

ORIGINAL RESEARCH

Open access

A Graph Attention Network Framework for Surgical Site Infection Prediction Integrating Intraoperative Variables, Surgeon Experience, and Comorbidity Graphs

Elena Petrova^{1*}, Ivan Georgiev¹, Nikolay Stoyanov², Petar Kolev¹

Abstract

Surgical site infections (SSIs) affect 2–20% of surgical procedures and are a major source of postoperative morbidity, prolonged hospitalization, readmission, mortality, and healthcare costs, making prevention a key priority. Existing prediction tools such as the NNIS index and SENIC score depend on a limited set of clinical variables including wound class, ASA status, and operative duration, while failing to capture complex interactions among patients, surgeons, and comorbidities. To address this limitation, we propose a graph attention network (GAT) framework that represents each surgical case as a heterogeneous graph composed of patient, surgeon, and comorbidity nodes, with intraoperative variables included as features and attention mechanisms used to learn the most influential relationships. This approach models relational dependencies such as the interaction between surgeon experience, patient conditions, and comorbidity combinations, enabling more accurate and context-aware SSI risk prediction to support personalized preventive interventions.

Keywords Graph attention networks, Surgical site infection, Surgeon experience, Comorbidity graphs, Intraoperative variables, Postoperative complications

*Correspondence:

Elena Petrova
elena.petrova@gmail.com

¹ Department of Healthcare Intelligence Systems, Medical University of Sofia, Sofia, Bulgaria

² Department of Clinical AI Engineering, Technical University of Sofia, Sofia, Bulgaria

Introduction

Surgical site infections are classified as superficial incisional, deep incisional, or organ/space infections, with incidence varying substantially by procedure type from approximately 1% in clean surgeries to over 20% in contaminated or emergency operations [1, 2]. The consequences of SSI extend beyond the immediate postoperative period to include prolonged hospital stays averaging 7-11 additional days, doubled readmission rates, increased mortality risk ranging from two to eleven times that of uninfected patients, and substantial healthcare costs estimated at \$20,000 to \$35,000 per infection episode [3,

4]. These clinical and economic burdens have motivated decades of research into risk factor identification and prediction model development, yet SSI rates have remained stubbornly persistent in many surgical populations [5, 6].

Current risk prediction tools used in clinical practice, such as the NNIS index and various institutional scoring systems, typically incorporate only three to five variables including wound contamination class, ASA score, procedure duration, and occasionally body mass index or diabetes status [4, 7]. These models, while simple to calculate at the bedside, fundamentally ignore two critical

domains: the substantial evidence that surgeon experience metrics including annual case volume and years of practice independently predict surgical outcomes, and the network of interacting comorbidities that modify infection risk in nonlinear ways [8-10]. Machine learning approaches applied to SSI prediction have shown improved performance over traditional logistic regression, but most remain tabular models that treat each patient as an independent observation rather than modeling relational structure [11-14].

Table 1 contrasts the proposed graph attention framework with both traditional SSI scoring systems and non-relational machine learning models to clarify the distinct analytical advantage of relational surgical risk modeling.

Table 1. Conceptual Comparison of Prediction Paradigms for Surgical Site Infection Risk Stratification

Dimension	Traditional SSI Scores (e.g., NNIS/SENIC-type logic)	Conventional Tabular Machine Learning	Proposed Heterogeneous GAT Framework
Fundamental analytical unit	Isolated patient with a few manually selected predictors	Isolated patient represented as a flattened feature vector	Relational surgical episode composed of patient, surgeon, comorbidity nodes
Representation of surgeon influence	Usually omitted or crudely proxied	Added as a fixed scalar feature if available	Explicit surgeon node with dynamic experience/outcome attributes
Representation of comorbidity burden	Simple counts or binary indicators	Multiple binary covariates treated independently unless handcrafted interactions are added	Separate comorbidity nodes to preserve diagnostic and relational contributions

Modeling of interactions	Limited to predefined additive logic	Possible but dependent on implicit nonlinear learning in tabular space	Directly learned through attention weights, message passing across patient, surgeon, comorbidity relations
Capacity to capture nonlinear clinical synergy	Low	Moderate	High, especially for complex dependent combinations of comorbidity burden, operative exposure, surgeon experience
Handling of intraoperative variables	Often restricted to duration or wound class	Included as standard predictors	Integrated patient-level features interpreted in relation to neighbors
Interpretability structure	Rule-based but clinically shallow	Variable importance often global and feature-centric	Attention-derived, relational, and edge-sensitive explanations
Clinical meaning of prediction	Score or probability with limited causal narrative	Probability with partial feature attribution	Probabilistic structural explanations of whether risk is driven by surgeon, comorbidity constellation, operative exposure, or their interactions
Adaptability over time	Low; static score design	Moderate; retraining required	Highly adaptive surgeon- and graph-based

			represent can evol accumul case his
Suitability for perioperative intervention design	Broad risk flagging only	Better ranking but limited relational insight	Stronger for target interven because of risk decompo across cli actiona domai
Main conceptual limitation	Under-specifies clinical complexity	Treats observations as independent cases	Requires data infrastru and gra construc discipli

Graph neural networks, and specifically graph attention networks (GATs), offer a paradigm shift by explicitly modeling entities as nodes and their relationships as edges, then learning through message passing how information propagates across the graph structure [15, 16]. Unlike conventional graph convolutional networks that apply fixed weights to neighbors, GATs compute attention coefficients that learn which neighboring nodes are most relevant for prediction, providing inherent interpretability about which relationships drove a given risk estimate [17, 18]. This attention mechanism is particularly well-suited to surgical risk prediction because it can identify whether a patient's SSI risk is driven primarily by their own comorbidities, by their surgeon's inexperience, by intraoperative factors, or by interactions among these domains [19, 20].

This article presents a conceptual framework for SSI prediction using a heterogeneous graph attention network that integrates three node types: patient nodes carrying demographic and clinical features, surgeon nodes encoding experience metrics, and comorbidity nodes representing individual diagnoses. Intraoperative variables are incorporated as additional features on patient nodes, while edges connect each patient to their operating surgeon and to all comorbidities present. The remainder of this paper is organized as follows.

Background

Surgical site infection risk factors

Established risk factors for SSI span preoperative patient characteristics including diabetes mellitus, obesity defined as body mass index greater than 30 kg/m², active smoking, immunosuppressive medication use, and prior surgical site infection, as well as intraoperative variables including procedure duration exceeding the 75th percentile for that operation, wound contamination class (clean, clean-contaminated, contaminated, dirty), estimated blood loss, and transfusion requirement [2, 6, 21]. Postoperative factors such as wound care practices, glycemic control, and duration of surgical drain placement also contribute to risk, though these occur after the prediction window for preoperative risk stratification [22]. Meta-analyses have consistently identified diabetes with odds ratios for SSI ranging from 1.5 to 2.5, obesity with similar effect sizes, and dirty wound class with odds ratios exceeding 5.0 compared to clean procedures [2, 6].

Surgeon experience metrics

Surgeon experience is most commonly quantified through annual case volume, with systematic reviews demonstrating that higher-volume surgeons achieve lower complication rates, including SSI, across multiple surgical specialties including cardiac, urologic, hepatobiliary, and colorectal surgery [8, 9, 23, 24]. Additional experience metrics include years in independent practice following residency completion, fellowship training in a subspecialty, board certification status, and individual surgeon's historical SSI rate adjusted for case mix [10, 25, 26]. The volume-outcome relationship exhibits both a threshold effect, where very low-volume surgeons (fewer than 5-10 cases annually) have substantially higher complication rates, and a continuous gradient where increasing volume up to approximately 50-100 cases annually continues to show improved outcomes [8, 23, 24].

Comorbidity networks

Individual comorbidities do not operate independently but rather interact in complex networks where the presence of multiple conditions can amplify infection risk synergistically beyond additive effects [27, 28]. Common comorbidities relevant to SSI risk include diabetes mellitus (impaired wound healing and immune function), obesity (poor vascularity of adipose tissue and technical difficulty), chronic obstructive pulmonary disease (impaired

oxygenation and coughing impairment), chronic kidney disease (immune dysfunction and fluid management challenges), and immunosuppression from medications or disease states [2, 17, 22]. The interaction between diabetes and obesity, for example, produces higher SSI risk than either condition alone, while the combination of immunosuppression with any contaminated wound class dramatically elevates risk [6, 21].

Graph neural networks

Graph neural networks operate by propagating information between connected nodes through iterative message passing, where each node aggregates features from its neighbors to update its own embedding [15, 16]. Graph convolutional networks (GCNs) perform this aggregation using fixed, normalized weights based on node degrees, whereas graph attention networks introduce learnable attention coefficients that allow the model to weight neighbors differently depending on their features [15, 18]. The attention mechanism computes a normalized importance score for each neighbor-edge pair using a shared learnable weight matrix, then aggregates neighbor features weighted by these attention coefficients, enabling the model to focus on the most informative relationships for the prediction task [17, 20].

Framework Overview

High-level architecture

The proposed framework constructs a heterogeneous graph for each surgical episode containing three node types: a single patient node, one surgeon node representing the operating surgeon, and multiple comorbidity nodes for each diagnosis present in the patient's medical history. Edges connect the patient node to the surgeon node and to each comorbidity node, while no direct edges exist between comorbidity nodes or between surgeon nodes across different patients. This graph serves as input to a multi-head graph attention network that learns node embeddings through attention-based aggregation, followed by a readout function and multilayer perceptron classifier to output a probability of SSI within 30 postoperative days.

Figure 1 shows the proposed heterogeneous graph attention network framework, illustrating how patient features, intraoperative variables, surgeon experience, and

comorbidity structure are integrated into an interpretable relational model for surgical site infection risk prediction.

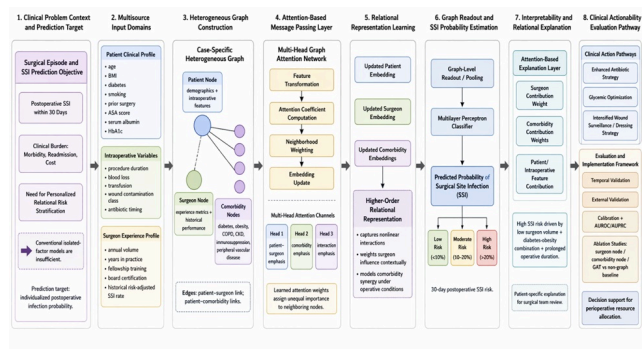


Figure 1. Heterogeneous Graph Attention Network Architecture for Relational Surgical Site Infection Prediction

Core assumptions

The framework assumes availability of structured electronic health record data containing patient demographics, preoperative diagnosis codes, intraoperative variables recorded in standard anesthesia or surgical documentation, and reliable surgeon identifiers that can be linked across cases. Additional assumptions include consistent coding of wound contamination class, complete capture of surgical duration and estimated blood loss, and availability of at least 12 months of historical data to compute surgeon volume metrics and prior SSI rates. For hospitals with low surgical volumes per surgeon, the framework may require pooling across multiple sites or using Bayesian approaches to stabilize volume estimates.

Design principles

Three design principles guide the framework: relational reasoning that explicitly models the connections between patients, surgeons, and comorbidities rather than treating them as independent features; attention-based interpretability that enables clinicians to understand which relationships drove each prediction; and multi-source integration that combines preoperative, intraoperative, and surgeon-level information without requiring manual feature engineering of interactions. The framework prioritizes clinical actionability by designing risk stratification thresholds that correspond to evidence-based interventions such as additional antibiotic dosing, enhanced glucose monitoring, or wound care protocols.

Graph Construction

Patient nodes

Each surgical case is represented as a single patient node with features extracted from preoperative and intraoperative data, including age in years, body mass index, diabetes status (binary), current smoking status, history of prior surgery at the same anatomic site, wound contamination class encoded as one-hot categorical variable, and American Society of Anesthesiologists physical status score [2, 7]. Additional patient features may include preoperative serum albumin as a marker of nutritional status, hemoglobin A1c for diabetic patients, and white blood cell count as a baseline inflammatory marker [4, 21]. All continuous features are normalized to zero mean and unit variance across the training dataset to facilitate stable gradient-based optimization.

Surgeon nodes

Each distinct surgeon in the dataset is represented as a single node with features that capture experience metrics computed from historical surgical data, including annual case volume averaged over the preceding 12 months, years in independent practice since residency completion, board certification status in the relevant specialty, fellowship training in a subspecialty (binary), and the surgeon's historical SSI rate adjusted for case mix using a rolling window of the prior 50-100 cases [8-10]. Surgeon node features are updated dynamically as new surgical cases are added to the dataset, allowing the framework to adapt to improving surgeon experience or changing practice patterns over time [24, 26].

Comorbidity nodes

Individual comorbidities are represented as separate nodes, each with a fixed feature vector containing the prevalence of that comorbidity in the training population and a learned embedding that captures the comorbidity's baseline association with SSI risk [27, 28]. Comorbidities included in the graph are selected based on established SSI risk factors and International Classification of Diseases (ICD) diagnosis codes, including diabetes mellitus (E08-E13), obesity (E66), chronic obstructive pulmonary disease (J44), chronic kidney disease (N18), immunosuppression (D84, Z79.8), and peripheral vascular disease (I73.9) [2, 17, 22]. Edges connect each patient node to every comorbidity node for which that patient has a qualifying diagnosis code in the 12 months preceding surgery.

Graph Attention Network

Attention mechanism

For each node in the heterogeneous graph, the attention mechanism computes normalized importance scores for all neighboring nodes by applying a shared learnable weight matrix to transform node features, then using a single-layer feedforward neural network to produce unnormalized attention coefficients [15, 18]. These coefficients are normalized across each node's neighborhood using the softmax function, producing attention weights that sum to one and indicate the relative importance of each neighbor for predicting SSI from the target node's perspective. The attention weights are computed separately for each node and each attention head, allowing the model to learn that for some patients, the surgeon node may be highly influential, while for others, specific comorbidity nodes may dominate the risk signal.

Multi-head attention

Multi-head attention stabilizes the learning process and enables the framework to capture different types of relationships simultaneously by applying K independent attention mechanisms in parallel, each with its own learnable parameters, then concatenating or averaging the resulting feature representations [15, 18]. In the context of SSI prediction, one attention head might specialize in learning patient-surgeon relationships that capture volume-outcome effects, a second head might focus on patient-comorbidity interactions that identify high-risk comorbidity combinations, and a third head might learn surgeon-comorbidity indirect relationships where specific surgeons have systematically different outcomes for patients with particular comorbidities [16, 20]. The multi-head architecture provides a form of ensemble learning within the graph attention network without substantially increasing model complexity.

Node embedding update

The node embedding update proceeds by computing, for each node, a weighted sum of its neighbors' transformed features using the attention weights from each head, applying a nonlinear activation function (typically exponential linear unit or rectified linear unit), and optionally adding a self-connection to preserve the node's original features [1, 3]. Multiple graph attention layers are stacked to enable higher-order relationship learning, where the first layer aggregates information from immediate neighbors

(directly connected comorbidities and the surgeon), and subsequent layers propagate information from neighbors-of-neighbors (e.g., other patients of the same surgeon or patients sharing comorbidities) [18, 19]. This multi-layer architecture allows the framework to learn that a patient's SSI risk may be influenced not only by their own surgeon's volume but also by the complication rates of that surgeon across other patients with similar comorbidity profiles.

Node Feature Integration

Intraoperative variables

Intraoperative variables are integrated as additional features on the patient node, including total procedure duration in minutes from incision to closure, estimated blood loss in milliliters, binary indicator for intraoperative transfusion of packed red blood cells, wound contamination class (clean, clean-contaminated, contaminated, dirty), and timing of prophylactic antibiotic administration relative to incision [3, 5, 7]. These variables are recorded in real time during surgery and represent modifiable factors that can be targeted for quality improvement, such as reducing procedure duration through efficient surgical techniques or ensuring antibiotic redosing for prolonged operations [2, 4]. The framework processes these intraoperative features alongside static preoperative patient characteristics, enabling the attention mechanism to learn interactions such as how prolonged operative time may increase SSI risk more substantially for patients with diabetes or obesity than for otherwise healthy patients [6, 21].

Surgeon Experience Metrics

Surgeon experience metrics are encoded as features on the surgeon node, including annual case volume categorized as low (fewer than 20 cases per year), medium (20-50 cases), or high (more than 50 cases), years in independent practice post-residency, completion of fellowship training in the relevant surgical subspecialty, and the surgeon's risk-adjusted SSI rate computed from historical cases using a rolling window of the preceding 50 procedures or 12 months, whichever is larger [8-10]. These metrics are updated dynamically as new surgical cases accrue, allowing the framework to reflect improving surgeon performance over time and to capture volume thresholds beyond which additional experience yields diminishing returns [23, 24, 26]. The attention mechanism can learn that low surgeon volume should be weighted more heavily for complex procedures or for patients with

multiple comorbidities, while for low-risk patients undergoing routine procedures, surgeon experience may contribute minimally to the predicted SSI probability [25].

SSI Prediction

Readout and classification

After applying multiple graph attention network layers, the framework produces final node embeddings for the patient node, surgeon node, and all comorbidity nodes, which are then aggregated through a readout function that pools information across the entire graph to produce a fixed-dimensional graph-level representation [15, 16]. The readout operation can take multiple forms including mean pooling (averaging all node embeddings), sum pooling (adding all node embeddings), or attention-based pooling that learns which nodes are most informative for the prediction task [17, 18]. This graph-level representation is passed through a multilayer perceptron with one or two hidden layers and a sigmoid output activation to produce a predicted probability of surgical site infection occurring within 30 days postoperatively, with the model trained using binary cross-entropy loss and class weights to address the inherent class imbalance where SSI is the minority outcome [12, 14].

Risk stratification

The continuous predicted SSI probability is mapped to discrete risk categories using thresholds derived from the training distribution and calibrated to clinical actionability, with typical stratification defining low risk as predicted probability below 2-3%, moderate risk as 3-10%, and high risk as above 10% [4, 5, 7]. For patients classified as high risk, the framework recommends specific perioperative interventions including administration of additional preoperative antibiotic doses, enhanced postoperative wound monitoring with daily inspection, extended duration of prophylactic antibiotics for 24 rather than 12 hours, and stricter glycemic control protocols targeting blood glucose below 180 mg/dL [2, 3]. Moderate-risk patients may benefit from targeted interventions such as chlorhexidine gluconate washes or negative pressure wound therapy dressings, while low-risk patients receive standard perioperative care without additional resource allocation [13, 22].

Interpretability

Attention visualization

The attention coefficients learned by the graph attention network provide direct interpretability by revealing which neighboring nodes most influenced the prediction for any given patient, enabling visualization of the relational risk factors that drove the SSI probability estimate [15, 18]. For an individual prediction, the framework can output a heatmap showing the attention weight assigned to the surgeon node (indicating how strongly the surgeon's experience influenced risk), to each comorbidity node (identifying which specific diagnoses contributed most), and to the patient's own features (capturing baseline risk from demographics and intraoperative variables) [17, 19]. These attention weights can be aggregated across a cohort of patients to discover population-level patterns, such as whether certain surgeons systematically receive high attention weights from patients with specific comorbidity profiles, suggesting systematic quality gaps that could be addressed through targeted training or protocol changes [8, 20].

Table 2 consolidates the relational mechanisms that the graph attention model is intended to capture and links each mechanism to its corresponding clinical interpretation and intervention pathway.

Table 2. Relational Mechanisms, Attention Targets, and Clinical Consequences in the Proposed SSI Graph Framework

Relational mechanism within the framework	Primary node/edge structure involved	What the attention mechanism is expected to learn	Clinical interpretation high level weight
Surgeon volume amplification of baseline patient risk	Patient–surgeon edge	Whether low case volume materially increases infection vulnerability for a given patient profile	The operation surgeon experience meaningfully modifies patient's beyond baseline comorbidity
Comorbidity-specific risk concentration	Patient–comorbidity edges	Which individual diagnoses are most influential	A specific diagnosis such as diabetes immunosuppression

		for the current prediction	is a dominant infection
Comorbidity synergy rather than additive burden	Multiple patient–comorbidity edges viewed jointly	Whether combinations such as diabetes plus obesity carry disproportionate importance	Risk arises from interactions structure, not merely the sum of diagnoses
Operative exposure interacting with physiologic vulnerability	Patient node features plus comorbidity-linked context	Whether prolonged duration, blood loss, or transfusion become more important in particular comorbidity settings	Intraoperative stressors are more dangerous in susceptible biologic contexts
Surgeon effect conditioned by case complexity	Patient–surgeon edge plus patient-node feature set	Whether surgeon experience is most influential in technically difficult or contaminated procedures	Experience matters more selectively than uniformly across all cases
Historical surgeon performance as a contextual prior	Surgeon node attributes	Whether prior risk-adjusted SSI outcomes remain informative after patient adjustment	Persistent surgeon-level quality contributes to current assessment
Attention redistribution across domains	All immediate patient neighbors	Which domain dominates the final prediction for an individual case	The model distinguishes between surgery-driven, surgery-driven multimorbidity-driven risk
Population-level aggregation of attention patterns	Cohort-level summaries of node and edge weights	Whether recurrent institutional patterns	Certain surgical procedure comorbidity constellations

		emerge across many surgeries	systema concentr
--	--	------------------------------	------------------

Clinical utility

The clinical utility of attention-based interpretability lies in generating patient-specific explanations that can be communicated to the surgical team at the point of care, for example: "This patient is at high risk for SSI (predicted probability 14%) primarily due to low surgeon volume for this procedure (attention weight 0.62) and the combination of diabetes with obesity (combined comorbidity attention 0.28)" [5, 14, 16]. These explanations enable clinicians to understand not just the risk magnitude but also its drivers, facilitating targeted interventions such as assigning a higher-volume surgeon to the case, optimizing glycemic control preoperatively, or implementing enhanced infection prevention protocols for this specific patient-surgeon combination [7, 9]. The framework can also produce counterfactual explanations by simulating how the predicted probability would change if a different surgeon performed the procedure or if a modifiable comorbidity were better controlled, supporting shared decision-making between surgeons and patients [4, 10].

Evaluation Strategy

Prediction metrics

The framework should be evaluated using metrics appropriate for binary classification with class imbalance, including area under the receiver operating characteristic curve (AUROC) for overall discriminative ability, area under the precision-recall curve (AUPRC) which is more informative when SSI prevalence is low (typically 2-10%), calibration assessed via Brier score and calibration plots, and sensitivity and specificity at clinically relevant risk thresholds [5, 12, 13]. For surgical site infection prediction, the minimum acceptable performance threshold should be an AUROC of at least 0.75 for external validation, given that traditional models such as NNIS achieve AUROC values in the 0.60-0.70 range and machine learning models have reported values from 0.70 to 0.85 depending on procedure type and dataset size [4, 7, 14]. Calibration is equally important as discrimination, because poorly calibrated probabilities may lead to inappropriate resource allocation if predicted risks do not match observed frequencies across risk strata [5, 29].

Validation protocols

Validation of the framework requires temporal splitting where the model is trained on surgeries performed during an earlier time period (e.g., calendar years 1-3) and tested on surgeries from a subsequent period (year 4), ensuring that the evaluation simulates prospective deployment where future cases are unseen during training [4, 5, 14]. External validation on data from a different hospital or healthcare system is essential to assess generalizability, as surgeon volume distributions, comorbidity coding practices, and SSI surveillance protocols vary substantially across institutions [12, 29]. For temporal validation, the framework should demonstrate stability of performance across time, with particular attention to whether model performance degrades as surgeon experience improves or as perioperative protocols change, necessitating periodic model retraining or online learning approaches [7, 13].

Ablation studies

Ablation studies are necessary to quantify the contribution of each framework component, including removal of surgeon nodes (forcing the model to predict SSI using only patient and comorbidity information), removal of comorbidity nodes (using only patient and surgeon features), replacement of the graph attention network with a graph convolutional network to assess the value of attention, and replacement of the graph-based architecture with a logistic regression or random forest model using the same flattened features [15, 16, 18]. Expected results from ablation studies include demonstration that adding surgeon nodes improves AUROC by at least 0.03-0.05 compared to patient-only models, that attention provides interpretability without substantial performance degradation relative to GCNs, and that the full graph model outperforms logistic regression by a margin of 0.05-0.10 in AUROC across validation datasets [7, 19, 20]. These analyses identify which data elements are most critical for prediction and guide minimum data collection requirements for hospitals considering implementation [4, 17].

Conclusion

This conceptual framework proposes a graph attention network that integrates intraoperative variables, surgeon experience metrics, and patient comorbidity graphs to predict postoperative surgical site infection risk, moving beyond traditional tabular models by explicitly modeling the relational structure of the surgical episode. By representing

patients, surgeons, and comorbidities as nodes in a heterogeneous graph with edges capturing clinical relationships, the framework enables attention-based learning of how these entities interact to influence infection risk, including nonlinear and synergistic effects that conventional models cannot capture.

The key advantages of this framework include its relational reasoning capacity that mirrors clinical cognition, where experienced surgeons implicitly consider how patient comorbidities interact with their own technical approach; its attention-based interpretability that generates patient-specific explanations linking predicted risk to modifiable factors; and its multi-source integration that combines preoperative, intraoperative, and surgeon-level data without requiring manual specification of higher-order interaction terms. These advantages position the framework as a potential decision support tool for preoperative risk stratification and perioperative resource allocation.

Several limitations must be addressed before clinical implementation. The framework requires linked surgeon-patient data that may not be readily available in all electronic health record systems, particularly in settings where surgical schedules do not reliably document the primary operating surgeon. Hospitals with low surgical volumes may have insufficient data to estimate stable surgeon volume metrics or attention coefficients, requiring transfer learning from larger centers or Bayesian hierarchical models that pool information across surgeons. Variability in comorbidity coding practices across institutions and over time may affect generalizability, necessitating standardized coding guidelines or natural language processing approaches to extract comorbidities from clinical notes.

We call for implementation of this framework on existing surgical registries including the National Surgical Quality Improvement Program (NSQIP), the Society of Thoracic Surgeons (STS) database, and local institutional data sources to enable large-scale validation of the relational approach to SSI prediction. Prospective deployment studies should assess not only predictive accuracy but also clinical utility, including whether attention-based explanations change surgeon behavior, whether risk-stratified interventions reduce observed SSI rates, and whether the framework achieves acceptable usability and trust among surgical teams. If successful, the relational graph attention paradigm could be extended beyond SSI to predict other postoperative complications such as venous thromboembolism, acute kidney injury, and unplanned readmission.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 01 Feb 2024 Revised: 24 May 2024 Accepted: 24 Jun 2024

Published online: 20 January 2025

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Gillespie BM, Harbeck E, Rattray M, Liang R, Walker R, Latimer S, et al. Worldwide incidence of surgical site infections

in general surgical patients: a systematic review and meta-analysis of 488,594 patients. *Int J Surg*. 2021;95:106136. <https://doi.org/10.1016/j.ijsu.2021.106136>.

Xu Z, Qu H, Gong Z, Kanani G, Zhang F, Ren Y, et al. Risk factors for surgical site infection in patients undergoing colorectal surgery: A meta-analysis of observational studies. *PLoS One*. 2021;16(10):e0259107. <https://doi.org/10.1371/journal.pone.0259107>.

Fernandez-Moure JS, Wes A, Kaplan LJ, Fischer JP. Actionable risk model for the development of surgical site infection after emergency surgery. *Surg Infect (Larchmt)*. 2021;22(2):168-73. <https://doi.org/10.1089/sur.2020.171>.

Al Mamlook RE, Wells LJ, Sawyer R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *Am J Infect Control*. 2023;51(5):544-50. <https://doi.org/10.1016/j.ajic.2022.09.020>.

van Boekel AM, van der Meijden SL, Arbous SM, Nelissen RG, Veldkamp KE, Nieswaag EB, et al. Systematic evaluation of machine learning models for postoperative surgical site infection prediction. *PLoS One*. 2024;19(12):e0312968. <https://doi.org/10.1371/journal.pone.0312968>.

Zhao D, Liang GH, Pan JK, Zeng LF, Luo MH, Huang HT, et al. Risk factors for postoperative surgical site infections after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Br J Sports Med*. 2023;57(2):118-28. <https://doi.org/10.1136/bjsports-2021-105132>.

Chen T, Liu C, Zhang Z, Liang T, Zhu J, Zhou C, et al. Using machine learning to predict surgical site infection after lumbar spine surgery. *Infect Drug Resist*. 2023;16:5197-207. <https://doi.org/10.2147/IDR.S421171>.

Akmaz B, van Kuijk SM, Nia PS. Association between individual surgeon volume and outcome in mitral valve surgery: a systematic review. *J Thorac Dis*. 2021;13(7):4500-12. <https://doi.org/10.21037/jtd-20-3310>.

Van den Broeck T, Oprea-Lager D, Moris L, Kailavasan M, Briers E, Cornford P, et al. A systematic review of the impact of surgeon and hospital caseload volume on oncological and nononcological outcomes after radical prostatectomy for nonmetastatic prostate cancer. *Eur Urol*. 2021;80(5):531-45. <https://doi.org/10.1016/j.eururo.2021.06.032>.

Bruins HM, Veskimäe E, Hernandez V, Neuzillet Y, Cathomas R, Comperat EM, et al. The importance of hospital and surgeon volume as major determinants of morbidity and

mortality after radical cystectomy for bladder cancer: a systematic review and recommendations by the European Association of Urology Muscle-invasive and Metastatic Bladder Cancer Guideline Panel. *Eur Urol Oncol*. 2020;3(2):131-44. <https://doi.org/10.1016/j.euo.2019.11.005>.

Kuo PJ, Wu SC, Chien PC, Chang SS, Rau CS, Tai HL, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. *Oncotarget*. 2018;9(17):13768-82. <https://doi.org/10.18632/oncotarget.24403>.

Chen KA, Joisa CU, Stem JM, Guillem JG, Gomez SM, Kapadia MR. Improved prediction of surgical-site infection after colorectal surgery using machine learning. *Dis Colon Rectum*. 2023;66(3):458-66. <https://doi.org/10.1097/DCR.0000000000002564>.

Lu K, Tu Y, Su S, Ding J, Hou X, Dong C, et al. Machine learning application for prediction of surgical site infection after posterior cervical surgery. *Int Wound J*. 2024;21(4):e14607. <https://doi.org/10.1111/iwj.14607>.

Hopkins BS, Mazmudar A, Driscoll C, Svet M, Goergen J, Kelsten M, et al. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. *Clin Neurol Neurosurg*. 2020;192:105718. <https://doi.org/10.1016/j.clineuro.2020.105718>.

Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv*. 2017;arXiv:1710.10903. <https://doi.org/10.48550/arXiv.1710.10903>.

Sun Z, Yin H, Chen H, Chen T, Cui L, Yang F. Disease prediction via graph neural networks. *IEEE J Biomed Health Inform*. 2021;25(3):818-26. <https://doi.org/10.1109/JBHI.2020.3004844>.

Zhang XM, Liang L, Liu L, Tang MJ. Graph neural networks and their current applications in bioinformatics. *Front Genet*. 2021;12:690049. <https://doi.org/10.3389/fgene.2021.690049>.

Johnson R, Li MM, Noori A, Queen O, Zitnik M. Graph artificial intelligence in medicine. *Annu Rev Biomed Data Sci*. 2024;7(1):345-68. <https://doi.org/10.1146/annurev-biodatasci-092123-114708>.

Tariq A, Lancaster L, Elugunti P, Siebeneck E, Noe K, Borah B, et al. Graph convolutional network-based fusion model to predict risk of hospital acquired infections. *J Am Med Inform Assoc*. 2023;30(6):1056-67.

Song K, Park H, Lee J, Kim A, Jung J. COVID-19 infection inference with graph neural networks. *Sci Rep*. 2023;13(1):11469.
<https://doi.org/10.1038/s41598-023-38323-5>.

Peng XQ, Sun CG, Fei ZG, Zhou QJ. Risk factors for surgical site infection after spinal surgery: a systematic review and meta-analysis based on twenty-seven studies. *World Neurosurg*. 2019;123:e318-e329.
<https://doi.org/10.1016/j.wneu.2018.11.158>.

Roberts DJ, Nagpal SK, Stelfox HT, Brandys T, Corrales-Medina V, Dubois L, et al. Risk factors for surgical site infection after lower limb revascularization surgery in adults with peripheral artery disease: protocol for a systematic review and meta-analysis. *JMIR Res Protoc*. 2021;10(9):e28759.
<https://doi.org/10.2196/28759>.

Chikwe J, Toyoda N, Anyanwu AC, Itagaki S, Egorova NN, Boateng P, et al. Relation of mitral valve surgery volume to repair rate, durability, and survival. *J Am Coll Cardiol*. 2017;69(19):2397-406.
<https://doi.org/10.1016/j.jacc.2017.02.026>.

Franchi E, Donadon M, Torzilli G. Effects of volume on outcome in hepatobiliary surgery: a review with guidelines

proposal. *Glob Health Med*. 2020;2(5):292-7.
<https://doi.org/10.35772/ghm.2020.01063>.

Ju IE, Trieu D, Chang SB, Mungovan SF, Patel MI. Surgeon experience and erectile function after radical prostatectomy: a systematic review. *Sex Med Rev*. 2021;9(4):650-8.
<https://doi.org/10.1016/j.sxmr.2021.04.003>.

Trieu D, Ju IE, Chang SB, Mungovan SF, Patel MI. Surgeon case volume and continence recovery following radical prostatectomy: a systematic review. *ANZ J Surg*. 2021;91(4):521-9.
<https://doi.org/10.1111/ans.16503>.

Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457-i466.

Li X, Wang Y, Wang D, Yuan W, Peng D, Mei Q. Improving rare disease classification using imperfect knowledge graph. *BMC Med Inform Decis Mak*. 2019;19(Suppl 5):238.
<https://doi.org/10.1186/s12911-019-0955-6>.

Arvind V, Kim JS, Oermann EK, Kaji D, Cho SK. Predicting surgical complications in adult patients undergoing anterior cervical discectomy and fusion using machine learning. *Neurospine*. 2018;15(4):329-37.
<https://doi.org/10.14245/ns.1836194.097>.