Predicting Rupture Risk in Cerebral Arteriovenous Malformations

Salma Habib[1] · Niranjani Komethagavel[1] · Dhivya Pradeep[1] · Jiaqiang Wang[1] · Jaden Tepper[1] · Annaka Saffron[1] · Hailey Lee[1] · Amanda Kwok[1] · Sidharth Raghavan[1] · Iyinoluwa Tugbobo[1] · Beiya Xu[1] · Joseph L Greenstein[1] · Casey Overby Taylor[1] · Siamak Ardekani[1] · James Feghali[2] · Ali Uneri[1]

[1]Johns Hopkins Whiting School of Engineering · [2]Johns Hopkins School of Medicine, Baltimore MD

JOHNS HOPKINS MEDICINE · THE JOHNS HOPKINS HOSPITAL · JOHNS HOPKINS WHITING SCHOOL of ENGINEERING

Introduction

Brain arteriovenous malformations (bAVMs) are complex vascular lesions where arteries connect directly to veins, bypassing the capillary network. They are the **leading cause of intracerebral hemorrhage** in young adults. Existing clinical scoring systems fail to capture non-linear feature interactions or predict hemorrhage timing.

2 - 4% Annual Hemorrhage Risk

20% Mortality Post-hemorrhage

40% Functionally Impaired in 1 Yr

Current Standard: R²eD AVM Score (Feghali, 2019)
Predict hemorrhagic presentation (cross-sectional only). Does not estimate timing of future rupture
AUC=0.685, 5 features (race, deep location, AVM size, deep venous drainage, monoarterial supply)

Objectives

We aim to improve rupture risk prediction in **bAVMs** by applying **machine learning to a large IRB-approved clinical cohort (~1,065 patients)** and addressing two fundamental clinical questions.

1. **Will this patient rupture?** *(classification)*
2. **When is the rupture risk highest?** *(time-to-event analysis)*

Patient Cohort

Classification Cohort

- Johns Hopkins Hospital bAVM Registry, N = 1065
  - Ruptured, N = 436
    - Re-ruptured, N = 46
  - Unruptured, N = 629
    - Ruptured during follow-up, N = 47
    - Remained unruptured, N = 582

Time to Event Analysis Cohort

Patients without follow-up data were excluded.

Initial Status	Patients (N)	Rupture Events (during follow-up)	Median Follow-Up
Unruptured	617	39	0.68 yr
Ruptured	405	40	0.44 yr
Total	1,022	79	0.54 yr

Methods

AIM 1 Classification Will this patient rupture?	AIM 2 Time-to-Event When is risk highest?
Iterative feature addition beyond R²eD features (tested combinations to maximize AUC. 5-fold cross-validation throughout)	Exhaustive combinatorial search with **480K–600K** subsets (size 2–12), no univariate gate. Top 50 confirmed with nested CV (**5×3 folds**).

MODELS

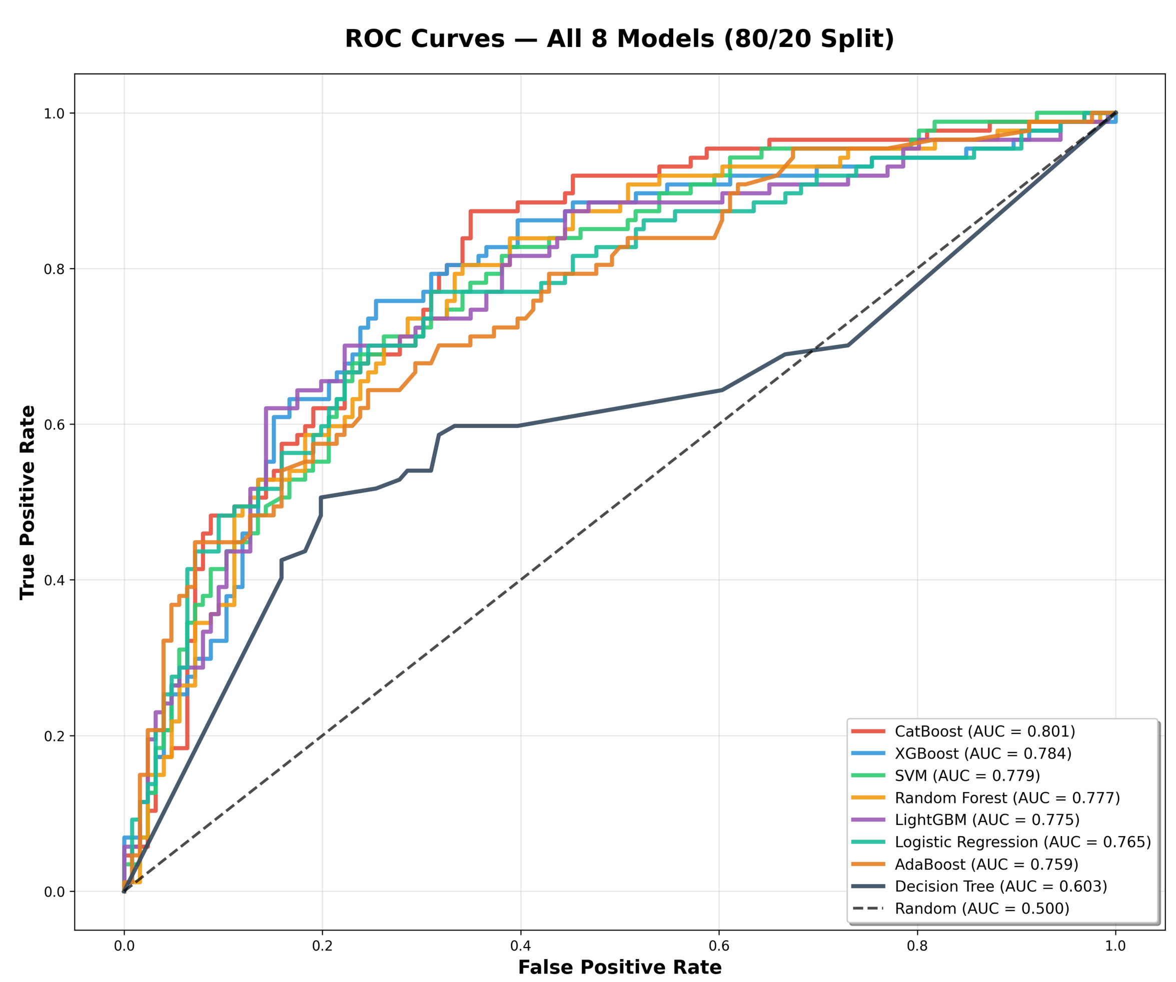
• CatBoost • SVM • LightGBM • Logistic Regression	• Random Forest • AdaBoost • XGBoost • Decision Tree	• Cox PH • Ridge Cox • Gradient Boosting	• Extra Survival Trees • Unified Ridge Cox (HRs) • Random Survival Forest

EVALUATION

AUC, Sensitivity, Specificity, Brier, SHAP	C-index, Calibration, DCA

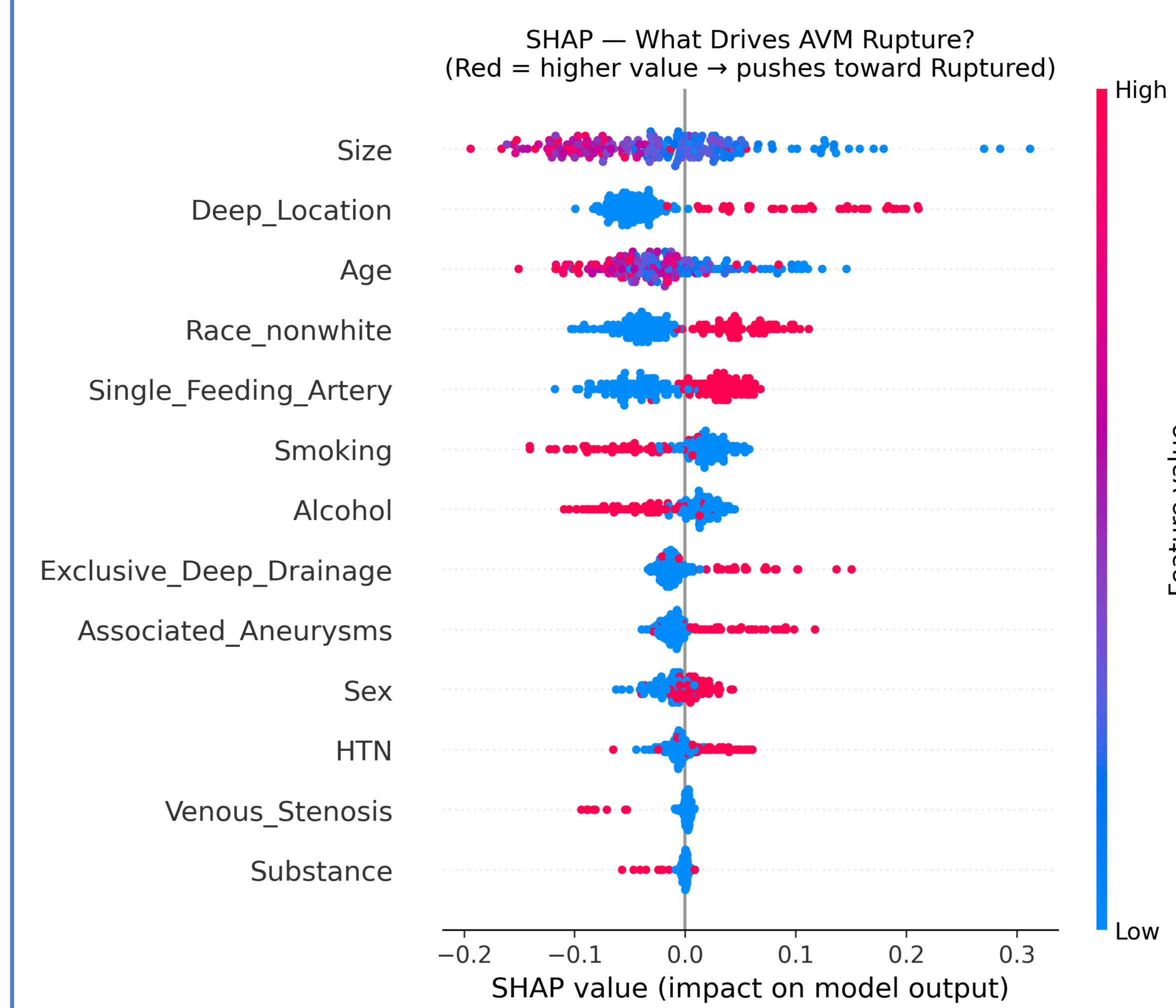
Classification

ROC Curve



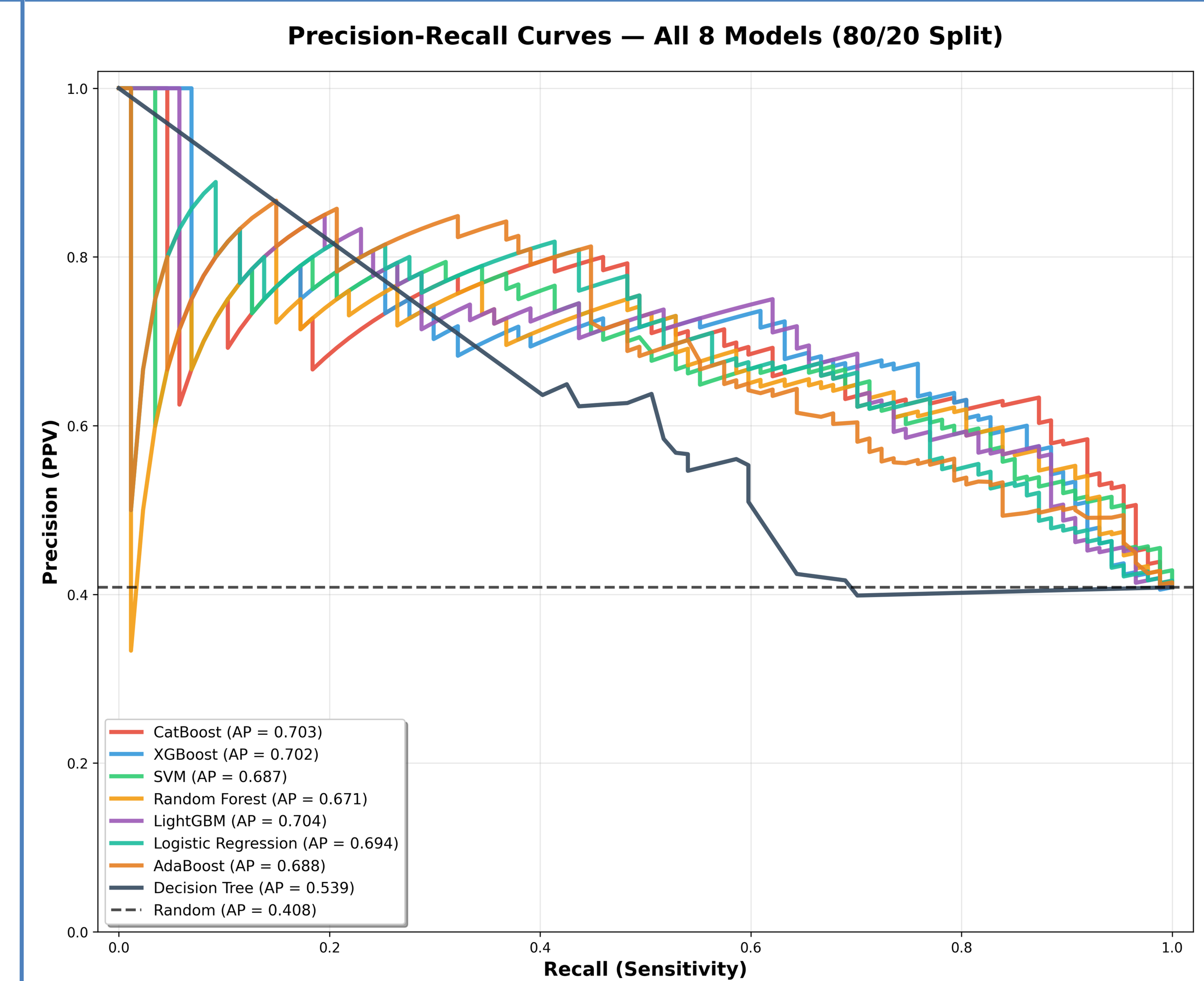
Receiver operating characteristic curves for all eight classical ML classifiers on the 20% held-out test set (n=213). CatBoost achieved the highest discrimination (AUC=0.801), representing an absolute improvement of 11.4 percentage points over the R²eD baseline (AUC=0.687; DeLong p=0.034).

SHAP Analysis



Directional feature importance for the CatBoost classifier. The top five predictors were nidus size (smaller → higher rupture probability), deep location, age at diagnosis, non-White race, and single feeding artery.

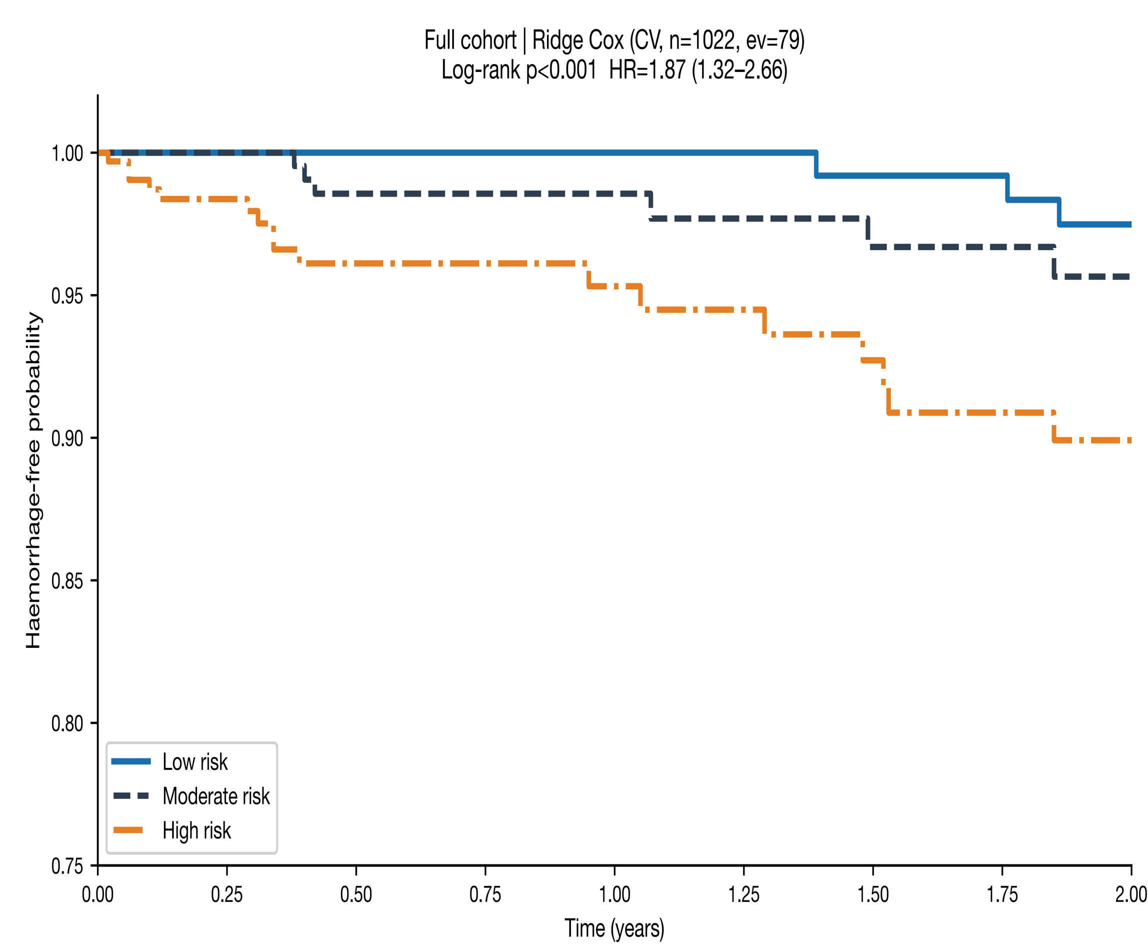
Precision-Recall Curve



Precision–recall curves for all eight classifiers on the 20% held-out test set. The best model (CatBoost, AP=0.70) substantially exceeded the prevalence-based reference line (0.41), indicating retained precision across the clinically relevant recall range.

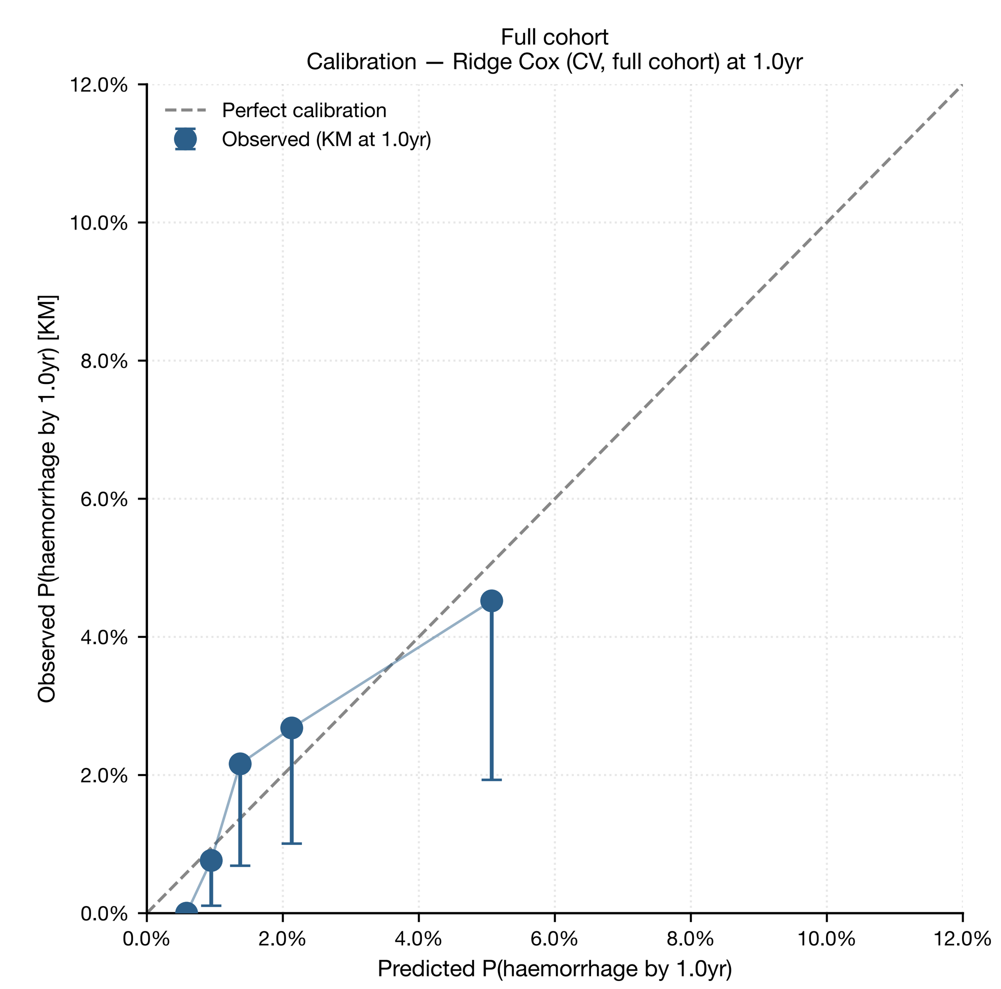
Time to Event Analysis

Kaplan Meier on Full Cohort



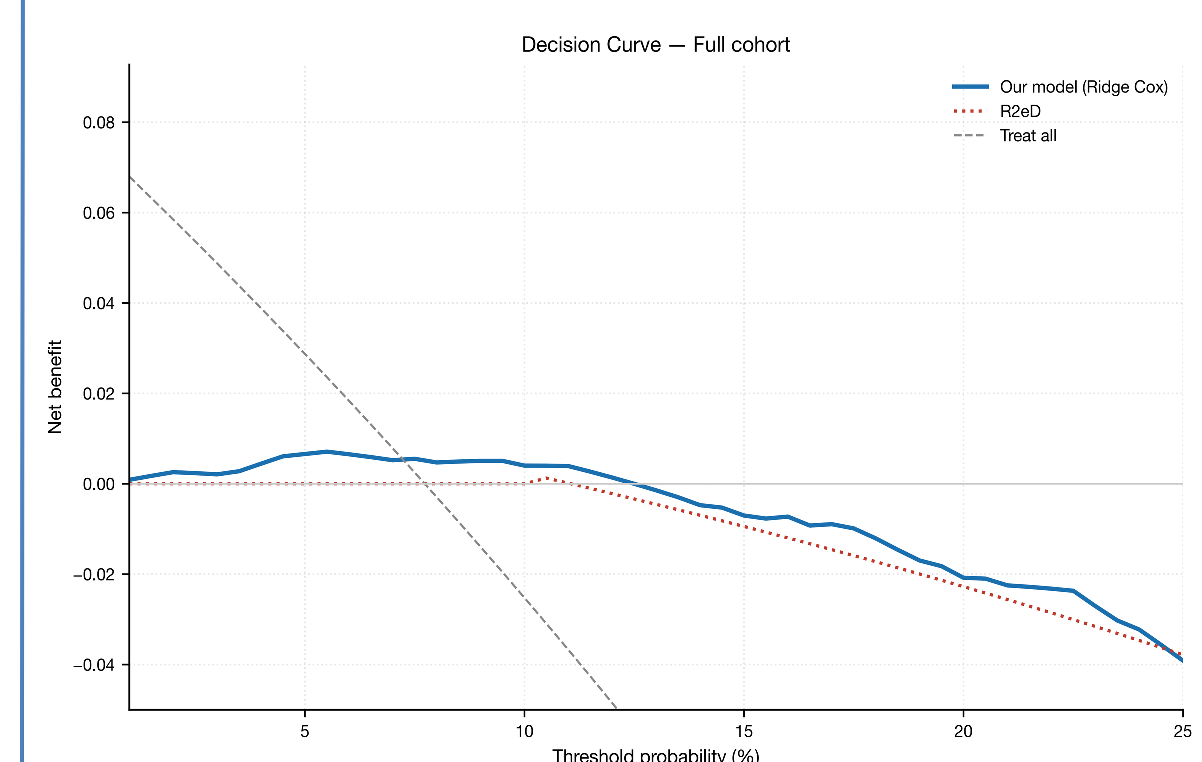
Hemorrhage-free survival stratified by model-predicted 1-year risk into three tertiles (low, moderate, high). Risk groups were significantly separated (log-rank p<0.001; HR=1.87, 95% CI 1.32–2.66), confirming the model's ability to discriminate longitudinal hemorrhage risk.

Calibration Plot



Predicted versus observed 1-year hemorrhage probabilities on the full cohort, assessed at decile-level bins. Predictions closely tracked the line of identity, indicating well-calibrated absolute risk estimates suitable for direct clinical communication.

Decision Curve



Decision-curve analysis comparing the Ridge Cox model against treat-all and treat-none strategies across threshold probabilities of 1–25%. The model yielded higher net clinical benefit throughout the range relevant to intervention decisions in bAVM management.

Conclusion

- ML-based models outperform the current clinical standard (R2eD AVM) for rupture risk prediction in bAVM patients (**AUC 0.687 → 0.801**)
- Using **13 extended clinical and angiographic features**, **CatBoost** achieved the best classification performance **(AUC 0.801, sensitivity 87.4%)** , with AVM size, deep brain location, and age as dominant predictors.
- Time-to-event analysis **stratified patients into low-, moderate-, and high-risk groups** with significant survival differences **(log-rank p < 0.001)**, enabling estimation of hemorrhage timing.

Future work:

- **Multimodal Integration**: Incorporate MRA imaging via a CNN-based pipeline with tabular clinical features.
- **External Validation**: Validate models on an independent bAVM cohort to assess generalizability and clinical transferability beyond the JHMI cohorts.
- **Ensemble Modeling**: Combine top-performing classical and deep learning models to improve robustness and generalization across populations