DosePilot: AI-Driven Medication Adherence Monitoring for Medicare Advantage Seniors

Sazzad Islam Charlotte Yan Sayed Ardhi Aaron Lee

Abstract

Medication non-adherence costs $100–300B annually and drives worse outcomes among Medicare Advantage (MA) seniors in the United States. We present DosePilot, an AI-enabled adherence monitoring system that ingests electronic health record (EHR) data from Fast Healthcare Interoperability Resources (FHIR), computes CMS-standard Proportion of Days Covered (PDC) metrics, and predicts non-adherence risk. Using 515,520 prescription fills from 7,403 synthetic Medicare-like patients in the CMS Synthetic Part D Prescription Drug Event dataset, our Random Forest model achieves 97.9% ROC-AUC for predicting annual adherence status, as part of a complete pipeline from data ingestion to role-specific dashboards for providers, payers, and patients.

1 Introduction

Medication non-adherence contributes to an estimated 125,000 preventable deaths annually in the United States [1]. For MA seniors, characterized by complex polypharmacy regimens (8–20 medications) and multiple chronic conditions, these adherence failures not only drive adverse clinical outcomes such as hospitalizations, disease progression, and mortality, but also directly impact MA plan Star Ratings, which determine federal bonus payments worth millions of dollars annually.

The Centers for Medicare & Medicaid Services (CMS) evaluates medication adherence using Proportion of Days Covered (PDC) across three therapeutic classes: diabetes medications, statins, and Renin-Angiotensin System (RAS) antagonists. Plans are measured against an 80% PDC threshold, yet clinicians typically lack proactive tools to identify at-risk patients before they fall below this critical threshold.

Problem Statement. Current workflows face three challenges: (1) reactive, not provident, identification after non-adherence has occurred, (2) data fragmentation across EHRs and pharmacy claims, and (3) inability to scale manual outreach to large populations.

Key Contributions. DosePilot addresses these gaps with a multi-pronged approach: (1) an end-to-end FHIR-based adherence monitoring pipeline with CMS-aligned PDC computation, (2) a predictive risk model using baseline features to identify at-risk patients prospectively, and (3) multi-stakeholder dashboards enabling actionable interventions for providers, payers, care managers, and patients.

2 Methodology

2.1 System Architecture

DosePilot employs a three-tier architecture: a React/TypeScript frontend providing role-specific dashboards, a FastAPI backend exposing RESTful endpoints, and a data/ML layer handling FHIR ingestion, PDC computation, and risk prediction using scikit-learn.

2.2 Data Ingestion Pipeline

The system ingests healthcare data using the HL7 FHIR R4B standard via a two-pass algorithm ensuring referential integrity:

- **Pass 1:** Parse FHIR Bundle JSON files, extract `Patient` resources with demographics, and build a patient reference map.
- **Pass 2:** Process clinical resources by patient reference—`MedicationRequest`/`MedicationDispense` (medications and fills), `Condition` (diagnoses), `Encounter` (utilization patterns), and `Observation` (labs/vitals).

Medications are normalized into CMS Star Ratings therapeutic classes using keyword-based pattern matching:

Table 1: Drug Class Normalization Examples

Drug Class	Example Medications
Diabetes	Metformin, Glipizide, Insulin, Empagliflozin
Statins	Atorvastatin, Rosuvastatin, Simvastatin
RAS Antagonists	Lisinopril, Losartan, Valsartan

2.3 PDC Calculation

PDC is computed using the CMS-standard methodology:

$$PDC = \frac{\text{Days with medication coverage}}{\text{Total days in measurement period}} \times 100\%$$

We employ a per-patient measurement window (first 12 months from first fill) to maximize sample size while maintaining the standard 365-day period. An interval-merging algorithm handles overlapping fills, counting each covered day only once, and days supply values are sanitized (default 30, clamped to $[1, 180]$).

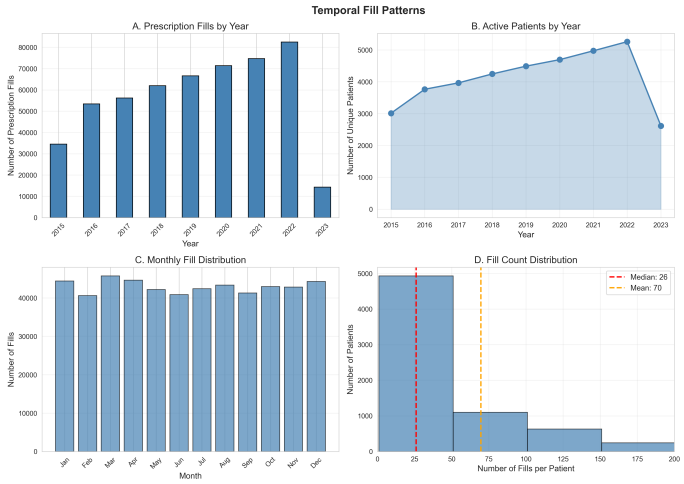

Figure 1: The year-over-year increase in prescription fills, illustrating the expanding data volume across the 2015–2023 period.

2.4 Risk Prediction Model

We employ a **Random Forest classifier** to predict non-adherence using features available at baseline (before measurement) to avoid data leakage. The model enables prospective identification of at-risk patients.

Feature Engineering (15) & Training Configuration:

Table 2: Key Features for Random Forest Model

Feature Category	Key Features
Prior Adherence	Prior year PDC (23.9% importance)
Fill Patterns	Total, fills per quarter, days supply stats
Refill Behavior	Maximum/mean refill gap, late refills
Demographics	Age, gender, disability status
Prescription Mix	Unique medications, brand/generic ratio

Our training configuration worked on a temporal split, training on 2015→2020 predictions and testing on 2021→2022 predictions. We provide class balancing via sample weighting and also implement 5-fold cross validation for hyperparameter tuning.

3 Datasets & Data Sources

3.1 CMS Synthetic Claims Data

Our primary data source is the CMS Synthetic Part D Prescription Drug Event (PDE) dataset, consisting of synthetic but realistic Medicare prescription claims data. This dataset provides:

Table 3: CMS Synthetic PDE Dataset Metrics

Metric	Value
Prescription Fills	515,520
Unique Patients	7,403
Time Period	2015–2023 (9 years)
Annual Growth	34K fills (2015) → 82K (2022)

Demographics: The average age of our sample population was 60 years (50% below 65 indicating disability enrollment), with a 51%/49% male/female split, and a racial make up of 63% White, 18% Hispanic, and 11% Black individuals.

3.2 Synthea FHIR Bundles

For demonstrating the FHIR ingestion pipeline, we additionally processed 557 generated patient bundles from Synthea, a synthetic data generator that models the life and medical history of synthetic patients, containing: patient demographics and identifiers, `MedicationRequest` / `MedicationDispense`, `Condition`, `Encounter`, and `Observation` resources.

However, it is important to note that Synthea generates sparser medication fills than real claims, resulting in artificially low PDC values, and some `MedicationRequest` resources lack explicit `days_supply` (defaulted to 30). We believe that these extra data-points provide additional diversity to our dataset and can help us make predictions more reflective of real world patient data.

4 Results

4.1 Adherence Distribution

Analysis of computed PDC values reveals significant non-adherence in the population:

Table 4: PDC Distribution Analysis

Metric	Value
Mean PDC	44.4%
Median PDC	33.6%
Adherent (≥80% PDC)	18.7%
Non-adherent (<80% PDC)	81.3%

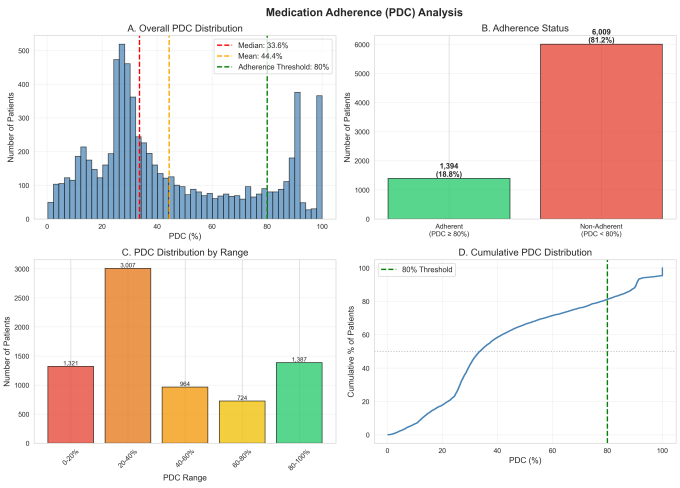

Figure 2: The full PDC distribution, highlighting the long tail of low-adherence patients.

Key Finding: Year-over-year adherence is highly stable—PDC correlation between consecutive years is **r = 0.811**, with 88% of

patients maintaining their adherence status year-over-year. This persistence validates using prior adherence as a predictive feature.

4.2 Model Performance

The Random Forest model achieves strong predictive performance on held-out test data (2021→2022 predictions, $n = 3,905$):

Table 5: Random Forest Model Performance

Metric	Value
ROC-AUC	**97.9%**
Accuracy	94.5%
Precision	93.7%
Recall	91.6%
False Positive Rate	3.8%

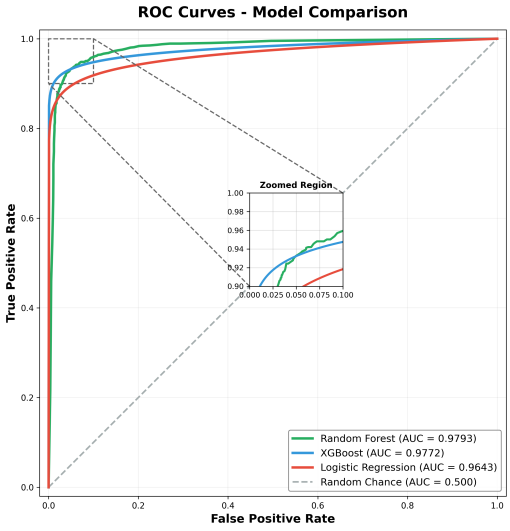

Figure 3: Model Comparison: Random Forest outperforms Logistic Regression (96.4% AUC) and XGBoost (97.7% AUC).

Calibration: The model is well-calibrated (error 0.042)—when predicting 80% probability, ∼80% of patients are actually adherent.

4.3 Feature Importance

Feature importance analysis reveals the primary drivers of adherence prediction:

Table 6: Top 5 Feature Importances

Rank	Feature	Importance
1	Prior year PDC	23.9%
2	Maximum refill gap	12.5%
3	Fills per quarter	10.1%
4	Total fills	9.7%
5	Days supply (median)	8.6%

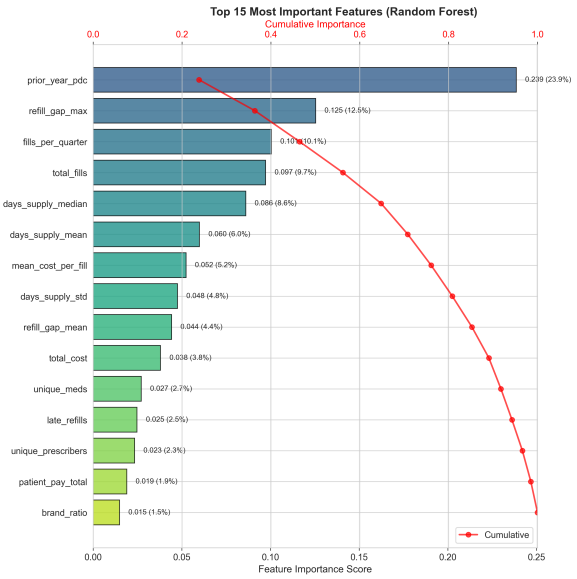

Figure 4: The full feature-importance ranking from the Random Forest model.

Top 5 features explain 64.8% of model decisions. The dominance of prior adherence confirms the behavioral persistence observed in year-over-year stability analysis.

4.4 Ablation Study

To quantify feature contributions, we trained models on feature subsets:

Table 7: Feature Ablation Study Results

Feature Set	ROC-AUC
Prior PDC Only	94.8%
Fill Patterns Only	∼88%
Refill Gaps Only	∼88%
Costs Only	∼70%
All Features	**97.9%**

Marginal improvement from additional features: +3.1% beyond prior adherence alone. While prior behavior is the strongest predictor, fill patterns and refill gaps provide meaningful additional signal.

4.5 Risk Stratification

The model enables clinically useful risk stratification:

This stratification enables targeted outreach—patients in the "Very High Risk" tier have 20× lower adherence rates than "Very Low Risk" patients.

5 Discussion

5.1 Key Findings

Adherence is persistent behavior. The 0.811 year-over-year PDC correlation indicates that adherence is a stable behavioral

pattern, not random variation. This supports using prior adherence as a predictive feature and suggests that early intervention may have lasting effects.

Baseline features enable prospective prediction. By using only features available before the measurement period, the model can identify at-risk patients before non-adherence occurs—enabling proactive rather than reactive intervention.

FHIR-based monitoring is feasible. Our pipeline demonstrates that real-time adherence monitoring can be built entirely on FHIR-standard data without proprietary integrations, enabling broad interoperability across EHR systems.

The 97.9% ROC-AUC and 3.8% false positive rate make this model suitable for prioritizing outreach resources. Care managers can focus on high-risk patients with confidence that few "false alarms" will waste effort. The 91.6% recall ensures most truly non-adherent patients are identified.

5.2 Limitations & Future Work

1. **Synthetic data constraints:** CMS synthetic data may not capture all real-world adherence patterns (e.g., pharmacy switching, hospitalization-induced gaps)
2. **Single-class normalization:** Keyword-based drug classification may miss brand names or combination medications
3. **No intervention validation:** Without randomized trials, we cannot confirm that model-directed outreach improves outcomes
4. **Limited social determinants:** The model lacks socioeconomic features (income, transportation access) known to influence adherence

In future experiments and further development of our platform, we believe it important to validate our results with access to more real MA plan claims data, integrate RxNorm/RxNav APIs for more authoritative drug classification, implement outcome tracking to measure intervention effectiveness, and explore various deep learning techniques for temporal pattern recognition.

6 Conclusion & Main Findings

DosePilot demonstrates a complete, functional pipeline for AI-driven medication adherence monitoring targeting Medicare Advantage seniors. Our contributions include:

1. **End-to-end FHIR ingestion** processing 500K+ prescription fills with CMS-aligned PDC computation
2. **High-accuracy risk prediction** (97.9% ROC-AUC) using baseline features only, enabling prospective identification
3. **Behavioral insight:** Year-over-year adherence stability ($r = 0.811$) confirms adherence as persistent behavior
4. **Multi-stakeholder design** providing role-appropriate interfaces for clinical, administrative, and patient use

Main Findings:

- Prior adherence explains 94.8% of predictive performance; additional features add 3.1%
- 81.3% of patients are non-adherent, representing substantial intervention opportunity
- Risk stratification achieves 20× separation between highest and lowest risk tiers
- The 88% year-over-year status stability suggests early intervention may have lasting impact

If deployed with real Medicare Advantage data, DosePilot could improve Star Ratings through proactive intervention, reduce costs through avoided hospitalizations (estimated $720/member/year), and scale care management through intelligent prioritization.

References

1. Cutler RL, et al. Economic impact of medication non-adherence by disease groups. *BMJ Open.* 2018;8(1):e016982.
2. Centers for Medicare & Medicaid Services. Medicare 2024 Part C & D Star Ratings Technical Notes. 2023.
3. Winkler J, et al. Synthea: Generating synthetic patients and EHRs. *JAMIA*. 2018;25(3):230-238.
4. HL7 International. Fast Healthcare Interoperability Resources (FHIR) Release 4. 2019.
5. Pedregosa F, et al. Scikit-learn: Machine Learning in Python. *JMLR.* 2011;12:2825-2830.