

Review

Not peer-reviewed version

A Clinical Decision Support System for Post-Surgical Cardiovascular Remote Monitoring

[Charalampia Pylarinou](#)*, [Efstratios Koletsis](#), [Francesk Mulita](#)*, [Vasileios Leivaditis](#), [Elias Liolis](#),
Lefteris Gortzis, Dimosthenis Mavrilas

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1768.v1

Keywords: retrieval-augmented generation; clinical decision support; evidence-based medicine; artificial intelligence; cardiovascular surgery; remote monitoring



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Review

A Clinical Decision Support System for Post-Surgical Cardiovascular Remote Monitoring

Charalampia Pylarinou ^{1,*}, Efstratios Koletsis ², Francesk Mulita ³, Vasileios Leivaditis ⁴, Elias Liolis ⁵, Lefteris Gortzis ⁶ and Dimosthenis Mavrilas ⁷

¹ Department of Mechanical and Aeronautical Engineering, University of Patras, Patras 26504, Greece

² Department of Cardiothoracic Surgery, University Hospital of Patras, Patras 26504, Greece

³ Department of Surgery, University Hospital of Patras, Patras 26504, Greece

⁴ Department of Cardiothoracic Surgery, Westpfalz-Klinikum, Kaiserslautern, Germany

⁵ Department of Oncology, University Hospital of Patras, Patras 26504, Greece

⁶ Research & Development, CAREPOI (™), Patras 26221, Greece

⁷ Department of Mechanical and Aeronautical Engineering, University of Patras, Patras 26504, Greece

* Correspondence: hpylarin@upatras.gr

Abstract

Background: Post-surgical cardiovascular monitoring places a heavy information burden on clinical teams, requiring rapid synthesis of patient history, intraoperative data, monitoring streams, and surgical outcomes evidence. Existing clinical decision support systems handle this integration poorly, and most offer little visibility into their reasoning. We present a Retrieval-Augmented Generation (RAG) architecture designed specifically for this domain, with a focus on evidence traceability and practical workflow integration. **Methods:** We developed a three-layer RAG architecture comprising: a retrieval layer to create 768-dimensional representations of clinical scenarios; an augmentation layer employing context-aware filtering and machine learning algorithms to integrate patient-specific data with retrieved evidence; and a generative layer using fine-tuned language models to synthesise actionable clinical recommendations. **Results:** An evaluation framework is proposed to assess the technical performance and clinical applicability of RAG architecture. The evaluation methodology encompasses technical validation of system components, assessment of clinical workflow integration potential, and analysis of interpretability features essential for healthcare deployment. **Conclusions:** We describe a RAG architecture for post-surgical cardiovascular monitoring in which every recommendation is linked to retrievable source documents, making the reasoning visible and challengeable. A structured evaluation framework is proposed to guide the system towards clinical validation.

Keywords: retrieval-augmented generation; clinical decision support; evidence-based medicine; artificial intelligence; cardiovascular surgery; remote monitoring

Introduction

Post-surgical cardiovascular care generates an information burden that current clinical decision support systems (CDSS) handle poorly.^{1,2} At the bedside, clinicians must rapidly reconcile patient history, intraoperative events, real-time monitoring streams, and a growing surgical outcomes literature—tasks that rule-based or opaque machine learning tools support only in part.³ When these tools do offer guidance, their reasoning is rarely visible, and this lack of transparency remains a practical barrier to adoption in cardiac surgery units.⁴

Retrieval-Augmented Generation (RAG) offers a different approach. Rather than encoding all knowledge in model weights, RAG queries a curated evidence base at inference time, so that each recommendation is grounded in retrievable, inspectable source documents.⁵ This traceability is

important for clinical adoption: a surgeon needs not just a risk estimate but a visible reasoning chain that can be examined and, if necessary, challenged.⁶

Deploying RAG in a clinical setting introduces its own constraints. Workflow integration, evidence quality control, and regulatory audit requirements all shape what an acceptable system looks like in practice.⁷ In cardiac surgery the pace of postoperative deterioration means that response latency matters as much as classification accuracy.⁸ The postoperative monitoring component of our system drew on the DETER algorithm,¹⁶ which showed that combining device-based telemetry with generative AI can produce actionable deterioration warnings early enough to alter management. Recent studies confirm that continuous wireless telemetry after cardiac surgery detects arrhythmias earlier than intermittent monitoring,¹⁷ and wearable-based early warning scores show promising performance as part of post-surgical escalation systems.¹⁸

We describe a purpose-built RAG architecture for post-surgical cardiovascular monitoring. The central design goal was traceability: every recommendation the system produces can be linked back to the specific retrieved evidence that generated it. A secondary goal was realistic integration with the perioperative workflows already in place at participating centres.

The study pursued five objectives: (1) design a RAG architecture suited to cardiovascular surgery decision support; (2) build domain-specific retrieval and evidence integration components; (3) make the reasoning behind AI-generated recommendations fully auditable; (4) propose an evaluation framework appropriate for a system at this stage of development; and (5) assess its potential to augment clinical decision-making while preserving physician oversight.

Methods

RAG Architecture Overview

The proposed RAG architecture for clinical decision support consists of three interconnected layers designed to address the specific requirements of post-surgical cardiovascular monitoring. Figure 1 illustrates the architecture that integrates domain-specific knowledge sources, employs clinical workflow-aware processing, and maintains complete interpretability throughout the decision support process.

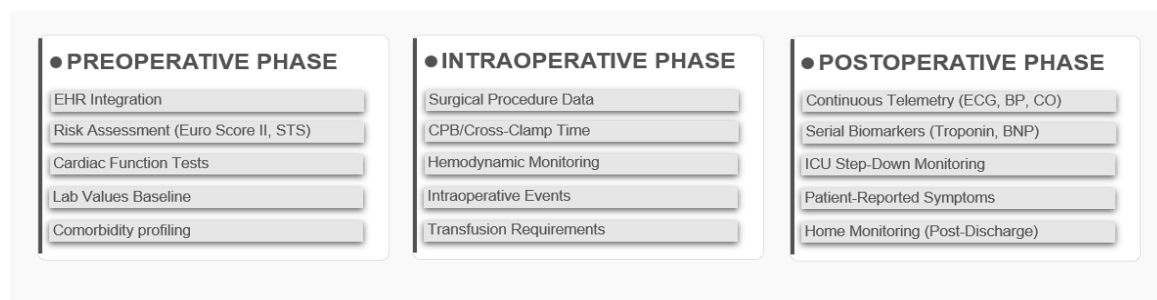


Figure 1. Overview of the RAG (Retrieval-Augmented Generation) architecture for clinical decision support in post-surgical cardiovascular remote monitoring, showing integration of multi-source knowledge retrieval with patient-specific data.

The system operates through a systematic workflow where clinical queries trigger retrieval of relevant evidence from multiple knowledge sources, followed by context-aware integration and augmentation of this evidence with patient-specific data, culminating in the generation of interpretable clinical recommendations with full evidence provenance. When the system generates a deterioration prediction, it simultaneously presents the "top-k relevant records" (e.g., similar surgical cases, complication management guidelines, outcome literature) that informed its assessment, enabling clinicians to scrutinize the internal logic and verify its alignment with their own clinical reasoning and established cardiovascular surgery evidence.

The overall pipeline is illustrated in Figure 2. Each layer has a defined input and output, and the interfaces between them are designed so that individual components can be tested and updated independently.



Figure 2. Retrieval-Augmented Generation (RAG) pipeline for post-surgical cardiovascular monitoring, illustrating the Input Layer data streams, Retrieval Layer knowledge retrieval, Augmentation Layer evidence integration, and Generative Layer clinical recommendation synthesis.

Input Layer

This layer is responsible for collecting comprehensive data streams across the surgical continuum as displayed in Table 1.

Table 1. Data sources for the post-surgical monitoring RAG system.

Preoperative Phase	Intraoperative Phase	Postoperative Phase
Structured data from Electronic Health Records (EHRs): cardiac history, comorbidities (diabetes, hypertension, renal dysfunction), medications, baseline cardiac function (ejection fraction, valve function)	Surgical procedure data: type of surgery (CABG, valve replacement, combination), CPB time, aortic cross-clamp time	Vital Signs: Wearable continuous telemetry (ECG, SpO ₂ , RR, temperature, blood pressure). Integration with CAREPOI (™) perioperative telemetry monitoring system via HL7/FHIR. Early detection of haemodynamic instability, arrhythmias, and respiratory compromise.
Preoperative diagnostic tests: ECG, echocardiography, cardiac catheterization, stress tests, laboratory values	Intraoperative events: arrhythmias, haemodynamic instability, transfusion requirements / Haemodynamic parameters during surgery / Laboratory trend monitoring: troponin, lactate, creatinine / Serial biochemical markers: troponin, BNP/NT-proBNP, lactate, creatinine, inflammatory markers	Alerts & Early Warning Scores: Clinical deterioration early warning scores (NEWS2, MEWS) computed continuously from wearable telemetry. DETER-type algorithms ¹⁶ provide real-time deterioration risk stratification, triggering escalation pathways and clinician notifications.
Risk assessment scores: EuroSCORE II, STS Predicted Risk of Mortality and Morbidity	Continuous biosignal monitoring from ICU/step-down telemetry: ECG (continuous), arterial blood pressure (invasive/non-invasive), cardiac	AI-Driven Deterioration Prediction: DETER-type algorithms ¹⁶ combining device-based biosignal telemetry with Generative AI for real-time

Patient-reported symptoms and functional status / Medication administration records	output monitoring, central venous pressure, pulmonary artery pressure (when available) Clinical assessments: physician and nursing notes, chest tube output, urine output, fluid balance / Imaging: chest X-rays, echocardiography when indicated / Anaesthesia records and intraoperative monitoring data	composite risk scoring; post-discharge readmission prediction; integration of troponin, BNP/NT-proBNP, and creatinine trends into deterioration models. Complications Detection & NLP Processing: SSI, DVT, sepsis, AKI detected via NLP on free text (BioBERT ¹⁹ , cTAKES ²⁰). Patient-reported symptoms (chest pain, dyspnoea, palpitations) extracted from recovery questionnaires and telehealth interactions. Wearable activity monitors and patient-reported outcome measures (PROMs) integrated for functional recovery tracking.
---	---	---

CABG: coronary artery bypass grafting; CPB: cardiopulmonary bypass; ECG: electrocardiogram; SpO₂: peripheral oxygen saturation; RR: respiratory rate; NEWS2: National Early Warning Score 2; MEWS: Modified Early Warning Score; SSI: surgical site infection; DVT: deep vein thrombosis; AKI: acute kidney injury; NLP: Natural Language Processing; BNP: B-type natriuretic peptide; NT-proBNP: N-terminal proBNP; PROM: patient-reported outcome measure.

Such data are drawn from CAREPOI (™) perioperative telemetry monitoring system⁹; the system connects to hospital HIS/EHR and intensive care unit (ICU) monitoring systems via HL7/Fast Healthcare Interoperability Resources (FHIR) interoperability standards, where applicable. Patient-reported symptoms during recovery are processed via two complementary NLP tools: BioBERT¹⁹, a biomedical domain pre-trained BERT model used to extract subjective yet essential inputs about chest pain, dyspnoea, palpitations, or other concerning symptoms from free-text patient narratives and telehealth consultation transcripts; and cTAKES (Apache Clinical Text Analysis and Knowledge Extraction System)²⁰, which is used in parallel for extracting clinical concepts—including symptoms, medications, procedures, and anatomical sites—from free text in EHRs through its named entity recognition, negation detection, and temporal reasoning pipelines. Together, BioBERT and cTAKES provide complementary coverage: BioBERT excels in sentence-level semantic understanding and entity classification across diverse biomedical text, while cTAKES provides structured clinical concept extraction mapped to standardised terminologies (SNOMED CT, RxNorm) essential for downstream ICD coding and clinical reasoning.

A rule-based system incorporating established post-surgical complication signatures—implemented using HL7 FHIR Clinical Reasoning (specifically the FHIR ClinicalImpression, ClinicalReasoning, and PlanDefinition resources)—flags abnormal parameter combinations for clinician review. This approach encodes published clinical guidelines (e.g., STS/ACC/AHA post-cardiac surgery complication criteria, ESC atrial fibrillation guidelines) as computable clinical decision logic within FHIR-native rule sets, enabling deterministic, interpretable, and standards-compliant identification of high-risk parameter constellations such as new-onset arrhythmia combined with troponin elevation, or fever with wound erythema and leukocytosis. The rule engine output is then passed to Transformer-based models to process sequential physiological data across the perioperative timeline.

BioBERT-based medical entity identification is used for entity recognition and sentence segmentation, abbreviation expansion, and negation detection as part of clinical text preprocessing. Transformers are highly parallelisable, making them computationally efficient for processing high-frequency continuous monitoring streams, especially in real-time intensive care unit (ICU)

environments. Together, these preprocessing steps give the DSS a structured, temporally ordered data representation—a prerequisite for accurate deterioration modelling. Integration ensures the DSS has a holistic, temporal understanding of the patient's surgical journey, crucial for accurate deterioration prediction.¹⁰

Retrieval Layer Design

The knowledge base used by the retrieval layer draws on three main sources: peer-reviewed cardiovascular surgery literature, anonymised institutional outcome data from participating centres, and current clinical practice guidelines. Combining these sources improves the specificity and clinical relevance of retrieved evidence across the perioperative continuum, including procedure-specific profiles for 97 cardiovascular surgical procedures.¹¹

Using provided application programming interfaces (APIs), the retrieval layer queries relevant internal surgical outcome databases and external cardiovascular literature in response to patient-specific deterioration risk queries. Pre-processing pipelines normalise diverse data formats (JSON, HL7, and FHIR) allowing seamless interoperability across perioperative systems. Dense vector search techniques using Sentence-BERT embeddings enable fast, semantically accurate query resolution.¹²

The system employs domain-adapted BERT models to create 768-dimensional vector representations of clinical scenarios and knowledge base content.¹³ The embedding process involves Clinical Query Processing, Knowledge Base Vectorisation, and a Retrieval Algorithm.

When the system generates a deterioration prediction, it simultaneously presents the “*top-k relevant records*” (e.g., similar surgical cases, complication management guidelines, outcome literature) that informed its assessment. RAG architecture allows clinicians to see the factual basis for the deterioration prediction, enabling them to scrutinise the internal logic and verify its alignment with their own clinical reasoning and established cardiovascular surgery evidence. This directly addresses the “*black box*” problem of many AI systems in critical care. This design choice ensures that the DSS's outputs are not just statistical predictions but are grounded in verifiable surgical outcome evidence, fostering trust and explainability.¹⁴ Implementation involves building and maintaining robust search indexes of surgical outcomes and embedding models trained in cardiovascular surgery literature.

The retrieval component implements semantic search across cardiovascular surgery literature, institutional outcome databases, and clinical practice guidelines to identify contextually relevant information for patient-specific risk assessment. The system employs dense vector representations using Sentence-BERT embeddings to encode patient presentations and retrieve similar prior cases from the indexed surgical knowledge bases. ClinicalBERT, trained on MIMIC-III clinical notes, may additionally be deployed for clinical note-specific retrieval tasks. Knowledge base construction encompasses multiple complementary data sources including peer-reviewed cardiovascular surgery literature from PubMed.

Vector embedding generation utilises domain-adapted BERT models pre-trained on cardiovascular surgery literature to create 768-dimensional representations of clinical scenarios. Patient presentations undergo systematic preprocessing including medical entity extraction, temporal sequence identification, and clinical context normalisation before embedding generation.

The retrieval process utilises cosine similarity scoring to identify relevant surgical outcomes, complication patterns, and evidence-based intervention strategies. Similarity threshold optimisation ensures high-quality evidence retrieval while maintaining computational efficiency. The system incorporates temporal relevance weighting to prioritise recent evidence while maintaining access to foundational clinical knowledge. A mechanism is in place to validate retrieved evidence through automated fact-checking against established clinical databases and expert-curated knowledge bases.

Augmentation Layer Implementation

Retrieved evidence undergoes systematic integration and contextualisation through machine learning algorithms trained on cardiovascular surgery outcomes data. Ensemble Meta-Learning was

used combining Gradient Boosting, with Random Forest and XGBoost so as to obtain cardiovascular surgery outcomes with expert annotation. Risk estimates derived from this ensemble approach include confidence intervals, which are reported alongside the recommendation to give clinicians a sense of prediction certainty.¹⁵ The augmentation step combines patient-specific risk factors, surgical complexity, institutional benchmarks, and guideline recommendations to produce individualised risk profiles and ranked intervention options.

Context-aware filtering removes irrelevant or contradictory information while preserving clinically significant evidence. Multi-source evidence reconciliation identifies consistent patterns across different knowledge sources and highlights potential conflicts requiring clinical judgment.

Patient-specific contextualisation adjusts general clinical evidence to account for individual patient characteristics including age, comorbidities, surgical complexity, and institutional factors. Personalisation algorithms weight evidence relevance based on similarity to the current patient's clinical profile. The augmentation layer generates structured clinical insights including risk factor identification, complication probability estimates, evidence-based intervention recommendations, and monitoring intensity suggestions.

Generative Layer Architecture

Large Language Models (LLMs)—initially a fine-tuned version of BERT (Bidirectional Encoder Representations from Transformers) specialised on cardiovascular surgery literature, followed by Gemini 2.5 Pro—process the retrieved surgical outcome information to synthesise deterioration risk predictions. These predictions include risk stratification scores for specific complications (myocardial infarction, arrhythmias, heart failure, infection, bleeding), clinical reasoning outputs explaining the prediction basis, and diagnostic/intervention suggestions. The generative output is grounded in the patient's specific surgical context, preoperative risk factors, intraoperative course, and current postoperative trajectory, aligning with validated post-surgical complication patterns to ensure clinical relevance and minimise AI hallucinations.

Model architecture employs a two-stage approach beginning with domain-specific fine-tuning of BERT on cardiovascular surgery literature and clinical guidelines. This specialised foundation model develops deep understanding of cardiovascular surgery terminology, complication patterns, and intervention strategies. Subsequently, integration with large language models enables sophisticated reasoning and natural language generation of clinical recommendations.

Training used anonymised records from multiple centres, covering a range of procedural types and patient demographics. The primary outcome labels were 30-day mortality, major adverse cardiac events, readmission, and functional recovery. Supervised fine-tuning was combined with reinforcement learning from human feedback to keep model outputs consistent with expert clinical judgment.

The generative component is designed to produce structured clinical recommendations including risk stratification scores with confidence intervals, evidence-based intervention recommendations with supporting citations, monitoring intensity guidelines tailored to predicted risk levels and anticipated complication signatures with temporal progression patterns.

Decision Support Integration

The system provides tailored clinical suggestions including urgent diagnostic testing recommendations (repeat ECG, echocardiography, cardiac catheterisation), medication adjustments (inotropes, antiarrhythmics, diuretics), intensification of monitoring frequency, specialist consultation recommendations (cardiology, cardiac surgery), and transfer to higher level of care (step-down to ICU).

Evaluation Methodology

The objective of this evaluation methodology is to provide a basic framework for assessing the proposed RAG-based clinical decision support system, recognising that comprehensive clinical validation would require extensive real-world deployment and outcome measurement beyond the scope of this initial system design study. Given the absence of actual clinical results at this stage, the evaluation framework focuses on technical feasibility assessment, expert validation of system components, and preliminary analysis of clinical applicability.

Technical Component Validation

The evaluation begins with systematic assessment of each RAG architecture component to ensure technical robustness and clinical relevance. For the retrieval layer, evaluation focuses on knowledge base completeness and quality, assessing coverage of major cardiovascular surgery guidelines and literature, currency of evidence sources, and accuracy of semantic similarity matching. Expert cardiovascular surgeons would review sample retrievals to validate clinical relevance and appropriateness of returned evidence for specific clinical scenarios. Response time benchmarking ensures compatibility with clinical workflow requirements, targeting sub-3-second response times for routine queries.

The augmentation layer undergoes evaluation for evidence integration quality and patient-specific contextualisation accuracy. Clinical experts assess the appropriateness of evidence filtering, contradiction resolution mechanisms, and personalisation algorithms through review of sample cases representing diverse patient profiles and clinical scenarios. Inter-rater reliability assessment among clinical reviewers provides quantitative validation of system performance consistency. The generative layer evaluation emphasises clinical appropriateness and safety of generated recommendations, with expert panels reviewing sample outputs for medical accuracy, completeness, and alignment with established clinical guidelines.

Interpretability and Trust Assessment

A critical evaluation component focuses on the system's interpretability features, essential for clinical adoption and regulatory compliance. Clinical users evaluate evidence provenance completeness, reasoning chain clarity, and confidence scoring accuracy through structured review sessions. The assessment examines whether clinicians can effectively trace recommendation origins, understand the logical basis for AI-generated suggestions, and appropriately weight recommendations based on confidence indicators. Expert evaluation of alternative option presentation ensures comprehensive decision support rather than singular recommendations.

Comparative analysis against existing clinical decision support tools provides context for system performance evaluation. While full clinical outcome comparison requires extensive prospective studies, initial assessment can compare recommendation quality, evidence citation completeness, and user interface effectiveness against established cardiovascular surgery risk assessment tools such as EuroSCORE II and STS risk calculators.

Clinical Workflow Integration Analysis

Preliminary workflow integration assessment examines system compatibility with existing clinical processes and electronic health record systems. Healthcare informatics specialists and clinical users review system interfaces, data integration requirements, and workflow modification needs through structured evaluation sessions. This assessment identifies potential implementation barriers, training requirements, and technical infrastructure needs for successful deployment.

User experience evaluation with cardiovascular surgeons, cardiologists, and critical care physicians provides insights into system usability, clinical relevance, and acceptance potential. Through controlled demonstration sessions using representative clinical scenarios, evaluators assess system response appropriateness, interface intuitiveness, and perceived clinical value. Qualitative

feedback gathering identifies refinement priorities and implementation considerations for future deployment phases.

Safety and Risk Assessment

Given the high-stakes nature of post-surgical cardiovascular care, safety assessment constitutes a fundamental evaluation component. Expert clinical panels review system outputs for potential adverse recommendations, contraindicated suggestions, or clinically inappropriate guidance. Safety evaluation protocols include assessment of recommendation conservatism, appropriate uncertainty expression, and robust handling of edge cases or unusual clinical presentations.

Risk assessment examines system robustness under various conditions including incomplete data, conflicting evidence sources, and atypical patient presentations. Technical risk evaluation addresses system reliability, data security, and privacy protection mechanisms essential for healthcare deployment. This preliminary safety assessment provides foundation for more comprehensive risk management protocols required for clinical implementation.

Expected Evaluation Outcomes

The expected outcomes are: validation that each architectural component functions within specification; expert confirmation that retrieved evidence is clinically relevant; and identification of the infrastructure and training requirements for site deployment. Specific technical targets are retrieval precision >90% for relevant evidence, query response time <3 seconds, and a clinical appropriateness rating of >85% from expert review of generated recommendations.

Clinical assessment outcomes focus on interpretability validation, workflow compatibility confirmation, and identification of user training and support requirements. Safety evaluation establishes preliminary risk profiles and identifies additional safeguards needed for clinical deployment. The evaluation provides foundation for future clinical validation studies while demonstrating system readiness for pilot implementation in controlled clinical environments.

Discussion

Clinical Relevance and Contribution

To our knowledge, this is the first RAG architecture designed specifically for post-surgical cardiovascular decision support. The central contribution is not the RAG framework itself, which is well established in natural language processing, but rather its adaptation to the data types, evidence sources, and clinical constraints of cardiac surgery. The three-layer design makes the reasoning process transparent: clinicians can see which retrieved cases and guidelines shaped each recommendation, rather than receiving an unexplained risk score.

In post-surgical care, delayed recognition of deterioration has measurable consequences for mortality and morbidity. Interpretability is not only a regulatory requirement in this context—it is a clinical one. Surgeons and intensivists are more likely to act on a recommendation when they can see why it was made and which patients or outcomes it draws on.

Methodological Contributions

The proposed RAG architecture demonstrates several methodological innovations relevant to healthcare AI applications. The 768-dimensional semantic embedding approach enables sophisticated matching between clinical scenarios and relevant evidence, going beyond keyword-based search to capture conceptual similarity between clinical situations. The context-aware evidence integration addresses the challenge of personalising population-level evidence to individual patient characteristics, a critical requirement for effective clinical decision support.

Integrating multiple knowledge sources—published literature, guideline repositories, and institutional outcome data—requires active quality management; contradictory evidence between

sources is surfaced rather than silently resolved. The provenance tracking built into the generative layer means that every output carries a citation trail, which also supports post-hoc audit if a recommendation is later questioned.

Evaluation Approach and Path to Clinical Validation

The proposed evaluation methodology provides a structured approach to assessing RAG-based clinical decision support systems, addressing both technical performance and clinical applicability. While this initial evaluation focuses on system design validation and expert assessment, it establishes foundation for comprehensive clinical validation studies necessary for widespread deployment.

The emphasis on interpretability assessment and safety evaluation reflects the unique requirements of healthcare AI applications, where system transparency and risk management are paramount. The evaluation framework could serve as a template for assessing other clinical AI applications, providing standardised approaches to technical validation, clinical relevance assessment, and safety evaluation.

Implementation and Future Work

The modular architecture design facilitates gradual implementation and iterative refinement based on clinical feedback and performance assessment. The system's integration with existing electronic health record systems and clinical workflows ensures practical feasibility while minimising disruption to established care processes. Future development priorities include expansion to additional clinical specialties, integration with real-time physiological monitoring systems, and development of adaptive learning mechanisms that continuously improve based on clinical outcomes.

The evaluation methodology provides a pathway for systematic validation and refinement. Future priorities include multi-institutional validation studies, cost-effectiveness analysis, and investigation of long-term impact on clinical outcomes.

Limitations and Considerations

Several limitations should be acknowledged in this system design and evaluation approach. The focus on post-surgical cardiovascular monitoring may limit immediate generalisability to other clinical domains, though the architectural principles could be adapted for broader applications. The evaluation methodology is detailed but resource-intensive; it will require close collaboration with clinical sites and cannot be fully executed within a single-centre study.

The absence of actual clinical deployment results limits conclusions about real-world effectiveness and safety, highlighting the need for prospective validation studies. The system's dependence on high-quality clinical data and expert annotation for optimal performance represents an implementation consideration that must be addressed through robust data governance and quality assurance procedures.

Conclusions

We have described a RAG-based clinical decision support system for post-surgical cardiovascular monitoring, with a design emphasis on evidence traceability and workflow compatibility. The three-layer architecture—retrieval, augmentation, generative—allows each component to be validated and refined independently, and the proposed evaluation framework provides concrete performance targets across technical, clinical, and safety dimensions.

The system is not yet clinically deployed, and the evaluation framework described here sets out what will need to be demonstrated before that step is taken. The next phase of this work will involve single-site piloting alongside existing systems, with the primary aim of establishing whether the AI-generated recommendations are accurate enough and fast enough to be useful in a real post-surgical

setting. If that evidence is positive, expansion to additional centres and cardiac surgery subspecialties would follow.

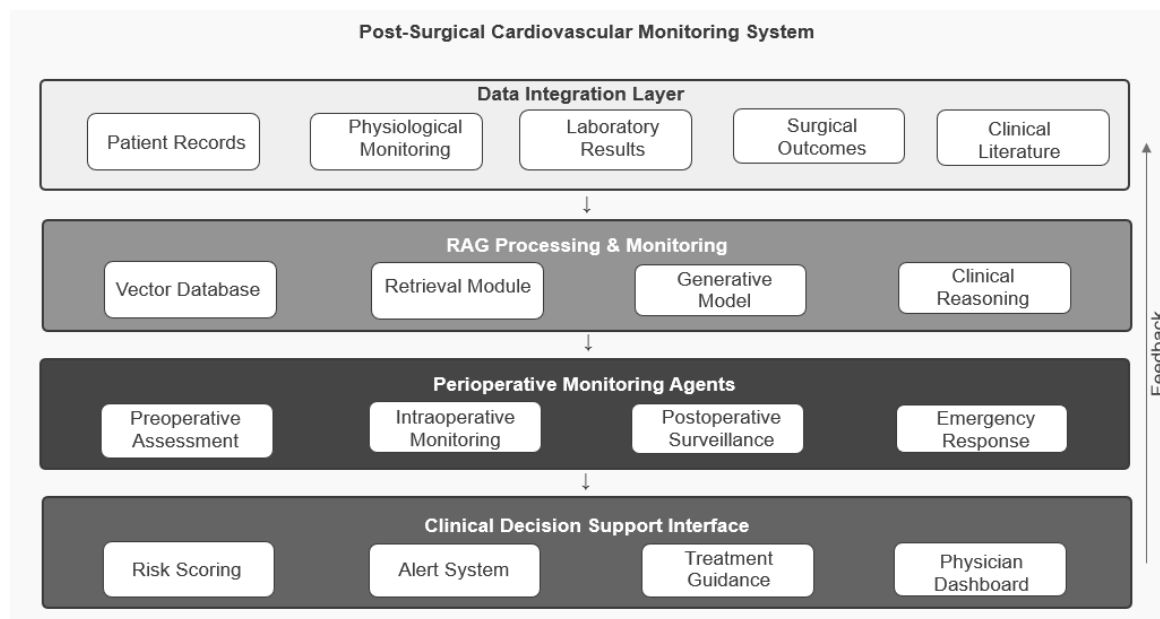


Figure 3. Conceptual three-layer architecture of the RAG-based clinical decision support system: Retrieval Layer → Augmentation Layer → Generative Layer. The system integrates cardiovascular surgery knowledge bases, real-time patient monitoring data, and clinician feedback loops to support evidence-based post-surgical decision making.

Acknowledgments: The authors acknowledge the clinical teams and healthcare institutions that provided expertise during system design and evaluation planning. We thank the technical staff for their contributions to system development and the clinical advisory committee for their guidance on evaluation methodology design. This work was partly supported by the CAREPOI organisation.

Ethics Approval and Consent to Participate: This study presents a system design and evaluation methodology. Future clinical validation studies will be conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines, with ethics approval obtained from institutional review boards prior to study commencement.

Conflicts of Interest: The authors declare no conflicts of interest relevant to this work. The CAREPOI organisation provided partial support for this research (use/validation through the CAREPOI telemetry system) but had no role in the study design, methodology, analysis, or interpretation of results.

Funding Information: This research is supported by internal institutional funding from the University of Patras and CAREPOI organisation. No external funding was received for this system design study.

Data Availability: As this is a system design study, no clinical datasets were generated or analysed. Future validation studies will make anonymised data available through appropriate repositories in accordance with institutional and regulatory requirements.

Artificial Intelligence Disclosure: AI-based tools were not used for manuscript preparation. The AI models described in this paper are components of the system under study.

References

1. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *Circulation*. 2018;138(20):e618–e651. doi:10.1161/CIR.0000000000000617
2. Landesberg G, Beattie WS, Mosseri M, Jaffe AS, Alpert JS. Perioperative myocardial infarction. *Circulation*. 2009;119(22):2936–2944. doi:10.1161/CIRCULATIONAHA.108.828228
3. Devereaux PJ, Chan MT, Alonso-Coello P, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*. 2012;307(21):2295–2304. doi:10.1001/jama.2012.4502
4. Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg*. 2005;242(3):326–343. doi:10.1097/01.sla.0000179092.38798.ef
5. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Adv Neural Inf Process Syst*. 2020;33:9459–9474. arXiv:2005.11401
6. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259–265. doi:10.1038/s41586-023-05881-4
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
8. Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med*. 2014;190(6):649–655. doi:10.1164/rccm.201406-1022OC
9. CAREPOI. Care at any Point of Interest – perioperative telemetry monitoring system. Available at: <https://carepoi.com> (accessed 15 February 2025).
10. Dencker EE, Bonde A, Troelsen A, Sillesen M. Assessing the utility of natural language processing for detecting postoperative complications from free medical text. *BJS Open*. 2024;8(2):zrae020. doi:10.1093/bjsopen/zrae020
11. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform*. 2018;22(5):1589–1604. doi:10.1109/JBHI.2017.2767063
12. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics; 2019:3982–3992. doi:10.18653/v1/D19-1410
13. Zikos D. A system to design successful clinical decision support systems. In: *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*. New York: ACM; 2017:185–188. doi:10.1145/3056540.3056572
14. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–1358. doi:10.1056/NEJMr1814259
15. Bhavani SV, Carey KA, Gilbert ER, et al. Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. *Sci Rep*. 2023;13:1173. doi:10.1038/s41598-023-27418-5
16. Pylarinou C, Gortzis LG, Koletsis E, Leivaditis V, Mavrilas D. DETER: a clinical deterioration prediction algorithm to improve patient care with devices-based telemetry and generative AI. *J Comput Intell Biomed*. 2026 [Accepted for publication].
17. Pidoux J, Conus E, Blackman N, et al. Comparison of postoperative continuous wireless cardiac rhythm monitoring with traditional telemetry in cardiac surgery patients: the SMART-TEL study. *J Innov Card Rhythm Manag*. 2024;15(8):5997–6003. doi:10.19102/icrm.2024.15085
18. Bergamaschi L, Pizzi C, Gallitto E, et al. Wearable devices as part of postoperative early warning score systems: a scoping review. *J Clin Monit Comput*. 2025;39(1):233–244. doi:10.1007/s10877-024-01224-4

19. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234–1240. doi:10.1093/bioinformatics/btz682
20. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010;17(5):507–513. doi:10.1136/jamia.2009.001560

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.