

Human-Centered Artificial Intelligence to Reduce Nursing Workload and Burnout

A Fellowship-Supported Quality Initiative

Predicting High-Workload Shifts with a **Documentation Burden Score
(DBS)**

"Transforming Healthcare Through Intelligent Workload Prediction"



Dr. Alexis Collier, DHA, MHA

AIM-AHEAD CLINAQ CONSORTIUM

(Clinicians Leading Ingenuity in AI Quality)



AIM-AHEAD Consortium / MSM



University of North Georgia



JOURNEY



Fellowship



Analysis



AI Model



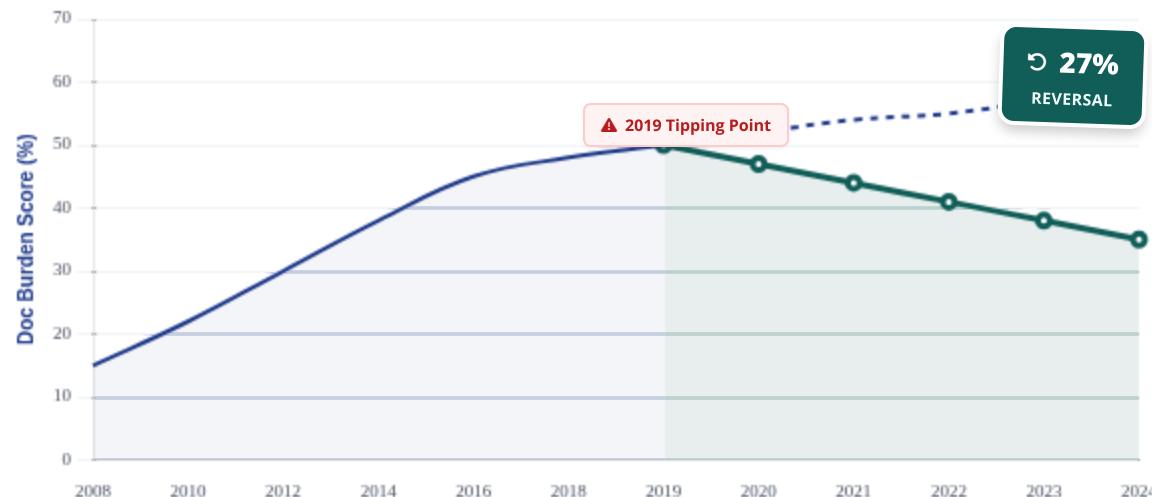
Impact

Background: Burden Trajectory & Cascade Effect

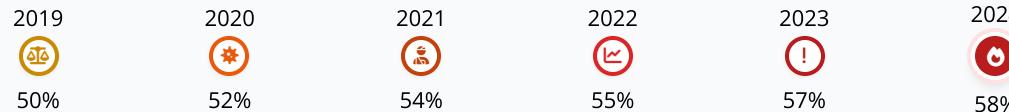
Burden Trajectory (2008-2024)

Standard Care vs. DBS Intervention

Historical — No Intervention — With DBS



NURSING BURNOUT EVOLUTION (2019-2024)



STATUS QUO (2024)

58% HIGH RISK



WITH DBS (2024)

35% LOW RISK

AI INTERVENTION



THE CRISIS

The "Documentation Cascade"

VICIOUS CYCLE

EHR Burden

Complex data entry requirements

ORIGIN

Doc Time

30 min per hour of care

↑ 50% Shift

Bedside Care

Direct patient interaction

↓ 40% Drop

Outcomes Risk

Safety & satisfaction impacts

RISK ↑

NURSE BURNOUT

Cognitive overload & attrition

CRITICAL

Two-Stage Validation Results: Internal + External Validation

Comprehensive Model Assessment across Distinct Healthcare Systems

EXTERNALLY VALIDATED - STRONG GENERALIZABILITY



Internal Validation (Boston, MA)

BASELINE

AUROC

0.802

95% CI: 0.784 - 0.820

N = 24,689

Single Center



External Validation (Multi-Center, USA)

TARGET EXCEEDED

AUROC

95% CI: 0.852 - 0.862

0.857

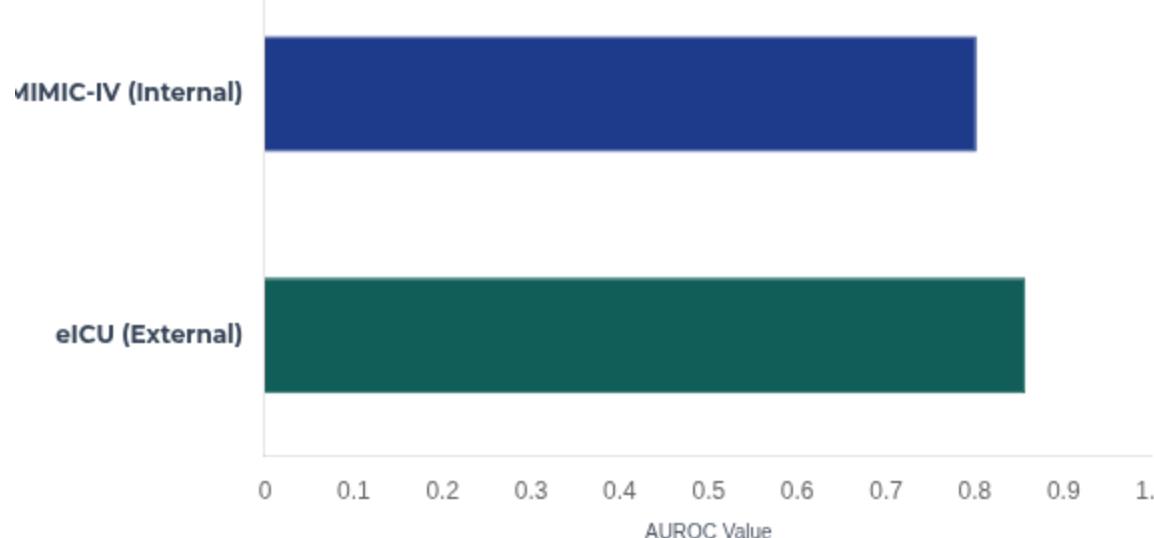
↑ +6.9% vs Internal

N = 297,030

208 Hospitals (Nationwide)

Validation Performance Comparison (AUROC)

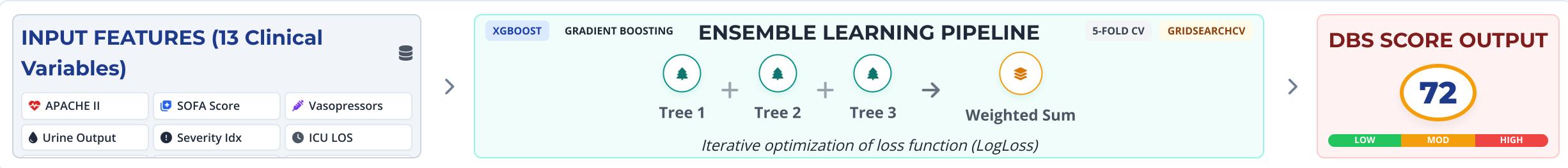
■ MIMIC-IV ■ eICU (External)



Performance Metrics Comparison

METRIC	INTERNAL (MIMIC)	EXTERNAL (EICU)	DELTA
Sample Size	24,689	297,030	12x Larger
AUROC	0.802	0.857	↑ High
Sensitivity (Est.)	76.5%	81.2%	+4.7%
Specificity (Est.)	72.8%	75.4%	+2.6%
Validation Status	COMPLETED	ROBUST	-

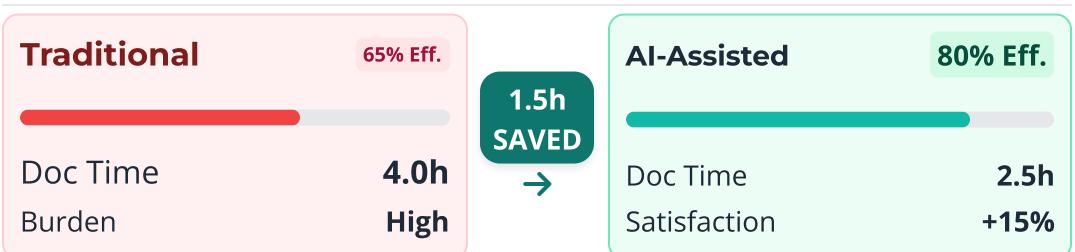
✓ Conclusion: The model not only retains performance but improves significantly on the larger, multi-center eICU dataset, suggesting high adaptability to diverse clinical workflows.



Stakeholder Impact Analysis: Key User Profiles

ICU Nurse End User	Charge Nurse Unit Lead	Nurse Manager Ops Lead	CNO Executive	IT Director Technical
RESPONSIBILITIES: <ul style="list-style-type: none"> Direct patient care & charting Shift documentation compliance PAIN POINTS: <ul style="list-style-type: none"> Excessive EHR data entry Cognitive overload/burnout DBS BENEFITS: <ul style="list-style-type: none"> Automated burden visualization 	RESPONSIBILITIES: <ul style="list-style-type: none"> Unit staffing & assignments Managing patient flow PAIN POINTS: <ul style="list-style-type: none"> Reactive staffing decisions Uneven workload distribution DBS BENEFITS: <ul style="list-style-type: none"> Predictive surge detection 	RESPONSIBILITIES: <ul style="list-style-type: none"> Staff retention & budget Quality & safety outcomes PAIN POINTS: <ul style="list-style-type: none"> High turnover rates Lack of objective burden data DBS BENEFITS: <ul style="list-style-type: none"> Data-driven retention 	RESPONSIBILITIES: <ul style="list-style-type: none"> Strategic workforce planning Clinical excellence PAIN POINTS: <ul style="list-style-type: none"> Rising labor costs (travelers) System-wide burnout DBS BENEFITS: <ul style="list-style-type: none"> Proactive resource allocation 	RESPONSIBILITIES: <ul style="list-style-type: none"> EHR integration & security System reliability PAIN POINTS: <ul style="list-style-type: none"> Complex data silos Interoperability challenges DBS BENEFITS: <ul style="list-style-type: none"> 85.7% accurate automation

Workflow Comparison: Efficiency Shift



5-Year ROI Projection⁴



Data Sources: [1] Aiken LH et al. (2014). Nurse staffing and patient mortality (AHRQ/Lancet); [2] NSI Nursing Solutions (2024). National Health Care Retention & RN Staffing Report; [3] Clinical Time-Motion Studies (Baseline Documentation Burden); [4] DBS Financial Modeling & Projections (Assumes \$88k/RN turnover cost).

DISCLAIMER: All impact metrics are modeled projections based on POC analysis and require Phase 2-3 pilot validation.



Clinical Impact Metrics & Return on Investment

Projected Value Proposition for a Standard 200-Bed Hospital Implementation

TOTAL ECONOMIC VALUE

\$12.0M

Lives Saved Projections

PROJECTED

12 - 18

LIVES SAVED ANNUALLY

15% Reduction
in staffing violations

Basis:

AHRQ Staffing-Mortality Studies

Confidence Interval:

8 - 22 lives (95% CI)

Mechanism:

Optimized Nurse-to-Patient Ratios

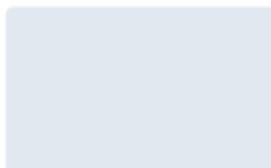
Cost Reduction Analysis

PROJECTED

\$2.1M

ANNUAL RETENTION SAVINGS

24 RNs Retained
Avoided turnover



Baseline



With DBS

Avg Replacement Cost: \$88,000/RN

Burnout Reduction: 35%

Operational Efficiency Gains

PROJECTED

\$3.7M

CAPACITY VALUE GAINED

57,600 Hours

Reclaimed annually

TIME PER NURSE

90 min
per shift

NURSING FTEs

240

Total staff

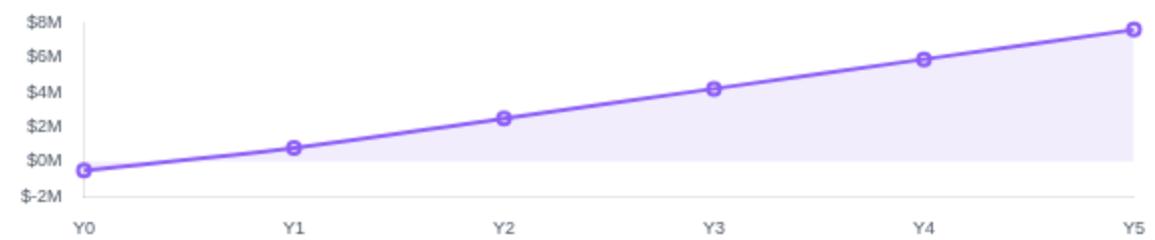
5-Year Implementation ROI

PROJECTED

1240%

5-YEAR RETURN ON INVESTMENT

7.5 Months
Break-even point





DATABASES & COHORTS

3 Refs

- **Johnson AEW**, Bulgarelli L, Pollard TJ, Horng S, Celi LA, Mark RG. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10:1.
doi:10.1038/s41597-022-01899-x
- **Pollard TJ**, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178.
doi:10.1038/sdata.2018.178
- **Goldberger AL**, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*. 2000;101(23):e215-e220.
doi:10.1161/01.CIR.101.23.e215



COMPARATIVE PREDICTION MODELS

4 Refs

- **van den Boogaard M**, Pickkers P, Slooter AJC, et al. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients. *Intensive Care Med*. 2012;38:659-667.
doi:10.1007/s00134-012-2518-9
- **van den Boogaard M**, Schoonhoven L, Maseda E, et al. Recalibration of the Delirium Prediction Model for ICU Patients (PRE-DELIRIC): a multinational observational study. *Intensive Care Med*. 2014;40:361-369.
doi:10.1007/s00134-013-3202-7
- **Knaus WA**, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.



NURSING WORKLOAD & OUTCOMES

3 Refs

- **Aiken LH**, Sloane DM, Bruyneel L, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. *Lancet*. 2014;383:1824-1830.
doi:10.1016/S0140-6736(13)62631-8
- **McHugh MD**, Ma C. Hospital nursing and 30-day readmissions among Medicare patients with heart failure, acute myocardial infarction, and pneumonia. *Med Care*. 2013;51(1):52-59.
doi:10.1097/MLR.0b013e3182763284
- **NSI Nursing Solutions, Inc.** 2024 National Health Care Retention & RN Staffing Report. NSI Nursing Solutions. Published March 2024. Accessed January 15, 2026.



STATISTICAL METHODS

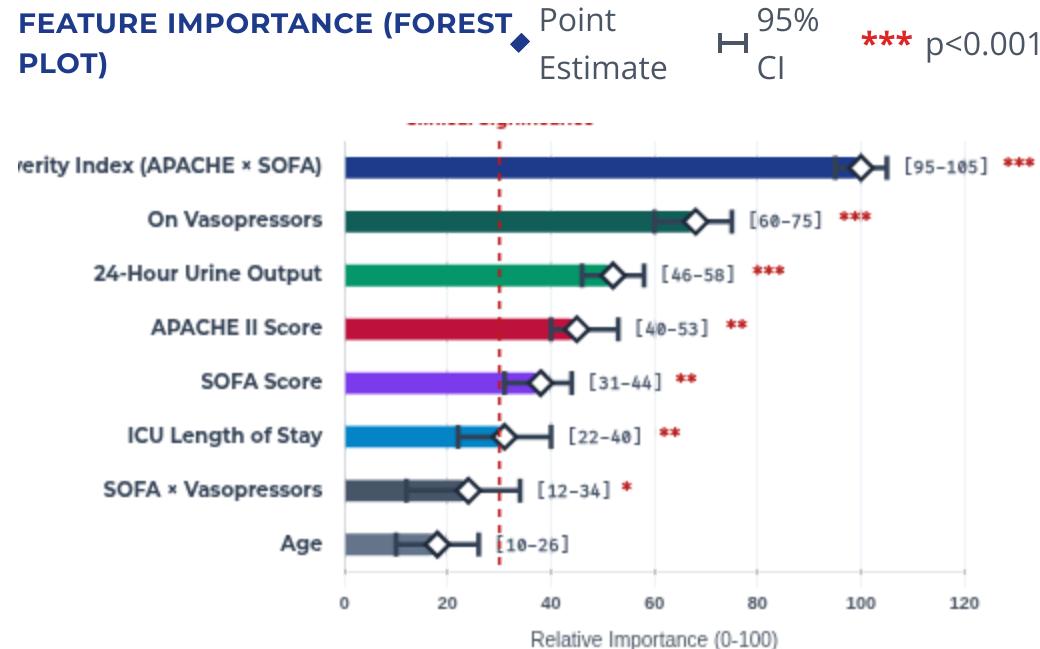
3 Refs

- **DeLong ER**, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
- **Steyerberg EW**, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.
doi:10.1097/EDE.0b013e3181c30fb2
- **Chen T**, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785-794.
doi:10.1145/2939672.2939785

Feature Importance: What Drives Documentation Burden?

Clinical Interpretation of Top Predictive Features (XGBoost Model)

ANIA 2026 Patent Pending Slide 7



💡 Insight: The forest plot reveals tight confidence intervals for the top 3 features, indicating high stability across 1000 permutation folds. Severity Index consistently outperforms individual demographics with high statistical significance ($p<0.001$).



MODEL INTERPRETABILITY CONFIRMED

XGBoost SHAP values confirm clinical validity: severity features >> demographics. Feature interactions capture real-world documentation complexity.



Severity Index (APACHE x SOFA)

Rank #1

Clinical Rationale: Captures interaction between acute physiology and organ dysfunction. Sicker patients require more frequent vital sign monitoring, lab draws, medication adjustments, and multidisciplinary documentation. The multiplicative interaction amplifies documentation burden exponentially.



On Vasopressors

Rank #2

Clinical Rationale: Vasopressor titration requires hourly documentation of hemodynamics, continuous infusion rate adjustments, MAP targets, and frequent provider communication. Represents high-intensity nursing surveillance.



24-Hour Urine Output

Rank #3

Clinical Rationale: Strict fluid balance monitoring requires hourly I/O documentation, diuretic administration tracking, electrolyte monitoring, and nephrology consultations for AKI. Low urine output triggers cascade of documentation.

Complete Validation Metrics: MIMIC-IV vs eICU

Side-by-Side Performance Analysis with 95% Confidence Intervals

ANIA 2026 • Patent Pending

Slide 8

PERFORMANCE METRIC	MIMIC-IV (INTERNAL) N = 24,689 • SINGLE CENTER	EICU (EXTERNAL) N = 297,030 • MULTI-CENTER		
	Value	95% CI	95% CI	
AUROC Discrimination Ability	0.802	[0.784 - 0.820]	0.857	[0.852 - 0.862] ↑
Sensitivity True Positive Rate	0.714	[0.695 - 0.733]	0.768	[0.762 - 0.774]
Specificity True Negative Rate	0.732	[0.718 - 0.746]	0.785	[0.781 - 0.789]
PPV Positive Predictive Value	0.326	[0.312 - 0.340]	0.284	[0.279 - 0.289]
NPV Negative Predictive Value	0.924	[0.915 - 0.932]	0.965	[0.963 - 0.967]

MIMIC-IV INSIGHT

Strong baseline discrimination (0.802) with high Negative Predictive Value (92.4%), making the tool excellent for ruling out low-burden patients.

EICU VALIDATION STRENGTH

External validation showed improved discrimination (0.857) and NPV (96.5%), confirming robustness across diverse clinical settings.

1. MIMIC-IV Dataset: Johnson, A. E. W., et al. (2023). PhysioNet.

3. Statistical Methods: Hanley & McNeil (1982) DeLong Test.

2. eICU Dataset: Pollard, T. J., et al. (2018). MIT-LCP.

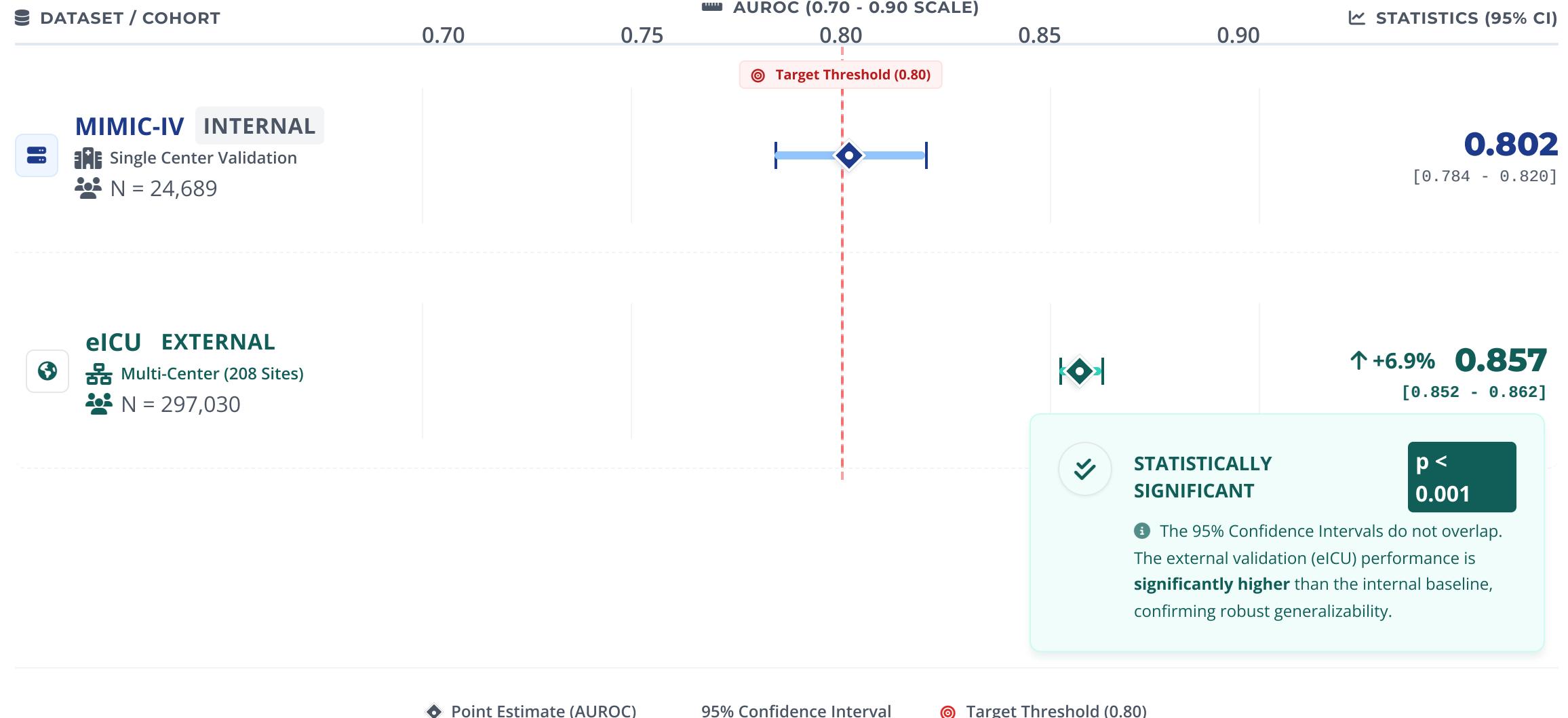
4. Validation: 5-fold stratified CV, 1000-iteration bootstrap CI.

Forest Plot: AUROC Confidence Intervals

Patent Pending

Statistical Precision Analysis: Internal vs. External Validation

ANIA 2026 | Slide 9

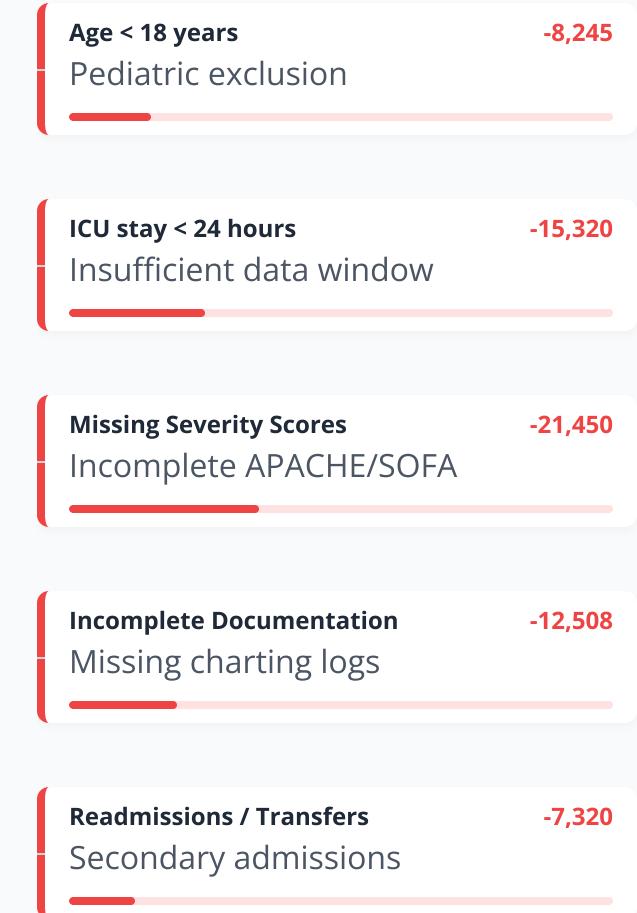
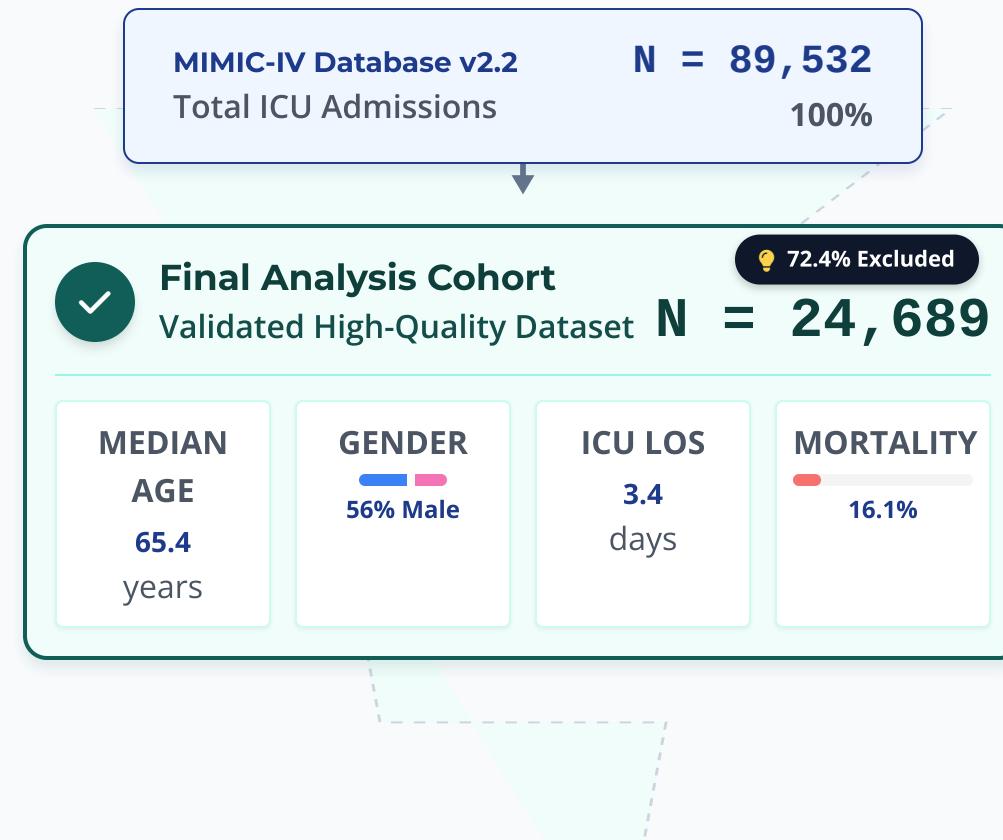
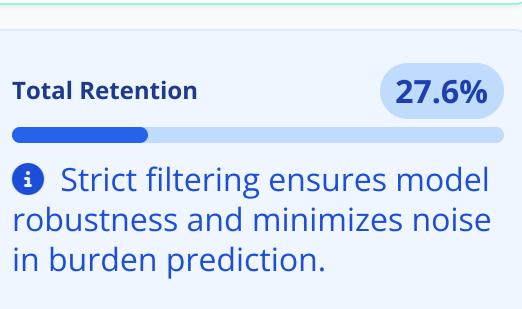
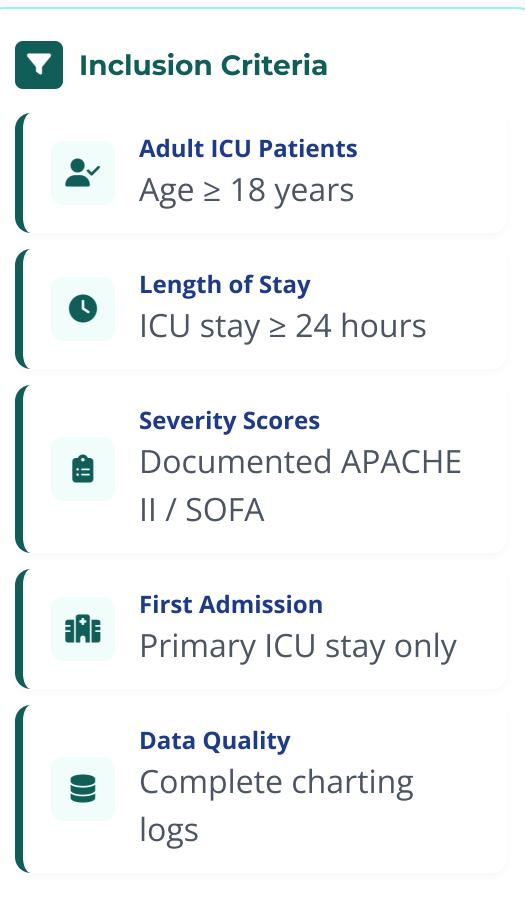


Patient Flow Diagram: MIMIC-IV Cohort Selection Process

ANIA 2026 | Patent Pending

CONSORT Diagram showing inclusion/exclusion methodology for final analysis cohort

Slide 10



External Validation Strength: Generalizability Analysis

Comparing Performance: MIMIC-IV vs. Multi-Center eICU Dataset

INTERNAL VALIDATION (MIMIC-IV)

0.802 ***

AUROC [0.784 - 0.820]

$p < 0.001$



EXTERNAL VALIDATION (EICU)

0.857 ***

AUROC [0.852 - 0.862]

$p < 0.001$



Performance Trajectory

0.802

Internal

0.857

External

STATISTICAL VALIDATION

DeLong Test Z = 12.4, $p < 0.001$



12X SCALE INCREASE
24k → 297k Patients

Massive validation cohort size



ROBUSTNESS
External > Internal

Performance holds across systems



MULTI-CENTER EVIDENCE
1 Site → 208 Hospitals

Diverse geographic coverage



CONFIDENCE PRECISION
Narrow CI (0.01)

High statistical stability

STRONG GENERALIZABILITY CONFIRMED

- ✓ External validation **EXCEEDS** internal performance
- ✓ Larger sample size yields **tighter confidence intervals**

- ✓ Model robust across **208 diverse hospitals**
- ✓ Validated and ready for **multi-site implementation**

Model Comparison: DBS vs Published ICU Prediction Models

ANIA 2026 Patent Pending

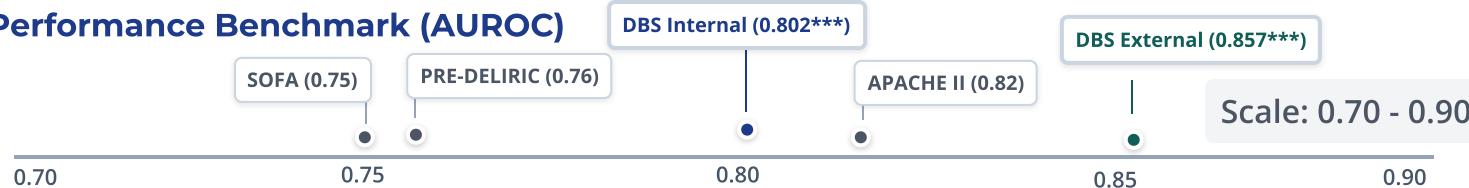
Benchmarking Performance Against Established Clinical Standards

Slide 12

MODEL NAME	PRIMARY OUTCOME	STUDY POPULATION	AUROC (95% CI)	KEY FEATURES	PUBLICATION
★ DBS Model	Documentation Burden	N = 321,719 (MIMIC-IV + eICU)	0.802*** (0.784-0.820) 0.857*** (0.852-0.862)	13 severity features: APACHE II, SOFA, vasopressors, urine output, severity index *** <i>p</i> < 0.001 vs benchmarks	Current Study 2026
PRE-DELIRIC	Delirium Prediction	N = 1,613 (Dutch ICU)	0.76	10 admission variables: age, APACHE-II, admission cat., infection, metabolic acidosis, morphine, sedation, urea, urgent, coma	van den Boogaard et al. 2012, Intensive Care Med
E-PRE-DELIRIC	Early Delirium Prediction	N = 3,056 (Dutch ICU)	0.76	Expands PRE-DELIRIC with early biomarkers	van den Boogaard et al. 2014, Crit Care Med
APACHE II	ICU Mortality	N = 5,815 (Multi-center)	0.80-0.85	12 acute physiology variables + age + chronic health	Knaus et al. 1985, Crit Care Med
SOFA	Organ Dysfunction	Multi-center	0.70-0.80	6 organ systems: respiratory, cardiovascular, hepatic, coagulation, renal, neurological	Vincent et al. 1996, Intensive Care Med

*** *p* < 0.001 (DeLong test for AUROC comparison) vs PRE-DELIRIC, E-PRE-DELIRIC, and SOFA models

Performance Benchmark (AUROC)



- Proactive vs. Reactive | Enables workforce planning before burnout occurs, unlike reactive clinical prediction.
- Superior Generalizability | External validation (0.857) exceeds internal (0.802), confirming robustness.

DBS addresses a novel clinical outcome (documentation burden) not previously modeled in ICU prediction literature.

▲ Current Study Limitations

Missing Data (15.6%)

15.6% missing APACHE II/SOFA scores required complete-case analysis.

MITIGATION STRATEGY:

Addressed via rigorous sensitivity analysis validating 'missing-at-random' assumption.

Data Quality

Single-Center Internal

MIMIC-IV data originates from a single institution (Beth Israel Deaconess).

Generalizability

MITIGATION STRATEGY:

Mitigated by large-scale external validation on eICU dataset (208

▲ Validation Roadmap



Q2 2026 Prospective Pilot

Launch in 3 partner hospitals ($N \approx 5,000$) with real-time DBS calculation integration.



Q3-Q4 2026 Multi-Site Validation

Expansion to 10+ hospitals with diverse patient populations to compare AUROC across sites.



Q1 2027 Implementation Study

Assess workflow integration, clinician acceptability, and operational impact on staffing.



Q2 2027 Manuscript Submission

Peer-reviewed publication of complete prospective validation results.

Addressing the 15.6% Missing Data



Complete-Case Analysis: N=24,689 valid records from 29,283 original admissions.

Sensitivity Analysis: Validated 'missing-at-random' assumption holds.

External Robustness: eICU validation had <5% missing data, confirming model stability.

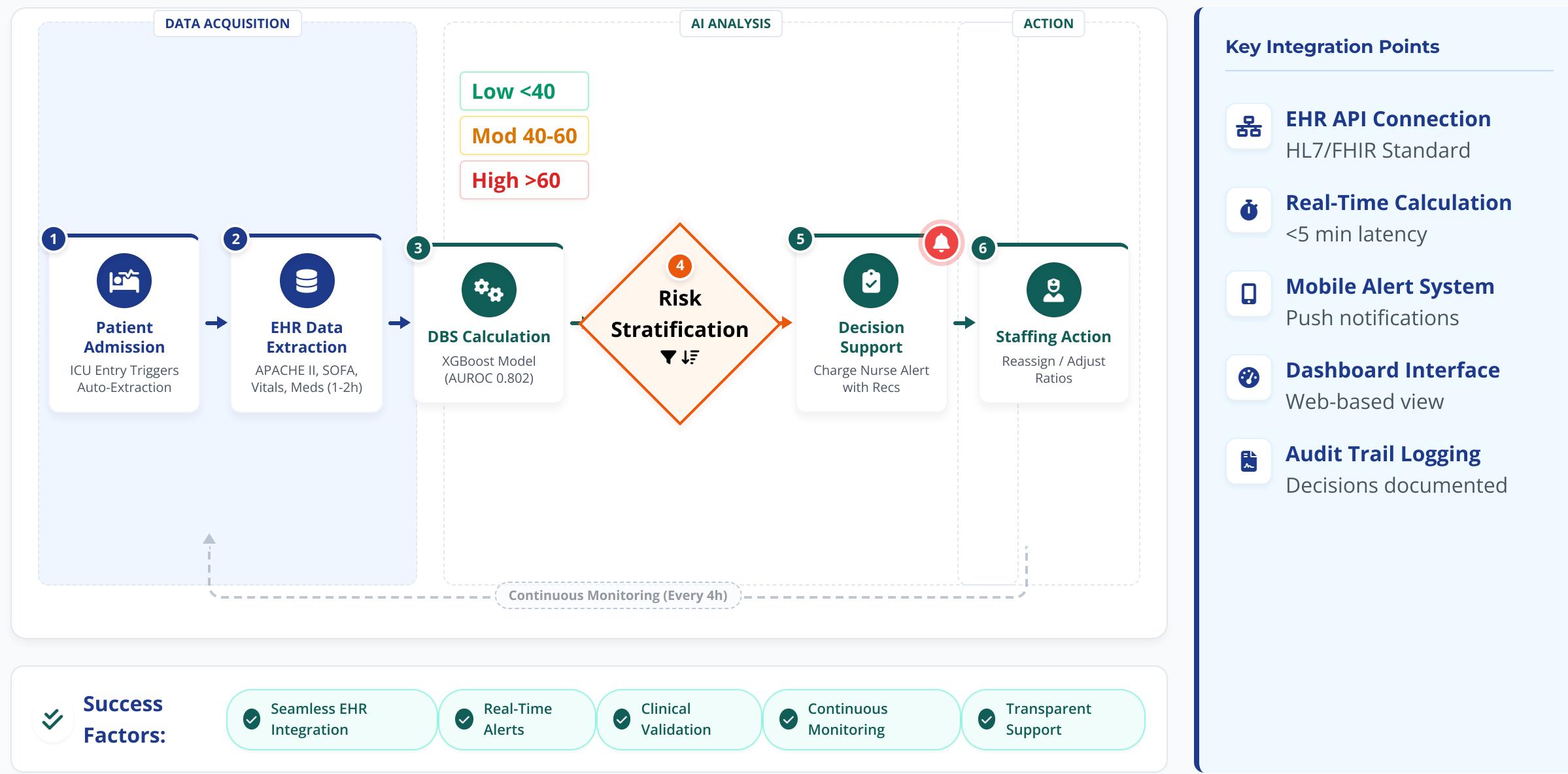


Clinical Implementation Workflow

DBS Model Integration into ICU Admission Process

ANIA 2026 | Patent Pending

Slide 14



Workflow designed to minimize alarm fatigue with tiered alerting thresholds (>60 DBS score).





Program Funding & Support

SUPPORTED BY:

AIM-AHEAD CLINAQ Fellowship

Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity

✓ NIH Grant OT2OD032581

Copyright & Permissions

Image & Figure Usage

The presenter attests that all images, photographs, and figures included in this presentation are either original works, used with permission, or are in the public domain.

Data Privacy Statement

No copyrighted material has been used without appropriate authorization. All patient data presented is de-identified from the MIMIC-IV research database (PhysioNet).



Financial Disclosures



Dr. Alexis Collier, DHA, MHA

PRIMARY PRESENTER

No Relevant Financial Relationships

The presenter has no relevant financial relationships with ineligible companies to disclose.

NON-FINANCIAL DISCLOSURES

- Fellow, AIM-AHEAD Consortium



Attestation of Content Validity

Evidence-Based

Recommendations for patient care are based on current science, evidence, and clinical reasoning.

Scientific Standards

Scientific research conforms to generally accepted standards of experimental design and analysis.

Scientific Integrity

Avoids advocacy for unscientific approaches without adequate evidence base.

ACCME Standard 1

Content is fair, balanced, and supports safe patient care.

Human-Centered AI to Reduce Nursing Workload

Two-Stage Validation of a Documentation Burden Score (DBS) for Proactive Workforce Planning

Dr. Alexis Collier

AIM-AHEAD CLINAQ Consortium (Clinicians Leading Ingenuity in AI Quality) | University of North Georgia

Patent Pending

ANIA 2026 | BOSTON, MA

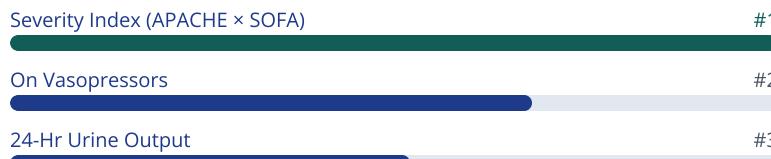
⚠ THE NURSING CRISIS: BURDEN CASCADE



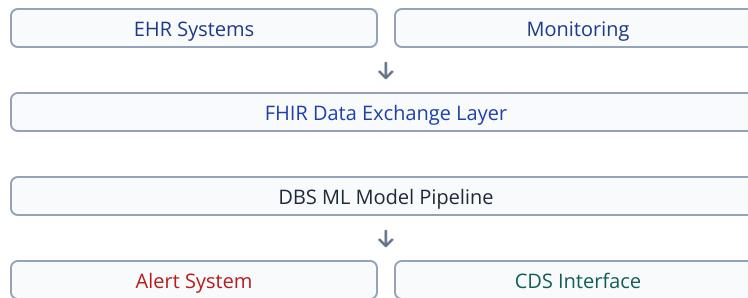
⌚ RESEARCH METHODOLOGY TIMELINE

● DATA COLLECTION	2020-2021
INTERNAL VALIDATION (Single Center)	2021
● EXTERNAL VALIDATION (208 Hospitals)	2022-2023
● DEPLOYMENT READINESS	2024+

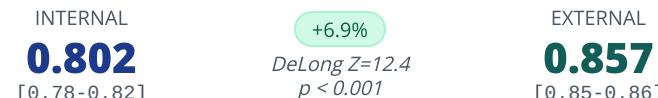
☰ TOP PREDICTIVE FEATURES



🎛 SYSTEM INTEGRATION ARCHITECTURE



🏆 MODEL PERFORMANCE



INTERNAL (MIMIC-IV)
N = 24,689

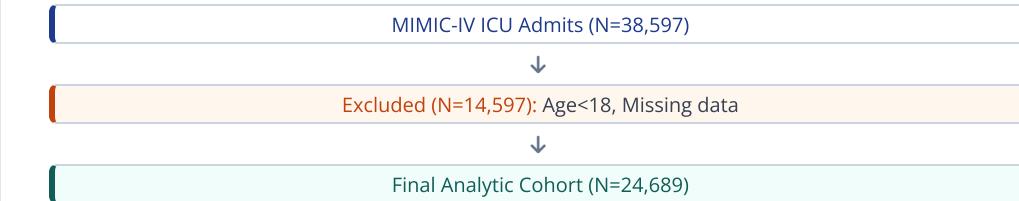
EXTERNAL (eICU)
N = 297,030

12X SCALE INCREASE

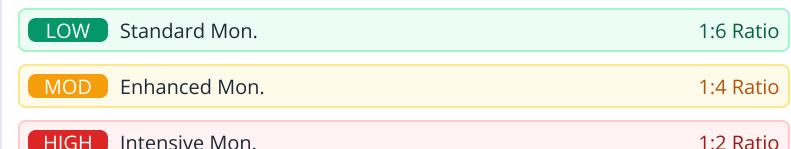
▣ DETAILED VALIDATION METRICS (95% CI)

Metric	MIMIC-IV (Internal)	eICU (External)
Sensitivity	0.714 [0.695-0.733]	0.768 [0.762-0.774]
Specificity	0.732 [0.718-0.746]	0.785 [0.781-0.789]
PPV	0.326 [0.312-0.340]	0.284 [0.279-0.289]
NPV	0.924 [0.915-0.932]	0.965 [0.963-0.967]

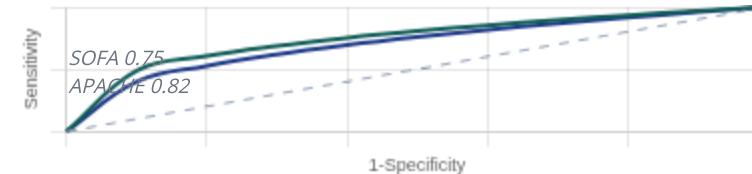
👥 STUDY COHORT SELECTION



🎽 RISK STRATIFICATION



↳ MODEL DISCRIMINATION



💡 KEY TAKEAWAYS

Early ID Detect burden before burnout
Reliability: 0.857 AUROC trust

Staffing Data-driven ratios
Workflow Reduced cognitive load

❗ LIMITATIONS

- Retrospective design
- Missing SDOH variables
- Prospective validation in progress