Scientific Journals
An international open-access platform

Hanlin Letters

Article

OPEN ACCESS

From Predictive Feasibility to Embedded Surgical Intelligence: Implementing Anastomotic Leak Risk Stratification in Routine Clinical Workflow

Youssef Khaled,[1] Mahmoud Tarek[2]

[1] Department of Computer Science, South Valley University, Qena–Safaga Road, Qena 83523, Egypt;
[2] Department of Information Systems, Damietta University, New Damietta City, Damietta 34517, Egypt;

Abstract

Risk stratification tools in surgery often demonstrate statistical promise long before they can alter care delivery. Anastomotic leak prediction is a clear example: the complication is infrequent enough to challenge model development, severe enough to justify surveillance, and temporally dynamic enough that a single static score rarely matches clinical reality. The distance between a proof-of-concept classifier and a clinically integrated decision support system is therefore not a matter of packaging alone. It is a problem of representation, calibration, workflow timing, accountability, and organizational adaptation. This paper develops a technical and translational framework for moving anastomotic leak prediction from retrospective model performance toward operational use in perioperative care. The argument is that clinical implementation requires reframing the task from isolated binary classification to sequential risk estimation under uncertainty, embedded within a bounded alerting environment and linked to explicit downstream actions. The paper examines phenotype fidelity, temporal data design, multimodal model structure, threshold governance, calibration maintenance, interface design, human oversight, subgroup reliability, and prospective evaluation strategy. It proposes a workflow-centered architecture in which risk estimation is repeatedly updated across preoperative, intraoperative, and early postoperative phases, while preserving interpretability and auditability at each stage. It also outlines monitoring and validation mechanisms required for safe deployment under changing patient mix and documentation patterns. The central claim is modest: meaningful implementation is achievable only when predictive performance, clinical utility, and sociotechnical fit are treated as a single design problem rather than as separate stages delegated to different teams.

1. Introduction

Anastomotic leak remains one of the most difficult postoperative complications to manage because it compresses several forms of uncertainty into a narrow clinical window [1]. Its onset is often preceded by nonspecific physiologic perturbation rather than a single definitive sign, yet the consequences of delayed recognition can be substantial. This combination makes the problem attractive for data-driven risk stratification. A clinician rarely observes the full multivariate trajectory of inflammatory response, hemodynamic instability, perioperative context, and baseline nutritional reserve as a formally integrated signal. A computational system can do so, at least in principle, by aggregating information that is fragmented across electronic health records, perioperative flowsheets, laboratory tables, and care-location transitions. The interest of the problem, however, is not exhausted by whether a model can discriminate between future leak and non-leak cases. The deeper question is whether that discrimination can be converted into a usable intervention pathway without increasing cognitive burden, generating counterproductive false alarms, or obscuring responsibility for action.

A recent proof-of-concept study reported that structured early postoperative electronic health record variables could support meaningful anastomotic leak stratification, with discriminative performance in the mid-0.7 AUC range, informative contributions from postoperative white blood cell count, ICU admission, lactate, surgery type, and preoperative albumin, and a clear acknowledgment that external validation, subgroup reliability, and clinical workflow implementation remained unresolved, as Sadanandan (2024) argued [2]. That statement is impor-

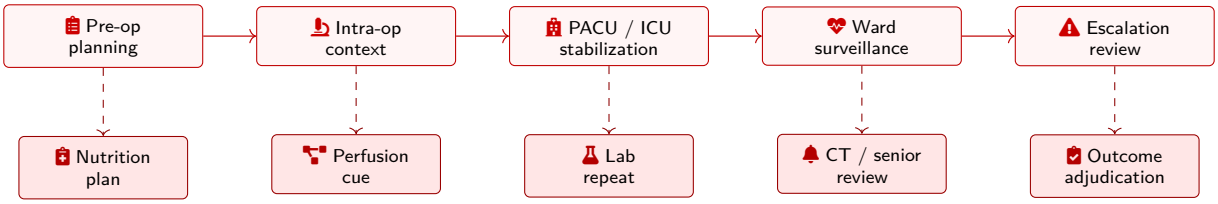

Figure 1: Workflow-centered perioperative state map for leak surveillance. The score is emitted at bounded clinical junctions, so each phase carries a distinct information set and a distinct action space rather than a single static notion of risk.

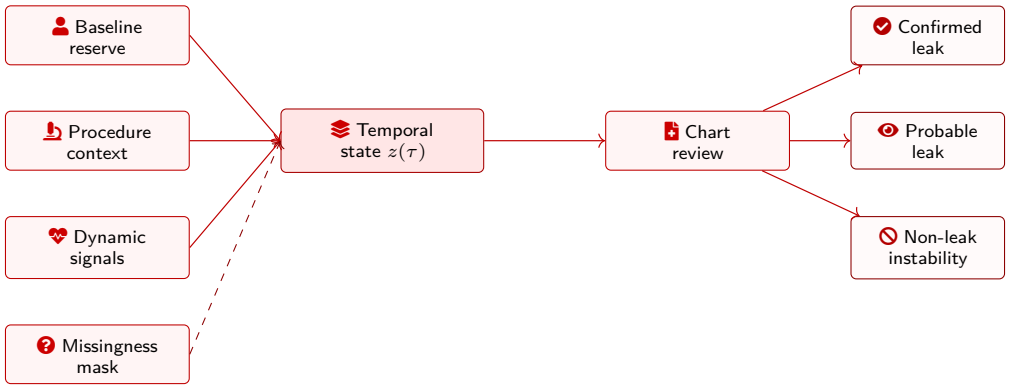

Figure 2: Representation and phenotype fidelity are handled as parallel design tasks. Stable perioperative inputs build the evolving patient state, while adjudicated endpoint refinement separates true leak from concern-driven postoperative instability that can otherwise contaminate training and calibration.

tant not because it settles the field, but because it sharply exposes the translational gap. The field no longer needs another abstract declaration that predictive signal exists in perioperative data. It needs a design grammar for turning signal into action under the constraints of routine hospital work.

That translational gap is often underestimated because proof-of-concept modeling studies and implementation studies are described as successive points on a linear pipeline. In reality, the relationship is recursive. Choices made during model construction already determine much of the eventual implementability. A model trained on variables that are available only after extensive manual abstraction, or one that depends on institution-specific coding conventions, is not merely inconvenient to deploy; it is structurally misaligned with clinical workflow. By contrast, a model deliberately engineered around stable, routinely captured, semantically interpretable variables has already taken an implementation-oriented form before any interface exists. The difference is not cosmetic. It affects threshold governance, recalibration strategy, clinician trust, and the possibility of multicenter validation.

This paper therefore treats implementation not as a downstream software engineering exercise but as a formal extension of statistical design. The prediction target is recast as a sequence of decisions distributed across time, rather than a single retrospective label attached to a hospitalization. The central unit of analysis becomes the clinical workflow state: preoperative planning, intraoperative progression, immediate postoperative stabilization, ward surveillance, escalation, and adjudication. Each state reveals different data modalities, tolerates different levels of uncertainty, and supports different interventions [3]. A useful system must therefore maintain coherence across states while allowing the meaning of the score to evolve. A probability issued before incision, another issued in the recovery unit, and a third issued on postoperative day one should not be treated as interchangeable artifacts. They arise from different information sets and imply different response options.

The objective here is not to propose a universal implementation recipe. Hospitals differ in staffing, surgical case mix, escalation pathways, radiology access, and patterns of alert fatigue. Instead, the goal is to formalize the minimal technical and organizational requirements for moving from predictive feasibility to embedded clinical use. Those requirements include faithful phenotype construction, temporally aligned data representation, inter-

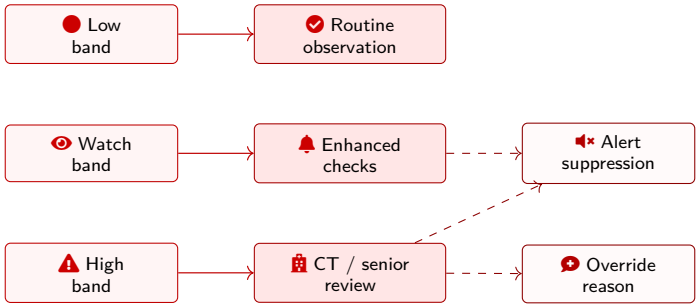

Figure 3: Threshold governance is action-linked rather than score-centric. Each risk band maps to a bounded response, while suppression and override channels prevent repetitive alarms and preserve clinician accountability inside the escalation ladder.

Phase	Data Inputs	Primary Objective	Output
Preoperative	Demographics, albumin, comorbidity	Risk counseling, optimization	Baseline risk
Intraoperative	Procedure type, duration, contamination	Contextual risk refinement	Adjusted estimate
Immediate Postop	Lactate, ICU admission, vitals	Placement, monitoring intensity	Early risk score
Ward Surveillance	Labs, trends, instability signals	Escalation decisions	Updated probability
Escalation	Imaging, senior review	Confirm/exclude complication	Action decision

Table 1: Workflow-aligned risk estimation across perioperative phases.

pretable multimodal model structure, calibrated and monitored outputs, action-linked threshold policies, subgroup accountability, and prospective evaluation under realistic use conditions. The resulting argument is intentionally neutral. It does not assume that deployment is automatically beneficial, nor that stronger discrimination alone will justify adoption. It proposes that implementation becomes rational only when the prediction system is explicitly designed as part of a bounded decision environment.

2. Translational Problem Definition

The first step in implementation is to formulate the task correctly. In many retrospective studies, the outcome is expressed as a binary label indicating whether an anastomotic leak occurred within a fixed postoperative horizon. That formulation is analytically convenient, but it obscures the operational meaning of prediction. A clinical team does not act on a label revealed after the hospitalization has ended. It acts on partial information under time pressure. The implementation problem is therefore better understood as sequential risk estimation coupled to intervention selection. At any clinical time τ, the system observes an evolving history $H(\tau)$ and emits a risk estimate that influences surveillance intensity, imaging thresholds, antimicrobial decisions, or senior review. The consequence of the estimate depends not only on its numerical value, but on when it is generated and what action space is open at that moment.

This sequential framing also clarifies why temporal alignment is not a secondary issue. A score generated too late may still be statistically accurate while clinically unhelpful. A score generated too early may be clinically attractive while underinformed [4]. Implementation requires locating the risk estimate at points where there is both sufficient signal and sufficient room for action. In anastomotic leak surveillance, those points are typically distributed across three perioperative windows. The preoperative phase supports optimization of nutrition, diversion planning, and candid discussion of risk. The intraoperative phase supports interpretation of perfusion, tension, contamination, and technical difficulty. The early postoperative phase supports intensified monitoring, low-threshold imaging, and early multidisciplinary review. Each window is not simply a later version of the same prediction problem. It is a different problem with different observability and utility.

A rigorous implementation-oriented definition therefore starts with a state-space view of the patient. Let $x_i(\tau) \in \mathbb{R}^p$ denote the observed structured covariates for patient i at workflow time τ, and let $m_i(\tau) \in \{0,1\}^p$ denote the missingness mask. The latent postoperative state is not directly observed, but it can be treated as a dynamic process influenced by baseline physiology, operative char-

Feature Type	Examples	Clinical Meaning	Risk Role
Baseline	Age, albumin, comorbidity	Physiologic reserve	Prior risk
Procedural	Surgery type, urgency	Operative stress	Context modifier
Dynamic	WBC, lactate, vitals	Recovery trajectory	Signal update
Missingness	Unordered labs	Care process signal	Implicit info
Location	ICU vs ward	Monitoring intensity	Severity proxy

Table 2: Factorized representation of perioperative data.

Threshold Level	Probability Range	Recommended Action	Context
Low	$< \theta_1$	Routine observation	All phases
Moderate	$\theta_1 - \theta_2$	Enhanced labs/monitoring	Postoperative
High	$> \theta_2$	Imaging or escalation	Ward/ICU
Critical	Rapid rise	Immediate senior review	Any phase
Resolved	Declining risk	De-escalation	Follow-up

Table 3: Risk thresholds linked to clinical action strata.

acteristics, tissue response, and emerging complications. A compact formulation is

$$\begin{aligned} s_i(\tau+1) &= As_i(\tau) + Bx_i(\tau) + Cm_i(\tau) + \varepsilon_i(\tau) \\ y_i &= \mathbb{I}(\exists \tau \le T : \phi(s_i(\tau)) > 0), \end{aligned} \tag{1}$$

where $s_i(\tau)$ is a latent clinical state, A, B, C are transition operators, $\varepsilon_i(\tau)$ is process noise, y_i is the leak indicator within horizon T, and ϕ maps latent state to failure propensity. The point of the expression is not to claim access to a true mechanistic model, but to formalize that leak risk emerges from a temporal process rather than a static table of predictors.

Phenotype fidelity enters this formulation at the level of y_i. An implementation program cannot rely indefinitely on a loosely defined retrospective label if the eventual clinical tool will be judged by real-time actionability. For anastomotic leak, the operational phenotype should ideally separate confirmed leak, probable leak under conservative management, and non-leak postoperative instability. The problem is not semantic fastidiousness. Mislabeling affects calibration and policy learning directly. If the training label conflates drain-related concerns, postoperative ileus, occult sepsis, and true anastomotic failure, then the model may learn an escalation propensity rather than a leak propensity. In silent retrospective evaluation this confusion can remain hidden; in live care it generates ambiguous alerts that are difficult to trust.

The translational definition must therefore make three commitments. The first is temporal commitment: every feature must be indexed by the time it becomes knowable in routine care. The second is semantic commitment: each variable must have a stable interpretation across institutions, or else be accompanied by a harmonization plan. The third is policy commitment: the prediction should exist only where there is a credible downstream action. Without these commitments, the move from proof-of-concept to workflow implementation is ill-posed. One does not merely deploy a classifier. One designs a timed, interpretable, capacity-constrained intervention instrument.

3. Data Representation and Phenotype Fidelity

Clinical implementation places unusual pressure on data representation because the same variable has distinct meanings depending on when and how it enters the record. White blood cell count, lactate, albumin, operative duration, and care location are not just scalar quantities; they are the visible traces of measurement processes embedded in workflow. A value is observed because someone ordered it, a patient reached a location where monitoring occurred, or concern had already risen enough to generate additional documentation. If a model ignores this fact, it can inadvertently treat healthcare process intensity as physiology. That is not always undesirable, since process variables can be informative, but implementation requires that the system distinguish between biologic signal and observation signal well enough to remain interpretable [5].

A useful representation begins by factorizing the perioperative record into baseline, procedural, and dynamic components. Baseline information includes age, comorbidity burden, nutritional markers, anthropometrics, prior operations, and other variables that are stable or slow-moving before the index procedure. Procedural information includes surgery type, urgency, approach, contamina-

Metric	Definition	Purpose	Frequency
AUC	Discrimination ability	Ranking quality	Periodic
Calibration slope	Predicted vs observed	Reliability	Continuous
ECE	Avg. probability error	Calibration detail	Rolling
Alert rate	Alerts per patient	Burden control	Daily
PPV	True alert precision	Actionability	Weekly

Table 4: Core performance and monitoring metrics in deployment.

Drift Type	Signal	Cause	Response
Covariate	Feature shift	Case-mix change	Reweighting
Missingness	Lab absence change	Ordering behavior	Feature audit
Calibration	Prob. mismatch	Prevalence shift	Recalibration
Performance	Metric decline	Model aging	Retraining
Action-induced	Outcome timing shift	Intervention effects	Re-evaluation

Table 5: Drift categories and corresponding mitigation strategies.

tion context, diversion status, and operative timing. Dynamic information includes postoperative laboratory trajectories, early hemodynamic instability, vasopressor use, level of care, fluid balance, and subsequent interventions. The implementation advantage of such a factorization is that it matches workflow states. Baseline data support preoperative counseling. Procedural data support intraoperative contextualization. Dynamic data support postoperative surveillance. A single monolithic feature vector may yield acceptable discrimination, but it obscures where in the care process the model is actually learning.

This factorization also allows explicit handling of variable availability. Let $x_i^{(b)}$, $x_i^{(p)}$, and $x_i^{(d)}(\tau)$ denote baseline, procedural, and dynamic features. One can construct an evolving representation

$$z_i(\tau) = W_b x_i^{(b)} + W_p x_i^{(p)} + W_d x_i^{(d)}(\tau) + W_m m_i(\tau) + b, \tag{2}$$

where W_b, W_p, W_d, and W_m are learned operators and $m_i(\tau)$ captures missingness. The explicit missingness term matters in implementation because absence of data is often clinically meaningful. A preoperative albumin may be missing because a patient arrived emergently. A postoperative lactate may be available only because instability prompted measurement. If the deployment environment changes laboratory ordering patterns, then the model may drift even when physiologic distributions remain stable. Treating missingness as part of the signal allows the monitoring layer to identify such dependencies rather than letting them remain implicit.

Phenotype fidelity is equally central. In retrospective proof-of-concept work, it is common to define leak using diagnosis codes combined with supporting postoperative actions. That may be reasonable in an initial study, but implementation requires a phenotype that can survive scrutiny from surgeons, informaticians, and quality officers. A clinically useful phenotype should distinguish between target condition, surrogate concern, and downstream response [6]. A patient who undergoes postoperative imaging because of rising inflammatory markers but has no leak should not be assimilated to a confirmed leak merely because the workflow resembles leak workup. Otherwise the system becomes self-referential: it predicts clinician concern because clinician concern is embedded in the label.

The most defensible path is staged phenotype refinement. At the earliest stage, coded retrospective definitions can support initial model construction. At the next stage, chart adjudication should validate a stratified sample of positive and negative cases to estimate label noise. At the implementation stage, a prospective adjudication committee should define the operational endpoint against which live performance is monitored. That endpoint should be aligned with intervention rationale. If the system is intended to trigger early imaging, then the relevant endpoint may be actionable leak or actionable occult deterioration rather than leak alone. The distinction matters because a system that increases early imaging may improve time-to-detection without reducing overall leak incidence. The deployment metric must be capable of registering such an effect.

Representation fidelity also requires attention to correlation structure. Perioperative data are highly collinear. Inflammatory markers, organ dysfunction signals, and care-location shifts co-move because they reflect common latent stress processes. A naive model may exploit these

Role	Responsibility	Interaction	Accountability
Surgeon	Final decision-making	Reviews high-risk alerts	Action taken
Nursing	Monitoring escalation	Triggers concerns	Documentation
Resident	Initial response	Executes protocols	Communication
Informatics	System maintenance	Monitors drift	Updates model
Governance	Policy oversight	Reviews outcomes	Approval

Table 6: Role-based accountability in clinical deployment.

Stage	Design	Objective	Output
Silent	Hidden scoring	Validate pipeline	Baseline metrics
Pilot	Limited visibility	Assess usability	Uptake data
Controlled	Unit rollout	Measure impact	Workflow changes
Expansion	Broader use	Generalization	System scaling
Trial	Comparative study	Causal effect	Outcome evidence

Table 7: Staged prospective evaluation pathway.

correlations opportunistically, but an implementation-oriented model should do so deliberately. One useful approach is low-rank decomposition of the centered feature matrix $X \in \mathbb{R}^{n \times p}$,

$$X = U\Sigma V^\top$$
$$X_r = U_r \Sigma_r V_r^\top, \tag{3}$$

where r is a rank chosen to preserve major latent structure while reducing noise and instability. The point is not dimensionality reduction for its own sake. It is to identify whether the predictive system is driven by a small number of coherent physiologic axes or by brittle institution-specific combinations. If the former, generalization becomes more plausible. If the latter, multicenter implementation is likely to fail unless features are redesigned.

Finally, data representation for implementation should preserve temporal granularity without overwhelming the interface [7]. Clinical teams do not need every raw measurement displayed, but the system should know whether the risk score rose because of a transient laboratory blip, persistent hemodynamic deterioration, or accumulation of moderate abnormalities. Accordingly, the internal representation can use summary operators such as recent slope, local variability, deviation from patient-specific baseline, and measurement density, while the outward-facing explanation layer presents only the most clinically interpretable contributors. The design principle is asymmetry: the model may process rich temporal structure internally, but the interface should disclose only what a clinician can realistically interpret at the point of decision.

4. Statistical Learning and Sequential Decision Architecture

A proof-of-concept model is often selected by comparing a small set of algorithms on a held-out sample and retaining the configuration with the strongest apparent discrimination. That procedure is acceptable for exploratory work, yet it is insufficient when the objective is implementation. The implemented system must satisfy at least five properties simultaneously: it must support stage-specific inference, preserve interpretability at the level of clinical concepts, tolerate missingness and workflow heterogeneity, maintain calibration under prevalence shifts, and allow threshold adaptation without retraining the full model. These requirements favor an architecture that is modular rather than monolithic.

A useful decomposition separates the system into representation, scoring, calibration, and policy layers. The representation layer converts raw perioperative variables into a temporally indexed latent state. The scoring layer maps that state to an uncalibrated risk logit. The calibration layer transforms the logit into an estimated probability. The policy layer converts probability into action recommendations conditional on workflow state, service capacity, and clinician overrides. Such a decomposition makes implementation tractable because each layer can be monitored independently. If performance degrades, one can ask whether the issue lies in missing data patterns, score instability, calibration drift, or policy mismatch. End-to-end black-box systems make that diagnostic separation much harder.

At the scoring layer, a pragmatic strategy is to combine additive clinical structure with limited interaction

Equity Dimension	Risk	Indicator	Mitigation
Data availability	Missing labs	Missingness rate	Imputation strategy
Subgroup performance	Calibration gaps	PPV/TPR diff	Recalibration
Access to care	Uneven escalation	Action rates	Policy adjust
Measurement bias	Process variation	Ordering patterns	Monitoring
Override disparity	Differential trust	Override logs	Training

Table 8: Equity considerations in predictive system deployment.

depth. Let $z_i(\tau)$ denote the latent feature state. Then a semi-structured predictor can be written as

$$\eta_i(\tau) = \beta^\top z_i(\tau) + \sum_{k=1}^{K} f_k(z_{ik}(\tau)) + \sum_{(k,\ell)\in\mathcal{I}} g_{k\ell}(z_{ik}(\tau), z_{i\ell}(\tau)), \tag{4}$$

with calibrated probability $\hat{p}_i(\tau) = \sigma(\eta_i(\tau))$, where σ is the logistic map. The linear term preserves global interpretability, the univariate functions f_k allow nonlinear effects such as threshold behavior, and the sparse interaction set $\mathcal{I}$ captures clinically plausible combinations such as inflammatory marker elevation with hemodynamic compromise or low albumin with operative complexity. This form avoids the false choice between purely linear modeling and opaque high-capacity architectures. It is expressive enough for perioperative risk while still permitting explanation at the level of named variables and interactions.

Sequential updating is also essential [8]. The system should not discard prior assessments when new data arrive. Instead, it should propagate uncertainty over time. One simple approach is recursive state updating,

$$\begin{aligned} h_i(\tau) &= \Psi(h_i(\tau-1), z_i(\tau)) \\ \eta_i(\tau) &= \Gamma(h_i(\tau), w(\tau)), \end{aligned} \tag{5}$$

where $h_i(\tau)$ is an internal memory state and $w(\tau)$ indexes workflow context such as preoperative clinic, operating room exit, ICU recovery, or ward surveillance. The functions Ψ and Γ need not be deep recurrent networks. They can be designed as transparent state aggregators that update clinically meaningful summaries: accumulated inflammatory burden, instability persistence, measurement density, and deviation from expected postoperative recovery. The key is that the score becomes path-aware. A patient whose lactate transiently rises and normalizes should not be treated equivalently to a patient with persistent upward drift, even if the most recent scalar value is the same.

Implementation also requires explicit treatment of uncertainty. In most retrospective studies, the model emits a point probability and is judged against an observed label. In live care, however, uncertainty about the prediction may be as important as the prediction itself. A narrow probability interval around a moderate risk estimate can justify action more readily than a highly unstable estimate produced from sparse observations. One may therefore maintain an uncertainty envelope around $\hat{p}_i(\tau)$, obtained through bootstrap aggregation, conformal prediction, or Bayesian posterior approximation. The interface can then present not only the central estimate but also its confidence width. This is particularly helpful in low-prevalence settings, where small logit changes can materially affect posttest probability.

The decision architecture should further encode monotonic clinical constraints where appropriate. If rising lactate, persistent leukocytosis, or increasing vasopressor requirement is known to be nondecreasing in concern within a narrow postoperative context, then the model can be constrained to avoid implausible local reversals. Such constraints do not enforce a simplistic causal doctrine. They merely prevent the model from exploiting spurious correlations that would be difficult to defend in deployment. A technically convenient formulation is to impose partial monotonicity on designated variables in the additive or tree-based components, thereby reducing variance and increasing clinician acceptability.

A model designed for workflow use should also support decomposition of contribution [9]. Let the final calibrated logit be expressed as a sum of named terms,

$$\begin{aligned} \ell_i(\tau) &= \ell_0 + \ell_i^{\text{base}} + \ell_i^{\text{proc}} + \ell_i^{\text{dyn}}(\tau) \\ \hat{p}_i(\tau) &= \sigma(\ell_i(\tau)). \end{aligned} \tag{6}$$

The terms correspond to baseline reserve, procedural context, and dynamic postoperative response. This decomposition can be shown to the clinician in natural language rather than raw coefficients. The implementation value is substantial. It communicates whether a patient is high risk because the operation was intrinsically difficult,

because baseline nutritional status was poor, or because the postoperative trajectory is deviating from expectation. The same numerical probability means different things in each of those cases and calls for different actions.

Another important architectural choice concerns ensembling. Bagging and blending can improve stability and discrimination, but they complicate interpretation and maintenance. In implementation, ensemble diversity is valuable only if the component models bring distinct and monitorable inductive biases. A heterogeneous ensemble that mixes linear calibration-aware models with interaction-sensitive tree-based models can be defensible if the deployment stack preserves component-level logging. A blind ensemble that marginally improves retrospective metrics at the expense of opaque behavior is less attractive. The relevant criterion is not whether the ensemble wins by a few hundredths of AUC; it is whether the resulting system remains governable under drift, retraining, and audit.

Finally, the scoring system must be embedded in a policy layer that translates probability into recommendation. Let θ_w denote a workflow-state-specific threshold, and let $\mathcal{R}_w$ denote the recommendation set in state w. Then

$$r_i(\tau) = \begin{cases} \text{routine}, & \hat{p}_i(\tau) < \theta_{w_1}, \\ \text{enhanced surveillance}, & \theta_{w_1} \le \hat{p}_i(\tau) < \theta_{w_2}, \\ \text{diagnostic escalation}, & \hat{p}_i(\tau) \ge \theta_{w_2}, \end{cases} \tag{7}$$

with thresholds adapted to local capacity and action cost. This form makes explicit that an implemented model is not just a score generator. It is an engine for assigning patients to surveillance strata. The statistical architecture and the intervention ladder must therefore be co-designed. Otherwise even a well-performing model becomes an isolated probability stream with no stable operational meaning.

5. Workflow-Centered System Design

A clinical workflow is not simply a sequence of timestamps. It is an allocation system for attention, authority, and response capacity. Any predictive instrument introduced into it competes with preexisting habits of triage and interpretation [10]. Implementation therefore requires an architecture that maps the statistical system onto actual care pathways rather than onto an idealized linear timeline. For anastomotic leak surveillance, that architecture must at minimum account for preoperative planning, intraoperative documentation, postoperative recovery-unit stabilization, ICU or ward disposition, daytime and overnight review patterns, imaging logistics, and escalation to senior decision-makers. A risk score that arrives outside these decision junctions will either be ignored or generate redundant work.

A practical workflow-centered design uses event-triggered inference rather than continuous background scoring. Trigger events may include case scheduling, surgical closure, postoperative admission to ICU or ward, availability of a critical laboratory panel, nursing concern escalation, or abrupt change in vasopressor exposure. Event-triggered inference reduces unnecessary computation and preserves semantic clarity: each emitted score can be interpreted relative to a known workflow transition. The preoperative score informs planning; the immediate postoperative score informs placement and monitoring intensity; the postoperative day one score informs whether deviation from expected recovery warrants escalation. By tying predictions to events, the system becomes legible to clinicians and easier to audit.

The data pipeline for such a design should be deliberately conservative. A production implementation should ingest only variables with stable source systems, unambiguous timestamps, and low dependence on free-text manual entry unless text models have been separately validated. In many hospitals, the safest initial configuration will therefore privilege structured demographics, procedure descriptors, laboratory results, medication exposure, vital sign summaries, and care-location transitions. The reason is not that richer data are unhelpful. It is that deployment failure often begins with brittle data plumbing rather than poor model logic. An implementation that depends on parsing institution-specific operative prose before the record is finalized is exposed to hidden failure modes that are hard to detect in real time.

Inference services should be isolated from the transactional electronic record while remaining tightly synchronized with it. A common pattern is to maintain a read-only feature store fed by timestamped extracts from source tables, a stateless scoring service that consumes event-based feature bundles, and a logging layer that records inputs, outputs, thresholds, recommendations, overrides, and subsequent outcomes. The separation is valuable for governance. It allows model versioning and retrospective reconstruction of what the system knew at the moment a recommendation was issued. That capability is indispensable in high-stakes review after a missed complication or an

unnecessary escalation. Without comprehensive logging, post hoc discussion quickly becomes speculative.

User interaction design must be equally explicit [11]. The score should not appear as a floating isolated number buried in a generic dashboard. It should be attached to a question that matches the clinician's current task, such as whether postoperative surveillance intensity should increase or whether imaging threshold should be lowered. The display should provide three elements only: a calibrated probability band, the dominant clinically interpretable contributors, and the recommended action stratum with an explanation of what that stratum means operationally. This restraint matters. An overloaded explanation panel can paradoxically reduce trust by forcing clinicians to parse computational details during already constrained decision windows.

Alert logic deserves particular care. It is tempting to treat any score above threshold as a notification event, but this approach rapidly generates fatigue. Implementation should instead use escalation-aware suppression rules. A patient already under enhanced surveillance should not trigger repetitive alerts unless the score crosses a materially higher boundary or new high-consequence contributors emerge. Likewise, scores should decay in salience if the relevant action has already been taken, such as cross-sectional imaging performed or senior review completed. This converts the system from a repetitive alarm source into a context-sensitive escalation assistant. The policy should be explicit, logged, and revisable because alert burden is one of the most immediate determinants of adoption.

Workflow fit also depends on accountability. Every recommendation must have an identifiable recipient and an expected response path. A score visible to everyone can effectively be owned by no one. In many settings, the best recipient will differ by workflow stage. Preoperative outputs belong with the surgeon and perioperative planning team. Immediate postoperative outputs may belong with the anesthesiology or ICU team. Ward-phase escalation outputs may belong with the primary surgical service, nursing escalation chain, and covering residents. The implementation should therefore define responsibility matrices rather than assume that visibility implies action [12]. This is not bureaucratic excess; it is the condition under which prospective evaluation can determine whether the tool altered care.

The system design should also preserve a clear override and feedback channel. Clinicians must be able to document why a recommendation was not followed, whether because of contraindication, competing diagnosis, known chronic laboratory abnormality, or direct intraoperative knowledge unavailable to the model. These override reasons are not merely medicolegal safeguards. They form a rich source of implementation data. Persistent override patterns may reveal threshold miscalibration, missing covariates, or service-specific workflow mismatch. A tool that cannot learn from structured disagreement is unlikely to mature beyond its initial deployment state.

Most importantly, workflow-centered design requires that the implementation be modest in scope at first. The initial live system should not attempt to solve every phase of leak detection simultaneously. A narrow, high-integrity use case, such as supporting early postoperative escalation decisions within a defined patient cohort, is preferable to a diffuse multitask platform. The reason is methodological. A constrained deployment yields clearer estimates of actionability, override behavior, alert burden, and subgroup performance. Once these are understood, broader perioperative integration becomes more defensible. The path from proof-of-concept to routine use is therefore not a leap from retrospective AUC to hospital-wide automation. It is a controlled narrowing of purpose followed by staged broadening under evidence.

6. Calibration, Thresholding, and Drift Surveillance

Discrimination is often the headline metric in predictive modeling, but implementation succeeds or fails much more often on calibration and threshold governance. A surgical team can work productively with an imperfectly discriminative model if the reported probabilities are stable, interpretable, and linked to rational action thresholds. By contrast, a highly discriminative but poorly calibrated model can be operationally hazardous because it distorts the implied urgency of response. In a low-prevalence complication such as anastomotic leak, small miscalibrations in the moderate-risk range can materially alter the number of patients escalated for imaging or enhanced review. Calibration is therefore not a secondary refinement; it is the bridge between model output and clinical meaning.

A deployment-ready system should treat calibration as a separate learned object [13]. Let $\tilde{p}_i(\tau)$ denote the raw score from the predictor. A simple yet effective calibrated estimate is

$$\hat{p}_i(\tau) = \sigma\left(\frac{\operatorname{logit}(\tilde{p}_i(\tau)) - \alpha_w}{T_w}\right), \tag{8}$$

where α_w and T_w are workflow-state-specific shift and temperature parameters. The state indexing matters because the same raw score can have different prevalence context and measurement structure in preoperative and postoperative phases. More flexible monotone calibration maps can be used where sample size permits, but the main principle remains: calibration must respect workflow state and be maintained independently of the core predictor. This makes recalibration feasible even when full retraining is undesirable or impossible.

Thresholding should likewise be state-specific and utility-aware. A universal threshold across all perioperative contexts is rarely coherent. In the preoperative clinic, a moderate estimated probability may justify optimization or diversion discussion because action costs are relatively low and time remains available. In the recovery unit, the same probability may justify only enhanced surveillance. On the ward at night, however, if imaging access is constrained and staff bandwidth is limited, the same estimated risk may need a different response ladder. Accordingly, thresholds should be selected by maximizing local expected utility rather than by a generic metric such as F1 alone. If π_1 denotes prevalence, u_{TP}, u_{FP}, u_{FN}, and u_{TN} denote utilities, and t is the threshold, then the deployment objective can be expressed as

$$J(t) = \pi_1 \,\mathrm{TPR}(t)\, u_{\mathrm{TP}} + (1-\pi_1)\,\mathrm{TNR}(t)\, u_{\mathrm{TN}} - \pi_1 \,\mathrm{FNR}(t)\, u_{\mathrm{FN}} - (1-\pi_1)\,\mathrm{FPR}(t)\, u_{\mathrm{FP}}. \quad (9)$$

The point is not to pretend that exact utilities are easy to specify. It is to require that threshold choice be an explicit policy decision shaped by local priorities rather than an arbitrary inheritance from retrospective validation.

Because prevalence and documentation processes evolve, calibration cannot be assumed stable after deployment. Drift surveillance must therefore monitor both covariate shift and performance shift. Covariate drift can be detected through changes in feature means, covariance structure, or missingness profiles. One convenient multivariate diagnostic is

$$D_t^2 = (\mu_t - \mu_0)^\top \Sigma_0^{-1} (\mu_t - \mu_0), \quad (10)$$

where μ_0 and Σ_0 are baseline mean and covariance, and μ_t is the current deployment-window mean. This diagnostic is not sufficient by itself, since clinically harmless changes can also raise the statistic, but it provides an early warning that the inference population may no longer match the development population. Missingness drift should be tracked separately because changes in laboratory ordering behavior can affect score behavior before physiologic distributions visibly move [14].

Performance drift should be monitored using delayed outcome linkage and rolling calibration assessment. A deployment dashboard should estimate calibration slope, intercept, expected calibration error, subgroup-specific positive predictive value, and threshold-trigger frequency over moving windows. In low-incidence settings, confidence intervals will be wide, so the monitoring plan should prespecify minimum event counts and Bayesian or shrinkage-based stabilization where needed. Silent accumulation of miscalibration is especially dangerous because a tool may appear operationally quiet while progressively losing meaning. If recalibration is needed, the logging layer must support versioned recalibration without obscuring which calibration map was in force for each patient encounter.

Subgroup surveillance is part of calibration governance rather than a separate ethics appendix. A model can be well calibrated on average while systematically underestimating or overestimating risk in specific groups defined by age, sex, urgency status, procedure type, care location, or measurement density. Implementation should therefore monitor group-conditioned calibration and threshold behavior. A simple disparity functional is

$$\Delta_g = |\mathrm{TPR}_g - \mathrm{TPR}_{\mathrm{ref}}| + \lambda |\mathrm{PPV}_g - \mathrm{PPV}_{\mathrm{ref}}|, \quad (11)$$

where λ weights the relative importance of sensitivity and alert precision disparities. The group definitions should be clinically and operationally meaningful, not merely demographic by convention. In anastomotic leak surveillance, procedure subtype and care pathway may matter as much as race or sex for practical subgroup reliability.

Another underappreciated issue is action-induced distribution shift. Once the model is live, its own recommendations can alter downstream data generation. If high-risk patients are imaged earlier or monitored more intensely, then the apparent incidence and timing of detected leaks may change. That is not a flaw; it is a sign that the system is interacting with care. But it complicates post-deployment evaluation. Drift surveillance must therefore distinguish between harmful performance deterioration and beneficial process-mediated outcome shift. A drop in late leak detections accompanied by an increase in early imaging and stable calibration among action-eligible pa-

tients may indicate success rather than failure. Monitoring plans that ignore this feedback loop are likely to misread the implementation trajectory.

For these reasons, calibration and monitoring should be institutionalized as ongoing operations rather than one-time validation tasks [15]. A model that enters workflow without a specified recalibration authority, review cadence, and retirement trigger is not implemented in any serious sense. It is merely installed. Clinical implementation begins only when there is a standing mechanism for checking whether the probability still means what the interface claims it means.

7. Human Factors, Governance, and Equity

No perioperative prediction system becomes clinically real by producing a score alone. It becomes real when clinicians incorporate that score into their reasoning without losing clarity about responsibility. In surgery, this requirement is particularly demanding. Decision authority is distributed across attending surgeons, fellows, residents, anesthesiologists, intensivists, nurses, and radiology pathways, each operating under time pressure and partial information. A prediction tool that ignores this ecology of expertise is likely to be perceived either as a nuisance or as an encroachment. The implementation challenge is therefore sociotechnical. The system must fit not only the data pipeline, but also the norms by which clinicians recognize concern, escalate uncertainty, and justify deviation from routine care.

Trust in this setting is rarely a function of model complexity alone. Clinicians do not necessarily reject nonlinear models because they are mathematically intricate; they reject outputs whose practical meaning is unclear. A system that can explain risk in terms of recognizable contributors, trajectory deviation, and recommended response level has a better chance of being accepted than a nominally simpler model presented as an unexplained percent. Explanations should therefore be framed around counterfactual interpretability that is meaningful to care teams. Instead of saying that a latent variable increased the risk score, the interface should say that the present estimate is driven primarily by persistent leukocytosis, higher-than-expected lactate trajectory, ICU-level postoperative instability, or poor preoperative reserve. The explanation need not be exhaustive. It needs to be compatible with clinical language and auditable after the fact.

Governance structures should be established before live activation rather than after the first adverse event. A deployment committee should include surgical, nursing, informatics, quality, ethics, and operational stakeholders, but its authority must be concrete. It should approve the target cohort, action ladder, alert recipients, override taxonomy, recalibration plan, subgroup monitoring strategy, and retirement criteria. Without such governance, disputes about whether an alert should have been followed will be adjudicated informally and inconsistently [16]. With governance, the organization can distinguish between model error, interface error, workflow mismatch, and reasonable clinician override. That distinction matters if the tool is to remain useful after inevitable edge cases.

Equity concerns enter at multiple levels. The most obvious is differential performance across patient groups. Yet implementation also raises equity questions through measurement processes and action pathways. If one subgroup is systematically less likely to have preoperative nutritional data, another more likely to be admitted emergently, and another more likely to receive ICU-level postoperative monitoring, then the model may assign uncertainty and escalation in uneven ways even when global performance appears acceptable. Equity analysis must therefore examine not only outcome discrimination and calibration, but also data availability, threshold crossing rates, override patterns, and downstream intervention access. A prediction tool is not equitable merely because it excludes protected attributes. It must be equitable in the sequence of consequences it generates.

A further governance concern is automation bias. In a low-frequency but high-consequence complication, there is a temptation either to defer excessively to the score or to reject it categorically after a few false positives. Both responses are operationally harmful. Implementation should instead support bounded reliance. The interface can do this by presenting the system as a surveillance adjunct rather than an adjudicator, by requiring acknowledgment of high-risk alerts without forcing rigid action, and by preserving structured override reasons. Over time, override data can reveal whether clinicians are appropriately integrating the tool or merely clicking through it. This is one of the strongest arguments for logging disagreement as carefully as agreement.

Training also matters, though it should be targeted rather than theatrical. Clinicians do not need lectures on machine learning theory. They need a precise understanding of when the score updates, what variables it uses, what the action strata mean, how to document overrides, and what not to infer from the probability. Training should explicitly address failure modes [17]. For example, a low estimated probability does not exclude other postoperative

complications, and a high estimated probability does not prove leak in the absence of corroborating evaluation. Stating these limitations at the outset protects the system from being used as a pseudo-diagnostic oracle.

Organizational legitimacy is strengthened when the implementation is transparent about intended benefit and bounded scope. A narrowly targeted postoperative surveillance tool is easier to govern than a vague promise of surgical intelligence. Claims should remain neutral. The tool may improve timeliness of review, consistency of escalation, or identification of patients who warrant closer attention. It may not reduce leak incidence directly, and it may not generalize uniformly across all colorectal pathways. This modest framing is not a rhetorical compromise. It is part of trustworthy implementation. Clinical teams are more likely to use a system whose stated purpose matches what it actually delivers.

Finally, equity and governance converge in the question of accountability after error. When a patient flagged as low risk later deteriorates, or when an unnecessary escalation occurs, review should ask structured questions: Was the phenotype appropriate? Were the necessary data available at the time? Was the score calibrated in that subgroup? Did the interface present the recommendation clearly? Was the override reasonable? Did organizational constraints impede response? Treating every miss as model failure and every false alarm as threshold failure will produce shallow learning. Implementation matures only when adverse cases are understood as interactions between algorithm, data, workflow, and institution.

8. Prospective Evaluation and Organizational Adoption

A system is not clinically implemented when it has been connected to the electronic record and made visible on a dashboard. It is implemented when its use changes care in a measurable and governable way. That distinction places prospective evaluation at the center of translation. The evaluation strategy should be staged, beginning with silent deployment, progressing to limited-scope visible use, and only then considering broader activation. Silent deployment is not optional window dressing. It establishes the real-time behavior of the data pipeline, score latency, event-trigger fidelity, threshold crossing rates, and subgroup exposure patterns before clinician behavior is altered. Many apparent model failures are uncovered at this stage as timestamp inconsistencies, feature leakage, or unstable variable mappings that retrospective analysis concealed.

The next stage is controlled visible deployment [18]. A reasonable design is a service- or unit-level stepped introduction, preferably with prespecified target cohorts and stable action pathways. The goal is not merely to see whether clinicians like the tool. It is to estimate how often recommendations are acknowledged, how often they lead to action, what kinds of override occur, and whether alert burden remains within acceptable bounds. Implementation outcomes should include uptake, fidelity, time-to-review, time-to-imaging when indicated, overnight escalation patterns, and clinician-reported comprehensibility. These are not softer substitutes for patient outcomes. They are the intermediate variables through which patient outcomes become interpretable.

For causal evaluation of workflow impact, quasi-experimental designs are often more practical than immediate patient-level randomization. If introduction occurs at the unit or team level over time, one may estimate changes using a panel model such as

$$Y_{it} = \alpha_i + \gamma_t + \theta D_{it} + \varepsilon_{it}, \tag{12}$$

where Y_{it} is an implementation or patient outcome for unit or patient i at time t, α_i captures unit or patient fixed effects, γ_t captures temporal shocks, and D_{it} indicates active tool exposure. Depending on the outcome, the model can be embedded in generalized linear forms or survival formulations. The crucial issue is not the exact estimator, but the recognition that live deployment changes behavior, so evaluation must address secular trends, service heterogeneity, and feedback effects.

Patient-centered outcomes require special caution. Leak incidence itself may be too blunt or delayed to serve as the sole early success criterion. A system may improve timeliness of recognition, reduce time-to-imaging in appropriate patients, or improve consistency of postoperative review without immediately changing overall leak rate. Conversely, a drop in observed leak rate could reflect altered coding or intensified early diversion rather than truly improved healing. The evaluation framework should therefore include layered outcomes: process outcomes, surveillance outcomes, diagnostic timing, rescue-related outcomes, and only then final clinical endpoints such as severe sepsis, unplanned reoperation, or mortality. This layered structure is especially important when the prediction tool primarily influences detection pathways rather than the biology of anastomotic healing itself.

Organizational adoption depends not only on evidence but also on the economics of attention. Even a

statistically reasonable tool may fail if it consumes senior review time without a proportional increase in actionable findings. Resource modeling should therefore accompany evaluation. One can estimate expected daily alert volume, number needed to review, radiology demand generated per threshold stratum, and incremental nursing or surgical workload [19]. These estimates should be reported alongside discrimination and calibration. A hospital deciding whether to sustain the system cares not only about whether the model is correct, but about whether the operational burden is acceptable relative to the expected benefit. A technically elegant model that saturates overnight escalation channels will not survive institutional review.

Retrospective proof-of-concept studies also tend to underplay the importance of maintenance. Organizational adoption requires a versioning plan that distinguishes between minor recalibration, moderate threshold adjustment, and major model revision. Each level should have defined approval pathways. A model whose thresholds change silently every few weeks will erode clinician trust. On the other hand, a frozen model that remains untouched as workflow and case mix evolve will gradually lose relevance. Maintenance governance is therefore part of adoption, not post-adoption housekeeping. It determines whether clinicians experience the tool as stable enough to learn and adaptive enough to remain useful.

Longer-term adoption may justify a randomized or stepped-wedge trial once silent and limited visible deployment have shown acceptable fidelity and burden. At that stage, the relevant comparison is not between model and no model in the abstract, but between model-supported workflow and usual care under defined response protocols. Such a trial should stratify by procedure complexity and workflow setting, include subgroup calibration monitoring throughout, and preserve structured capture of override reasons. The trial does not need to prove that every patient outcome improves immediately. It needs to estimate whether model-supported escalation produces a net favorable shift in clinically meaningful pathways without creating unacceptable burden or inequity.

In this sense, organizational adoption is not the endpoint of implementation. It is the condition under which implementation becomes testable. A proof-of-concept result demonstrates plausibility. A prospective deployment demonstrates operability. A properly governed trial demonstrates whether operability translates into worthwhile change. Only after all three are present can one credibly describe an anastomotic leak prediction system as part of routine clinical workflow rather than as a promising computational adjunct [20].

9. Conclusion

Moving anastomotic leak prediction from proof-of-concept to clinical workflow implementation requires more than incremental gains in retrospective discrimination. It requires a shift in problem definition from static binary classification to sequential decision support under bounded resources and evolving uncertainty. That shift immediately changes what counts as good modeling practice. Temporal alignment, phenotype fidelity, calibration, threshold policy, subgroup reliability, and interface legibility become first-order design constraints rather than downstream refinements. A system that performs well in retrospective evaluation but cannot explain its recommendations, tolerate missingness drift, or map its outputs onto explicit clinical actions is not close to implementation, even if its summary metrics appear competitive.

The technical path is therefore best understood as modular and workflow-centered. Representation should reflect baseline, procedural, and dynamic postoperative information separately. Scoring should allow nonlinear risk accumulation while preserving clinically interpretable contribution structure. Calibration should be maintained as an independent operational layer. Policy thresholds should be state-specific and utility-aware. Monitoring should treat drift, subgroup calibration, and action-induced distribution shift as routine phenomena rather than exceptional failures. These are not embellishments added after a model has been chosen. They are the architecture of implementation itself.

The organizational path is equally structured. Event-triggered inference, clear recipient accountability, conservative data plumbing, suppression-aware alert design, override capture, and staged prospective evaluation are the components through which a statistical system becomes a usable instrument in care. Governance matters because errors and false alarms arise from the joint behavior of model, data, workflow, and institution. Equity matters because reliability is expressed through access to measurement, threshold crossing, and response pathways, not only through aggregate performance averages. A neutral conclusion follows from these observations. Data-driven leak risk stratification can plausibly support perioperative care, but only when prediction, action, and oversight are designed together. Clinical implementation is not a final packaging step after modeling is complete. It is the broader technical problem within which modeling takes its operational meaning [21].

References

[1] N. El-Sourani, C. Kechagia, F. Alfarawan, A. Troja, and M. Bockhorn, "Classification and evaluation of anastomotic leaks after esophageal surgery—a tertiary university experience," *European Surgery*, vol. 54, pp. 80–85, 4 2021.

[2] B. Sadanandan, "Data-driven risk stratification for anastomotic leak: A proof-of-concept study using electronic health records," *Journal of Data Science, Predictive Analytics, and Big Data Applications*, vol. 9, no. 6, pp. 1–27, 2024.

[3] M. Hussain, "Tp7.2.10 the un-abating debate of using drains following colorectal anastomosis: a systematic review," *British Journal of Surgery*, vol. 108, 10 2021.

[4] M. Hofeldt and B. Richmond, "Elective robotic partial colon and rectal resections: Series of 170 consecutive robot procedures involving the da vinci xi robot by a community general surgeon," 1 2023.

[5] Y. Li, J.-J. Deng, and J. Jiang, "Relationship between body mass index and short-term postoperative prognosis in patients undergoing colorectal cancer surgery.," *World journal of clinical cases*, vol. 11, pp. 2766–2779, 4 2023.

[6] J. C. Benedetto, L. W. Cohen, and M. I. Abraham, "Transhiatal blunt esophagectomy for carcinoma of the esophagus.," *The Journal of the American Osteopathic Association*, vol. 90, pp. 54–54, 1 1990.

[7] J. Cwaliński, J. Paszkowski, F. Lorek, P. Samborski, M. Kucharski, H. Michalak, and T. Banasiewicz, "Minimally invasive treatment of postoperative fistulas, leakages, and perforations of the upper gastrointestinal tract: a single-center observational study.," *Wideochirurgia i inne techniki maloinwazyjne = Videosurgery and other miniinvasive techniques*, vol. 18, pp. 655–664, 12 2023.

[8] G. A. Cornejo, N. Jelen, and A. Gangemi, "Robot-assisted revisional surgery of gastric greater curvature plication to roux-en-y gastric bypass.," *Obesity surgery*, vol. 29, pp. 1434–1435, 2 2019.

[9] I. U. I. Nasir, M. F. Shah, S. Panteleimonitis, N. Figueiredo, and A. Parvaiz, "Spotlight on laparoscopy in the surgical resection of locally advanced rectal cancer: Multicenter propensity score match study.," *Annals of coloproctology*, vol. 38, pp. 307–313, 8 2021.

[10] C.-S. Liao, C.-W. Tu, C.-C. Chiang, and M.-C. Shieh, "Early total gastrectomy in emphysematous gastritis," *Advances in Digestive Medicine*, vol. 7, pp. 227–230, 3 2020.

[11] K. Adams and S. Papagrigoriadis, "Creation of an effective colorectal anastomotic leak early detection tool using an artificial neural network.," *International journal of colorectal disease*, vol. 29, pp. 437–443, 12 2013.

[12] L. Siragusa, G. Pellino, B. Sensi, Y. Panis, V. Bellato, J. Khan, and G. S. Sica, "Ambulatory laparoscopic colectomies: a systematic review.," *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland*, vol. 25, pp. 1102–1115, 2 2023.

[13] R. F. Coelho, K. J. Palmer, B. Rocco, R. R. Moniz, S. Chauhan, M. A. Orvieto, G. Coughlin, and V. R. Patel, "Early complication rates in a single-surgeon series of 2500 robotic-assisted radical prostatectomies: Report applying a standardized grading system," *European urology*, vol. 57, pp. 945–952, 2 2010.

[14] B. K. Sah, Z. Yang, L. Jian, Y. Chao, L. Chen, Z. Huan, Y. Min, and Z. Z. Gang, "Early diagnosis of anastomotic leakage after gastric cancer surgery," 4 2020.

[15] S. Vural, O. Civil, M. Kement, Y. E. Altuntas, N. Okkabaz, C. Gezen, M. C. Haksal, E. Gundogan, and M. Oncel, "Risk factors for early postoperative morbidity and mortality in patients underwent radical surgery for gastric carcinoma: a single center experience.," *International journal of surgery (London, England)*, vol. 11, pp. 1103–1109, 9 2013.

[16] S. Gavini, S. Kadiyala, and V. V. Reddy, "Choledochoduodenostomy: Our experience in 50 consecutive patients," *International journal of scientific research*, vol. 5, 8 2016.

[17] J. Tumnan, W. Suwanthanma, and C. Euanorasetr, "Risk factors for anastomotic leakage in open colorectal surgery: An analysis of 558 patients," *Journal of the Medical Association of Thailand*, vol. 100, pp. 47–, 11 2017.

[18] Y. M. Ozgun, V. Oter, E. Pişkin, M. K. Çolakoğlu, O. Aydin, and B. Bostanci, "Effects of distal pancreatectomy and splenectomy on outcomes in patients undergoing cytoreductive surgery and hyperthermic intraperitoneal chemotherapy with cc/0 resection," *Akademik Gastroenteroloji Dergisi*, vol. 20, pp. 104–111, 8 2021.

[19] N. Sugamata, T. Okuyama, E. Takeshita, H. Oi, Y. Hakozaki, S. Miyazaki, M. Takada, T. Mitsui, T. Noro, H. Yoshitomi, and M. Oya, "Surgical site infection after laparoscopic resection of colorectal cancer is associated with compromised long-term oncological outcome.," *World journal of surgical oncology*, vol. 20, pp. 111–111, 4 2022.

[20] M. Rutegård and J. Rutegård, "Anastomotic leakage in rectal cancer surgery : The role of blood perfusion," *World journal of gastrointestinal surgery*, vol. 7, pp. 289–292, 11 2015.

[21] V. Butnari, A. Mansuri, S. Kaul, J. Huang, and R. Nirooshun, "Tu4.2 robotic surgery for colorectal cancer: a single-center experience," *British Journal of Surgery*, vol. 109, 8 2022.