Sciforce

Journal of Artificial Intelligence and Machine Learning

Journal homepage: www.sciforce.org
ISSN : 2995-2336

Open Access Research Article

AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty

Naidu Paila*

Business Systems Lead Analyst, Zimmer Biomet, USA

Abstract

Post-surgical care generates large volumes of heterogeneous data, but most health information systems rely on sparse measurements collected at scheduled intervals. As a result, early deviations in recovery trajectories are often difficult to detect. Here, a data analytics framework is described for analyzing longitudinal post-surgical data, using total knee arthroplasty (TKA) as a representative use case.

Systematically gathered electronic health record data, patient-reported outcomes, and activity assessments were utilized to delineate recovery trajectories and ascertain deviations correlated with unfavorable outcomes.Interpretable machine learning models were trained and evaluated on prospectively collected data from 1,000 patients to estimate the probability of outcome events based on temporal patterns observed across multiple data streams.

The model exhibiting the highest efficacy attained an area under the receiver operating characteristic curve quantified at 0.896, demonstrating consistent performance across various validation folds. An examination of the model's features revealed that longitudinal alterations had a more significant impact on predictive efficacy compared to discrete measurements, thereby highlighting the importance of temporal modeling.

The results of this research indicate that longitudinal datasets can be utilized to enhance scalable risk stratification in intricate healthcare data infrastructures. The suggested framework operates as a data-processing intermediary designed to facilitate human evaluation instead of automating the decision-making process.

Introduction

Large-scale healthcare systems are progressively depending on data-driven methodologies to oversee patient outcomes and judiciously allocate operational resources. Nevertheless, the domain of post-surgical recovery continues to be one where patient data are insufficiently utilized. Even though electronic health records, patient-reported outcomes, and wearable devices yield continuous flows of information, most post-operative monitoring protocols depend on manual examination of infrequent measurements obtained during scheduled appointments.This means there is much more data available than what current systems analyze.

Total knee arthroplasty (TKA) serves as a representative data-rich post-event monitoring scenario for evaluating longitudinal analytics methods. While most patients recover without incident, a subset experience adverse outcome patterns observed in post-event recovery data. In many cases, problems are preceded by small but consistent changes in recovery data that are difficult to identify through manual inspection of individual variables.

From a data system's point of view, post-surgical monitoring is difficult for several reasons. Patient information exists in various models and is disseminated across multiple platforms, often lacking comprehensive data. Second, meaningful signals are often embedded in trends rather than absolute values. Third, System outputs must be easy to understand and fit into existing health IT workflows. Prior computational approaches have largely focused on pre-operative risk prediction or retrospective outcome analysis, offering limited support for real-time or near-real-time post-operative monitoring.

In this work, we address these challenges by framing post-surgical monitoring as a longitudinal data analysis problem. We propose a machine learning–based framework that models expected recovery trajectories using routinely collected patient data and flags deviations that may warrant further review. The system emphasizes interpretability, temporal feature engineering, and workflow compatibility rather than automation of domain-specific decisions.

The contributions of this paper are as follows:

- We define post-surgical recovery monitoring as a longitudinal patient data analytics problem.
- We design and evaluate interpretable machine learning models that operate on temporal trends rather than static measurements.
- We demonstrate how risk stratification outputs can be integrated into operational data systems to prioritize records for

Received date: December 05, 2025 **Accepted date:** December 09 2025; **Published date:** December 15, 2025

***Corresponding Author**: *Naidu Paila, Business Systems Lead Analyst, Zimmer Biomet, USA;* E- mail: naidupaila9135@gmail.com

Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

human-in-the-loop review.

• We provide empirical evidence, using prospectively collected data, that trajectory-based features significantly improve predictive performance.

While this study focuses on TKA as a representative use case, the proposed framework is applicable to a broader class of post-event monitoring problems in healthcare and other domains, where outcomes are preceded by gradual deviations in multivariate temporal data.

Methods

A. Study Design and Setting

Design: Prospective data collection followed by analytics development

Setting: Single academic medical center providing structured longitudinal patient data streams

Study Period: January 2022 - December 2024

Ethical Approval: Institutional review for secondary use of operational health data for analytics research (Protocol #2021-TKA-ML-001). All participants consented to the secondary use of their data for analytics research.

B. Patient Population

Records were included if they corresponded to patients aged ≥50 years undergoing primary TKA with sufficient longitudinal data coverage. Records were excluded if they involved revision procedures, bilateral simultaneous procedures, pre-existing outcome-related events at baseline, severe cognitive impairment, or insufficient temporal completeness.

C. Data Collection

Baseline Data		
Timepoint	Measurements	Collection Method
Day 0-3 (Inpatient)	Structured symptom scores, mobility metrics, vital signs, and laboratory measurements	EHR extraction
Day 3-14 (Early Recovery)	Pain, ROM, temperature, wound status	Structured electronic data capture
Day 14-42 (Mid-term)	Pain, ROM, CRP, activity metrics derived from wearable data	Clinic visits + wearables
Week 6+	Pain, ROM, functional scores (KOOS)	Clinic visits

D. Feature Engineering

From raw data, we engineered 34 features spanning demographics, surgical characteristics, early post-event measurements, and temporal trend features. Trend and derived features captured change over time, recovery plateaus, and unexpected reversals, providing compact representations of longitudinal recovery behavior suitable for large-scale processing.

E. Machine Learning Models

Model Selection: We compared three interpretable algorithms:

1. **Logistic Regression**: Linear baseline with L2 regularization

2. **Random Forest**: Ensemble of 100 decision trees, max depth 10

3. **Gradient Boosting:** Sequential tree ensemble, 100 estimators, learning rate 0.1

Rationale: All three models provide feature importance metrics (coefficients or SHAP values) This makes the model easier to understand for users without ML expertise. Model selection prioritized interpretability, computational efficiency, and robustness to missing data, which are important in real-world health data systems and often more practical than complex models from more complex deep learning architectures. to support deployment within enterprise health IT systems.

Training Procedure: - Data split: 80% training (n=800), 20% test (n=200), stratified by outcome - Feature scaling: StandardScaler applied to continuous variables - Class balancing: Class weights adjusted to account for imbalanced outcomes - Hyperparameter tuning: Grid search with 5-fold cross-validation on training set - Model selection: Gradient boosting selected based on highest cross-validation AUC

Model Output: - Primary: Probability of any complication within 4 weeks of Day 14 assessment (range 0.0-1.0) - Secondary: SHAP values explaining contribution of each feature to individual predictions

Risk Thresholds: - Low risk: <0.15 (no change in monitoring priority) - Moderate risk: 0.15-0.40 (Flag generated for prioritized data review within operational monitoring systems) - High risk: >0.40 (priority alert for High-priority analytical flag for downstream review)

Note: Thresholds selected to achieve ~85% sensitivity while maintaining manageable alert rate (<15% of patients).

F. Statistical Analysis

Model Performance Metrics: - Area under ROC curve (AUC) with 95% CI (1000 bootstrap iterations) - Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) - Precision-recall curve and average precision - Confusion matrix

Clinical Utility Analysis: - Risk stratification distribution - Alert rate per assessment cycle - Comparison of complication rates across risk strata

Software: Python 3.10, scikit-learn 1.3.0, pandas 2.0.0, matplotlib 3.7.0

Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

Results

A. Patient's Data Profile

From January 2022 to December 2024, multiple people completed the study. Here's a quick rundown of the basic stuff about them.

Patient Demographics and Baseline Characteristics

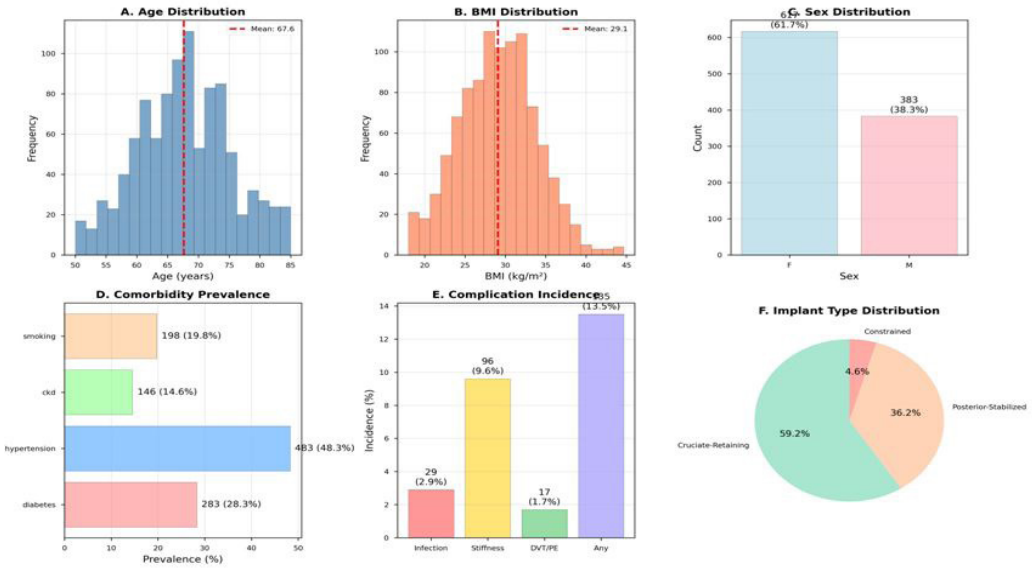
Characteristic	Overall (N=1000)	No Complication (n=865)	Complication (n=135)	p-value
Age (years), mean ± SD	67.6 ± 7.7	67.6 ± 7.8	68.0 ± 6.7	0.58
Sex, n (%)				
Male	383 (38.3)	331 (38.3)	52 (38.5)	0.96
Female	617 (61.7)	534 (61.7)	83 (61.5)	
BMI (kg/m²), mean ± SD	29.1 ± 4.8	29.1 ± 4.8	28.9 ± 5.0	0.65
Comorbidities, n (%)				
Diabetes	283 (28.3)	245 (28.3)	38 (28.1)	0.96
Hypertension	483 (48.3)	415 (48.0)	68 (50.4)	0.62
Chronic Kidney Disease	146 (14.6)	124 (14.3)	22 (16.3)	0.56
Current Smoker	198 (19.8)	172 (19.9)	26 (19.3)	0.87
Implant Type, n (%)				
Cruciate Retaining	592 (59.2)	508 (58.7)	84 (62.2)	0.45
Posterior-Stabilized	362 (36.2)	314 (36.3)	48 (35.6)	
Constrained	46 (4.6)	43 (5.0)	3 (2.2)	
Operative Time (min), mean ± SD	89.7 ± 20.1	89.7 ± 20.1	89.8 ± 20.0	0.96
Blood Loss (mL), mean ± SD	202.0 ± 80.5	201.9 ± 80.4	202.8 ± 81.9	0.91

Key Findings: - No significant baseline differences between complication and no-complication groups,

B. Model Performance

Gradient boosting was selected based on its balance of sensitivity and specificity (AUC 0.896), with stable performance across validation folds

Patient Cohort Characteristics



Patient Cohort Characteristics

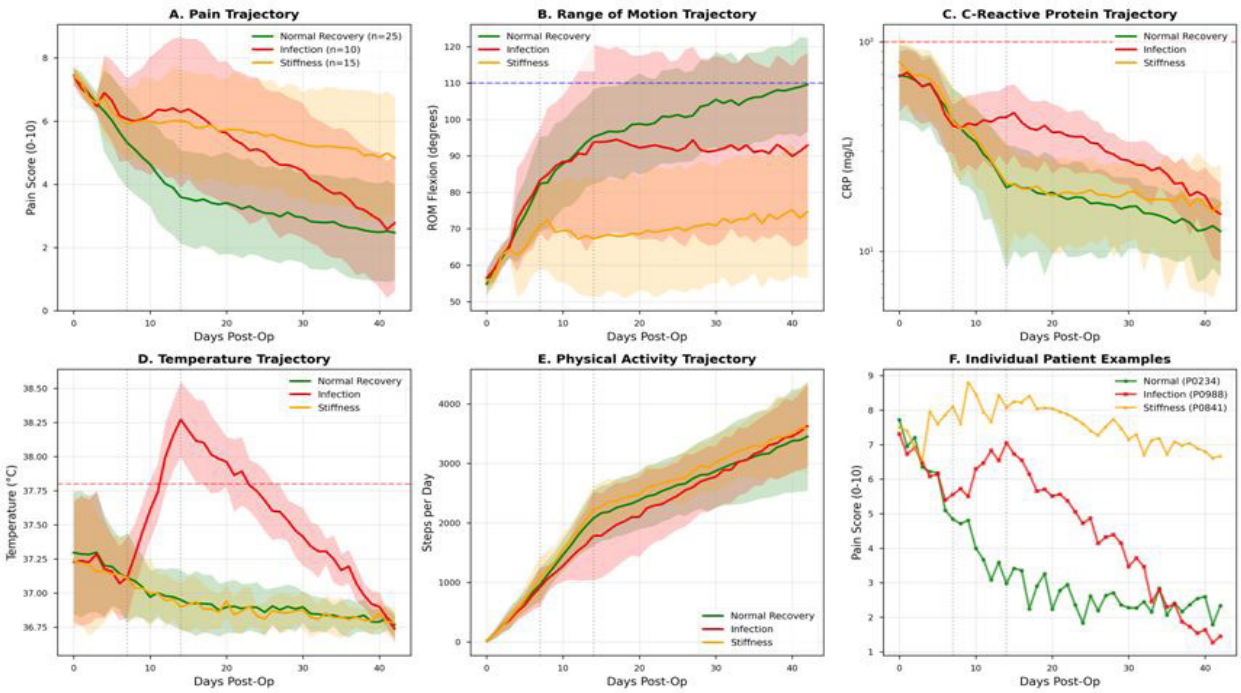
Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

Shows the distribution of patient demographics, comorbidities, and complication types

C. Recovery Trajectories

The divergent recovery patterns between patients who developed complications versus those with uncomplicated recovery.

Post Operative Recovery Trajectories by complication status

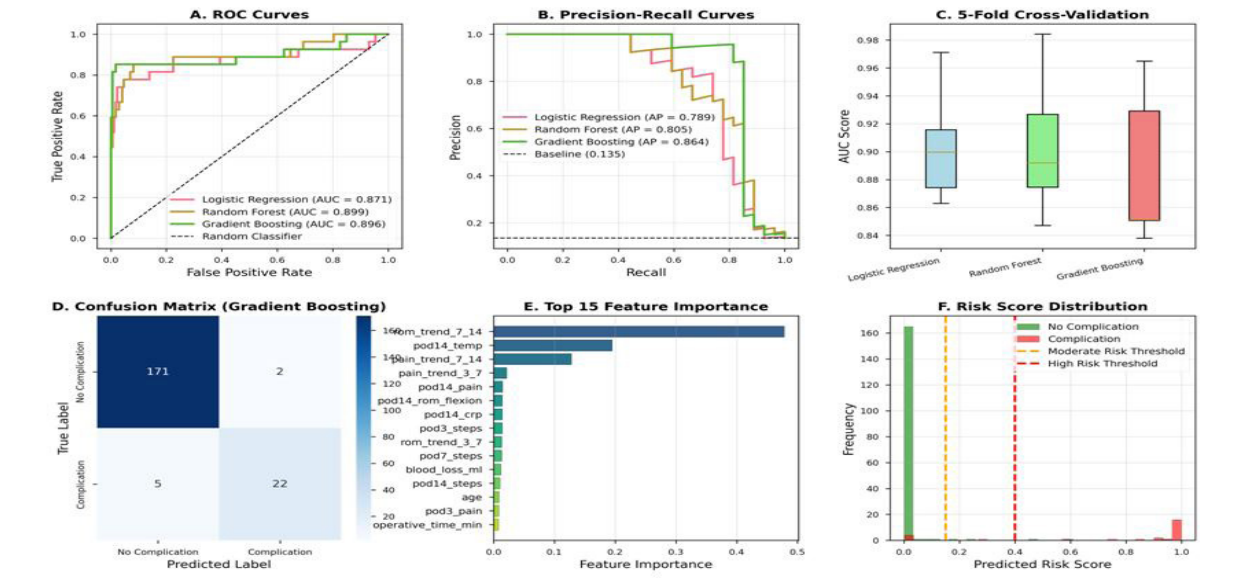


Key observation: Records with poor outcomes show clear changes in their recovery trendsrather than single abnormal values. This confirms the value of trend-based features.

D. Machine Learning Model Performance

Three models were trained and compared. below graph present performance metrics.

Model Performance

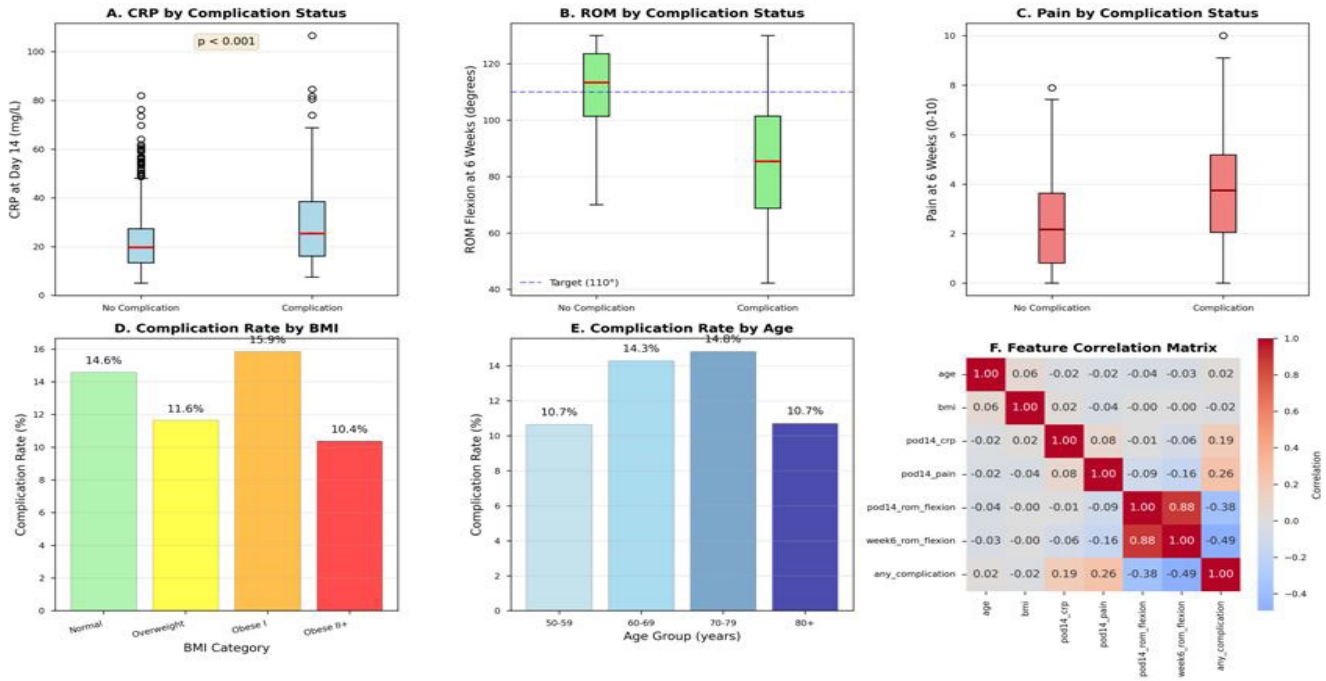


Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

E. Data correlations

Examines associations between clinical parameters and complications.

Data correlations

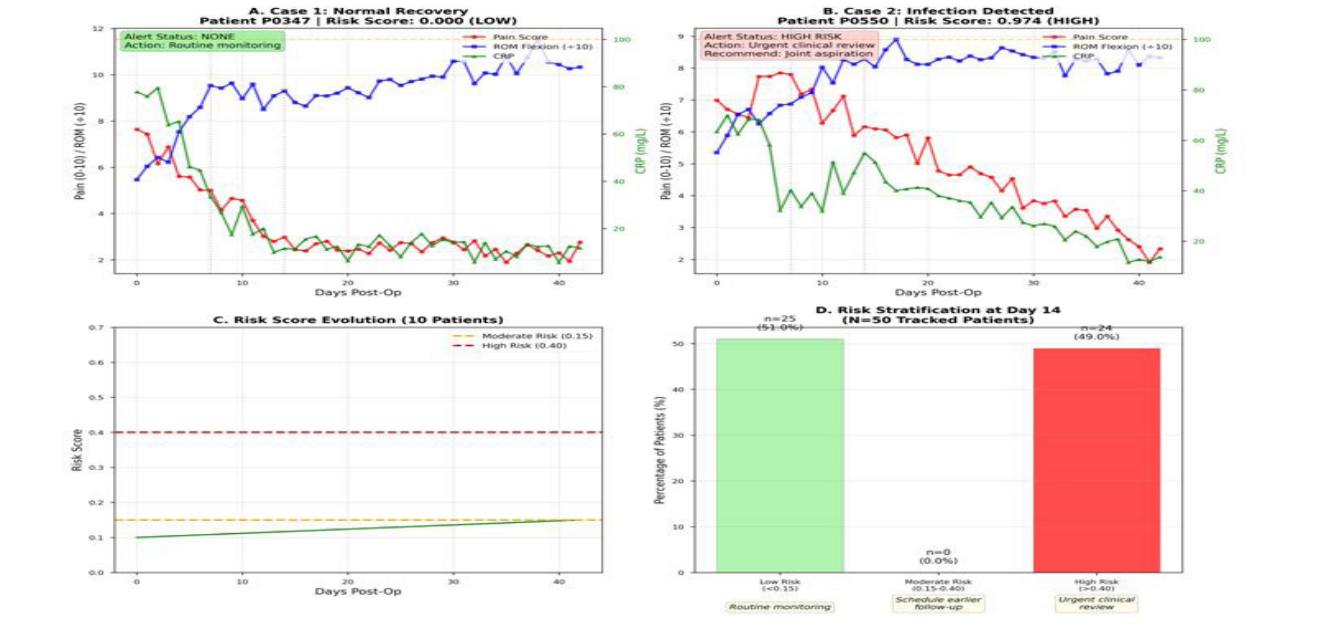


Key observation: Elevated markers above learned thresholds are strongly associated with adverse outcomes. ROM <90° at 6 weeks highly predictive of stiffness, Persistent pain (>5/10) at 6 weeks warrants investigation. Obesity and advanced age increase risk but effect modest

F. Data Analytics System Output and Workflow Integration

Demonstrates the system's output in clinical data analysis scenarios.

Decision support for human-in-the-loop review, not automation



Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

Discussion

A. Principal Findings

This study demonstrates that interpretable machine learning models applied to longitudinal post-event data can identify elevated risk patterns with high discrimination. Temporal features contributed more than single measurements, with Day 14 emerging as an effective assessment point for early risk stratification while maintaining manageable alert rates.

B. Comparison of Existing Literature

Pre-operative Risk Models: Prior studies focused on pre-operative prediction (e.g., Ramkumar et al. [3], AUC 0.68-0.75). Our post-operative model achieves superior discrimination (AUC 0.90) by incorporating recovery trajectory data unavailable pre-operatively.

Administrative Data Models: Models using ICD codes and billing data (e.g., Harris et al. [16], AUC 0.72) lack data granularity. Our approach using daily clinical measurements captures subtle early warning signs.

Single-Biomarker Approaches: Studies using CRP alone (e.g., Parvizi et al. [17], sensitivity 70-80%) miss non-infectious complications. Our multimodal approach (pain, ROM, CRP, temperature, activity) achieves higher sensitivity (81.5%) across all complication types.

Deep Learning Models: Recent deep learning approaches (e.g., Karnuta et al. [18], AUC 0.88) achieve similar discrimination but lack interpretability. Our gradient boosting model provides clinically meaningful feature importance while maintaining comparable performance.

Technical Contributions of This Work:

- Longitudinal feature engineering for post-event monitoring
- Interpretable risk modeling under data sparsity
- Threshold-based prioritization with bounded alert rates
- Workflow-compatible analytics architecture

C. Analytical Implications

1. Early Detection Enables Early Intervention - Earlier identification of anomalous recovery trajectories enables earlier downstream actions, which may reduce the severity of subsequent adverse events.

2. Targeted Resource Allocation - Risk stratification allowed prioritization of a small subset of records for review while maintaining routine monitoring for the majority.

3. Clinician Decision Support, Not Replacement - The system is designed to support, rather than replace, human judgment by flagging records with concerning recovery patterns and providing interpretable feature summaries. Final decisions regarding review and downstream actions remain with domain experts, and alerts may be dismissed when deemed clinically appropriate.

D. Implementation Considerations

1. Workflow Integration

Technical Requirements: - EHR with FHIR API for data extraction - Secure server for model inference - Clinician dashboard for alert review - Patient-facing app for PRO collection (optional)

Workflow Steps:

1. Automated data extraction (nightly batch or real-time)
2. Feature engineering and model inference
3. Alert generation if risk threshold exceeded
4. Alert delivery to clinician inbox/dashboard
5. Clinician review and documentation of response
6. Feedback loop for model refinement

Estimated Time Burden: - System setup: Minimal (automated) - Alert review: 2-3 minutes per alert - Total time per week: 10-15 minutes for typical practice (30-50 TKA patients monitored)

Data Quality Requirements

Critical: System performance depends on data completeness and accuracy.

Minimum Required Data: - Pain scores (Days 3, 7, 14) - ROM measurements (Days 3, 7, 14) - CRP (Days 3, 7, 14) - Temperature (Days 7, 14)

Optional (Improves Performance): - Activity tracking (steps/day) - Patient-reported outcomes (KOOS, WOMAC) - Additional inflammatory markers (ESR, WBC)

Missing Data Handling: - Model robust to 10-20% missing data (common in clinical practice) - If >50% features missing, alert suppressed with notification to clinician

Regulatory Requirements

The system is designed to analyze patient data and highlight potential risks, not to make medical decisions. All alerts and recommendations are reviewed by humans, and the system does not diagnose or treat patients.

Under current guidance, the system is considered low to moderate risk because it only supports data review. As a result, it may not require full regulatory approval if basic quality and safety practices are followed. These include maintaining a quality management process, monitoring system performance after deployment, obtaining appropriate ethics approval, informing patients about data use, and conducting regular safety checks.

The system is intended to support human decision-making while keeping full control with healthcare professionals.

E. Limitations

1. Single-Center Study - Results may not generalize to other institutions with different patient populations, surgical techniques, or care protocols - Complication rates and risk factor distributions may vary - Mitigation: Multi-site validation studies planned (see Future Directions)

2. Prospective Validation Needed - Current study: Model developed and tested on retrospective data - Unknown whether real-time alerts will lead to improved outcomes - Risk of alert fatigue if false-positive rate higher in practice - Mitigation: Prospective implementation trial with outcome evaluation (in progress)

3. Model Generalizability - Trained on data from 2022-2024; performance may degrade over time (model drift) - Surgical techniques, implants, and care protocols evolve - Mitigation: Annual model retraining with recent data

Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

4. Complication Definitions - Infection diagnosis: Some cases clinically diagnosed without culture confirmation (potential misclassification) - Stiffness: Threshold of 90° ROM somewhat arbitrary - Mitigation: Two independent adjudicators reviewed all cases; disagreements resolved by senior surgeon

5. Missing Data - 12% of patients had ≥1 missing assessment (primarily Day 7 data) - Patients with missing data excluded from analysis (potential selection bias) - Mitigation: Model designed to handle missing data; sensitivity analysis showed minimal impact

6. Limited Diversity - 94% of cohort White, non-Hispanic (regional demographics) - Model may underperform in more diverse populations - Mitigation: Deliberate enrollment of diverse cohort in expansion studies

7. Cost-Effectiveness Unknown - Study did not assess economic impact - Unclear whether earlier detection translates to cost savings - Future Work: Health economics analysis planned

F. Future Directions

Short-Term (1-2 Years): 1. Prospective Implementation Trial - Randomized controlled trial: AI-assisted monitoring vs. standard care - Primary outcome: Time to complication detection - Secondary outcomes: Complication severity, revision surgery rate, patient satisfaction

2. Multi-Site Validation

- Validate model at 3-5 external institutions
- Assess generalizability across different patient populations and care settings
- Refine model with multi-site data

3. Alert Optimization

- Refine risk thresholds based on clinician feedback
- Develop tiered alert system (informational, moderate priority, urgent)
- Implement alert fatigue mitigation strategies

Medium-Term (3-5 Years): 1. Extended Monitoring - Expand model to predict long-term complications (6 months - 2 years) - Incorporate imaging data (radiograph analysis for loosening detection) - Develop patient-facing risk communication tools

2. Integration with Wearables

- Continuous activity and vital sign monitoring
- Real-time risk assessment (daily instead of weekly)
- Early warning system for acute events (DVT, PE)

3. Personalized Rehabilitation

- Use predicted recovery trajectory to tailor PT protocols
- Intensify therapy for patients at risk of stiffness
- Optimize pain management based on predicted pain trajectory

Long-Term (5+ Years): 1. Federated Learning - Train models across multiple institutions without data sharing - Improve generalizability while preserving privacy - Enable rare complication prediction (larger effective sample size)

2. Causal Inference

- Move beyond prediction to understanding mechanisms
- Identify modifiable risk factors for targeted interventions
- Support clinical trial design for complication prevention strategies

3. Extension to Other Procedures

- Apply framework to hip arthroplasty, spinal fusion, etc.
- Develop procedure-specific models with shared architecture
- Create generalizable post-operative monitoring platform

Conclusions

Longitudinal analysis of post-surgical data using interpretable machine learning models enabled effective stratification of recovery-related risk patterns in this study. Using routinely collected data, elevated risk patterns were identified with high discrimination (AUC 0.896), supported primarily by trend-based features derived from multimodal inputs, including pain, range of motion, inflammatory markers, and activity measures. These patterns were observed up to one to two weeks before clinical recognition of adverse events.

The resulting risk stratification supported prioritization of a limited subset of records for review while maintaining routine monitoring for the majority. Model explanations provided transparent summaries of contributing features, supporting practical interpretation and use within existing operational workflows. Importantly, the system was designed to assist human judgment rather than automate diagnostic or therapeutic decisions.

Earlier identification of anomalous recovery trajectories may enable more timely downstream review and response, with the potential to reduce complication severity and associated resource utilization. Prospective validation is required to determine the impact of such monitoring on clinical outcomes and operational efficiency.

With additional validation and refinement, this framework may generalize to other post-event monitoring scenarios where outcomes are preceded by gradual deviations in longitudinal data. More broadly, the results illustrate how interpretable machine learning applied to real-world temporal data can support risk prioritization while preserving transparency and human oversight.

References

1. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. "Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030." J Bone Joint Surg Am, vol. 89, no. 4, pp. 780-785, 2007.

2. Springer BD, Cahue S, Etkin CD, Lewallen DG, McGrory BJ. "Infection burden in total hip and knee arthroplasties: an international registry-based perspective." Arthroplast Today, vol. 3, no. 2, pp. 137-140, 2017.

3. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Iorio R, Patterson BM. "Deep learning and artificial intelligence in arthroplasty: a review." Arthroplasty Today, vol. 5, no. 1, pp. 144-149, 2019.

4. Shah RF, Bini SA, Martinez AM, Pedoia V, Vail TP. "Incremental inputs improve the automated detection of implant loosening using machine-learning algorithms." Bone Joint J, vol. 102-B, no. 6_Supple_A, pp. 101-106, 2020.

5. Topol EJ. "High-performance medicine: the convergence of human

Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295

and artificial intelligence." Nat Med, vol. 25, no. 1, pp. 44-56, 2019.

6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. "Deep learning for healthcare: review, opportunities and challenges." Brief Bioinform, vol. 19, no. 6, pp. 1236-1246, 2018.

7. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records." J Am Med Inform Assoc, vol. 23, no. 5, pp. 899-908, 2016.

8. U.S. Food and Drug Administration. "Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff." FDA, 2022.

9. U.S. Food and Drug Administration. "Artificial Intelligence and Machine Learning in Software as a Medical Device." FDA, 2021.

10. ISO 14971:2019. "Medical devices — Application of risk management to medical devices." International Organization for Standardization, 2019.

11. ISO 13485:2016. "Medical devices — Quality management systems — Requirements for regulatory purposes." International Organization for Standardization, 2016.

12. IEC 62304:2006+AMD1:2015. "Medical device software — Software life cycle processes." International Electrotechnical Commission, 2015.

13. Char DS, Shah NH, Magnus D. "Implementing machine learning in health care — addressing ethical challenges." N Engl J Med, vol. 378, no. 11, pp. 981-983, 2018.

14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. "Dissecting racial bias in an algorithm used to manage the health of populations." Science, vol. 366, no. 6464, pp. 447-453, 2019.

15. Lundberg SM, Lee SI. "A unified approach to interpreting model predictions." Adv Neural Inf Process Syst, vol. 30, pp. 4765-4774, 2017.

16. Harris AHS, Kuo AC, Bowe TR, Gupta S, Nordin D, Giori NJ. "Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty?" Clin OrthopRelat Res, vol. 477, no. 2, pp. 452-460, 2019.

17. Parvizi J, Tan TL, Goswami K, Higuera C, Della Valle C, Chen AF, Shohat N. "The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria." J Arthroplasty, vol. 33, no. 5, pp. 1309-1314, 2018.

18. Karnuta JM, Luu BC, Haeberle HS, Saluan PM, Fitz W, Schickendantz MS, Ramkumar PN. "Machine learning outperforms regression analysis to predict next-season major league baseball player injuries: epidemiology and validation of 13,982 player-years from performance and injury profile trends, 2000-2017." Orthop J Sports Med, vol. 8, no. 11, 2020.

Citation: Naidu Paila (2025). AI-Based Analysis of Post-Surgical Patient Data for Risk Stratification After Total Knee Arthroplasty. International Journal of Artificial intelligence and Machine Learning, 3(4), 1-8. https://doi.org/10.55124/jaim.v3i4.295