In-Hospital Mortality Prediction on SUPPORT2

Xinxin Zhou

Data Science Institute | Brown University

Github: https://github.com/NoraZhouXX/1030-Final-Project-SUPPORT2

12/12/2025

Midterm Recap

Problem:

- Predict whether a patient will die in-hospital after day 3 using data available by the end of day 3.

Why day 3?

- Data are routinely available by the end of day 3.
- Still early enough for triage, goals-of-care, and resource planning.

Task type:

- Binary classification

Collected From:

- UCI ML Repository: ICU/IMCU patients from multiple U.S. hospitals.

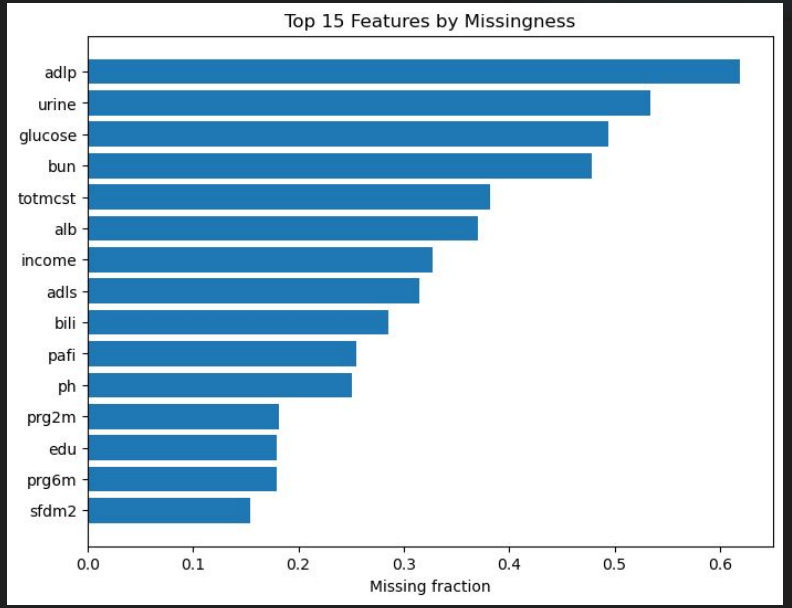
Dataset:

- 9105 rows; 48 columns (mix of continuous, categorical, ordinal).

Midterm Recap

Difficulty & Preprocessing:

- 33/48 have missing values
- Excluded (post-day-3): costs/charges, death, surv2m/6m, `d.time`, etc.
- Categorical Features: SimpleImputer + OneHot Encoder
- Ordinal Features: SimpleImputer + Ordinal Encoder
- Continuous Features: StandardScaler + XGBoost + Pattern-reduced features



Cross-validation & data split

- Repeated random split (5 seeds) to capture uncertainty from splitting
- Within each split: train & val used for model selection, held-out test for final evaluation only (60-20-20)
- Evaluation Metrics: ROC-AUC
    - robust to class imbalance

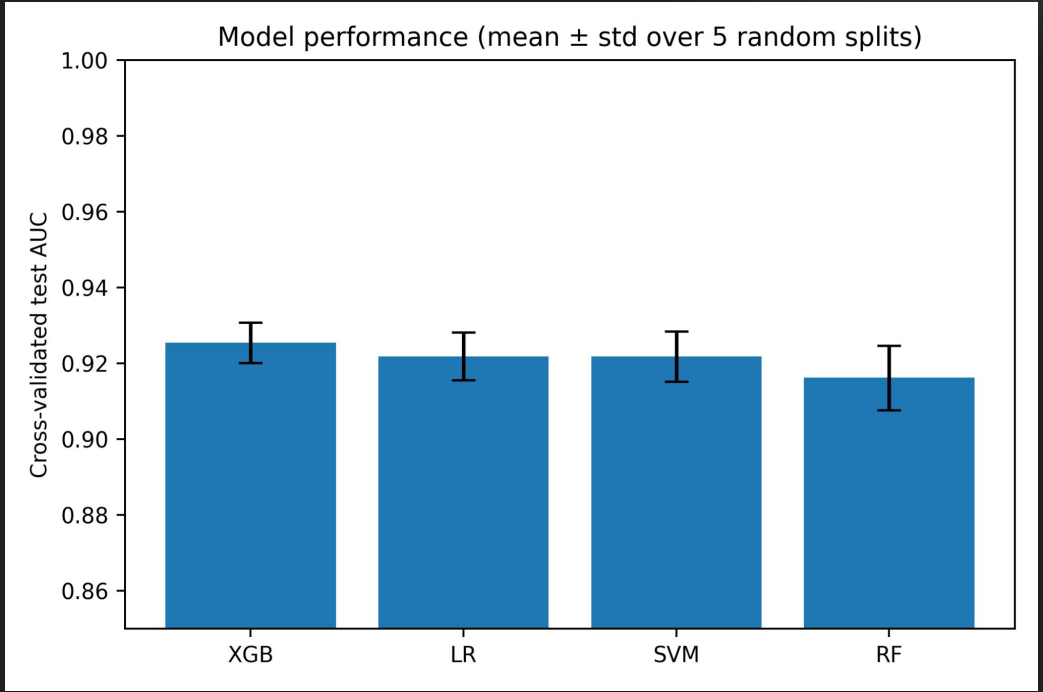
Models & hyperparameters

Model	Missing Handling	CV Strategy	Hyperparameter Tuned
Logistic Regression	Pattern-reduced	Manual grid-search	C ∈ {0.01, 0.1, 1, 10, 100}
Linear SVM	Pattern-reduced	Manual grid-search	C ∈ {0.01, 0.1, 1, 10, 100}
Random Forest	Pattern-reduced	Manual grid-search	max_depth = {1, 3, 10, 30, 100}, max_features = {0.5, 0.75, 1.0}
XGBoost	Native missing	5-fold KFold + early stopping	max_depth ∈ {1, 3, 10, 30, 100}, reg_alpha ∈ {0.1, 1, 10}, reg_lambda ∈ {0.1, 1, 10}, n_estimators = 10000 with early stopping

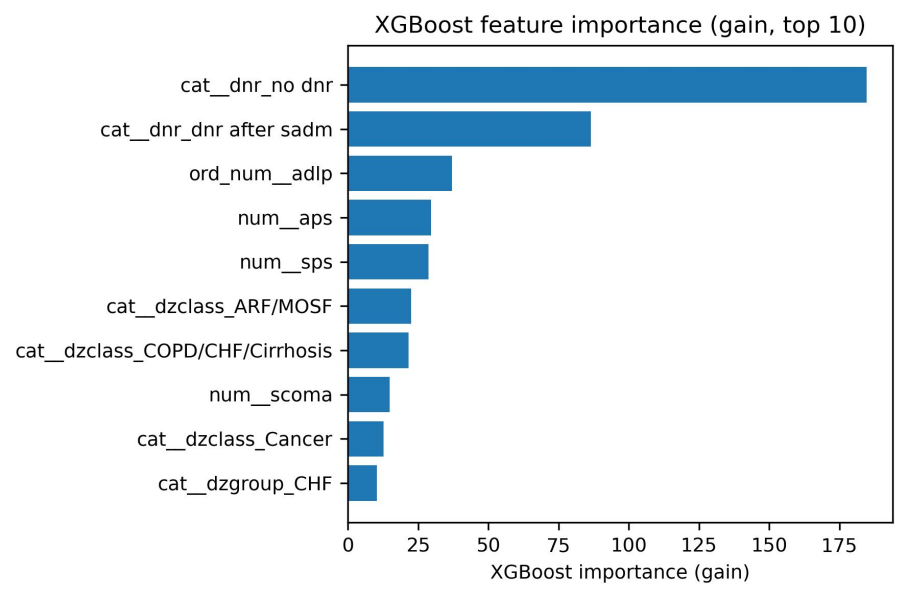
Result

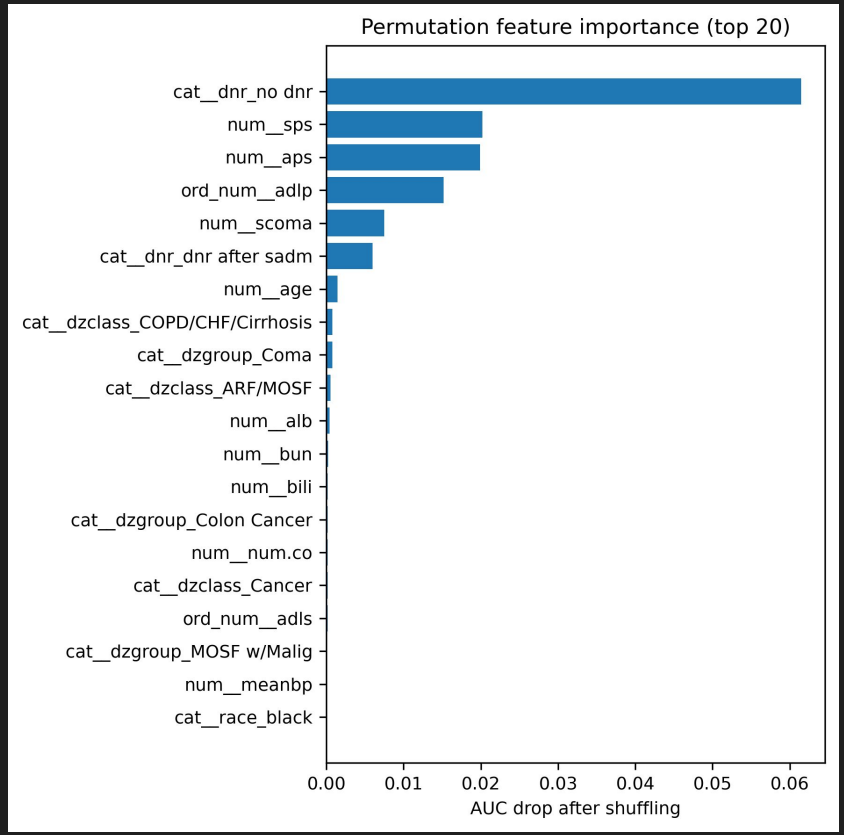
- Baseline majority: AUC: 0.50
- All models: AUC ~0.92
- XGB slightly better AUC with lowest variance across splits

Model	AUC (mean)	AUC (std)
LR	0.9219	0.0063
SVM	0.9219	0.0066
RF	0.9162	0.0085
XGBoost	0.9255	0.0053



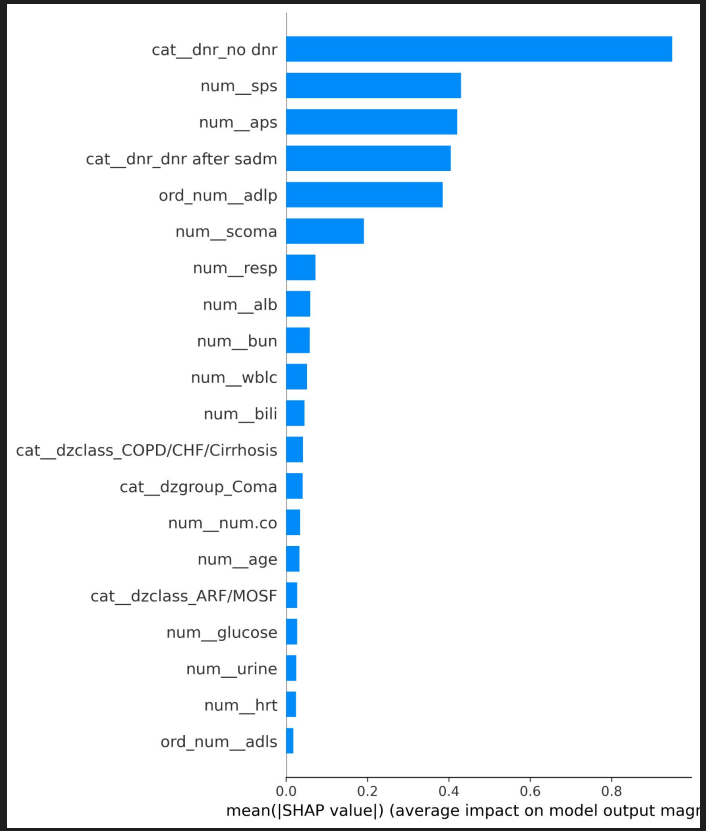
Global importance: XGB vs permutation

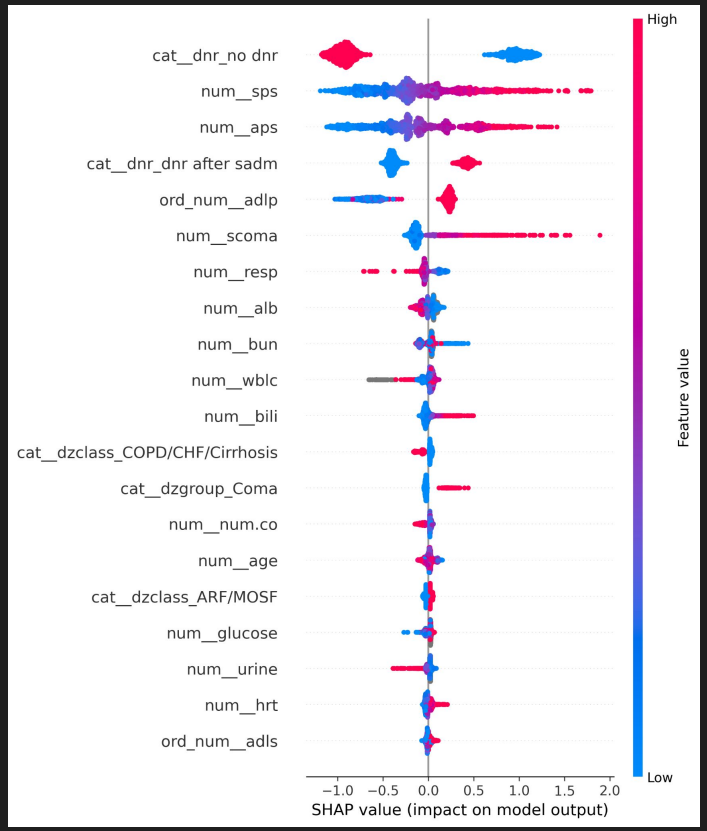




- DNR dominates across both metrics
- Acute physiology (aps, scoma, sps) and functional status (aldp) consistently top 10
- Age is important but less than acute indicators once physiology is included

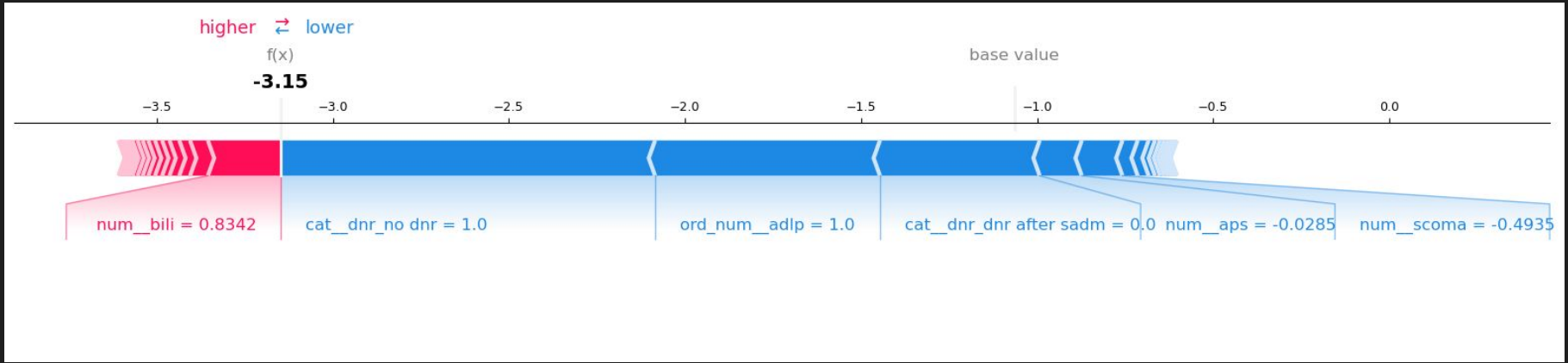
Global importance: SHAP summary





SHAP global importance largely agrees with XGBoost's gain-based and permutation importances

Local explanation: SHAP force plot



- This SHAP force plot shows a low-risk patient, f(x) well below the baseline, so predicted death risk is very small.
- A slightly high bilirubin pushes risk up a bit (red).
- Several features pull risk down strongly (blue): no DNR order, better functional status (`aldp`), and relatively mild APS and coma scores.

Results

- Model consistently places DNR as top feature, which is clinically sensitive but signals strong association with in-hospital mortality

- Acute physiology (aps, scoma, pafi, bun…) and functional status (aldp) behave as expected and dominate risk once measured

- Age and some chronic conditions become relatively less important once we account for acute physiology and DNR

Outlook

- Try Better Models
  - Gradient boosting with monotonic constraints
- Collect more patient data / engineer with the current features
- Deployment Direction:
  - Convert model into a bedside risk score

QUESTIONS ?

Thank you!

Reference

The SUPPORT Principal Investigators. (1995). A controlled trial to improve care for seriously ill hospitalized patients:
The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT).
Journal of the American Medical Association, 274(20), 1591–1598.
https://doi.org/10.1001/jama.1995.03530200027032