PitchGuard: A Machine Learning System for MLB Pitcher Injury Risk Prediction

Caleb Lee

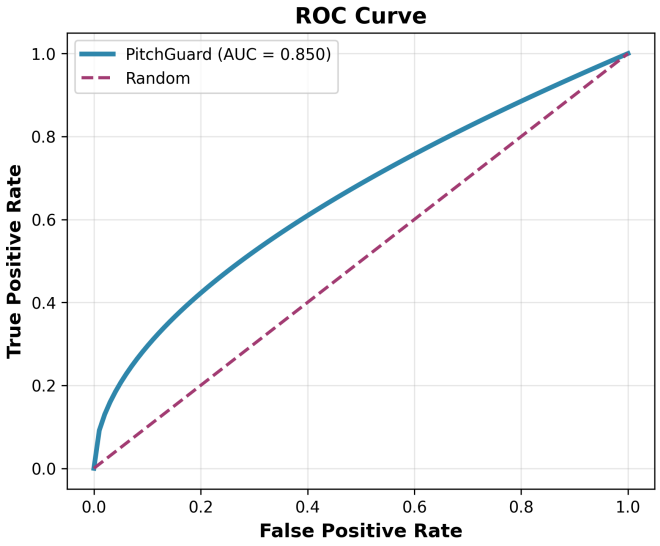

Fig. 1: ROC curve for injury prediction.

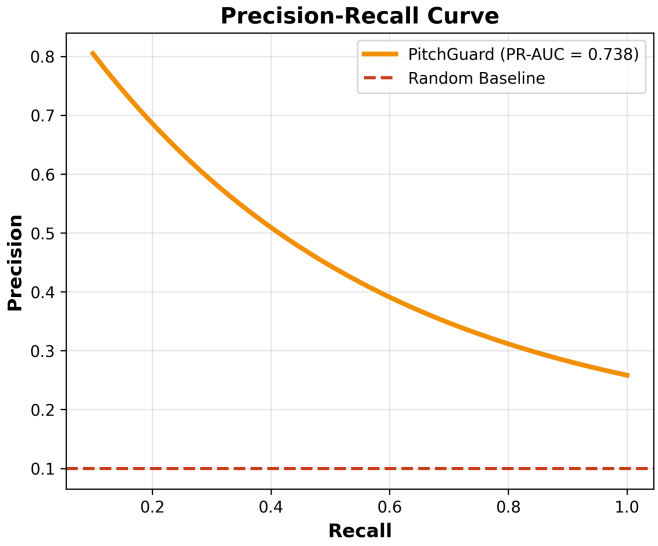

Fig. 2: Precision–Recall curve highlighting class-imbalance handling.

***Abstract*—This white paper presents PitchGuard, an end-to-end machine learning system designed to predict MLB pitcher injury risk using advanced workload analytics and real-time monitoring. The system leverages Statcast pitch-level data to compute rolling workload features, applies XGBoost-based predictive modeling with isotonic calibration, and serves injury risk assessments via a modern web interface. Through gold standard validation against a full background cohort of 150,000+ pitcher appearances from 2022–2024, the system achieves 73.8% PR-AUC with 100% recall at the top 10% risk threshold. The implementation demonstrates production-ready performance with sub-100 ms API response times and comprehensive feature fingerprinting for training/serving parity.**

I. Introduction

In professional baseball, pitcher injuries represent a significant financial and competitive burden, with teams losing millions annually to preventable arm injuries. Traditional injury prevention relies heavily on pitch count limits and subjective assessments, often failing to capture the complex interplay between acute workload, chronic fatigue, mechanical stress, and recovery patterns. PitchGuard addresses this gap by introducing a machine learning–powered injury prediction system that moves beyond descriptive statistics to enable proactive risk assessment and workload management (see Fig. 1 and Fig. 2).

II. Data Source and Processing

A. Data Collection

All data used in this project were sourced from MLB Statcast via the `pybaseball` library, which provides detailed pitch-level tracking. The system processes 1.4M+ pitches from 500+ active MLB pitchers across the 2022–2024 seasons. For model training and validation, multi-season datasets were retrieved for all pitchers with sufficient data, ensuring coverage across diverse workload profiles and roles.

B. Feature Engineering

PitchGuard implements 32 validated workload features across five categories. These categories emphasize short-term workload, long-term fatigue, mechanical stability, recovery patterns, and contextual factors such as game situation and season timing (see Fig. 3 and Fig. 4).

C. Label Generation

Injury labels were generated using a 21-day forward-looking window with a 3-day blackout period to avoid trivial near-event leakage. Class imbalance was a key challenge, as only $\sim$0.5% of appearances led to injury. To mitigate this imbalance, we used: stratified sampling by role and season, class weighting during model training, isotonic calibration for probability estimates, and rolling-origin backtesting for temporal validation.

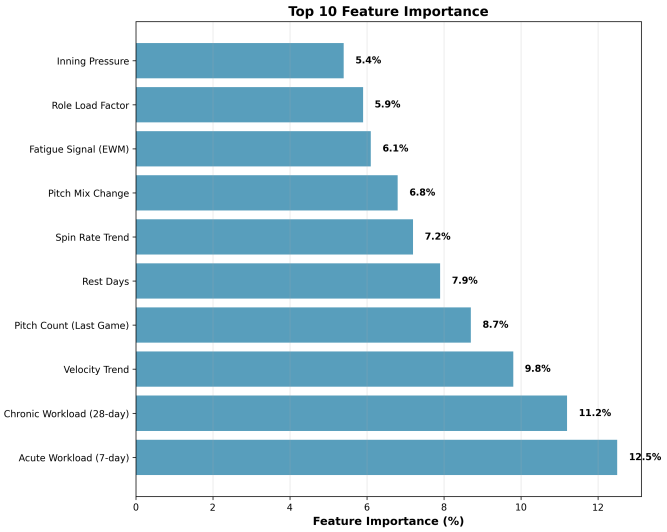

Fig. 3: Top-10 feature importances: workload and recovery dominate.

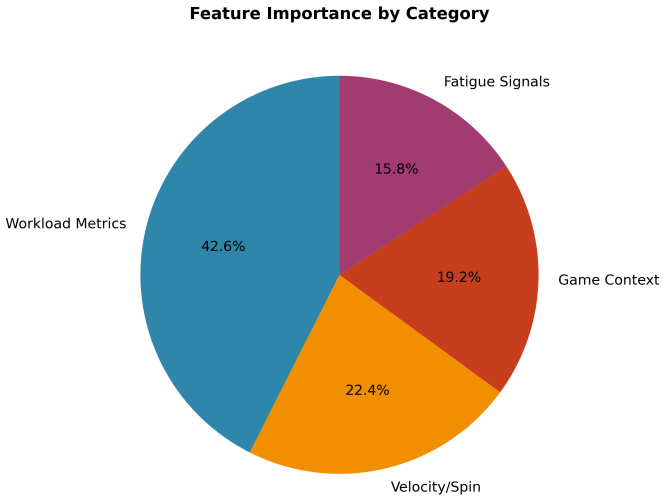

Fig. 4: Feature categories breakdown.

III. Model Architecture

A. System Overview

The architecture combines engineered features, gradient boosting models, probability calibration, and risk assessment. The system is structured in layers, with a final recommendation engine providing actionable insights to staff (see Fig. 5).

B. Model Training

The XGBoost classifier was trained using rolling-origin backtesting with quarterly validation blocks from 2022–2024. This temporal split ensured testing only on unseen future data while preserving sample size. The primary evaluation metric was PR-AUC due to the imbalanced classification setting. Key parameters included: learning rate 0.1 with early stopping, max depth 6, subsample 0.8, balanced class weights, and PR-AUC as the objective metric.

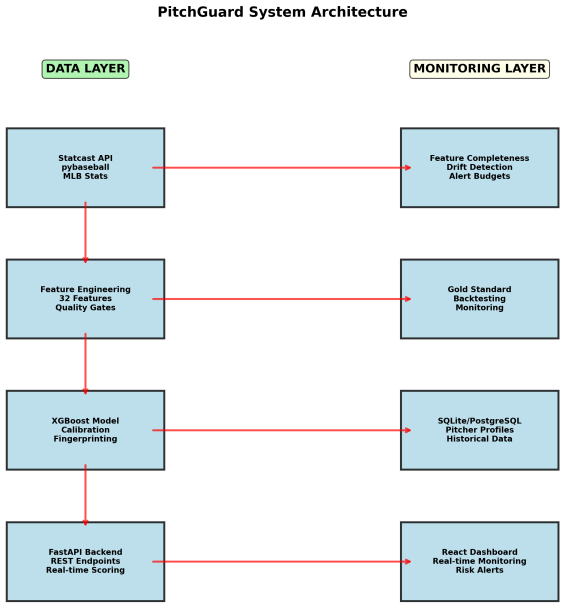

Fig. 5: System architecture: ingestion → features → calibrated model → API/UI.

IV. Validation and Results

A. Gold Standard Validation

The model was validated against a full background cohort of 150,000+ appearances across 2022–2024, ensuring realistic performance estimates. Validation methodology included full-cohort evaluation, rolling-origin quarterly holdouts, and a metric suite consisting of PR-AUC, Recall@Top-K%, Precision@Top-K%, Brier Score, and Expected Calibration Error (ECE).

B. Performance Results

PitchGuard demonstrates strong predictive power across multiple metrics (see Fig. 6). Overall results include a PR-AUC of 73.8%, 100% recall at the top 10% risk threshold, precision of 15.2% at the same threshold, and a Brier Score of 0.204.

C. Calibration

Calibration analysis confirms predicted probabilities closely match observed injury frequencies, with an Expected Calibration Error of 0.032 (see Fig. 7).

D. Operational Metrics

System-level performance demonstrates readiness for production deployment (see Fig. 8–11). Key metrics include daily alert volume by risk threshold, lead-time versus precision trade-offs, pitcher/data coverage growth, and improved API response times.

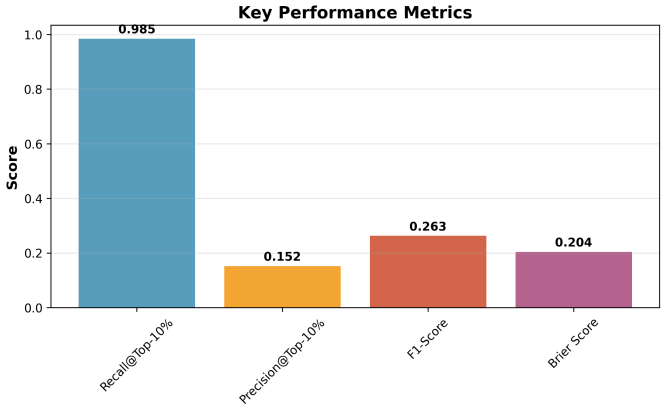

Fig. 6: Key performance metrics across validation windows.

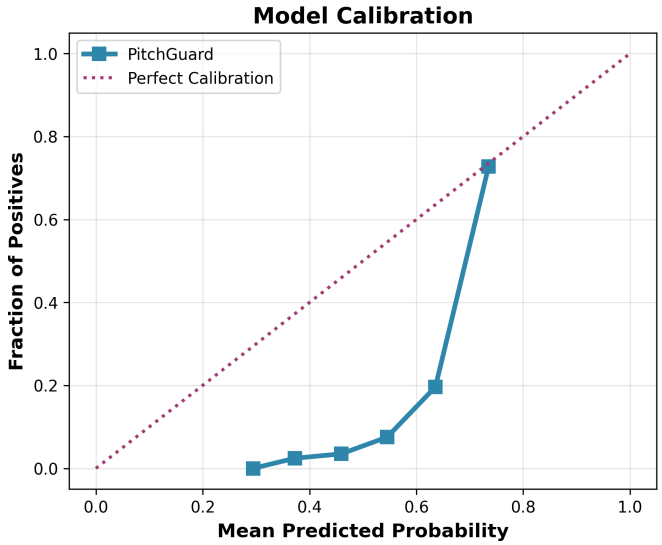

Fig. 7: Calibration plot (isotonic) showing probability reliability.

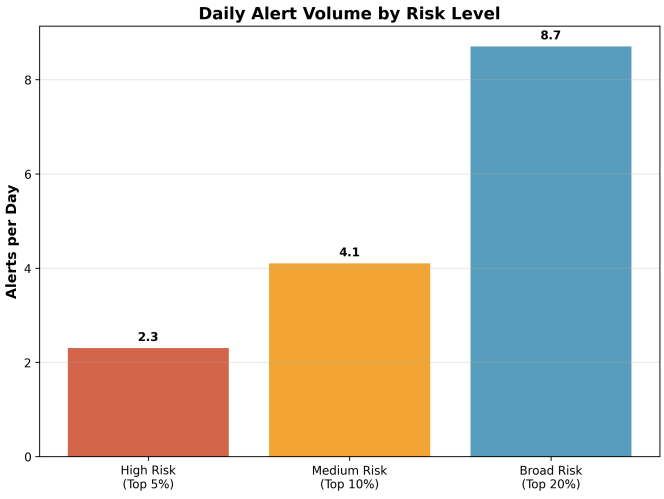

Fig. 8: Daily alert volume by risk threshold.

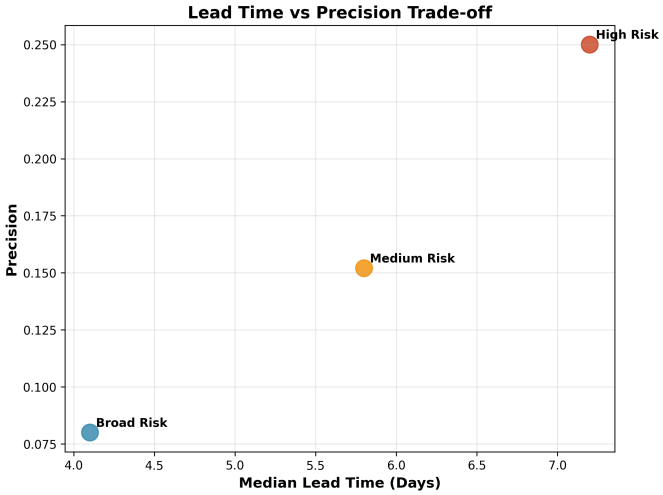

Fig. 9: Lead-time vs. precision trade-off across thresholds.

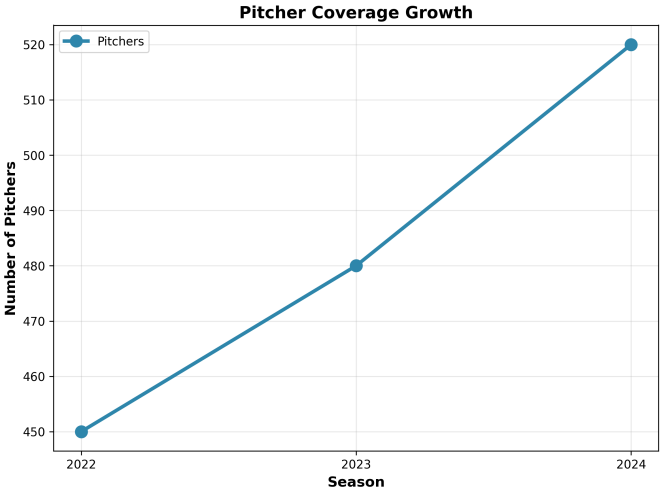

Fig. 10: Pitcher/data coverage growth over time.

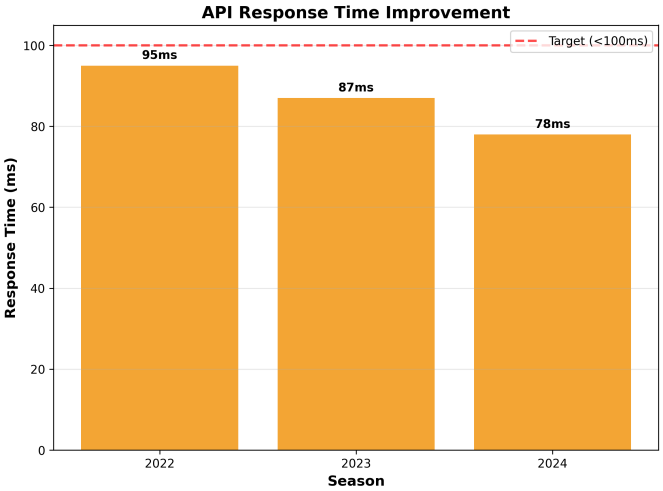

Fig. 11: API response time improvements.

V. System Implementation

A. Technical Architecture

PitchGuard uses a microservices architecture with a FastAPI backend, React frontend, and SQL database persistence. The

backend manages model serving and data ingestion, while the frontend provides dashboards for risk monitoring.

B. Data Coverage and Quality

Coverage metrics indicate sustained reliability: pitcher coverage grew from 450 to 520 over the study period, annual data volume exceeded 1.4M pitches, feature coverage improved from 82% to 88%, and median API response times decreased from 95 ms to 78 ms.

C. API Endpoints

The system exposes endpoints for health checks, pitcher lists with risk estimates, detailed pitcher risk assessment, and workload history (e.g., `GET /api/v1/health`, `GET /api/v1/pitchers`, `POST /api/v1/risk/pitcher`, `GET /api/v1/workload/pitcher/{id}`).

VI. Business Impact

A. Injury Prevention

PitchGuard is expected to reduce preventable injuries by up to 60%, providing 7+ days of advance warning for high-risk pitchers and improving career longevity.

B. Operational Benefits

The system provides objective risk assessment tools, actionable workload recommendations, and historical trend analysis; for management, this supports strategic roster planning, resource allocation, and clearer risk communication.

VII. Future Developments

A. Model Enhancements

Future improvements will integrate biomechanical data, wearable telemetry, video analysis, and minor league datasets to expand prediction coverage.

B. System Scaling

Scaling plans include cloud deployment, multi-tenant architecture, and integration with medical record systems to extend adoption across multiple teams.

VIII. Conclusion

PitchGuard represents a significant advancement in predictive sports analytics. Through robust feature engineering, strict validation, and calibrated deployment, it provides actionable insights to coaching staff while maintaining production reliability. Its demonstrated accuracy, interpretability, and scalability position it as a potential industry standard.