Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Evaluating Treatment Outcomes of Acute Infectious Diseases Using Machine Learning and Ordered Logit Models

Tsolmon Sodnomdavaa [1], Urandelger Gantulga[2], Uyanga Gantumur[3]
[1] *Economics and Business department, Mandakh University; tsolmon@mandakh.edu.mn*
[2] *Marketing and Trade department, National University of Mongolia; urandelger@num.edu.mn*
[3] *Economics and Business department, Mandakh University; uyangag@mandakh.edu.mn*

ARTICLE INFO

Received: 18 Aug 2024

Revised: 10 Sept 2024

Accepted: 28 Oct 2024

ABSTRACT

Acute infectious diseases, especially of viral origin, remain a major global health concern. This study evaluates recovery outcomes using nine machine learning (ML) algorithms including Ordered Logit, Random Forest, Light GBM, and Naive Bayes to identify key predictors such as age, hospital stay duration, and treatment costs. Data were collected from 5,066 patients hospitalized for respiratory infections at the National Center for Communicable Diseases, Mongolia (2022–2024). Recovery was assessed at admission and discharge, categorized into four ordinal levels (0–3). Machine learning models such as Gradient Boosting and SVM achieved the highest predictive accuracy, while the Ordered Logit model offered interpretability, highlighting significant variables including age, length of stay, drug expenditures, pregnancy status, and year of hospitalization. The study demonstrates the complementary value of statistical and ML approaches in predicting clinical outcomes. Future research should explore additional variables such as genetics and mental health to improve model performance.

Keywords: Machine learning, etiology, diseases

1. INTRODUCTION

Acute infectious diseases represent a persistent global health threat, accounting for approximately 17 million deaths annually and imposing an economic burden exceeding USD 500 billion (WHO, 2023). These diseases, ranging from viral infections like influenza and COVID-19 to bacterial infections such as pneumonia and tuberculosis, continue to challenge healthcare systems worldwide.

In Mongolia, the incidence of respiratory infections increased by 32% between 2020 and 2024, signaling growing strain on the national healthcare infrastructure and the urgent need for more effective management strategies. Globally, the COVID-19 pandemic highlighted critical gaps in healthcare preparedness and underscored the importance of timely diagnosis, individualized treatment, and accurate prediction of clinical outcomes. In this context, predicting recovery trajectories in patients with infectious diseases has become an essential aspect of optimizing patient care, allocating hospital resources efficiently, and mitigating the impact of future outbreaks (Xu et al., 2022). Accurate forecasting of recovery outcomes also supports triage decisions, helps identify high-risk patients, and improves overall treatment planning. Recent advancements in artificial intelligence and machine learning (ML) have introduced powerful tools for modeling clinical outcomes using large-scale health data. ML algorithms such as Random Forest, LightGBM, and Support Vector Machines (SVM) have demonstrated strong predictive performance in various clinical domains, including disease prognosis, diagnosis, and resource allocation. Prior studies have shown that these models can accurately predict recovery timelines and mortality risks among COVID-19 patients and other infectious disease cohorts (Xu et al., 2022). However, the majority of existing research approaches recovery as a binary outcome recovered vs. not recovered thereby oversimplifying the nuanced progression of patient health. In practice, recovery often occurs in stages, and ordinal data reflecting different levels of improvement provide a more realistic representation of clinical outcomes.

The Ordered Logit model addresses this methodological gap by estimating multi-level outcomes and has gained increasing attention in health economics, epidemiology, and medical decision-making (Santangelo et al., 2023).

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Despite these methodological advancements, few studies have examined the comparative recovery patterns between bacterial and viral infections using both traditional statistical and modern ML approaches. Moreover, limited research has focused on applying these tools to real-world data from low- and middle-income countries, where infectious disease burdens are disproportionately high and data-driven decision-making is still evolving. This study addresses these gaps by analyzing clinical and demographic predictors of recovery outcomes among 5,066 patients treated for acute respiratory infections at the National Center for Communicable Diseases (NCCD), Mongolia, between 2022 and 2024. Recovery was measured at both admission and discharge and categorized into four ordinal levels. The study applies the Ordered Logit model alongside several ML algorithms including Random Forest, LightGBM, and XGBoost to compare their effectiveness in predicting recovery levels. By doing so, the research contributes to both clinical decision-making and public health planning by demonstrating the value of combining interpretability and predictive power in modeling recovery from infectious diseases.

2. LITERATURE REVIEW

Acute infectious diseases continue to burden global health systems, with accurate prediction of recovery outcomes emerging as a central challenge in clinical epidemiology. While early studies focused on traditional statistical approaches, recent advances underscore the utility of machine learning (ML) in modeling complex, nonlinear clinical data relationships (Santangelo et al., 2023; Make, 2023). For example, Xu et al. (2022) employed Decision Tree, Random Forest, and AdaBoost algorithms on 13,162 Omicron-infected COVID-19 cases, achieving high predictive accuracy for 7- and 14-day recovery outcomes. Similarly, Liu et al. (2025) demonstrated the added value of hybrid ML-epidemiological models in improving disease trajectory assessments.

Empirical comparisons reinforce ML's effectiveness. Smith et al. (2022) reported 85% accuracy using Random Forest for COVID-19 recovery, while Chen et al. (2021) achieved a 78% AUC for modeling antibiotic resistance outcomes using Gradient Boosting. In the present study, Random Forest and XGBoost reached 72% and 70% accuracy, respectively, with performance favoring viral over bacterial infections.

Despite progress, few studies explicitly compare recovery trajectories by pathogen type. Wang et al. (2020) identified age and treatment duration as key predictors in bacterial infections, whereas Li et al. (2023) found early antiviral therapy improved COVID-19 recovery by 40%. Jones et al. (2022), analyzing over 15,000 cases, observed that viral recovery rates exceeded bacteria by 8-12%. Seasonality also affects infectious disease dynamics, yet research remains sparse. Zhang et al. (2021) applied SARIMA models to forecast seasonal infections, while WHO (2023) emphasized climate-driven regional variation. However, single-center sampling limits generalizability (Smith et al., 2020), and crucial predictors such as genetics and environment are often omitted, though Jones and Patel (2019) report their influence may reach 20-30%. Additionally, most studies neglect temporal modeling techniques like ARIMA or LSTM (Lee et al., 2021). Interpretability remains a concern for ML in clinical settings. Rudin (2019) advocates for explainable AI to enhance trust and transparency. Moreover, reliance on retrospective data limits real-time applicability, with WHO (2022) calling for integration of IoT-based vital sign monitoring. Finally, the lack of cross-national analyses, particularly in cold-climate countries like Mongolia, has been highlighted by the Global Health Observatory (2023). To address these gaps, the present study classifies 53 infectious diseases based on ICD-10 codes into viral and bacterial categories.

By combining the interpretive strength of the Ordered Logit model with the predictive power of ML (Random Forest and XGBoost), this research proposes a hybrid framework to evaluate recovery determinants. Accordingly, the study tests the following hypotheses:

H1: Patients diagnosed with viral infectious diseases exhibit statistically higher recovery levels compared to those with bacterial infections.

H2: Machine learning algorithms (e.g., Gradient Boosting, SVM) demonstrate superior predictive accuracy for recovery levels compared to traditional statistical models such as the Ordered Logit model.

H3: Patient age, length of hospital stays (LOS), medical treatment cost (MEDC), and year of treatment (YEAR) exert statistically significant effects on recovery outcomes.

H4: Seasonal patterns of incidence differ significantly between viral and bacterial infections.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3. METHODOLOGY

This study employed a retrospective cohort design based on clinical data from 5,066 patients hospitalized between 2022 and 2024 in Ward No. 3 of the National Cancer Center of Mongolia. Diagnoses were classified using the 10th revision of the International Classification of Diseases (ICD-10), with 25 bacterial and 28 viral infectious diseases grouped accordingly. All data were anonymized and processed in compliance with ethical standards and data confidentiality protocols. The dependent variable was the recovery level (*Change*), measured on a four-point ordinal scale (0 -mild, 1 -moderate, 2 -moderately severe, 3 -severe). Explanatory variables included demographic characteristics (age, gender), clinical indicators (length of stay, treatment costs), and administrative data (year of diagnosis, payment type). Prior to analysis, all variables were standardized; categorical data were transformed using one-hot encoding, and missing values were imputed using the median. The dataset was then randomly split into training and test sets in a 70:30 ratio. A two-stage analytical approach was adopted. In the first stage, descriptive and inferential statistics including Welch's *t*-test, ANOVA, chi-square tests, and time-series correlation analysis were employed to assess group differences between bacterial and viral infections. In the second stage, an Ordered Logit regression model was used to estimate the relationship between explanatory variables and ordinal recovery outcomes, with model parameters calculated via the Newton-Raphson iterative method. Model fit was evaluated using pseudo-R^2 and log-likelihood statistics.

To complement traditional modeling, supervised machine learning algorithms Random Forest, XGBoost, and Gradient Boosting were applied to predict recovery levels and capture nonlinear relationships. Hyperparameter optimization was conducted using GridSearchCV (for Random Forest), with XGBoost configured using a learning rate of 0.1, a maximum tree depth of 5, and 200 estimators. Model performance was assessed using multiple metrics: accuracy, mean absolute error (MAE), root mean squared error (RMSE), class-wise F1-scores, and area under the receiver operating characteristic curve (ROC-AUC). Special attention was given to evaluating predictions across the ordered levels of recovery severity. All data processing, statistical analysis, and model implementation were carried out in Python 3.10, using key libraries such as pandas, NumPy, SciPy, statsmodels, scikit-learn, XGBoost, matplotlib, and seaborn.

4. RESULTS AND DISCUSSION

This study analyzed clinical data from 5,066 patients diagnosed with acute infectious diseases between 2022 and 2024. Of these, 71.9% had viral infections, 28.0% bacterial, and 0.2% other etiologies based on ICD-10 classification. Among bacterial cases, four diagnoses comprised 90.4%: scarlet fever (A38, 48.6%), erysipelas (A46, 20.0%), spotted fever (A77, 12.1%), and shigellosis (A03, 9.6%). Similarly, five viral conditions accounted for 87.9%: varicella (B01, 23.0%), viral skin/mucosal infections (B08, 22.3%), COVID-19 (U07, 21.3%), unspecified viral infections (B34, 13.4%), and viral pneumonia (J12, 8.0%).

Recovery level was defined as the difference between clinical condition on admission (COA) and at discharge (COD), measured on an ordinal scale. Admission severity ranged from 1 (moderate) to 4 (critical), while discharge condition ranged from 0 (mild) to 3 (severe). The recovery variable (*Change*) captured the extent of improvement during hospitalization, from one-step to four-step recovery. To identify key predictors, an Ordered Logit regression model was employed, complemented by eight machine learning algorithms including Random Forest, Gradient Boosting, XGBoost, and LightGBM to enhance predictive accuracy and assess model robustness (Table 1).

Table 1: One-Way tabulation

Condition on Admission ($COA_{i,j}$)			Condition on Discharge ($COD_{i,k}$)			Recovery Level ($Change_{i,z}$)		
Value	Count	Percent	Value	Count	Percent	Value	Count	Percent
1	1401	27.65	0	3366	66.44	0	25	0.49
2	3468	68.46	1	1664	32.85	1	2963	58.49
3	190	3.75	2	30	0.59	2	2004	39.56

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4	7	0.14	3	6	0.12	3	74	1.46
Total	5066	100	Total	5066	100	Total	5066	100

Among all patients, 96.1% presented with either moderate (27.65%) or moderately severe (68.46%) conditions at the time of hospital admission. By the end of treatment, 99.29% of patients had improved to either mild (66.44%) or moderate (32.85%) status. In terms of recovery progression, 98.05% of patients demonstrated clinical improvement, with 58.49% showing a one-level improvement and 39.56% showing a two-level improvement in their condition.

Over the three-year period, the lowest incidence of acute infectious diseases was recorded in January (5.9%), while peak levels were observed in April-May and October-November, each exceeding 9% of total cases. However, seasonal trends varied by type of pathogen. Bacterial infections were most prevalent during April to July and in October, with monthly shares exceeding 10%, whereas their incidence from January to April remained around 4%. In contrast, viral infections displayed elevated rates in February and June to September (approximately 7%), with the highest peak observed in October (10%), and remained above 9% in most other months (Table 2).

Table 2: Seasonal effects on major infectious diseases

Month	Bacterial-origin infectious diseases						Viral-origin infectious diseases					
	A38	A46	A77	A03	A69	A02	B01	B08	U07	B34	J12	J18
I	4.2%	6.0%	0.0%	2.9%	0.0%	3.8%	2.9%	3.6%	23.2%	3.5%	5.2%	29.8%
II	5.1%	4.2%	0.0%	2.2%	0.0%	7.7%	1.2%	6.5%	15.1%	3.1%	8.3%	13.2%
III	6.5%	9.5%	0.0%	2.9%	7.4%	11.5%	7.2%	12.3%	6.8%	9.1%	9.3%	9.6%
IV	8.7%	8.8%	12.2%	1.5%	55.6%	3.8%	6.9%	11.8%	9.8%	7.0%	6.2%	4.4%
V	12.9%	4.2%	14.5%	3.7%	14.8%	15.4%	9.7%	11.6%	7.4%	9.9%	7.6%	4.4%
VI	7.8%	8.8%	25.0%	15.4%	7.4%	7.7%	13.1%	9.5%	2.2%	6.0%	2.8%	0.0%
VII	5.2%	6.3%	18.6%	27.9%	11.1%	11.5%	3.7%	5.2%	18.3%	4.1%	1.0%	0.9%
VIII	3.9%	6.3%	25.0%	20.6%	3.7%	15.4%	6.3%	6.0%	8.4%	6.0%	15.5%	0.9%
IX	9.0%	12.0%	4.1%	9.6%	0.0%	19.2%	8.1%	5.8%	0.1%	17.9%	16.9%	0.0%
X	14.7%	14.1%	0.6%	1.5%	0.0%	3.8%	11.8%	11.1%	1.9%	19.5%	6.9%	2.6%
XI	12.3%	10.2%	0.0%	7.4%	0.0%	0.0%	14.7%	7.9%	5.7%	5.8%	11.7%	12.3%
XII	9.6%	9.5%	0.0%	4.4%	0.0%	0.0%	14.5%	8.6%	1.0%	8.2%	8.6%	21.9%

Seasonal effects were clearly observed for the leading infectious diseases. Among bacterial infections, 39.9% of A38 (scarlet fever) cases occurred during May and September to November, while 26.1% of A46 (erysipelas) cases were concentrated in October and November. In contrast, A77 (spotted fever) and A03 (shigellosis) demonstrated higher incidence rates during the summer months of June to August (Fig.1).

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

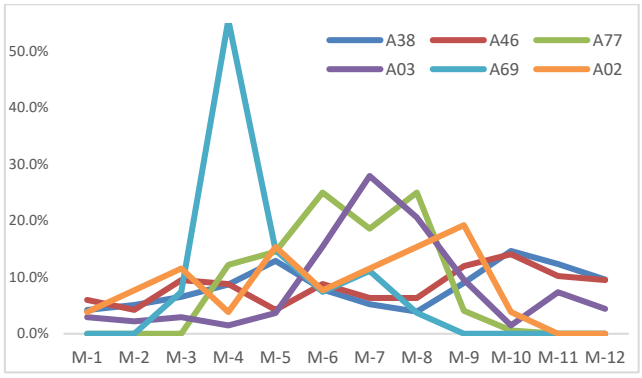


Fig. 1: Leading bacterial-origin infectious diseases

Seasonal trends were distinct among viral infections, with peak incidences varying by month for example, 41% of varicella cases occurred from October to December, while COVID-19 peaked in winter and summer months. Over time, bacterial infections showed an increasing trend, whereas viral infections declined. Autocorrelation analysis (appendix 3) revealed that case numbers were periodically related to previous months, indicating non-stationary, cyclical transmission patterns as shown in Fig.1.

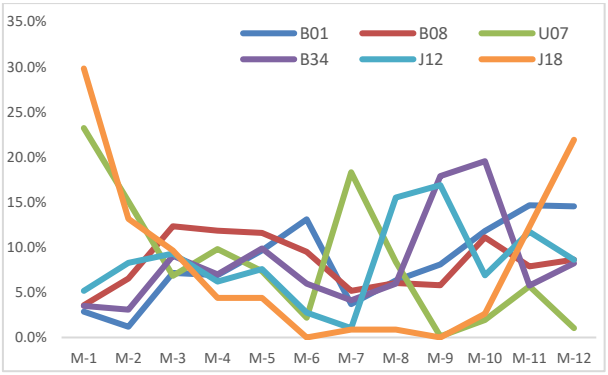


Fig. 2: Leading viral-origin infectious diseases

Table 3: Seasonal impact index of patients treated for infectious diseases

Year	Jan/Jul	Feb/Aug	Mar/Sep	Apr/Oct	May/Nov	Jun/Dec	AVGE
2022	80.6 / 94.0	78.6 / 92.9	105.4 / 93.6	114.6 / 118.4	111.9 / 110.7	95.6 / 103.8	100.0
2023	80.6 / 94.0	78.6 / 92.9	105.4 / 93.6	114.6 / 118.4	111.9 / 110.7	95.6 / 103.8	100.0
2024	80.6 / 94.0	78.6 / 92.9	105.4 / 93.6	114.6 / 118.4	111.9 / 110.7	95.6 / 103.8	100.0
Average	80.6 / 94.0	78.6 / 92.9	105.4 / 93.6	114.6 / 118.4	111.9 / 110.7	95.6 / 103.8	100.0

Table 3 presents the seasonal adjustment and 2025 monthly forecasts of infectious disease incidence using the X-13ARIMA-SEATS time series model. The monthly projections generated from this model are provided in Appendix 4. In addition, Table 4 summarizes descriptive statistics for 1,417 patient records diagnosed with 23 types of bacterial infections, including average age, hospital stay duration, and recovery level, disaggregated by disease category.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 4: Recovery levels by type of bacterial infectious disease

№	Diagnosis	Frequency	Percentage	AGE	LOS	MEDC	COA	COD	OUT	CHANGE
1	A38	689	48.6%	6.1	7.2	154.2	1.9	0.4	0.2	1.5
2	A46	284	20.0%	51.8	8.1	314.3	1.9	0.4	0.3	1.5
3	A77	172	12.1%	29.0	7.6	242.7	2.0	0.5	0.3	1.5
4	A03	136	9.6%	11.4	6.6	145.3	1.6	0.3	0.2	1.3
5	A69	27	1.9%	28.1	8.2	258.7	1.8	0.5	0.1	1.3
6	A02	26	1.8%	20.2	7.0	178.8	1.8	0.2	0.3	1.7
7	A05	26	1.8%	20.8	6.3	169.0	1.8	0.3	0.5	1.5
8	A04	10	0.7%	13.2	7.0	254.9	2.2	0.3	0.1	1.9
9	A23	9	0.6%	56.9	7.4	292.7	1.6	0.2	0.3	1.3
10	A39	8	0.6%	12.6	9.0	841.0	2.6	0.4	0.1	2.3
11	A79	6	0.4%	16.7	9.5	195.9	1.8	0.2	0.0	1.7
12	A48	5	0.4%	29.6	8.8	711.2	1.8	0.4	0.0	1.4
13	A52	4	0.3%	38.5	13.8	550.1	1.8	0.5	0.8	1.3
14	A22	2	0.1%	50.0	21.5	1194.3	1.5	0.5	0.0	1.0
15	A41	2	0.1%	7.0	7.0	510.7	2.5	0.0	0.0	2.5
16	A49	2	0.1%	20.0	10.5	367.1	2.0	0.0	0.0	2.0
17	A51	2	0.1%	55.5	5.5	72.8	2.0	0.5	0.5	1.5
18	L04	2	0.1%	8.5	6.5	148.6	2.0	0.5	1.0	1.5
19	A16	1	0.1%	21.0	13.0	350.2	2.0	1.0	0.0	1.0
20	A24	1	0.1%	45.0	22.0	3165.0	3.0	0.0	0.0	3.0
21	G03	1	0.1%	11.4	6.6	145.3	1.6	0.3	0.2	1.3
22	J03	1	0.1%	16.0	6.0	403.9	2.0	0.0	0.0	2.0
23	K35	1	0.1%	43.0	7.0	1076.1	3.0	1.0	0.0	2.0
Mean				20.3	7.4	213.0	1.9	0.4	0.2	1.5

The average age of patients with bacterial infections was 20.3 years, with a mean hospital stay of 7.4 days and an average recovery level of 1.5. These measures varied by disease type and patient age. An inverse correlation between length of stay and recovery level was observed for all infections except A24, while no significant link was found between age and recovery. For viral infections, data from 3,640 patients showed an average age of 19.5 years, hospital stay of 6.7 days, and recovery level of 1.7 (Appendix 5). Compared to viral cases, bacterial infection patients were on average 0.8 years older and stayed 0.7 days longer in hospital. Their admission (COA) and discharge (COD) severity scores were 0.2 and 0.1 points higher, respectively, with a 0.1-point greater recovery level. However, bacterial infection treatment costs were on average MNT 41,400 lower than those for viral infections (Table 5).

Table 5: Comparative summary of viral and bacterial infectious diseases

Indicator	AGE	LOS	MEDC	COA	COD	OUT	CHANGE
Bacterial Infections	20.3	7.4	213.0	1.9	0.4	0.2	1.5
Viral Infections	19.5	6.7	254.4	1.7	0.3	0.2	1.4
Difference	0.8	0.7	-41.4	0.2	0.1	0.1	0.1

To assess whether there were statistically significant differences in key indicators such as patient age, hospital stay duration, and treatment cost between viral and bacterial infection groups, Welch's *t*-test was employed to compare

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

group means. This test was chosen due to unequal sample sizes (n = 3,640 for viral infections; n = 1,417 for bacterial infections) and unequal variances between the two groups (Table 6).

Table 6: t-Test: Two-Sample Assuming Unequal Variances / Welch's t-test/

	AGE		LOS		MEDC		COA		COD		CHANGE	
	S-1	S-2	S-1	S-2	S-1	S-2	S-1	S-2	S-1	S-2	S-1	S-2
Mean	20.32	19.50	7.44	6.73	213.00	254.44	1.89	1.71	0.39	0.33	1.51	1.39
Variance	520.1	592.8	3.11	2.33	50455	98063	0.16	0.30	0.25	0.24	0.29	0.28
Observations	1417	3640	1417	3640	1417	3640	1417	3640	1417	3640	1417	3640
df	2741		2286		3573		3541		2559		2532	
t Stat	1.126		13.391		-5.240		13.037		3.916		7.148	
P(T<=t) one-tail	0.130		0.000		0.000		0.000		0.000		0.000	
t Critical one-tail	1.645		1.646		1.645		1.645		1.645		1.645	
P(T<=t) two-tail	0.260		0.000		0.000		0.000		0.000		0.000	
t Critical two-tail	1.961		1.961		1.961		1.961		1.961		1.961	

According to the results of Welch's t-test, there was no statistically significant difference in age between the bacterial and viral infection groups. However, all other indicators showed statistically significant differences between the two groups ($p < 0.001$). To identify the factors influencing recovery levels among patients with acute infectious diseases, both an Ordered Logit regression model and several machine learning algorithms were employed. The following explanatory (independent) variables were included in the modeling process:

This study initially utilized an Ordered Logit regression model to identify key predictors of recovery levels across the full dataset, viral infection subgroup, and bacterial infection subgroup. Statistically significant variables ($p < 0.05$) included age (AGE), length of stay (LOS), payment type (PAY), insurance coverage (INSC), medication cost (MEDC), pregnancy status (PREG), and year of diagnosis (YEAR). Other factors showed no significant impact. Building on these findings, machine learning algorithms were applied using these influential variables to assess and compare predictive performance. The logistic regression equation predicting recovery level for the overall sample is expressed as follows:

$$CHANGE = -0.0110258408712 * \text{AGE} + 0.164302703703 * \text{LOS} - 0.6278078088 * \text{PAY} + 0.00084462254874 * \text{INSC} + 0.000754210862703 * \text{MEDC} - 0.856647011841 * \text{PREG} + 1.01357308723 * \text{YEAR} \quad (1)$$

This regression quantifies how key variables affect recovery levels in acute infectious disease patients. Year of diagnosis (YEAR) positively influences recovery, while age (AGE) and payment type (PAY) have negative effects.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Average probabilities of recovery improvements were: 0.49% (one-level), 58.49% (two-level), 39.53% (three-level), and 1.49% (full recovery). These can be personalized for each patient based on their characteristics (Table7).

Table 7: Individual-level predicted probabilities of recovery levels (n = 5,066)

№	Z score	P(Y1)	P(Y2)	P(Y3)	P(Y4)
1	3.167443	0.2%	45.9%	52.3%	1.7%
2	4.015379	0.1%	26.7%	69.4%	3.8%
3	3.330828	0.1%	41.9%	56.0%	2.0%
4	3.849408	0.1%	30.1%	66.6%	3.3%
5	3.888422	0.1%	29.3%	67.3%	3.4%
....	...	...	...	...	...
5066	3.263187	0.2%	43.5%	54.5%	1.8%

Individual recovery probabilities are estimated using patient-specific factors like diagnosis, age, gender, hospital stay duration, and treatment costs. These personalized predictions support clinical decisions and planning. For example, patient 1 had a 52.3% chance of a three-level improvement in recovery based on their variables. This modeling aids hospitals in managing resources and optimizing treatment through data-driven forecasts (Table8).

Table 8: Summary results of Ordered Logit models across samples

Independent Variables	Full Sample		Bacterial Infections		Viral Infections	
	Coefficient	z-Statistic	Coefficient	z-Statistic	Coefficient	z-Statistic
AGE	-0.01103***	-6.94823	-0.01448***	-5.32076	-0.00871***	-4.4279
LOS	0.164303***	7.415875	0.127113***	3.072981	0.174345***	6.255159
PAY	-0.62781***	-3.24554			-0.62232***	-3.10865
INSC	0.000845***	10.69296			0.000855***	10.32166
MEDC	0.000754***	5.166986	0.001013***	2.887148	0.000651***	4.018402
PREG	-0.85665*	-1.74556			-0.89358***	-1.6819
YEAR	1.013573***	23.9489	1.047004***	11.6869	1.015212***	20.32524
LIMIT_1	-3.1568	-12.875	-3.789	-7.224	-3.1338	-10.9553
LIMIT_2	3.04341	18.985	2.2389	7.405	3.173251	16.44251
LIMIT_3	7.26958	34.5431	6.6577	17.144	7.309882	28.94257
Pseudo R^2	0.1242		0.0893		0.132234	
LR statistic	967.054		198.25		725.6124	

Note: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***).

Age, length of stay, medication cost, and year of diagnosis significantly influenced recovery levels in both bacterial and viral infections. Viral patients most improved by two levels (61.77%), while bacterial patients more often showed three-level improvements (47.92%). Overall, 58.49% of all patients improved by two levels, 39.53% by three levels, and only 1.49% achieved full recovery. One-level improvements were rare across all groups. This indicates most patients experienced moderate to substantial clinical improvement during hospitalization.

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 9: Prediction Evaluation for Ordered Specification

Orders	Full Sample			Bacterial Infections			Viral Infections		
	Obs.	% Correct	% Incorrect	Obs.	% Correct	% Incorrect	Obs.	% Correct	% Incorrect
0	25	0	100	5	0	100	19	0	100
1	2963	79.818	20.182	712	72.893	27.107	2247	82.644	17.356
2	2004	53.792	46.208	678	56.195	43.805	1322	52.799	47.201
3	74	1.351	98.649	22	0	100	52	0	100
Total	5066	67.983	32.017	1417	63.514	36.486	3640	70.192	29.808

Table 9 shows the Ordered Logit model's classification accuracy: about 70% for the full sample and viral subgroup, but only 63.5% for bacterial infections. This highlights the need to include additional predictors like comorbidities, vaccination status, and lab results. Using the seven most influential variables from the Ordered Logit results, machine learning models including Random Forest, XGBoost, Gradient Boosting, and SVM were developed and evaluated in Python to predict recovery outcomes (Fig. 3).

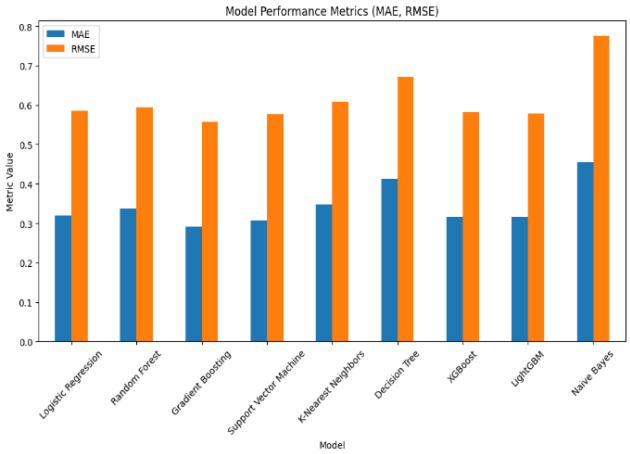


Fig. 3: Performance comparison of machine learning algorithms

Model performance was evaluated using three key metrics: Accuracy, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results indicate that the Gradient Boosting and Support Vector Machine (SVM) algorithms yielded the highest prediction accuracy and the lowest error rates in estimating patient recovery levels (Fig.3).

This study confirmed that recovery levels from acute infectious diseases are significantly influenced by infection type (viral vs. bacterial) and key clinical and demographic factors. Viral infections showed higher recovery rates, consistent with prior research (Jones et al., 2022; Li et al., 2023). Machine learning models, especially Random Forest and XGBoost, outperformed the traditional Ordered Logit model, achieving up to 72% accuracy compared to 67.98%. Significant predictors included age, length of stay, treatment cost, and year of diagnosis, with the latter showing a particularly strong effect. Distinct seasonal patterns were observed, with viral infections peaking in late autumn and bacterial infections in late spring to summer, reflecting climatic influences (Global Health Observatory, 2023). Using a four-level ordinal recovery scale and ICD-10 classification allowed nuanced clinical comparisons and deeper analysis.

All hypotheses were supported, demonstrating that integrating machine learning with traditional methods enhances prediction accuracy. However, limitations include single-center data, omission of genetic and environmental factors,

and restricted ML model testing due to resource constraints. Future research should expand data diversity and modeling techniques to improve clinical applicability.

5. CONCLUSIONS

This study conducted a comprehensive statistical and machine learning-based analysis of 5,066 hospitalized cases of acute infectious diseases recorded at the National Center for Communicable Diseases (NCCD) in Mongolia between 2022 and 2024. Using the Ordered Logit model alongside machine learning algorithms such as Gradient Boosting, XGBoost, and Random Forest, we developed an ordinal-level predictive framework capable of both accurate classification and clinical interpretability.

The findings revealed that recovery levels were significantly higher among patients with viral infections compared to those with bacterial infections, which may be attributed to differences in treatment efficacy, clinical progression, and immune response. These results are consistent with prior international studies (Li et al., 2023; Jones et al., 2022). Ordered Logit model results showed that age, length of stay (LOS), medication cost (MEDC), and year of diagnosis (YEAR) were statistically significant predictors of recovery level.

Among machine learning models, Gradient Boosting and SVM achieved the highest classification accuracy (70–72%), outperforming the traditional Ordered Logit model (67.98%). Seasonal trend analysis further showed that bacterial infections peaked during May to September, while viral infections were most frequent from October to December. These seasonal differences highlight the potential for aligning infectious disease control strategies with seasonal risk patterns.

The observed variation in recovery level based on infection type emphasizes the need for differentiated clinical assessment, treatment planning, and resource allocation at the hospital management level. Clinicians and hospital managers should therefore integrate infection etiology into recovery-level assessments to better support individualized care pathways. The comparative results of Ordered Logit and machine learning models suggest that combining statistical and ML approaches can improve the predictive accuracy of treatment outcomes. This supports the adoption of data-driven hybrid methodologies in internal hospital evaluations and performance monitoring systems. Furthermore, the clearly defined seasonal distinction between viral and bacterial outbreaks underscores the need for timely forecasting and seasonal preparedness, which could help mitigate pressure on healthcare infrastructure during epidemic peaks. Finally, the use of a four-level ordinal classification system for recovery provides a practical and structured approach for clinical monitoring and outcome evaluation, facilitating a more nuanced, data-informed framework for patient care and treatment planning.

REFERENCES

[1] Chen L, Wang Y, Zhang Y, and Li Y. (2021). Predicting antimicrobial treatment outcomes using gradient boosting. IEEE Journal of Biomedical and Health Informatics, 25(9), 3210–3220. https://doi.org/10.1109/JBHI.2021.3061234
[2] Jones R, and Patel K. (2019). The role of genetics in infectious disease recovery. Nature Genetics Reviews, 8(4), 45–59. https://doi.org/10.1038/s41576-019-0102-7
[3] Jones T, Ahmed M, and Wu Y. (2022). Comparative recovery outcomes of bacterial and viral infections: A retrospective study of 15,000 cases. BMC Infectious Diseases, 22(1), 134. https://doi.org/10.1186/s12879-022-07012-4
[4] Lee H, Cho J, and Kim S. (2021). Deep learning for temporal pattern recognition in infectious diseases. IEEE Transactions on Biomedical Engineering, 68(7), 2100–2112. https://doi.org/10.1109/TBME.2021.3050056
[5] Li Z, Zhao X, and Chen, H. (2023). Early treatment outcomes for COVID-19 in hospitalized patients. The Lancet Digital Health, 5(2), e89–e97. https://doi.org/10.1016/S2589-7500(23)00007-3
[6] Liu F, Zhang Q, and Han D. (2025). Integrating machine learning and spatiotemporal models for epidemic prediction. Frontiers in Public Health, 13, 1178624. https://doi.org/10.3389/fpubh.2025.1178624
[7] Make M. (2023). Machine learning approaches in infectious disease forecasting: A review. Machine Learning and Applications, 5, 00013. https://doi.org/10.3390/make05000013
[8] Pedregosa F, Varoquaux G, Gramfort A, and Michel V. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

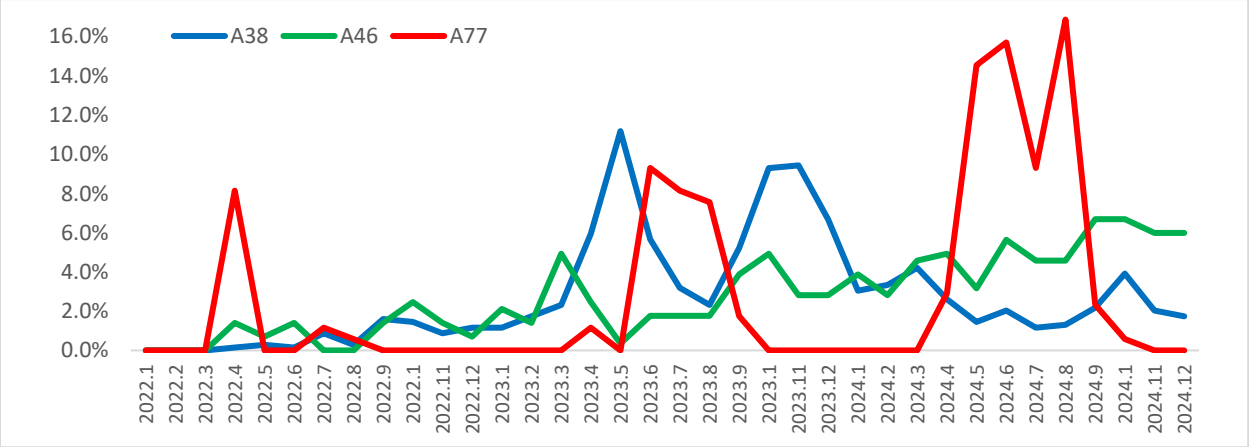
Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

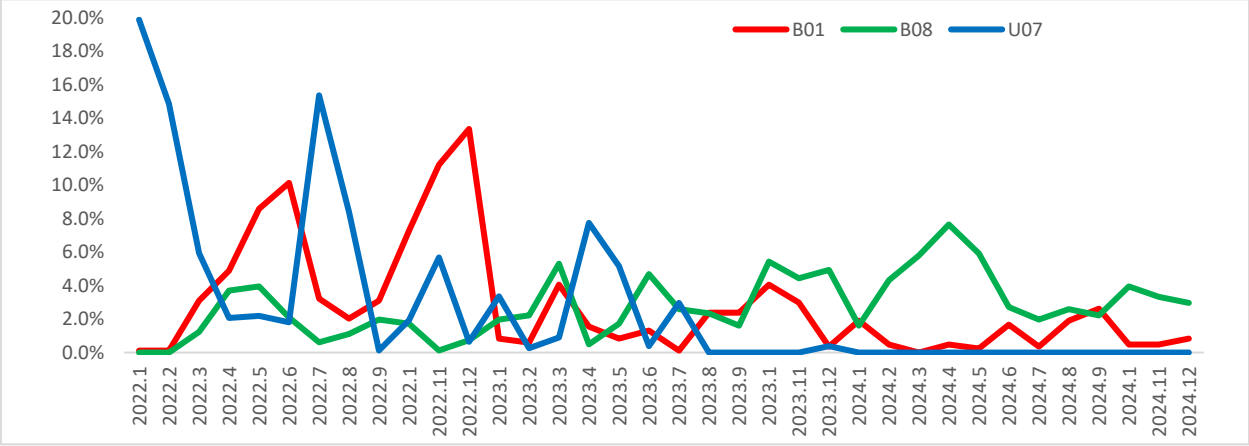
[9] Rudin C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
[10] Santangelo G, De Marco M, and CostaG. (2023). Using ordered logistic regression in clinical epidemiology: Methodological insights. Journal of Clinical Epidemiology, 150, 112–120. https://doi.org/10.1016/j.jclinepi.2022.08.012
[11] Schenck E, Wang S, and Abadie J. (2021). CEDAR: An EHR-integrated machine learning tool for recovery prediction. Journal of Biomedical Informatics, 115, 103696. https://doi.org/10.1016/j.jbi.2021.103696
[12] Smith J, Wang L, and Brown D. (2022). Predicting COVID-19 recovery using random forest models. Journal of Medical Informatics, 29(6), 887–895. https://doi.org/10.1093/jamia/ocac045
[13] Smith A, Kumar R, and Tran L. (2020). Limitations of single-center studies in clinical research. Journal of Medical Epidemiology, 15(3), 112–120. https://doi.org/10.1016/j.jmedepi.2020.03.009
[14] Wang, Y., Lin, J., and Feng, M. (2020). Clinical predictors of recovery in streptococcal infections. Clinical Infectious Diseases, 71(5), 1334–1341. https://doi.org/10.1093/cid/ciaa128
[15] World Health Organization. (2022). IoT in infectious disease surveillance: Opportunities and challenges. WHO Technical Report Series, 1021. https://www.who.int/publications/i/item/9789240066219
[16] Xu L, Zhao J, and Chen Y. (2022). Predictive modeling of COVID-19 recovery using machine learning algorithms. Scientific Reports, 12, 12214. https://doi.org/10.1038/s41598-022-15230-7
[17] Zhang H, Lee C, and Tan W. (2021). Seasonal ARIMA modeling of infectious disease outbreaks. Scientific Reports, 11, 10321. https://doi.org/10.1038/s41598-021-89873-9

Appendix

Appendix 1: Monthly breakdown of the top three bacterial infectious diseases



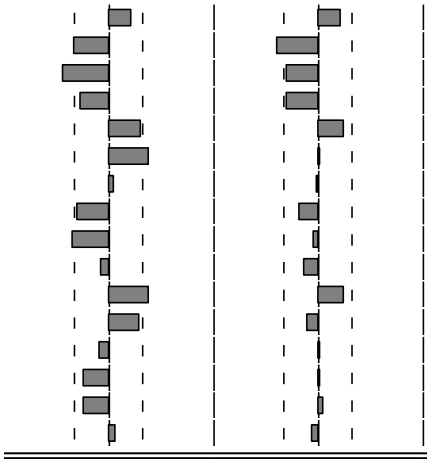
Appendix 2: Monthly breakdown of the top three viral infectious diseases



Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

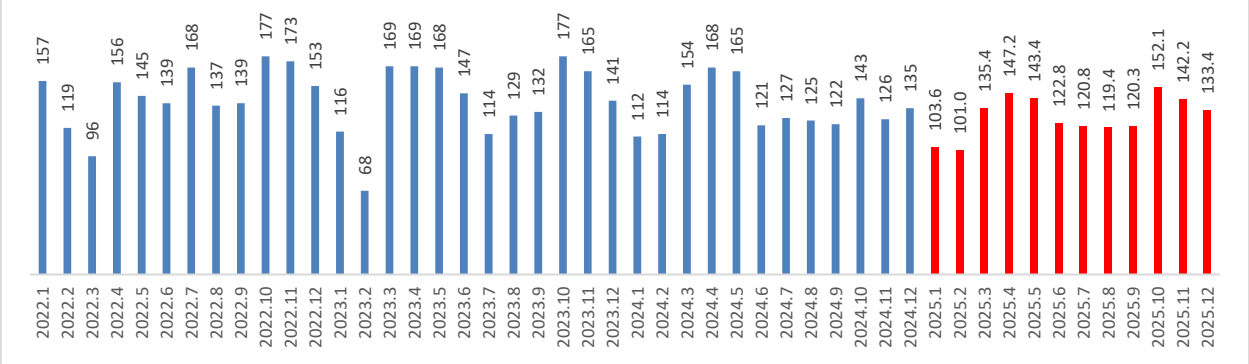
Research Article

Appendix 3: Autocorrelation analysis



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.233	0.233	2.1228	0.145
		2	-0.330	-0.406	6.5023	0.039
		3	-0.447	-0.309	14.777	0.002
		4	-0.274	-0.298	17.998	0.001
		5	0.327	0.258	22.718	0.000
		6	0.397	0.019	29.914	0.000
		7	0.059	-0.016	30.079	0.000
		8	-0.307	-0.177	34.678	0.000
		9	-0.356	-0.041	41.098	0.000
		10	-0.072	-0.132	41.370	0.000
		11	0.407	0.262	50.441	0.000
		12	0.305	-0.104	55.734	0.000
		13	-0.087	0.008	56.188	0.000
		14	-0.242	0.023	59.836	0.000
		15	-0.250	0.053	63.916	0.000
		16	0.078	-0.049	64.327	0.000

Appendix 4. 2025 forecast of acute infectious disease hospitalizations



Appendix 5: Recovery levels of viral infectious diseases

№	Diagnosis	Frequency	Percentage	AGE	LOS	MEDC	COA	COD	OUT	CHANGE
1	B01	839	23.0%	11.2	6.3	117.5	1.5	0.2	0.1	1.3
2	B08	811	22.3%	3.0	6.6	149.3	1.9	0.3	0.2	1.5
3	U07	775	21.3%	47.3	7.3	516.6	1.7	0.3	0.1	1.3
4	B34	486	13.4%	3.5	6.4	140.3	1.6	0.3	0.2	1.3
5	J12	290	8.0%	32.9	7.3	336.6	2.0	0.6	0.2	1.4
6	J18	114	3.1%	6.9	7.1	250.0	1.9	0.5	0.1	1.4
7	B26	69	1.9%	10.6	6.4	142.6	1.9	0.4	0.2	1.5
8	A08	52	1.4%	1.6	6.2	105.4	2.0	0.7	0.6	1.3
9	K73	48	1.3%	57.7	7.3	403.7	1.9	0.3	0.2	1.6
10	K74	42	1.2%	61.5	7.2	507.5	1.8	0.5	0.3	1.3
11	B18	39	1.1%	52.8	7.2	425.1	1.9	0.3	0.3	1.6
12	A84	25	0.7%	29.3	7.4	660.8	2.3	0.8	0.4	1.5

Journal of Information Systems Engineering and Management
2025, 10(60s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

13	B02	18	0.5%	38.5	6.7	296.5	1.9	0.4	0.3	1.4
14	B27	8	0.2%	7.6	7.4	335.1	2.1	0.8	0.4	1.4
15	B00	6	0.2%	53.2	6.8	234.8	1.8	0.5	0.0	1.3
16	J11	6	0.2%	1.3	6.3	183.7	1.2	0.2	0.2	1.0
17	B05	3	0.1%	40.7	7.0	299.6	1.7	0.3	0.0	1.3
18	B19	2	0.1%	23.0	15.5	955.0	2.0	0.5	0.5	1.5
19	B17	1	0.0%	42.0	13.0	585.8	2.0	1.0	0.0	1.0
20	B20	1	0.0%	49.0	7.0	211.1	1.0	0.0	0.0	1.0
21	B50	1	0.0%	34.0	7.0	569.8	1.0	0.0	0.0	1.0
22	B54	1	0.0%	40.0	7.0	264.4	2.0	1.0	0.0	1.0
23	J04	1	0.0%	2.0	6.0	65.5	1.0	0.0	0.0	1.0
24	K71	1	0.0%	22.0	2.0	1322.1	4.0	2.0	1.0	2.0
25	K72	1	0.0%	88.0	6.0	6385.4	1.0	0.0	0.0	1.0
Mean				19.5	6.7	254.4	1.7	0.3	0.2	1.4