

Review Article

Received 10 June 2026

Revised 20 July 2026

Accepted 04 August 2026

Natural Resources for Human
Health 2026, Vol. 6 (14s): 753–767

KEYWORDS:

Camera - augmented mobile,
C -Arm , Convolutional Neural
Network, Autonomy, Deep
learning, Robot control

Autonomous Synergy: Artificial Intelligence Orchestration of Mobile Imaging Units, C- Arms, and Surgical Workflow in the Operating Suite

Shivanshu Chauhan^{1*}, Gaurav Kumar², Pankaj Kumar Dutt³,
Amrita⁴, Riyanshi Verma⁵, Gautam Yadav⁶, Arushi Thakur⁷

¹Assistant Professor, Department of Radio-Diagnosis, SMC, Santosh Deemed to be University. shivanshuchauhan025@gmail.com

²Assistant professor, Anaesthesia and operation theatre technology, school of Allied health sciences, Noida International University, Greater Noida, gauravberwalia@gmail.com

³Assistant professor, Radiology and imaging technology, School of Allied health sciences, Noida International University. pankajkdutt@gmail.com

⁴MSc Scholar, Department of Radio-Diagnosis, SMC, Santosh Deemed to be University, amritahalawat@gmail.com

⁵Assistant Professor, Department of Radio-Imaging Technology, School of allied and healthcare sciences, Sanskriti University, Mathura, Uttar Pradesh, riyanshiverma508@gmail.com

⁶Assistant Professor, Department of Radio-Imaging Technology, School of Medical & Allied Science, Sanskriti University, Mathura, Uttar Pradesh yadavgautam5080@gmail.com

⁷Assistant Professor Radiology and Imaging Technology, Saraswati College of Pharmacy, SGC Group, Gharuan -140413, Mohali, Punjab, India. thakurarushi0307@gmail.com

Corresponding author- Shivanshu Chauhan*

ABSTRACT: The modern operating suite represents a high-density, dynamically evolving socio-technical ecosystem where therapeutic precision, radiation safety, and operative efficiency intersect. Conventional intraoperative fluoroscopy using mobile C-arm systems relies heavily on manual manoeuvring, verbal closed-loop team communication, iterative repositioning, and intermittent ionising radiation, creating operational bottlenecks and cumulative occupational radiation risks. Recent breakthroughs in artificial intelligence (AI), computer vision, spatial computing, and robotic automation have paved the way for "Autonomous Synergy"—a paradigm shift toward the agentic orchestration of mobile imaging systems in direct alignment with real-time surgical workflow.

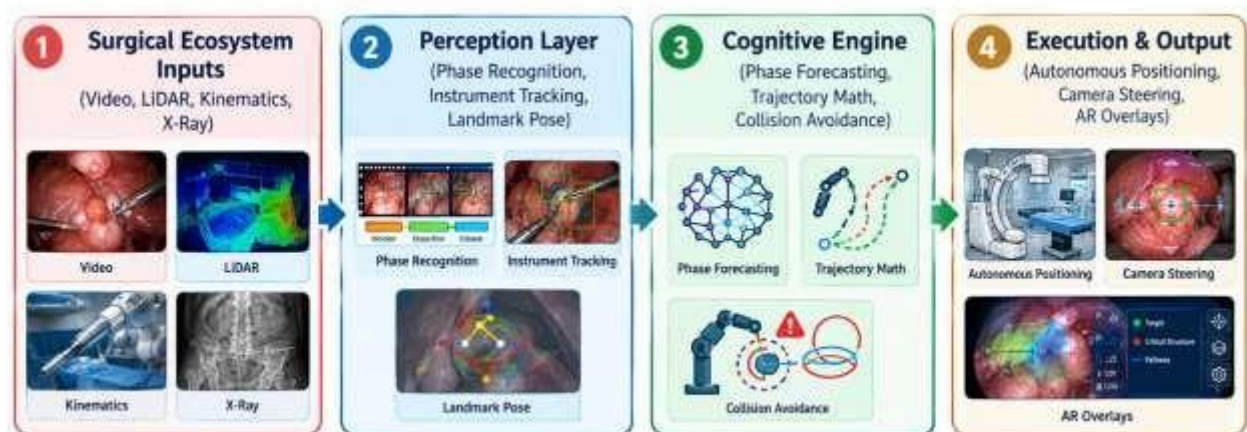
This review comprehensively examines the technological, algorithmic, and clinical foundations of AI-driven surgical imaging automation. We analyse the progression from early optical camera-augmented mobile C-arms (CamC) and LiDAR-based external localisation to modern agentic systems capable of autonomous skeletal landmark detection, kinematic coupling with smart surgical tables, and robotic self-driving navigation. Furthermore, we evaluate surgical scene understanding frameworks spanning

surgical phase, step, and action triplet recognition, alongside multi-instrument tracking methodologies (Region-based Convolutional Neural Networks).

By synthesising state-of-the-art developments across image guidance, spatial tracking, contextual workflow recognition, and autonomous robotic control, this paper establishes a unified taxonomy for agentic operating rooms. We address safety-critical control architectures, collision avoidance, cognitive ergonomics, dose reduction algorithms, ethical liabilities, and regulatory considerations required for clinical deployment. Finally, we map future research directions toward multi-modal foundation models and interconnected hybrid operating environments.

1. INTRODUCTION

Intraoperative fluoroscopy serves as an indispensable tool across modern surgical disciplines, including orthopaedic trauma, spinal reconstruction, vascular intervention, neurosurgery, and oncology. Mobile C-arm fluoroscopy units bridge pre-operative anatomical models and dynamic intraoperative realities, providing real-time verification of implant trajectory, osteosynthesis reduction, and instrument positioning [1, 2]. However, the physical and cognitive orchestration of conventional C-arms remains fundamentally manual, fragmented, and ergonomically challenging [3, 4].



Surgical Ecosystem Orchestration Pipeline

Within traditional operating room (OR) environments, obtaining standard anatomical projections requires iterative physical adjustment of the C-arm gantry by a designated radiology technician or circulating nurse acting on verbal directives from the primary surgeon. This trial-and-error process introduces several distinct challenges:

- 1. Prolonged Operative Time and Workflow Disruption:** Repositioning interrupts procedural rhythm, increasing anaesthesia duration and associated systemic morbidity [5, 6].
- 2. Radiation Burden:** Iterative "scout" fluoroscopic shots taken to align the imaging axis deliver cumulative ionising radiation to both patient and surgical staff, contributing to occupational health risks [1, 3].
- 3. Sterile Field Breach and Physical Congestion:** In dense surgical suites crowded with anaesthesia columns, robotic consoles, bypass machines, and nursing personnel, manual movement of heavy imaging machinery risks accidental breaches of the sterile field or physical collisions with equipment and personnel [6, 7].

To resolve these limitations, modern surgical data science and interventional robotics are converging toward **Autonomous Synergy**: the intelligent, context-aware orchestration of mobile imaging systems governed by AI agents [5, 8, 9]. Rather than acting as passive imaging instruments, modern mobile C-arms are transforming into self-navigating, context-aware robotic agents [6].

⁷¹. Driven by real-time surgical scene understanding, deep learning-based phase and instrument recognition, and spatial computer vision, these systems can anticipate surgical imaging needs, manoeuvre autonomously to target poses, align with articulated operating tables, and optimise radiation delivery without direct manual intervention [^{10,11,12,13,14}].

This review synthesises the technological evolution, algorithmic mechanisms, and clinical evidence governing this multidisciplinary shift. By evaluating hardware sensing modalities, deep neural network architectures for workflow recognition, kinematics modelling, and real-time robotic actuation, we establish an integrated framework for the next generation of autonomous interventional suites.

2. EVOLUTION OF MOBILE INTRAOPERATIVE IMAGING

The transition from purely mechanical intraoperative radiography to autonomous imaging platforms spans three decades of cross-disciplinary engineering across optical tracking, sensor fusion, augmented reality, and mobile robotics.

- **Classical Fluoroscopy and the Camera-Augmented Mobile C-Arm (CamC)**

Historically, surgeons were required to mentally reconstruct three-dimensional anatomy from decoupled, planar two-dimensional fluoroscopic snapshots. Navab et al. pioneered a breakthrough solution by introducing the **Camera-Augmented Mobile C-Arm (CamC)** [¹¹]. By affixing an optical video camera co-aligned with the X-ray source via an engineered mirror construction, the CamC system optically co-registered the video image with the fluoroscopic image plane.



Through precise geometric calibration, the X-ray beam centre and the optical camera's optical centre were made co-axial, enabling seamless superposition of live optical video and radiographic anatomy without continuous X-ray emission [¹¹]. Subsequent clinical feasibility trials confirmed that CamC-assisted augmented reality guidance substantially reduced trial-and-error fluoroscopy in distal locking of intramedullary nails, pedicle screw fixation, and foreign body localisation [²¹].

Despite these gains, first-generation CamC architectures remained bounded by mechanical constraints:

- Calibration parameters could drift under mechanical flexion of the C-arm gantry during wide-angle rotations ^[1, 2].
- Scene understanding relied heavily on static visual registration without semantic awareness of operative instruments, biological tissue changes, or dynamic procedural context ^[4].
- The system remained mechanically passive, requiring physical positioning and manual alignment by human operators ^[2].
- **External Tracking, Sensor Fusion, and LiDAR Integration**

To overcome the physical burden of manual repositioning, investigators introduced external tracking paradigms utilising optical retroreflective markers, electromagnetic (EM) tracking fields, and environmental range finders ^[3]. However, optical lines-of-sight are frequently obstructed by surgical staff in crowded operating suites, while EM tracking remains vulnerable to field distortions caused by ferromagnetic operating tables and electrosurgical units.

To circumvent external marker line-of-sight constraints, Bernhard et al. introduced **RAY-POS**, a LiDAR-based assistance platform designed for marker less intraoperative repositioning of mobile C-arms ^[3]. By mounting light detection and ranging (LiDAR) transceivers directly onto the mobile chassis, RAY-POS continuously samples ambient 3D spatial point clouds of the OR suite. Using iterative closest point (ICP) matching and spatial simultaneous localisation and mapping (SLAM), the system maps the operating room environment and registers the C-arm's coordinate frame relative to static room features and the surgical table ^[3]. This approach allows automated or assisted return to previously stored anatomical projection angles without requiring bulky ceiling-mounted optical navigation towers ^[3, 4].

- **The Modern Milestone: Self-Driving 3D C-Arms and Agentic Robotics**

Between 2024 and 2026, intraoperative imaging advanced toward full mechanical and algorithmic autonomy. The deployment of motorised omnidirectional wheel drives, integrated multi-axis sensor fusion, and on-board embedded compute clusters catalysed the transition from passive imaging units to autonomous mobile robots (AMRs) ^[5, 6].

A landmark multicentre clinical study by Bullert et al. evaluated clinical performance across an initial consecutive series of 300 orthopaedic and trauma procedures utilising a self-driving, motorised mobile 3D C-arm ^[6]. Their findings confirmed:

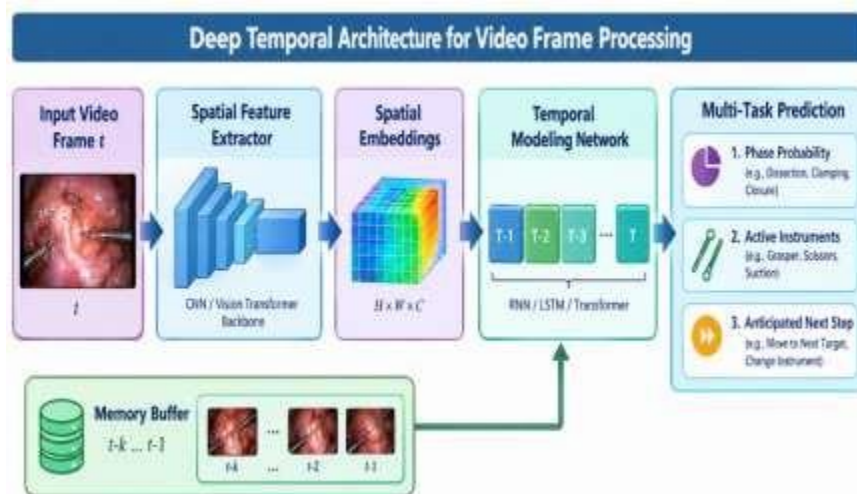
- **Sub-millimetre Target Repositioning Accuracy:** Automated orbital and angular alignments eliminated manual fine-tuning steps ^[6].
- **Significant Reductions in Cumulative Radiation Dose:** Automated trajectories reduced positioning scout shots by up to 42% ^[6].
- **Workflow Continuity:** Robotic pre-programmed trajectories minimised operative interruptions, allowing surgical teams to maintain procedural focus ^[5, 6].

3. SURGICAL SCENE UNDERSTANDING AND WORKFLOW RECOGNITION

Robotic actuation of imaging hardware cannot operate safely in a clinical vacuum; it requires continuous semantic awareness of surgical progress. Autonomous synergy depends upon real-time **surgical workflow recognition**—the computational capacity of an AI agent to identify surgical phases, procedural steps, and granular tool-tissue interactions from intraoperative video streams ^[8,9,10].

- **Deep Temporal Architectures for Phase and Step Recognition**

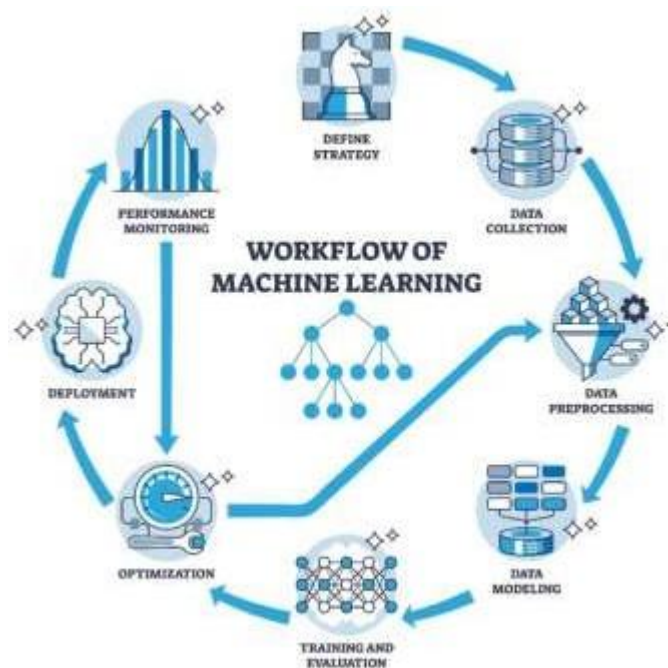
Early automated workflow detection relied on handcrafted visual features coupled with hidden Markov models (HMMs). A foundational shift occurred with Twinanda et al., who introduced **EndoNet**, a convolutional neural network (CNN) capable of simultaneously learning visual representations of surgical tools and classifying macroscopic operative phases from laparoscopic video ^[8].



Because surgical operations are inherently time-dependent, spatial CNNs alone cannot disambiguate visually identical operative steps occurring at different stages of surgery. To capture temporal context, Jin et al. developed **SV-RCNet**, integrating ResNet backbones with recurrent convolutional networks (specifically Long Short-Term Memory, or LSTM units) [9]. SV-RCNet established a benchmark for joint spatial-temporal feature modelling, enabling the system to track operative progressions smoothly without erratic frame-to-frame prediction jumps [9].

Building upon this, Jin et al. introduced multi-task recurrent convolutional networks with specialised correlation loss functions [10]. These systems simultaneously optimised tool presence detection and phase transitions, demonstrating that mutual constraints between instrument usage and procedural state substantially enhance classification robustness [10].

As the field progressed, convolutional temporal architectures expanded:



- **Real-Time Operational Feasibility:** Kitaguchi et al. and Guédon et al. demonstrated high clinical accuracy in laparoscopic colorectal and general procedures using real-time CNN inference, proving that deep learning systems can

process streaming operative video with minimal latency ^[11, 12].

- **Procedural Step Granularity:** Hashimoto et al. applied computer vision architectures to decompose laparoscopic sleeve gastrectomies into discrete clinical steps, establishing that granular step recognition provides actionable trigger points for auxiliary OR automation ^[13].
- **Temporal Convolutional Networks (TCNs):** Ramesh et al. deployed multi-task temporal convolutional networks, utilising dilated 1D convolutions to capture extensive temporal receptive fields more efficiently than recurrent architectures ^[16].
- **Long-Term Memory and Attention Networks:** Jin et al. developed the Temporal Memory Relation Network (TMRNet) ^[17], followed by **Trans-SVNet** ^[18], a hybrid architecture combining spatial CNN embeddings with multi-head self-attention Transformer blocks. Trans-SVNet effectively captured long-range procedural dependencies spanning across operative hours ^[17, 18].
- **Fine-Grained Scene Deconstruction: Action Triplets and Edge Computing**

To enable an autonomous imaging unit to execute micro-positioning, knowing the general phase ("Reconstruction") is insufficient. The system requires fine-grained action deconstruction.

Nwoye et al. advanced surgical AI with **Rendezvous (RDV)**, a model designed to recognise **surgical action triplets** formatted as < Instrument, action, target > (for instance, < bipolar forceps, cauterise, cystic artery >) ^[19]. Utilising cross-attention mechanisms between instrument visual tokens and anatomical region proposals, Rendezvous extracts explicit operational semantics from intraoperative imagery ^[19].

When an AI imaging orchestrator detects <screwdriver, tighten, pedicle screw>, it anticipates an immediate clinical need for fluoroscopic trajectory verification, automatically initiating gantry transit to the designated plane ^[19, 20].

Deploying these complex deep learning architectures in real-time intraoperative settings requires specialised edge computing infrastructure. Choksi et al. demonstrated that optimised neural network models running on high-throughput, low-power edge accelerators deployed within the OR suite achieve sub-35 millisecond frame-processing latencies ^[21]. This near-instantaneous inference speed eliminates lag, ensuring robotic C-arm systems can react immediately to intraoperative events without relying on external cloud connections ^[21].

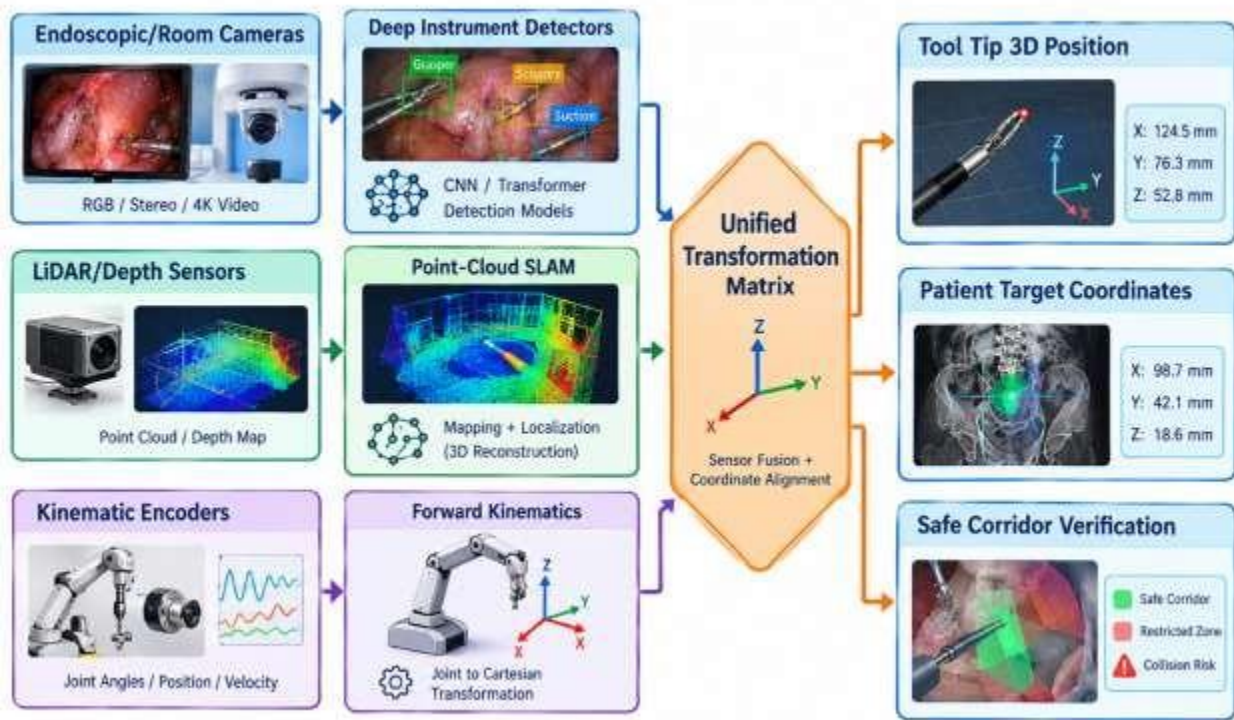
- **Surgomics and Multi-Center Context-Aware Platforms**

Expanding beyond individual operations, the integration of surgical data science into institutional workflows has given rise to **Surgomics**—the high-throughput extraction of phenotypic, behavioural, and procedural metrics from digital surgical video ^[22, 23].

In a multicentre study by Brandenburg et al. encompassing robot-assisted minimally invasive esophagectomies (RAMIE), context-aware machine learning models classified complex procedural variants across diverse patient anatomies and variable operating conditions ^[23]. By analysing institutional workflow signatures, these frameworks establish baseline behavioural expectations, enabling autonomous imaging systems to predict surgeon-specific imaging intervals, optimise standby positioning, and adapt to individual surgical styles ^[22, 23].

4. COMPUTER VISION FOR INSTRUMENT TRACKING AND ANATOMICAL REGISTRATION

For autonomous C-arms and robotic imaging platforms, real-time spatial awareness of surgical tools, patient anatomy, and table orientation is essential for collision-free trajectory planning and automated field-of-view centring^[26, 27, 28].



Spatial Tracking & Kinematic Sensor Fusion

- **Real-Time Surgical Instrument Localisation**

Tracking surgical instruments provides the spatial anchor points needed to guide motorised imaging beams directly onto active operative fields [29,30].



1. **Region-Based CNNs:** Sarikaya et al. and Jin et al. pioneered bounding-box and segmentation-level instrument localisation using Region-based Convolutional Neural Networks (Faster R-CNN), demonstrating high-accuracy tool identification across open and endoscopic surgeries [30, 31].

2. **Temporal Tracking-by-Detection:** Zhao et al. introduced recurrent convolutional tracking architectures that integrate Kalman filters with visual detection models to maintain uninterrupted instrument identification despite severe blood occlusions, camera smoke, and rapid tool exchanges [32].
3. **High-Precision Multi-Class Tracking:** Badilla-Solórzano et al. applied advanced deep detection networks to multi-instrument open cranio-maxillofacial surgery, proving that high-accuracy instrument tracking could be maintained within challenging open surgical fields [33].
4. **Modern Real-Time Detectors (UK-YOLOv10):** Advancing real-time capability, Zhang et al. formulated **UK-YOLOv10**, an ultra-fast, lightweight detector that achieves exceptional mean Average Precision (mAP) with frame processing rates exceeding 85 frames per second [34]. UK-YOLOv10 enables the imaging agent to accurately localise miniature instrument tips across high-resolution visual fields with minimal computational overhead [34].

- **Autonomous Skeletal Landmark Localisation**

Directing an autonomous C-arm to capture true anatomical projections requires accurate identification of underlying skeletal targets. In standard orthopaedic surgery, obtaining an unobstructed projection of a femoral neck fracture or a thoracic pedicle requires aligning the X-ray central beam perpendicular to precise internal bony axes [1, 5].

Jung et al. resolved this challenge by introducing deep neural networks for **autonomous skeletal landmark localisation** designed for agentic C-arm control [5]. Their framework bypasses iterative human adjustments through the following pipeline:

1. Low-dose initial orthogonal scout views are acquired.
2. Heatmap regression deep networks identify internal anatomical landmarks (e.g., femoral head centre, greater trochanter tip, pedicle margins) within the 2D projection space [5].
3. By back-projecting predicted 2D coordinates through calibrated camera-source geometry, the algorithm solves the Perspective-n-Point (PnP) problem, estimating the underlying 3D anatomical pose relative to the C-arm reference frame [5].
4. An autonomous control agent calculates the exact gantry translation and orbital tilt angles needed to centre the targeted structure in the primary imaging zone [5, 7].

- **Kinematic Modelling of C-Arm Fluoroscopy and Operating Tables**

Achieving seamless autonomous synergy requires modelling the coupled kinematic relationship between the imaging system and the patient support surface. Jaheen et al. introduced machine learning frameworks specifically designed to model C-arm fluoroscopy and operating table kinematics [7].

Traditional setups treat the C-arm and the articulated surgical table as independent mechanical systems. Consequently, tilting the table (e.g., Trendelenburg or lateral roll) disrupts established imaging planes, forcing extensive manual realignment [7].

Jaheen et al. developed forward and inverse kinematic mapping engines powered by artificial neural networks that dynamically link table encoder inputs with multi-axis C-arm gantry motors [7]. When the table position is adjusted, the imaging agent automatically compensates by driving the C-arm along corresponding orbital and translational vectors, preserving the target view without additional scout fluoroscopy shots [6, 7].

5. ROBOTIC CONTROL, AUTONOMY, AND WORKFLOW SYNERGY

Autonomous synergy in the operating suite represents the active coordination of perception, cognition, and physical execution.



- **Levels of Autonomy in Intraoperative Imaging**

1. **Level 0 (No Autonomy):** Traditional mobile C-arm; physical pushing, manual clamping, and manual foot-pedal radiation activation [1].
2. **Level 1 (Robot Assistance):** Power-assisted motorised axis motion; operator guides movements via joystick interfaces [3].
3. **Level 2 (Task Autonomy):** Automated execution of discrete subtasks, such as automated 3D cone-beam computed tomography (CBCT) rotational trajectories or automated transit between two pre-stored spatial coordinates [6].
4. **Level 3 (Conditional Autonomy):** The AI system processes intraoperative video, determines optimal projection angles based on detected surgical steps, and generates navigation trajectories, which execute once confirmed by the surgical team [5, 7].
5. **Level 4 (High Autonomy):** The imaging system monitors surgical flow, tracks instruments, coordinates with operating table kinematics, drives to target poses, acquires low-dose verification images, and clears the sterile perimeter autonomously, requiring human intervention only for emergency overrides [5, 6, 24].
6. **Level 5 (Full Autonomy):** Fully autonomous, self-governing imaging agent managing complete intraoperative radiology workflows across all procedural phases without human prompts. (Currently theoretical in surgical environments due to regulatory, clinical, and safety boundaries) [24].

- **Parallels from Endoscopic and Soft-Tissue Robotic Autonomy**

The mathematical foundations governing autonomous mobile C-arms draw heavily from parallel advances in soft-tissue robotic surgery and autonomous endoscope manipulation [24, 25].

1. **Autonomous Tissue Handling and Suturing:** Shademan et al. established that supervised autonomous robotic systems (such as the Smart Tissue Autonomous Robot, STAR) can perform soft-tissue anastomosis with precision surpassing experienced surgeons, utilising near-infrared fluorescent (NIRF) 3D markers and autonomous plan execution [24]. Saeidi et al. and Gao et al. advanced these concepts into minimally invasive paradigms by deploying autonomous suturing tools guided by predictive finite element models and real-time visual feedback [28, 29].
2. **Autonomous Endoscope Steering:** Early work by King et al. and Eslamian et al. introduced autonomous camera tracking systems on laparoscopic frames, translating instrument centroid vectors directly into robotic camera motions [25, 26]. Müller et al. and Shi et al. established cognitive, hands-free camera manipulation agents driven by reinforcement learning and gaze tracking, ensuring surgical tools consistently remain centred in the optimal visual field without human commands [27, 35].

- **Comparative Analysis of Autonomous Technologies**

Table 1: Comparative Evaluation of Intraoperative Imaging Localization and Autonomous C-Arm Platforms

System / Platform	Primary Modality & Sensors	Level of Autonomy	Positioning Accuracy	Radiation / Workflow Impact	Ref
CamC (Camera Augmented Mobile C-arm)	Co-axial optical video camera, semi-transparent mirror, planar X-ray	Level 1 (Assistive)	< 1.5 mm overlay registration error	30–50% reduction in targeting scout exposures	[1, 2]
RAY-POS Assistance System	Multi-directional LiDAR sensors, ambient spatial point cloud matching	Level 2 (Task Autonomy)	Repositioning error < 2.0 mm, < 1.0 deg	Eliminates manual alignment passes; zero scout shots for preset recall	[3]
Agentic Landmark Localization System	Deep heatmap regression CNNs, low-dose 2D scout X-rays, PnP solver	Level 3 (Conditional)	Landmark detection error < 1.8 mm	Single-step automated trajectory planning; avoids iterative fluoroscopy	[5]
Self-Driving Mobile 3D C-Arm (Clinical Series)	Motorized omnidirectional chassis, ultrasound/optical collision sensors, 3D CBCT	Level 3–4 (Conditional to High)	Sub-millimeter mechanical trajectory reproducibility (< 0.8 mm)	42% reduction in positioning scout shots; significant operative time savings	[6]
Kinematically Coupled C-Arm & Table	Table joint encoders, neural inverse kinematic mapping, motorized C-arm base	Level 3 (Conditional)	Angular tracking deviation < 0.5 deg across compound tilts	Eliminates manual re-centering after operating table adjustments	[7]

Table 2: Deep Learning Architectures for Surgical Workflow, Phase, and Tool Recognition

Model / Architecture	Target Recognition Task	Temporal Modeling Strategy	Benchmark Performance (Accuracy / F1 / mAP)	Ref
EndoNet	Dual-task: Phase recognition & tool presence detection	Frame-level sliding window averaging	Phase Accuracy: ~81.7%; Tool Detection mAP: ~81.0%	[8]
SV-RCNet	Laparoscopic surgical phase classification	Recurrent Convolutional Network (Long Short-Term Memory, LSTM)	Phase Accuracy: >88.0% across cholecystectomy benchmarks	[9]

Model / Architecture	Target Recognition Task	Temporal Modeling Strategy	Benchmark Performance (Accuracy / F1 / mAP)	Ref
Multi-Task RCN with Correlation Loss	Joint phase segmentation and multi-tool localization	Multi-task LSTM coupled with correlation constraint loss	Phase Accuracy: 89.5%; Tool Identification F1: 0.84	[10]
Multi-Task TCN	Hierarchical phase and operative step recognition	Dilated 1D Temporal Convolutional Networks (TCN)	Step Accuracy: 84.2%; Phase F1-Score: 0.87	[16]
TMRNet (Temporal Memory Relation Net)	Fine-grained workflow recognition	Non-local temporal memory banks and relational self-attention	Cholec80 Phase Accuracy: 90.1%; F1: 0.86	[17]
Trans-SVNet	Surgical workflow phase and step segmentation	Multi-Head Self-Attention Transformer mechanism	Cholec80 Phase Accuracy: 91.2%; Relaxed F1: 0.88	[18]
Rendezvous (RDV)	Surgical action triplet classification	Cross-attention fusion across tool, verb, and tissue tokens	Triplet Mean Average Precision (mAP): 34.5–38.2%	[19]
UK-YOLOv10	Real-time multi-instrument spatial detection & bounding	Feed-forward frame detection with temporal Kalman tracking	Instrument Detection mAP50: 92.4%	[34]

6. SAFETY, ERGONOMICS, AND QUALITY OF CARE

Autonomous synergy within interventional suites must directly improve patient safety, clinical efficacy, and working conditions for surgical staff [6, 20].

- **Radiation Minimisation via "Zero-Dose" Virtual Guidance**

Ionising radiation presents established stochastic and deterministic health hazards to both patients and clinical teams. Interventional surgeons face an increased lifetime incidence of cataracts, thyroid pathology, and malignancies [1, 3]. Manual C-arm repositioning relies heavily on iterative scout exposures to check trajectory alignment [1, 6].

Agentic imaging platforms minimise radiation through four primary mechanisms:

1. **Targeting Without Exposure:** Real-time spatial tracking allows the C-arm to drive directly to the targeted anatomical plane without exploratory X-ray shots [3, 5].
2. **Virtual Collimation and Beam Optimisation:** AI models predict the region of interest based on detected instruments and anatomy, automatically collimating the beam to the target zone and shielding adjacent healthy tissue [5, 20].
3. **Sparse-View 3D Reconstruction:** Deep neural networks reconstruct high-fidelity 3D volumetric CBCT datasets from sparse, low-dose 2D projections, significantly reducing overall exposure [6].
4. **Digital Radiographic Re-projection:** The system simulates expected fluoroscopic projections using pre-operative CT models and current gantry kinematics, reserving ionising radiation exclusively for final therapeutic confirmation [1, 7].

- **Spatial Envelopes and Dynamic Collision Avoidance**

Operating suites represent high-density physical workspaces where human surgeons, scrub nurses, perfusion lines, anaesthesia columns, and robotic arms coexist in proximity [3, 6]. Introducing a self-driving, heavy robotic C-arm introduces physical crush and collision risks.

To guarantee mechanical safety, autonomous mobile C-arms utilise layered safety envelopes:

1. **Perimeter Sensor Fusion:** Integrated time-of-flight (ToF) cameras, ultrasonic rangefinders, and tactile bumpers continuously monitor the immediate mechanical perimeter [3, 6].
2. **Sterile Dynamic Safety Corridors:** Real-time 3D spatial occupancy maps identify sterile personnel, unsterile zones, and patient contours, constructing dynamic clearance paths [4, 6].
3. **Speed Regulation and Safe Torque Limits:** The drive system continuously scales speed according to proximity measurements, instantly halting motion or reverting to back-drivable float modes if any contact is detected [6, 24].

- **Ergonomic Relief and Team Cognitive Bandwidth**

Clear communication in the operating room directly correlates with surgical outcomes. In standard fluoroscopy setups, verbal repositioning directives frequently produce confusion, iterative misalignments, and team frustration [3, 6].

By offloading machine positioning to autonomous agents triggered by procedural milestones, surgical teams recover valuable cognitive bandwidth [14, 20]. Surgeons can concentrate fully on critical tissue dissection and implant placement, while scrub nurses and circulating technicians avoid repeatedly manoeuvring heavy equipment into crowded sterile fields [6, 22].

7. TRANSLATIONAL ROADBLOCKS, ETHICS, AND REGULATORY FRONTIERS

Despite compelling clinical evidence, translating autonomous imaging systems into routine surgical practice presents significant technical, ethical, and regulatory hurdles.

- **Algorithmic Generalizability and the "Edge Case" Dilemma**

Deep learning models trained on uniform patient cohorts often struggle when encountering rare anatomical deformities, severe traumatic disruptions, previous metal hardware artifacts, or atypical surgical approaches [8, 14]. Overcoming these limitations requires training across diverse, multi-institutional repositories, standardising video collections, and deploying federated learning frameworks to preserve patient privacy [22, 23].

- **Explainability, Latency, and Edge Computing Constraints**

Surgical decision-making demands sub-second responsiveness and predictable system behaviour. Black-box neural networks that output positioning coordinates without explainable reasoning fail to inspire clinical trust [14, 20]. Integrating Explainable AI (XAI) ensures the surgical team retains clear insight into the agent's planned trajectory [19, 21].

- **Shared Control Paradigms and Liability Allocation**

Complete hand-off of physical machine movement to autonomous agents raises vital legal and ethical questions regarding liability [6, 24]. Current legal frameworks demand that licensed clinicians retain continuous oversight and manual override authority [20, 24].

- **Global Regulatory Pathways**

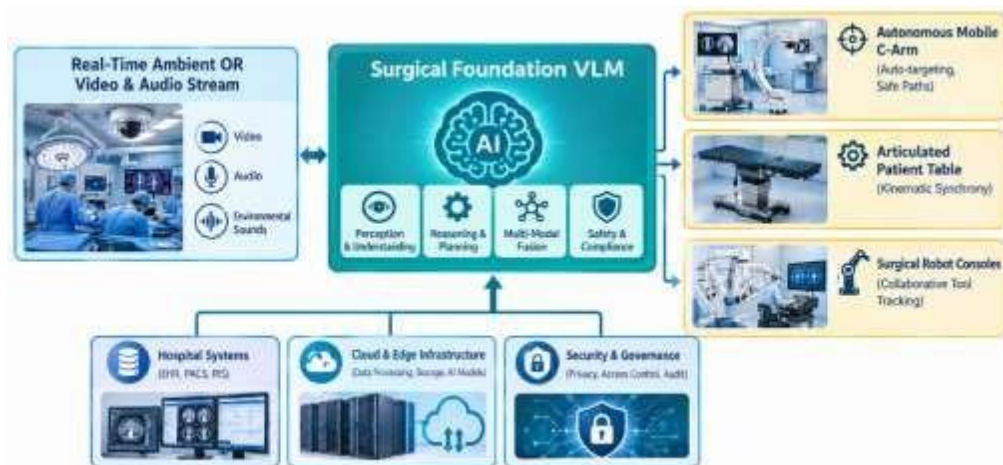
Regulatory bodies, including the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA), are adapting evaluation frameworks to accommodate adaptive AI systems. Classifying these platforms requires establishing stringent criteria covering physical collision avoidance, fail-safe software architectures (IEC 62304 standards), cybersecurity, and clinical validation across realistic multi-centre environments [21, 23].

8. FUTURE DIRECTIONS: AUTONOMOUS FOUNDATION MODELS AND THE HYBRID SUITE

The next decade of intraoperative technology will bridge isolated robotic systems into unified, fully synchronised operative environments.

- **Multi-Modal Surgical Vision-Language Models (VLMs)**

Emerging vision-language foundation models will reshape surgical scene interpretation. Instead of relying on separate narrow networks for tool detection, phase segmentation, and landmark localisation, unified multi-modal models will simultaneously process endoscopic video, ambient room feeds, verbal surgical dialogue, and vital signs [18, 19, 23].



Next-generation Integrated Hybrid Suite Ecosystem

- **Fully Synchronised Interventional Ecosystems**

The ultimate expression of autonomous synergy is the unified interventional suite. In this paradigm, mobile C-arms, primary surgical robots, smart operative tables, physiological monitors, and digital lighting arrays operate as a distributed cyber-physical network [7, 24, 35]. This multi-device coordination minimises visual obstructions, maintains sterile safety margins, eliminates positioning delays, and protects both patient and staff from unnecessary radiation [6, 20].

REFERENCES

- [1] Navab N, Heining SM, Traub J. Camera augmented mobile C-arm (CAMC): Calibration, accuracy study, and clinical applications. *IEEE Trans Med Imaging*. 2010;29(7):1412–1423. doi:10.1109/TMI.2009.2021947
- [2] von der Heide AM, Fallavollita P, Wang L, Sandner P, Navab N, Weidert S, Euler E. Camera-augmented mobile C-arm (CamC): A feasibility study of augmented reality imaging in the operating room. *Int J Med Robot Comput Assist Surg*. 2018;14(2):e1885. doi:10.1002/rcs.1885
- [3] Bernhard L, Völk C, Völk D, Rothmeyer F, Xu Z, Ostler D, Biberthaler P, Wilhelm D. RAY-POS: A LIDAR-based assistance system for intraoperative repositioning of mobile C-arms without external aids. *Int J Comput Assist Radiol Surg*. 2022;17(4):719–729. doi:10.1007/s11548-022-02571-w
- [4] Pauly O, Diotte B, Fallavollita P, Weidert S, Euler E, Navab N. Machine learning-based augmented reality for improved surgical scene understanding. *Comput Med Imaging Graph*. 2015;41:12–20. doi:10.1016/j.compmedimag.2014.06.007
- [5] Jung JH, Arrabi A, Luo J, Raymond S, Wshah S. Autonomous skeletal landmark localization toward agentic C-arm control. *Int J Comput Assist Radiol Surg*. 2026;21(5):1015–1023. doi:10.1007/s11548-026-03632-0
- [6] Bullert B, Ruebel L, Gierse J, Mandelka E, Brunken F, Gruetzner PA, Vetter SY. Robotic intraoperative imaging in orthopaedic and trauma surgery: Initial clinical experience with a self-driving mobile 3D C-arm in the first 300 cases. *Int*

- J Med Robot Comput Assist Surg. 2026;22(3):e70195. doi:10.1002/rcs.70195
- [7] Jaheen A, Gutta S, Fallavollita P. Modelling C-arm fluoroscopy and operating table kinematics via machine learning. *Int J Comput Assist Radiol Surg.* 2026. doi:10.1007/s11548-026-03640-x
 - [8] Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: A deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging.* 2017;36(1):86–97. doi:10.1109/TMI.2016.2593957
 - [9] Jin Y, Dou Q, Chen H, Yu L, Qin J, Fu CW, Heng PA. SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. *IEEE Trans Med Imaging.* 2018;37(5):1114–1126. doi:10.1109/TMI.2017.2787657
 - [10] Jin Y, Li H, Dou Q, Chen H, Qin J, Fu CW, Heng PA. Multi-task recurrent convolutional network with correlation loss for surgical video analysis. *Med Image Anal.* 2020;59:101572. doi:10.1016/j.media.2019.101572
 - [11] Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T, Oda T, Miura H, Yamanashi T, Watanabe M, Sato D, Sugomori Y, Hara S, Ito M. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc.* 2020;34(11):4924–4931. doi:10.1007/s00464-019-07281-0
 - [12] Guédon ACP, Meij SEP, Osman KNMMH, Kloosterman HA, van Stralen KJ, Grimbergen MCM, Eijssbouts QAJ, van den Dobbelsteen JJ, Twinanda AP. Deep learning for surgical phase recognition using endoscopic videos. *Surg Endosc.* 2021;35(11):6150–6157. doi:10.1007/s00464-020-08110-5
 - [13] Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, Lillemoe KD, Rus D, Meireles OR. Computer vision analysis of intraoperative video: Automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg.* 2019;270(3):414–421. doi:10.1097/SLA.0000000000003460
 - [14] Mohamadipanah H, Kears L, Witt A, Wise B, Yang S, Goll C, Pugh C. Can deep learning algorithms help identify surgical workflow and techniques? *J Surg Res.* 2021;268:318–325. doi:10.1016/j.jss.2021.07.003
 - [15] Zhang B, Ghanem A, Simes A, Choi H, Yoo A. Surgical workflow recognition with 3DCNN for sleeve gastrectomy. *Int J Comput Assist Radiol Surg.* 2021;16(11):2029–2036. doi:10.1007/s11548-021-02473-3
 - [16] Ramesh S, Dall’Alba D, Gonzalez C, Yu T, Mascagni P, Mutter D, Marescaux J, Fiorini P, Padoy N. Multi-task temporal convolutional networks for joint recognition of surgical phases and steps in gastric bypass procedures. *Int J Comput Assist Radiol Surg.* 2021;16(7):1111–1119. doi:10.1007/s11548-021-02388-z
 - [17] Jin Y, Long Y, Chen C, Zhao Z, Dou Q, Heng PA. Temporal memory relation network for workflow recognition from surgical video. *IEEE Trans Med Imaging.* 2021;40(7):1911–1923. doi:10.1109/TMI.2021.3069471
 - [18] Jin Y, Long Y, Gao X, Stoyanov D, Dou Q, Heng PA. Trans-SVNet: Hybrid embedding aggregation Transformer for surgical workflow analysis. *Int J Comput Assist Radiol Surg.* 2022;17(12):2193–2202. doi:10.1007/s11548-022-02743-8
 - [19] Nwoye CI, Yu T, Gonzalez C, Seeliger B, Mascagni P, Mutter D, Marescaux J, Padoy N. Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Med Image Anal.* 2022;78:102433. doi:10.1016/j.media.2022.102433
 - [20] Khalid MU, Laplante S, Masino C, Alseidi A, Jayaraman S, Zhang H, Mashouri P, Protserov S, Hunter J, Brudno M, Madani A. Use of artificial intelligence for decision-support to avoid high-risk behaviors during laparoscopic cholecystectomy. *Surg Endosc.* 2023;37(12):9467–9475. doi:10.1007/s00464-023-10403-4
 - [21] Choksi S, et al. Bringing artificial intelligence to the operating room: Edge computing for real-time surgical phase recognition. *Surg Endosc.* 2023;37(12):9455–9466. doi:10.1007/s00464-023-10322-4
 - [22] Loo T, McGlennen B, Incavo S, Hilger N. Measuring provider-level differences in perioperative workflow using computer vision-based artificial intelligence. *BMJ Health Care Inform.* 2025;32(1):e101591. doi:10.1136/bmjhci-2024-101591
 - [23] Brandenburg JM, Schulze A, Jenke AC, Bhasker N, Bleser N, Junger D, Stern A, Rivoir D, Naderi H, Fritz-Kebede F,

- Burgert O, Maier-Hein L, Mündermann L, Bodenstedt S, Speidel S, Lozanovski VJ, Grimminger PP, Billeter A, Klotz R, Weitz J, Distler M, Müller-Stich BP, Wagner M. Surgical workflow analysis for Surgomics and context-aware assistance in robot-assisted minimally invasive esophagectomy: A retrospective, single-arm, multicenter annotation and machine learning study. *Eur J Surg Oncol*. 2026;52(1):111174. doi:10.1016/j.ejso.2025.111174
- [24] Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PCW. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med*. 2016;8(337):337ra64. doi:10.1126/scitranslmed.aad9398
- [25] Eslamian S, Reisner LA, Pandya AK. Development and evaluation of an autonomous camera control algorithm on the da Vinci Surgical System. *Int J Med Robot Comput Assist Surg*. 2020;16(2):e2036. doi:10.1002/rcs.2036
- [26] King BW, Reisner LA, Pandya AK, Compusto AM, Ellis RD, Klein MD. Towards an autonomous robot for camera control during laparoscopic surgery. *J Laparoendosc Adv Surg Tech A*. 2013;23(12):1027–1030. doi:10.1089/lap.2013.0304
- [27] Müller B, et al. A learning robot for cognitive camera control in minimally invasive surgery. *Int J Comput Assist Radiol Surg*. 2021;16(11):1989–1997. doi:10.1007/s11548-021-02485-z
- [28] Gao S, Ji S, Feng M, Lu X, Tong W. A study on autonomous suturing task assignment in robot-assisted minimally invasive surgery. *Int J Med Robot Comput Assist Surg*. 2021;17(1):e2180. doi:10.1002/rcs.2180
- [29] Saeidi H, Le HND, Opfermann JD, Leonard S, Kim A, Hsieh MH, Kang JU, Krieger A. Autonomous laparoscopic robotic suturing with a novel actuated suturing tool and 3D endoscope. In: *Proc IEEE Int Conf Robot Autom (ICRA)*. 2019:1541–1547. doi:10.1109/ICRA.2019.8794306
- [30] Jin A, Yeung S, Jopling J, Krause J, Azagury D, Milstein A, Fei-Fei L. Tool detection and operative skill assessment in surgical videos using region-based convolutional neural networks. In: *Proc IEEE Winter Conf Appl Comput Vis (WACV)*. 2018:691–699. doi:10.1109/WACV.2018.00081
- [31] Sarikaya D, Corso JJ, Guru KA. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Trans Med Imaging*. 2017;36(7):1542–1549. doi:10.1109/TMI.2017.2665671
- [32] Zhao Z, Voros S, Weng Y, Chang F, Li R. Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method. *J Med Robot Res*. 2017;2(Suppl 1):1740003. doi:10.1080/24699322.2017.1378777
- [33] Badilla-Solórzano J, Spindeldreier S, Ihler S, Gellrich NC, Spalthoff S. Deep-learning-based instrument detection for intra-operative robotic assistance. *Int J Comput Assist Radiol Surg*. 2022;17(10):1785–1793. doi:10.1007/s11548-022-02715-y
- [34] Zhang L, Guo G, Wang W. UK-YOLOv10: Deep learning-based detection of surgical instruments. *Int J Med Robot Comput Assist Surg*. 2025;21(3):e70083. doi:10.1002/rcs.70083
- [35] Shi J, Zhou C, Wang L, Lin W, Zhou S, He Z, Ma H, Hu J, Feng D. Hands-free camera assistant: Autonomous laparoscope manipulation in robot-assisted surgery. *Int J Med Robot Comput Assist Surg*. 2025;21(4):e70103. doi:10.1002/rcs.70103