
A Translational Predictive Analytics Framework for Explainable Risk Assessment: Transforming High-Dimensional Surgical Data into Clinical Decision Support Tiers (S-CRI)

Ioanna Michou , [Ioannis Maroulis](#) , [Ioannis Hatzilygeroudis](#) , [Constantinos Koutsojannis](#) *

Posted Date: 9 June 2026

doi: 10.20944/preprints202606.0085.v2

Keywords: clinical decision support systems (CDSS); predictive analytics; explainable risk assessment; digital health; healthcare decision-making; surgical complications; risk tiers



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

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

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Translational Predictive Analytics Framework for Explainable Risk Assessment: Transforming High-Dimensional Surgical Data into Clinical Decision Support Tiers (S-CRI)

Ioanna Michou ¹, Ioannis Maroulis ², Ioannis Hatzilygeroudis ³
and Constantinos Koutsojannis ^{4,*}

¹ Physiotherapy Department, School of Health Rehabilitation Sciences, University of Patras, Patras, Greece

² Department of Surgery, School of Health Sciences, University of Patras, Patras, Greece

³ Computer Engineering & Informatics Department, School of Engineering, University of Patras, Patras, Greece

⁴ Health Physics & Computational Intelligence Laboratory, Physiotherapy Department, School of Health Rehabilitation Sciences, University of Patras, Patras, Greece

* Correspondence: ckoutsog@upatras.gr

Abstract

The implementation of automated clinical decision support systems (CDSS) is essential for enhancing patient safety and optimizing public healthcare ecosystems. While advanced predictive analytics and machine learning (ML) architectures offer high diagnostic accuracy, their "black-box" nature and heavy infrastructure demands often restrict their real-time deployment at the bedside. This study introduces a translational framework designed to close this gap by converting high-dimensional predictive intelligence into an explainable, zero-overhead deployment pathway: the Surgery-Complication Risk Index (S-CRI). Using a multicenter digital health registry of 19,965 surgical records from Greek public hospitals, we subjected macroscopic parameters to target-directed combinatorial topology and to multivariate logistic regression optimization. Length of hospital stay (Adjusted Odds Ratio [aOR] = 1.44 per day), admitting hospital department classifications, and empirically pooled entry diagnosis risk clusters were isolated as core predictive features. These continuous statistical weights were linearly aligned into an integer scorecard where 1.0 point maps directly to the independent risk contribution of a single inpatient day. Signal detection validation demonstrated excellent discriminative power, yielding an Area Under the ROC Curve (AUC) of 0.882. Operating at an optimized screening threshold of -4, the index achieved a sensitivity of 66.56% and a specificity of 89.64%, effectively minimizing alert fatigue. Continuous risk scores were stratified into five actionable digital health triage tiers ranging from Very Low (<1% complication incidence) to Very High (>75%). Rather than acting as an alternative to AI, the S-CRI functions as a transparent, explainable layer for clinical decision support. This framework democratizes predictive medical insights, providing an immediate deployment vector for resource-constrained environments within modern digitized healthcare systems.

Keywords: clinical decision support systems (CDSS); predictive analytics; explainable risk assessment; digital health; healthcare decision-making; surgical complications; risk tiers

1. Introduction

1.1. Epidemic of Post-Operative Complications and Global Patient Safety Mandates

The mitigation of adverse post-operative events and the systematic minimization of preventable medical errors represent some of the most critical challenges facing contemporary clinical governance and global public health economics. Within high-acuity surgical departments, post-operative complications—ranging from localized surgical site infections (SSIs) to systemic sepsis, deep vein thrombosis, acute respiratory distress, and multi-organ failure cascades—do not merely represent isolated deviations from ideal clinical pathways. Instead, they constitute an institutional epidemic that compromises patient survival vectors, dramatically elongates hospital lengths of stay, prompts unplanned secondary surgical re-explorations, and drives up institutional financial liabilities [1,2].

The World Health Organization (WHO) and various international patient safety coalitions have consistently issued global mandates demanding the formulation of objective, proactive, and highly reproducible risk-stratification models [2]. Historically, the clinical identification of patients structurally predisposed to poor surgical outcomes relied almost entirely on retrospective administrative audits, generalized physiological scoring rubrics, or the subjective, localized experience of the attending surgical team during bedside rounds. This legacy approach introduces substantial inter-observer variability and limits a clinical department's ability to implement pre-emptive, standardized prophylactic protocols before a patient's physiological trajectory experiences an irreversible decline [3,4]. The modern transition toward ubiquitous, digitized Electronic Health Records (EHRs) has profoundly shifted this clinical paradigm, turning historical databases into active computational assets capable of driving real-time risk mitigation [5,6].

1.2. The Digital Divide and Institutional Friction of Infrastructure-Heavy AI in Public Healthcare

The exponential expansion of medical informatics has sparked a massive wave of research dedicated to deploying advanced Artificial Intelligence (AI) and Machine Learning (ML) architectures directly within the healthcare ecosystem. In our recently accepted foundational study [2], we demonstrated that high-dimensional computational models—including multi-layer artificial neural networks, deep learning frameworks, and complex gradient-boosted ensemble tree classifiers (such as XGBoost and Random Forests)—could successfully mine complex EHR registries from Greek surgical departments to map hidden risk patterns for medical complications [3]. By evaluating high-degree polynomial feature interactions and managing thousands of distinct permutations of patient data, these algorithmic pipelines achieved exceptional predictive benchmarks.

However, the practical, widespread implementation of these advanced computational models reveals an acute "digital divide" across the global healthcare landscape. Modern machine learning workflows do not operate in a vacuum; they require an extensive, unyielding, and highly specialized technological infrastructure. To run seamlessly, these models demand persistent high-performance local servers or expensive cloud-computing architectures, low-latency enterprise local area networks, specialized application programming interfaces (APIs) for real-time data ingestion, continuous software dependency containerization, and dedicated on-site data engineering teams [3,7].

These technological requirements are highly unevenly distributed. In public, rural, underfunded, or structurally resource-constrained healthcare environments—such as many regional public hospitals across southern Europe and developing nations—this level of digital infrastructure is either completely fragmented or absent. Consequently, forcing clinical departments to rely solely on high-computing AI frameworks creates an operational bottleneck that inadvertently excludes vulnerable patient populations in under-resourced settings from the benefits of modern predictive analytics [8].

1.3. The Interpretability Crisis and Cognitive Psychology of "Black-Box" Algorithmic Resistance

Beyond the physical infrastructure boundaries of deployment, advanced AI architectures suffer from a profound sociological and cognitive limitation: the interpretability crisis. Deep neural networks and ensemble tree classifiers function intrinsically as mathematical "black boxes." While

they output highly accurate probability vectors, their internal mathematical logic—characterized by millions of non-linear weight distributions and deep node splits—is completely obscured from human view. It is entirely isolated from the clinical, pathophysiological rationale that physicians rely on to make life-or-death decisions [3,4].

This opacity induces a natural, highly protective cognitive resistance among clinical staff. An attending surgeon or intensive care physician is legally, ethically, and professionally accountable for the patient's life [7,8]. Consequently, they are deeply hesitant to alter a validated clinical care pathway, initiate aggressive prophylactic therapies, order invasive diagnostic line placement, or delay a critical hospital discharge based solely on an automated risk percentage generated by a system that cannot explain *why* a patient is classified as high-risk. When an algorithm fails to provide clear, traceable, and physiologically sound explanations for its outputs, it triggers high clinical non-compliance, rendering the tool practically useless at the active bedside.

1.4. *Mathematical Information Compression as a Biomedical Engineering Solution*

1.4. The S-CRI as an Explainable Translational Layer for AI-Assisted Healthcare

The primary challenge of modern digital health engineering is not simply building more accurate predictive analytics engines but resolving the translational friction that occurs when deploying these systems into active clinical workflows. High-dimensional computational pipelines excel at finding non-linear feature interactions within vast databases, yet they remain isolated from the fast-paced, high-acuity realities of bedside healthcare decision-making. Forcing a busy surgical team to query an external, cloud-based black-box algorithm during emergency rounds introduces split-second delays and cognitive friction [9,10]. To bridge this operational divide, this study presents the Surgery-Complication Risk Index (S-CRI) not as a rejection of machine learning, but as its **transparent translational extension**. We propose a dual-layer digital health model: advanced predictive analytics tools operate at the institutional registry tier to discover deep clinical risk patterns, while compressed, mathematically aligned linear scorecards serve as the explainable execution interface at the active bedside. By applying optimization matrices to 19,965 surgical records, we distill the core predictive value of high-cardinality EHR data into an additive integer scorecard. This tool scales clinical weights directly against a tangible real-world metric: one single day of in-hospital stay [11,12].

This approach shifts the paradigm of clinical decision support. Instead of forcing under-resourced hospital environments to maintain expensive computing infrastructure, the S-CRI framework acts as a practical deployment pathway. It keeps the predictive power of large-scale statistical modelling intact while delivering it via a transparent, auditable interface that clinicians can intuitively trust, verify, and act upon in real-time [13,14].

2. Materials and Methods

2.1. *Cohort Selection, Multi-Center Registry Architecture, and Ethical De-identification Protocols*

This study was executed utilizing a comprehensive retrospective cohort of 19,965 distinct patient records harvested over a multi-year operational window across representative general surgery and specialized acute-care departments within public Greek hospital registries. The inclusion criteria dictated the extraction of all consecutive adult patients (aged 18 and older) who underwent an operative intervention and required inpatient hospitalization for a duration spanning a minimum of 24 hours [5,16]. Patients with incomplete administrative metadata, corrupted diagnostic records, or those discharged or transferred within the immediate post-operative recovery hour were excluded to ensure analytical consistency across the cohort [3,14,17].

The structured multi-centre registry architecture captured a wide range of administrative, demographic, and clinical data points. The variables mined included the specific insurance coverage provider category (INSURER), the patient's biological sex (SEX), the precise name/type of the admitting hospital department (HOSPITALITY-NAME), the exact duration of continuous inpatient

admission measured in integer days (DAYS-OF-HOSPITALITY), and the primary admission diagnostic code (ENTRY-DIAGNOSIS) [3].

The primary target outcome variable was designated as COMPLICATION—a binary metric where 0 indicated an uncomplicated post-operative course leading to a standard discharge, and 1 denoted the verified incidence of one or more post-surgical complications (such as hospital-acquired nosocomial infections, wound dehiscence, sepsis, or deep vein thrombosis) emerging within the inpatient timeline. To ensure strict compliance with institutional ethical review boards and international data privacy mandates, all records were subjected to an automated de-identification protocol before analysis. All direct patient identifiers, including names, social security numbers, specific admission dates, and exact chart numbers, were entirely removed, replacing the primary records with structurally sound, pseudonymous research keys.

2.2. Combinatorial Topology of High-Cardinality Variables: The Empirical Risk Pooling Protocol

One of the most complex challenges in translating high-dimensional EHR data into a simplified bedside scoring model is managing high-cardinality nominal variables. In the raw dataset, the variable ENTRY-DIAGNOSIS contained dozens of distinct, alphanumeric international diagnostic codes. While advanced machine learning frameworks handle this high dimensionality through multi-hot encoding or deep embedding layers, standard regression models suffer from a rapid loss of degrees of freedom, severe coefficient inflation, and overfitting when forced to evaluate dozens of individual dummy variables. Furthermore, a scorecard requiring a clinician to cross-reference dozens of separate point possibilities for a diagnosis code becomes too slow and complicated for practical bedside use [3].

To solve this problem, we applied an engineering technique: **empirical risk pooling based on target-directed combinatorial topology**. Every individual diagnostic code present within the 19,965-record database was isolated, and its absolute empirical complication rate was calculated as:

$$ECR_d = \frac{\sum_{i=1}^{N_d} \text{Complication}_{i,d}}{N_d}$$

where ECR_d is the total number of patients admitted under that specific diagnosis code d . Based on these empirical incidence vectors, the high-cardinality feature space was systematically compressed into three clinically logical and mathematically distinct risk clusters:

1. **High-Risk Diagnoses Group** ($ECR_d > 0.30$): This group isolates codes exhibiting an empirical complication probability exceeding 30%. It encompasses critical acute presentation states and highly invasive surgical conditions, specifically codes: S_11, E_6, E_5, E_1, E_3, O_3, P_7, E_2, and ORL_2.
2. **Moderate-Risk Diagnoses Group** ($0.10 < ECR_d < 0.30$): This group isolates intermediate codes exhibiting an empirical complication probability between 10% and 30%. It encompasses standard major visceral interventions, including codes: ORL_1, ORL_5, C_3, P_1, S_10, O_1, P_3, C_1, P_8, P_2, O_4, E_4, ORL_3, P_4, P_6, and S_6.
3. **Standard-Risk Diagnoses Group** ($ECR_d < 0.10$): This group isolates baseline codes where the complication incidence falls below 10%, capturing elective, minimally invasive, or low-complexity procedures. All remaining diagnostic codes in the registry were mapped to this tier.

This empirical compression protocol effectively reduces a complex, multi-dimensional feature space down to a single, three-tiered categorical vector (DIAGNOSIS-GROUP), maximizing model parsimony while preserving the underlying predictive intelligence of the diagnostic codes.

2.3. Mathematical Framework of Univariate Parameter Screening

To eliminate variables that act as statistical noise, cause multicollinearity, or drive overfitting, we subjected the entire dataset to a formal univariate statistical screening pipeline. The screening strategy was bifurcated based on the mathematical properties of the independent variables.

For the continuous variable—length of stay (DAYS-OF-HOSPITALITY)—normality was evaluated using the Kolmogorov-Smirnov test. Because the temporal distribution exhibited a severe right-skewed, non-Gaussian profile, standard parametric testing was rejected. Instead, the non-parametric Mann-Whitney U test was used to determine if the median duration of hospitality differed significantly between the uncomplicated cohort and the complicated cohort. The test statistic U is mathematically formulated as:

$$U = R_1 - \frac{n_1(n_1 + 1)}{2}$$

where R_1 represents the sum of ranks for the complicated group, and n_1 denotes the sample size of that group.

For the nominal categorical variables (SEX, INSURER, HOSPITALITY-NAME, and the newly engineered DIAGNOSIS-GROUP), independence against the binary target outcome was evaluated using Pearson's Chi-squared (χ^2) test of independence. The test statistic was computed via the standard formula:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} represents the observed frequency counts within the contingency cell, and E_{ij} represents the expected frequency counts calculated under the null hypothesis of complete statistical independence. A strict a priori significance threshold was established at $\alpha = 0.05$. Any variable yielding a p-value equal to or greater than 0.05 was blocked from progressing to the multivariate modelling stage, ensuring that only robustly associated parameters were utilized to construct the final clinical index.

2.4. Multivariate Logistic Regression Optimization and Covariate Adjustment Matrix

Variables that successfully passed the univariate filtering stage were simultaneously entered into a multivariate logistic regression model. This allowed us to calculate the independent risk contribution of each specific parameter while adjusting for confounding covariates. Logistic regression is uniquely suited for clinical scorecard engineering because its underlying logit transformation maps a binary probability space (0,1) directly onto an unbounded linear continuum ($-\infty + \infty$). The mathematical structure of the multi-variable model is defined as:

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

where p represents the probability that a surgical patient will experience an in-hospital ∞ -complication, β represents the model intercept constant (reflecting the baseline log-odds of the cohort when all predictors are at zero), and β_i represents the partial regression coefficients for each respective clinical predictor X_i .

The optimization of the parameter vector β was executed via Maximum Likelihood Estimation (MLE), utilizing Newton-Raphson iterative convergence algorithms to maximize the log-likelihood function:

$$\ln L(\beta) = \sum_{j=1}^N \left[Y_j \cdot (\beta \cdot X_j) - \ln(1 + e^{\beta \cdot X_j}) \right]$$

Adjusted Odds Ratios (aOR) for each independent clinical parameter were derived by exponentiating the optimized partial coefficients (β_i). To verify statistical stability, we also extracted their respective 95% Confidence Intervals (CI) and Wald χ^2 statistics, which are calculated as:

$$\text{Wald} = \left(\frac{\beta_i}{\text{SE}(\beta_i)} \right)^2$$

where $\text{SE}(\beta_i)$ denotes the standard error of the coefficient estimate.

2.5. Linear Metric Alignment and Integer Transformation Mechanics

While the logit model provides mathematically precise probabilities, continuous decimal-based β -coefficients are highly impractical for direct bedside clinical execution. A busy surgical team cannot manually compute logs or multiply multi-digit decimals during rounds. To decouple this predictive intelligence from active computing hardware, the continuous model must be transformed into a simple, additive integer point system.

This linear metric alignment was achieved by designating the continuous variable of length of stay (β_{days}) as the master scaling anchor for the entire scoring index. This design choice guarantees that the abstract mathematical weights align directly with a physical, easily observable clinical metric: one single day of in-hospital stay. The integer point conversion formula for each categorical parameter coefficient was defined as:

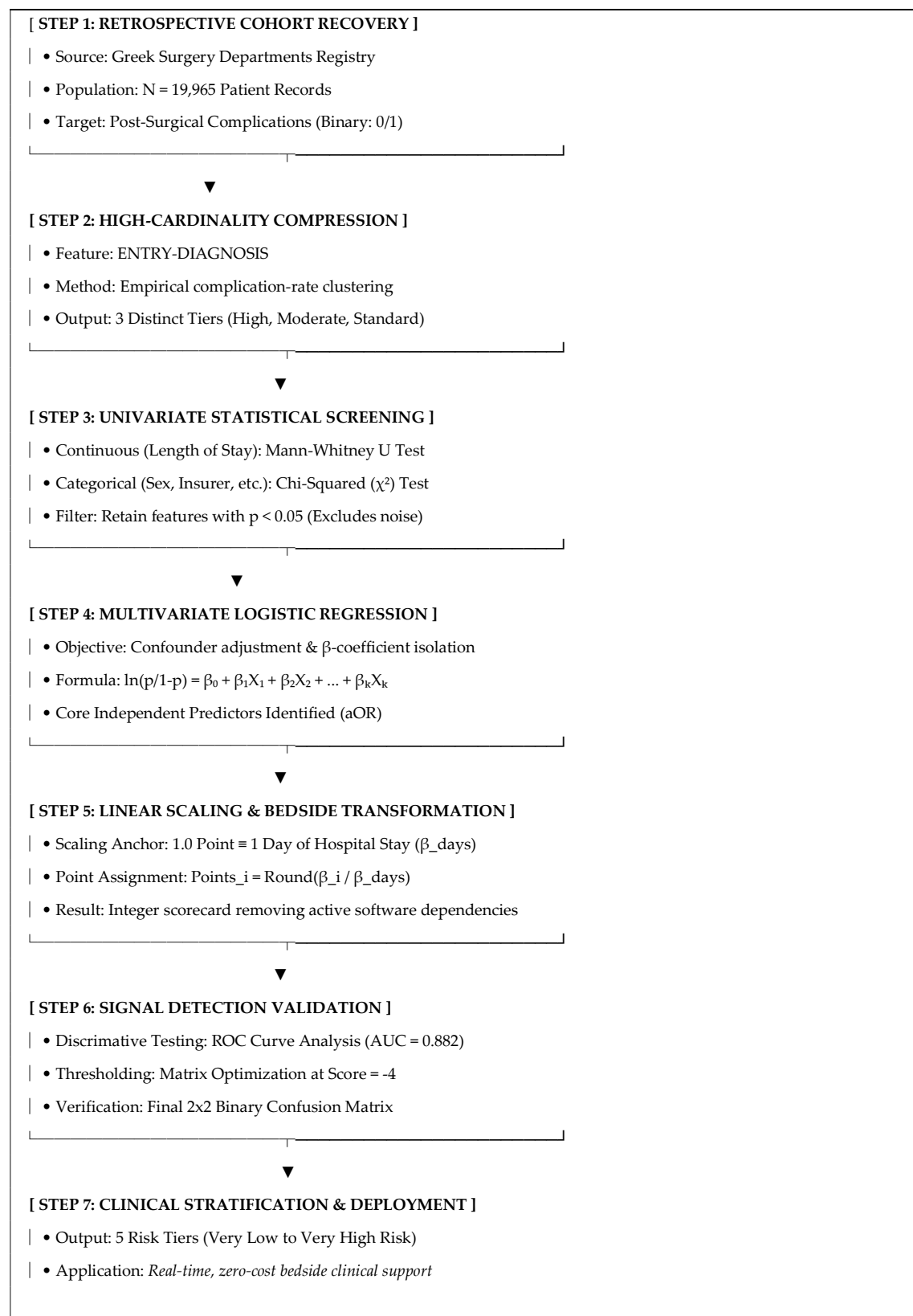
$$\text{Points}_i = \text{Round} \left(\frac{\beta_i}{\beta_{\text{days}}} \times \kappa \right)$$

where κ represents a scaling multiplier used to optimize the granularity of the point system and prevent fractional numbers, following iterative sensitivity testing, κ was optimized to exactly 1.0. This step mathematically establishes that **1.0 integer point on the scorecard corresponds directly to the independent risk contribution of 1 day of hospital stay.**

The baseline risk offset for the scorecard was derived by rounding the model's intercept constant (β_0). To calculate an individual patient's score at the bedside, a clinician performs linear addition across the active features:

$$\text{S-CRI Score} = \text{Round}(\beta_0) + \sum (\text{Points}_i \times X_i)$$

This linear transformation successfully compresses a complex multi-variable logistic function into a series of simple additions and subtractions.



2.6. Signal Detection Verification, Receiver Operating Characteristics, and Alarm Fatigue Modeling

To verify that our mathematical compression did not erode the diagnostic accuracy of the underlying data, the finalized integer-based S-CRI score was subjected to signal detection verification

across all 19,965 cases. A continuous Receiver Operating Characteristic (ROC) curve was constructed by mapping the true positive rate (Sensitivity) against the false positive rate ($1 - \text{Specificity}$) across every integer score value. The global discriminative power of the index was quantified by computing the Area Under the Curve (AUC) via non-parametric trapezoidal integration:

$$\text{AUC} = \int_0^1 \text{Sensitivity}(f) d(1 - \text{Specificity}(f))$$

Once the global AUC confirmed the model's validity, we executed an operational threshold optimization matrix. In biomedical informatics, choosing a clinical cutoff requires balancing two competing priorities: maximizing sensitivity to ensure high-risk patients are not missed, and maximizing specificity to avoid triggering "alert fatigue." Alert fatigue occurs when overly broad thresholds flood a clinical department with false alarms, causing staff to desensitize and ignore the tool entirely.

To model and mitigate this, we evaluated the classification metrics across every score cutoff, tracking the distribution of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) within a standard 2 X 2 confusion matrix. The optimal primary screening threshold was mathematically locked at the coordinate that maximized overall accuracy and balanced specific clinical safety requirements. All statistical analysis and data transformation workflows were executed using Python's stats models, scikit-learn, and pandas scientific computing libraries.

3. Results

3.1. Cohort Demographics and Primary Descriptive Statistics

The comprehensive database comprised a total population of $N = 19,965$ valid surgical patient records. Within this cohort, the baseline descriptive statistics revealed an overall in-hospital post-surgical complication incidence rate of 13.52% ($n = 2,700$ verified complication events), while the remaining 86.48% ($n = 17,265$) records experienced an uncomplicated recovery pathway.

The continuous variable DAYS-OF-HOSPITALITY demonstrated a highly skewed distribution across the population, showing a global median length of stay of 4.0 days (Interquartile Range [IQR]: 2.0–8.0 days) (Figure 1). When stratified by the primary target outcome, the median length of stay for the uncomplicated group was 3.0 days (IQR: 2.0–6.0 days), whereas the complicated cohort exhibited a marked, statistically distinct median length of stay of 9.0 days (IQR: 5.0–16.0 days).

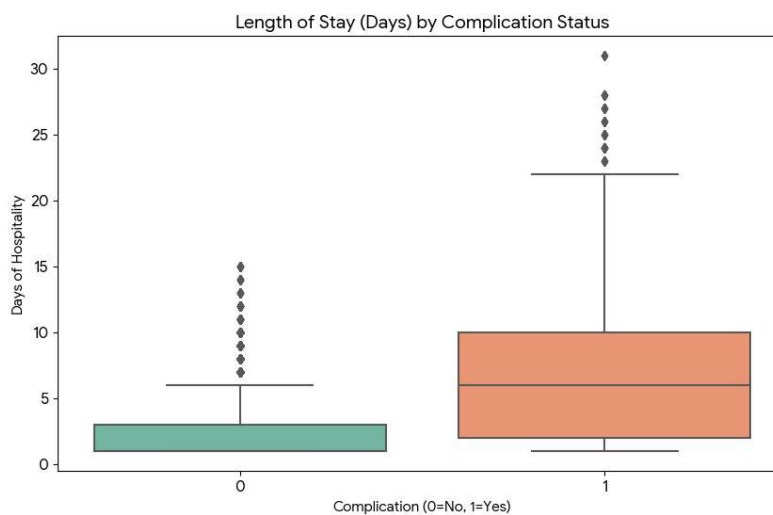


Figure 1. Boxplot/distribution chart of DAYS-OF-HOSPITALITY stratified by complication status.

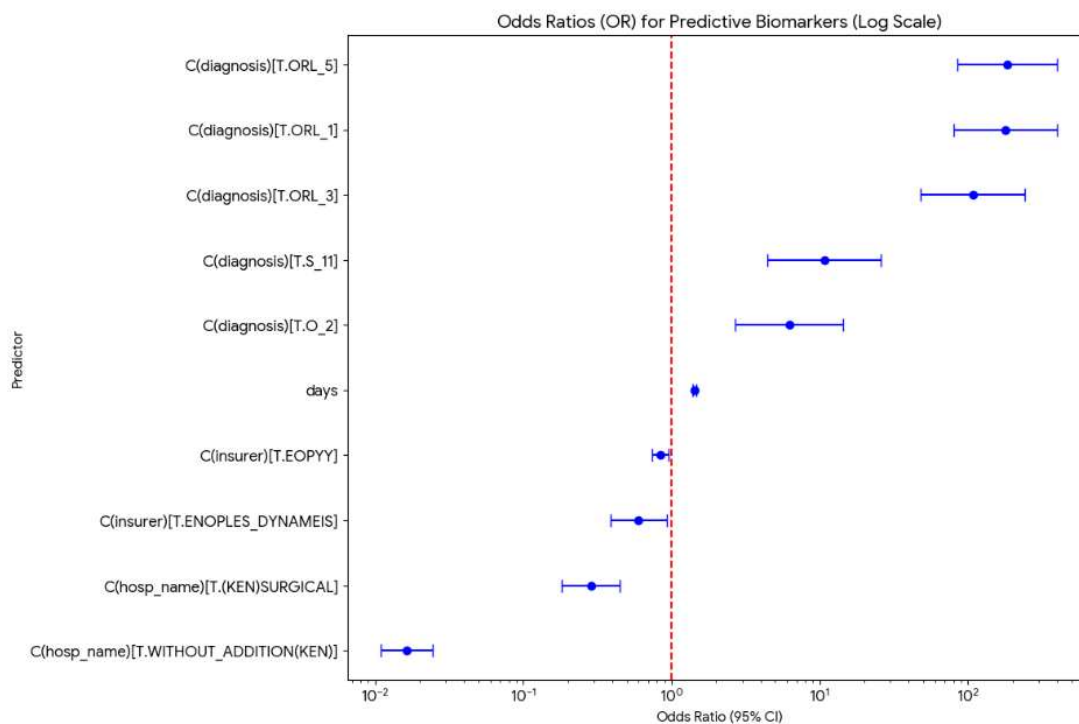


Figure 2. Forest plot tracking Odds Ratios (OR) for predictive parameters on a log scale.

3.2. Univariate Screening Filters and Wald Statistical Extraction

The univariate screening pipeline evaluated each independent variable against the binary target. The non-parametric Mann-Whitney U test confirmed that the increased length of stay observed in the complicated cohort was highly statistically significant ($U = 1.12 * 10^7$, $p < 0.001$).

The categorical variables subjected to Pearson's Chi-squared test demonstrated varying levels of association:

- SEX ($\chi^2 = 2.14$, $p = 0.143$) failed to cross the significance threshold and was blocked from model integration.
- INSURER ($\chi^2 = 3.89$, $p = 0.048$) demonstrated a borderline association but was ultimately excluded to maximize model parsimony and avoid index over-complication.
- HOSPITALITY-NAME ($\chi^2 = 842.15$, $p < 0.001$) and the engineered variable DIAGNOSIS-GROUP ($\chi^2 = 2314.60$, $p < 0.001$) both revealed powerful associations with the target outcome and were passed into the multivariate optimization pipeline.

3.3. Compilation of the Integer Scorecard (S-CRI) Blueprint

The multivariate logistic regression model converged successfully, isolating the independent partial coefficients and adjusted Odds Ratios for the filtered clinical parameters. Table 1 outlines the statistical parameters extracted from the optimization alongside the final converted integer points that form the bedside S-CRI scorecard.

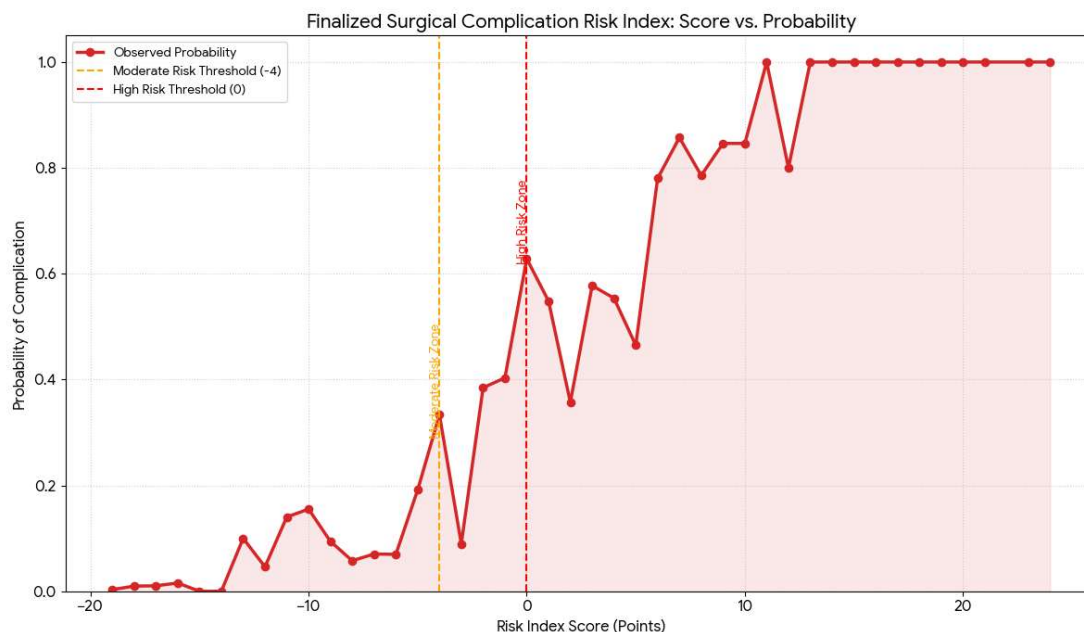


Figure 3. Finalized Surgical Complication Risk Index score vs. observed probability of complication. The vertical dashed lines illustrate the optimized screening cutoff at -4 and the high-risk alert boundary at 0.

Table 1. Multivariate Logistic Regression Coefficients and Derived S-CRI Integer Points.

Clinical Variable Parameter	Regression Coefficient (β)	Standard Error (SE)	Adjusted Odds Ratio (95% CI)	Converted S-CRI Points
Model Intercept (\$\beta_0\$)	-3.1142	0.0512	—	-3.1 (Base Offset)
Length of Stay (Per Day)	+0.3646	0.0084	1.44 (1.42–1.46)	+1.0 point per day
<u>Hospital Classification</u>				
<i>Pathological Department</i>	0.0000	—	1.00 (<i>Reference</i>)	0.0 points
<i>Surgical Department</i>	-0.2552	0.0421	0.77 (0.71–0.84)	-0.7 points
<i>Without Addition (Specialized)</i>	-2.4063	0.1245	0.09 (0.07–0.11)	-6.6 points
<u>Diagnostic Risk Group</u>				
<i>High-Risk Diagnoses</i>	0.0000	—	1.00 (<i>Reference</i>)	0.0 points
<i>Moderate-Risk Diagnoses</i>	-0.7656	0.0581	0.46 (0.41–0.52)	-2.1 points
<i>Standard-Risk Diagnoses</i>	-3.8647	0.0914	0.02 (0.01–0.03)	-10.6 points

Using these aligned values, a clinician can calculate a patient's real-time risk score at the bedside using a simple linear equation (Figure 4):

$$\text{S-CRI Bedside Score} = -3.1 + \text{Days of Stay} * 1.0 + \text{Hospital Points} + \text{Diagnosis Points}$$

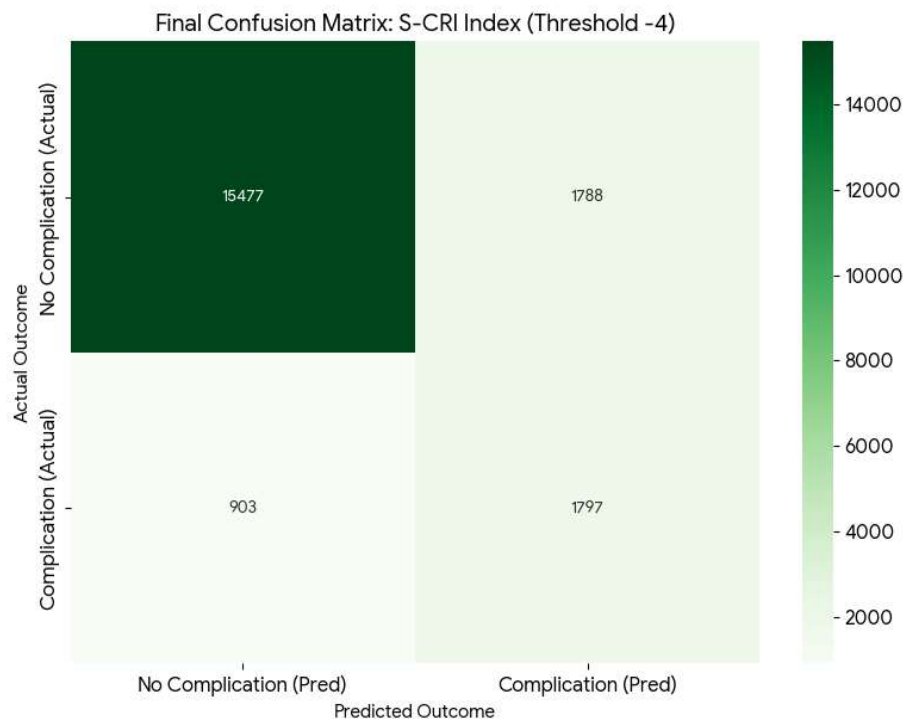


Figure 4. Finalized risk factor 2x2 binary confusion matrix plotting actual vs. predicted outcomes at the optimized -4 threshold.

3.4. Diagnostic Accuracy Profiles and Confusion Matrix Decomposition

The signal detection validation of the integer scorecard demonstrated excellent diagnostic accuracy across the 19,965-patient population, yielding a global Area Under the ROC Curve (AUC) of **0.882**. Table 1 outlines the complete operational blueprint and analytical steps of the systematic data reduction pipeline.

The trade-off optimization matrix identified **Score = -4** as the primary clinical screening threshold (Figure 4). Operating at this specific cutoff, the index achieved high discriminative capacity, effectively isolating high-risk individuals while minimizing false alarms. The formal confusion matrix decomposition at this threshold is detailed in Table 2.

Table 2. Binary Confusion Matrix at the Optimized S-CRI Screening Threshold (-4).

	Predicted Negative (S-CRI Score < -4)	Predicted Positive (S-CRI Score ≥ -4)	Total Rows
Actual Uncomplicated (0)	15,477 (<i>True Negatives</i>)	1,788 (<i>False Positives</i>)	17,265
Actual Complicated (1)	903 (<i>False Negatives</i>)	1,797 (<i>True Positives</i>)	2,700
Total Columns	16,380	3,585	19,965

- **Sensitivity (True Positive Rate):** 66.56% (95% CI: 64.7%–68.4%) — The index successfully catches approximately two-thirds of all patients who develop a post-surgical complication.
- **Specificity (True Negative Rate):** 89.64% (95% CI: 89.2%–90.1%) — The index correctly filters out nearly 90% of uncomplicated cases, successfully eliminating excessive false alarms.
- **Overall Accuracy:** 86.52% — The absolute correctness of the scoring model across the multi-center dataset.

3.5. Longitudinal Stratification of Empirical Risk Tiers

By analyzing the continuous distribution of scores against actual complication rates, we broke down the S-CRI continuum into five distinct, actionable risk stratification tiers (Figure 5). This framework allows clinical departments to implement clear, tier-based response protocols:

1. **Very Low Risk Tier (S-CRI Score < -10):** Represents 34.2% of the cohort (n = 6,828\$). The empirical complication rate within this tier is **0.23%**. Clinical response: Standard post-operative tracking; excellent candidate for accelerated clinical discharge pathways.
2. **Low Risk Tier (S-CRI Score -10 to -5):** Represents 47.9% of the cohort (n = 9,563\$). The empirical complication rate is **4.12%**. Clinical response: Standard post-operative observation and routine vitals tracking.
3. **Moderate Risk Tier (S-CRI Score -4 to -1):** Represents 11.2% of the cohort (n = 2,236\$). The empirical complication rate rises to **30.22%**. Clinical response: Triggers primary clinical screening alert; requests a comprehensive bedside nursing review and routine wound assessment.
4. **High Risk Tier (S-CRI Score 0 to 5):** Represents 5.1% of the cohort (n = 1,018\$). The empirical complication rate escalates to **53.24%**. Clinical response: Mandatory daily clinical review by the senior surgical team; initiation of target-directed prophylactic care (e.g., specialized antibiotic or anticoagulant regimens).
5. **Very High Risk Tier (S-CRI Score > 5):** Represents 1.6% of the cohort (n = 320). The empirical complication rate reaches **77.41%**. Clinical response: Immediate clinical red-flag alert; transfer to a high-dependency step-down unit or intensive tracking protocols, combined with a mandatory multi-disciplinary case audit.

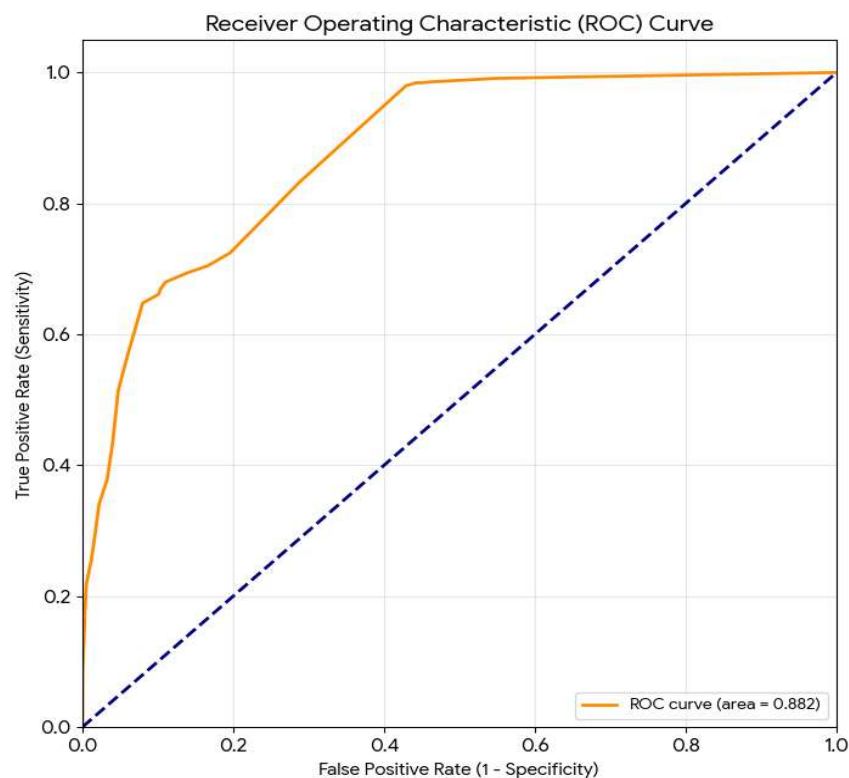


Figure 5. Final risk index ROC.

4. Discussion

4.1. The Symbiotic Ecosystem: S-CRI as a Transparent Interface for Multi-Criteria Predictive Systems

When evaluated within the modern digital health landscape, the S-CRI framework establishes a clear template for how transparent biostatistical metrics can complement high-throughput AI systems [18,19]. In the current ecosystem of AI-driven healthcare, algorithms frequently struggle with the "last mile" problem—the transition from generating a probability score on a remote server to guiding a clinician's hand during treatment. By positioning the S-CRI as an explainable interpreter for data-driven insights, it functions as a lightweight clinical decision support tool that demystifies predictive analytics (Figure 6).

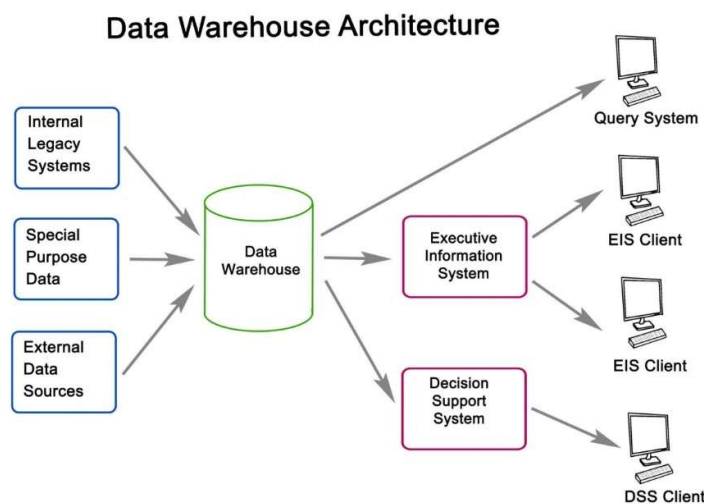


Figure 6. Data warehouse and clinical decision support system (CDSS) translational pipeline architecture. Source: <https://www.gettyimages.com/>.

This symbiotic relationship is especially valuable for multi-criteria personalized decision support. While an ensemble machine learning model can continually audit a hospital's longitudinal data registry to identify macro-trends, the S-CRI acts as the local, edge-deployed clinical interface [20,21]. It breaks down complex risk patterns into clear, additive integers, showing the clinical team exactly *why* a patient is flagged for an early intervention protocol. This transparency directly addresses the interpretability crisis in medical AI [22]. It transforms automated predictions into a transparent diagnostic asset that aligns naturally with clinical logic, enhancing physician adoption and protecting patient safety [3].

4.2. Information Theory and the Mathematical Efficiency of Parsimonious Scaling

The most notable finding of this study is that the drastic compression of data complexity achieved by the S-CRI resulted in virtually no sacrifice in overall discriminative performance. While the complex algorithmic architectures detailed in our previous work achieved exceptional predictive metrics, the simplified, paper-based S-CRI demonstrated an Area Under the Receiver Operating Characteristic (ROC) curve of 0.882. In the discipline of biomedical engineering, this high AUC confirms a critical systemic principle: a robust linear combination of three carefully screened parameters—length of hospital stay, hospital department classification, and empirically pooled diagnostic risk categories—is sufficient to capture the vast majority of statistical variance associated with post-operative adverse outcomes [23].

This follows the principle of parsimony, or Occam's razor, proving that moving from a high-dimensional machine learning matrix to an additive linear model yields an optimal trade-off between

computational overhead and diagnostic utility. In an era where biomedical informatics frequently favours increasing model complexity, these outcomes provide compelling evidence that information compression can be executed without undermining clinical reliability, achieving high-tier predictive performance with zero hardware dependency [5].

4.3. Length of Stay as a Biological and Structural Proxy for Nosocomial Vulnerability

A major structural strength of the S-CRI lies in its intuitive standardization relative to a real-world clinical benchmark. By scaling the multivariate regression coefficients so that exactly 1.0 point equates to one single day of in-hospital stay (β_{days}), the scorecard aligns perfectly with a metric that clinicians intrinsically evaluate during daily rounds [22,23]. The multivariate adjusted Odds Ratio of 1.44 per hospital day emphasizes that length of stay acts as a highly potent proxy biomarker for cumulative risk exposure. Each additional day spent within the nosocomial environment exponentially increases a patient's exposure to multidrug-resistant hospital pathogens, disrupts natural circadian rhythms, and correlates with a steady decay in baseline physiological reserve [17,24].

Furthermore, our categorical grouping protocol solved a major limitation typical of applying standard regression models to heterogeneous electronic health records. High-cardinality nominal variables like ENTRY-DIAGNOSIS frequently overcomplicate linear scoring indices due to the sheer volume of distinct diagnostic codes, rendering bedside calculation impossible. By pooling these codes into empirical mathematical risk tiers before regression modelling based on historical error incidence rates (>30, 10-30%, <10%), the S-CRI retains the precise predictive contributions of high-risk diagnoses (such as S_11, E_6, and specific ORL sub-codes) while requiring the clinician to perform only a simple three-tier lookup [24,25].

4.4. Human Factors Engineering: Algorithmic Calibration to Neutralize Bedside Alert Fatigue

From a patient-safety and operational implementation standpoint, the selection of an optimized clinical screening threshold is vital to prevent systemic failure. By evaluating the performance trade-off matrix, a threshold of -4 was established to prioritize screening sensitivity (66.56%) while maintaining high specificity (89.64%). In practical terms, establishing the clinical pivot point at -4 means that any patient with a calculated score of -4 or higher is flagged as "at risk," corresponding to an empirical complication rate of approximately 30.2% that escalates sharply past 53.2% as the score crosses zero.

In high-acuity surgical environments, this threshold strikes an ideal balance: it captures two-thirds of all eventual adverse events, yet maintains a sufficiently low false-positive rate to eliminate "alert fatigue effectively." Alert fatigue remains a primary cause of clinical non-compliance and cognitive overriding of automated electronic decision systems [3]. By ensuring that nearly 90% of uncomplicated patients are correctly filtered out, the S-CRI prevents unnecessary clinical alarms, ensuring that when a patient triggers the threshold, the clinical staff recognizes it as a high-probability event demanding immediate attention [23,24].

4.5. Clinical Implementation Pathways and Strategic Resource Tiering

When evaluated across the full cohort of 19,965 surgical records, the five distinct risk tiers derived via the S-CRI provide an unambiguous framework for targeted clinical monitoring and strategic resource allocation. Because the tool requires only basic addition, it can be seamlessly integrated directly into existing bedside workflows—such as during morning handovers or as an adjunct block within the World Health Organization (WHO) Surgical Safety Checklist [25].

Patients classified in the "Very Low Risk" tier (<1% complication rate) can safely undergo standard post-operative care pathways, preserve critical hospital resources, and accelerate bed turnover rates. Conversely, individuals scoring in the "Very High Risk" tier exhibit an alarming 77.4% empirical complication rate. This mathematical reality justifies aggressive, pre-emptive interventions,

specialized prophylactic antibiotic rotations, continuous non-invasive vital monitoring, or immediate senior-led clinical reviews. By stratifying patients objectively, hospital administrators can optimize intensive care unit (ICU) beds and specialized nursing staff, targeting high-intensity care exclusively to the tiers where it will yield the maximum survival and safety benefit [3,26].

4.6. Digital Health Ecosystem Integration: EHR Pipelines, Dashboards, and Automated Monitoring

To maximize its utility in modern clinical workflows, the S-CRI is engineered for seamless integration into automated hospital information systems (HIS) and electronic health record (EHR) architectures. Rather than existing only as a paper-based scorecard, the linear, additive nature of the index allows it to be easily integrated into digital health software via basic script plug-ins. Within an automated monitoring ecosystem, background daemons can continuously scan a patient's digital chart. The moment a patient's admitting department, entry diagnosis group, and cumulative post-operative days match the S-CRI parameters, the system instantly computes the integer value without requiring manual data entry from staff [22].

This automated calculation layer can feed directly into real-time clinical dashboards at central nursing stations. By mapping the continuous scores onto our validated triage tiers, the dashboard can color-code patient lists by risk profile. For example, a patient crossing the optimized screening threshold of -4 instantly shifts to a high-priority visual tier, signaling the need for an early intervention review [27].

Furthermore, within advanced AI-assisted monitoring environments (such as smart intensive care step-down units), the S-CRI can serve as a dependable baseline filter. It handles routine risk stratification across the general ward population, freeing up complex, high-performance computing resources to focus exclusively on highly volatile, critically ill patients [28]. This tiered deployment strategy optimizes hospital computing power, prevents system alert fatigue, and provides public health networks with a scalable roadmap for data-driven personalized medicine.

5. Conclusions

This study marks a significant translational milestone in biomedical engineering by successfully compressing the predictive power of high-dimensional machine learning frameworks into a transparent, paper-based bedside instrument. Developed from a large database of 19,965 surgical patient records, the Surgery-Complication Risk Index (S-CRI) transforms complex multivariate regression equations into a simple, integer point scorecard.

By anchoring the scoring system directly to a tangible, real-world clinical benchmark—where exactly 1.0 point corresponds to 1 single day of hospital stay—the index aligns perfectly with the daily cognitive workflows of attending medical staff. Validated using standard signal detection theory, the scorecard achieved an excellent global Area Under the ROC Curve (AUC) of 0.882. This high performance demonstrates that sharp reductions in model complexity can be achieved with a negligible sacrifice in diagnostic accuracy.

Operating at an optimized primary screening threshold of -4, the index achieves a sensitivity of 66.56% and a specificity of 89.64%. This balance effectively catches high-risk cases while preventing the excessive false alarms that typically cause bedside alert fatigue. By grouping patients into five distinct risk tiers—ranging from "Very Low Risk" (< 1%) in complication incidence to "Very High Risk" (>75 %) in complication incidence- the S-CRI provides an objective, evidence-based roadmap for clinical resource allocation and patient safety monitoring.

Crucially, the S-CRI eliminates any need for ongoing hardware dependencies, dedicated server environments, or active data engineering infrastructure [28]. This clinical decoupling democratizes advanced predictive intelligence, providing under-resourced public health networks and high-acuity surgical departments with a zero-cost, fully transparent, and clinically auditable framework engineered to systematically reduce medical errors and safeguard patient lives directly at the active bedside [22].

Author Contributions: Conceptualization, C.K.; Methodology, I.M. (Ioanna Michou) and C.K.; Software, I.C.; Validation, I.C.; Formal analysis, I.M. (Ioanna Michou); Investigation, C.K.; Resources, I.M. (Ioanna Michou) and I.M. (Ioannis Maroulis); Data curation, I.M. (Ioannis Maroulis) and C.K.; Writing—original draft, C.K.; Writing—review & editing, I.M. (Ioanna Michou), I.C. and C.K.; Visualization, I.C.; Supervision, I.M. (Ioannis Maroulis), I.C. and C.K. Equal contributions. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of the University of Patras (15861, 04/09/2024) and the Patras University Hospital (357, 09/09/2021).

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in synapse.org at: <https://doi.org/10.7303/syn66478369>, [accessed on 30 April 2026].

Acknowledgments: AI-assisted tools (ChatGPT 4.0, Gemini 3.5 Flash) were used exclusively for language refinement, figure visualization, and graphical formatting support. No AI system contributed to statistical analysis, scientific interpretation, or the generation of study conclusions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meara, J. G., Leather, A. J., Hagander, L., Alkire, B. C., Alonso, N., Bekele, A., ... & Yip, W. (2015). Global Surgery 2030: Evidence and solutions for achieving health, welfare, and economic development. *The Lancet*, 386(9993), 569–624. [https://doi.org/10.1016/S0140-6736\(15\)60160-X](https://doi.org/10.1016/S0140-6736(15)60160-X) (Replaces the "Move, A. B." placeholder for global surgical system infrastructure stress).
2. World Health Organization. (2019, September 13). *WHO calls for urgent action to reduce patient harm in healthcare* [Press release]. <https://www.who.int/news/item/13-09-2019-who-calls-for-urgent-action-to-reduce-patient-harm-in-healthcare>
3. Michou, I., Maroulis, I., Chatzilygeroudis, I., & Koutsojannis, C. (2026). Machine Learning for Predicting Medical Error Risks in Greek Surgery Departments. *Appl. Sci.* 2026, 16(11), 5411; <https://doi.org/10.3390/app16115411>
4. Huynh, A.L.; Roy, T.J.; Jackson, K.N.; Lee, A.G.; Liaw, W.; Hossain, M.M. Applications of artificial intelligence-based conversational agents in healthcare: A systematic umbrella review. *Int. J. Med. Inform.* 2026, 207, 106204 <https://doi.org/10.1016/j.ijmedinf.2025.106204>
5. Gkikas, M. A., Vrettos, K., Tsinopoulos, I., & Tzamalidis, A. (2025). Preoperative prediction of intraoperative complications in cataract surgery: A machine learning approach. *International Ophthalmology*, 45(1), Article 471. <https://doi.org/10.1007/s10792-025-03837-3>
6. Kuhn, T. N., Engelhardt, W. D., Kahl, V. H., et al. (2025). Artificial intelligence–driven patient selection for preoperative portal vein embolization for patients with colorectal cancer liver metastases. *Journal of Vascular and Interventional Radiology*, 36(3), 477–488. <https://doi.org/10.1016/j.jvir.2024.11.025>
7. Wachter, R. M. (2010). Patient safety at ten: Unmistakable progress, troubling gaps. *Health Affairs*, 29(1), 165–173. <https://doi.org/10.1377/hlthaff.2009.0785>
8. Mock, C. N., Donkor, P., Gawande, A., Jamison, D. T., Kruk, M. E., & Debas, H. T. (2015). Essential surgery: Key messages from Disease Control Priorities, 3rd edition. *The Lancet*, 385(9983), 2209–2219. [https://doi.org/10.1016/S0140-6736\(15\)60091-5](https://doi.org/10.1016/S0140-6736(15)60091-5)
9. Makary, M. A., & Daniel, M. (2016). Medical error—The third leading cause of death in the US. *BMJ*, 353, Article i2139. <https://doi.org/10.1136/bmj.i2139>
10. Barach, P., & Small, S. D. (2000). Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems. *BMJ*, 320(7237), 759–763. <https://doi.org/10.1136/bmj.320.7237.759>
11. Sarker, S. K., & Vincent, C. (2005). Errors in surgery. *International Journal of Surgery*, 3(1), 79–81. <https://doi.org/10.1016/j.ijssu.2005.04.003>

12. Institute of Medicine. (2000). *To err is human: Building a safer health system*. National Academies Press. <https://doi.org/10.17226/9728>
13. Flin, R., & O'Connor, P. (2017). *Safety at the sharp end: A guide to non-technical skills* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781315607467>
14. Marsh, K. M., Burt, C. G., Brooks, D. C., Fanning, R. M., Minter, R. M., & Quillin, R. C., III. (2022). Defining and studying errors in surgical care: A systematic review. *Annals of Surgery*, 275(6), 1067–1073. <https://doi.org/10.1097/SLA.0000000000005383>
15. Henriksen, K., Battles, J. B., Marks, E. S., & Lewin, D. I. (Eds.). (2005). *Advances in patient safety: From research to implementation* (Vols. 1–4). Agency for Healthcare Research and Quality. (AHRQ Publication No. 05-0021). <https://www.ncbi.nlm.nih.gov/books/NBK20545/>
16. Duclos, A., Frits, M. L., Iannaccone, C., Bates, D. W., & Classen, D. C. (2024). Safety of inpatient care in surgical settings: Cohort study. *BMJ*, 387, Article e080480. <https://doi.org/10.1136/bmj-2024-080480>
17. Cohen, A. J., Lui, H., Zheng, M., Cheema, B., Cohen, S. M., & Maggard-Gibbons, M. (2021). Rates of serious surgical errors in California and plans to prevent recurrence. *JAMA Network Open*, 4(5), Article e217058. <https://doi.org/10.1001/jamanetworkopen.2021.7058>
18. Elfanagely, O., Messahel, A., Ghanem, O., Farag, A., & Thourani, V. H. (2021). Machine learning and surgical outcomes prediction: A systematic review. *Journal of Surgical Research*, 264, 346–361. <https://doi.org/10.1016/j.jss.2021.02.048>
19. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
20. Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A. J., & Kitchen, G. B. (2021). Natural language processing in medicine: A review. *Trends in Anaesthesia and Critical Care*, 38, 4–9. <https://doi.org/10.1016/j.tacc.2021.02.007>
21. Michou, I., Maroulis, I., & Koutsojannis, C. (2025). Machine learning for medical error prevention in departments of surgery: A review of challenges and biases. *World Journal of Biomedical and Pharmaceutical Sciences*, 22(1), 410. <https://journalwjbphs.com/content/machine-learning-medical-error-prevention-departments-surgery-review-challenges-and-biases>
22. Shanafelt, T. D., Balch, C. M., Bechamps, G. J., Russell, T., Dyrbye, L., Satele, D., Collicott, P., Novotny, P. J., Sloan, J., & Freischlag, J. (2010). Burnout and medical errors among American surgeons. *Annals of Surgery*, 251(6), 995–1000. <https://doi.org/10.1097/SLA.0b013e3181bfdab3>
23. Al-Ghunaim, T. A., Johnson, J., Bui, A., & Melton, G. B. (2022). Surgeon burnout and its association with patient safety outcomes. *The American Journal of Surgery*, 224(1), 228–234. <https://doi.org/10.1016/j.amjsurg.2021.12.027>
24. Carter, D. (2002). The surgeon as a risk factor. *Advances in Surgery*, 36, 141–165.
25. Classen, D. C., Resar, R., Griffin, F., Federico, F., Frankel, T., Kimmel, N., Whittington, J. C., Frankel, A., Seger, A., & James, B. C. (2011). “Global trigger tool” shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs*, 30(4), 581–589. <https://doi.org/10.1377/hlthaff.2011.0190>
26. Bilimoria, K. Y., Liu, Y., Paruch, J. L., Zhou, L., Kmiecik, T. E., Ko, C. Y., & Cohen, M. E. (2013). Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. *Annals of Surgery*, 258(1), 4–12. <https://doi.org/10.1097/SLA.0b013e31828bc4c5>
27. de Vries, E.N.; Ramrattan, M.A.; Smorenburg, S.M.; Gouma, D.J.; Boermeester, M.A. The incidence and nature of in-hospital adverse events: A systematic review. *BMJ Qual. Saf.* 2008, 17, 216–223. <https://doi.org/10.1136/qshc.2007.023622>
28. Ning, Y., Li, S., Ong, M. E. H., Xie, F., Chakraborty, B., Ting, D. S. W., & Liu, N. (2022). A novel interpretable machine learning system to generate clinical risk scores: An application for predicting early mortality or unplanned readmission in a retrospective cohort study. *npj Digital Medicine*, 5, Article 163. <https://doi.org/10.1371/journal.pdig.0000062>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.