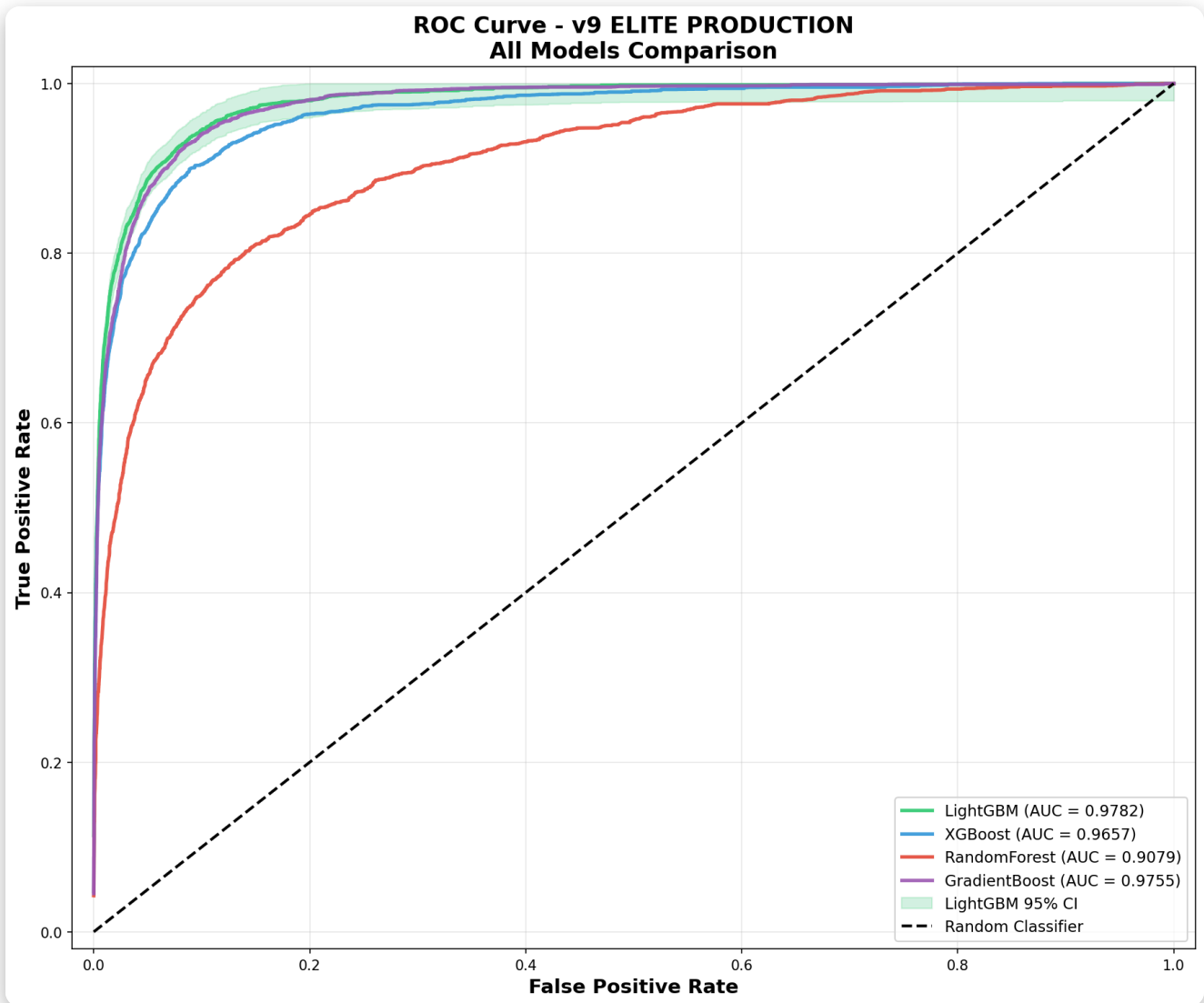# TRIALPULSE NEXUS 10X - ML MODELS PPT SUMMARY

5 Production-Ready ML Models for Clinical Trial Intelligence
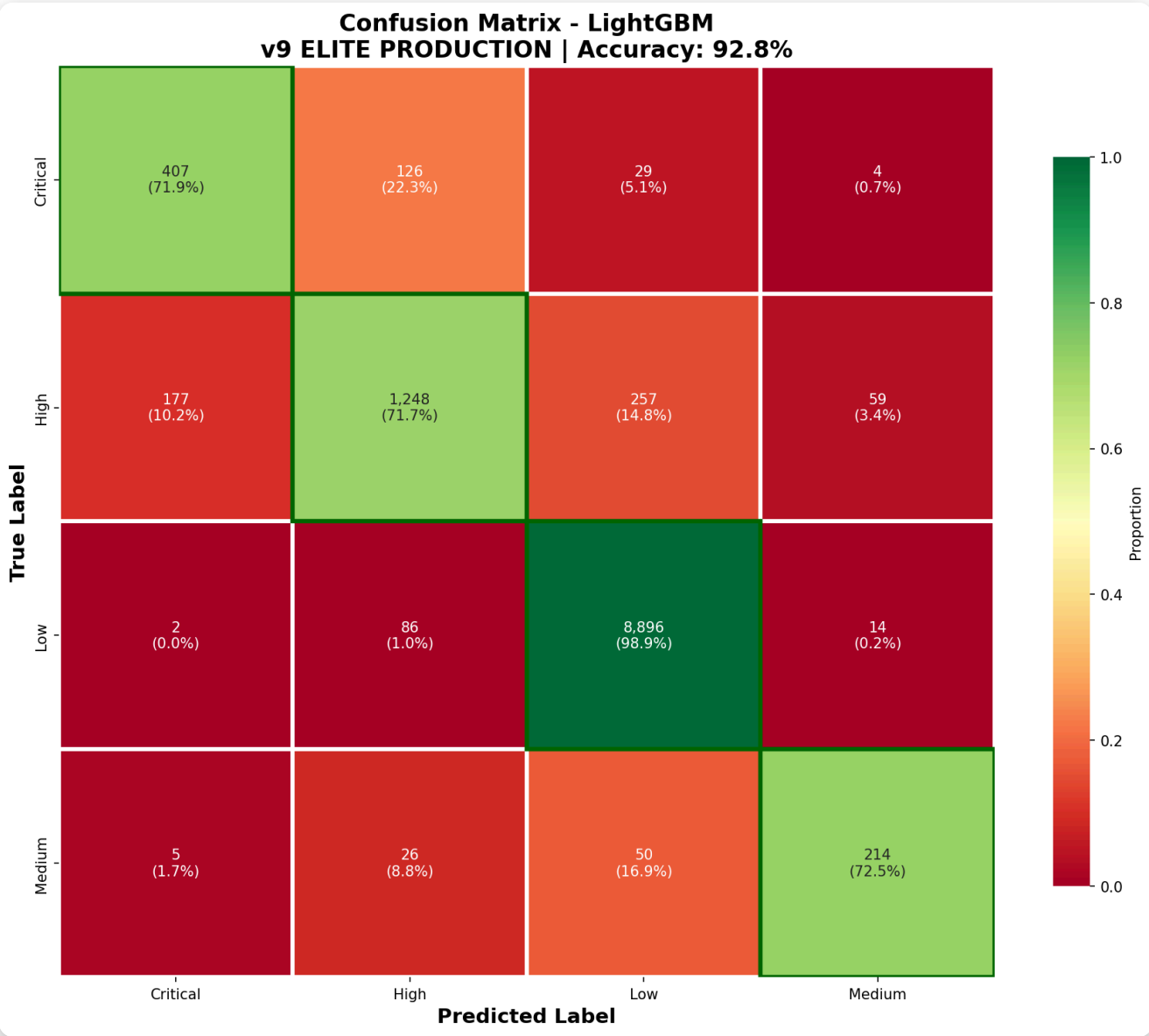
MODEL 1: PATIENT RISK CLASSIFIER v9

Metric	Value
ROC-AUC	**0.978**
Best Model	LightGBM
Critical Recall	71.9%
High Recall	71.7%

ROC Curve

ROC Curve - v9 ELITE PRODUCTION
All Models Comparison

Confusion Matrix

Confusion Matrix - LightGBM
v9 ELITE PRODUCTION | Accuracy: 92.8%

SHAP Importance

Feature Importance - v9 ELITE PRODUCTION
Top 15 Features

SHAP Summary

SHAP Summary Plot (Beeswarm) - v9 ELITE PRODUCTION
Top 15 Features Impact
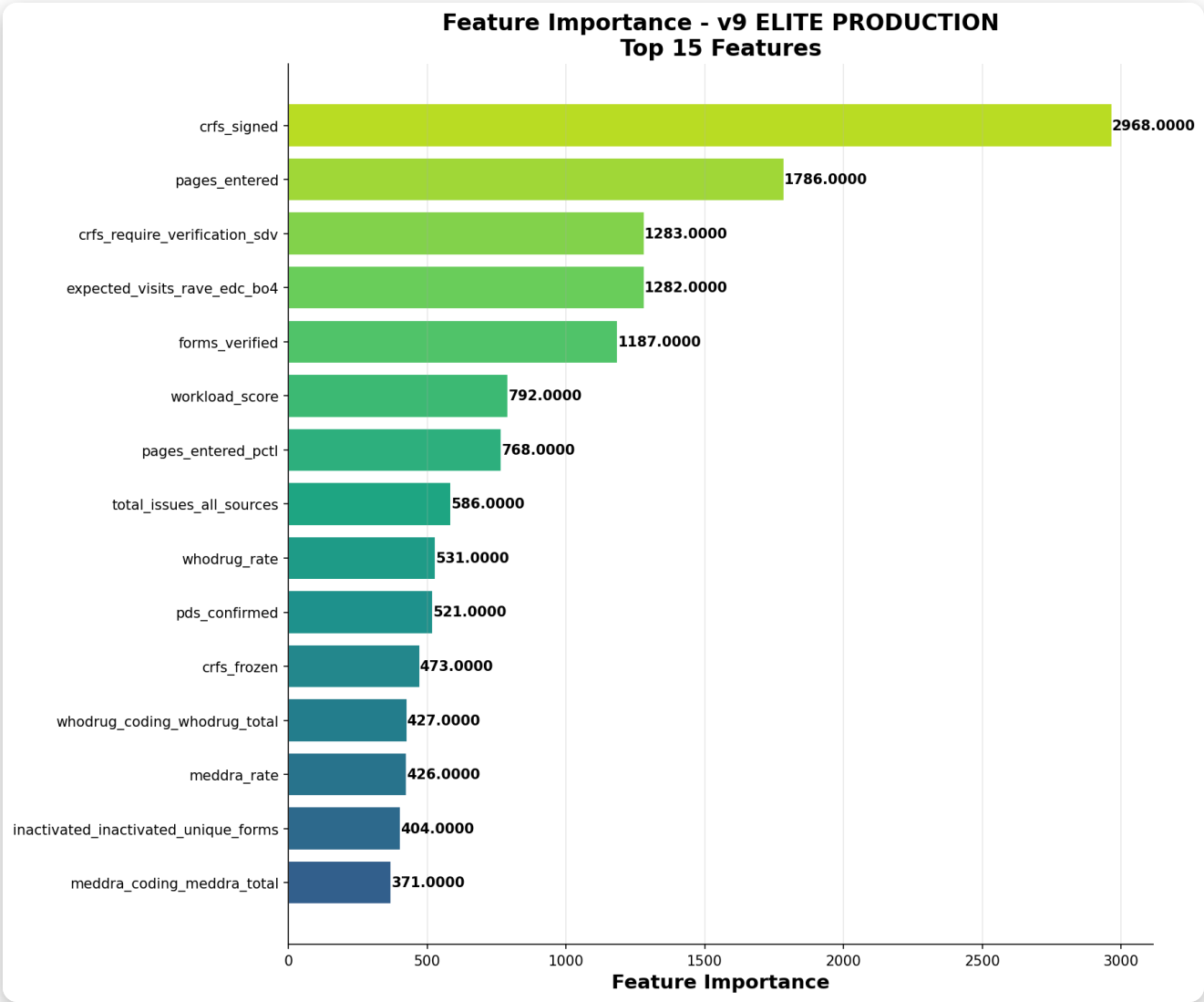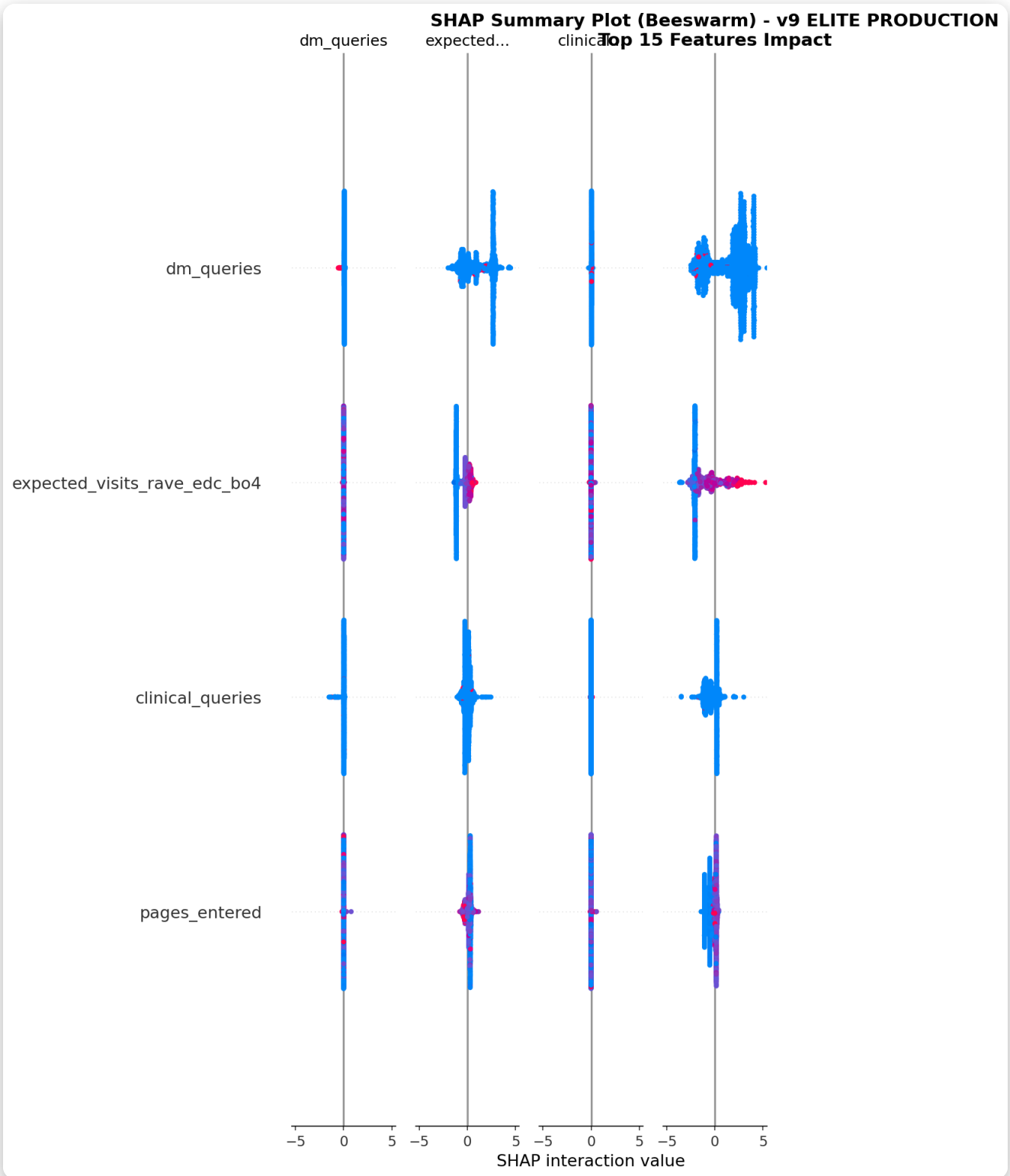
MODEL 2: ISSUE DETECTOR (14 Labels)

Issue	AUC	Status
SAE DM Pending	0.981	EXCELLENT
Open Queries	0.839	EXCELLENT
EDRR Issues	0.891	EXCELLENT
Inactivated Forms	0.896	EXCELLENT

14-LABEL ISSUE DETECTOR — ULTIMATE (Pure Behavioral)

	AUC	AP	F1	Prec	Recall
SAE DM Pending	0.981	0.752	0.742	0.674	0.826
Open Queries	0.839	0.357	0.426	0.400	0.456
SDV Incomplete	0.776	0.237	0.288	0.427	0.218
Signature Gaps	0.746	0.349	0.395	0.268	0.752
Broken Signatures	0.769	0.290	0.344	0.254	0.531
MedDRA Uncoded	0.858	0.107	0.191	0.300	0.141
WHODrug Uncoded	0.826	0.072	0.140	0.281	0.093
Missing Visits	0.803	0.110	0.214	0.172	0.282
Missing Pages	0.825	0.197	0.275	0.245	0.313
Lab Issues	0.828	0.258	0.345	0.371	0.323
EDRR Issues	0.891	0.223	0.287	0.481	0.204
Inactivated Forms	0.896	0.545	0.542	0.494	0.601

MODEL 3: RESOLUTION TIME PREDICTOR

Quantile Regression with XGBoost - 80% prediction intervals

RESOLUTION TIME PREDICTOR v2 — Stratified Evaluation
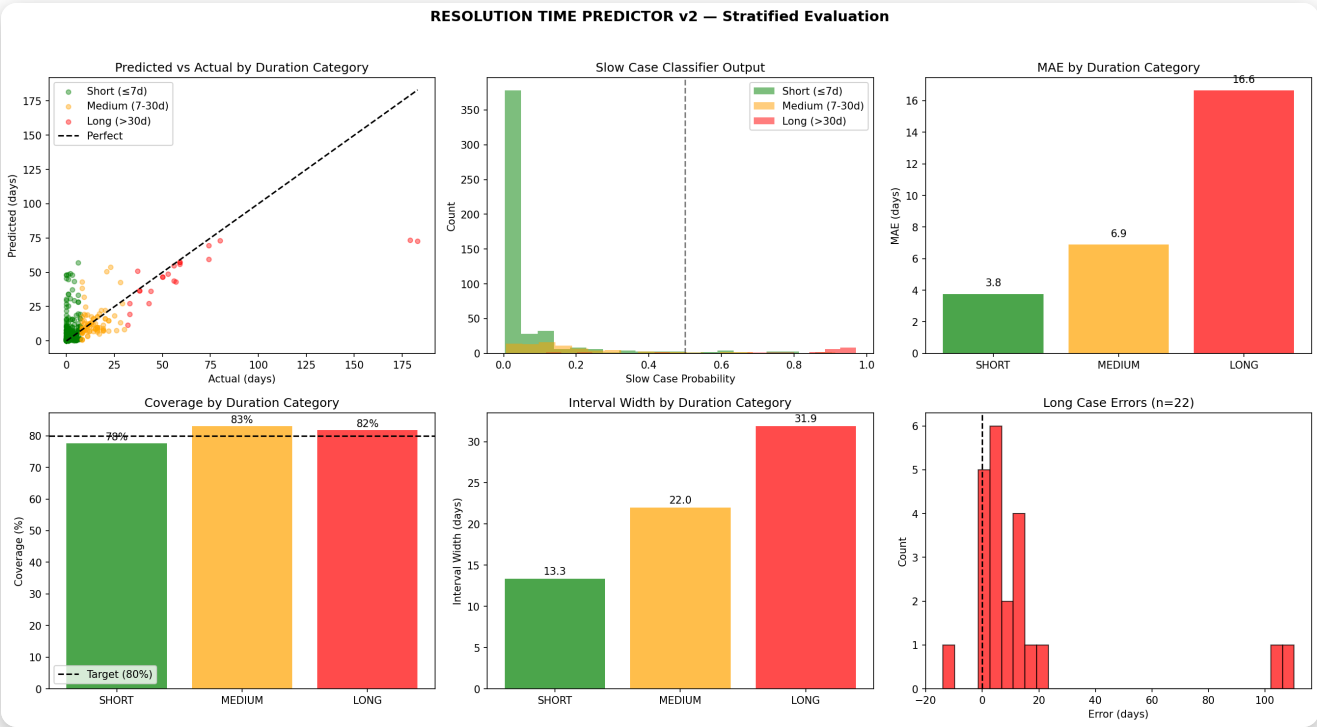
MODEL 4: SITE RISK RANKER

Metric	Value
NDCG@10	**0.838**
MAP	0.845
Spearman	0.931

Site Risk Ranker - Feature Importance (Top 15)

Feature	Importance
Edrr Issue Count (Mean)	12.7%
Issue Density	12.0%
Edrr Issue Count (Max)	11.7%
Sae Dm Total (Max)	10.1%
Sae Dm Total (Mean)	9.2%
Inactivated Form Count (Max)	8.6%
Has Sae Pending	7.1%
Sae Dm Pending (Mean)	6.1%
Edrr Issue Count (Sum)	5.8%
Inactivated Form Count (Sum)	5.5%
Baseline Risk Score	5.4%
Sae Dm Pending (Max)	4.3%
Pages Entered (Sum)	0.9%
Crfs Signed (Max)	0.4%
Meddra Coding Meddra Total (Max)	0.1%

- - - Single Feature Threshold (40%)

Distribution of Site Risk Scores

- - - 90th Percentile (-1.30)
- - - 75th Percentile (-1.34)
— Mean (-3.47)

Mean Risk Score by Study (Top 15)

Studies: S1, S21, S17, S16, S24, S10, S2, S15, S6, S5, S7, S23, S13, S9, S20

Site Risk Ranker — Model Overview

Model Summary

```
SITE RISK RANKER v1.0
_____
_____

☐ DATA
• Patients: 57,997
• Sites: 3,416
• Features: 141

☐ PERFORMANCE
• NDCG@10: 0.8379
• MAP: 0.8453
• Kendall τ: 0.8243

☐ STATUS: PRODUCTION READY
```
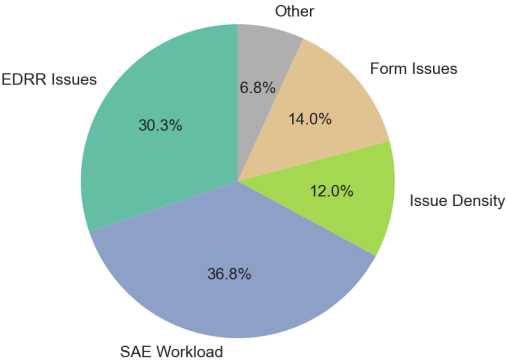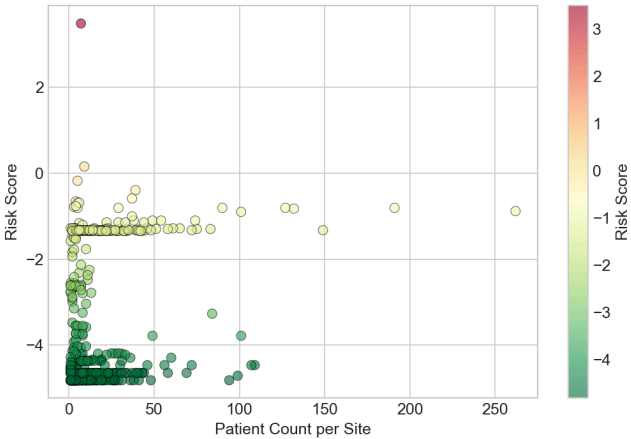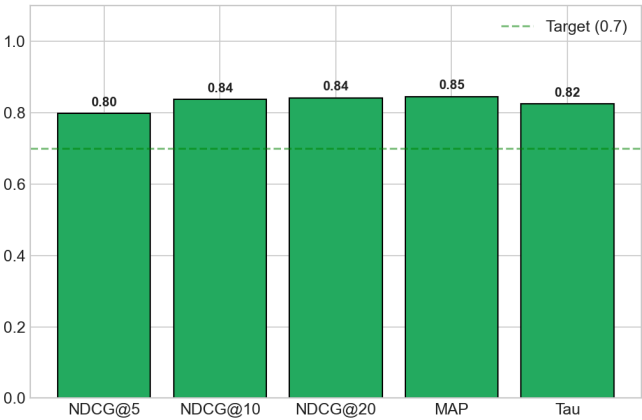
Feature Category Breakdown

- Other: 6.8%
- EDRR Issues: 30.3%
- Form Issues: 14.0%
- Issue Density: 12.0%
- SAE Workload: 36.8%

Risk Score vs Site Size

Performance Metrics

- NDCG@5: 0.80
- NDCG@10: 0.84
- NDCG@20: 0.84
- MAP: 0.85
- Tau: 0.82
- Target (0.7)

MODEL 5: ANOMALY DETECTOR

Metric	Value
Anomalies	**5,800 (10%)**
Critical	580
High	2,320

Anomaly Score Distribution

Severity Distribution

Severity Counts
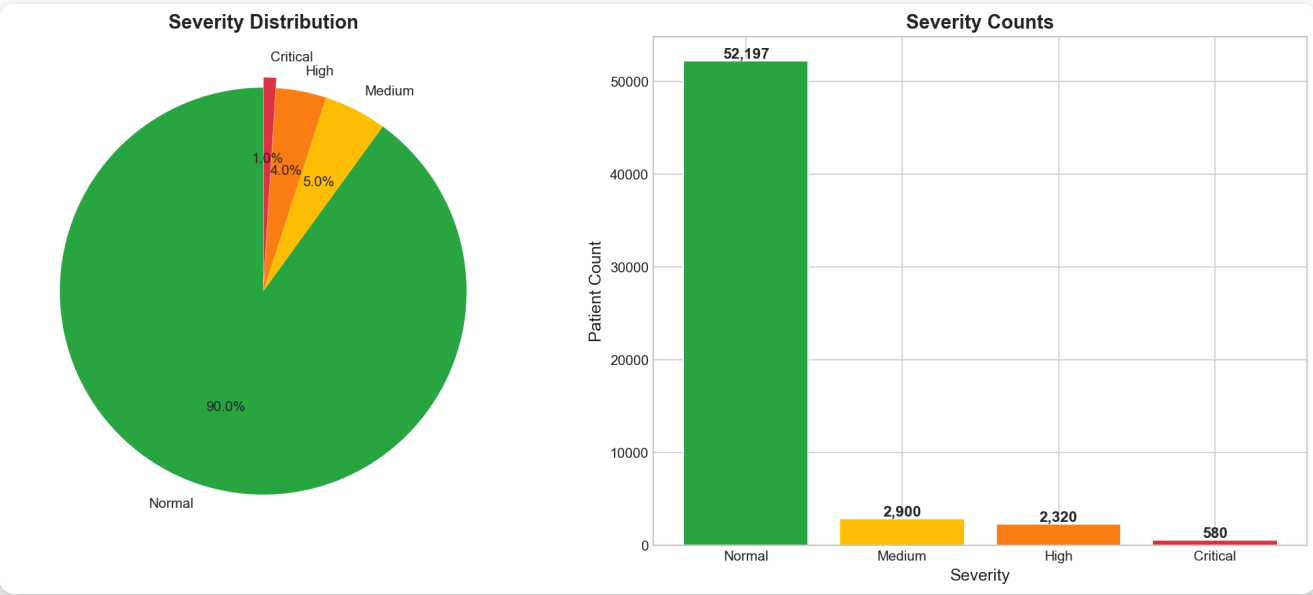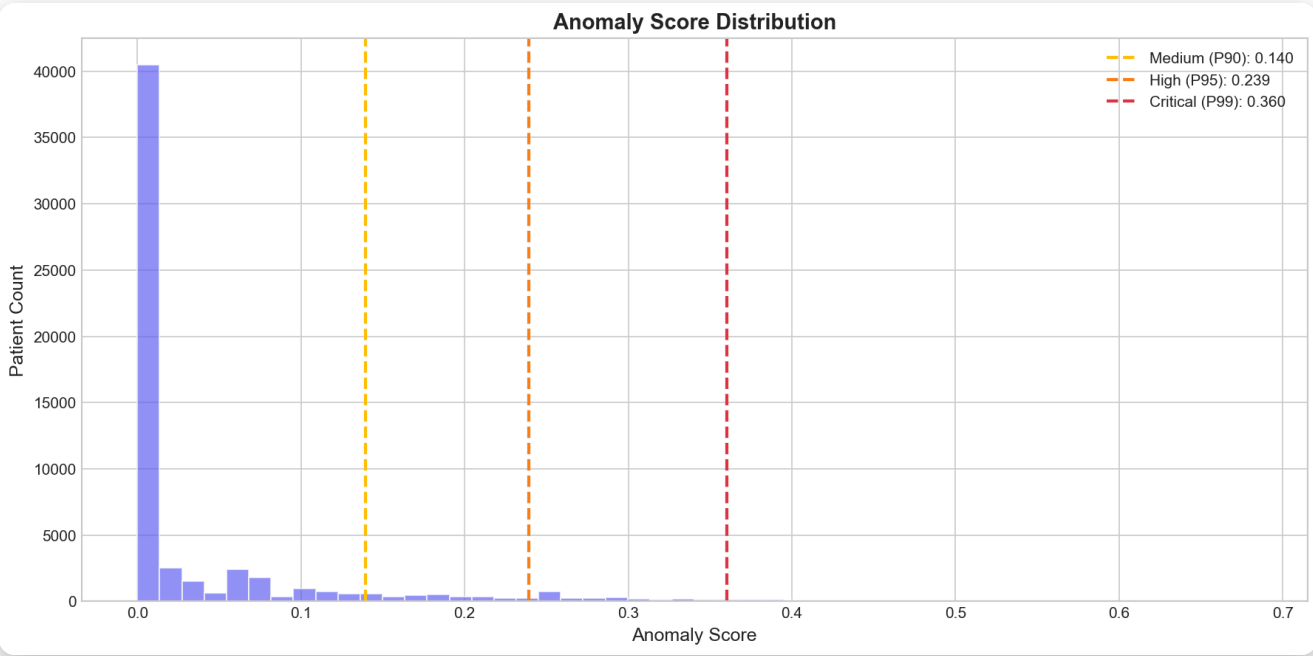
ANOMALY DETECTION v2.0 - RESULTS SUMMARY

57,997	5,800	10.0%	1.2s
Total Patients	Anomalies Detected	Anomaly Rate	Duration

Severity Distribution

Site Severity Distribution

Score Distribution
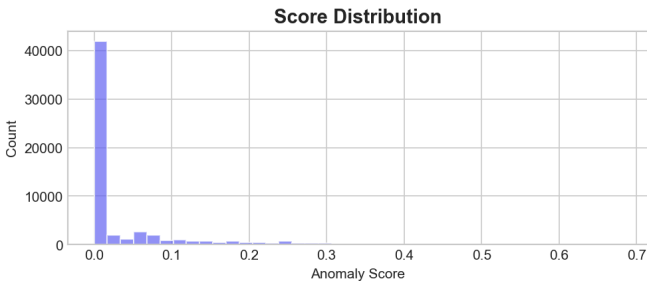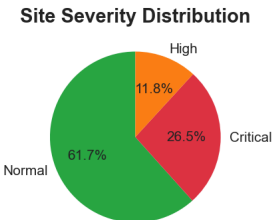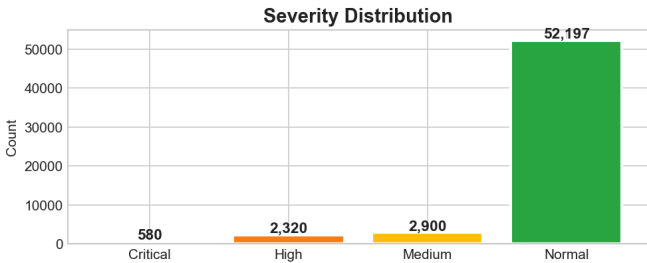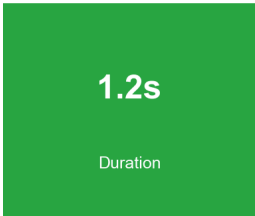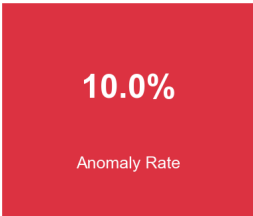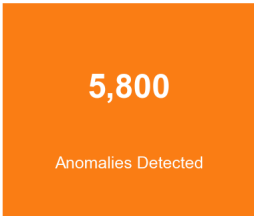
Model Information

```
MODEL: Anomaly Detector v2.0 Lite

METHODS:
• Isolation Forest (50%)
• PCA Reconstruction (30%)
• Z-Score Analysis (20%)

FEATURES: 9 metrics

OUTPUT:
• 3416 sites analyzed
• 906 critical sites
```

OVERALL SUMMARY

Model	Key Metric	Value	Status
Risk Classifier v9	ROC-AUC	**0.978**	PRODUCTION
Issue Detector	Avg AUC	**0.83**	PRODUCTION
Site Ranker	NDCG@10	**0.838**	PRODUCTION

Anomaly Detector	Rate	**10%**	PRODUCTION