International Journal of Applied Sciences, Research, and Innovation

OPEN ACCESS

FULL PAPER | IJASRI | OPEN

Causal State Estimation and Information-Theoretic Surveillance for Early Detection of Anastomotic Leak After Gastrointestinal Reconstruction Procedures

Sidi Mohamed[1], **Ahmed Salem**[2], **Cheikh Ould**[3]

[1]Département de Génie Informatique, Institut Supérieur des Technologies de Rosso, Avenue du Fleuve Sénégal, Rosso 5160, Mauritania

[2]Unité de Mathématiques et Informatique Appliquée, École Normale Supérieure de Nouadhibou, Boulevard Maritime, Nouadhibou 7050, Mauritania

[3]Département des Systèmes d' Information, Université des Sciences, de Technologie et de Médecine de Nouakchott, Campus Universitaire, Tevragh-Zeina, Nouakchott 5026, Mauritania

Abstract

Anastomotic leak persists as a difficult postoperative complication because the local process that causes it begins before the bedside evidence becomes distinctive enough to force clinical certainty. The immediate challenge is therefore not only classification of who will eventually receive a leak diagnosis, but reconstruction of when the latent failure process has started to diverge from an uncomplicated trajectory. Early detection is difficult because the informative content of postoperative data is distributed across several channels that differ in timing, precision, and clinical meaning. Laboratory abnormalities, subtle hemodynamic instability, delayed physiologic recovery, altered nursing observations, operative descriptors, drain characteristics, and imaging cues often appear as weak fragments rather than as a single decisive signal. This paper develops a technical framework in which early leak detection is formalized as a partially observed, sequential, multimodal inference problem governed by biological latency, intervention-dependent observability, and asymmetric decision costs. The discussion integrates latent-state modeling, information-theoretic feature construction, causal adjustment for treatment effects, continuous-time representation learning, and action-oriented calibration. Particular emphasis is placed on the distinction between diagnosis and detectability, on the fact that missingness and measurement timing contain clinical signal, and on the need to evaluate surveillance systems by lead time, alarm burden, and transportability rather than by global discrimination alone. The resulting perspective treats postoperative surveillance as a dynamic control problem in which algorithmic estimates, clinical responses, and data acquisition strategies co-evolve. Within that frame, early detection becomes a question of identifying when accumulating evidence has crossed an actionable threshold before overt decompensation has become the principal source of prediction.

1. Introduction

The practical difficulty of anastomotic leak does not arise merely from its clinical seriousness, but from the temporal mismatch between where the complication begins and where clinicians can reliably observe it [24]. Mechanical strain, inadequate perfusion, bacterial contamination, ischemia at the bowel edge, impaired collagen deposition, edema, and host inflammatory dysregulation occur locally at or near the anastomotic site. Yet direct observation of that local environment is uncommon after the abdomen has been closed and the patient has entered postoperative recovery. Clinicians infer the status of the anastomosis indirectly through laboratory results, vital signs, analgesic needs, drain character, physical examination, delayed bowel function, imaging, and care-process escalations such as transfer to higher-acuity units. Each of these channels is imperfect. Each is also affected by therapeutic actions, documentation style, and baseline operative context. The central analytic problem is therefore not equivalent to ordinary postoperative risk prediction. It is the reconstruction of an evolving hidden process from asynchronous, intervention-modified, and only partially informative observations.

For that reason, the phrase early detection should be treated carefully. In many studies, leak is handled as a binary endpoint observed after clinical recognition or coding. That approach is useful for benchmarking but incomplete for bedside surveillance. A patient does not become biologically leak-positive only when reoperation occurs or a computed tomography report is finalized. Rather, a latent interval usually exists during which the local failure process is active, but the available evidence is still equivocal. The length of that interval differs across procedures, body compartments, diversion status, source control conditions, and host reserve. Some patients develop contained pelvic leaks with muted systemic expression, whereas others declare rapidly through diffuse contamination or septic physiology. If the purpose of modeling is to support earlier imaging, closer review, or lower thresholds for source control, then the relevant target is not simply

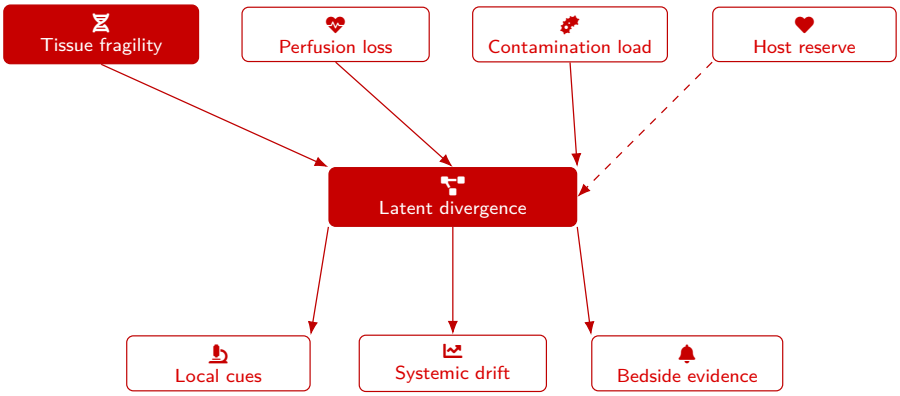


Figure 1. Biophysical leak genesis is shown as a hidden divergence process. Mechanical and biological vulnerabilities act upstream, while bedside evidence appears only after latent failure has already projected into local and systemic channels.

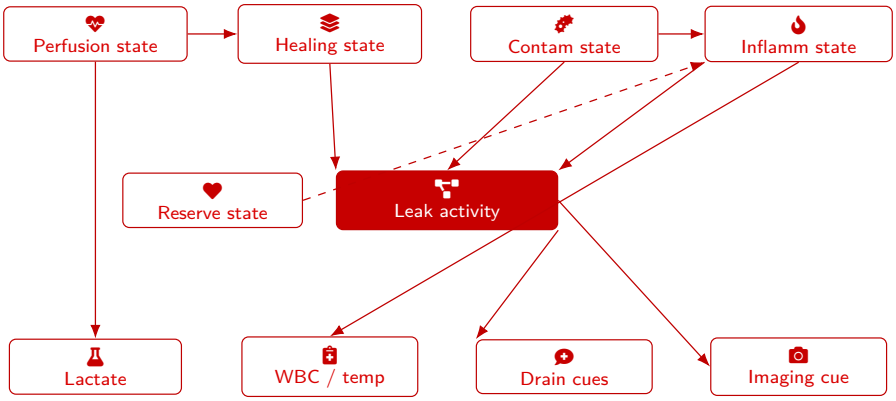


Figure 2. The latent state is decomposed into interacting components rather than a single score. Early observability depends on how perfusion, healing, contamination, inflammation, and reserve couple before they appear in laboratory, drain, or imaging channels.

eventual leak status but the earliest time at which the posterior probability of meaningful anastomotic failure becomes high enough to justify a changed management posture.

A useful empirical point of departure is that routinely collected data from the first postoperative day already appear to contain nontrivial signal for anastomotic leak, with postoperative white blood cell count, ICU admission, postoperative lactate, surgery type, and preoperative albumin carrying prominent importance among early features, and with overall discrimination suggesting that a clinically meaningful portion of leak-related information exists before overt diagnosis, as described by Sadanandan (2024) [20]. The significance of that observation is methodological rather than rhetorical. It implies that the earliest informative layer may already be embedded in standard perioperative records, even if that signal is weak, nonlinear, and heavily confounded by treatment intensity. The question then becomes how to extract that signal without mistaking surveillance density for biology, without collapsing dynamic trajectories into static covariates, and without allowing the model to exploit features that become available only after clinicians have effectively identified the problem themselves.

The distinction between susceptibility and detectability is foundational. Susceptibility concerns the conditions that make a leak more likely to occur: malnutrition, vascular disease, steroid exposure, hostile pelvic anatomy, long operative duration, emergent contamination, technical difficulty, and low tissue perfusion [16]. Detectability concerns whether the manifestations of a leak can be inferred from data that have already accumulated. These are related but not interchangeable. A patient with high baseline susceptibility can recover without complication if tissue repair progresses adequately and contamination remains controlled. A patient with low baseline susceptibility can nevertheless develop a leak if postoperative perfusion becomes compromised or technical failure emerges at the anastomosis. Early detection models therefore require a layered view. Baseline variables define an initial prior over healing vulnerability, while postoperative trajectories update that prior as evidence accumulates. Much of the clinical value lies in the update, not the prior alone.

Once framed this way, early detection becomes a problem at the boundary of physiology, statistics, and decision theory. Physiology determines which local states can give rise to observable systemic changes. Statistics determines how weak, irregular, and noisy observations can be fused into a coherent posterior distribution. Decision theory determines when that posterior should trigger intensified monitoring, diagnostic imaging, procedural intervention, or continued observation. A system can be statistically impressive and clinically unhelpful if its alerts occur only after bedside suspicion is already obvious. It can also be operationally disruptive if it is sensitive to common postoperative instability but insuffi-

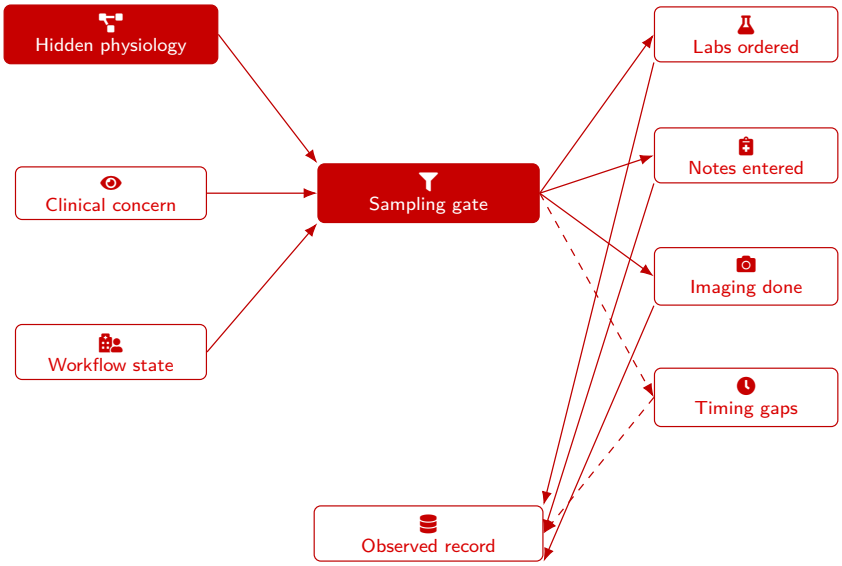


Figure 3. The postoperative record is generated by a care process, not by a neutral sensor. Measurement presence, timing, and omission carry mixed information about physiology, bedside suspicion, and institutional workflow.

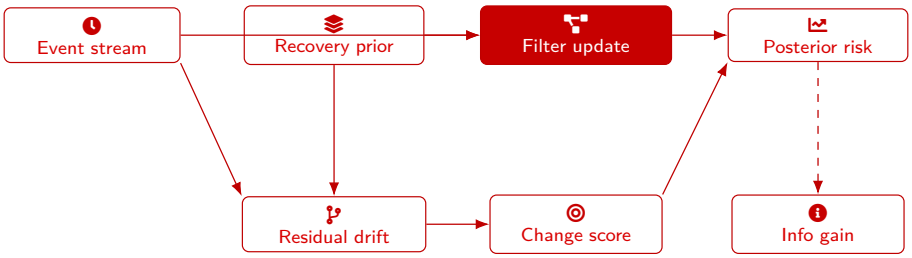


Figure 4. Sequential surveillance updates leak belief recursively as new events arrive. Residual deviation from expected recovery, structural change accumulation, and expected information gain jointly shape the decision surface before overt decompensation.

ciently specific for leak. A credible surveillance system must therefore be evaluated not only for discrimination, but for temporal usefulness, calibration, subgroup robustness, false-alarm burden, and explainability in the presence of changing care pathways.

This paper develops a technical perspective organized around that layered problem. The first aim is to articulate how local anastomotic failure becomes transformed into incomplete postoperative signals and why those signals are intrinsically heterogeneous. The second aim is to formalize postoperative surveillance as inference on a latent dynamical system observed through irregular multichannel measurements. The third aim is to examine how causal structure enters the problem, since therapeutic actions both respond to and reshape the physiology being monitored. The fourth aim is to connect probabilistic output to concrete alarm policies that are sensitive to clinical cost asymmetry and human oversight. The final aim is to outline what constitutes rigorous validation when the true objective is earlier action rather than retrospective endpoint prediction. The paper remains neutral in tone because the available evidence supports neither dismissal nor overstatement [3]. There is a plausible path to better postoperative leak surveillance, but it depends on respecting the fact that detectability, treatment, documentation, and workflow are mathematically entangled.

2. Biophysical Genesis and Latent Failure Dynamics

The earliest stages of anastomotic failure are biophysical rather than semantic. They involve local oxygen delivery, microvascular adequacy, tissue apposition, mechanical loading, collagen synthesis, bacterial burden, and the permeability of the healing interface. These processes are difficult to measure directly after surgery, yet they generate the posterior signals on which bedside reasoning depends. It is therefore useful to formalize leak emergence as a latent trajectory rather than as a single event. Let a hidden state summarize healing integrity, regional perfusion, contamination pressure, inflammatory activation, and systemic reserve. This state may move gradually away from the manifold of uncomplicated healing even while conventional observation remains close to normal. During that divergence interval, the anastomosis is not yet clinically obvious as failed, but the probability of downstream decompensation is increasing. Early detection, in a strict sense, means recognizing that divergence before the trajectory reaches a region where diagnostic certainty depends on severe systemic deterioration or imaging-confirmed contamination.

A mechanistic decomposition helps clarify why leak is so difficult to observe early. The first component is structural

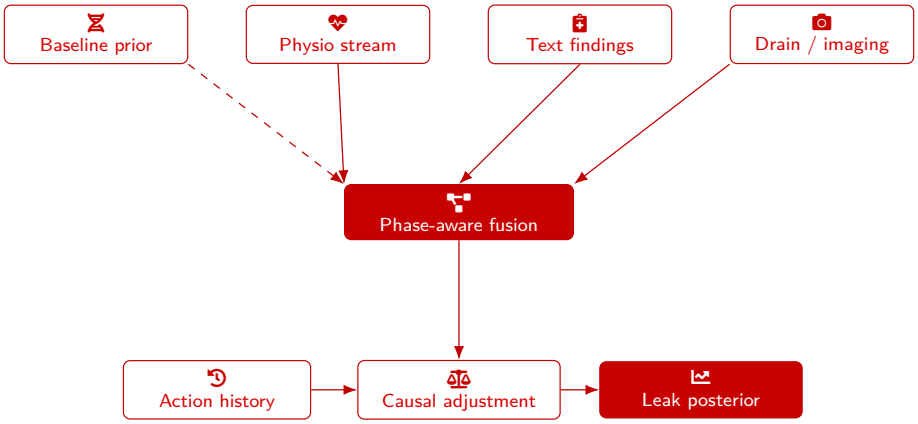


Figure 5. The multimodal stack preserves causal role as well as modality type. Baseline context shapes the prior, dynamic streams provide evolving evidence, and treatment history is handled explicitly before the final posterior is emitted.

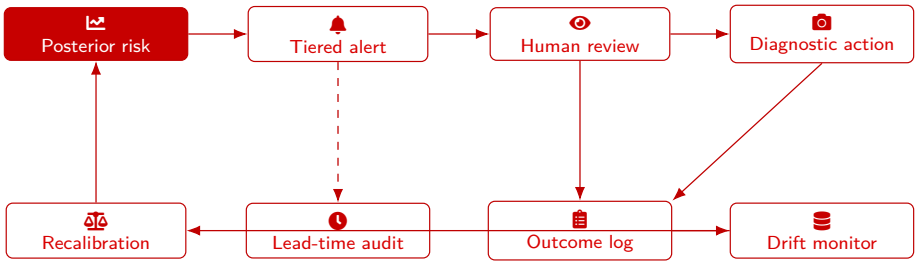


Figure 6. Risk estimates become useful only after they enter an operational loop. Tiered alerts, clinician adjudication, diagnostic escalation, audit of lead time, and continuous recalibration are all required for stable deployment under site and workflow shift.

vulnerability, determined by tissue quality, tension, vascular supply, and technical configuration. The second component is biological resilience, determined by inflammatory competence, nutritional reserve, perfusion responsiveness, and microbial containment. The third component is propagation, meaning the extent to which local failure escapes containment and begins to alter systemic physiology. A tiny defect can remain subclinical if contamination is limited and drainage is adequate. A similar defect can progress rapidly if the local environment is ischemic or feculent and host reserve is poor. Because these components interact, the same measured abnormality can imply different things across patients. Persistent tachycardia after a technically straightforward right hemicolectomy in a well-nourished patient may be more concerning than the same heart rate after a prolonged pelvic dissection accompanied by epidural-related vasodilation and postoperative pain. The mapping from biology to bedside data is therefore strongly contextual.

This contextual dependence can be expressed through coupled latent variables. Let h_t represent healing integrity, p_t regional perfusion sufficiency, c_t contamination burden, i_t inflammatory activation, and r_t physiological reserve. A simplified evolution can be written as [6]

$$\begin{aligned} h_{t+\Delta} &= h_t + \Delta\big(\alpha_p p_t - \alpha_c c_t - \alpha_i i_t\big) + \eta_t \\ i_{t+\Delta} &= i_t + \Delta\big(\beta_c c_t + \beta_h(1-h_t) - \beta_r r_t\big) + \zeta_t \\ \ell_t &= \sigma(\gamma_0 + \gamma_h(1-h_t) + \gamma_i i_t + \gamma_c c_t)\,. \end{aligned} \tag{1}$$

Here ℓ_t denotes instantaneous leak activity on a probabilistic scale, while η_t and ζ_t capture unmodeled stochastic influences. The equations are not intended as exact physiology. Their value lies in showing that inflammatory activation can rise because healing integrity falls, contamination expands, or reserve is insufficient, and that several distinct pathways can produce similar bedside patterns. A surveillance model that ignores these couplings may attribute leak to single predictors when the clinically relevant signal is actually the interaction among latent processes.

The notion of reserve deserves special emphasis because it governs how quickly local failure becomes globally observable. Two patients with comparable anastomotic defects may exhibit different clinical trajectories solely because one can buffer contamination and maintain hemodynamic stability longer. Reserve includes cardiopulmonary resilience, immunologic capacity, renal clearance, nutritional substrate availability, and ability to mount or suppress inflammatory cascades. In a high-reserve patient, systemic laboratory and vital-sign deviations may remain small despite meaningful local failure, making early detection more dependent on subtle trajectory anomalies or local information such as drain change. In a low-reserve patient, the same local perturbation may propagate quickly into global instability, increasing detectability but also narrowing the intervention window. Thus, the detectability of leak is not monotonic in severity alone. It is jointly determined by severity and the buffering properties of the host.

Mechanical factors further complicate the latent dynamics. Anastomoses exist within moving, distending, and pressure-sensitive structures. Edema, postoperative ileus, intraluminal

Table 1. Latent State Components and Clinical Interpretation

Symbol	Meaning	Clinical Proxy	Role
h_t	Healing integrity	Tissue viability	Structural state
p_t	Perfusion	Lactate, perfusion notes	Support factor
c_t	Contamination	Drain output, imaging	Injury driver
i_t	Inflammation	WBC, , CRP	Response signal
r_t	Reserve	Frailty, albumin	Buffer capacity

Table 2. Observation Modalities and Signal Characteristics

Modality	Timing	Noise Level	Clinical Meaning
Labs	Intermittent	Moderate	Systemic response
Vitals	Continuous	High	Instability trends
Text notes	Irregular	Variable	Context + suspicion
Drains	Sparse	Low	Local signal
Imaging	Delayed	Low	High specificity

pressure, coughing, vasopressor-associated microcirculatory changes, and mobilization can alter tension and perfusion across the healing interface. Some of these influences are transient and reversible; others push the tissue environment toward irreversible dehiscence. From a surveillance perspective, the important implication is that leak-related signal may emerge through failed recovery rather than through abrupt collapse. A lactate that clears slowly, a white blood cell count that remains elevated longer than expected, or a drain output that changes character without dramatic hemodynamic instability may each reflect a gradual failure of local equilibrium. Consequently, a model centered only on threshold exceedance will underutilize information encoded in prolonged deviation from expected recovery kinetics.

The temporal shape of inflammatory activation is also informative. Standard postoperative physiology already includes leukocytosis, mild fever, fluid shifts, and pain-related sympathetic activation. What distinguishes leak is often not the presence of these features, but their persistence, trajectory curvature, and co-occurrence with other abnormalities [1]. A white blood cell count that peaks early and declines may be compatible with uncomplicated recovery, whereas a second rise after initial normalization carries a different implication. Similar logic applies to lactate, temperature, heart rate, and bowel function. The practical consequence is that the latent process should be estimated relative to an individualized recovery baseline rather than to population-wide cutoffs. Biological interpretation improves when a patient's observed data are compared with what would have been expected for that patient under uncomplicated healing.

Contained and free leaks generate different observability profiles. A free leak may quickly provoke peritonitis, hemodynamic instability, and sepsis, producing a more obvious systemic signature. A contained leak, especially in the pelvis or near a drain, may evolve through prolonged ileus, low-grade inflammatory drift, subtle pain escalation, or intermittent fever. The latter cases are often the most consequential for early surveillance because they occupy the ambiguous zone in which proactive imaging or procedural drainage can materially change the clinical course. Models that are trained disproportionately on dramatic presentations may detect advanced cases well while remaining weak precisely where early recognition is most valuable. The latent-state formulation helps here because it allows the trajectory toward leak to have multiple modes, rather than assuming that all events travel through a single physiological pathway.

One should also note that leak is not independent of other postoperative syndromes. Pneumonia, bleeding, myocardial injury, urinary sepsis, aspiration, pancreatitis, ileus, and fluid overload each perturb parts of the same measurement space. The biophysical latent state is therefore embedded within a broader postoperative state space that includes several competing pathologies. In such an environment, specificity cannot come from isolated values alone. It must come from the sequence, combination, and persistence of changes relative to operative context. That insight is often more important than the choice of algorithm. If the underlying representation collapses all postoperative abnormality into a single severity axis, then leak detection will be confounded by the very processes that clinicians struggle to differentiate at the bedside.

The main lesson from the biophysical layer is that leak should be treated as an evolving hidden process with variable propagation into observable physiology. Early evidence is typically weak because the system is still close to the manifold of normal recovery. Yet weak evidence need not be useless if it is interpreted through appropriate temporal and contextual structure. The aim of modeling is therefore not to discover a universal signature that appears identically across patients, but to estimate whether the current postoperative trajectory is departing from individualized healing dynamics in a way that is mechanistically consistent with anastomotic failure [17].

3. Observation Process, Sampling Bias, and Identifiability

Once the local physiological process is acknowledged as hidden, the next question becomes how the postoperative record samples it. The answer is not straightforward because clinical data are not collected by a neutral measurement device. They are generated within a care process that reacts to emerging concern, to unit-level routines, to staffing patterns, and to the availability of diagnostic resources. A lactate value exists

Table 3. Sources of Bias in Postoperative Data

Source	Mechanism	Effect	Example
Measurement bias	Selective testing	Inflated signal	Lactate ordered
Treatment confounding	Actions alter data	Mask severity	Fluids reduce lactate
Documentation bias	Subjective notes	Leakage of suspicion	"Concern for leak"
Sampling frequency	Irregular timing	Hidden signal	ICU vs ward

Table 4. Sequential Modeling Approaches

Method	Strength	Limitation	Use Case
Kalman filter	Efficient	Linear assumption	Smooth trends
Particle filter	Flexible	Expensive	Nonlinear states
RNN models	Data-driven	Opaque	Multimodal fusion
Neural ODE	Continuous time	Complex training	Irregular sampling

because someone ordered it or because a protocol required it. A note exists because a clinician had something worth documenting or because a workflow demanded an entry. Imaging appears because suspicion or routine surveillance justified it. Missingness therefore contains information, but that information mixes biological state with institutional behavior. This dual role of the observation process is one of the central reasons retrospective prediction often overstates deployable performance.

Let x_t denote hidden physiological state and $y_t^{(k)}$ the observation in modality k at time t. Let $m_t^{(k)}$ indicate whether that modality is observed. A simple representation is

$$\begin{aligned} y_t^{(k)} &= \mu_k(x_t, a_t, c_t) + \epsilon_t^{(k)} \\ m_t^{(k)} &\sim \text{Bernoulli}\Big(\sigma\big(\delta_k^\top q_t\big)\Big) \\ q_t &= [x_t, a_t, w_t]. \end{aligned} \tag{2}$$

Here a_t denotes actions already taken and w_t denotes workflow state such as unit type, monitoring protocol, or clinician concern level. The crucial point is that $m_t^{(k)}$ depends not only on hidden physiology but also on care delivery. Thus the absence of a laboratory test is not just missing data in the ordinary statistical sense. It is itself a response of the care system to the current situation. If one ignores that dependence, the model may treat measurement density as pure signal for leak, when in reality it partly reflects clinician behavior. Such confounding may still appear predictive in retrospective data, but it weakens transportability and can generate self-fulfilling alerts after deployment.

Irregular timing adds another layer of complexity. Structured models often assume regularly spaced samples and use interpolation or last-observation-carried-forward to fill gaps. In postoperative surveillance, that assumption is clinically awkward. The interval between measurements is meaningful. A long gap may indicate stability, discharge trajectory, or low unit acuity. A short gap may indicate bedside concern, protocolized monitoring, or recent intervention. It is therefore better to represent the data as an event sequence with marked time intervals rather than as a regular grid [22]. The event representation preserves the fact that one measurement appearing four hours after the previous one conveys different information than the same value appearing thirty minutes later. Continuous-time models, marked temporal point processes, and neural state-space systems are all mathematically better aligned with this structure than static tabular encodings.

Identifiability becomes problematic when leak-sensitive variables are also treatment-sensitive. Suppose a patient receives fluid boluses and antibiotics because clinicians suspect deterioration. Subsequent vital signs and laboratory values then reflect both the underlying latent pathology and the intervention. If the model sees only the post-intervention values, it may underestimate latent severity because therapy has already partly attenuated the signal. If it sees the interventions as features, it may learn that antibiotic initiation or ICU transfer is predictive, which is true in retrospective data but partly circular. The prediction system must therefore decide whether it is estimating current untreated biological risk, treated risk under current management, or risk conditional on the care trajectory actually taken. These are different quantities. Most models silently estimate the third because they are trained on observational records. That is acceptable for pragmatic surveillance only if one recognizes its limits and handles actions explicitly.

Documentation channels differ in semantic density. Tabular data such as white blood cell count, lactate, temperature, heart rate, urine output, and vasopressor dose provide relatively direct quantitative traces of physiology. Clinical text contains richer local observations but also carries diagnostic suspicion in explicit linguistic form. A note stating concern for leak is not merely a predictor; it is evidence that clinicians have already formed a hypothesis. The same issue arises with radiology reports. If a model is allowed to ingest all documentation up to the recorded diagnosis time, it may perform well by using language that effectively encodes human recognition. For early detection, the observation process must therefore be censored in time and semantically filtered where possible. Otherwise the model measures whether clinicians can name the problem, not whether the data support earlier recognition than ordinary care.

Drain-related information occupies an intermediate category between local and systemic signal. A change in drain character, feculent output, or unexpected volume may be closer to the anastomosis than any blood test, but it is also highly contingent on whether a drain exists, where it lies, how it is documented, and how routinely it is assessed [9]. The presence

Table 5. Feature Types for Early Detection

Feature Type	Example	Temporal Role	Importance
Baseline	Albumin	Prior	Moderate
Trend	Rising WBC	Dynamic	High
Residual	Deviations	Relative	High
Event-based	ICU transfer	Discrete	Contextual
Interaction	WBC + HR	Coupled	High

Table 6. Information Gain from Clinical Actions

Action	Cost	Info Gain	Use Timing
Repeat labs	Low	Moderate	Early uncertainty
Bedside exam	Low	Moderate	Continuous
CT imaging	High	High	Elevated risk
Drain check	Low	High	Local suspicion

of a drain itself reflects operative judgment and case complexity. Thus drain features can be highly informative and highly biased simultaneously. Similar comments apply to intraoperative perfusion imaging. When present, it offers direct insight into tissue viability, yet its availability is procedure- and institution-dependent. The observation process is therefore not just incomplete; it is differentially rich across cases. Multimodal models must handle that heterogeneity without converting modality availability into a surrogate for hospital or surgeon identity.

The issue of identifiability can also be phrased geometrically. Let the hidden postoperative state space include regions corresponding to uncomplicated recovery, hemorrhage, cardiopulmonary deterioration, infectious complications other than leak, and leak itself. Observations are noisy projections from this hidden space into modality-specific measurement space. Some projections overlap heavily. Early leak may project onto a region nearly indistinguishable from ileus with inflammatory stress, while later leak may move into a more separable region. The goal of early detection is to exploit those dimensions along which the leak trajectory diverges earliest. These dimensions may not correspond to the largest variance in the data. They may be low-amplitude combinations of timing, persistence, and cross-modal correlation. Consequently, unsupervised dimensionality reduction that maximizes explained variance is not guaranteed to preserve the information most relevant for early leak inference.

A practical response to these challenges is to model individualized expected recovery and then analyze residual structure. For patient i, let $\hat{\mu}_{it}$ denote the expected measurement vector under uncomplicated recovery, conditioned on baseline and operative context. The residual $e_{it} = y_{it} - \hat{\mu}_{it}$ then captures deviation from expected course. This changes the problem from detecting absolute abnormality to detecting unexpected abnormality. A modest leukocytosis may be unremarkable after a long contaminated operation but concerning after a clean, short reconstruction with otherwise stable physiology. Residual modeling also helps with transportability because expected-recovery models can be recalibrated by site more easily than raw risk models whose thresholds are entangled with local monitoring intensity.

Another important consequence of the observation-process view is that performance metrics should distinguish between biological and documentation lead time. An alert generated before the diagnosis code appears may still be late if the relevant note, imaging order, or clinical suspicion already existed [23]. Therefore, a proper evaluation timeline should track operative end time, first abnormal trend, first escalated bedside concern, first diagnostic test specific to leak, formal diagnosis, and intervention. This fuller timeline allows one to ask whether the model identifies cases before existing clinical escalation or merely around the same moment using parallel evidence. Without such temporal granularity, the concept of early detection becomes ambiguous and can be overstated.

The observational layer therefore imposes a demanding but tractable requirement: the model must treat data acquisition as part of the signal, yet avoid mistaking care-process artifacts for universal biology. That balance is difficult, but it is central. A system that ignores the observation process will often be numerically impressive in development and disappointing in deployment. A system that explicitly represents measurement timing, missingness, action history, and individualized deviation has a better chance of identifying the weak but genuine structure that precedes overt leak recognition.

4. Sequential State Estimation and Information Gain

Once leak surveillance is formulated as inference on a hidden dynamical process, the most natural mathematical object becomes the filtered posterior over current latent state. The question is not whether the patient belongs to a fixed class, but what the present distribution over healing integrity and leak activity looks like given all observations so far. In this setting, sequence order matters intrinsically. A high white blood cell count after declining values means something different from the same count on a monotonic downward trajectory. A lactate level that fails to normalize after resuscitation has a different implication from a brief postoperative spike that resolves. Early detection is therefore best treated through state estimation rather than through one-shot classification.

Let x_t denote the latent postoperative state, u_t the intervention history, c static context such as procedure type and baseline reserve, and o_t the multimodal observation at time

Table 7. Alarm Policy Tiers

Tier	Risk Range	Action	Goal
Low	< 0.2	Observe	Stability
Moderate	0.2–0.5	Re-evaluate	Reduce uncertainty
High	0.5–0.8	Imaging	Diagnose early
Critical	> 0.8	Escalate care	Immediate action

Table 8. Evaluation Metrics for Surveillance

Metric	Definition	Strength	Limitation
AUC	Discrimination	Standard	Ignores timing
Lead time	Alert advance	Clinical value	Label dependent
Calibration	Prob accuracy	Decision support	Sensitive to shift
Alert burden	Alerts/day	Workflow impact	Context-specific

t. A continuous-time or discretized state-space representation can be written as

$$\begin{aligned} x_{t+\Delta} &= F_\Delta x_t + G_\Delta u_t + b(c) + \xi_t \\ o_t &= H_t x_t + J_t u_t + \varepsilon_t \\ p(x_t \mid \mathcal{F}_t) &\propto p(o_t \mid x_t) \int p(x_t \mid x_{t-\Delta}) p(x_{t-\Delta} \mid \mathcal{F}_{t-\Delta}) dx_{t-\Delta}. \end{aligned} \tag{3}$$

This compact expression encompasses a broad family of filters. Linear-Gaussian assumptions yield Kalman-type updates. Nonlinear physiology and non-Gaussian noise motivate particle filters, variational state-space models, neural ordinary differential equations, or recurrent latent-variable systems. The specific estimator matters less than the conceptual separation between biological evolution and observation formation. That separation allows uncertainty to be updated recursively as new data arrive, which is essential when evidence is sparse and interventions have already modified the trajectory.

The posterior quantity of direct clinical interest is often the instantaneous or near-term leak probability $\pi_t = \mathbb{P}(L_t = 1 \mid \mathcal{F}_t)$, where L_t is a binary latent indicator of meaningful anastomotic failure. Yet in practice, it is often preferable to estimate a graded state reflecting proximity to clinically relevant dehiscence rather than a sharp binary variable [4]. Binary formulations are attractive for decision thresholds, but they can encourage the model to learn abrupt transitions where the underlying process is gradual. A continuous latent severity scale permits more realistic representation of preclinical deterioration, after which binary alarm rules can be applied. This approach also handles label ambiguity more gracefully because chart-verified diagnosis times are only imperfect proxies for biological onset.

Hazard-based modeling offers another useful perspective. Let τ be the latent onset time of meaningful leak activity. Then one may define a time-varying hazard

$$\begin{aligned} \lambda_t &= \lambda_0(t) \exp\Big(\theta^\top \phi_t + \omega^\top c\Big) \\ R_t &= 1 - \exp\Big(-\int_0^t \lambda_s ds\Big) \\ \pi_t &= \mathbb{P}(\tau \le t \mid \mathcal{F}_t). \end{aligned} \tag{4}$$

Here ϕ_t encodes trajectory summaries, cross-modal interactions, and observation-process features accumulated up to time t. The value of this framing is that it explicitly treats onset as a time-indexed event and makes it possible to distinguish the risk that onset has already occurred from the risk that diagnosis will eventually be made. For postoperative surveillance, that distinction is clinically meaningful. A patient can have high probability that leak activity has begun while still lacking definitive imaging, and that interval is exactly where early detection aims to add value.

However, onset labels are rarely observed directly. Recorded diagnosis times are delayed and influenced by clinician decisions, imaging access, and documentation. One response is to treat the diagnosis indicator as a delayed observation of the latent onset process rather than as the onset itself. If d_t denotes the time at which leak becomes formally recognized, then training can proceed by allowing a learned delay distribution between onset and documentation. Hidden Markov models, interval-censored survival models, or positive-unlabeled objectives are all reasonable tools for that purpose. The key is to prevent the system from learning that the complication starts only at the moment of formal recognition. Without this adjustment, models optimized on delayed labels tend to overvalue late-stage features that are poor instruments for genuinely early surveillance.

An information-theoretic view becomes helpful when deciding which measurements carry the most incremental value at a given moment. Suppose the model can recommend additional tests or closer surveillance. The relevant quantity is no longer only the current posterior π_t, but the expected reduction in uncertainty if a particular observation were obtained. Let a denote a candidate information-gathering action, such as repeat lactate testing, contrast-enhanced imaging, or senior bedside review with structured documentation [15]. One may define the expected information gain

$$\begin{aligned} I_t(a) &= \mathbb{E}\big[D_{\mathrm{KL}}\big(p(L_t \mid \mathcal{F}_{t+1}^a) \,\|\, p(L_t \mid \mathcal{F}_t)\big) \big| \mathcal{F}_t\big] \\ &= \mathbb{E}\left[\sum_{\ell \in \{0,1\}} p(\ell \mid \mathcal{F}_{t+1}^a) \log \frac{p(\ell \mid \mathcal{F}_{t+1}^a)}{p(\ell \mid \mathcal{F}_t)}\right]. \end{aligned} \tag{5}$$

This quantity formalizes the intuitive idea that some tests are most useful not because they directly treat the patient, but because they sharply clarify whether the patient is drifting

Table 9. Transportability Challenges

Factor	Variation Source	Risk	Mitigation
Workflow	Hospital policy	Bias shift	Recalibration
Data density	ICU vs ward	Signal distortion	Time-aware models
Documentation	Style differences	Leakage	Text filtering
Protocols	ERAS adoption	Drift	Continuous monitoring

Table 10. Multimodal Fusion Strategies

Strategy	Mechanism	Advantage	Limitation
Early fusion	Concatenation	Simple	Loses structure
Late fusion	Separate models	Robust	Weak interaction
Attention	Weighted focus	Adaptive	Complex
Hierarchical	Context-aware	Clinically aligned	Design effort

toward leak. An early surveillance system is more clinically credible when it can identify not only elevated risk but also which next observation is expected to be most informative relative to its cost and invasiveness.

The signal relevant for early detection often lies in deviations from expected uncomplicated recovery rather than in raw measurements. For that reason, sequential models should estimate two related posteriors: one over the latent pathology state and one over the latent recovery baseline. Let $\bar{x}_t$ represent the counterfactual trajectory under uncomplicated healing for the same patient and operative context. The residual process $r_t = x_t - \bar{x}_t$ then captures pathological divergence. Filtering on r_t rather than on x_t alone can materially improve early sensitivity because it suppresses large but expected postoperative perturbations. A physiologically ordinary first-day leukocytosis after major pelvic surgery should not dominate the posterior if it is compatible with the predicted recovery course. By contrast, mild but persistent deviation across several channels may become more prominent because it is unexpected.

Change-point formulations are also useful. Rather than asking when the posterior leak probability exceeds a threshold, one may ask whether the postoperative trajectory has undergone a structural transition from expected healing dynamics to leak-compatible dynamics. Let $p_0(o_t \mid \mathcal{F}_{t^-})$ denote predictive density under uncomplicated recovery and $p_1(o_t \mid \mathcal{F}_{t^-})$ under leak-perturbed recovery. Then cumulative evidence for a change can be summarized by

$$S_t = \max_{1 \le \kappa \le t} \sum_{s=\kappa}^{t} \log \frac{p_1(o_s \mid \mathcal{F}_{s^-})}{p_0(o_s \mid \mathcal{F}_{s^-})}$$
$$T_A = \inf\{t : S_t \ge A\}. \tag{6}$$

This perspective is advantageous because it naturally encodes persistence. A single noisy laboratory value rarely moves the cumulative evidence far, whereas sustained divergence across time steadily raises S_t. For bedside care, that property is desirable. It mirrors the way experienced clinicians become more concerned when several minor irregularities persist or align rather than when one isolated value briefly crosses a threshold.

Representation learning can be inserted into this sequential framework by using self-supervised objectives that predict future observations under uncomplicated recovery and then measuring forecast residuals. If a model learns the ordinary postoperative dynamics of heart rate, inflammatory markers, renal function, and unit transitions, then unusual forecasting error itself may become a signal for early complication. Leak detection can then be posed as a second-stage problem on the residual trajectory and its structure. This two-step view has an appealing logic. The first stage learns the geometry of normal recovery. The second stage asks whether the patient is leaving that geometry in a leak-compatible direction [10]. The approach is especially relevant when leak cases are rare but uncomplicated recoveries are abundant.

The deeper implication of sequential state estimation is that probability alone is not enough. The model must also quantify uncertainty arising from sparse data, irregular observation, and subgroup-specific extrapolation. A probability of 0.35 generated from dense multimodal evidence is different from the same probability generated from a sparse ward record with few measurements. In the first case, the appropriate action may be immediate imaging. In the second, it may be intensified measurement to reduce uncertainty. Uncertainty therefore interacts directly with surveillance policy. Methods such as ensembles, Bayesian filtering, approximate posterior distributions, and conformal prediction are not ancillary technical embellishments. They are central because early detection often happens in the region where evidence is still ambiguous.

The sequential perspective also reshapes what a useful output looks like. Instead of a once-daily risk score, the system can provide a trajectory of posterior leak probability, estimated divergence from expected recovery, dominant sources of information gain, and confidence intervals reflecting data sparsity. Such output is more aligned with the temporal nature of the clinical question. The bedside team is not deciding in a vacuum whether the patient will ever leak. It is deciding whether the current direction of travel justifies changed management before the next ordinary checkpoint. State estimation, hazard modeling, information gain, and change-point logic together form a coherent mathematical basis for that decision.

5. Multimodal Representation and Causal Adjustment

The postoperative leak problem is multimodal because its earliest clues occupy different channels with different causal roles. Structured physiology represents the systemic consequences of local failure. Operative descriptors encode preconditions for poor healing. Text captures clinical nuance, local examination findings, and evolving concern. Imaging and perfusion assessments may reveal spatial or vascular detail unavailable in standard records. The critical challenge is not only to combine these modalities, but to combine them without erasing their distinct temporal and causal meanings [7]. A high-quality representation for early surveillance must be able to recognize that a rising inflammatory trajectory, an operative note suggesting marginal tissue perfusion, and a new description of foul drain output should not be treated as interchangeable inputs, even if all three push the posterior toward concern.

A modular representation strategy is therefore preferable to indiscriminate concatenation. Structured time series can be encoded by continuous-time recurrent units, temporal convolutions, or state-space layers that preserve event timing and sampling intervals. Operative notes and postoperative text can be embedded by language models that are adapted to perioperative semantics but constrained by temporal censoring. Imaging or intraoperative perfusion signals can be encoded by spatial models whose outputs are projected into a common latent clinical space. Fusion then occurs through gated or cross-attentive mechanisms conditioned on elapsed time, baseline context, and uncertainty. In such a design, the model is not forced to treat every modality as equally reliable at every moment. Instead, it learns when a modality is informative, when it is missing in a meaningful way, and when it may be unavailable because clinicians have not yet escalated the workup.

The representation should also preserve the distinction between local and systemic evidence. Operative text that mentions tenuous blood supply or challenging pelvic exposure is not evidence that leak has occurred; it is evidence that the anastomosis began with higher vulnerability. By contrast, a persistent postoperative lactate elevation or secondary inflammatory rise is evidence that the latent trajectory may already be diverging. These two types of information should occupy different positions in the model: one as prior-shaping context, the other as dynamic evidence. If they are simply pooled into a single vector, the model may overstate the significance of baseline vulnerability at the expense of real-time change, or it may ignore important prior structure when interpreting modest postoperative deviations. A causally coherent architecture lets baseline features initialize the prior distribution over latent healing dynamics, after which dynamic measurements update that distribution.

Text introduces special opportunities and risks. Surgeons and nurses often describe phenomena that are not easily captured in structured fields, such as escalating abdominal distension, drain odor, tenderness pattern, or a qualitative sense that the patient is not following the expected course. Such observations can be highly informative for early detection. Yet clinical language also directly transmits suspicion, and that suspicion is itself an intervention in the care process. A note that explicitly mentions concern for leak or orders a scan is not merely passive evidence; it marks a point at which human decision-making has already identified the possibility. Therefore, the safest text strategy for model development is to use temporally early notes, to strip explicit diagnostic labels when possible, and to decompose language into factual findings, uncertainty cues, and clinician assessment. This encourages the model to learn the language of postoperative deviation rather than the language of post hoc recognition [5].

Multimodal learning must also confront treatment confounding. A patient who receives aggressive fluids, broad-spectrum antibiotics, or ICU-level monitoring will generate a different data stream than an otherwise identical patient managed expectantly. Those actions both respond to perceived risk and alter the physiology later observed. If the model ignores treatment history, it conflates treated and untreated trajectories. If it uses treatment variables as ordinary predictors, it may learn that clinician concern itself is the signal. The causal target must therefore be chosen deliberately. One option is pragmatic prediction of leak under the observed management pathway. Another is estimation of latent leak status net of current treatment effects. The former aligns with operational forecasting; the latter aligns more closely with biological surveillance. A sophisticated system may need both quantities: one to forecast likely outcomes under current care, another to estimate underlying untreated severity.

A causal formalization clarifies the issue. Let L_t denote latent leak activity, A_t interventions, Y_t observations, and H_t the observed history. Then

$$\begin{aligned} Y_t &= f(L_t, A_t, H_{t^-}) + \epsilon_t \\ A_t &= g(H_t, \nu_t) \\ \Psi_t &= \mathbb{E}\big[L_t \mid do(A_{0:t} = a^\star_{0:t}), H_t\big]\,. \end{aligned} \tag{7}$$

The observational record identifies $\mathbb{E}[L_t \mid H_t]$ only indirectly and under assumptions. Yet the expression shows why action history must be modeled: current observations are downstream of both leak activity and prior intervention. Inverse-propensity weighting, marginal structural models, g-computation, representation balancing, and doubly robust objectives can each be used to reduce bias when estimating the impact of interventions on observability or when separating concern-driven actions from biology-driven changes. None eliminates unmeasured confounding completely, but each is preferable to silently ignoring the feedback loop.

A related challenge is label leakage through near-diagnostic modalities. Imaging, drain cultures, and operative text may contain highly specific information that becomes available only once clinicians have effectively entered the diagnostic phase. Training on these modalities without careful time censoring yields impressive retrospective performance but weak genuine early detection. One way to preserve their value without contaminating the task is to use them as auxiliary supervision rather than as synchronous predictors. For example, when postoperative imaging is available, its interpretation can supervise latent representations of earlier structured data, teaching the model which earlier patterns tend to precede spatial evidence of leak. In this semi-supervised arrangement, late-specific modalities shape the representation without being allowed to dominate the immediate prediction in a way that would be unavailable in real use [2].

Cross-modal attention can be made clinically more coherent by conditioning it on uncertainty and phase of recovery.

During the first hours after surgery, operative context and perfusion-related measurements may deserve higher weight because inflammatory variables are still dominated by routine stress response. Later, the relative weight may shift toward persistence of inflammatory deviation, hemodynamic instability, delayed gastrointestinal recovery, and note-derived local findings. A fixed fusion rule across all postoperative phases is unlikely to be optimal. The same is true across anatomical procedures. A low pelvic anastomosis and an upper gastrointestinal reconstruction do not project leak into the data stream in identical ways. Therefore, hierarchical context conditioning is advantageous: a global representation captures generic postoperative physiology, while procedure-specific adaptation layers adjust how modalities are interpreted.

The geometry of the learned latent space matters greatly. Ideally, uncomplicated recoveries should cluster along smooth trajectories parameterized by procedure and reserve, while leak-prone trajectories diverge in directions reflecting distinct phenotypes such as contained pelvic failure, diffuse septic contamination, or perfusion-dominant dehiscence. Contrastive objectives, temporal neighborhood preservation, and phase-conditioned triplet losses can help organize the space. Yet these techniques should be supervised by temporal validity. It is not enough that leak cases end up separate from nonleak cases; they must diverge early enough and in a way that is stable under site shift. Latent representations that separate cases only after clinicians have already escalated care are of limited bedside value.

Interpretability should arise from model structure rather than from retrospective explanation overlays alone. A useful multimodal surveillance system can expose how much of the current posterior comes from baseline vulnerability, from systemic physiological divergence, from documentation of local findings, and from care-process signals. It can also report which uncertainty sources dominate: sparse data, modality disagreement, subgroup extrapolation, or poor calibration in the current procedure class. Such decomposition is more informative than a generic heat map of feature importance because it aligns with clinical reasoning. Surgeons and intensivists do not usually think in terms of arbitrary predictor weights; they think in terms of whether the patient's current course is unexpected, whether the operative setting created fragility, and whether new observations support or contradict prior concern.

The multimodal and causal layers therefore converge on a single principle: representation learning for leak surveillance must preserve where information comes from, when it becomes available, and how care decisions have already shaped it. Models that flatten these distinctions may still obtain strong numerical results, but their predictions are harder to interpret and less likely to remain valid outside the development environment. By contrast, architectures that encode causal role, temporal phase, modality reliability, and intervention dependence are better positioned to infer leak early without confusing clinical suspicion, resource utilization, and underlying biology [8].

6. Alarm Policy, Utility, and Human Oversight

A surveillance model becomes clinically consequential only when its posterior estimates are converted into recommendations or actions. This conversion is not a secondary implementation detail. It is the stage at which probabilistic reasoning enters a workflow already constrained by staffing, diagnostic access, procedure-specific risk tolerance, and the consequences of false escalation. For anastomotic leak, the asymmetry of error is obvious. Delayed detection can culminate in sepsis, reoperation, organ failure, and prolonged hospitalization. Excessive false alarms can lead to unnecessary imaging, low-value antibiotic exposure, prolonged length of stay, invasive procedures, alarm fatigue, and eventual loss of trust in the system. The design of alarm policy must therefore be explicit, mathematically defensible, and sensitive to the fact that the optimal response to uncertainty is often information acquisition rather than immediate intervention.

Let $\mathcal{A}$ denote a set of available actions such as continued observation, repeat laboratory testing, intensified bedside review, cross-sectional imaging, interventional consultation, or urgent surgical reassessment. Given filtered information $\mathcal{F}_t$ and posterior leak probability π_t, an action rule can be written as

$$\begin{aligned} a_t^\star &= \arg\min_{a\in\mathcal{A}} \mathbb{E}[C(a, L_t, s_t) \mid \mathcal{F}_t] \\ &= \arg\min_{a\in\mathcal{A}} \sum_{\ell\in\{0,1\}} C(a,\ell,s_t)\,\mathbb{P}(L_t=\ell \mid \mathcal{F}_t). \end{aligned} \tag{8}$$

Here s_t denotes contextual state, including current physiological reserve, procedure class, diagnostic resource availability, and perhaps even discharge proximity. The same posterior probability can imply different optimal actions in different states. A moderate estimated probability in a frail patient with low reserve and rising inflammatory drift may justify immediate imaging. The same probability in a clinically robust patient with sparse data may justify repeat measurements and focused examination before exposing the patient to contrast and transport.

Because postoperative data are noisy, single-threshold policies tend to oscillate. A risk estimate may rise transiently after pain, mobilization, or fluid shifts and then fall again. To reduce instability, alarming should depend on cumulative or smoothed evidence rather than on an instantaneous score alone. One can define a persistence-weighted risk

$$\begin{aligned} \bar{\pi}_t &= \int_0^t \kappa(t-s)\pi_s ds \\ T^\star &= \inf\{t : \bar{\pi}_t \geq \Gamma(s_t)\}. \end{aligned} \tag{9}$$

The kernel κ discounts distant evidence while preserving the contribution of recent sustained concern, and $\Gamma(s_t)$ allows context-specific thresholds. This construction captures a clinically sensible principle: persistent moderate concern over several hours may be more actionable than a single brief spike to a higher value. It also helps align alerts with the way bedside teams reason, since they often become more worried about trajectories than about isolated numbers.

A dynamic programming perspective further clarifies the role of information gathering. Suppose the system can either continue observing, request an additional low-cost test, or trigger a high-cost diagnostic or therapeutic action. Then the value

function for optimal management can be expressed as [11]

$$V_t(\mathcal{F}_t) = \min\Big\{ c_o + \mathbb{E}[V_{t+1} \mid \mathcal{F}_t], \\ c_g + \mathbb{E}[V_{t+1} \mid \mathcal{F}_t, a_t = \text{gather}], \\ c_m \pi_t + c_f(1-\pi_t) \Big\}. \tag{10}$$

The first term corresponds to continued observation, the second to deliberate acquisition of more evidence, and the third to immediate major escalation such as imaging or urgent review with action-specific costs c_m and c_f. This representation is useful because it formalizes a point that is often implicit in clinical care: the best response to moderate posterior risk may be to acquire information rapidly rather than to act as though the diagnosis were already certain. For early leak detection, that intermediate layer is especially valuable because the complication is serious enough that waiting passively may be unsafe, yet uncertain enough that definitive intervention may still be disproportionate.

Alarm policy should ideally be tiered rather than binary. Low-tier alerts can prompt focused reassessment, earlier repeat measurements, or senior awareness. Intermediate-tier alerts can recommend imaging or interventional review. High-tier alerts can suggest immediate escalation or operative reconsideration. Such grading reduces the cognitive burden associated with all-or-none alarms and permits the model to respond proportionally to uncertainty. It also makes calibration more operational, because different posterior bands map onto different resource commitments. In that setting, calibration errors have concrete consequences. Overestimation in the intermediate tier may flood imaging pathways. Underestimation in the high-risk tier may delay source control. Therefore, a surveillance system should report not only discrimination but band-specific calibration and action yield.

Human oversight is not merely a safeguard; it is part of the control loop. Clinicians interpret the alert against information the model may not have captured well, such as nuanced physical examination, intraoperative memory, or evolving social constraints affecting diagnostic feasibility. At the same time, the model may detect weak multivariate divergence that is easy to miss in a busy ward environment. The most sensible arrangement is therefore cooperative, not substitutive. Alerts should invite structured confirmation or contradiction. The clinician can accept, defer, override, or annotate the signal. Those responses can then be logged and used to study whether the model is systematically overreacting to certain patterns or missing others. In other words, bedside adjudication is itself an informative observation that can refine future policy [12].

Trust enters the problem because repeated low-value alerts alter future behavior. A formally optimal threshold computed from retrospective costs may still be suboptimal once clinician adherence is considered. Excessive false positives decrease compliance, while a few timely accurate alerts can increase willingness to act on uncertainty. This means the effective cost function is partly endogenous to the history of model performance and user experience. Although difficult to encode precisely, one can approximate this by including alert burden and acceptance rate in the objective used to choose thresholds. Pragmatically, prospective silent deployment followed by staged rollout is often needed to estimate these behavioral effects before fully operationalizing alerts.

The timing of action relative to clinical workflow also matters. An alert that appears when no one is readily available to act, or after routine rounds have already passed, may have less effect than an alert delivered at a decision-rich moment. Similarly, the same alert may need different wording for nursing staff, residents, and attending surgeons. A bedside nurse might receive a message emphasizing reassessment of abdominal findings and drain quality, whereas the surgical team might receive a summary of risk trajectory and dominant contributing deviations. The content and recipient of the alarm are therefore elements of the policy, not just of the interface. Poor routing can convert a valid posterior estimate into inaction.

It is also important to distinguish between recommending diagnosis and recommending therapy. A system should be cautious about directly recommending invasive therapeutic actions unless prospectively validated for that purpose. In contrast, recommending intensified review or diagnostic imaging at elevated risk is more defensible because it acknowledges uncertainty while still responding to it. This is particularly appropriate in early leak surveillance, where the main benefit may come from shortening the interval between latent divergence and confirmatory evaluation. By narrowing that interval, the system may influence severity at diagnosis even if it does not prevent every leak event.

The alarm layer thus connects statistical inference to clinical consequences. Its design must respect persistence, uncertainty, asymmetry of error, workflow timing, and the fact that information gathering can itself be the optimal action. Without this layer, posterior probabilities remain analytically interesting but operationally incomplete. With it, early detection becomes a disciplined process of deciding when accumulating evidence is sufficiently strong, sufficiently persistent, and sufficiently actionable to justify changing the ordinary postoperative trajectory [18].

7. Transportability, Validation, and Adaptive Deployment

No matter how elegant the latent-state formulation or how refined the alarm policy, a surveillance model is not clinically credible unless it remains valid when moved across time, institutions, and user behavior. Anastomotic leak models are particularly vulnerable to transport failure because many predictors are entangled with workflow. Unit assignment, laboratory frequency, drain practice, documentation style, enhanced-recovery protocols, and threshold for imaging vary substantially across hospitals and even across surgical teams within the same hospital. These differences do not merely shift the marginal distributions of features; they alter the causal pathway by which leak becomes observable. As a result, retrospective single-center performance may say less about true generalizability than it initially appears.

Validation must therefore begin with the target estimand. If the model aims to estimate eventual leak diagnosis from data up to postoperative day one, then ordinary time-fixed discrimination has some meaning. If the goal is early detection of latent failure before current practice would ordinarily escalate, then evaluation must be aligned to that objective. The decisive metrics are lead time before clinical suspicion or confirmatory testing, sensitivity at clinically tolerable false-

alert rates, calibration within actionable risk bands, alert burden per patient-day, and yield of triggered diagnostic actions. Global area under the curve can still be reported, but it cannot serve as the primary criterion because it is insensitive to whether the signal arrived early enough to matter.

Temporal validation should therefore be constructed around postoperative landmarks and clinical events. Performance can be assessed at 6, 12, 24, and 48 hours after surgery, with strict censoring of inputs unavailable by each landmark. A more clinically rigorous approach is rolling prospective replay, in which each patient's record is reconstructed as an event stream and the model is evaluated on when it would have alerted relative to actual bedside suspicion, imaging order, diagnosis note, and intervention. Such replay makes it possible to separate administrative lead time from meaningful lead time. An alert that precedes the diagnosis code by ten hours but follows the first leak-directed scan order by eight hours is not genuinely early in the operational sense. This distinction is central to honest evaluation.

Calibration deserves equal weight because alarm policies depend on probability thresholds. Dynamic calibration should be checked across postoperative phases, not only in aggregate. A model can be well calibrated overall yet systematically overestimate risk in the immediate postoperative period and underestimate it later, or vice versa. Band-specific calibration is especially important when risk tiers map to different actions. Let B_b denote the set of predictions in band b and τ a prediction horizon [21]. Then a phase-specific calibration error can be expressed as

$$\begin{aligned} \mathrm{ECE}_\tau &= \sum_{b=1}^{B} \frac{|B_b|}{N} \left|\mathrm{freq}_\tau(B_b) - \mathrm{conf}(B_b)\right| \\ \mathrm{LeadTime} &= \frac{1}{N_+} \sum_{i:L_i=1} \left(T_i^{\mathrm{diag}} - T_i^{\mathrm{alarm}}\right)_+ . \end{aligned} \tag{11}$$

These quantities jointly describe whether the model is numerically trustworthy and whether it alerts with clinically useful advance notice. Reporting one without the other is insufficient.

External validation should ideally use internal-external cross-validation across several hospitals or surgical services. In such a design, the model is repeatedly trained on all but one site and tested on the held-out site. This approach reveals whether the representation has captured physiological structure or merely local practice. Temporal holdout is also necessary because protocols change. Enhanced-recovery pathways alter fluid management, feeding schedules, mobility, and expected laboratory trajectories. New perfusion technologies alter baseline operative risk. If the model degrades sharply across eras, it likely depends too heavily on historically contingent workflows. Site and era are therefore not nuisances to be averaged away. They are stress tests of whether the model has learned something transportable.

Fairness analysis should move beyond broad demographic comparisons of discrimination and examine the distribution of missed early detections, false alarms, and calibration errors across clinically relevant subgroups. Procedure type, diversion status, emergent versus elective setting, nutritional vulnerability, frailty, and anatomical location may matter as much as age or sex. A model could achieve similar overall discrimination across groups while still delivering shorter lead time in one subgroup and frequent low-value alerts in another. Such disparities matter because surveillance systems influence who receives diagnostic attention. If a subgroup systematically gets later alerts or more uncertain estimates, the system may amplify existing inequities in postoperative rescue.

Adaptive deployment requires monitoring for dataset shift after rollout. Once clinicians begin responding to alerts, the data-generating process changes. Repeat testing may become more frequent. Imaging may occur earlier [14]. Documentation may explicitly reference model concern. These adaptations can improve care and simultaneously destabilize the model. A responsible deployment therefore includes ongoing surveillance of feature distributions, missingness patterns, calibration curves, subgroup performance, and alert acceptance rates. When shifts are detected, recalibration or partial retraining may be necessary. This is not evidence of failure; it is a predictable consequence of introducing an active system into a learning health environment.

Prospective deployment should proceed in stages. Silent mode allows the institution to measure real-time calibration and alert burden without changing care. Assisted mode can then expose alerts to clinicians while leaving the response discretionary and heavily audited. Only after these stages should one assess outcome impact through pragmatic interventional designs. The relevant outcomes include time to leak-directed imaging, severity at diagnosis, unplanned ICU transfer, need for urgent reoperation, burden of unnecessary imaging, and clinician workload. It is possible that the main benefit of surveillance will be a reduction in severity at recognition rather than a reduction in leak incidence itself. Evaluation should be open to that possibility.

Digital infrastructure is another determinant of deployability. Real-time leak surveillance depends on reliable timestamps, robust mapping of operative end time, consistent patient identity resolution, normalization of laboratory units, handling of delayed note finalization, and identification of genuine missingness versus pipeline failure. These are often treated as engineering details, but they directly affect inference. A model expecting minute-level vital signs cannot be safely deployed if ward data arrive in delayed batches. A model using note-derived local findings cannot function well if notes are finalized many hours after bedside events. High-quality deployment therefore requires that the measurement architecture be aligned with the assumptions of the modeling architecture.

Transportability also depends on whether the model's dominant evidence sources remain clinically sensible after relocation. If the model originally relied heavily on ICU admission, lactate ordering frequency, or particular note phrases, then those features may fail elsewhere even if formal recalibration is applied. Explainability audits across sites can reveal whether the system is still basing alerts on coherent physiological divergence or whether it has shifted toward local documentation artifacts [13]. These audits are especially valuable in leak surveillance because care-process predictors are both strong and brittle. A system whose internal reasoning changes dramatically across hospitals is less trustworthy than one whose performance varies modestly but whose evidence structure remains stable.

The deeper message is that early leak detection cannot be validated in a purely static manner. The target, the labels,

the workflow, and the data stream all evolve. A robust model must therefore be treated as an adaptive instrument embedded in practice, not as a finished artifact whose retrospective score settles its clinical worth. Transportability, fairness, calibration, and postdeployment drift are not afterthoughts. They are constitutive of whether the system has any realistic chance of improving postoperative rescue.

8. Conclusion

The problem of early anastomotic leak detection is best characterized as sequential inference on a partially observed postoperative dynamical system. Local tissue failure develops before it becomes diagnostically obvious, and the evidence that eventually reveals it is distributed across heterogeneous channels that are themselves shaped by treatment, workflow, and institutional routine. For that reason, clinically useful surveillance requires more than a static risk score. It requires representation of latent healing dynamics, explicit modeling of observation timing and missingness, accommodation of treatment-induced confounding, and translation of posterior risk into graded action policies that acknowledge uncertainty and resource constraints.

A coherent technical framework follows from that premise. Baseline operative and patient factors define a prior over healing vulnerability. Dynamic physiological and contextual observations update that prior through continuous-time state estimation, hazard modeling, and change-detection logic. Multimodal evidence must be fused in a way that preserves causal role and temporal availability. Alarm policies must reward persistence and actionable lead time rather than instantaneous severity alone. Validation must focus on calibration, transportability, subgroup behavior, and the practical interval between model alert and ordinary clinical recognition. Deployment must be adaptive because clinician responses alter the future data stream and therefore the validity of the model itself.

Under this view, the central question is not whether postoperative records can predict eventual leak status in hindsight. The central question is whether accumulating evidence can identify clinically meaningful divergence from expected recovery before overt deterioration becomes the dominant source of signal. Answering that question demands methods that are biologically informed, statistically careful, and operationally realistic. Progress is likely to come from systems that treat physiology, data acquisition, causal structure, and bedside decision-making as components of the same surveillance problem rather than as isolated technical layers [19].

References

[1] Waleed Al-Khyatt, Charles A. Bull, Sherif Awad, and Javed Ahmed. Laparoscopic roux en-y gastric bypass using a modified retrocolic-supracolic approach: Outcomes from 300 patients. *World journal of surgery*, 40(8):1918–1924, 3 2016.

[2] Mohamed Mahmoud Ali and Ashraf Mohammad El-Badry. Influence of combined mechanical-chemical versus mechanical bowel preparation on anastomotic leak and surgical site infections after elective resection of left colon cancer. *International Surgery Journal*, 7(10):3192–, 9 2020.

[3] Jerry T Dang, Tiffany Vaughan, Valentin Mocanu, Hadika Mubashir, Juan S Barajas-Gamboa, Ricard Corcelles Codina, John Rodriguez, Shahzeer Karmali, and Matthew Kroh. Conversion of sleeve gastrectomy to roux-en-y gastric bypass: Indications, prevalence, and safety. *Obesity surgery*, 33(5):1486–1493, 3 2023.

[4] Carmen Dingemann, J. Dietrich, Jan Zeidler, J. Blaser, Jan-Hendrik Gosemann, Benno Ure, and Martin Lacher. Early complications after esophageal atresia repair: analysis of a german health insurance database covering a population of 8 million. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus*, 29(7):780–786, 4 2015.

[5] Silvio Däster, Savas D. Soysal, Luca Koechlin, Lea Stoll, Ralph Peterli, Markus von Flüe, and Ackermann C. Perioperative outcomes of esophageal cancer surgery in a mid-volume institution in the era of centralization. *Langenbeck's archives of surgery*, 401(6):787–795, 7 2016.

[6] K. P. Epari and Richard J. Cade. Hp20 oesophagectomy for tumours and dysplasia of the oesophagus and gastro-oesophageal junction: Experience of a single surgeon. *ANZ Journal of Surgery*, 77(s1), 4 2007.

[7] E. Espín, M. A. Ciga, Miguel Pera, and H. Ortiz. Oncological outcome following anastomotic leak in rectal surgery. *The British journal of surgery*, 102(4):416–422, 1 2015.

[8] Mark Gray, Jamie R. K. Marland, Alan F. Murray, David Argyle, and Mark A Potter. Predictive and diagnostic biomarkers of anastomotic leakage: A precision medicine approach for colorectal cancer patients. *Journal of personalized medicine*, 11(6):471–, 5 2021.

[9] Francesco Guerra, Giuseppe Giuliani, Diego Coletta, Marcello Boni, Fabio Rondelli, Paolo Bianchi, and Andrea Coratti. A meta-analysis of randomized controlled trials on the use of suction drains following rectal surgery. *Digestive surgery*, 35(6):482–490, 12 2017.

[10] Philip J. J. Herrod, H Boyd-Carson, Brett Doleman, J Trotter, S Schlichtemeier, G Sathanapally, J Somerville, John P. Williams, and Jonathan N. Lund. Quick and simple; psoas density measurement is an independent predictor of anastomotic leak and other complications after colorectal resection. *Techniques in coloproctology*, 23(2):129–134, 2 2019.

[11] Hiroaki Iwamoto, Yoshifumi Kadono, Hiroshi Yaegashi, Masashi Iijima, Shohei Kawaguchi, Takahiro Nohara, Kazuyoshi Shigehara, Kouji Izumi, and Atsushi Mizokami. Examination of necessity for pelvic drain placement after robot-assisted radical prostatectomy, 5 2021.

[12] Shifa A. Kalokhe, Rohit Sonawane, Shrutik Devdikar, and G. S. Narshetty. Single interrupted vs. continuous all layer closure in bowel anastomosis in emergency surgeries: a comparative study. *International Journal of Research in Medical Sciences*, 11(2):518–522, 1 2023.

[13] Shigeki Kusamura. Peer review report for: Hyperthermic intraperitoneal chemoperfusion with high dose oxaliplatin: Influence of perfusion temperature on postoperative outcome and survival [version 2; peer review: 1 approved, 2 approved with reservations], 11 2015.

[14] Yingjun Liu, Gangcheng Wang, Xiangbin Wan, Yong Cheng, Youcai Wang, and Xiaoyong Liu. Use of pedicled omentum in the prevention of anastomotic leakage after resection of obstructive colorectal cancer. *Chinese Journal of General Surgery*, 32(01):23–25, 1 2017.

[15] Yarman Mazni, Ardani Syafiuddin, and Agi Satria Putranto. Intraoperative pancreatic assessment in pancreaticoduodenectomy the correlation with pancreatic fistula formation. *The New Ropanasuri : Journal of Surgery*, 5(1):12–15, 6 2020.

[16] Michał Mik, Justyna Magdzinska, Lukasz Dziki, Marcin Tchórzewski, Radzisław Trzciński, and Adam Dziki. Relaparotomy in colorectal cancer surgery –do any factors influence the risk of mortality? a case controlled study. *International journal of surgery (London, England)*, 12(11):1192–1197, 9 2014.

[17] Christos Papadimitriou, Wilfried Martin, Athanasios E Dellis, Charalampos Deliveliotis, and Iraklis Mitsogiannis. Twenty-four-month efficacy of ureteroureterostomy combined with unilateral nephrostomy following radical cystectomy. *Cureus*, 14(10):e30478–e30478, 10 2022.

[18] Hans-Christian Pommergaard, Jesper Arnold Olsen, Stefan K. Burgdorf, and Michael Patrick Achiam. Laparoscopic versus right-sided hemicolectomy in cancer of colon therapy. *Ugeskrift for laeger*, 172(13):1034–1038, 3 2010.

[19] Dimitrios Raptis, Matthias Maak, Christian Krautz, Susanne Merkel, Maximilian Brunner, Abbas Agaimy, Arndt Hartmann, Sabine Semrau, Oliver Ott, Fietkau Rainer, Jürgen Siebler, Robert Grützmann, Werner Hohenberger, and Claus-Wilhelm Schildberg. Surgical and oncological outcome after extended lymph node dissection for carcinoma of the stomach and the esophagogastric junction: a retrospective analysis from an experienced single center. *Archives of medical science : AMS*, 20(1):124–132, 12 2021.

[20] Binesh 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*, 9(6):1–27, 2024.

[21] Iestyn M. Shapey, Angela Summers, Hussein Khambalia, Petros Yiannoullou, Catherine Fullwood, Neil A. Hanley, Titus Augustine, Martin K. Rutter, and David van Dellen. Donor insulin therapy in intensive care predicts early outcomes after pancreas transplantation. *Diabetologia*, 64(6):1375–1384, 3 2021.

[22] Eelco J. Veen, Jessica Steenbruggen, and Jan A. Roukema. Classifying surgical complications: A critical appraisal. *Archives of surgery (Chicago, Ill. : 1960)*, 140(11):1078–1083, 11 2005.

[23] Eran Zittan, Nathalie Wong-Chong, Grace W., Robin S. McLeod, Mark S. Silverberg, and Zane Cohen. Modified two-stage ileal pouch-anal anastomosis results in lower rate of anastomotic leak compared with traditional two-stage surgery for ulcerative colitis. *Journal of Crohn's & colitis*, 10(7):766–772, 3 2016.

[24] Mehmet Ali ÇAPARLAR, Şeref DOKÇU, and Salim DEMİRCİ. Loop ileostomy or loop transverse colostomy for resectable rectal cancers. *The European Research Journal*, 8(3):340–346, 5 2022.