Machine Learning-Based Prediction of 'real-world' Mortality and Prolonged Hospitalization in Upper Respiratory Infection Patients: A Retrospective Analysis and insights for policy makers

Doron Zamir MD[1,2], Yaniv S. Ovadia PhD RD[1], Matan Leventer BSc[2], Maxwell Swartz[3], Nadav Rapoport[2]

[1]Barzilai University Medical Center, Ashkelon, Israel; [2]Ben-Gurion University of Negev, Beer-Sheva, Israel;
[3]Tulane University School of Medicine, New Orleans, Louisiana   nadavrap@bgu.ac.il

BACKGROUND & OBJECTIVES

- Upper respiratory infections (URIs), including COVID-19, pose significant challenges to global health systems
- Early identification of factors predicting prolonged hospitalization and mortality is crucial for effective management

Objectives:
(a) Identify critical factors influencing clinical outcomes
(b) Support healthcare providers in resource allocation

METHODS

Dataset: 56,554 adult patients (2020-2023)
Admitted to Barzilai Medical Center

Machine Learning Models:
Random Forest, Gradient Boosting, k-NN, Decision Trees, MLP, Logistic Regression

Outcomes:
1. Mortality (in-hospital or 30-day post-discharge)
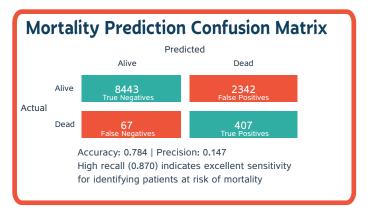2. Prolonged hospitalization (≥5 days)

KEY RESULTS

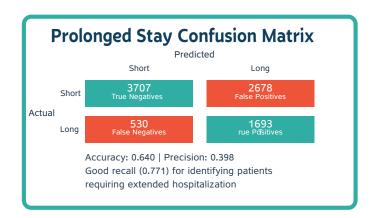🏆 Gradient Boosting: Best Performing Model for Both Outcomes

Mortality Prediction:
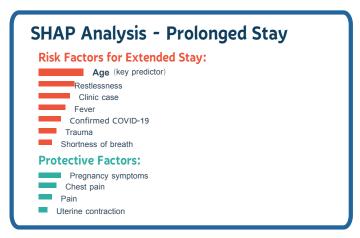ROC-AUC: 0.901 (0.894-0.909) | PR-AUC: 0.640 | Recall: 0.870

Prolonged Hospitalization:
ROC-AUC: 0.737 (0.731-0.743) | F1-Score: 0.525 | Recall: 0.771

Mortality Prediction Confusion Matrix

	Predicted	
	Alive	Dead
Actual Alive	8443 True Negatives	2342 False Positives
Actual Dead	67 False Negatives	407 True Positives

Accuracy: 0.784 | Precision: 0.147
High recall (0.870) indicates excellent sensitivity for identifying patients at risk of mortality

Prolonged Stay Confusion Matrix

	Predicted	
	Short	Long
Actual Short	3707 True Negatives	2678 False Positives
Actual Long	530 False Negatives	1693 True Positives

Accuracy: 0.640 | Precision: 0.398
Good recall (0.771) for identifying patients requiring extended hospitalization

SHAP Analysis - Mortality Prediction

Key Risk Factors (Positive SHAP):
- Age (most significant)
- Mobile intensive care unit
- Number of hospitalizations
- Pneumonia
- Function decline
- Shortness of breath
- COVID department

Protective Factors (Negative SHAP):
- Pain
- Chest pain
- Headache
- Fall, trauma, cut

SHAP Analysis - Prolonged Stay

Risk Factors for Extended Stay:
- Age (key predictor)
- Restlessness
- Clinic case
- Fever
- Confirmed COVID-19
- Trauma
- Shortness of breath

Protective Factors:
- Pregnancy symptoms
- Chest pain
- Pain
- Uterine contraction

Prospective Validation

Dataset: 275 patients (Sept 2023 - June 2025)
Mortality: ROC-AUC: 0.87, Recall: 0.94, F1: 0.37
Prolonged Stay: ROC-AUC: 0.49, F1: 0.55, Recall: 0.82

Conclusion & Implications

- ML effectively predicted mortality and prolonged hospitalizaiton
- Age was the most significant predictor for both outcomes
- Unexpected protective of headache and chest pain