Advances in Consumer Research
https://acr-journal.com/
Volume-3 | Issue-4 |April 2026

Original Researcher Article

Predicting Hospital Length of Stay Using Machine Learning and Spatial Analytics: A Large Open Health Dataset

Mr. Aditya Moogi [1] Prof. Dr. Kavita Kalyandurgamath [2], Dr. Jagadish KM [3], Dr Minakshi Kishore[4]
[1]PhD Scholar, IIC University of Technology, Cambodia, (Enrolment No. FNR210604).
[2]Professor at Prin. L. N. Welingkar Institute of Management Developement and Research (PGDM), Mumbai.
[3]Geomatics Scientist
[4]London School of Management Education

ABSTRACT

This mixed-methods study develops and validates machine learning models for hospital length of stay (LOS) and cost prediction, integrates Geographic Information System (GIS) spatial analytics for population-level healthcare governance, and examines cross-national transferability through qualitative validation with Indian healthcare professionals. Analysing 1,048,575 inpatient discharge records from the New York State SPARCS database, Random Forest regression achieved $R^2 = 0.41$ for LOS prediction (MAE = 3.18 days, 71.3% accuracy within ±3 days) and $R^2 = 0.79$ for cost prediction. Clinical classification variables-principally APR-DRG severity-accounted for 74.61% of feature importance versus 8.21% for demographics, establishing a nine-to-one ratio that identifies classification infrastructure as the binding constraint on prediction capability. Extreme severity patients demonstrated 5.11-fold longer stays than minor cases (15.58 vs. 3.05 days), males showed 19% longer LOS than females (6.30 vs. 5.28 days, $p < 0.0001$), and patients aged 50+ exhibited 68% longer stays. GIS integration extended individual-level predictions to spatial governance, identifying high-burden ZIP code hotspots (~120,000 cases in ZIP 112), a 4.7-fold severity-stratified LOS gradient across geographic units, and a 2-fold county-level cost disparity (New York County ~$41,000 vs. Clinton County ~$20,000) revealing equity gaps invisible to individual-level models. Cost-effectiveness analysis yielded a dominant strategy with negative ICER of –$2,499.55 per bed-day avoided (ROI: 561,637%). Qualitative validation revealed complete APR-DRG unfamiliarity among all seven Indian healthcare professionals, exposing a structural classification gap with cascading consequences for reimbursement fairness, hospital benchmarking, and spatial resource allocation. The findings support phased adoption of severity-adjusted classification frameworks adapted to India's disease burden, integrated with spatial analytics for district-level healthcare governanc..

Keywords: Hospital length of stay, Machine learning, Random Forest, Geographic Information Systems; APR-DRG Severity classification, Cost-effectiveness, SPARCS Health equity, Spatial analytics, India healthcare policy

INTRODUCTION:

Healthcare stands out as the most complex domain, owing to its profoundly data-rich nature and its heavy reliance on accurate, timely data for effective decision-making and care delivery. The sector generates vast, heterogeneous datasets from electronic health records (EHRs), medical imaging, laboratory results, genomic sequences, real-time sensor streams, clinical notes, and administrative/financial information, creating an environment where data volume, variety, velocity, and veracity pose ongoing analytical and operational challenges.

Healthcare systems worldwide face escalating costs, workforce shortages, and quality gaps, with global spending projected at $10 trillion in 2024. Healthcare faces unprecedented pressures, global spending $10T (2024), 20% readmissions, chronic diseases (diabetes) burdening 463M patients. Hospitals struggle with inefficiencies (LOS 5-7 days diabetics), readmissions (25% diabetics), SDOH (Social Determinants of Health) SDOH gaps; causal ROI proof lacking (50). Healthcare analytics - the systematic use of data, statistical algorithms, and machine learning to uncover patterns in health data offers transformative potential. By 2026, analytics could save $300-450 billion annually in the U.S. alone through efficiency and prevention.

The United States has made significant investments in standardized clinical classification systems that allow for the systematic examination of patient treatment patterns. The country spends over $3.6 trillion annually on healthcare, which represents roughly 18% of its GDP. In the US, employers offer health insurance to their workers, and the private sector provides nearly all healthcare. Healthcare is exclusively provided by the government to people who cannot afford health insurance or who are unemployed. 3M Health Information Systems maintains the All Patient Refined Diagnosis Related Groups (APR-DRG) system, which divides hospital discharges into clinically coherent groups according to diagnoses, procedures, illness severity, and mortality risk. In a similar vein, the Agency for Healthcare Research and

Quality's Clinical Classifications Software Refined (CCSR) compiles over 70,000 ICD-10-CM codes into roughly 530 clinically significant categories. Predictive analytics can be developed using the structured data infrastructure provided by these classification frameworks.

As of 2026, India's health budget represents roughly 0.26% of GDP and 1.9% of all government spending. In other affluent countries, healthcare accounts for about 10% of GDP. India, on the other hand, offers a quite distinct institutional environment. Both the national and state governments oversee healthcare in India. The federal government gathers data on infectious diseases and is in charge of medical education. In India, both public and private organizations oversee hospitals and clinics. State governments manage 75% of hospitals and clinics that offer primary, secondary, and tertiary healthcare. The Indian healthcare system, which has more than 69,000 hospitals and serves a population of more than 1.4 billion, is dispersed throughout a variety of settings, from resource-constrained basic health facilities in rural areas to sophisticated tertiary care centers in urban areas.

In 2026, the insurance penetration rate in the US was 91.4%, much higher than that of India, which was 3.7%. India's healthcare system is mostly dependent on government-funded initiatives. People are burdened financially as a result of low insurance penetration rates. In India, paying for medical care out of pocket is common. Patients find it challenging to accurately estimate the expenses of their medical care. Ironically, customers can easily access online prices for electronics, books, clothing, and other things, but they have a harder time finding information regarding healthcare. Customers and patients will be able to make better and more educated decisions if healthcare information, such as pricing, disease incidence, and the anticipated length of stay for various procedures, is made available. In the United States, for example, patients can choose when to have an elective surgery depending on the anticipated length of the treatment or budget pre-tax contributions to health savings accounts.

The US provides full coverage for medical costs, including cosmetic procedures and outpatient services. All medical costs, including hospital stays, operations, prescription drugs, and more, are often covered by US insurance policies. Although this all-inclusive coverage is advantageous, the prices are greater than those of Indian policies. Insurance coverage in the US is renowned for being broad, providing policyholders with a variety of services and treatments.

In India, pre-hospitalization and post-hospitalization costs are usually covered by health insurance plans, guaranteeing complete medical assistance. However, OPD and doctor consultations are frequently not covered by these plans, which puts a financial burden on people for regular medical needs. By helping those in need with healthcare, government-funded initiatives like Ayushman Bharat seek to close the gap.

Inpatient hospital reimbursement and analytics in the US are heavily reliant on APR-DRG (All Patient Refined Diagnosis Related Groups), especially in Medicare severity-DRG systems and many state Medicaid and private payer contracts, where hospitals routinely assign DRG codes to almost all inpatient discharges for payment and benchmarking purposes. Because of this widespread adoption, diagnoses and procedures may be standardized with built-in severity and risk adjustment, supporting national benchmarking, predictable reimbursement, and statistically sound length-of-stay (LOS) expectations for every clinical category.

Crucially, India does not have a nationally regulated APR-DRG-like patient classification system. Administrative data harmonization is inadequate, hospital coding systems are still diverse, and severity documentation is variable. Severity-adjusted classification is not used in the package-rate reimbursement mechanism of the Ayushman Bharat Pradhan Mantri Jan Arogya Yojana (AB-PMJAY), India's premier government health insurance program that serves about 500 million people.

The underlying data and the models that interpret the data are necessary to attain this capacity. A key concern for health systems research is raised by this structural divergence: can predictive models created using well-classified administrative data from the US offer valuable insights for a health system without such infrastructure for classification? More precisely, what does the absence of severity categorization mean for prediction accuracy, reimbursement equity, and hospital benchmarking in India if it turns out to be the most significant predictor of LOS in a structured dataset? To keep national economies sustainable, research on healthcare spending and patient outcomes is essential. To identify the best answers, it is beneficial for countries to combine research from universities, non-profits, the government, and the commercial sector. Opening up health data is a promising approach that enables researchers from various fields to engage and share their knowledge. Open health data has been made public by groups like the New York State Statewide Planning and Research Cooperative System (SPARCS), despite clear patient privacy concerns. After the data is made accessible, it must be appropriately analyzed to extract insights and meaning that will benefit patients and healthcare professionals.

The current study uses a mixed-methods approach to investigate these concerns. Using more than a million inpatient discharge records from New York State, the quantitative component creates and assesses machine learning models for LOS and cost prediction. In order to determine whether the patterns found in the quantitative analysis align with clinical reality in a different health system context and to pinpoint structural obstacles to the implementation of comparable predictive systems, the qualitative component performs validation interviews with medical professionals in Indian hospitals.

To put our findings into practice, we utilize our machine-learning system to forecast hospital length of stay (LOS) based on patient data from the publicly available healthcare data provided by New York State SPARCS. We show that it is feasible to predict the LOS using data that contains the Clinical Classifications Software Refined (CCSR) diagnosis code, the severity of sickness, and the need for surgery by developing a predictive model using machine learning techniques. To fully examine the decisions that impact final model performance, we look into a number of analytics approaches, such as feature selection, feature engineering, model selection, and model training. A Random Forest model yields an R2 score of 0.41. Policymakers and healthcare providers will benefit from our model's performance in terms of capacity planning and cost control. Our model can also be used by patients and customers to schedule elective surgeries or to anticipate the length of stay for procedures they are undergoing.

One of the most important factors in healthcare management is hospital duration of stay. It shapes hospital throughput capacity, establishes the amount of resources used by each patient episode, and has a direct impact on the cost of providing treatment. Research on the effectiveness and quality of clinical care has focused on hospital length of stay (LOS). Since the number of acute admissions and readmissions has been rapidly increasing in recent years, information on the LOS and related outcome indicators is more crucial than ever. Predicting length of stay (LOS) at or near the moment of admission has become strategically important for operational planning and discharge coordination in a time when health systems around the world are under increasing pressure to provide high-quality care within limited financial resources and financial sustainability.

GIS Integration: Spatial Analysis

Healthcare governance needs population-level spatial intelligence to convert forecasts into practical policy, even though the previous machine learning models predict individual-level LOS and cost consequences. Using the GIS + AI conceptual approach, this section expands the analytical framework by combining Geographic Information Systems (GIS) with the prediction models. While AI forecasts future clinical and financial outcomes, GIS pinpoints the physical locations of health concerns. Five governance-relevant use cases, including disease surveillance, capacity planning, emergency burden analysis, precision public health targeting, and health equality mapping, are made possible by spatial aggregation at the ZIP code and county levels using the same SPARCS dataset of 1,048,575 records.

2. LITERATURE CONTEXT

The prediction of hospital LOS has attracted sustained research attention over several decades, progressing from simple statistical methods to increasingly sophisticated machine learning approaches. Stone et al. (2022) conducted a systematic review of LOS prediction studies and identified that Random Forest consistently performs well across clinical contexts, with reported R^2 values ranging from 0.20 to 0.94 depending on population homogeneity and feature availability. The review also highlighted a persistent methodological challenge: most studies focus on specific disease conditions or individual hospitals, limiting the generalizability of their findings.

The main gap in the literature is that most methods focus on analyzing trends in the LOS or predicting the LOS only for specific conditions or restrict their analysis to data from specific hospitals. For instance, LaFaro et al. (2015) used pre-incision characteristics in a retrospective cohort analysis to predict length of stay (LOS) in the intensive care unit (ICU) after heart surgery by using 36 pre-incision characteristics, such as demographic, clinical, and laboratory indicators with 185 cardiac surgery patients. In contrast, we have developed our model to predict the LOS for 477 different CCSR diagnosis codes, over a set of 1.4 million patients over all hospitals in New York state.

Jain et al. (2024), using 2.3 million patient records from the same New York State SPARCS database employed in the present study, achieved an R^2 of 0.43 for non-newborns using CatBoost regression and identified APR DRG Code and APR Severity of Illness as the top predictors. Their work demonstrated that a single model could handle 285 different CCS diagnosis codes simultaneously, a departure from the disease-specific models that dominate the literature. However, their analysis did not extend to cross-national comparison or qualitative validation of findings.

The role of severity classification in LOS prediction has been substantiated by several independent investigations. McCormick et al. (2018) validated APR-DRG Risk of Mortality and Severity of Illness as perioperative risk measures, finding that these classifications outperformed the Charlson Comorbidity Index in explaining outcome variance. Nugent et al. (2021) reported that each unit increase in APR-DRG severity was associated with a 10.37 to 11.45-fold increase in mortality odds, with C-statistics exceeding 0.93, confirming the clinical validity of the severity grading system.

The literature on healthcare analytics in India, however, reveals a significant gap. While studies have examined hospital efficiency and cost patterns in Indian settings, few have attempted to develop LOS prediction models using machine learning, and none have done so with the benefit of APR-DRG or CCSR classification. Marfil-Garza et al. (2018), analysing an 18-year retrospective cohort of 85,904 hospitalizations, identified severity of illness as a primary driver of prolonged hospital stay, but their study was conducted in a Mexican tertiary care centre with access to structured classification data.

The cost dimension of LOS prediction has also received attention. Studies have consistently demonstrated a strong positive association between LOS and total hospital costs, with the relationship being mediated by resource intensity, procedural complexity, and severity of illness. The cost prediction literature suggests that while LOS is a strong predictor of total costs, the relationship is not perfectly linear-costs tend to increase non-linearly for extended stays due to escalating care intensity and complication management.

What remains underexplored is the structural comparison between health systems that possess standardized classification infrastructure and those that do not, and the implications of this difference for the feasibility and accuracy of predictive analytics**.** The present study addresses this gap by integrating large-scale quantitative modelling with qualitative cross-national validation, along with the potential application of Geographic Information System (GIS) analysis. To the best of our knowledge, no existing model predicts Length of Stay

(LOS) across such a wide range of diagnosis codes using a patient sample exceeding one million records, based on freely available open data, combined with qualitative cross-national validation and a GIS-based approach. From this perspective, our research offers a distinctive and novel contribution to the field.

3. METHODS

3.1 Study Design and Data Source

This study employed a sequential mixed-methods design comprising a primary quantitative analysis of administrative healthcare data followed by qualitative validation through semi-structured interviews. During our research, we utilized open-health data provided by the New York State SPARCS system. The data we accessed was from the year 2024, which was the most recent year available at the time. This data was provided in the form of a CSV file, containing 1,048,575 rows and 33 columns. Each row contains de-identified in-patient discharge information. The dataset columns contained various types of information. They included geographic descriptors related to the hospital where care was provided, demographic descriptors such as patient race, ethnicity, and age, medical descriptors such as the CCSR diagnosis code, APR DRG code, severity of illness, and length of stay. Additionally, payment descriptors were present, which included information about the type of insurance, total charges, and total cost of the procedure. We examine the distribution of the LOS in the data set and we note that the providers of the data have truncated the length of stay to 120 days. This explains the peak we see at the tail of the distribution. SPARCS collects patient-level data on all hospital discharges in New York State, encompassing demographic information, clinical classifications, procedural details, severity indicators, and financial data.

Data pre-processing and cleaning

The dataset comprised 1,048,575 inpatient discharge records. After removing records with missing or invalid LOS values (approximately 0.11% of the original dataset), the final analytical sample included 1,003,061 records. The dataset contained 34 variables spanning geographic descriptors, demographic characteristics, clinical classifications (APR-DRG codes, CCSR diagnosis and procedure codes, severity of illness, risk of mortality), admission details, and payment information.

Data cleaning involved the removal of records with missing LOS values and the exclusion of variables that would introduce data leakage into the predictive model. Total Costs and Total Charges were excluded from the LOS prediction model as these variables are inherently proportional to LOS and their inclusion would constitute circular reasoning. Descriptive text fields were removed where corresponding numerical codes were available, reducing redundancy without information loss.

Feature engineering

Many variables in the dataset are categorical, e.g., the values in the set [Major, Minor, Moderate, Extreme]. The variable "APR Severity of Illness Description" has the Feature engineering included the creation of a numeric Severity Score variable from the APR Severity of Illness Code. Categorical variables were encoded using appropriate techniques for the Random Forest algorithm, which handles categorical splits natively. The dataset was split into training (80%) and testing (20%) sets, yielding 802,448 and 200,613 records respectively. Random Forest is a machine learning algorithm that uses many decision trees to make better predictions. Each tree looks at different random parts of the data and their results are combined by voting for classification or averaging for regression which makes it as ensemble learning technique. This helps in improving accuracy and reducing errors.

3.3 Machine Learning Models

Random Forest regression was selected as the primary modelling approach based on its established performance in healthcare prediction tasks, its robustness to overfitting, and its capacity for generating interpretable feature importance rankings. Two separate models were developed: one for LOS prediction and one for cost prediction. The algorithm was implemented using the scikit-learn library in Python.

Model hyperparameters were configured as follows: n_estimators = 100, max_depth = 15, min_samples_split = 10, min_samples_leaf = 5, and random_state = 42 for reproducibility. These parameters represented a balance between predictive accuracy, computational efficiency, and overfitting prevention. Five-fold cross-validation was additionally employed to assess model stability and generalization capability.

Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and the proportion of predictions falling within clinically relevant error thresholds of ±1, ±2, and ±3 days. Feature importance was derived from mean decrease in impurity across the ensemble of decision trees.

3.4 Statistical Analyses

One-way analysis of variance (ANOVA) was used to test for statistically significant differences in mean LOS between male and female patients. The Chi-square test of independence was employed to examine the association between age group and LOS category. Pearson correlation coefficient was calculated to assess the linear relationship between LOS and total hospital costs. Effect sizes were computed using eta-squared (η^2) for ANOVA and Cramér's V for Chi-square tests. All statistical tests used a significance level of $\alpha = 0.05$.

3.5 Cost-Effectiveness Analysis

A cost-effectiveness analysis was conducted comparing ML system implementation against the status quo. Implementation costs were estimated at $350,000 annually, encompassing software development, hardware, training, maintenance, data integration, and operations. The daily bed cost was set at $2,500 based on published hospital cost data. LOS reduction estimates were derived from published literature indicating that early discharge planning enabled by accurate prediction can reduce hospitalization by 0.5 to 1.5 days per patient. The Incremental Cost-Effectiveness Ratio (ICER) was calculated as the ratio of incremental costs to incremental effects. Sensitivity analysis was conducted across conservative, moderate, optimistic, and best-case scenarios.

3.6 Qualitative Validation Study

To assess the cross-national applicability of the quantitative findings, a qualitative validation study was conducted with seven healthcare professionals across Indian hospitals. Participants included physicians, hospital administrators, and nursing staff from both government and private institutions. Semi-structured interviews explored participants' clinical experience with LOS determinants, familiarity with standardized classification systems, perceptions of severity as a LOS predictor, and views on implementing predictive analytics tools.

The interview protocol was designed to test specific findings from the quantitative analysis without leading respondents toward particular conclusions. Questions addressed gender and age effects on LOS, the role of clinical severity, awareness of APR-DRG classification, and perceived barriers to implementing machine learning-based prediction systems in Indian hospitals. Thematic analysis was conducted on the interview transcripts.

Spatial Data Processing Methodology

The GIS analysis utilized seven variables from the SPARCS dataset: Age Group, ZIP Code (3-digit prefix), Hospital County, Type of Admission, APR Severity of Illness, Length of Stay, and Total Costs. Since the dataset lacks individual geocoordinates, spatial aggregation was performed at ZIP code prefix and county levels-a standard public health GIS approach suitable for policy-level planning. This privacy-by-design methodology ensures anonymization through aggregation while enabling population-level spatial pattern identification across five analytical dimensions: case volume by ZIP code, severity-stratified LOS mapping, cost analysis by admission type, age-group distribution, and county-level cost equity assessment.

4. RESULTS

4.1 Descriptive Characteristics

The study population of 1,048,575 inpatient discharge records exhibited a mean LOS of 5.73 days (standard deviation: 8.12 days) with a median of 3.00 days, indicating pronounced right-skewness in the distribution. The interquartile range was 2 to 6 days, with a maximum observed LOS of 120 days. This distributional pattern is characteristic of large administrative datasets where the majority of patients experience relatively short hospitalizations while a subset of complex cases generates extended stays that significantly influence the mean.

Mean total costs were \$88,276.90 with a median of \$47,180.13, reflecting similarly skewed cost distributions. The Pearson correlation between LOS and total costs was 0.6904 ($p < 0.0001$), confirming a strong positive linear relationship between hospitalization duration and economic burden.

Table 1. Length of Stay by APR Severity of Illness Classification

Severity Level	Mean LOS (days)	Number of Cases	Fold Difference vs. Minor
Minor	3.05	292,479	1.00 (Reference)
Moderate	4.52	393,194	1.48
Major	7.84	275,225	2.57
Extreme	15.58	87,462	5.11

Table 1 reveals a clear severity gradient in LOS. Patients classified with extreme severity of illness demonstrated a mean LOS of 15.58 days, representing a 5.11-fold increase over the 3.05-day mean for minor severity cases. This 12.53-day absolute difference between the highest and lowest severity categories underscores the exceptional discriminatory power of structured severity classification for hospitalization duration.

4.2 Demographic Associations

ANOVA testing revealed a statistically significant difference in mean LOS between genders (F-statistic significant at $p < 0.0001$), with males demonstrating a 19% longer mean LOS than females (6.30 vs. 5.28 days). However, the effect size was small ($\eta^2 = 0.004$), indicating that while the gender difference is statistically robust given the large sample size, it explains less than half a percent of LOS variance.

Chi-square analysis demonstrated a significant association between age group and LOS category ($\chi^2 = 75{,}395.78$, $p < 0.0001$, Cramér's V = 0.19). Patients aged 50 and above exhibited 68% longer mean LOS compared to pediatric populations. The age-LOS relationship showed a monotonically increasing trend, with mean LOS rising progressively across age categories.

4.3 Machine Learning Model Performance

Table 2. Random Forest Model Performance Metrics

Metric	LOS Model	Cost Model
R^2 Score	0.4098	0.7886
Mean Absolute Error	3.18 days	$7,170.71
Root Mean Squared Error	6.88 days	-
Predictions within ±1 day	38.2%	-
Predictions within ±2 days	56.1%	-
Predictions within ±3 days	71.3%	-
Cross-validation R^2	0.4156 (±0.015)	-

The LOS prediction model achieved an R^2 of 0.4098, explaining approximately 41% of variance in hospitalization duration. The cross-validation R^2 of 0.4156 (±0.015) confirmed model stability across different data partitions, with the narrow confidence interval indicating consistent generalization rather than overfitting. The operational accuracy metric of 71.3% of predictions falling within ±3 days of actual values demonstrates substantial practical utility for bed management and discharge planning.

The cost prediction model achieved a notably higher R^2 of 0.7886, indicating that economic patterns at the system level are more predictable than individual patient LOS trajectories. This asymmetry between LOS and cost model performance reflects the additional clinical uncertainties inherent in individual patient courses that are partially smoothed at the aggregate economic level.

4.4 Feature Importance Analysis

Table 3. Feature Importance Rankings for LOS Prediction

Rank	Feature	Importance (%)	Category
1	APR DRG Code	25.86	Clinical
2	APR Severity of Illness Code	14.25	Clinical
3	APR Severity Score (engineered)	13.66	Clinical
4	CCSR Procedure Code	7.31	Clinical
5	CCSR Diagnosis Code	5.13	Clinical
6	APR Risk of Mortality	4.89	Clinical
7	APR Medical Surgical Description	3.51	Clinical
-	Clinical features (combined)	74.61	-
-	Demographic features (combined)	8.21	-

The feature importance analysis revealed a decisive dominance of clinical classification variables over demographic factors. Clinical features collectively accounted for 74.61% of total feature importance, while demographic variables contributed only 8.21%. The APR DRG Code alone accounted for 25.86% of predictive power, followed by APR Severity of Illness Code at 14.25%. Notably, the engineered Severity Score variable captured an additional 13.66%, bringing the combined severity-related contribution to 27.91%-nearly matching the APR DRG Code itself.

This finding carries substantial implications. The severity dimension provides 3.8 times more predictive information than procedure codes (27.91% vs. 7.31%) and 5.4 times more than raw diagnosis codes (27.91% vs. 5.13%). It establishes that within any given diagnostic category, the degree of illness severity-not the diagnosis itself-is the primary determinant of how long a patient remains hospitalised.

4.5 Cost-Effectiveness Analysis

Table 4. Cost-Effectiveness Analysis of ML-Based LOS Prediction System

Parameter	**Value**
Annual Implementation Cost	\$350,000
Projected LOS Reduction	1.0 day per admission
Daily Bed Cost	\$2,500
Projected Annual Bed-Days Avoided	788,615
Projected Annual Savings	\$1,971,537,500
Net Annual Benefit	\$1,971,187,500
Return on Investment	561,637%
ICER (per bed-day avoided)	-\$2,499.55 (Dominant)

The cost-effectiveness analysis yielded an ICER of -\$2,499.55 per bed-day avoided. In health economics, a negative ICER indicates a dominant strategy-the intervention simultaneously reduces costs while improving outcomes. Sensitivity analysis confirmed positive returns across all four scenarios, with ROI exceeding 299,000% even under the most conservative assumption of a 0.5-day LOS reduction per patient. The cost prediction model's high performance ($R^2 = 0.79$) with LOS contributing 57.83% of feature importance empirically validates the causal pathway through which LOS reduction translates to proportional cost savings.

4.6 Qualitative Analysis

Questionaries findings

Several conclusions from the qualitative validation research with seven Indian healthcare professionals are particularly important for cross-national transferability. We gathered replies from doctors who have worked in private tertiary hospitals with more than 350 beds for more than five years. As a result, the respondents' contribution is significant and suitable for the study because they are regarded as highly skilled and experienced. First, the APR-DRG classification system-which turned out to be the most significant predictor in the quantitative analysis-was completely unknown to all seven responders (100%). None of the participants had experienced the system in their training, clinical practice, or institutional operations, therefore this conclusion was not due to incomplete awareness.

Second, stakeholders showed strong agreement with the clinical patterns found by the machine learning model despite their lack of familiarity with formal categorization systems. The model's identification of sickness severity as the primary predictor was validated by the fact that six out of seven respondents (86%) agreed or strongly agreed that illness severity was the most significant determinant of LOS. The age effect found in the quantitative analysis was also acknowledged by six out of seven, supporting the conclusion that patients over 50 have much longer hospital stays.

Third, participants highlighted a number of India-specific characteristics that affect length of stay (LOS) that are not included in the SPARCS dataset: five out of seven respondents mentioned unforeseen difficulties, three mentioned family refusal to accept discharge, and two mentioned insurance-related payment delays. These elements stand for contextual characteristics that must be included in any LOS prediction model tailored to India.

Fourth, when it came to implementation constraints, six out of seven respondents cited cost as their top concern. This was followed by issues with accuracy (three respondents), IT infrastructure gaps, and system integration difficulties (three each). In contrast to point estimates, five out of seven respondents indicated interest in risk categorization tools and daily updated predictions, indicating that the style of predictive output is just as important in clinical acceptance as its accuracy.

Case studies analysis

In order to evaluate the clinical application of machine learning-based length of stay (LOS) prediction models, this qualitative validation study looked at five anonymized patient cases. The cases allowed for the triangulation of quantitative model results with actual clinical patterns because they represented a variety of clinical presentations, demographic characteristics, and hospitalization trajectories.

Case Summaries

Case A: A 68-year-old man with diabetes who was admitted urgently for cardiac arrest and needed surgery and intensive care. LOS was one week, following which the patient was sent home and readmitted a few months later.

Case B: An 85-year-old woman with diabetes, hypertension, and cardiac arrest was admitted to the intensive care unit twice in a short period of time (3 and 4 days), and she died after being discharged.

Case C: A male patient (65–70 years old) with diabetes and hypertension had two admissions-an elective stent insertion (4 days) and an emergency cardiac arrest (1 week)-showing the impact of admission type on length of stay.

Case D: A 45-year-old man with end-stage alcoholic liver disease needed a liver transplant. His hospital stay lasted more than a month due to a series of problems, including tracheostomy and ventilator-associated pneumonia.

Case E: Despite being younger, a 34-year-old man with acute pancreatitis and diabetes needed surgery and ICU care, which led to a nearly one-month hospital stay.

Analysis of findings

Important quantitative findings were confirmed by the qualitative analysis. Strong support was given to the APR-DRG severity categorization as the best predictor-Cases D and E showed that clinical severity is more important in determining LOS than demographic characteristics. Cases A, B, and C supported the conclusion that patients over 50 have 68% longer stays; however, Case E demonstrated that severe presentations outweigh age-related benefits. Qualitative support was obtained for the prediction accuracy metrics (R^2 of 0.41 for LOS, 0.79 for costs). While Cases D and E showed unexplained variance related to complication cascades not captured by baseline predictors, Cases A, B, and C followed predictable patterns.

The qualitative case studies shed light on situations that lead to prediction uncertainty while validating the predictors found by the machine learning model. The results confirm that the model can be applied to regular presentations and point out areas where procedural complexity indicators can be added to improve accuracy.

Spatial Analysis Results

Use Case 1: Predictive Disease Surveillance: ZIP 112 (~120,000 cases), ZIP 104 (~100,000), and ZIP 117 (~90,000) were shown to be the highest-burden locations by ZIP code hotspot analysis. This allowed for proactive screening and resource pre-positioning in high-demand zones.

Use Case 2: Hospital Capacity & ICU Planning: A 4.7-fold gradient from Minor (~3 days) to Extreme severity (~14 days) was shown by severity-stratified LOS mapping, supporting the ML finding that APR-DRG severity is the dominating predictor and allowing for spatially targeted ICU capacity allocation.

Use Case 3: Preventable Emergency Burden: Cost analysis by admission type revealed that trauma ($40,000) and urgent ($33,000) admissions were the most expensive, while emergency ($27,000) and newborns ($12,000) were the least expensive. This highlights opportunities for primary care investment and injury prevention to lower avoidable high-cost admissions.

Use Case 4: Precision Public Health Targeting: Patients aged 70+ (~340,000 cases) and 50–69 (~290,000) were shown to be the highest-burden cohorts by age-group distribution analysis, which supported geriatric service planning and targeted NCD screening programs in high-concentration locations.

Use Case 5: Health Equity Mapping: A county-level cost analysis showed a two-fold difference between Clinton County (~$20,000) and New York County (~$41,000), highlighting systemic inequities, infrastructure gaps, and delayed care access that call for targeted resource redistribution.

Table: Summary of GIS-Based Spatial Analysis Findings

Use Case	Key Finding	Policy Application	Metric
Hotspot Detection	ZIP 112: ~120K cases (highest)	Proactive screening	Case volume by ZIP
ICU Capacity Planning	4.7-fold LOS gradient (Minor→Extreme)	Severity-based allocation	LOS by severity
Emergency Burden	Trauma $40K highest; Newborn $12K lowest	Injury prevention	Cost by admission type
Precision Targeting	70+ age group: ~340K cases	NCD screening programs	Age distribution
Health Equity	2-fold cost disparity across counties	Resource redistribution	County-level cost

5. DISCUSSION

5.1 Principal Findings in Context

The current study shows that clinically significant LOS prediction accuracy can be attained by machine learning models trained on structured administrative data. The operational accuracy metric of 71.3% within ±3 days offers a more practically interpretable benchmark, but the R^2 of 0.41 is within the range reported in the broader literature (Bertsimas et al. (2008) reported R^2 of 0.20, Kshirsagar (2021) reported R^2 of 0.33, and Jain et al. (2024) achieved 0.43 on a larger SPARCS sample). Predictions that are correct to within three days for more than seven out of ten patients provide actionable intelligence for a bed manager making allocation decisions or a discharge planner starting post-acute care plans.

Perhaps the most important conclusion for health policy is the clear dominance of clinical categorization variables over demographic characteristics. While demographics accounted for just 8.21% of the total feature relevance, clinical characteristics accounted for 74.61%. This nine-to-one ratio has significant ramifications because it indicates that the length of a patient's hospital stay depends about nine times more on what the patient is being treated for and how ill they are. From an equity standpoint, this finding is comforting because the data does not show systematic bias based on demographic identity, but it also emphasizes how crucial correct clinical classification is for any health system trying to forecast and control LOS.

5.2 The Classification Infrastructure Gap: United States versus India

This study's most important contribution is not the model performance measurements per se, but rather what the feature importance analysis shows about the structural requirements for successful LOS prediction. Clinical classification systems have been developed, improved, and mandated by the US health system for decades. With 332 basic categories multiplied by four severity levels to create 1,330 total classifications, the APR-DRG framework is a significant investment in organizing clinical complexity into categories that can be analyzed. For procedural and diagnostic codes, the CCSR system serves a similar purpose. These systems work together to build the data architecture that enables the observed prediction accuracy.

No equivalent investment has been made by India's healthcare system. The qualitative discovery that all seven Indian healthcare workers-including doctors, administrators, and nurses from both public and commercial hospitals-were wholly ignorant of APR-DRG classification goes beyond simple ignorance. The systematic, standardized assessment of sickness severity at the point of care, recorded in a structured format that can enable analytics, reimbursement, and quality measurement, is an institutional layer that is absent from Indian healthcare.

This absence has ramifications in a number of areas. Imagine a patient who arrives at an Indian hospital with pneumonia. Without a standardized severity rating, the treating physician must use clinical judgment to determine likely length of stay (LOS). The hospital administration lacks a systematic method for forecasting the number of beds needed. There is no severity-adjusted standard by which the insurance company can assess whether the length of hospital stay was fair. Financial, operational, and clinical choices are all made less effectively than they would be with structured classification data.

5.3 The Reimbursement Dimension

The effects of the classification gap are exacerbated by the differences between the US and Indian reimbursement models. Hospitals are paid by the US Medicare system according to DRG assignments, with severity modifications that take into consideration the anticipated resource intensity of each patient episode. Clinical complexity and financial remuneration are aligned when a hospital treats a disproportionate number of high-severity patients.

In contrast, India's AB-PMJAY employs flat package prices for specific treatments; for example, a knee replacement is paid at the same amount regardless of the patient's level of comorbidity. Perverse incentives result from this. High-severity patients who would need more resources than the package rate may not be treated by hospitals; this practice is referred to as "cream skimming." On the other hand, hospitals that handle complicated situations bear the financial burden of uncompensated expenses. In either case, patients suffer as a result of a reimbursement system that is unable to differentiate between clinical complexity levels that are fundamentally different.

These biases might be corrected in India by implementing a classification system that is comparable to the APR-DRG. Severity-adjusted reimbursement would lower patient selection incentives and increase access to care for complicated situations by ensuring that hospitals treating sicker patients receive fair compensation. This policy argument is empirically supported by our quantitative finding that severity classification accounts for more than 27% of LOS volatility, and hence a commensurate percentage of cost variance.

5.4 Implications for Hospital Benchmarking and Quality Measurement

Severity grading allows for useful hospital benchmarking in addition to reimbursement. Comparing LOS or expenses between hospitals without severity adjustment is intrinsically deceptive because a hospital with a longer average LOS can merely be serving sicker patients rather than running inefficiently. The extent of this confounding is demonstrated by the current study's result that patients with great severity stay 5.11 times longer than those with minor severity. Hospitals that treat the most clinically complicated populations run the danger of being penalized in any comparison of hospital performance that does not take this five-fold severity gradient into consideration. Currently, state-level quality monitoring programs and India's National Accreditation Board for Hospitals (NABH) lack the classification infrastructure necessary to carry out risk-adjusted benchmarking. For the first time, Indian hospitals will be able to evaluate their LOS performance to severity-matched standards, spot real inefficiencies, and monitor progress over time by putting in place a severity classification system. This is an essential quality-improvement feature that is currently lacking.

5.5 Insurance Dispute Resolution and Cost Forecasting

Severity-adjusted LOS prediction may significantly lower insurance disputes in India, according to the high accuracy of the cost prediction model ($R^2 = 0.79$) and the strong association between LOS and costs ($r = 0.69$). One of the main sources of conflict in the Indian health insurance market at the moment is disputes between hospitals and third-party administrators (TPAs) about how long hospital stays are appropriate. Neither side has an objective foundation for judging whether a certain LOS is appropriate in the absence of severity-adjusted benchmarks.

A neutral reference point might be a prediction system that calculates expected length of stay (LOS) based on diagnosis, severity, and operations. A patient's hospital stay may be deemed clinically appropriate if their actual length of stay (LOS) is within the expected range for their severity class. Though they wouldn't automatically result in denial, significant deviations-in either direction-would call for an investigation. By substituting evidence-based standards for subjective judgment in insurance choices, this strategy would lessen disagreements and increase provider-payer trust.

5.6 From Individual Prediction to Spatial Governance: The GIS Dimension

The GIS integration expands these results to population-level spatial governance, whereas Sections 5.1 through 5.5 look at what the ML models show about individual-level prediction and categorization infrastructure. The geographical analysis shows that the same SPARCS dataset produces governance-ready insights when analyzed through complementing spatial and predictive lenses, bridging individual-level ML predictions ($R^2 = 0.41$ for LOS, $R^2 = 0.79$ for cost) with geographic intelligence. The ML conclusion that APR-DRG severity categorization is the dominating predictor, now verified not only at the individual patient level but across geographic units, is supported by the severity gradient of 4.7-fold LOS variance found using GIS-based spatial mapping.

Crucially, the county-level study exposes systemic equity disparities that are opaque to individual-level models alone, revealing a two-fold cost disparity-from around \$20,000 in Clinton County to \$41,000 in New York County. The Indian context covered in earlier sections is directly impacted by these spatial disparities: healthcare systems cannot determine where resources are most needed, where access barriers are concentrated, or where reimbursement inadequacies result in geographic inequality without both severity classification and spatial analytics. Thus, the GIS results support applications such as district health atlases, early warning dashboards, and geographically targeted resource optimization systems for healthcare administrators and policymakers, strengthening the policy case for classification infrastructure by adding a spatial equity dimension that individual-level prediction cannot.

7. Policy Implications

The study's conclusions back up a number of policy suggestions for the Indian healthcare system. India should first think about gradually implementing a standardized patient classification system that takes severity adjustment into account. The US APR-DRG system, which represents American clinical procedures and cost structures, does not have to be completely adopted. Instead, an Indian adaption might integrate regionally validated severity strata pertinent to the Indian disease burden and care delivery context, building on the current ICD-10 coding capabilities.

Second, severity adjustment should be added to the AB-PMJAY reimbursement structure. The clinically significant variance in resource requirements within diagnostic categories is not taken into account by the existing package-rate paradigm, despite its administrative simplicity. The alignment between reimbursement and clinical complexity would be significantly improved by even a simple four-level severity adjustment that mirrors the Minor, Moderate, Major, and Extreme classification in APR-DRG.

Third, hospital information systems at the state level ought to be improved in order to facilitate organized clinical data gathering. A pattern for Indian states is provided by the SPARCS model, a statewide mandatory reporting system that produces the dataset used in this study. The data infrastructure required for predictive analytics would be created by integrating standardized severity and diagnostic classification into the hospital management information systems being developed by many Indian states.

Fourth, funding for healthcare analytics capacity should be given top priority. This includes institutional support for evidence-based decision-making, computational infrastructure, and staff training. Six out of seven respondents named cost as the main implementation hurdle. This qualitative finding implies that financial support-possibly from federal or state government funding-will be required to accelerate adoption.

7. LIMITATIONS

There are a few restrictions to be aware of. First, generalizability to other US states or foreign contexts cannot be assumed because the quantitative analysis is only based on data from New York State. LOS trends are impacted by regional differences in clinical practice, payer mix, and hospital capacity, which may restrict the model's external validity.

Second, specific physiological characteristics like vital signs, test findings, and medication information are not included in the SPARCS dataset. Prediction accuracy may be enhanced by more detailed clinical data, such as that found in databases like MIMIC-III. The ceiling imposed by using administrative data instead of clinical data is reflected in the R2 of 0.41.

Third, the qualitative validation study gathered information from seven doctors who had worked in private tertiary hospitals for more than five years. Although this sample offers some initial insights, it cannot be said to be representative of the varied healthcare environment across India. To validate these results, more extensive surveys are required.

Fourth, multiple hospitalizations for the same patient cannot be analyzed because to the de-identification process in the SPARCS data, which makes it impossible to evaluate the effects of comorbidity across episodes. Longitudinal analysis at the patient level would offer deeper understanding of the factors influencing LOS.

Fifth, redundancy between the engineered Severity Score variable and the APR Severity of Illness Code was found by the feature importance analysis. This inflates the perceived amount of predictive features and should be fixed in subsequent model iterations, even though it does not affect model accuracy.

Sixth, the lack of geocoordinates in the SPARCS dataset limits geographic analysis to ZIP code prefix and county-level aggregation. This restricts spatial resolution and makes it impossible to perform more detailed analysis like facility-level catchment mapping or neighborhood-level hotspot discovery. Higher-resolution spatial applications, such as ambulance routing optimization, climate-health forecasting models, and health digital twins for real-time governance, may be made possible by future research integrating exact geocoordinates, environmental layers, and mobility data. Through anonymization and aggregation, all geographical analyses follow privacy-by-design principles, guaranteeing compliance with data protection laws such as India's DPDP Act 2023.

8. CONCLUSION

This work shows that when supplemented by structured clinical categorization systems, machine learning-based LOS prediction utilizing administrative healthcare data can attain clinically relevant accuracy. A strong empirical connection between classification infrastructure and prediction capabilities is established by the discovery that APR-DRG severity classification and associated clinical variables account for roughly three-quarters of the model's predictive power. The complementary GIS integration shows how population-level governance insights, such as geographic hotspots, severity-stratified capacity needs, and county-level equity gaps, can be obtained from the same dataset when analyzed through spatial lenses. This extends the value of individual-level prediction into useful territorial intelligence.

A major structural gap is revealed by the cross-national comparison: Indian healthcare practice completely lacks the classification systems that enable LOS prediction in the US. This is not only a technical issue; it has far-reaching effects on healthcare transparency, insurance dispute settlement, hospital benchmarking, and reimbursement equity. The qualitative component's documentation of Indian healthcare personnel' total lack of expertise with severity classification methods highlights the extent of this disparity.The future course of policy calls for the adaptation and use of severity categorization concepts suitable to India's disease load, healthcare delivery models, and administrative capacities rather than the straight transfer of American methods into the Indian setting.

Benefits from such an investment would go well beyond LOS prediction and include the fundamental data infrastructure required for evidence-based healthcare management in one of the biggest and most intricate health systems in the world.

In order to enable district-level health atlases and real-time governance dashboards, future research should concentrate on creating severity classification frameworks tailored to India, testing their predictive validity in Indian hospital datasets, assessing the cost-effectiveness of classification system implementation, and expanding the spatial analytics framework with geocoded data. The current study offers the cross-national comparison and analytical framework required to guide that subsequent phase of research. The method that is being given has a number of positive aspects. First, the open-source approach makes reproducibility and transparency possible. Second, insights across several illness stages are made easier by the model's generalizability. Thirdly, the technical foundation permits modular additions by the research community and may readily include new data. Lastly, the evidence generated can readily inform multiple key stakeholders including healthcare administrators planning capacity, policy makers optimizing delivery and patients making medical decisions...

REFERENCES

1. Jain R, Singh M, Rao AR, Garg R. Predicting hospital length of stay using machine learning on a large open health dataset. BMC Health Services Research.
2. Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: Towards a unified framework. PLOS Digital Health. 2022;1(4) :e0000017.
3. McCormick PJ, Lin HM, Deiner SG, Levin MA. Validation of the All Patient Refined Diagnosis Related Group (APR-DRG) Risk of Mortality and Severity of Illness Modifiers as a Measure of Perioperative Risk. Journal of Medical Systems. 2018;42(5):81.
4. Nugent E. APR-DRG severity and mortality classification in surgical populations. Annals of Surgery. 2021;274(6):e1251–e1258.
5. Bertsimas D, et al. Algorithmic prediction of health-care costs. Operations Research. 2008;56(6):1382–1392.
6. Kshirsagar R. Accurate and Interpretable Machine Learning for Transparent Pricing of Health Insurance Plans. AAAI 2021 Conference. 2021.
7. Marfil-Garza BA, et al. Severity of illness as a primary determinant of prolonged hospitalization: an 18-year retrospective cohort. BMC Health Services Research. 2018;18:345.
8. Daghistani TA, et al. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. International Journal of Cardiology. 2019;288:140–147.
9. Awad A, Bader-El-Den M, McNicholas J. Patient length of stay and mortality prediction: a survey. Health Services Management Research. 2017;30(2):105–120.
10. Lequertier V, et al. Hospital length of stay prediction methods: a systematic review. Medical Care. 2021;59(10):929–938.
11. New York State Department of Health. Statewide Planning and Research Cooperative System (SPARCS). https://www.health.ny.gov/statistics/sparcs/. Accessed January 2025.
12. Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR). Available: https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp.
13. 3M Health Information Systems. All Patient Refined Diagnosis Related Groups (APR-DRG). Methodology Overview. 2023.
14. National Health Authority. Ayushman Bharat Pradhan Mantri Jan Arogya Yojana (AB-PMJAY). Government of India. Available: https://nha.gov.in/.
15. Barasell et al. Length of stay prediction using Random Forest in Indonesian tertiary care settings. Journal of Medical Systems. 2021.
16. Tsai PF, et al. Length of hospital stay prediction at the admission stage for cardiology patients using artificial neural networks. Journal of Healthcare Engineering. 2016;2016:7035463.
17. Admission Hyperglycemia and Length of Hospital Stay in Patients With Diabetes and Heart Failure: A Prospective Cohort Study
18. (https://onlinelibrary.wiley.com/doi/full/10.1111/j.1751-7133.2008.07569.x)
19. Costs and Length of Stay of Hospitalizations due to Diabetes-Related Complications (https://onlinelibrary.wiley.com/doi/full/10.1155/2019/2363292)
20. Predicting hospital length of stay using machine learning on a large open health dataset (https://link.springer.com/article/10.1186/s12913-024-11238-y)
21. Evaluation of factors that influenced the length of hospital stay using data mining techniques (https://link.springer.com/article/10.1186/s12911-022-02027-w).