dynamic range, small image lag (<1%), and excellent linearity [74]. Published DQE values for FPDs are typically reported in radiographic operation at exposures of the order of μGy per image. For CsI/a-Si indirect FPDs, Granfors reported that representative zero-frequency values reach $\text{DQE}(0) \approx 0.77$ [75]. Measuring DQE under fluoroscopic conditions is complicated by temporal dynamics of the system; as a result, the DQE values typically quoted for an FPD describe its radiographic operation rather than its fluoroscopic operation. At fluoroscopic exposures, additive electronic readout noise becomes a non-negligible fraction of the per-frame quantum signal, but with adequate lag correction, the FPD retains most of its radiographic DQE: Granfors et al. measured $\text{DQE}(0) \approx 0.77$ at 150 nGy/frame—essentially unchanged from radiographic conditions—and a reduction in less than 15% even at 5 nGy/frame and 0.5 cycles/mm. When the same FPD was compared with a state-of-the-art image intensifier/CCD chain using identical methodology, the FPD retained higher DQE at radiographic exposures but converged to approximately equivalent DQE at fluoroscopic exposures (~ 8 nGy/frame), so the FPD's DQE advantage over a traditional image intensifier is preserved

in radiography but largely closes at fluoroscopic levels [75]. To circumvent this, modern systems utilize pulsed fluoroscopy, which delivers radiation in high-intensity bursts. By increasing the dose-per-pulse while maintaining a low total exposure, the detector operates in a higher-signal regime that bridges the performance gap between fluoroscopy and radiography. Ultimately, this combination of digital architecture, high frame rates, and geometric accuracy makes FPD-based fluoroscopy an ideal tool for the rigors of tumor tracking. By 2014, the technology had matured sufficiently for integration into modern proton therapy systems [76], transitioning these advanced motion management strategies from the laboratory to the treatment room.

3. The Evolution of FGPT in Particle Therapy

The early usage of fluoroscopy in particle therapy can be traced back to the pioneering efforts at the Harvard Cyclotron Laboratory (HCL) in the 1970s, necessitated by the unique challenges of treating ocular melanoma. In an effort to expand the then-novel proton therapy to broader anatomic sites, Suit and colleagues proposed that the precision of the Bragg peak could offer a conservative alternative to enucleation (surgical removal of the eye) for patients with choroidal melanoma [77]. This concept was validated pre-clinically by Constable et al. [78], and the first human cases of choroidal malignant melanoma were treated by Gragoudas et al. at HCL shortly thereafter [2]. While previous successes with treating intracranial targets using protons relied on rigid skeletal fixation—where immobilization of the skull guaranteed target stability—the eye retains significant independent mobility, requiring active motion management. In the initial protocol, the patient was asked to maintain voluntary fixation while clinicians monitored eye position through a Closed-Circuit TV (CCTV) system; this method allowed the treatment to be paused by a clinician using a hand-switch if eye movement exceeding 0.5 mm was observed [2]. However, this method could not detect an unlikely but crucial source of motion: the head and the eye moving in opposite directions, which would cause tumor misalignment while leaving the visual landmarks on the TV monitor unchanged. This limitation was resolved with the integration of a fluoroscopy system [79,80]. In this improved system, tantalum rings sutured near the tumor served as radio-opaque surrogates, allowing for direct fluoroscopic monitoring of the tumor itself. The combination of patient gaze fixation, image-based surrogate tracking, and beam suspension on detected motion remains the operational template for ocular proton therapy at the dedicated 60–70 MeV beamlines in current clinical use [3]. At the CATANA Centre, for example, the patient fixes on a small red dot delineated by a laser light to establish the treatment gaze angle, while “during the proton therapy the patient’s eye is monitored by a video camera, the position of the pupil marked on the display so that the tiniest movements can be immediately detected and irradiation can be suspended” [4].

The real-time visualization of internal anatomy during radiation therapy was first demonstrated in photon treatment by Leong et al. [81]. Using a setup that coupled a fluorescent screen to a high-speed camera, they captured images at 30 frames per second to visualize the motion of the oral cavity and soft palate during nasopharyngeal treatments. Building on this foundation, Ohara et al. at the University of Tsukuba pioneered the first respiratory-gated photon irradiation for metastatic lung tumors in the late 1980s [82]. By the 1990s, the National Institute of Radiological Sciences (NIRS) in Japan had adapted these gated irradiation concepts for carbon-ion therapy [83]. Although this implementation utilized orthogonal fluoroscopy to monitor internal organ motion, the gating signal itself was not yet derived from the radiographic images. Instead, an infrared LED placed on the patient’s chest wall functioned as a respiratory surrogate, triggering the carbon-ion beam via a synchrotron RF-knockout extraction mechanism [84]. While particle therapy centers were establishing these first gated protocols, commercial photon therapy systems were

simultaneously maturing to integrate automated motion monitoring into daily practice. Platforms such as CyberKnife (Accuray, Sunnyvale, CA, USA) [85] and ExacTrac (Brainlab, Munich, Germany) [86] adopted room-mounted orthogonal X-ray imagers to monitor target motion directly during delivery. This parallel development in the photon sector provided the first robust, commercial-grade evidence that internal target tracking could significantly improve the precision of external beam delivery. More recently, MR-Linac systems integrating diagnostic-quality MRI directly into the treatment unit have entered routine clinical use. The 0.35 T ViewRay MRIdian [87] and the 1.5 T Elekta Unity [88] both acquire continuous cine MRI during beam delivery, providing direct real-time soft-tissue visualization of the tumor with no ionizing imaging dose.

In particle therapy, gated treatments triggered by fluoroscopy have evolved along two complementary trajectories: marker-based tracking, which utilizes implanted fiducial markers as surrogate targets, and marker-less tracking, which localizes the tumor or anatomical structures directly from native image contrast. To date, only marker-based approaches have achieved commercial maturity and routine clinical deployment [76]. In contrast, marker-less tracking systems remain largely confined to institutional research environments. The following sections review the development and status of both methodologies, beginning with the history of the marker-based fluoroscopy gating system developed in Japan.

3.1. Marker-Based FGPT and Hitachi Real-Time Gated Particle Therapy (RGPT)

The foremost commercial implementation of marker-based internal gating in particle therapy is the Real-time Gated Particle Therapy (RGPT) by Hitachi Ltd. (Tokyo, Japan) (Figure 2). This solution represents the direct evolution of extensive clinical research and technological innovation from a collaboration with Hokkaido University. The system adapts the pioneering tumor-tracking technologies called Real-Time Tumor-Tracking Radiotherapy (TRTT) originally established for conventional photon radiotherapy [89–95] to meet the rigorous demands of synchrotron-based proton delivery.

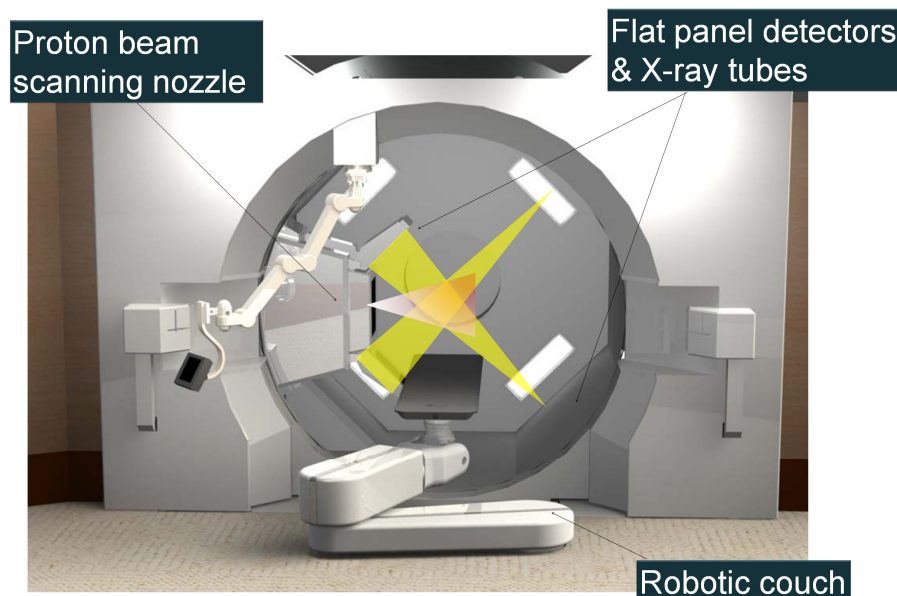


Figure 2. Hitachi RGPT Hitachi RGPT system on a rotating gantry. The horizontal orange beam represents the proton particle beam. The two mutually orthogonal yellow beams represent the on-treatment kV fluoroscopic imaging beams used for real-time fiducial-marker tracking; the corresponding kV X-ray tubes and flat-panel detectors are mounted directly on the gantry so that the imaging geometry rotates with the treatment beam. Reproduced from Shimizu et al. [76] under the terms of the Creative Commons Attribution License (CC BY 4.0).

RTRT was the first implementation of fluoroscopy-triggered gated treatment. It was first implemented at Hokkaido University using a linear accelerator (Linac) in the late 1990s. In this system, patients were implanted with round gold fiducial markers of 2 mm diameter. During treatment, two sets of diagnostic fluoroscopes capture tumor motion in real-time. The locations of the fiducial markers were automatically processed using Otsu's thresholding algorithm [96], and radiation was triggered when the gold marker was located in the planned position. The initial fluoroscopy imaging chain was based on image intensifiers and could take images every 0.03 s. The system achieved a tracking accuracy of 1.5 mm for tumors moving at speeds up to 40 mm/s [97]. To optimize the signal and minimize patient dose, the Linac pulse was synchronized to pause during X-ray acquisition, while the image intensifiers were activated only during the X-ray pulses [90]. This in-house tracking system also enabled researchers to perform detailed studies of tumor motion and hysteresis driven by respiratory and cardiac cycles [48].

The integration of real-time tracking was significantly aided by the evolution of beam delivery hardware. In passive scattering systems, the presence of large field-shaping devices at the nozzle often physically obstructed or narrowed the field of view of imaging devices. The transition to dedicated PBS systems removed these bulky accessories, creating the spatial clearance required to install fluoroscopy panels without compromising their imaging geometry or limiting their visibility of the target [98].

The successful implementation of respiratory-gated particle therapy requires first overcoming a layer of technical complexity greater than that found in photon therapy, primarily due to the specific constraints of the beam delivery mechanism. The challenge differs depending on accelerator architecture. In a cyclotron-based system, the magnetic field remains fixed and produces a continuous beam similar to a linear accelerator (Linac); as a result, the duty cycle and efficiency of gated irradiation are comparable to those of standard photon therapy. In contrast, a synchrotron system operates on a pulsed cycle, where the magnetic field must be ramped in synchronization with the increasing energy of the particles to maintain a stable circular trajectory. Beam extraction is restricted to the "flat top" phase of the magnet excitation pattern, making delivery efficiency highly sensitive to the synchronization between the synchrotron cycle and the patient's respiratory phase. In a simulation study utilizing actual patient motion traces, Tsunashima et al. [99] found that a fixed magnet excitation cycle would increase average treatment times by a factor of three, suggesting that a variable magnet excitation pattern is essential for efficient gated irradiation. Addressing this, Hitachi Ltd., in collaboration with Hokkaido University, developed a "beam waiting" function that extends the flat top phase, enabling multiple gated irradiations within a single synchrotron cycle [100]. Using patient trajectory data collected from the RTRT system during the photon treatments, Matsuura et al. [101] determined the optimal gate window to be 2 mm for gated proton treatment. This series of investigations led to the clinical implementation of Real-time Gated Particle Therapy (RGPT) with Hitachi [76]. A hardware evolution in this system was the replacement of the image intensifier and optical coupling systems—standard in the previous photon-based RTRT system—with a modern FPD-based imager. This transition improved space efficiency, enabling the integration of orthogonal imaging units directly into the gantry alongside the spot-scanning nozzle. The system utilizes fluoroscopy to track implanted fiducials at 30 frames per second (FPS) with a tracking accuracy of 1 mm. The synchrotron operation cycle (injection, acceleration, waiting for the first gate, extraction, and deceleration) ranges from 2 to 7 s. While the beam waiting function improves efficiency by enabling multiple gates per cycle, the wait timer is limited to 200 ms to ensure stability [76]. The system latency—defined as the duration between the generation of the pulsed X-ray beam and the resultant proton beam-on/off—is set to 66 ms. Early clinical evaluations indicated

that treatment time lengthening was manageable, ranging between 1.22 and 1.72 times the standard duration [76].

Two principal configurations of RGPT systems have been implemented clinically by Hitachi, each with distinct geometric and operational characteristics. In the first configuration, the X-ray source and flat-panel detector are mounted directly on the rotating treatment gantry, such that the fluoroscopic projection angle is inherently coupled to the treatment beam direction. In such cases, the treatment planning needs to carefully balance dosimetric optimality against the visibility of the fiducial markers on the fluoroscopy projection, as at certain gantry angles, overlying bony structures or other high-density anatomy may obscure fiducial markers, compromising tracking reliability. In the second configuration, the X-ray imaging system is fixed to the room infrastructure, with the X-ray tube installed on the ceiling and the detector panel mounted on the floor, resulting in fluoroscopic projections that remain geometrically constant irrespective of the treatment beam angle. Both configurations are being implemented at Mayo Clinic Florida, where the gantry-mounted system is deployed on the proton therapy gantry and the room-fixed configuration is installed at the carbon-ion fixed-beam port. For pulsed-fluoroscopy operation specifically, the Hitachi PROBEAT RGPT system operates at 70–125 kVp with tube currents up to 99 mA, and selectable pulse rates of 1, 7.5, 15, or 30 pulses per second chosen according to the target's motion characteristics—typically 1 PPS for quasi-static targets such as prostate and 15–30 PPS for respiratory-driven liver, lung, and pancreatic lesions [102]. The pulse width is fixed at the installation level and is set within 2–3 ms. Beam quality of the X-ray imaging is specified by half-value-layer (HVL) acceptance testing, with a minimum permissible first HVL of ≥ 2.5 mm Al at 70 kVp rising to ≥ 5.4 mm Al at 150 kVp.

Figure 3 illustrates the general RGPT clinical workflow, from fiducial marker insertion to treatment delivery. Figure 4 shows a typical RGPT field treatment workflow, detailing patient setup, template preparation, and gated treatment.



Figure 3. General RGPT clinical workflow.

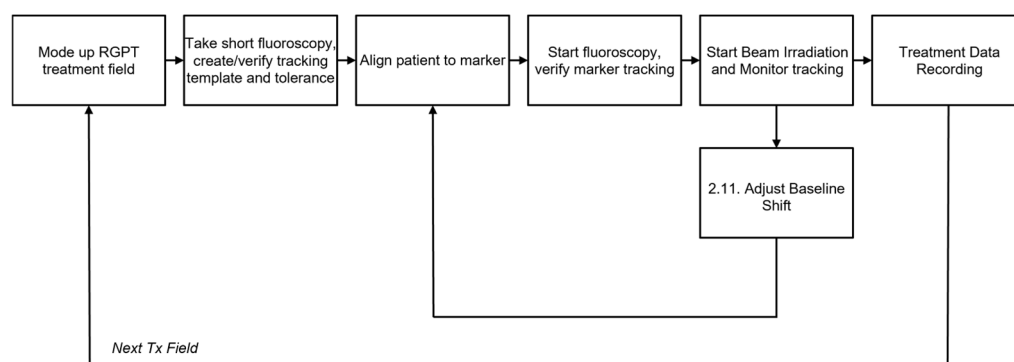


Figure 4. Typical RGPT workflow in a treatment.

The dosimetric benefits of RGPT have been demonstrated in multiple simulation and validation studies. Shimizu et al. showed that RGPT dramatically improved dose conformity for hepatocellular carcinoma: while free-breathing spot-scanning achieved successful dose delivery (95–107% CTV coverage) in only 9/48 and 0/48 motion scenarios for two patients, RGPT achieved 48/48 and 42/48, respectively, while also reducing mean liver dose by approximately 50% for smaller tumors [76]. For lung tumors, simulation

studies have demonstrated that RGPT with a ± 2 mm gating width—i.e., the proton beam is permitted to remain on while the fiducial marker stays within a 2 mm radius of its planned 3D position—reduces dose error without substantially extending treatment time. [101]. Yamada et al. validated the reliability of RGPT dose delivery through log-data-based dose reconstruction across 168 fractions in eight liver cancer patients, confirming that delivered doses closely matched planned distributions [103]. Clinical outcomes are just beginning to emerge. Nishioka et al. reported the first prospective study of RGPT for prostate cancer, demonstrating 88.9% five-year biochemical relapse-free survival with early adverse event rates (8.9% $\geq$ grade 2) that were non-inferior to conventional proton therapy [104]. While these results confirm the safety and feasibility of RGPT, comparative clinical outcome studies directly demonstrating superiority over non-gated approaches for moving tumors remain limited, highlighting an opportunity for future investigation.

Comprehensive commissioning and quality assurance protocols for RGPT have recently been published. Chen et al. described the commissioning of the Hitachi PROBEAT system at Johns Hopkins University, demonstrating that dose delivery to moving targets passed 3%/3 mm gamma analysis and that plan delivery uncertainty could be maintained within 2 mm [105]. Tan et al. reported the first published RGPT-specific commissioning and QA procedure, detailing six commissioning measurements including imaging quality, imaging dose, marker tracking accuracy, gating latency, tracking fidelity for irregularly shaped fiducial markers, and dosimetry. Their results showed gating latencies of 119.5 ms and 50.0 ms for beam-off and beam-on, respectively, with daily marker localization accuracy consistently below 0.2 mm [102]. Koh et al. presented the first comprehensive Failure Modes and Effects Analysis (FMEA) for RGPT following AAPM TG-100 guidelines, identifying 96 potential failure modes across the clinical workflow and highlighting irregular patient breathing as the highest-risk process [106]. Additional physics investigations have addressed treatment time efficiency, with Yoshimura et al. analyzing whether gated spot-scanning delivery can be completed within standard 30 min session times [107], and robustness evaluation methodology, with Lee et al. comparing algorithms for determining maximum allowable CTV shifts during RGPT for prostate cancer [108].

The imaging dose associated with continuous fluoroscopy during real-time tumor tracking needs to be considered in clinical implementation [109]. Shirato et al. reported that skin surface dose from a single fluoroscope in RTRT ranged from 29 to 1182 mGy/h depending on kVp and pulse width settings, concluding that precise dose estimation and reduction strategies are essential as lung RTRT treatment longer than 30 min of fluoroscopy can result in clinically significant cumulative imaging exposure [110,111]. Imaging dose can be reduced by using the lowest FPS setting suitable for the disease site, for example, the prostate often requires 1 FPS, whereas the lung may require 15 FPS. Postprocessing enhancement may further reduce the imaging exposure; for example, Miyamoto et al. showed that imaging dose can be reduced through motion-compensated image filters while maintaining tracking accuracy comparable to high-dose imaging [112].

The interaction between fluoroscopy and particle delivery creates a dual-challenge environment where each system can potentially compromise the accuracy of the other. While secondary radiation from the treatment beam can introduce noise into the imaging data, the inverse is also true: scattered fluoroscopic X-rays can infiltrate the dose monitor (DM) and be mistakenly recorded as proton monitor units (MU). Currently, the standard approach to mitigate this signal contamination is to momentarily pause during each fluoroscopy pulse to ensure that X-ray scatter is not active while the DM is recording, a method known as Interrupted Continuous Delivery (ICD). ICD reduces delivery efficiency, especially for particle beams delivered by Dose Driven Continuous Scanning (DDCS), intended for high-dose rate delivery. Recent investigation, however, suggests that ICD may

be unnecessary for modern high-dose rate systems [113]. Yamanaka et al. found that a proton-beam current greater than 2 MU/s keeps target dose deviations within 1% of the planned distribution; the threshold for the organ at risk was 1 MU/s [113]. Regarding the converse effect where the particle beam affects the fluoroscopy imaging system, Terunuma et al. evaluated treatment beam-induced secondary-radiation background on the quality of fluoroscopy and found that 1.25% of the pixels are affected by sparse spikes and impact the mean-pixel-intensity variation remaining below 1% [114]. This may not affect visual perception, but it may affect the robustness of machine learning algorithms, which could be sensitive to out-of-distribution noises.

Although Hitachi's PROBEAT/RGPT is the only commercially deployed system that performs continuous fluoroscopic intra-fraction tumor tracking in particle therapy, recent proton therapy vendors all provide substantial in-room kV imaging infrastructure (summarized in Table 1) on which comparable real-time tracking workflows could, in principle, be built.

Table 1. In-room kV imaging configurations of recent proton therapy systems.

Vendor/System	Gantry Rotation	kV Imaging Configuration	Volumetric Imaging	Reference
Hitachi PROBEAT (RGPT)	360°	(A) Gantry-mounted X-ray tubes + flat panels; (B) Room-fixed ceiling-mounted tubes + floor detectors	kV-CBCT and stereoscopic kV-kV planar	[76]
Varian ProBeam 360°	360° (±190°)	Two orthogonal kV imaging chains integrated into a gantry; 40–140 kV, 0.4–1000 mAs	Gantry-mounted kV-CBCT (80–140 kV) with full or half rotation	[115]
IBA Proteus PLUS/ONE	220°	Room-fixed kV-kV stereoscopic pair (two floor tubes 60° apart) + retractable gantry-mounted kV tube + detector	Gantry-mounted kV-CBCT	[116]
Mevion S250i HYPERSCAN	360°	Room-fixed: two orthogonal a-Si flat panels on ceiling rails	Separate medPhoton ImagingRing CBCT (imaging isocenter offset 50 cm)	[117]

Despite its clinical utility, marker-based tracking is subject to several limitations, ranging from complications associated with fiducial placement to significant dosimetric uncertainties. The most common complication is pneumothorax; Laurent et al. [118] reported incidence rates of 15% and 16.2% across two patient groups, while Geraghty et al. [119] observed a rate of 27% (226 out of 846 patients), and Collins et al. [120] reported rates as high as 30%. Pulmonary hemorrhage (hemoptysis) is another known risk. Furthermore, marker migration can compromise targeting precision [121]. The study by Kitamura et al. concluded that a planning target volume margin should be used to account for the registration uncertainty caused by marker migration [93]. The effect of fiducial marker migration was also studied by Shirato et al. [122], Imura et al. [123], and Van der Voort van Zyp et al. [124]. To mitigate marker migration, typically three non-collinear markers are used for 3D tracking [120], and Imura et al. suggest delaying tumor tracking radiotherapy until at least five days after insertion [125]. A rare but serious complication called tumor track seeding has also been documented by Patel et al. [126], who reported a case where

a new tumor nodule developed around a gold fiducial, likely due to the needle dragging malignant cells through the track. However, they noted that while seeding may be common, implantation metastasis is rare because lung cancer cells generally have low growth potential in the pleural space. Finally, operational limitations also exist: not all patients are eligible for fiducial placement.

From a dosimetric perspective, high-density markers can cause Hounsfield Unit (HU) artifacts that introduce calculation uncertainties [13], which is particularly critical for particle therapy. Newhauser et al. found that a 2.5 mm tantalum marker used in proton therapy for uveal melanoma could create a dose shadow ranging from 22% to 80% [127]. Giebeler et al. demonstrated via Monte Carlo simulation that dose perturbation depends on marker size, orientation, and distance from the beam's end of range [128]; they observed dose perturbations of 31% for large markers and 23% for medium markers in lateral opposed pair treatments. Habermehl et al. recommend utilizing only thin markers (<0.5 mm) or low-Z materials for hadron therapy to minimize these effects [129]. Matsuura et al. recommended utilizing 1.5 mm markers to avoid Tumor Control Probability (TCP) reduction [130]. Although they suggested that utilizing multiple-field irradiation could mitigate the underdosing effects caused by larger diameter markers, this could compromise the ability to spare surrounding critical organs.

Given the significant invasiveness and dosimetric uncertainties associated with fiducial markers, the ideal gating solution would be a non-invasive, marker-less tracking approach. Such methods aim to derive the gating signal directly from internal anatomical features, utilizing structures like the diaphragm as a surrogate or, most optimally, tracking the tumor mass itself to ensure precise beam delivery without the risks of implantation.

3.2. Marker-Less FGPT

Marker-less tumor tracking for particle therapy was pioneered at the National Institute of Radiological Sciences (NIRS) in Japan to support their carbon-ion PBS system [131–133]. NIRS launched a clinical trial in March 2015 with 10 patients (five lung and five liver), representing the first clinical application of marker-less tumor tracking for liver cancer in particle therapy. Treatments employed respiratory gating near end-expiration (typically phases T30–T70, but when motion within specific phase windows was deemed too rapid, T30 or T70 could be adaptively excluded to maintain stable gating). The initial implementation [131] reported an overall gating positional error less than 2.2 mm at the 95% confidence level, although the study used a non-traditional but clinically pragmatic definition of gating error—error was set to zero as long as the delivered CTV remained within the planned PTV. Under this definition, larger PTV margins can intrinsically yield smaller reported gating errors. Tracking accuracy was <2.5 mm, although the fraction of frames achieving a Tracking Registration Error (TRE) < 1 mm varied across patients (typically ~48–70%). The author showed that compared with simulated external gating (30% duty cycle based on external surrogates), the internal fluoroscopic approach produced substantially better gating accuracy, particularly for large-amplitude tumor motion. Figure 5 shows the marker-less tumor tracking system for carbon-ion pencil beam scanning treatment developed at NIRS, reproduced from Mori et al. [131] with permission.

The underlying tracking methodology integrates multiple-template-matching with machine learning augmentation to achieve robust target localization [134]. The multiple-template-matching framework relies on a library of “reference snapshots” constructed from pre-treatment fluoroscopy, where each template is explicitly associated with specific tumor coordinates and a respiratory phase [135,136]. During treatment, live fluoroscopic frames are compared against this library to identify the closest matches, hence infer the tumor's position. To account for inter-fractional variations in breathing compared to the

baseline sequence, the algorithm permits small translational shifts in the incoming image to maximize similarity. Rather than selecting a single best match, a voting strategy is employed to enhance robustness, aggregating tumor positions from all templates that exceed an empirical similarity threshold, typically between 85 and 95 training sets of approximately 60 images—roughly two breathing cycles—the multiple-template matching method achieved a reported tracking accuracy of ~ 3 mm. An interesting technical detail in the algorithm is that the respiratory phase is determined by the intensity of the fluoroscopic images: Berbeco et al. observed that fluoroscopic intensity fluctuates predictably with respiration, typically appearing darker during exhalation and brighter during inhalation, allowing for the definition of the phase via an extracted intensity waveform [137].

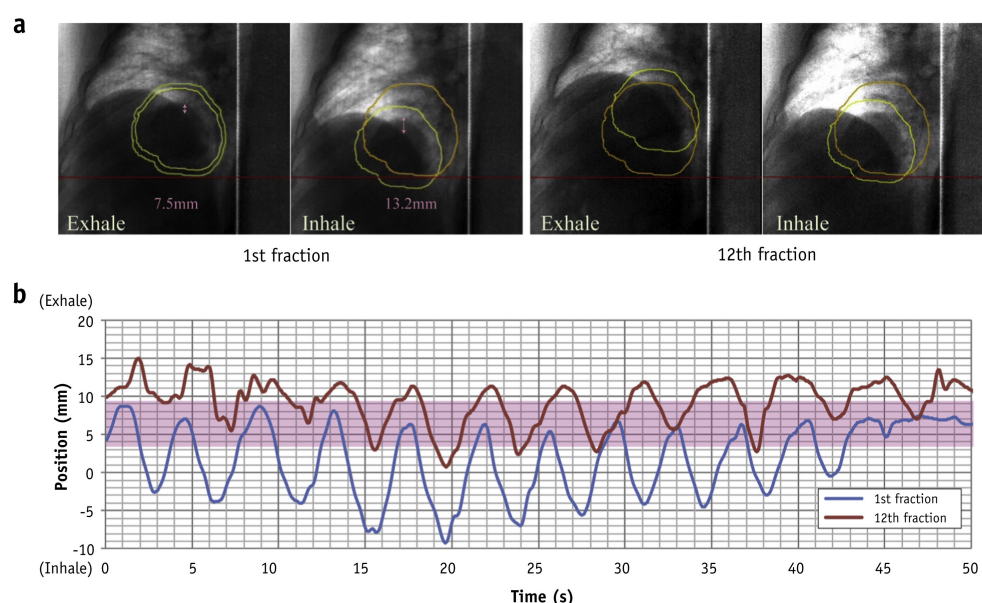


Figure 5. Marker-less tracking of lung tumor during carbon-ion therapy at NIRS reproduced from Reference [131] with permission. (a) Fluoroscopic images at exhale and inhale phases of the 1st and 12th treatment fractions. The green and orange contours delineate the PTV, and the yellow contour shows the CTV detected by the marker-less tracking algorithm; the carbon-ion beam is turned on whenever the tracked CTV lies inside the PTV; (b) Superior-inferior position of the tracked CTV as a function of time during the 1st (blue) and 12th (red) fractions. The pink shaded band marks the range of tracked CTV positions for which the carbon-ion beam is turned on.

The workflow originally developed at NIRS comprises the following steps:

- 4D-DRR generation: Create digitally reconstructed radiographs (DRRs) for each respiratory phase intended for irradiation, with the CTV projected onto each DRR.
- DFPD acquisition and registration: Before treatment, dynamic flat-panel detector (DFPD) fluoroscopy video is acquired over several breathing cycles; register each DFPD image to the corresponding 4D-DRR using a 2D–2D registration algorithm.
- Manual verification: An oncologist and physicist review the projected CTV positions on DFPD images at each phase and adjust as needed to ensure ground-truth accuracy.
- Optimization: Automatically optimize the number of templates, similarity thresholds/scores, confidence metrics, and the machine learning dictionary to finalize tracking parameters.
- Tracking: During treatment, an instantaneous fluoroscopy image is analyzed by multiple-template matching to determine tumor location and trigger a gating signal accordingly.

A critical phase of this process involves manual verification, where an oncologist and physicist review the projected Clinical Target Volume (CTV) positions on DFPD images to ensure ground-truth accuracy before automatically optimizing tracking parameters. However, this “human-in-the-loop” requirement presents a significant operational bottleneck, as providing manual ground-truth tumor positions for training images can take up to 10 min per patient [131], thereby significantly limiting clinical throughput. Recent research at NIRS has increasingly focused on leveraging these neural networks to replace manual labeling and optimize model preparation [133,138–143]. Deep learning represents a transformative opportunity to overcome efficiency hurdles in marker-less tracking by automating feature extraction and ground-truth generation. By eliminating the need for manual curation, deep learning algorithms can streamline the preparation of template libraries and machine learning models. This advancement is particularly critical as therapy systems transition toward more efficient delivery modes where the concurrent operation of imaging and treatment beams demands highly robust, automated tracking to maintain dosimetric accuracy. The shift toward deep learning-based marker-less tracking could be the key to unleashing the full potential of marker-less FGPT. Table 2 provides a concise summary of the marker-based and marker-less FGPT systems discussed in the preceding text.

Table 2. Summary of marker-based and marker-less FGPT systems discussed in the manuscript.

Author	Institution	Modality	Highlight	Ref
A. Commercial Marker-Based FGPT (Hitachi RGPT)				
Shimizu et al. 2014	Hokkaido University	Proton PBS	CTV coverage 48/48 vs. 9/48 free-breathing (HCC); ~50% liver dose reduction	[76]
Nishioka et al. 2024	Hokkaido University	Proton PBS	88.9% 5-yr bRFS for prostate cancer; ≥ 2 AE rate 8.9%	[104]
Chen et al. 2024	Johns Hopkins University	Proton PBS	Commissioning: dose delivery passed 3%/3 mm gamma; plan uncertainty within 2 mm	[105]
Tan et al. 2024	National Cancer Centre Singapore	Proton PBS	First RGPT-specific commissioning and QA report	[102]
Koh et al. 2025	National Cancer Centre Singapore	Proton PBS	Workflow FMEA	[106]
B. Clinical Marker-Less FGPT (NIRS Carbon-ion)				
Mori et al. 2016 and 2019; Hirai et al. 2016; Sakata et al. 2016 and 2020	NIRS (QST), Japan	Carbon-ion PBS	First marker-less gated PBS clinical trial for Carbon-ion therapy	[131–134,138]
C. NIRS Follow-Up AI Research for Marker-Less Tracking				
Hirai et al. 2019	NIRS (QST), Japan	Carbon-ion PBS	DNN tumor probability map	[139]
Hirai et al. 2020; Mori et al. 2020 and 2023; Takahashi et al. 2020	NIRS (QST), Japan	Carbon-ion PBS	Deep learning follow-up studies: regression models, synthetic fluoroscopy, patient-specific training	[140–143]

Marker-based and marker-less fluoroscopic tracking deliver different patient imaging doses on the same vendor system at identical kVp, mA, and pulse-width settings. The reason is geometric: marker-based tracking follows a small fiducial marker, so the minimum X-ray FOV is set by the fiducial's motion amplitude alone. Marker-less tracking follows the soft-tissue tumor and typically requires anatomical context for the registration; the minimum FOV is therefore the geometric sum of the projected tumor diameter, the motion amplitude, and an algorithm-dependent context margin, and is systematically larger than the marker-based minimum. Because patient dose-area product scales with FOV area at fixed exposure parameters, the imaging-dose advantages reported for marker-based RGPT do not transfer one-to-one to marker-less FGPT.

4. Harnessing AI for Marker-Less Tumor Tracking: Opportunities and Challenges

As of May 2026, no commercially deployed FGPT system uses artificial intelligence for soft-tissue tumor localization during particle therapy. The only mature commercial FGPT—Hitachi's PROBEAT RGPT performs fiducial localization by traditional template matching. The first and, to our knowledge, only clinical implementation of marker-less fluoroscopic tracking of soft tissue in particle therapy is the NIRS carbon-ion system pioneered by Mori and colleagues [131–134], which uses multiple-template matching with machine learning augmentation rather than an end-to-end deep neural network. The deep-learning algorithms surveyed in the remainder of this chapter, including the direct follow-ups to the NIRS work [139–142], are preclinical research-stage demonstrations. Accordingly, the present chapter reviews algorithms for marker-less tumor tracking in the broader context of fluoroscopic image registration. Many of these methods were originally developed for photon radiotherapy but are directly transferable to particle therapy, since the underlying kV fluoroscopic imaging chain is essentially identical across modalities.

4.1. Central Challenge for Marker-Less Tumor Tracking Using Projection Images

Fiducial marker tracking via kV imaging has demonstrated high accuracy across multiple clinical studies [144,145], even for lower-quality MV images [146]. This proven reliability has paved the way for the commercial implementation of RGPT [76]. However, the mandatory use of implanted fiducials imposes significant procedural burdens—including the risks associated with invasive insertion and marker migration—limiting the technique's applicability to specific anatomical sites. Marker-less tumor tracking, which localizes the target directly from native anatomical contrast to trigger gating signals, represents the ultimate paradigm shift in FGPT. While this approach holds the potential to expand the benefits of motion management to a broader patient population and disease sites, it presents significant technical challenges. The central challenge of marker-less FGPT is to develop algorithms that achieve “human-level” robustness: the capacity to reliably identify and track target structures despite the vast anatomical and image-quality variability encountered in clinical practice. Despite the rapid development of deep learning research and numerous proposals, achieving this degree of generalized robustness remains the foremost unsolved challenge for FGPT.

Accurate tumor tracking during radiation therapy requires robust real-time 2D/3D image-registration algorithms capable of aligning target structures on the 2D projection image acquired during treatment. Medical image registration has been the subject of extensive methodological development, spanning traditional optimization-based approaches [147–150] to more recent deep learning methods. An effective registration algorithm must be able to focus selectively on structures of clinical interest while ignoring irrelevant features—a task that is intuitive for humans but remains a significant challenge for computers. This

challenge is particularly acute for fluoroscopic projection images, where three-dimensional (3D) anatomy collapses into a single two-dimensional plane, creating complex patterns of tissue overlap. Although trained clinicians readily identify target anatomy within these projections, replicating this semantic understanding computationally has proven difficult.

Traditional registration algorithms search for the extremum of an image similarity metric [151]—such as Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC), or Mutual Information (MI)—as a function of registration parameters. While these metrics have achieved some success, significant limitations remain. The extremum of a similarity metric does not necessarily correspond to alignment of the clinically relevant structure, leading to erroneous registrations when the objects to be aligned are far apart or when irrelevant interfering features (e.g., bone projections over a lung tumor) dominate the image. This limitation is evident when positioning patients using orthogonal kV projection images at the treatment console. Automatic registration frequently fails if there is a large initial anatomic shift between the kV image and the reference digitally reconstructed radiograph (DRR). Currently, radiation therapists often need to perform manual pre-registration to bring landmarks into closer alignment before running the auto-registration. The problem is compounded when structures of interest are obscured by overlapping anatomy. Recently, deep learning has renewed hope for addressing these challenges by enabling algorithms to learn which structures to prioritize, though robust solutions remain an active area of research.

4.2. Emerging Deep Learning Algorithms

Over the past decade, deep learning has fundamentally transformed the research landscape of medical image analysis. The success of deep convolutional neural networks on natural image classification—achieving superhuman accuracy on certain benchmarks [152]—has inspired numerous efforts to develop systems that could assist or even replace humans in medical image analysis. Tremendous progress has been made across image-based diagnosis [153], image segmentation [154,155], image registration [156–159], and image reconstruction [160,161]. In radiation oncology, this is perhaps most evident in auto-segmentation, where deep learning models now delineate organs and tumor volumes with speed and precision that dramatically streamline clinical workflows.

The breakthrough success of AlexNet [162] on the 2012 ImageNet classification task revitalized the pursuit of human-level pattern recognition using deep neural networks in all aspects of industry, including the healthcare system. Deep neural network models are machine learning models parameterized by a large number of adjustable parameters that can be optimized (trained) using a large set of examples. A fundamental insight of deep learning, often referred to as the “scaling law”, is that the ability of deep learning models to generalize to new examples improves consistently as both the number of trainable parameters and the volume of training data increase. Notably, AlexNet, which brought revolution to the computer vision world, utilized an architecture structurally similar to LeNet [163] developed a decade earlier for written digit recognition, except that the number of parameters is orders of magnitude greater and the model was trained on a dataset order of magnitude larger. Three contributing factors drove the performance leap observed in AlexNet compared to the LeNet era of the late 1980s. The first factor is algorithmic efficiency: the calculus-based backpropagation algorithm accelerates parameter updates by at least a factor of 10 million compared to previous training methods. Secondly, hardware improvements over the intervening decades yielded a million-fold speed up in computation. Thirdly, efficient parametrization of the neural network using convolutional neural networks (CNNs) matches the inductive bias of natural images, reducing the number of parameters to be trained by a factor of millions compared to fully connected neural

networks. These three factors multiplied, giving rise to the possibility of training very large models over very large datasets in the 2010s.

The historical trajectory of deep learning in medical image analysis closely mirrors developments in mainstream AI research, with methodologies and architectures originally designed for natural images frequently being transferred or adapted for medical applications. This parallel development is evident in the adoption of the VGG network [164]—a leading architecture in the 2014 ImageNet challenge—to achieve state-of-the-art performance in melanoma diagnosis [165] for the ISIC-2016 skin cancer classification challenge. Similarly, the Inception v3 architecture [166] developed in 2015 enabled dermatologist-level skin cancer classification [153]. The integration of ResNet [167] led to an architecture that secured the highest average classification performance across three skin cancer categories in the ISIC-2017 challenge [168]. In the domain of semantic segmentation, the fully convolutional network (FCN) introduced by Long et al. [169] established the encoder–decoder framework, which directly inspired U-Net [170]—the building block for numerous medical image segmentation models [154,155]. Generative Adversarial Networks (GANs) [171], introduced in 2014, experienced explosive adoption across a wide range of medical tasks, including disease classification, auto-segmentation, image registration, image reconstruction, and cross-modality image synthesis [172–176]. The attention mechanism [177], initially proposed for neural machine translation, achieved remarkable success in language modeling through the transformer architecture [178]. Its adaptation for computer vision, the Vision Transformer (ViT) [179], was rapidly integrated into medical image analysis as a potent alternative to traditional CNN-based architectures [180–184]. Most recently, Denoising Diffusion Probabilistic Models (DDPM) [185–187] have surpassed GANs in popularity for image generation tasks, quickly becoming a focal point in medical image related research [188–191].

There has been tremendous progress in applying deep learning to medical image registration [156–159]. Deep learning models for image registration typically fall into the following frameworks: methods based on direct parameter regression, methods based on segmentation, methods based on image synthesis, and others. Table 3 summarizes the algorithms reviewed in Section 4.2.

Table 3. Quantitative summary of AI-based registration and marker-less tracking methods, grouped by Section 4.2 paradigm taxonomy.

Ref	Author	Image	Speed on GPU	Highlight
A. Direct parameter/DVF regression				
[192]	de Vos 2017—DIRNet deformable	Cardiac cine MRI	<50 ms	Spatial transformer-based deformable registration.
[193]	de Vos 2019—DLIR (affine + deformable)	Cardiac MRI; Chest CT	<40 ms	Coarse-to-fine spatial transformer-based deformable image registration.
[194]	Li 2018	Brain MRI	~50 ms	Jointly optimize the spatial transformer and the fully convolutional network (FCN).
[195,196]	VoxelMorph (Balakrishnan 2018 and 2019)	Brain MRI	~24 s	UNet + STN.
[197]	CycleMorph (Kim 2021)	Brain MRI; Liver CECT	~1 s/pair	Topology regularization via cycle consistency.
[188]	DiffuseMorph (Kim 2022)	Brain MRI; Cardiac MRI	<1 s	Diffusion model for deformable registration. Iterative.

Table 3. Cont.

Ref	Author	Image	Speed on GPU	Highlight
B. Segmentation-based				
[144]	Mylonas 2019	kV fluoroscopy, prostate	~9 ms	Prostate fiducial detection.
[198]	Roggen 2020	kV projection, spine SBRT	~0.5 s	ResNet, Mask R-CNN, Faster R-CNN. vertebra bone-based surrogate.
[199]	He 2022	kV projection (Varian), pancreas	~30 ms	Stent as a surrogate for pancreatic tumor motion. Perceptual Attention UNet.
[200]	Edmunds 2019	CBCT projections, lung	~0.5 s	Mask R-CNN; Diaphragm surrogate. Worse at lateral angles.
[139]	Hirai 2019	kV fluoroscopy, lung + liver Carbon-ion	<40 ms	4DCT-derived DRRs. Predict Target Probability Map (TPM).
[141]	Takahashi 2020	kV fluoroscopy, lung phantom	32.5 ms	Patient-specific FCN. Phantom proof of concept.
[201]	Terunuma 2018	kV fluoroscopy, lung	25 ms	“Importance recognition”: bone suppression.
[202]	Terunuma 2023	kV fluoroscopy, lung	8 ms	Attention heatmaps for explainability.
[203]	Huang 2024	Simulated kV, lung	170 ms/frame	Patient-specific Retina U-Net.
[204]	Mylonas 2025	kV projections, prostate	~10 ms	cGAN prostate segmentation. Trained on synthetic kV from planning data. Patient-specific model.
C. Image synthesis-based				
[205]	Lei 2020	kV proj → 3D CT, lung SBRT	<1 s/volume	TransNet GAN.
[206]	He 2021	kV projections, spine SBRT	~0.1 s	ResNetGAN spine-only decomposition to suppress soft tissue.
[207]	Fu 2023	kV projections, lung	~50 ms	Pix2Pix sDTI Target-only decomposed image suppresses anatomy.
[208]	Fu 2025	kV intra-fraction, lung	~50 ms	First clinical sDTI deployment.
[209]	Madden 2024	Simulated kV, pancreas SBRT	N/A	CBCT-DRR for better domain match for on-treatment tracking.
[210]	Ahmed 2025	kV intra-fraction, pancreas	~29 ms	cGAN CBCT-DRR fine-tuning.
[211,212]	Yan 2024/2025	Color fluoroscopy, lung	179.8 ms	DUCK-Net trained on DRRs.
D. Other methods				
[213]	Wang 2020	kV CBCT projections, lung	~20 ms	CRNN (CNN + RNN); RNN exploits the temporal continuity of projections.
[214]	Grama 2023	kV during VMAT, lung SBRT	~30 ms	Siamese network
[215]	Mok 2022	Brain MRI (atlas)	<0.1 s	ViT
[216]	Xu 2024	Stereoscopic kV (CyberKnife), lung	Real-time	Zero-shot Pre-trained DNN + template matching; uncertainty measure.

Methods based on direct parameter regression: Direct parameter regression constitutes one of the earliest paradigms in AI-based image registration, wherein neural networks predict the coordinates of the tumor or its bounding box. Neural network-based region proposal [217] and its variants [218,219] have been applied to pancreas [220,221], lung [222], and fiducial markers [223]. For this type of network, the models are trained to minimize a combination of regression loss and classification loss. CNN-based regression was also used to produce the Deformation Vector Field (DVF) [140,224,225]. The introduction of the Spatial Transformer Network (STN) by Jaderberg et al. [226] provided a popular differentiable module capable of explicitly warping an image within a neural network architecture. This mechanism has been widely adopted in medical physics for image registration. De Vos et al. utilized STNs for affine and deformable registration [192,193], while Li et al. applied the framework to non-rigid registration tasks [194]. The VoxelMorph model [195,196]

integrated STN-based warping to enable rapid learning of deformable registration fields. A variation called CycleMorph [197] incorporated cycle-consistency constraints into this architecture to improve topological preservation. With the advent of diffusion model, DVF generated by a diffusion model has also been proposed [188].

Methods based on segmentation: Segmentation-based methods have emerged as a popular paradigm due to the success of U-Net [170] for pixel-level delineation of targets or surrogates [155]. The idea of this approach is to segment the structure of interest from the projection images and then use the segmentation for image registration. This has been applied to track fiducial markers [144], spine [198], pancreatic stent [199], as well as soft tissue such as diaphragm [200] or tumor itself [139,141,201–204]. A marker-less lung tumor tracking algorithm was developed by Hirai et al. [139] at NIRS in order to improve upon their existing multiple-template matching algorithm. In their deep learning model, a tumor probability map (TPM) was predicted from a structure similar to a U-Net. The trained model has an inference time of less than 40 ms, thus available for real-time tracking at 15 FPS. Tracking accuracy was tested for 10 patients (five lung and five liver) and found to have an average tracking error less than 2 mm. The drawback, as the authors pointed out, is that it was trained with treatment-planning 4DCT data hence may not be able to capture changes between simulation and treatment [139].

Methods based on image synthesis: Image synthesis became popular with the success of GAN [171]. The remarkable ability of GAN for super-resolution [227] or style transfer [228] fits naturally with the wish of medical physicists to improve image quality or to match images from different modalities. For this reason, GAN became the indispensable component for numerous models [172–176]. In the context of image registration on a 2D projection domain, techniques now exist to synthesize volumetric information in less than 1 s from single projection images to assist with registration [205]. ResNet-based GANs have been employed to decompose kV images for enhanced registration for spine SBRT [206]. Fu et al. [207] built a patient-specific model to convert on-treatment KV projections into synthetic DRR images with enhanced tumor visibility for downstream on-treatment tumor tracking [208]. DRR generated from pre-treatment CBCT was also proposed to improve marker-less tracking for pancreas SBRT [209]. In the context of tracking the pancreas, Ahmed et al. published a contour prediction model based on patient-specific fine-tuning of a GAN-based population model [229]. The model was able to achieve a 95-percentile Hausdorff Distance (HD95) within 5 mm for 90% of the tests, and the inference speed is about 30 ms. Tracking based on prostate segmentation on KV projection was proposed by Mylonas et al. [204]. Yan et al. published tracking from fluoroscopy images from a color image intensifier using synthetic DRR generation [211,212].

Other methods: Besides the above three dominant approaches, the Recurrent Neural Network (RNN) has been used to predict tumor motion from transponder signals [213]. Siamese networks have introduced robust patient-specific similarity learning [214]. More recently, vision transformers have been incorporated for affine registration [215]. A Zero-shot learning framework proposed by Xu et al. combines a traditional image similarity measure with a pre-trained deep neural network for template matching [216]. Their method also provides an uncertain measure for the prediction.

4.3. Challenges of the Current Deep Learning Algorithm

Although there has been tremendous progress in applying deep learning to medical image registration [156–158], the application of these powerful tools to medicine carries uniquely high stakes. Unlike natural image classification, where an error amounts to a mislabeled photograph, radiation therapy is a mission-critical domain in which mistakes

translate directly to patient harm—underdosed tumors or overdosed organs at risk. This demands a standard of reliability far exceeding conventional performance benchmarks.

Deep learning approach in the present form has both advantages and disadvantages. One advantage of the deep learning solution is that the algorithm can be adapted to incorporate domain knowledge through training examples, unlike conventional algorithms, which use all-purpose image similarity metrics not specialized for the underlying tasks. In addition, deep learning models allow registration parameters to be directly predicted through a feed-forward neural network, resulting in an algorithm superior in speed compared to traditional methods that involve iterative optimization. A disadvantage of deep learning, however, is that a feed-forward neural network is opaque in what features are used and provides no explanations or uncertainty measure of its prediction. This black box approach creates a trust barrier that hinders the wide adoption of the technology. Conventional algorithms, on the other hand, are more transparent in this aspect, as the numerical value of the similarity metric provides its “explanation” for the prediction.

At the heart of the deep learning approach is the notion of generalizability [230]. The hope is that a model trained on a finite dataset can perform sufficiently well on unseen data. Because medical and natural images vary greatly at the pixel level—due to differences in viewpoint, exposure, detector noise, and artifacts—a model that generalizes well must capture high-level abstractions while remaining robust to irrelevant low-level variations. Numerous deep learning models have been proposed for image registration [156–158], but their robustness has not been thoroughly studied. Some intriguing aspects about deep learning raise concerns about the intelligence of these algorithms.

Adversarial Examples: It is a surprising failure mode where insignificant noises can cause a well-trained deep learning model to fail [231,232]. Adversarial Examples raise questions about whether deep learning algorithms have gained the required “intelligence” to perform mission-critical tasks such as image registration for stereotactic body radiation therapy (SBRT). Given the fact that deep neural networks have achieved superhuman performance on image classification tasks [152], it is tempting to assume that these models have acquired perceptual features analogous to those used by humans. Thus, it came as a surprise when adversarial examples were first discovered [231]. The authors found that input altered by negligible noises can cause a machine learning model to fail on tasks that were initially performed correctly. These can be considered as optical illusions for an AI system. Often, such attacking noises are insignificant and imperceptible to human observers. Nevertheless, it has been shown that adversarial examples exist in all popular modern deep learning vision systems. In a recent study, Jo et al. discovered that convolutional neural networks tend to learn superficial statistical clues rather than high-level abstracts [233]. In their experiment, a well-trained convolutional neural network generalizes poorly to natural images smoothed by a Fourier filter, losing up to 28% of its generalization accuracy. In another study, Azulay et al. found that a one-pixel shift or one-pixel rescaling of an image can result in a dramatic change in a network’s output [234]. Adversarial examples are also present in diagnostic AI and image segmentation models [235–239]. For example, Ozbulak et al. showed that a genuine breast image, initially classified as 95% cancerous by a deep learning model, can be perturbed by small adversarial noises, resulting in a prediction at the other extreme—classifying the tissue as healthy with 99% confidence [236]. In an experiment on brain tumor segmentation over MRI images, Cheng et al. found scenarios where adversarial examples cause significant drops in contouring accuracy for a UNet-based deep neural network [237].

Hallucination: Another aspect that raises concern about the robustness of AI-based algorithms is the phenomenon called hallucination. AI hallucination refers to the phenomenon where artificial intelligence systems generate outputs that appear plausible and

confident but are factually incorrect, fabricated, or unsupported by the input data or reality. In large language models (LLMs), hallucinations manifest as fluent, coherent text containing false information, fabricated citations, or statements that contradict established facts—often delivered with unwarranted confidence [240,241]. In medical imaging and computer vision, hallucinations are well documented and present as false structures, phantom lesions, or anatomical features that appear visually realistic but do not exist in the ground truth, potentially compromising diagnostic accuracy [242–246]. The underlying causes of hallucinations are not completely understood: they may arise from biased or insufficient training data, the intrinsic probabilistic nature of deep learning models, limited understanding of context or visual features, or the model's tendency to extrapolate beyond its training distribution. Regardless of the application domain, hallucinations share a common characteristic—they produce outputs that are plausible but incorrect, posing significant risks in high-stakes applications such as healthcare, legal, and scientific contexts where accuracy is paramount.

Peculiar Data Augmentation Requirement: Analyzing the way deep neural networks for image registration were trained also reveals some intriguing aspects that do not represent the robustness of human intelligence. For example, for regression-based registration models, the networks are typically trained with a pair of fixed and moving images and asked to produce the registration parameters (translations, rotations, etc.) that would bring the fixed and moving images into alignment. For such a feed-forward network, one finds that the model trained using pairs of input images with relative shifts up to 10 pixels will often make large test-time errors on input images with larger shifts, e.g., 15 pixels. The lack of geometric generalizability raises questions as to whether this type of network has learned to identify relevant high-level features to perform reliable registrations. Currently, this type of neural network needs to be trained on a dataset augmented significantly to cover all possible geometric transformations, leading to a training set that grows polynomially with the range of registration parameters. For instance, for a 3D registration network that predicts six parameters (three shifts and three angles), doubling the range of prediction requires a 64-fold augmentation of the training set. To avoid overfitting, the neural network model must be made larger (in terms of the number of parameters). Hence, the model must be trained for a much longer time, all to cover trivial geometric transformations that seem unnecessary to humans.

Lack of Geometric Stability: Testing whether registration error remains independent of initial misalignment represents one of the benchmarks for evaluating both the intelligence and the robustness of a registration algorithm. An algorithm exhibiting true translation-invariant behavior would demonstrate consistent accuracy regardless of the magnitude of initial misalignment—mirroring the human capacity for correspondence identification across arbitrary displacements. To date, however, no method has consistently demonstrated this capability. Most deep learning-based algorithms exhibit increasing registration error as initial misalignment grows, a pattern distressingly similar to the behavior of classical optimization-based approaches. This vulnerability suggests that despite architectural advances, current methods have not fundamentally solved the geometric robustness problem inherent to registration in the projection domain. Achieving reliable, initial-alignment-independent performance remains an open question facing 2D/3D image registration on the projection domain.

The peculiarities of the aforementioned aspects in current AI-based algorithms, and the fact that despite numerous marker-less tracking models published in the literature, robust real-world clinical implementations remain rare, underscore the inherent challenges of current deep learning approaches. Robust medical image analysis is intrinsically predicated on a 3D understanding of the physical world. This is particularly evident in the interpretation of 2D projection images; human experts navigate this data by leveraging an

implicit 3D mental model, which provides the cognitive basis for anatomically decoupling overlapping tissues even when they appear superimposed in the 2D plane. While the prevailing success of artificial intelligence in the industry is currently exemplified by large language models (LLMs), their underlying methodology does not translate seamlessly to the visual domain. LeCun has emphasized that the dimensionality and variance of image pixel space are astronomical compared to those of discrete language tokens [247]. Whereas LLMs perform stochastic prediction over a finite vocabulary—effectively navigating a massive probability table—the infinite variations in pixel intensities render such a “next-pixel” prediction approach computationally and theoretically intractable for computer vision. The pursuit of “true intelligence” in medical imaging hinges on the underlying representation of the images. Balestrierio et al. showed that learning paradigms prioritizing pixel-level reconstruction accuracy may be counterproductive, as the pursuit of superficial fidelity can degrade the underlying semantic representations learned by the model [248]. From this perspective, a GAN, ViT, or DDPM-based model developed with a loss function modified to improve pixel-level precision for the reconstructed images would be fundamentally limited in capturing the true representation of the 3D model. To capture the authentic variance of the physical world, LeCun advocates for a fundamental architectural shift toward world models—systems that perform prediction in latent space rather than raw pixel space [249–251]. Similarly, Li maintains that a foundational breakthrough yet to happen in AI is sophisticated 3D reasoning [252]. Without these advancements, deep learning models remain limited in their capacity to internalize the complex, 3D physical constraints requisite for high-stakes medical interventions.

4.4. Emerging Directions

Several directions are emerging in the field of AI and may lead to addressing the limitations of conventional deep learning in medical image registration, offering potential pathways toward more robust and clinically deployable solutions.

4.4.1. Patient-Specific Models

The inherent difficulty in training a universal model capable of generalizing across an entire patient population is evidenced by the proliferation of patient-specific models instead for motion management in the recent literature [141,202,203,207,209,210,253,254]. Given the nearly infinite variations in image morphology—including tumor size, texture, and shape—a population-level model may exceed the capacity of current deep learning architectures [247]. Consequently, patient-specific approaches are favored, as they operate under a significantly reduced learning burden by constraining the domain of visual variance to an individual’s unique anatomy. However, critical concerns persist regarding the clinical implementation of these models, particularly concerning temporal generalizability and anatomical drift; if a model is trained exclusively on planning CT images, it may fail to generalize to inter-fraction anatomical changes occurring by the day of treatment. This lack of generalizability to “on-the-day” anatomy remains a pervasive hurdle for robust motion tracking. To mitigate these discrepancies, the ideal approach would be training a treatment fraction-specific model based on images acquired immediately prior to delivery (e.g., daily CBCT). The primary barrier of this approach is the high computational cost of training. Current deep learning models typically require hours or days of GPU processing—resources often limited in a clinical hospital environment, making the training of sophisticated patient-specific registration models within minutes a major technical challenge. This necessitates the development of entirely new, rapidly adaptable model architectures.

In a runtime-trained system, the model weights used to treat a given patient are generated minutes before beam-on and are unique to that patient. Therefore, they cannot

be pre-approved as a fixed algorithm under the current Pharmaceuticals and Medical Devices Agency (PMDA) or Food and Drug Administration (FDA) frameworks. We expect that clinical translation will require a regulatory model analogous to treatment-planning QA in radiation oncology: the training procedure and its verification harness are cleared once, but each patient's model undergoes a per-patient acceptance test signed off by a qualified medical physicist before treatment. This acceptance test should produce human-judgeable artifacts, such as a predicted-versus-actual tumor overlay on held-out fluoroscopy, explainability maps, and uncertainty estimations. Responsibility for the per-patient model would then be shared among the device manufacturer, the institution, and the certifying physicist. A detailed regulatory framework for this paradigm will need to be developed collaboratively before routine clinical use.

4.4.2. Incorporating Explainability

For AI to gain trust from human users, it must be able to explain the underlying reason for its prediction. This provides a mechanism for clinicians to assess whether a model's decision aligns with sound clinical reasoning and to identify cases where the algorithm may be relying on spurious correlations rather than meaningful features. Explainability has become a central topic of discussion as AI algorithms are increasingly adopted in daily life and high-stakes decision-making [255–258]. Regulatory frameworks, including the European Union's General Data Protection Regulation (GDPR) and AI Act, have codified a "right to explanation" [255] for automated decisions that significantly affect individual—underscoring that transparency is not merely a technical desideratum but a legal and ethical imperative. Equally important is the capacity of an AI algorithm to provide a confidence level or uncertainty measure, particularly for deployment in healthcare settings where erroneous predictions carry significant harm [259–262]. The current incarnation of deep learning architectures, however, has overwhelmingly emphasized point-estimate prediction accuracy while neglecting uncertainty quantification. Most feed-forward neural networks produce deterministic outputs without any indication of whether a given prediction lies within the model's domain of competence or represents extrapolation into unfamiliar territory. This omission is problematic: a model may output a confident-appearing prediction even when the input deviates substantially from the training distribution, providing no warning to the end user. In clinical practice, such silent failures can be catastrophic. A robust AI system should not only provide accurate predictions but also "know what it does not know"—flagging cases of high epistemic uncertainty for human review rather than presenting all outputs with false equivalence. Achieving this dual capacity for accurate prediction and reliable uncertainty estimation represents a prerequisite for the trustworthy integration of AI into real-world clinical workflows.

Several efforts have been made to incorporate uncertainty estimation into medical image-registration frameworks. Within the context of traditional statistical learning, Risholm et al. introduced a non-rigid Bayesian registration framework that estimates the posterior distribution on both deformation and elastic parameters, enabling visualization of registration uncertainty for neurosurgical guidance [263]. Le Folgoc et al. investigated uncertainty quantification under a sparse Bayesian model, implementing reversible jump Markov Chain Monte Carlo sampling to characterize the posterior distribution of transformations [264]. With the advent of deep learning, Dalca et al. developed VoxelMorph-diff, a probabilistic diffeomorphic registration framework built on convolutional neural networks capable of providing uncertainty estimates alongside registration outputs [265]. Khawaled et al. proposed NPBDREG, a fully non-parametric Bayesian framework employing stochastic gradient Langevin dynamics to characterize the posterior distribution without restrictive Gaussian assumptions [266]. Gong et al. introduced an uncertainty

learning approach for unsupervised deformable registration that predicts registration uncertainty guided by reconstruction error [267]. More recently, Rivetti et al. developed a CNN-based model capable of jointly performing deformable image registration while predicting its associated uncertainty [268], and Xu et al. proposed models that incorporate learned similarity measures with uncertainty prediction [216]. Despite these advances, the capacity to provide reliable uncertainty estimates remains underdeveloped across most registration algorithms. A continued emphasis on equipping models with principled uncertainty quantification is essential, not only for enabling rigorous robustness evaluation but also for fostering the clinical trust and regulatory acceptance necessary for widespread adoption in mission-critical applications such as image-guided radiation therapy.

Beyond its role in building clinical trust, explainability can serve as a concrete enabler of regulatory approval for AI-driven FGPT systems. Saliency and attention maps allow regulators to verify that a deep learning marker-less tracker attends to anatomically meaningful structures such as the diaphragm contour and the tumor mass—rather than spurious correlations, thereby strengthening the equivalence argument when existing template-matching systems serve as the predicate device. Concept architectures, in which predictions are routed through human-interpretable intermediate variables such as diaphragm position or tumor centroid, enable performance specifications to be written in clinically meaningful terms rather than as opaque input–output behavior. Counterfactual analyses identifying the minimum input change that would alter a gating decision support systematic hazard identification by directly probing the decision boundary, connecting naturally to the adversarial robustness concerns discussed above. Uncertainty quantification provides the per-prediction confidence signal needed to operationalize selective-prediction safety nets: the system can abstain from autonomous gating and defer to the operator when confidence falls below a validated threshold. Together, these XAI techniques transform explainability from a desirable property into a practical toolkit for navigating the regulatory pathway toward clinically deployable marker-less FGPT.

A fully specified stress-test protocol for AI-based FGPT is beyond the scope of this review and will require a dedicated standardization effort. We propose, however, that any such protocol should encompass at least the following four categories of testing. (i) Geometric robustness testing, in which the model is challenged with controlled perturbations of the input geometry—initial-position offsets, rotations, scaling, and image translations—to characterize how tracking accuracy degrades with departures from the planning-time baseline. (ii) Image-quality robustness testing, in which the model is challenged with reduced fluoroscopic exposure (lower mAs and pulse width), increased Poisson and electronic noise, scatter and beam-hardening artifacts, anatomical occluders such as implanted hardware, and acquisition characteristics representative of vendors and institutions not seen during training, in order to characterize generalization beyond the training distribution. (iii) Adversarial robustness testing for the deep neural network itself, including both worst-case adversarial perturbations bounded and clinically realistic distractors such as decoy high-contrast structures (ribs, spinal hardware, surgical clips). The goal is to quantify the gap between in-distribution accuracy and behavior under inputs that are visually plausible but pathologically constructed. (iv) Explainability testing, in which post hoc explanations of the model's predictions (e.g., saliency or attribution maps) are evaluated for consistency under repeated and slightly perturbed inputs, for fidelity to the model's actual decision pathway, and for plausibility against clinician-annotated regions of interest, to guard against models that achieve apparent accuracy by relying on anatomically irrelevant features. We anticipate that future community standards will define quantitative pass thresholds for each category.

4.4.3. Emerging Deep Learning Paradigm

While patient-specific models may offer a pragmatic solution to immediate clinical needs, real change needs to come from a more intelligent model that is able to capture the underlying representation. Some new models occurred in the industry serve as the next candidates for medical physicists to develop more robust algorithms.

Geometric Deep Learning: The historic success of convolutional neural networks (CNNs) is largely attributable to their built-in translational equivariance, where shifting an input causes a corresponding shift in feature maps, enabling efficient weight sharing [269]. However, standard CNNs are inherently limited by their inability to respect more complex geometric structures, such as rotations, scalings, or the non-Euclidean curved geometry of anatomical surfaces. Geometric deep learning addresses these limitations by extending neural architectures to manifolds, graphs, and point clouds, encoding arbitrary symmetries directly into the network design [269–272]. Group equivariant neural networks, for instance, guarantee that outputs transform predictably under various geometric operations, which may help to build a principled framework for image registration that respects geometric constraints. Suliman et al. applied geometric deep learning to surface registration [273]. Greer et al. published a registration framework based on geometric deep learning principles with a coordinate attention mechanism [274]. Graph Neural Networks (GNNs) represent a parallel advancement in this domain by treating anatomical structures as nodes within a connected graph, allowing for the explicit modeling of spatial relationships and biomechanical constraints. This was exemplified by the work of Shao et al., who utilized GNNs for real-time liver tumor localization from sparse X-ray projections [275]. By performing node-based perceptual feature pooling on a liver surface mesh, their model predicts boundary DVFs, which then drive a biomechanical model to estimate internal tumor motion. The model is applied to liver motion estimation.

World Models: Beyond specialized geometric priors, a broader paradigm shift is occurring through the development of world models designed to learn general-purpose representations of physical reality rather than task-specific mapping functions. In the context of 3D understanding, Neural Radiance Fields (NeRF) have introduced continuous volumetric functions capable of synthesizing novel views from sparse observations [276]. This move toward “spatial intelligence” is being spearheaded by researchers at the World Labs, aiming to build large-scale world models capable of perceiving and reasoning about 3D environments. Their Marble platform, which generates spatially consistent 3D worlds from 2D inputs, suggests a future where medical simulations are driven by models that possess an inherent understanding of 3D space. Such architecture offers a path toward medical AI that can conceptualize the complex, 3D physical objects in the human body. Adaptations of NeRF have been used in medical image synthesis. MedNeRF has shown that these continuous representations can reconstruct CT-like volumes from limited X-ray projections by disentangling surface shape from internal depth [277]. On parallel development, Gaussian splatting has emerged as a high-speed alternative that represents scenes via explicit 3D Gaussian primitives, enabling rapid sparse-view CT reconstruction and multimodal surface modeling [278–280]. Another approach toward world models, advocated by LeCun, is the Joint Embedding Predictive Architecture (JEPA) that provides a foundational framework for autonomous machine intelligence that learns world models through passive observation [247]. Unlike generative models that struggle with the high dimensionality of raw pixels, JEPA operates in a restricted representation space, predicting the “latent state” of missing or future observations. Implementations such as I-JEPA [249] and V-JEPA [250,281] have demonstrated that this latent-space reasoning allows models to focus on semantically meaningful features while ignoring irrelevant noise [249]. By pre-training on massive video

datasets, these models achieve state-of-the-art motion understanding and can even transfer learned representations to zero-shot active control tasks [250].

4.4.4. Inference Time Optimization Using Energy-Based Model (EBM)

Most contemporary deep learning approaches for medical image registration rely on feed-forward neural networks to directly regress displacement fields or registration parameters in a single, constant-time pass. However, these static architectures are often ill-suited for the high-dimensional variability of clinical imagery. EBMs represent a framework designed as a strategy to bypass the limitations of traditional statistical learning [282]. While standard machine learning is often constrained by the requirement to model normalized probability distributions—a task that some argued to have “handicapped” machine learning—the EBM approach discards probabilistic normalization. Instead, the model learns a scalar energy function trained to assign minimum energy values to predictions that are compatible with observed reality and higher energy to those that are not. A critical advantage of EBMs is their capacity for inference-time optimization. Unlike traditional non-EBMs that produce outputs via a fixed computational budget, EBMs iteratively search for the energy minimum. This process mirrors human expert behavior in image registration; a radiologist or medical physicist does not take a constant time to align complex images. Instead, they devote a variable, indefinite amount of time to “reasoning” through difficult cases where anatomy is significantly deformed or occluded. This mechanism aligns with the cognitive transition from “System 1” (fast, intuitive) to “System 2” (slow, deliberate) thinking [283]. By shifting from rigid, one-pass regression to an iterative energy-minimization paradigm, EBMs may allow the model to handle wider variations and anomalies. This flexibility ensures that the output is not just a statistical guess but an optimized solution, ultimately bringing significantly more robustness to real-time clinical workflows.

4.4.5. Hardware Improvements

Beyond algorithmic advances, hardware innovations offer complementary pathways to improved tumor localization. Dual-energy (DE) imaging exploits the energy-dependent attenuation of X-rays to decompose images into tissue-specific components, enabling bone suppression that dramatically improves soft-tissue tumor visibility in fluoroscopic tracking. Dhont et al. provided foundational work on dual-energy CT applications in radiotherapy, demonstrating improved electron density estimation and tissue characterization for dose calculation accuracy [284]. Menten et al. demonstrated that dual-energy imaging could enhance automated lung tumor tracking for real-time adaptive radiotherapy by generating radiographs with reduced bone visibility, improving tracking success from 90.7% in single-energy images to 99.9% in dual-energy frames [285]. More recently, Haytmyradov et al. developed a benchtop fast-kV switching dual-energy fluoroscopy system for marker-less tumor tracking, demonstrating the feasibility of real-time DE imaging on clinical linear accelerators with on-board imaging systems [286]. These hardware-based approaches circumvent many limitations of software-only bone suppression algorithms by directly acquiring the physics-based information needed for tissue decomposition, though they require specialized imaging equipment and careful optimization of energy pairs and weighting factors.

Imaging hardware is only one half of the hardware story for real-time marker-less AI-based FGPT; the other half is the computing hardware that runs the deep network on every fluoroscopic frame within the system’s hard latency budget. AAPM Task Group 264 defines real-time system latency in radiotherapy as ≤ 500 ms [287], and the deployed Hitachi RGPT precedent achieves a total system latency of approximately 100 ms [76,102,105]. At the 15–30 fps frame rates typical of FGPT fluoroscopy, the per-frame interval is 33–67 ms;

after image acquisition, transmission, and beam-control signaling are accounted for, the budget remaining for the deep-network forward pass is on the order of 200–300 ms. Closing this gap requires GPU and AI-accelerator hardware deployed at the gantry and inference-runtime optimization. A separate compute concern arises in the patient-specific paradigm proposed in Section 4.4.1: per-patient model training must compress from the hours-to-days typical of research workflows to minutes that fit between simulation and treatment, which adds a high-throughput on-site training infrastructure to the clinical equipment list. Computing is, therefore, not a secondary engineering detail but a primary determinant of which algorithmic ideas will reach a treatment vault.

5. Conclusions

Fluoroscopy-guided motion management is essential for realizing the full dosimetric advantages of particle therapy. This review has traced the evolution from image intensifiers to flat-panel detectors, which have enabled FGPT solutions such as Hitachi's RGPT. Marker-based tracking has demonstrated clinical feasibility, but the invasiveness of fiducial implantation, associated complications, and dosimetric perturbations limit broader adoption. Marker-less tracking—directly localizing tumors from native anatomical contrast—remains the ideal but technically challenging solution due to the superposition of 3D anatomy onto two-dimensional projections. Deep learning offers the most promising path forward, with numerous architectures proposed for tumor localization in projection images. However, robust clinical translation remains elusive due to concerns about generalizability, adversarial susceptibility, and potential for hallucinated outputs—failure modes unacceptable in radiation therapy.

The path toward reliable, clinically deployable marker-less FGPT will likely require a convergence of advances: continued refinement of deep learning architectures with built-in geometric and physical constraints, hardware innovations such as dual-energy imaging for improved soft-tissue contrast, rigorous validation frameworks that test robustness across the full spectrum of clinical variability, and regulatory pathways that address the unique challenges of AI-driven real-time treatment adaptation. As particle therapy continues to expand globally, solving the marker-less tracking challenge will be essential to unlocking the full therapeutic potential of this technology for patients with thoracic and abdominal malignancies.

Author Contributions: Conceptualization, F.L., K.M.F. and C.J.B.; investigation, F.L., K.M.F. and C.J.B.; writing—original draft preparation, F.L.; writing—review and editing, F.L., K.M.F. and C.J.B.; visualization, F.L.; supervision, K.M.F. and C.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AC	Abdominal Compression
ADC	Analog-to-Digital Converter
AI	Artificial Intelligence
a-Se	Amorphous Selenium
a-Si	Amorphous Silicon
a-Si:H	Hydrogenated Amorphous Silicon
BH	Breath Hold
CaW	Calcium Tungstate
CBCT	Cone Beam CT
CCD	Charge-Coupled Devices
CCTV	Closed-Circuit TV
CNN	Convolutional Neural Network
CsI	Cesium-Iodide
CsI:Na	Sodium activated cesium iodide
CsI:Tl	Thallium-Doped Cesium Iodide
CTV	Clinical Target Volume
DDCS	Dose Driven Continuous Scanning
DDPM	Denoising Diffusion Probabilistic Models
DE	Dual-Energy
DFPD	Dynamic Flat-Panel Detector
DM	Dose Monitor
DQE	Detective Quantum Efficiency
DRR	Digitally Reconstructed Radiograph
DVF	Deformation Vector Field
EBM	Energy Based Model
FCN	Fully Convolutional Network
FGPT	Fluoroscopy-Guided Particle Therapy
FMEA	Failure Modes and Effects Analysis
FOV	Field of View
FPD	Flat-Panel Detector
GAN	Generative Adversarial Networks
GNN	Graph Neural Network
HCL	Harvard Cyclotron Laboratory
HD95	95-Percentile Hausdorff Distance
HU	Hounsfield Unit
ICD	Interrupted Continuous Delivery
II	Image Intensifier
ITV	Internal Target Volume
JEPA	Joint Embedding Predictive Architecture
kV	Kilovoltage
Linac	Linear Accelerator
LLM	Large Language Model
MEE	Multiple Energy Extraction
MI	Mutual Information
MU	Monitor Units
MV	Megavoltage
NCC	Normalized Cross-Correlation
NeRF	Neural Radiance Fields
NIRS	National Institute of Radiological Sciences
PBS	Pencil Beam Scanning
PCR	Phase Controlled Rescanning

PS	Passive Scattering
PSI	Paul Scherrer Institute
PTV	Planning Target Volume
RGPT	Real-Time Gated Particle Therapy
RGSC	Respiratory Gating for Scanners
RNN	Recurrent Neural Network
RPM	Real-Time Position Management
RTRT	Real-Time Tumor-Tracking Radiotherapy
SBRT	Stereotactic Body Radiation Therapy
SEE	Single Energy Extraction
SEER	Surveillance, Epidemiology, and End Results
SGRT	Surface Guided Radiation Therapy
SOBP	Spread-Out Bragg Peak
SSD	Sum of Squared Differences
STN	Spatial Transformer Network
TCP	Tumor Control Probability
TFT	Thin-Film Transistor
TPM	Tumor Probability Map
TRE	Tracking Registration Error
ViT	Vision Transformer
WEPL	Water Equivalent Path Length
ZnCdS	Zinc-Cadmium-Sulfide
ZnCdS:Ag	Silver-Activated Zinc-Cadmium-Sulfide

References

1. Wilson, R.R. Radiological use of fast protons. *Radiology* **1946**, *47*, 487–491. [CrossRef]
2. Gragoudas, E.S.; Goitein, M.; Koehler, A.M.; Verhey, L.; Tepper, J.; Suit, H.D.; Brockhurst, R.; Constable, I.J. Proton irradiation of small choroidal malignant melanomas. *Am. J. Ophthalmol.* **1977**, *83*, 665–673. [CrossRef] [PubMed]
3. Spatola, C.; Liardo, R.L.E.; Milazzotto, R.; Raffaele, L.; Salamone, V.; Basile, A.; Foti, P.V.; Palmucci, S.; Cirrone, G.A.P.; Cuttone, G.; et al. Radiotherapy of conjunctival melanoma: Role and challenges of brachytherapy, photon-beam and protontherapy. *Appl. Sci.* **2020**, *10*, 9071. [CrossRef]
4. Milazzotto, R.; Liardo, R.L.E.; Privitera, G.; Raffaele, L.; Salamone, V.; Arena, F.; Pergolizzi, S.; Cuttone, G.; Cirrone, G.A.P.; Russo, A.; et al. Proton beam radiotherapy of locally advanced or recurrent conjunctival squamous cell carcinoma: Experience of the catana centre. *J. Radiother. Pract.* **2022**, *21*, 97–104. [CrossRef]
5. PTCOG. Available online: <https://ptcog.online/facilities-world-map> (accessed on 2 March 2026).
6. Mori, S.; Zenklusen, S.; Knopf, A.C. Current status and future prospects of multi-dimensional image-guided particle therapy. *Radiol. Phys. Technol.* **2013**, *6*, 249–272. [CrossRef]
7. Yu, Z.H.; Lin, S.H.; Balter, P.; Zhang, L.; Dong, L. A comparison of tumor motion characteristics between early stage and locally advanced stage lung cancers. *Radiother. Oncol.* **2012**, *104*, 33–38. [CrossRef] [PubMed]
8. Mori, S.; Chen, G.T.Y.; Endo, M. Effects of intrafractional motion on water equivalent pathlength in respiratory-gated heavy charged particle beam radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2007**, *69*, 308–317. [CrossRef]
9. Langen, K.M.; Jones, D.T.L. Organ motion and its management. *Int. J. Radiat. Oncol. Biol. Phys.* **2001**, *50*, 265–278. [CrossRef]
10. Webb, S. Motion effects in (intensity modulated) radiation therapy: A review. *Phys. Med. Biol.* **2006**, *51*, R403. [CrossRef]
11. Keall, P.J.; Mageras, G.S.; Balter, J.M.; Emery, R.S.; Forster, K.M.; Jiang, S.B.; Kapatoes, J.M.; Low, D.A.; Murphy, M.J.; Murray, B.R.; et al. The management of respiratory motion in radiation oncology report of AAPM task group 76a. *Med. Phys.* **2006**, *33*, 3874–3900. [CrossRef] [PubMed]
12. Bert, C.; Durante, M. Motion in radiotherapy: Particle therapy. *Phys. Med. Biol.* **2011**, *56*, R113–R144. [CrossRef]
13. Mori, S.; Knopf, A.C.; Umegaki, K. Motion management in particle therapy. *Med. Phys.* **2018**, *45*, e994–e1010. [CrossRef] [PubMed]
14. Pakela, J.M.; Knopf, A.; Dong, L.; Rucinski, A.; Zou, W. Management of motion and anatomical variations in charged particle therapy: Past, present, and into the future. *Front. Oncol.* **2022**, *12*, 806153. [CrossRef] [PubMed]
15. Pedroni, E.; Bacher, R.; Blattmann, H.; Böhringer, T.; Coray, A.; Lomax, A.; Lin, S.; Munkel, G.; Scheib, S.; Schneider, U.; et al. The 200-mev proton therapy project at the paul scherrer institute: Conceptual design and practical realization. *Med. Phys.* **1995**, *22*, 37–53. [CrossRef] [PubMed]

16. Li, H.; Dong, L.; Bert, C.; Chang, J.; Flampouri, S.; Jee, K.-W.; Lin, L.; Moyers, M.; Mori, S.; Rottmann, J.; et al. AAPM task group report 290: Respiratory motion management for particle therapy. *Med. Phys.* **2022**, *49*, e50–e81. [\[CrossRef\]](#)
17. Paganetti, H. *Proton Therapy Physics*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2025.
18. ICRU. *Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU Report 50)*; ICRU: Bethesda, MD, USA, 1999.
19. Haberer, T.; Becher, W.; Schardt, D.; Kraft, G. Magnetic scanning system for heavy ion therapy. *Nucl. Instrum. Methods Phys. Res. A* **1993**, *330*, 296–305. [\[CrossRef\]](#)
20. Zenklusen, S.M.; Pedroni, E.; Meer, D. A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new gantry 2 at psi. *Phys. Med. Biol.* **2010**, *55*, 5103–5121. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Shen, J.; Tryggestad, E.; Younkin, J.E.; Keole, S.R.; Furutani, K.M.; Kang, Y.; Herman, M.G.; Bues, M. Technical note: Using experimentally determined proton spot scanning timing parameters to accurately model beam delivery time. *Med. Phys.* **2017**, *44*, 5081–5088. [\[CrossRef\]](#)
22. Rietzel, E.; Bert, C. Respiratory motion management in particle therapy. *Med. Phys.* **2010**, *37*, 449–460. [\[CrossRef\]](#)
23. Dowdell, S.; Grassberger, C.; Sharp, G.C.; Paganetti, H. Interplay effects in proton scanning for lung: A 4d monte carlo study assessing the impact of tumor and beam delivery parameters. *Phys. Med. Biol.* **2013**, *58*, 4137–4156. [\[CrossRef\]](#)
24. Yock, A.D.; Mohan, R.; Flampouri, S.; Bosch, W.; Taylor, P.A.; Gladstone, D.; Kim, S.; Sohn, J.; Wallace, R.; Xiao, Y.; et al. Robustness analysis for external beam radiation therapy treatment plans: Describing uncertainty scenarios and reporting their dosimetric consequences. *Pract. Radiat. Oncol.* **2019**, *9*, 200–207. [\[CrossRef\]](#)
25. Knopf, A.; Bert, C.; Heath, E.; Nill, S.; Kraus, K.; Richter, D.; Hug, E.; Pedroni, E.; Safai, S.; Albertini, F.; et al. Special report: Workshop on 4d-treatment planning in actively scanned particle therapy—Recommendations, technical challenges, and future research directions. *Med. Phys.* **2010**, *37*, 4608–4614. [\[CrossRef\]](#)
26. Knopf, A.; Nill, S.; Yohannes, I.; Graeff, C.; Dowdell, S.; Kurz, C.; Sonke, J.-J.; Biegun, A.K.; Lang, S.; McClelland, J.; et al. Challenges of radiotherapy: Report on the 4d treatment planning workshop 2013. *Phys. Med.* **2014**, *30*, 809–815. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Trnková, P.; Knäusl, B.; Actis, O.; Bert, C.; Biegun, A.K.; Boehlen, T.T.; Furtado, H.; McClelland, J.; Mori, S.; Rinaldi, I.; et al. Clinical implementations of 4d pencil beam scanned particle therapy: Report on the 4d treatment planning workshop 2016 and 2017. *Phys. Med.* **2018**, *54*, 121–130. [\[CrossRef\]](#)
28. Nohadani, O.; Seco, J.; Bortfeld, T. Motion management with phase-adapted 4d-optimization. *Phys. Med. Biol.* **2010**, *55*, 5189–5202. [\[CrossRef\]](#)
29. Kardar, L.; Li, Y.; Li, X.; Li, H.; Cao, W.; Chang, J.Y.; Liao, L.; Zhu, R.X.; Sahoo, N.; Gillin, M.; et al. Evaluation and mitigation of the interplay effects of intensity modulated proton therapy for lung cancer in a clinical setting. *Pract. Radiat. Oncol.* **2014**, *4*, e259–e268. [\[CrossRef\]](#)
30. Graeff, C. Motion mitigation in scanned ion beam therapy through 4d-optimization. *Phys. Med.* **2014**, *30*, 570–577. [\[CrossRef\]](#)
31. Liu, W.; Schild, S.E.; Chang, J.Y.; Liao, Z.; Chang, Y.-H.; Wen, Z.; Shen, J.; Stoker, J.B.; Ding, X.; Hu, Y.; et al. Exploratory study of 4d versus 3d robust optimization in intensity modulated proton therapy for lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *95*, 523–533. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zhang, Y.; Huth, I.; Wegner, M.; Weber, D.C.; Lomax, A.J. An evaluation of rescanning technique for liver tumour treatments using a commercial pbs proton therapy system. *Radiother. Oncol.* **2016**, *121*, 281–287. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Zhang, Y.; Huth, I.; Weber, D.C.; Lomax, A.J. A statistical comparison of motion mitigation performances and robustness of various pencil beam scanned proton systems for liver tumour treatments. *Radiother. Oncol.* **2018**, *128*, 182–188. [\[CrossRef\]](#)
34. Engwall, E.; Glimelius, L.; Hynning, E. Effectiveness of different rescanning techniques for scanned proton radiotherapy in lung cancer patients. *Phys. Med. Biol.* **2018**, *63*, 095006. [\[CrossRef\]](#)
35. Klimpki, G.; Zhang, Y.; Fattori, G.; Psoroulas, S.; Weber, D.C.; Lomax, A.; Meer, D. The impact of pencil beam scanning techniques on the effectiveness and efficiency of rescanning moving targets. *Phys. Med. Biol.* **2018**, *63*, 145006. [\[CrossRef\]](#)
36. Phillips, M.H.; Pedroni, E.; Blattmann, H.; Boehringer, T.; Coray, A.; Scheib, S. Effects of respiratory motion on dose uniformity with a charged particle scanning method. *Phys. Med. Biol.* **1992**, *37*, 223. [\[CrossRef\]](#)
37. Schätti, A.; Zakova, M.; Meer, D.; Lomax, A.J. Experimental verification of motion mitigation of discrete proton spot scanning by re-scanning. *Phys. Med. Biol.* **2013**, *58*, 8555. [\[CrossRef\]](#)
38. Li, Y.; Kardar, L.; Li, X.; Li, H.; Cao, W.; Chang, J.Y.; Liao, L.; Zhu, R.X.; Sahoo, N.; Gillin, M.; et al. On the interplay effects with proton scanning beams in stage iii lung cancer. *Med. Phys.* **2014**, *41*, 021721. [\[CrossRef\]](#)
39. Knopf, A.-C.; Hong, T.S.; Lomax, A. Scanned proton radiotherapy for mobile targets—The effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Phys. Med. Biol.* **2011**, *56*, 7257–7271. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Grassberger, C.; Dowdell, S.; Lomax, A.; Sharp, G.; Shackleford, J.; Choi, N.; Willers, H.; Paganetti, H. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2013**, *86*, 380–386. [\[CrossRef\]](#) [\[PubMed\]](#)

41. Furukawa, T.; Inaniwa, T.; Sato, S.; Tomitani, T.; Minohara, S.; Noda, K.; Kanai, T. Design study of a raster scanning system for moving target irradiation in heavy-ion radiotherapy. *Med. Phys.* **2007**, *34*, 1085–1097. [\[CrossRef\]](#)
42. Furukawa, T.; Inaniwa, T.; Sato, S.; Shirai, T.; Mori, S.; Takeshita, E.; Mizushima, K.; Himukai, T.; Noda, K. Moving target irradiation with fast rescanning and gating in particle therapy. *Med. Phys.* **2010**, *37*, 4874–4879. [\[CrossRef\]](#)
43. Mori, S.; Furukawa, T.; Inaniwa, T.; Zenklusen, S.; Nakao, M.; Shirai, T.; Noda, K. Systematic evaluation of four-dimensional hybrid depth scanning for carbon-ion lung therapy. *Med. Phys.* **2013**, *40*, 031720. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Andersson, K.M.; Edvardsson, A.; Hall, A.; Enmark, M.; Kristensen, I. Pencil beam scanning proton therapy of hodgkin's lymphoma in deep inspiration breath-hold: A case series report. *Tech. Innov. Patient Support Radiat. Oncol.* **2020**, *13*, 6–10. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Habermehl, D.; Debus, J.; Ganten, T.; Ganten, M.-K.; Bauer, J.; Brecht, I.C.; Brons, S.; Haberer, T.; Haertig, M.; Jäkel, O.; et al. Hypofractionated carbon ion therapy delivered with scanned ion beams for patients with hepatocellular carcinoma—Feasibility and clinical response. *Radiat. Oncol.* **2013**, *8*, 59. [\[CrossRef\]](#)
46. Gelover, E.; Deisher, A.J.; Herman, M.G.; Johnson, J.E.; Kruse, J.J.; Tryggestad, E.J. Clinical implementation of respiratory-gated spot-scanning proton therapy: An efficiency analysis of active motion management. *J. Appl. Clin. Med. Phys.* **2019**, *20*, 99–108. [\[CrossRef\]](#)
47. Hanley, J.; Debois, M.M.; Mah, D.; Mageras, G.S.; Raben, A.; Rosenzweig, K.; Mychalczak, B.; Schwartz, L.H.; Gloeggler, P.J.; Lutz, W.; et al. Deep inspiration breath-hold technique for lung tumors: The potential value of target immobilization and reduced lung density in dose escalation. *Int. J. Radiat. Oncol. Biol. Phys.* **1999**, *45*, 603–611. [\[CrossRef\]](#)
48. Seppenwoolde, Y.; Shirato, H.; Kitamura, K.; Shimizu, S.; van Herk, M.; Lebesque, J.V.; Miyasaka, K. Precise and real-time measurement of 3d tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2002**, *53*, 822–834. [\[CrossRef\]](#)
49. Cerviño, L.I.; Chao, A.K.Y.; Sandhu, A.; Jiang, S.B. The diaphragm as an anatomic surrogate for lung tumor motion. *Phys. Med. Biol.* **2009**, *54*, 3529. [\[CrossRef\]](#)
50. Mori, S.; Asakura, H.; Kandatsu, S.; Kumagai, M.; Baba, M.; Endo, M. Magnitude of residual internal anatomy motion on heavy charged particle dose distribution in respiratory gated lung therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2008**, *71*, 587–594. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Mori, S.; Takei, Y.; Shirai, T.; Hara, Y.; Furukawa, T.; Inaniwa, T.; Tanimoto, K.; Tajiri, M.; Kuroiwa, D.; Kimura, T.; et al. Scanned carbon-ion beam therapy throughput over the first 7 years at national institute of radiological sciences. *Phys. Med.* **2018**, *52*, 18–26. [\[CrossRef\]](#)
52. Riboldi, M.; Orecchia, R.; Baroni, G. Real-time tumour tracking in particle therapy: Technological developments and future perspectives. *Lancet Oncol.* **2012**, *13*, e383–e391. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Zhang, P.; Happersett, L.; Burleson, S.; Oh, J.H.; Elsayegh, A.; Leong, B.; Thor, M.; Damato, A.; Jackson, A.; Cervino, L.; et al. Reduction of postradiation therapy urinary toxicity via intrafractional megavoltage-kilovoltage prostate location monitoring. *Int. J. Radiat. Oncol. Biol. Phys.* **2025**, *121*, 261–268. [\[CrossRef\]](#)
54. Zhang, Y.; Trnkova, P.; Toshito, T.; Heijmen, B.; Richter, C.; Aznar, M.; Albertini, F.; Bolsi, A.; Daartz, J.; Bertholet, J.; et al. A survey of practice patterns for real-time intrafractional motion-management in particle therapy. *Phys. Imaging Radiat. Oncol.* **2023**, *26*, 100439. [\[CrossRef\]](#)
55. Bushberg, J.T. *The Essential Physics of Medical Imaging*, 4th ed.; Wolters Kluwer: Philadelphia, PA, USA, 2021.
56. Balter, S. Fluoroscopic technology from 1895 to 2019 drivers: Physics and physiology. *Med. Phys. Int.* **2019**, *7*, 111–140.
57. Lopez, P.D. Fluoroscopy history, evolution, and technological advancements: A narrative review. *J. Med. Imaging Radiat. Sci.* **2024**, *55*, 347–353. [\[CrossRef\]](#)
58. Seibert, J.A. Flat-panel detectors: How much better are they? *Pediatr. Radiol.* **2006**, *36*, 173–181. [\[CrossRef\]](#)
59. Roehrig, H.; Dallas, W.; Ovitt, T.; Lamoreaux, R.; Vercillo, R.; McNeill, K. *A High Resolution X-Ray Imaging Device*; SPIE: Bellingham, WA, USA, 1989.
60. Wang, J.; Blackburn, T.J. The AAPM/rsna physics tutorial for residents. *Radiographics* **2000**, *20*, 1471–1477. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Fujieda, I.; Nelson, S.; Street, R.A.; Weisfield, R.L. Radiation imaging with 2d a-si sensor arrays. In *Proceedings of the 1991 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSSMIC)*, Santa Fe, NM, USA, 2–9 November 1991; IEEE: Piscataway, NJ, USA, 1991; pp. 1882–1886. [\[CrossRef\]](#)
62. Antonuk, L.E.; Boudry, J.; Huang, W.; McShan, D.L.; Morton, E.J.; Yorkston, J.; Longo, M.J.; Street, R.A. Demonstration of megavoltage and diagnostic x-ray imaging with hydrogenated amorphous silicon arrays. *Med. Phys.* **1992**, *19*, 1455–1466. [\[CrossRef\]](#)
63. Schiebel, U.; Conrads, N.; Jung, N.; Weibrecht, M.; Wiecek, H.; Zaengel, T.; Powell, M.; French, I.; Glasse, C. *Fluoroscopic X-Ray Imaging with Amorphous Silicon Thin-Film Arrays*; SPIE: Bellingham, WA, USA, 1994.
64. Graeve, T.; Li, Y.; Fabians, A.; Huang, W. *High-Resolution Amorphous Silicon Image Sensor*; SPIE: Bellingham, WA, USA, 1996.

65. Antonuk, L.E.; Yorkston, J.; Huang, W.; Siewerdsen, J.H.; Boudry, J.M.; El-Mohri, Y.; Marx, M.V. A real-time, flat-panel, amorphous silicon, digital x-ray imager. *Radiographics* **1995**, *15*, 993–1000. [[CrossRef](#)]
66. Antonuk, L.E.; Yorkston, J.; Huang, W.; Sandler, H.; Siewerdsen, J.H.; El-Mohri, Y. Megavoltage imaging with a large-area, flat-panel, amorphous silicon imager. *Int. J. Radiat. Oncol. Biol. Phys.* **1996**, *36*, 661–672. [[CrossRef](#)] [[PubMed](#)]
67. Pisani, L.; Lockman, D.; Jaffray, D.; Yan, D.; Martinez, A.; Wong, J. Setup error in radiotherapy: On-line correction using electronic kilovoltage and megavoltage radiographs. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *47*, 825–839. [[CrossRef](#)]
68. Jaffray, D.A.; Siewerdsen, J.H.; Wong, J.W.; Martinez, A.A. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2002**, *53*, 1337–1349. [[CrossRef](#)] [[PubMed](#)]
69. Powell, M.J.; French, I.D.; Hughes, J.R.; Bird, N.C.; Davies, O.S.; Glasse, C.; Curran, J.E. Amorphous silicon image sensor arrays. *MRS Online Proc. Libr.* **1992**, *258*, 1127–1137. [[CrossRef](#)]
70. Hoheisel, M.; Arques, M.; Chabbal, J.; Chaussat, C.; Ducourant, T.; Hahm, G.; Horbaschek, H.; Schulz, R.; Spahn, M. Amorphous silicon x-ray detectors. *J. Non-Cryst. Solids* **1998**, *227–230*, 1300–1305. [[CrossRef](#)]
71. Orth, R.C.; Wallace, M.J.; Kuo, M.D. C-arm cone-beam ct: General principles and technical considerations for use in interventional radiology. *J. Vasc. Interv. Radiol.* **2008**, *19*, 814–820. [[CrossRef](#)]
72. Gupta, R.; Grasruck, M.; Suess, C.; Bartling, S.H.; Schmidt, B.; Stierstorfer, K.; Popescu, S.; Brady, T.; Flohr, T. Ultra-high resolution flat-panel volume ct: Fundamental principles, design architecture, and system characterization. *Eur. Radiol.* **2006**, *16*, 1191–1205. [[CrossRef](#)]
73. Baba, R.; Konno, Y.; Ueda, K.; Ikeda, S. Comparison of flat-panel detector and image-intensifier detector for cone-beam ct. *Comput. Med. Imaging Graph.* **2002**, *26*, 153–158. [[CrossRef](#)]
74. Ning, R.; Chen, B.; Yu, R.; Conover, D.; Tang, X.; Ning, Y. Flat panel detector-based cone-beam volume ct angiography imaging: System evaluation. *IEEE Trans. Med. Imaging* **2000**, *19*, 949–963. [[CrossRef](#)]
75. Granfors, P.R.; Aufrichtig, R.; Possin, G.E.; Giambattista, B.W.; Huang, Z.S.; Liu, J.; Ma, B. Performance of a amorphous silicon flat panel x-ray detector designed for angiographic and r&f imaging applications. *Med. Phys.* **2003**, *30*, 2715–2726. [[PubMed](#)]
76. Shimizu, S.; Miyamoto, N.; Matsuura, T.; Fujii, Y.; Umezawa, M.; Umegaki, K.; Hiramoto, K.; Shirato, H. A proton beam therapy system dedicated to spot-scanning increases accuracy with moving tumors by real-time imaging and gating and reduces equipment size. *PLoS ONE* **2014**, *9*, e94971. [[CrossRef](#)] [[PubMed](#)]
77. Suit, H.D.; Goitein, M.; Tepper, J.; Koehler, A.M.; Schmidt, R.A.; Schneider, R. Exploratory study of proton radiation therapy using large field techniques and fractionated dose schedules. *Cancer* **1975**, *35*, 1646–1657. [[CrossRef](#)]
78. Constable, I.J.; Roehler, A.M. Experimental ocular irradiation with accelerated protons. *Invest. Ophthalmol.* **1974**, *13*, 280–287. [[PubMed](#)]
79. Gragoudas, E.S. Proton irradiation of choroidal melanomas. *Arch. Ophthalmol.* **1978**, *96*, 1583. [[CrossRef](#)]
80. Verhey, L.J.; Goitein, M.; McNulty, P.; Munzenrider, J.E.; Suit, H.D. Precise positioning of patients for radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **1982**, *8*, 289–294. [[CrossRef](#)]
81. Leong, J. Use of digital fluoroscopy as an online verification device in radiation therapy. *Phys. Med. Biol.* **1986**, *31*, 985. [[CrossRef](#)]
82. Ohara, K.; Okumura, T.; Akisada, M.; Inada, T.; Mori, T.; Yokota, H.; Calaguas, M.J.B. Irradiation synchronized with respiration gate. *Int. J. Radiat. Oncol. Biol. Phys.* **1989**, *17*, 853–857. [[CrossRef](#)] [[PubMed](#)]
83. Minohara, S.; Kanai, T.; Endo, M.; Noda, K.; Kanazawa, M. Respiratory gated irradiation system for heavy-ion radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *47*, 1097–1103. [[CrossRef](#)]
84. Noda, K.; Kanazawa, M.; Itano, A.; Takada, E.; Torikoshi, M.; Araki, N.; Yoshizawa, J.; Sato, K.; Yamada, S.; Ogawa, H.; et al. Slow beam extraction by a transverse rf field with am and fm. *Nucl. Instrum. Methods Phys. Res. A* **1996**, *374*, 269–277. [[CrossRef](#)]
85. Schweikard, A.; Glosser, G.; Bodduluri, M.; Murphy, M.J.; Adler, J.R. Robotic motion compensation for respiratory movement during radiosurgery. *Comput. Aided Surg.* **2000**, *5*, 263–277. [[CrossRef](#)]
86. Jin, J.-Y.; Yin, F.-F.; Tenn, S.E.; Medin, P.M.; Solberg, T.D. Use of the brainlab exactrac x-ray 6d system in image-guided radiotherapy. *Med. Dosim.* **2008**, *33*, 124–134. [[CrossRef](#)] [[PubMed](#)]
87. Mutic, S.; Dempsey, J.F. The viewray system: Magnetic resonance-guided and controlled radiotherapy. *Semin. Radiat. Oncol.* **2014**, *24*, 196–199. [[CrossRef](#)] [[PubMed](#)]
88. Raaymakers, B.W.; Jürgenliemk-Schulz, I.M.; Bol, G.H.; Glitzner, M.; Kotte, A.N.T.J.; van Asselen, B.; de Boer, J.C.J.; Bluemink, J.J.; Hackett, S.L.; Moerland, M.A.; et al. First patients treated with a 1.5 t mri-linac: Clinical proof of concept of a high-precision, high-field mri guided radiotherapy treatment. *Phys. Med. Biol.* **2017**, *62*, L41. [[CrossRef](#)]
89. Shirato, H.; Shimizu, S.; Shimizu, T.; Nishioka, T.; Miyasaka, K. Real-time tumour-tracking radiotherapy. *Lancet* **1999**, *353*, 1331–1332. [[CrossRef](#)]
90. Shirato, H.; Shimizu, S.; Kunieda, T.; Kitamura, K.; van Herk, M.; Kagei, K.; Nishioka, T.; Hashimoto, S.; Fujita, K.; Aoyama, H.; et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *48*, 1187–1195. [[CrossRef](#)] [[PubMed](#)]

91. Shimizu, S.; Shirato, H.; Kitamura, K.; Shinohara, N.; Harabayashi, T.; Tsukamoto, T.; Koyanagi, T.; Miyasaka, K. Use of an implanted marker and real-time tracking of the marker for the positioning of prostate and bladder cancers. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *48*, 1591–1597. [[CrossRef](#)]
92. Harada, T.; Shirato, H.; Ogura, S.; Oizumi, S.; Yamazaki, K.; Shimizu, S.; Onimaru, R.; Miyasaka, K.; Nishimura, M.; Dosaka-Akita, H. Real-time tumor-tracking radiation therapy for lung carcinoma by the aid of insertion of a gold marker using bronchofiberscopy. *Cancer* **2002**, *95*, 1720–1727. [[CrossRef](#)] [[PubMed](#)]
93. Kitamura, K.; Shirato, H.; Shimizu, S.; Shinohara, N.; Harabayashi, T.; Shimizu, T.; Kodama, Y.; Endo, H.; Onimaru, R.; Nishioka, S.; et al. Registration accuracy and possible migration of internal fiducial gold marker implanted in prostate and liver treated with real-time tumor-tracking radiation therapy (rttrt). *Radiother. Oncol.* **2002**, *62*, 275–281. [[CrossRef](#)]
94. Sazawa, A.; Shinohara, N.; Harabayashi, T.; Abe, T.; Shirato, H.; Nonomura, K. Alternative approach in the treatment of adrenal metastasis with a real-time tracking radiotherapy in patients with hormone refractory prostate cancer. *Int. J. Urol.* **2009**, *16*, 410–412. [[CrossRef](#)]
95. Kimura, S.; Miyamoto, N.; Sutherland, K.L.; Suzuki, R.; Shirato, H.; Ishikawa, M. Fundamental study on quality assurance (qa) procedures for a real-time tumor tracking radiotherapy (rttrt) system from the viewpoint of imaging devices. *J. Appl. Clin. Med. Phys.* **2021**, *22*, 165–176. [[CrossRef](#)] [[PubMed](#)]
96. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
97. Shirato, H.; Shimizu, S.; Kitamura, K.; Nishioka, T.; Kagei, K.; Hashimoto, S.; Aoyama, H.; Kunieda, T.; Shinohara, N.; Dosaka-Akita, H.; et al. Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor. *Int. J. Radiat. Oncol. Biol. Phys.* **2000**, *48*, 435–442. [[CrossRef](#)]
98. Shimizu, S.; Matsuura, T.; Umezawa, M.; Hiramoto, K.; Miyamoto, N.; Umegaki, K.; Shirato, H. Preliminary analysis for integration of spot-scanning proton beam therapy and real-time imaging and gating. *Phys. Med.* **2014**, *30*, 555–558. [[CrossRef](#)]
99. Tsunashima, Y.; Vedam, S.; Dong, L.; Umezawa, M.; Sakae, T.; Bues, M.; Balter, P.; Smith, A.; Mohan, R. Efficiency of respiratory-gated delivery of synchrotron-based pulsed proton irradiation. *Phys. Med. Biol.* **2008**, *53*, 1947. [[CrossRef](#)]
100. Umezawa, M.; Fujimoto, R.; Umekawa, T.; Fujii, Y.; Takayanagi, T.; Ebina, F.; Aoki, T.; Nagamine, Y.; Matsuda, K.; Hiramoto, K.; et al. Development of the compact proton beam therapy system dedicated to spot scanning with real-time tumor-tracking technology. *AIP Conf. Proc.* **2013**, *1525*, 360–363.
101. Matsuura, T.; Miyamoto, N.; Shimizu, S.; Fujii, Y.; Umezawa, M.; Takao, S.; Nihongi, H.; Toramatsu, C.; Sutherland, K.; Suzuki, R.; et al. Integration of a real-time tumor monitoring system into gated proton spot-scanning beam therapy: An initial phantom study using patient tumor trajectory data. *Med. Phys.* **2013**, *40*, 071729. [[CrossRef](#)]
102. Tan, H.Q.; Koh, C.W.Y.; Lew, K.S.; Yeap, P.L.; Chua, C.G.A.; Lee, J.K.H.; Wibawa, A.; Master, Z.; Lee, J.C.L.; Park, S.Y. Real-time gated proton therapy with a reduced source to imager distance: Commissioning and quality assurance. *Phys. Med.* **2024**, *122*, 103380. [[CrossRef](#)] [[PubMed](#)]
103. Yamada, T.; Takao, S.; Koyano, H.; Nihongi, H.; Fujii, Y.; Hirayama, S.; Miyamoto, N.; Matsuura, T.; Umegaki, K.; Katoh, N.; et al. Validation of dose distribution for liver tumors treated with real-time-image gated spot-scanning proton therapy by log data based dose reconstruction. *J. Radiat. Res.* **2021**, *62*, 626–633. [[CrossRef](#)] [[PubMed](#)]
104. Nishioka, K.; Hashimoto, T.; Mori, T.; Uchinami, Y.; Kinoshita, R.; Katoh, N.; Taguchi, H.; Yasuda, K.; Ito, Y.M.; Takao, S.; et al. A single-institution prospective study to evaluate the safety and efficacy of real-time image-gated spot-scanning proton therapy (rgpt) for prostate cancer. *Adv. Radiat. Oncol.* **2024**, *9*, 101464. [[CrossRef](#)] [[PubMed](#)]
105. Chen, H.; Gogineni, E.; Cao, Y.; Wong, J.; Deville, C.; Li, H. Real-time gated proton therapy: Commissioning and clinical workflow for the hitachi system. *Int. J. Part. Ther.* **2024**, *11*, 100001. [[CrossRef](#)]
106. Koh, W.Y.C.; Tan, H.Q.; Lew, K.S.; Kor, W.T.A.; Samsuri, N.A.B.; Chan, J.W.S.; Chua, C.G.A.; Lee, J.K.H.; Wibawa, A.; Master, Z.; et al. Real-time gated proton therapy: Introducing clinical workflow and failure modes and effects analysis (fmea). *Tech. Innov. Patient Support Radiat. Oncol.* **2025**, *34*, 100311. [[CrossRef](#)]
107. Yoshimura, T.; Shimizu, S.; Hashimoto, T.; Nishioka, K.; Katoh, N.; Inoue, T.; Taguchi, H.; Yasuda, K.; Matsuura, T.; Takao, S.; et al. Analysis of treatment process time for real-time-image gated-spot-scanning proton-beam therapy (rgpt) system. *J. Appl. Clin. Med. Phys.* **2020**, *21*, 38–49. [[CrossRef](#)]
108. Lee, J.K.H.; Lew, K.S.; Koh, C.W.Y.; Lee, J.C.L.; Bettiol, A.A.; Park, S.Y.; Tan, H.Q. Comparison of translation algorithms in determining maximum allowable ctv shifts for real-time gated proton therapy (rgpt) robustness evaluation in prostate cancers. *J. Appl. Clin. Med. Phys.* **2025**, *26*, e14543. [[CrossRef](#)]
109. Murphy, M.J.; Balter, J.; Balter, S.; BenComo, J.A., Jr.; Das, I.J.; Jiang, S.B.; Ma, C.-M.; Olivera, G.H.; Rodebaugh, R.F.; Ruchala, K.J.; et al. The management of imaging dose during image-guided radiotherapy: Report of the AAPM task group 75. *Med. Phys.* **2007**, *34*, 4041–4063. [[CrossRef](#)]
110. Shirato, H.; Oita, M.; Fujita, K.; Watanabe, Y.; Miyasaka, K. Feasibility of synchronization of real-time tumor-tracking radiotherapy and intensity-modulated radiotherapy from viewpoint of excessive dose from fluoroscopy. *Int. J. Radiat. Oncol. Biol. Phys.* **2004**, *60*, 335–341. [[CrossRef](#)]

111. Bertholet, J.; Knopf, A.; Eiben, B.; McClelland, J.; Grimwood, A.; Harris, E.; Menten, M.; Poulsen, P.; Nguyen, D.T.; Keall, P.; et al. Real-time intrafraction motion monitoring in external beam radiotherapy. *Phys. Med. Biol.* **2019**, *64*, 15TR01. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Miyamoto, N.; Ishikawa, M.; Sutherland, K.; Suzuki, R.; Matsuura, T.; Toramatsu, C.; Takao, S.; Nihongi, H.; Shimizu, S.; Umegaki, K.; et al. A motion-compensated image filter for low-dose fluoroscopy in a real-time tumor-tracking radiotherapy system. *J. Radiat. Res.* **2015**, *56*, 186–196. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Yamanaka, M.; Furutani, K.M.; Shirata, R.; Matsumoto, K.; Yamano, A.; Shimo, T.; Tokuyue, K.; Beltran, C.J. Dosimetric impact and safety of uninterrupted fluoroscopic-gated proton therapy. *Med. Phys.* **2026**, *53*, e70227. [\[CrossRef\]](#)
114. Terunuma, T.; Takada, K.; Takao, S.; Osugi, M.; Miyamoto, N.; Asano, S.; Moriya, S.; Sakae, T.; Sakurai, H. Characterization of secondary-radiation background in x-ray flat-panel detectors during scanning proton beam irradiation. *Med. Phys.* **2025**, *52*, e70121. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Tanokura, R.; Horita, M.; Matsumoto, M.; Miyahara, A.; Furukawa, S.; Ido, Y.; Matsumoto, A.; Ogawa, S.; Makino, W.; Matsui, Y.; et al. Beam commissioning of a new compact scanned proton therapy system with four different dose calculation algorithms. *J. Appl. Clin. Med. Phys.* **2026**, *27*, e70433. [\[CrossRef\]](#)
116. Pidikiti, R.; Patel, B.C.; Maynard, M.R.; Dugas, J.P.; Syh, J.; Sahoo, N.; Wu, H.T.; Rosen, L.R. Commissioning of the world's first compact pencil-beam scanning proton therapy system. *J. Appl. Clin. Med. Phys.* **2018**, *19*, 94–105. [\[CrossRef\]](#)
117. Vilches-Freixas, G.; Unipan, M.; Rinaldi, I.; Martens, J.; Roijen, E.; Almeida, I.P.; Decabooter, E.; Bosmans, G. Beam commissioning of the first compact proton therapy system with spot scanning and dynamic field collimation. *Br. J. Radiol.* **2020**, *93*, 20190598. [\[CrossRef\]](#)
118. Laurent, F.; Latrabe, V.; Vergier, B.; Montaudon, M.; Vernejoux, J.M.; Dubrez, J. Ct-guided transthoracic needle biopsy of pulmonary nodules smaller than 20 mm: Results with an automated 20-gauge coaxial cutting needle. *Clin. Radiol.* **2000**, *55*, 281–287. [\[CrossRef\]](#)
119. Geraghty, P.R.; Kee, S.T.; McFarlane, G.; Razavi, M.K.; Sze, D.Y.; Dake, M.D. Ct-guided transthoracic needle aspiration biopsy of pulmonary nodules: Needle size and pneumothorax rate. *Radiology* **2003**, *229*, 475–481. [\[CrossRef\]](#) [\[PubMed\]](#)
120. Collins, B.T.; Erickson, K.; Reichner, C.A.; Collins, S.P.; Gagnon, G.J.; Dieterich, S.; McRae, D.A.; Zhang, Y.; Yousefi, S.; Levy, E.; et al. Radical stereotactic radiosurgery with real-time tumor motion tracking in the treatment of small peripheral lung tumors. *Radiat. Oncol.* **2007**, *2*, 39. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Bhagat, N.; Fidelman, N.; Durack, J.C.; Collins, J.; Gordon, R.L.; Laberge, J.M.; Kerlan, R.K. Complications associated with the percutaneous insertion of fiducial markers in the thorax. *Cardiovasc. Interv. Radiol.* **2010**, *33*, 1186–1191. [\[CrossRef\]](#)
122. Shirato, H.; Harada, T.; Harabayashi, T.; Hida, K.; Endo, H.; Kitamura, K.; Onimaru, R.; Yamazaki, K.; Kurauchi, N.; Shimizu, T.; et al. Feasibility of insertion/implantation of 2.0-mm-diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2003**, *56*, 240–247. [\[CrossRef\]](#) [\[PubMed\]](#)
123. Imura, M.; Yamazaki, K.; Shirato, H.; Onimaru, R.; Fujino, M.; Shimizu, S.; Harada, T.; Ogura, S.; Dosaka-Akita, H.; Miyasaka, K.; et al. Insertion and fixation of fiducial markers for setup and tracking of lung tumors in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2005**, *63*, 1442–1447. [\[CrossRef\]](#)
124. van der Voort van Zyp, N.C.; Hoogeman, M.S.; van de Water, S.; Levendag, P.C.; van der Holt, B.; Heijmen, B.J.M.; Nuytens, J.J. Stability of markers used for real-time tumor tracking after percutaneous intrapulmonary placement. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *81*, e75–e81. [\[CrossRef\]](#)
125. Imura, M.; Yamazaki, K.; Kubota, K.C.; Itoh, T.; Onimaru, R.; Cho, Y.; Hida, Y.; Kaga, K.; Onodera, Y.; Ogura, S.; et al. Histopathologic consideration of fiducial gold markers inserted for real-time tumor-tracking radiotherapy against lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2008**, *70*, 382–384. [\[CrossRef\]](#)
126. Patel, Z.; Retrouvey, M.; Vingan, H.; Williams, S. Tumor track seeding: A new complication of fiducial marker insertion. *Radiol. Case Rep.* **2014**, *9*, 928. [\[CrossRef\]](#)
127. Newhauser, W.; Fontenot, J.; Koch, N.; Dong, L.; Lee, A.; Zheng, Y.; Waters, L.; Mohan, R. Monte carlo simulations of the dosimetric impact of radiopaque fiducial markers for proton radiotherapy of the prostate. *Phys. Med. Biol.* **2007**, *52*, 2937. [\[CrossRef\]](#)
128. Giebler, A.; Fontenot, J.; Balter, P.; Ciangaru, G.; Zhu, R.; Newhauser, W. Dose perturbations from implanted helical gold markers in proton therapy of prostate cancer. *J. Appl. Clin. Med. Phys.* **2009**, *10*, 63–70. [\[CrossRef\]](#)
129. Habermehl, D.; Henkner, K.; Ecker, S.; Jakel, O.; Debus, J.; Combs, S.E. Evaluation of different fiducial markers for image-guided radiotherapy and particle therapy. *J. Radiat. Res.* **2013**, *54*, i61–i68. [\[CrossRef\]](#) [\[PubMed\]](#)
130. Matsuura, T.; Maeda, K.; Sutherland, K.; Takayanagi, T.; Shimizu, S.; Takao, S.; Miyamoto, N.; Nihongi, H.; Toramatsu, C.; Nagamine, Y.; et al. Biological effect of dose distortion by fiducial markers in spot-scanning proton therapy with a limited number of fields: A simulation study. *Med. Phys.* **2012**, *39*, 5584–5591. [\[CrossRef\]](#) [\[PubMed\]](#)
131. Mori, S.; Karube, M.; Shirai, T.; Tajiri, M.; Takekoshi, T.; Miki, K.; Shiraishi, Y.; Tanimoto, K.; Shibayama, K.; Yasuda, S.; et al. Carbon-ion pencil beam scanning treatment with gated markerless tumor tracking: An analysis of positional accuracy. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *95*, 258–266. [\[CrossRef\]](#)

132. Mori, S.; Sakata, Y.; Hirai, R.; Furuichi, W.; Shimabukuro, K.; Kohno, R.; Koom, W.S.; Kasai, S.; Okaya, K.; Iseki, Y. Commissioning of a fluoroscopic-based real-time markerless tumor tracking system in a superconducting rotating gantry for carbon-ion pencil beam scanning treatment. *Med. Phys.* **2019**, *46*, 1561–1574. [[CrossRef](#)]
133. Sakata, Y.; Hirai, R.; Kobuna, K.; Tanizawa, A.; Mori, S. A machine learning-based real-time tumor tracking system for fluoroscopic gating of lung radiotherapy. *Phys. Med. Biol.* **2020**, *65*, 085014. [[CrossRef](#)]
134. Sakata, Y.; Hirai, R.; Taguchi, Y.; Mori, S. Marker-less tumor tracking for lung cancer by tumor image pattern learning. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *96*, E651. [[CrossRef](#)]
135. Cui, Y.; Dy, J.G.; Sharp, G.C.; Alexander, B.; Jiang, S.B. Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers. *Phys. Med. Biol.* **2007**, *52*, 6229–6242. [[CrossRef](#)]
136. Cui, Y.; Dy, J.G.; Alexander, B.; Jiang, S.B. Fluoroscopic gating without implanted fiducial markers for lung cancer radiotherapy based on support vector machines. *Phys. Med. Biol.* **2008**, *53*, N315–N327. [[CrossRef](#)]
137. Berbeco, R.I.; Mostafavi, H.; Sharp, G.C.; Jiang, S.B. Towards fluoroscopic respiratory gating for lung tumours without radiopaque markers. *Phys. Med. Biol.* **2005**, *50*, 4481–4490. [[CrossRef](#)] [[PubMed](#)]
138. Hirai, R.; Sakata, Y.; Taguchi, Y.; Mori, S. Regression model of tumor and diaphragm position for marker-less tumor tracking in carbon ion scanning therapy for hepatocellular carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *96*, E639–E640. [[CrossRef](#)]
139. Hirai, R.; Sakata, Y.; Tanizawa, A.; Mori, S. Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis. *Phys. Med.* **2019**, *59*, 22–29. [[CrossRef](#)]
140. Mori, S.; Hirai, R.; Sakata, Y. Simulated four-dimensional ct for markerless tumor tracking using a deep learning network with multi-task learning. *Phys. Med.* **2020**, *80*, 151–158. [[CrossRef](#)]
141. Takahashi, W.; Oshikawa, S.; Mori, S. Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: Proof of concept by phantom study. *Br. J. Radiol.* **2020**, *93*, 20190420. [[CrossRef](#)]
142. Mori, S.; Hirai, R.; Sakata, Y.; Tachibana, Y.; Koto, M.; Ishikawa, H. Deep neural network-based synthetic image digital fluoroscopy using digitally reconstructed tomography. *Phys. Eng. Sci. Med.* **2023**, *46*, 1227–1237. [[CrossRef](#)]
143. Hirai, R.; Sakata, Y.; Tanizawa, A.; Mori, S. Regression model-based real-time markerless tumor tracking with fluoroscopic images for hepatocellular carcinoma. *Phys. Med.* **2020**, *70*, 196–205. [[CrossRef](#)]
144. Mylonas, A.; Keall, P.J.; Booth, J.T.; Shieh, C.-C.; Eade, T.; Poulsen, P.R.; Nguyen, D.T. A deep learning framework for automatic detection of arbitrarily shaped fiducial markers in intrafraction fluoroscopic images. *Med. Phys.* **2019**, *46*, 2286–2297. [[CrossRef](#)]
145. Liang, Z.; Zhou, Q.; Yang, J.; Zhang, L.; Liu, D.; Tu, B.; Zhang, S. Artificial intelligence-based framework in evaluating intrafraction motion for liver cancer robotic stereotactic body radiation therapy with fiducial tracking. *Med. Phys.* **2020**, *47*, 5482–5489. [[CrossRef](#)]
146. Lin, W.-Y.; Lin, S.-F.; Yang, S.-C.; Liou, S.-C.; Nath, R.; Liu, W. Real-time automatic fiducial marker tracking in low contrast cine-mv images. *Med. Phys.* **2013**, *40*, 011715. [[CrossRef](#)] [[PubMed](#)]
147. Maintz, J.B.A.; Viergever, M.A. A survey of medical image registration. *Med. Image Anal.* **1998**, *2*, 1–36. [[CrossRef](#)] [[PubMed](#)]
148. Viergever, M.A.; Maintz, J.B.A.; Klein, S.; Murphy, K.; Staring, M.; Pluim, J.P.W. A survey of medical image registration—Under review. *Med. Image Anal.* **2016**, *33*, 140–144. [[CrossRef](#)] [[PubMed](#)]
149. Sotiras, A.; Davatzikos, C.; Paragios, N. Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* **2013**, *32*, 1153–1190. [[CrossRef](#)]
150. Oliveira, F.P.M.; Tavares, J.M.R.S. Medical image registration: A review. *Comput. Methods Biomech. Biomed. Eng.* **2014**, *17*, 73–93. [[CrossRef](#)] [[PubMed](#)]
151. Penney, G.P.; Weese, J.; Little, J.A.; Desmedt, P.; Hill, D.L.G.; Hawkes, D.J. A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE Trans. Med. Imaging* **1998**, *17*, 586–595. [[CrossRef](#)]
152. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv* **2015**, arXiv:1502.01852.
153. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
154. Fu, Y.; Lei, Y.; Wang, T.; Curran, W.J.; Liu, T.; Yang, X. A review of deep learning based methods for medical image multi-organ segmentation. *Phys. Med.* **2021**, *85*, 107–122. [[CrossRef](#)]
155. Azad, R.; Aghdam, E.K.; Rauland, A.; Jia, Y.; Avval, A.H.; Bozorgpour, A.; Karimijafarbigloo, S.; Cohen, J.P.; Adeli, E.; Merhof, D. Medical image segmentation review: The success of u-net. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10076–10095. [[CrossRef](#)] [[PubMed](#)]
156. Fu, Y.; Lei, Y.; Wang, T.; Curran, W.J.; Liu, T.; Yang, X. Deep learning in medical image registration: A review. *Phys. Med. Biol.* **2020**, *65*, 20TR01. [[CrossRef](#)] [[PubMed](#)]
157. Xiao, H.; Teng, X.; Liu, C.; Li, T.; Ren, G.; Yang, R.; Shen, D.; Cai, J. A review of deep learning-based three-dimensional medical image registration methods. *Quant. Imaging Med. Surg.* **2021**, *11*, 4895–4916. [[CrossRef](#)]

158. Chen, J.; Liu, Y.; Wei, S.; Bian, Z.; Subramanian, S.; Carass, A.; Prince, J.L.; Du, Y. A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Med. Image Anal.* **2025**, *100*, 103385. [\[CrossRef\]](#)
159. Liu, X.; Geng, L.-S.; Huang, D.; Cai, J.; Yang, R. Deep learning-based target tracking with x-ray images for radiotherapy: A narrative review. *Quant. Imaging Med. Surg.* **2024**, *14*, 2671–2692. [\[CrossRef\]](#)
160. Ahishakiye, E.; Van Gijzen, M.B.; Tumwiine, J.; Wario, R.; Obungoloch, J. A survey on deep learning in medical image reconstruction. *Intell. Med.* **2021**, *1*, 118–127. [\[CrossRef\]](#)
161. Yaqub, M.; Jinchao, F.; Arshid, K.; Ahmed, S.; Zhang, W.; Nawaz, M.Z.; Mahmood, T. Deep learning-based image reconstruction for different medical imaging modalities. *Comput. Math. Methods Med.* **2022**, *2022*, 8750648. [\[CrossRef\]](#) [\[PubMed\]](#)
162. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
163. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [\[CrossRef\]](#)
164. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.556.
165. Lopez, A.R.; Giro-i-Nieto, X.; Burdick, J.; Marques, O. Skin lesion classification from dermoscopic images using deep learning techniques. In Proceedings of the 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), Innsbruck, Austria, 20–21 February 2017; pp. 49–54. [\[CrossRef\]](#)
166. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. *arXiv* **2015**, arXiv:1512.00567. [\[CrossRef\]](#)
167. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; IEEE: Las Vegas, NV, USA, 2016. [\[CrossRef\]](#)
168. Matsunaga, K.; Hamada, A.; Minagawa, A.; Koga, H. Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. *arXiv* **2017**, arXiv:1703.03108. [\[CrossRef\]](#)
169. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *arXiv* **2014**, arXiv:1411.4038.
170. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597. [\[CrossRef\]](#)
171. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661. [\[CrossRef\]](#)
172. Alamir, M.; Alghamdi, M. The role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Comput. Surv.* **2022**, *55*, 96. [\[CrossRef\]](#)
173. Wang, Z.; Lorenz, G.; Zhang, Z.; Dekker, A.; Traverso, A. Applications of generative adversarial networks (gans) in radiotherapy: Narrative review. *Precis. Cancer Med.* **2022**, *5*, 37. [\[CrossRef\]](#)
174. Sindhura, D.N.; Pai, R.M.; Bhat, S.N.; Pai, M.M.M. A review of deep learning and generative adversarial networks applications in medical image analysis. *Multimed. Syst.* **2024**, *30*, 161. [\[CrossRef\]](#)
175. Heng, Y.; Yinghua, M.; Khan, F.G.; Khan, A.; Ali, F.; Alzubi, A.A.; Hui, Z. Survey: Application and analysis of generative adversarial networks in medical images. *Artif. Intell. Rev.* **2024**, *58*, 39. [\[CrossRef\]](#)
176. Hussain, J.; Båth, M.; Ivarsson, J. Generative adversarial networks in medical image reconstruction: A systematic literature review. *Comput. Biol. Med.* **2025**, *191*, 110094. [\[CrossRef\]](#) [\[PubMed\]](#)
177. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
178. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
179. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
180. Li, J.; Chen, J.; Tang, Y.; Wang, C.; Landman, B.A.; Zhou, S.K. Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives. *Med. Image Anal.* **2023**, *85*, 102762. [\[CrossRef\]](#)
181. Parvaiz, A.; Khalid, M.A.; Zafar, R.; Ameer, H.; Ali, M.; Fraz, M.M. Vision transformers in medical computer vision—A contemplative retrospection. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106126. [\[CrossRef\]](#)
182. Xia, K.; Wang, J. Recent advances of transformers in medical image analysis: A comprehensive review. *MedComm—Future Med.* **2023**, *2*, e38. [\[CrossRef\]](#)
183. Azad, R.; Kazerouni, A.; Heidari, M.; Aghdam, E.K.; Molaei, A.; Jia, Y.; Jose, A.; Roy, R.; Merhof, D. Advances in medical image analysis with vision transformers: A comprehensive review. *Med. Image Anal.* **2024**, *91*, 103000. [\[CrossRef\]](#)
184. Aburass, S.; Dorgham, O.; Al Shaqsi, J.; Abu Rumman, M.; Al-Kadi, O. Vision transformers in medical imaging: A comprehensive review of advancements and applications across multiple diseases. *J. Imaging Inform. Med.* **2025**, *38*, 3928–3971. [\[CrossRef\]](#) [\[PubMed\]](#)
185. Sohl-Dickstein, J.; Weiss, E.A.; Maheswaranathan, N.; Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. *arXiv* **2015**, arXiv:1503.03585. [\[CrossRef\]](#)

186. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *arXiv* **2020**, arXiv:2006.11239. [[CrossRef](#)]
187. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv* **2020**, arXiv:2011.13456.
188. Kim, B.; Han, I.; Ye, J.C. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *Proceedings of the 17th European Conference on Computer Vision*; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer Nature: Cham, Switzerland, 2022; pp. 347–364.
189. Qin, Y.; Li, X. Fsdiffreg: Feature-wise and score-wise diffusion-guided unsupervised deformable image registration for cardiac images. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2023*; Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 655–665.
190. Zhuo, Y.; Shen, Y. Diffusereg: Denoising diffusion model for obtaining deformation fields in unsupervised deformable image registration. *arXiv* **2024**, arXiv:2410.05234. [[CrossRef](#)]
191. Kazerouni, A.; Aghdam, E.K.; Heidari, M.; Azad, R.; Fayyaz, M.; Hacihaliloglu, I.; Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Med. Image Anal.* **2023**, *88*, 102846. [[CrossRef](#)] [[PubMed](#)]
192. de Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Staring, M.; Išgum, I. End-to-end unsupervised deformable image registration with a convolutional neural network. In *Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 204–212.
193. de Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Sookoti, H.; Staring, M.; Išgum, I. A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **2019**, *52*, 128–143. [[CrossRef](#)]
194. Li, H.; Fan, Y. Non-rigid image registration using self-supervised fully convolutional networks without training data. In *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, USA, 4–7 April 2018; pp. 1075–1078. [[CrossRef](#)]
195. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. An unsupervised learning model for deformable medical image registration. *arXiv* **2018**, arXiv:1802.02604. [[CrossRef](#)]
196. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 1788–1800. [[CrossRef](#)]
197. Kim, B.; Kim, D.H.; Park, S.H.; Kim, J.; Lee, J.-G.; Ye, J.C. Cyclemorph: Cycle consistent unsupervised deformable image registration. *Med. Image Anal.* **2021**, *71*, 102036. [[CrossRef](#)]
198. Roggen, T.; Bobic, M.; Givehchi, N.; Scheib, S.G. Deep learning model for markerless tracking in spinal sbirt. *Phys. Med.* **2020**, *74*, 66–73. [[CrossRef](#)] [[PubMed](#)]
199. He, X.; Cai, W.; Li, F.; Zhang, P.; Reyngold, M.; Cuaron, J.J.; Cerviño, L.I.; Li, T.; Li, X. Automatic stent recognition using perceptual attention u-net for quantitative intrafraction motion monitoring in pancreatic cancer radiotherapy. *Med. Phys.* **2022**, *49*, 5283–5293. [[CrossRef](#)]
200. Edmunds, D.; Sharp, G.; Winey, B. Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam ct projections: A convolutional neural network approach. *Biomed. Phys. Eng. Express* **2019**, *5*, 035005. [[CrossRef](#)] [[PubMed](#)]
201. Terunuma, T.; Tokui, A.; Sakae, T. Novel real-time tumor-contouring method using deep learning to prevent mistracking in x-ray fluoroscopy. *Radiol. Phys. Technol.* **2018**, *11*, 43–53. [[CrossRef](#)]
202. Terunuma, T.; Sakae, T.; Hu, Y.; Takei, H.; Moriya, S.; Okumura, T.; Sakurai, H. Explainability and controllability of patient-specific deep learning with attention-based augmentation for markerless image-guided radiotherapy. *Med. Phys.* **2023**, *50*, 480–494. [[CrossRef](#)]
203. Huang, L.; Kurz, C.; Freisleder, P.; Manapov, F.; Corradini, S.; Niyazi, M.; Belka, C.; Landry, G.; Riboldi, M. Simultaneous object detection and segmentation for patient-specific markerless lung tumor tracking in simulated radiographs with deep learning. *Med. Phys.* **2024**, *51*, 1957–1973. [[CrossRef](#)]
204. Mylonas, A.; Li, Z.; Mueller, M.; Booth, J.T.; Brown, R.; Gardner, M.; Kneebone, A.; Eade, T.; Keall, P.J.; Nguyen, D.T. Patient-specific prostate segmentation in kilovoltage images for radiation therapy intrafraction monitoring via deep learning. *Commun. Med.* **2025**, *5*, 212. [[CrossRef](#)] [[PubMed](#)]
205. Lei, Y.; Tian, Z.; Wang, T.; Higgins, K.; Bradley, J.D.; Curran, W.J.; Liu, T.; Yang, X. Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: A proof of concept study. *Phys. Med. Biol.* **2020**, *65*, 235003. [[CrossRef](#)] [[PubMed](#)]
206. He, X.; Cai, W.; Li, F.; Fan, Q.; Zhang, P.; Cuaron, J.J.; Cerviño, L.I.; Li, X.; Li, T. Decompose kv projection using neural network for improved motion tracking in paraspinal sbirt. *Med. Phys.* **2021**, *48*, 7590–7601. [[CrossRef](#)]
207. Fu, Y.; Fan, Q.; Cai, W.; Li, F.; He, X.; Cuaron, J.; Cervino, L.; Moran, J.M.; Li, T.; Li, X. Enhancing the target visibility with synthetic target specific digitally reconstructed radiograph for intrafraction motion monitoring: A proof-of-concept study. *Med. Phys.* **2023**, *50*, 7791–7805. [[CrossRef](#)]

208. Fu, Y.; Zhang, P.; Fan, Q.; Cai, W.; Pham, H.; Burleson, S.; Shaverdian, N.; Wu, A.J.; Cervino, L.I.; Moran, J.M.; et al. Intrafractional markerless lung tumor tracking: First clinical experience with ai-empowered target decomposition technique. *Int. J. Radiat. Oncol. Biol. Phys.* **2025**, *123*, S163–S164. [[CrossRef](#)]
209. Madden, L.; Ahmed, A.; Stewart, M.; Chrystall, D.; Mylonas, A.; Brown, R.; Nguyen, D.T.; Keall, P.; Booth, J. Cbct-drrs superior to ct-drrs for target-tracking applications for pancreatic sbirt. *Biomed. Phys. Eng. Express* **2024**, *10*, 035039. [[CrossRef](#)]
210. Ahmed, A.; Madden, L.; Stewart, M.; Mylonas, A.; Brown, R.; Metz, G.; Shepherd, M.; Coronel, C.; Ambrose, L.; Turk, A.; et al. Patient-specific markerless tracking of pancreatic-gtv and abdominal organs-at-risk using deep-learning. *Int. J. Radiat. Oncol. Biol. Phys.* **2025**, *123*, S163. [[CrossRef](#)]
211. Yan, Y.; Fujii, F.; Shiinoki, T. Marker-less lung tumor tracking from real-time color x-ray fluoroscopic images using cross-patient deep learning model. *Bioengineering* **2025**, *12*, 1197. [[CrossRef](#)]
212. Yan, Y.; Fujii, F.; Shiinoki, T.; Liu, S. Markerless lung tumor localization from intraoperative stereo color fluoroscopic images for radiotherapy. *IEEE Access* **2024**, *12*, 40809–40826. [[CrossRef](#)]
213. Wang, C.; Hunt, M.; Zhang, L.; Rimner, A.; Yorke, E.; Lovelock, M.; Li, X.; Li, T.; Mageras, G.; Zhang, P. Technical note: 3d localization of lung tumors on cone beam ct projections via a convolutional recurrent neural network. *Med. Phys.* **2020**, *47*, 1161–1166. [[CrossRef](#)]
214. Grama, D.; Dahele, M.; van Rooij, W.; Slotman, B.; Gupta, D.K.; Verbakel, W.F.A.R. Deep learning-based markerless lung tumor tracking in stereotactic radiotherapy using siamese networks. *Med. Phys.* **2023**, *50*, 6881–6893. [[CrossRef](#)] [[PubMed](#)]
215. Mok, T.C.W.; Chung, A.C.S. Affine medical image registration with coarse-to-fine vision transformer. *arXiv* **2022**, arXiv:2203.15216.
216. Xu, D.; Descovich, M.; Liu, H.; Lao, Y.; Gottschalk, A.R.; Sheng, K. Deep match: A zero-shot framework for improved fiducial-free respiratory motion tracking. *Radiother. Oncol.* **2024**, *194*, 110179. [[CrossRef](#)]
217. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [[CrossRef](#)]
218. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv* **2015**, arXiv:1506.01497. [[CrossRef](#)]
219. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. *arXiv* **2015**, arXiv:1506.02640.
220. Zhao, W.; Shen, L.; Han, B.; Yang, Y.; Cheng, K.; Toesca, D.A.S.; Koong, A.C.; Chang, D.T.; Xing, L. Markerless pancreatic tumor target localization enabled by deep learning. *Int. J. Radiat. Oncol. Biol. Phys.* **2019**, *105*, 432–439. [[CrossRef](#)] [[PubMed](#)]
221. Zhao, W.; Han, B.; Yang, Y.; Buyyounouski, M.; Hancock, S.L.; Bagshaw, H.; Xing, L. Incorporating imaging information from deep neural network layers into image guided radiation therapy (igrt). *Radiother. Oncol.* **2019**, *140*, 167–174. [[CrossRef](#)]
222. Zhou, D.; Nakamura, M.; Mukumoto, N.; Matsuo, Y.; Mizowaki, T. Feasibility study of deep learning-based markerless real-time lung tumor tracking with orthogonal x-ray projection images. *J. Appl. Clin. Med. Phys.* **2023**, *24*, e13894. [[CrossRef](#)]
223. Ahmed, A.M.; Gargett, M.; Madden, L.; Mylonas, A.; Chrystall, D.; Brown, R.; Briggs, A.; Nguyen, T.; Keall, P.; Kneebone, A.; et al. Evaluation of deep learning based implanted fiducial markers tracking in pancreatic cancer patients. *Biomed. Phys. Eng. Express* **2023**, *9*, 035008. [[CrossRef](#)] [[PubMed](#)]
224. Cao, X.; Yang, J.; Zhang, J.; Nie, D.; Kim, M.; Wang, Q.; Shen, D. Deformable image registration based on similarity-steered cnn regression. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*; Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 300–308.
225. Hu, Y.; Modat, M.; Gibson, E.; Li, W.; Ghavami, N.; Bonmati, E.; Wang, G.; Bandula, S.; Moore, C.M.; Emberton, M.; et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Med. Image Anal.* **2018**, *49*, 1–13. [[CrossRef](#)] [[PubMed](#)]
226. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. *arXiv* **2015**, arXiv:1506.02025. [[CrossRef](#)]
227. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv* **2016**, arXiv:1609.04802.
228. Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576. [[CrossRef](#)]
229. Ahmed, A.M.; Madden, L.; Stewart, M.; Chow, B.V.Y.; Mylonas, A.; Brown, R.; Metz, G.; Shepherd, M.; Coronel, C.; Ambrose, L.; et al. Patient-specific deep learning tracking for real-time 2d pancreas localisation in kv-guided radiotherapy. *Phys. Imaging Radiat. Oncol.* **2025**, *35*, 100794. [[CrossRef](#)] [[PubMed](#)]
230. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
231. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. *arXiv* **2013**, arXiv:1312.6199.
232. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. *arXiv* **2014**, arXiv:1412.6572.
233. Jo, J.; Bengio, Y. Measuring the tendency of cnns to learn surface statistical regularities. *arXiv* **2017**, arXiv:1711.11561. [[CrossRef](#)]

234. Azulay, A.; Weiss, Y. Why do deep convolutional networks generalize so poorly to small image transformations? *arXiv* **2018**, arXiv:1805.12177.
235. Finlayson, S.G.; Bowers, J.D.; Ito, J.; Zittrain, J.L.; Beam, A.L.; Kohane, I.S. Adversarial attacks on medical machine learning. *Science* **2019**, *363*, 1287–1289. [[CrossRef](#)] [[PubMed](#)]
236. Ozbulak, U.; Van Messem, A.; De Neve, W. Impact of adversarial examples on deep learning models for biomedical image segmentation. *arXiv* **2019**, arXiv:1907.13124. [[CrossRef](#)]
237. Cheng, G.; Ji, H. Adversarial perturbation on mri modalities in brain tumor segmentation. *IEEE Access* **2020**, *8*, 206009–206015. [[CrossRef](#)]
238. Apostolidis, K.D.; Papakostas, G.A. A survey on adversarial deep learning robustness in medical image analysis. *Electronics* **2021**, *10*, 2132. [[CrossRef](#)]
239. Maliamanis, T.V.; Apostolidis, K.D.; Papakostas, G.A. How resilient are deep learning models in medical image analysis? The case of the moment-based adversarial attack (mb-ada). *Biomedicines* **2022**, *10*, 2545. [[CrossRef](#)] [[PubMed](#)]
240. Maynez, J.; Narayan, S.; Bohnet, B.; McDonald, R. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020*; Jurafsky, D., Chai, J., Schluter, N., Tetreault, J., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 1906–1919. [[CrossRef](#)]
241. Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y.J.; Madotto, A.; Fung, P. Survey of hallucination in natural language generation. *ACM Comput. Surv.* **2023**, *55*, 1–38. [[CrossRef](#)]
242. Bhadra, S.; Kelkar, V.A.; Brooks, F.J.; Anastasio, M.A. On hallucinations in tomographic image reconstruction. *IEEE Trans. Med. Imaging* **2021**, *40*, 3249–3260. [[CrossRef](#)] [[PubMed](#)]
243. Xia, M.; Bayerlein, R.; Chemli, Y.; Liu, X.; Ouyang, J.; Lin, M.; El Fakhri, G.; Badawi, R.D.; Li, Q.; Liu, C. On hallucinations in artificial intelligence-generated content for nuclear medicine imaging (the dream report). *J. Nucl. Med.* **2025**, *67*, 166–174. [[CrossRef](#)]
244. Kim, S.; Tregidgo, H.F.J.; Figini, M.; Jin, C.; Joshi, S.; Alexander, D.C. Tackling hallucination from conditional models for medical image reconstruction with dynamicdps. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2025*; Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J., Eds.; Springer Nature: Cham, Switzerland, 2026; pp. 593–603.
245. Tivnan, M.; Yoon, S.; Chen, Z.; Li, X.; Wu, D.; Li, Q. Hallucination index: An image quality metric for generative reconstruction models. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2024: 27th International Conference, Marrakesh, Morocco, 6–10 October 2024*; Part X; Springer: Marrakesh, Morocco, 2024; pp. 449–458. [[CrossRef](#)]
246. Li, J.; Rosellon-Inclan, I.; Kutyniok, G.; Starck, J.-L. Chem: Estimating and understanding hallucinations in deep learning for image processing. *arXiv* **2025**, arXiv:2512.09806. [[CrossRef](#)]
247. LeCun, Y. A path towards autonomous machine intelligence version 0.9.2, 2022-06-27. *Open Rev.* **2022**, *62*, 1–62.
248. Balestrieri, R.; LeCun, Y. Learning by reconstruction produces uninformative features for perception. *arXiv* **2024**, arXiv:2402.11337. [[CrossRef](#)]
249. Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.; LeCun, Y.; Ballas, N. Self-supervised learning from images with a joint-embedding predictive architecture. *arXiv* **2023**, arXiv:2301.08243.
250. Assran, M.; Bardes, A.; Fan, D.; Garrido, Q.; Howes, R.; Komeili, M.; Muckley, M.; Rizvi, A.; Roberts, C.; Sinha, K.; et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv* **2025**, arXiv:2506.09985.
251. Balestrieri, R.; LeCun, Y. Lejepa: Provable and scalable self-supervised learning without the heuristics. *arXiv* **2025**, arXiv:2511.08544.
252. Li, F.-F. *From Words to Worlds: Spatial Intelligence is AI's Next Frontier*; Substack: San Francisco, CA, USA, 2025.
253. Zhou, D.; Nakamura, M.; Mukumoto, N.; Yoshimura, M.; Mizowaki, T. Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning. *Med. Phys.* **2022**, *49*, 1382–1390. [[CrossRef](#)]
254. He, X.; Cai, W.; Li, F.; Fan, Q.; Zhang, P.; Cuaron, J.J.; Cerviño, L.I.; Moran, J.M.; Li, X.; Li, T. Patient specific prior cross attention for kv decomposition in paraspinal motion tracking. *Med. Phys.* **2023**, *50*, 5343–5353. [[CrossRef](#)]
255. Goodman, B.; Flaxman, S. European union regulations on algorithmic decision-making and a “right to explanation”. *AI Mag.* **2017**, *38*, 50–57. [[CrossRef](#)]
256. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. Xai—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [[CrossRef](#)]
257. Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I.; Precise4Q consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [[CrossRef](#)] [[PubMed](#)]
258. Bhati, D.; Neha, F.; Amiruzzaman, M. A survey on explainable artificial intelligence (xai) techniques for visualizing deep learning models in medical imaging. *J. Imaging* **2024**, *10*, 239. [[CrossRef](#)] [[PubMed](#)]
259. Begoli, E.; Bhattacharya, T.; Kusnezov, D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat. Mach. Intell.* **2019**, *1*, 20–23. [[CrossRef](#)]

260. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R.; et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion* **2021**, *76*, 243–297. [\[CrossRef\]](#)
261. Faghani, S.; Moassefi, M.; Rouzrokh, P.; Khosravi, B.; Baffour, F.I.; Ringler, M.D.; Erickson, B.J. Quantifying uncertainty in deep learning of radiologic images. *Radiology* **2023**, *308*, e222217. [\[CrossRef\]](#)
262. Huang, L.; Ruan, S.; Xing, Y.; Feng, M. A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods. *Med. Image Anal.* **2024**, *97*, 103223. [\[CrossRef\]](#) [\[PubMed\]](#)
263. Risholm, P.; Pieper, S.; Samset, E.; Wells, W.M. Summarizing and visualizing uncertainty in non-rigid registration. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010, Berlin, Heidelberg, 20–24 September 2010*; Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 554–561.
264. Le Folgoc, L.; Delingette, H.; Criminisi, A.; Ayache, N. Sparse bayesian registration of medical images for self-tuning of parameters and spatially adaptive parametrization of displacements. *Med. Image Anal.* **2017**, *36*, 79–97. [\[CrossRef\]](#) [\[PubMed\]](#)
265. Dalca, A.V.; Balakrishnan, G.; Guttag, J.; Sabuncu, M.R. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Med. Image Anal.* **2019**, *57*, 226–236. [\[CrossRef\]](#)
266. Khawaled, S.; Freiman, M. Npbdreg: Uncertainty assessment in diffeomorphic brain mri registration using a non-parametric bayesian deep-learning based approach. *Comput. Med. Imaging Graph.* **2022**, *99*, 102087. [\[CrossRef\]](#) [\[PubMed\]](#)
267. Gong, X.; Khaidem, L.; Zhu, W.; Zhang, B.; Doermann, D. Uncertainty learning towards unsupervised deformable medical image registration. In *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: Los Alamitos, CA, USA, 2022; pp. 1555–1564. [\[CrossRef\]](#)
268. Rivetti, L.; Studen, A.; Sharma, M.; Chan, J.; Jeraj, R. Uncertainty estimation and evaluation of deformation image registration based convolutional neural networks. *Phys. Med. Biol.* **2024**, *69*, 115045. [\[CrossRef\]](#)
269. Cohen, T.S.; Welling, M. Group equivariant convolutional networks. *arXiv* **2016**, arXiv:1602.07576. [\[CrossRef\]](#)
270. Bronstein, M.M.; Bruna, J.; Cohen, T.; Velicković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv* **2021**, arXiv:2104.13478. [\[CrossRef\]](#)
271. Gerken, J.E.; Aronsson, J.; Carlsson, O.; Linander, H.; Ohlsson, F.; Petersson, C.; Persson, D. Geometric deep learning and equivariant neural networks. *Artif. Intell. Rev.* **2023**, *56*, 14605–14662. [\[CrossRef\]](#)
272. Lafarge, M.W.; Bekkers, E.J.; Pluim, J.P.W.; Duits, R.; Veta, M. Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Med. Image Anal.* **2021**, *68*, 101849. [\[CrossRef\]](#)
273. Suliman, M.A.; Williams, L.Z.J.; Fawaz, A.; Robinson, E.C. Unsupervised multimodal surface registration with geometric deep learning. *Med. Image Anal.* **2026**, *107*, 103821. [\[CrossRef\]](#) [\[PubMed\]](#)
274. Greer, H.; Tian, L.; Vialard, F.-X.; Kwitt, R.; Estepar, R.S.J.; Niethammer, M. Carl: A framework for equivariant image registration. *arXiv* **2024**, arXiv:2405.16738. [\[CrossRef\]](#)
275. Shao, H.-C.; Wang, J.; Bai, T.; Chun, J.; Park, J.C.; Jiang, S.; Zhang, Y. Real-time liver tumor localization via a single x-ray projection using deep graph neural network-assisted biomechanical modeling. *Phys. Med. Biol.* **2022**, *67*, 115009. [\[CrossRef\]](#) [\[PubMed\]](#)
276. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [\[CrossRef\]](#)
277. Corona-Figueroa, A.; Frawley, J.; Bond-Taylor, S.; Bethapudi, S.; Shum, H.P.H.; Willcocks, C.G. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. *arXiv* **2022**, arXiv:2202.01020.
278. Zha, R.; Lin, T.J.; Cai, Y.; Cao, J.; Zhang, Y.; Li, H. R²-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. *arXiv* **2024**, arXiv:2405.20693.
279. Cai, Y.; Liang, Y.; Wang, J.; Wang, A.; Zhang, Y.; Yang, X.; Zhou, Z.; Yuille, A. Radiative gaussian splatting for efficient x-ray novel view synthesis. *arXiv* **2024**, arXiv:2403.04116. [\[CrossRef\]](#)
280. Marzol, K.; Kolton, I.; Smolak-Dyżewska, W.; Kaleta, J.; Mazur, M.; Spurek, P. Medgs: Gaussian splatting for multi-modal 3d medical imaging. *arXiv* **2025**, arXiv:2509.16806.
281. Bardes, A.; Garrido, Q.; Ponce, J.; Chen, X.; Rabbat, M.; LeCun, Y.; Assran, M.; Ballas, N. Revisiting feature prediction for learning visual representations from video. *arXiv* **2024**, arXiv:2404.08471. [\[CrossRef\]](#)
282. Dawid, A.; LeCun, Y. Introduction to latent variable energy-based models: A path toward autonomous machine intelligence. *J. Stat. Mech. Theory Exp.* **2024**, *2024*, 104011. [\[CrossRef\]](#)
283. Kahneman, D. *Thinking, Fast and Slow*, 1st ed.; Farrar, Straus and Giroux: New York, NY, USA, 2011.
284. Dhont, J.; Verellen, D.; Poels, K.; Tournel, K.; Burghelée, M.; Gevaert, T.; Collen, C.; Engels, B.; Van Den Begin, R.; Bult, N.; et al. Feasibility of markerless tumor tracking by sequential dual-energy fluoroscopy on a clinical tumor tracking system. *Radiother. Oncol.* **2015**, *117*, 487–490. [\[CrossRef\]](#) [\[PubMed\]](#)
285. Menten, M.J.; Fast, M.F.; Nill, S.; Oelfke, U. Using dual-energy x-ray imaging to enhance automated lung tumor tracking during real-time adaptive radiotherapy. *Med. Phys.* **2015**, *42*, 6987–6998. [\[CrossRef\]](#)

286. Haytmyradov, M.; Mostafavi, H.; Wang, A.; Zhu, L.; Surucu, M.; Patel, R.; Ganguly, A.; Richmond, M.; Cassetta, R.; Harkenrider, M.M.; et al. Markerless tumor tracking using fast-kv switching dual-energy fluoroscopy on a benchtop system. *Med. Phys.* **2019**, *46*, 3235–3244. [[CrossRef](#)] [[PubMed](#)]
287. Keall, P.J.; Sawant, A.; Berbeco, R.I.; Booth, J.T.; Cho, B.; Cerviño, L.I.; Cirino, E.; Dieterich, S.; Fast, M.F.; Greer, P.B.; et al. AAPM task group 264: The safe clinical implementation of mlc tracking in radiotherapy. *Med. Phys.* **2021**, *48*, e44–e64. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.