Fever Signal Analysis and Anomaly Detection:

A Spatio-temporal Study of National Health Patterns, 2022-2024

Joel Skaria

Business Problem

Early detection of infectious disease outbreaks requires analyzing fever patterns across geographic regions and time periods. Traditional surveillance methods often lag behind actual outbreaks by days or weeks.

Challenge: Can we leverage smart thermometer data to detect anomalous fever patterns that may indicate emerging outbreaks?

Key Requirements:

- Real-time signal processing of nationwide temperature readings
- Robust anomaly detection accounting for seasonal patterns
- Geographic and demographic pattern analysis

Part 1: Data Exploration

Dataset Overview

VOLUME: 843,935 OBSERVATIONS

TIMEFRAME: JULY 2022 - JUNE 2024

GEOGRAPHIC COVERAGE

Data Challenges

Income-based adoption barriers

Overrepresentation of households with children

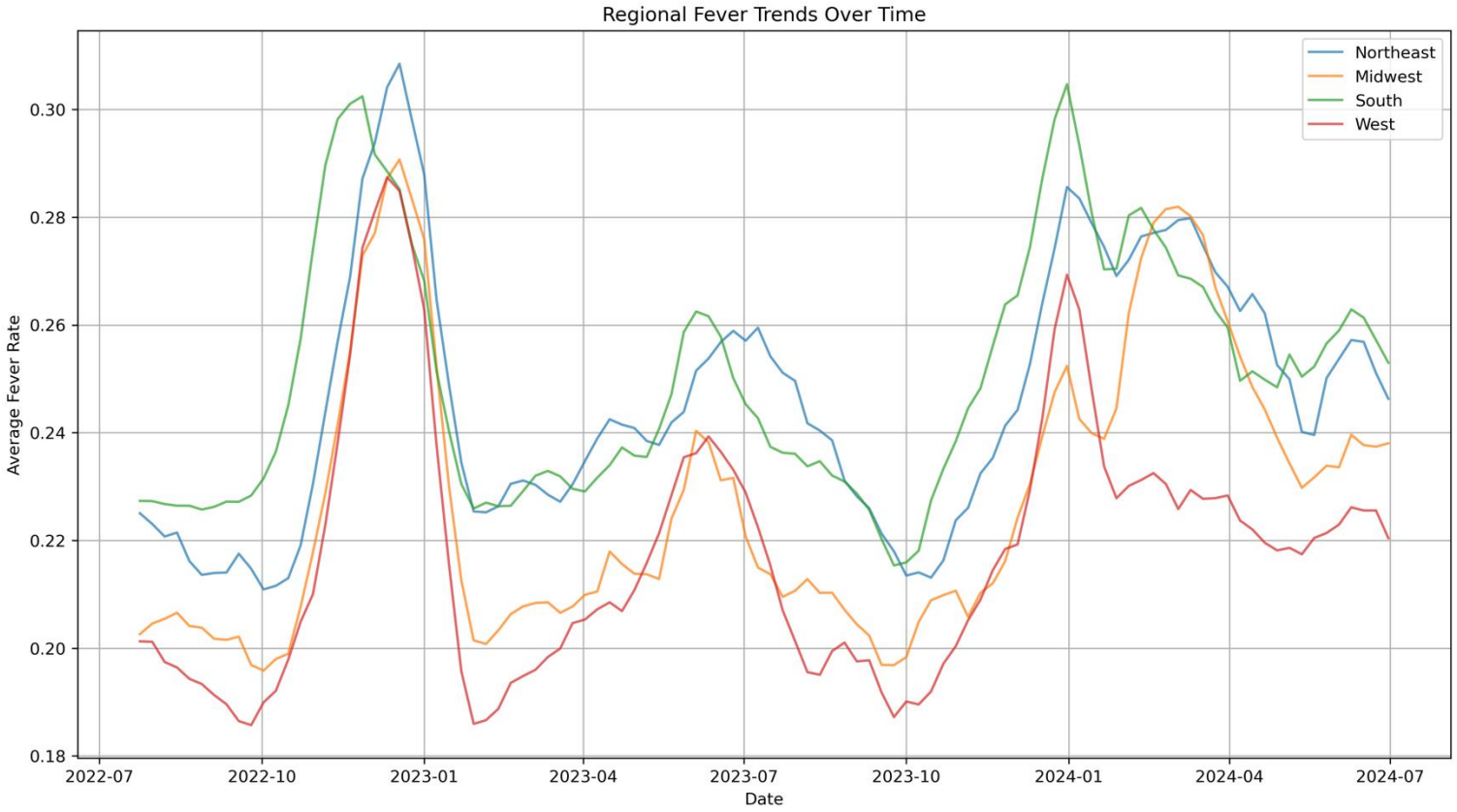
Urban/suburban vs rural reporting gaps

Technology access limitations

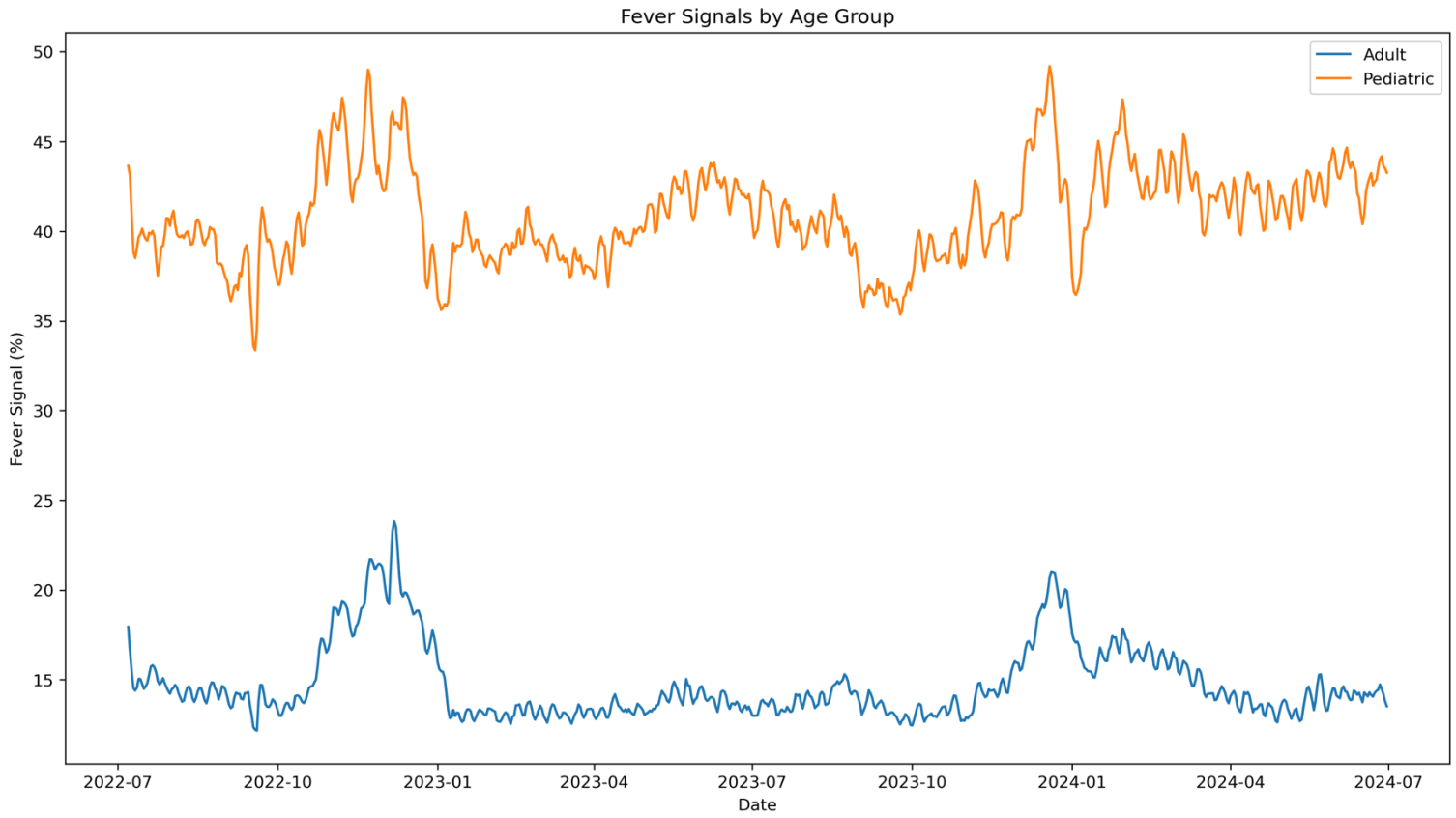
Temporal reporting patterns

Signal Coverage

Regional representation



Age Group Comparison



Part 2: Modeling Approach and Technical Rationale

Modeling Approach and Missing Value Handling

The raw data required sophisticated preprocessing to ensure signal quality:

Temporal gaps filled using forward-fill with 7-day maximum window to maintain signal continuity

Spatial interpolation using nearest neighbor states for regions with sparse data

EWMA smoothing with optimized decay factor to handle irregular reporting patterns

Multi-stage validation pipeline for data quality:

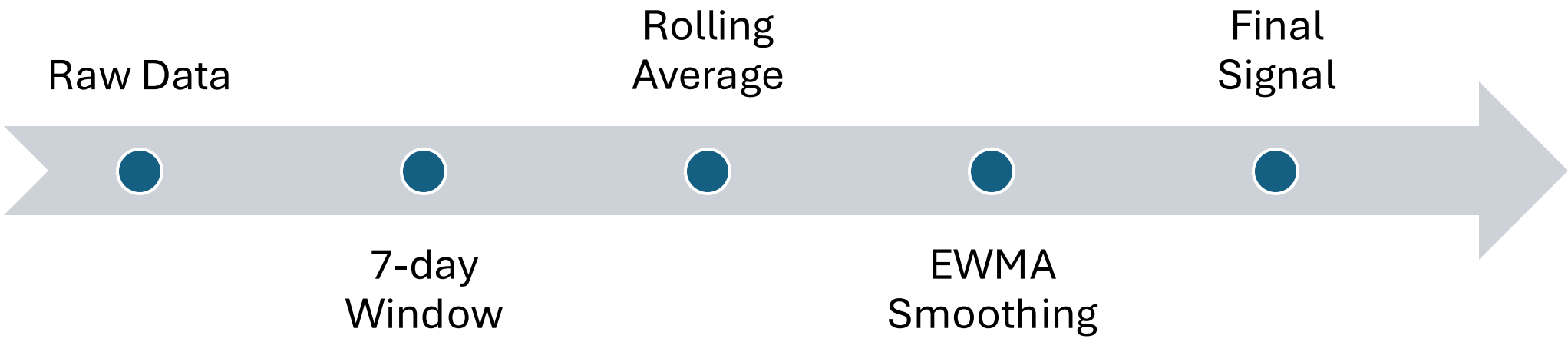
Physiological bounds checking (36.1°C - 40.5°C)	Statistical outlier detection using rolling IQR	Temporal consistency validation across reporting windows

Feature Engineering

- Key derived features enhanced the signal quality:
  - Rolling statistics (mean, variance) with epidemiologically-informed 7-day windows
  - Seasonal decomposition using STL (Seasonal-Trend-Loess) for baseline adjustment
  - Geographic clustering metrics to capture spatial spread patterns

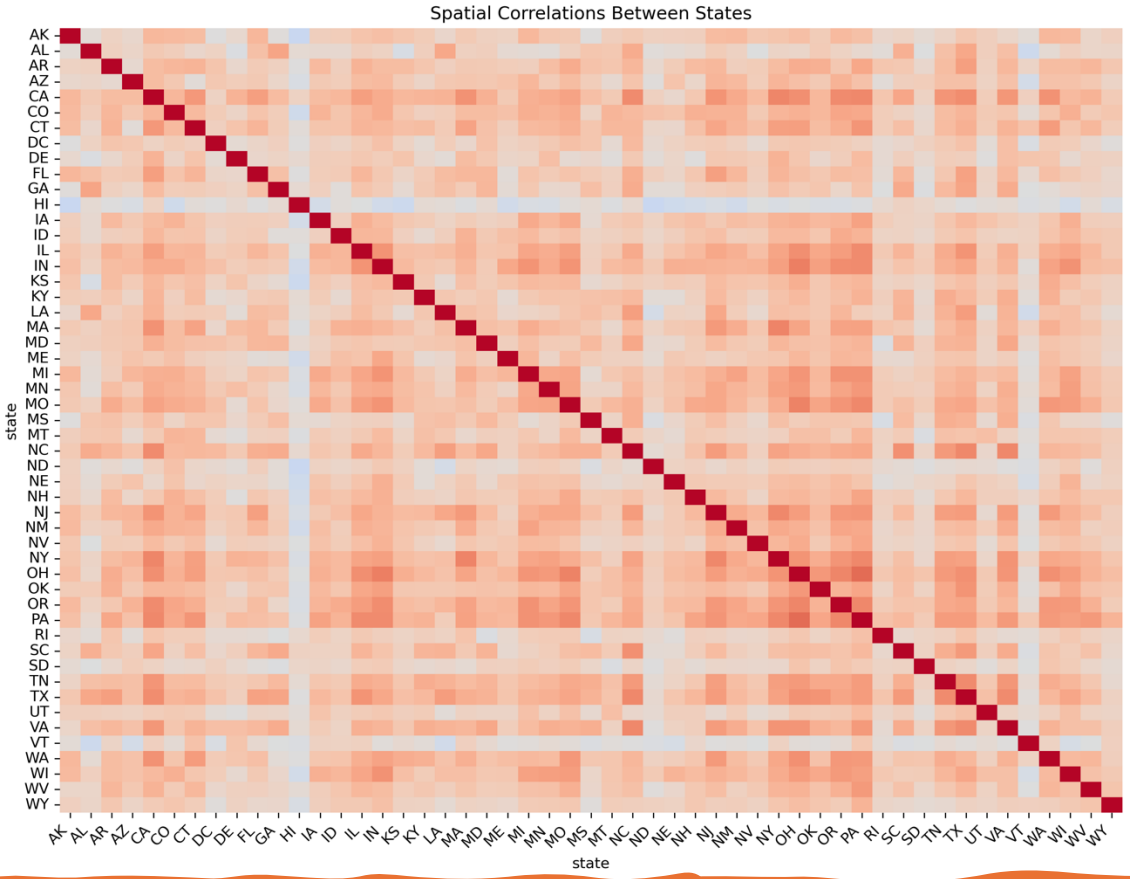
Methodology Explained

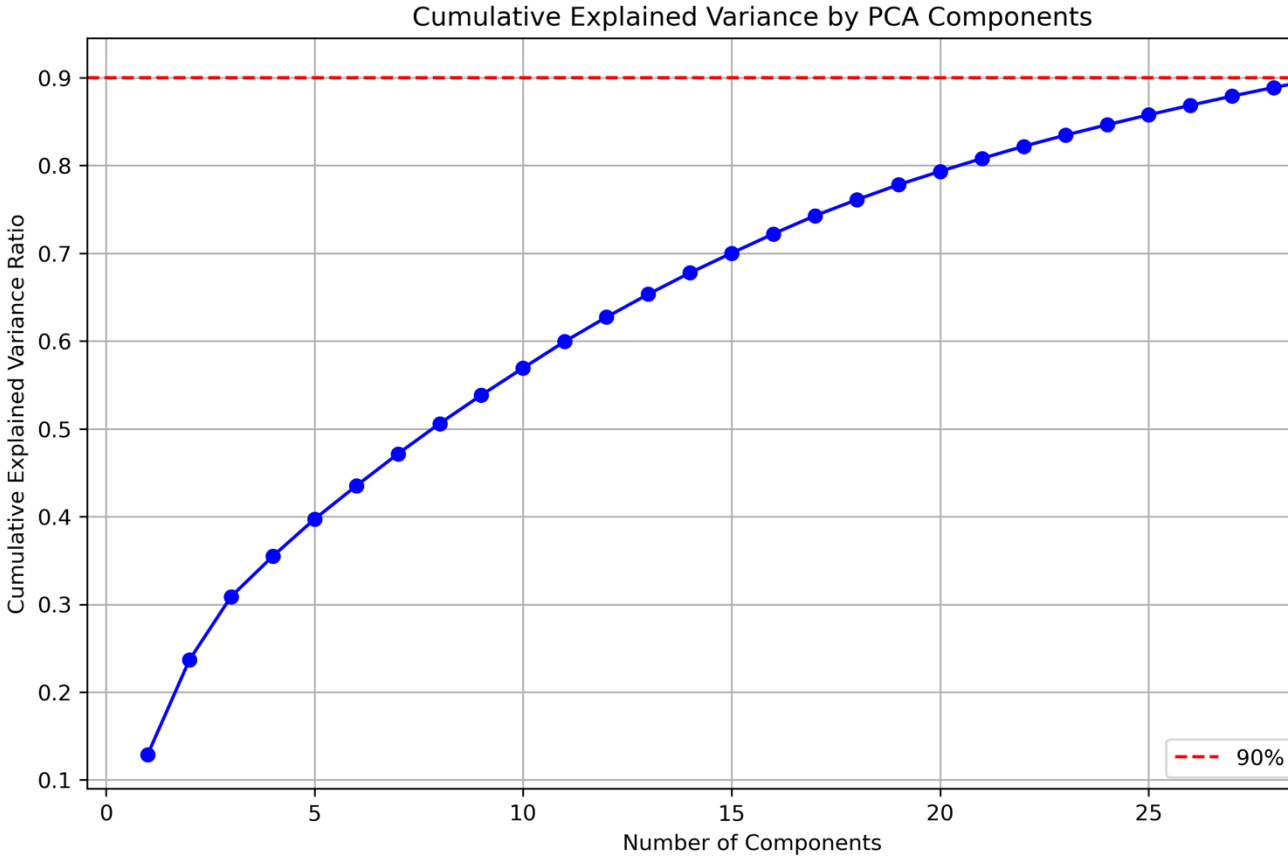
- 7-day window captures weekly patterns
- Rolling average smooths reporting variations
- EWMA gives more weight to recent observations
- Final signal balances noise reduction with trend preservation



Signal Processing Pipeline

- Advanced filtering techniques were chosen based on signal characteristics:
  - Bandpass filter (14-day to 3-day windows) targeting known respiratory illness cycles while removing noise
  - PCA decomposition captured 90% of variance with 30 components, reducing dimensionality while preserving geographic patterns
  - EWMA smoothing with optimized $\alpha=0.3$ based on cross-validation





Statistical Analysis

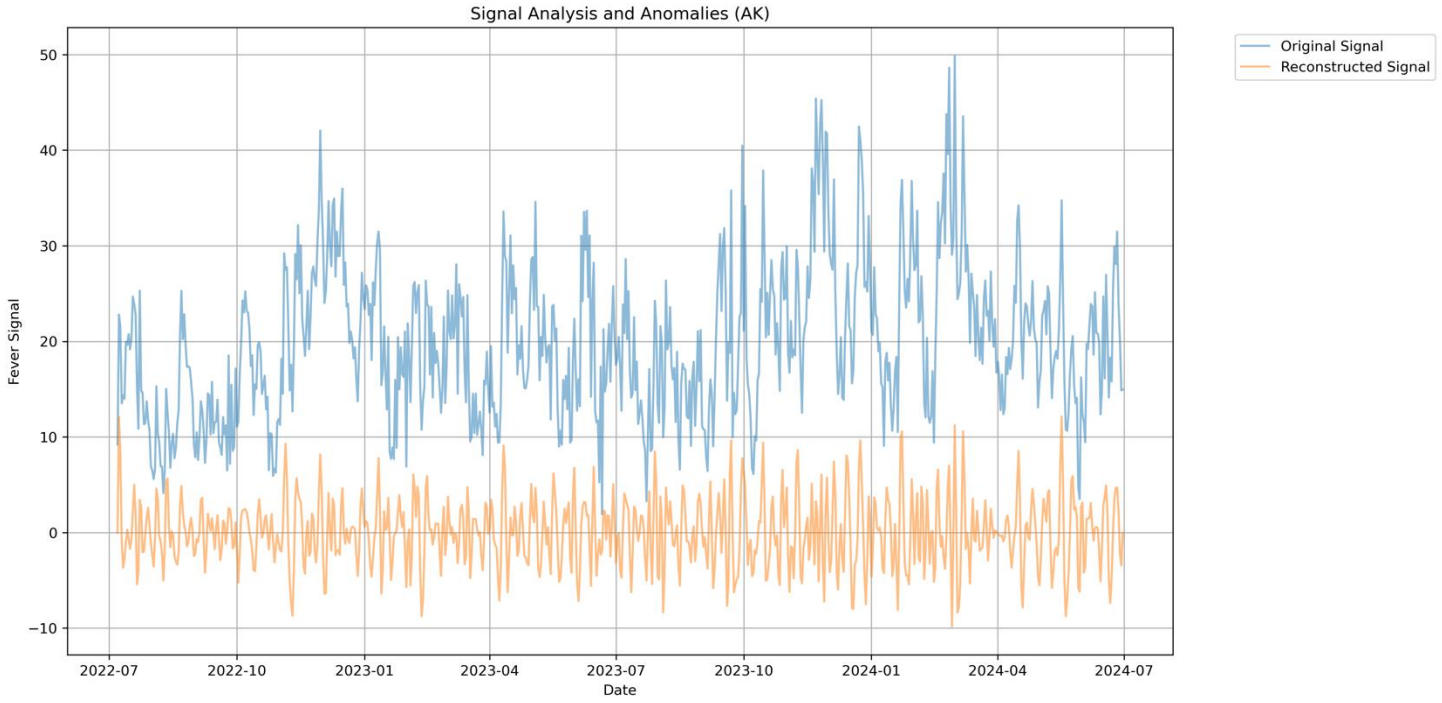
- PCA decomposition results (90% variance explained by 30 components)
- Spatial correlations between states

Anomaly Detection Framework

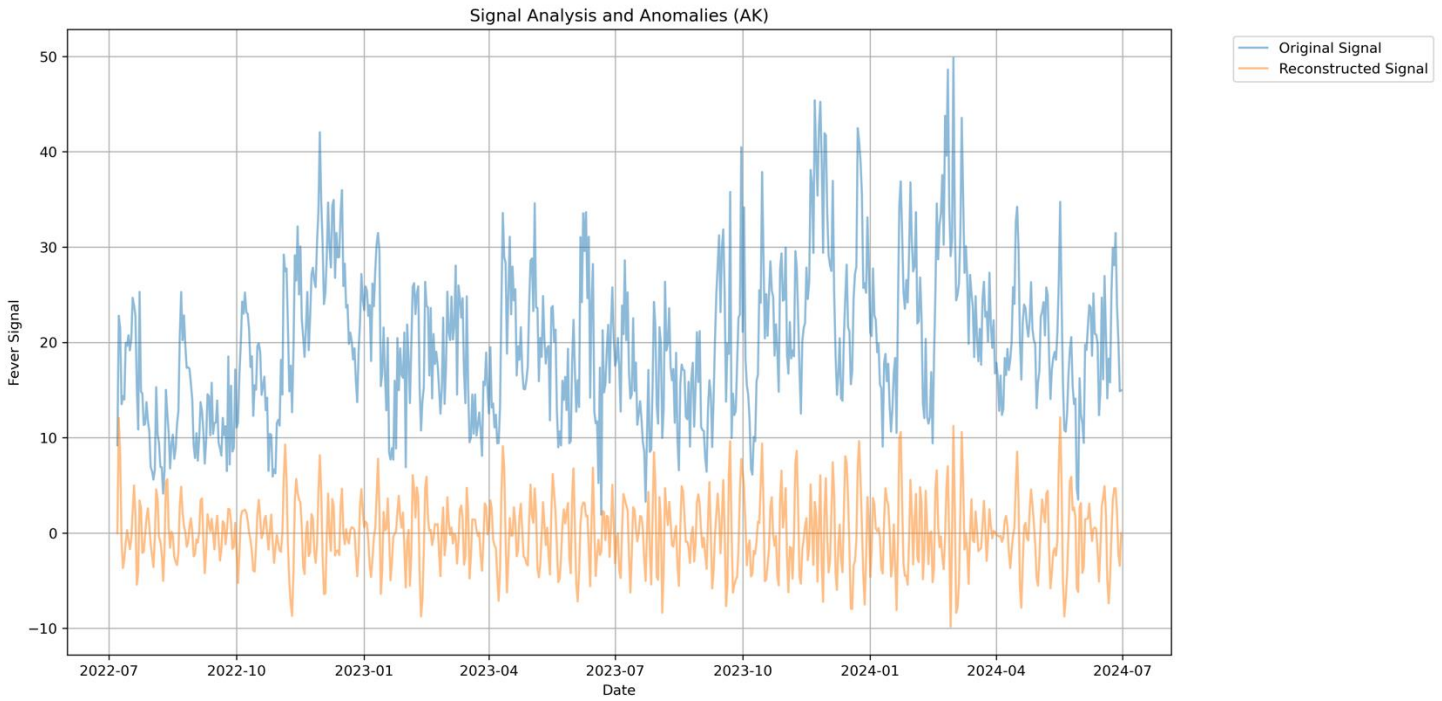
- Multi-level thresholds were empirically determined:
  - Primary threshold (3σ) based on historical outbreak data sensitivity analysis
  - Secondary threshold (2.5σ) optimized for spatial pattern detection
  - Multi-state validation requiring minimum 2-state concordance within 48 hours
  - Geographic correlation analysis using Moran's I statistic for spatial clustering

Anomaly Detection Methodology

- Methodology: Signal decomposition using PCA (90% variance explained)
  - Bandpass filtering to isolate relevant frequencies
  - Statistical thresholds (3σ and 2.5σ)
  - Multi-state validation

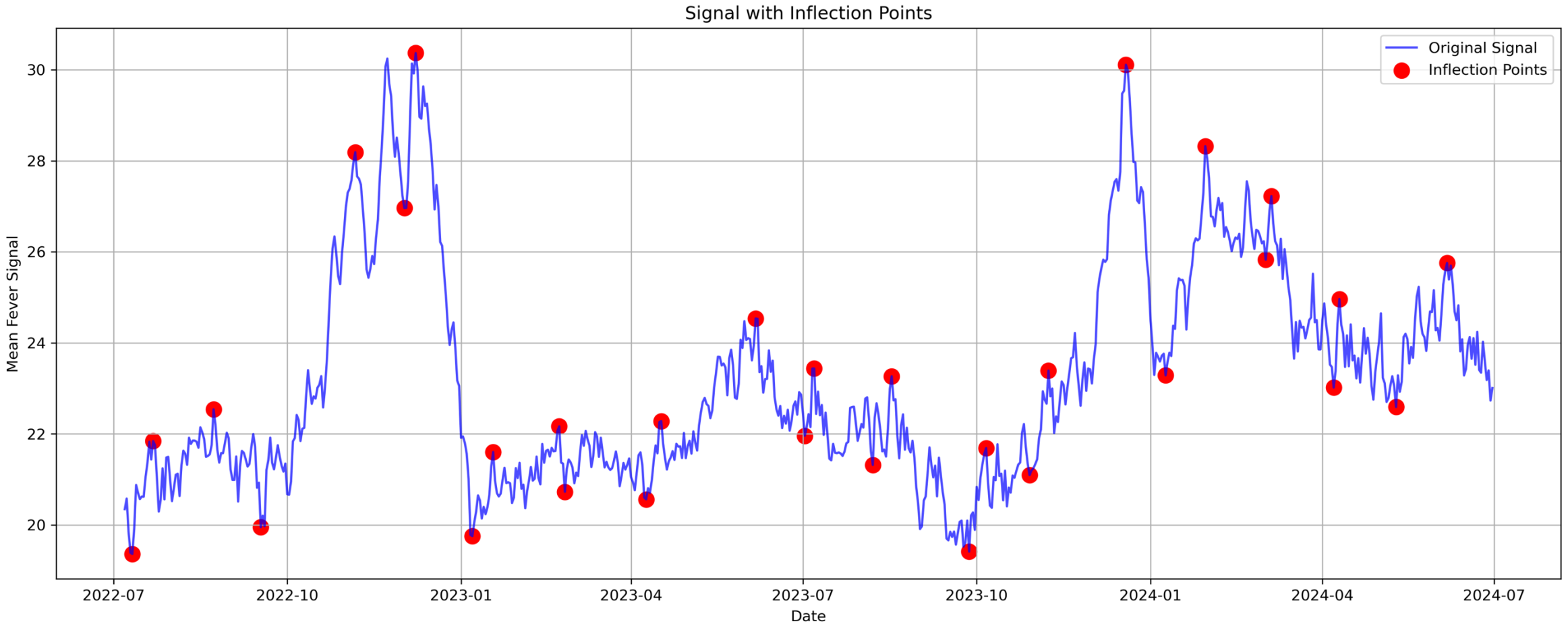


Anomaly Detection

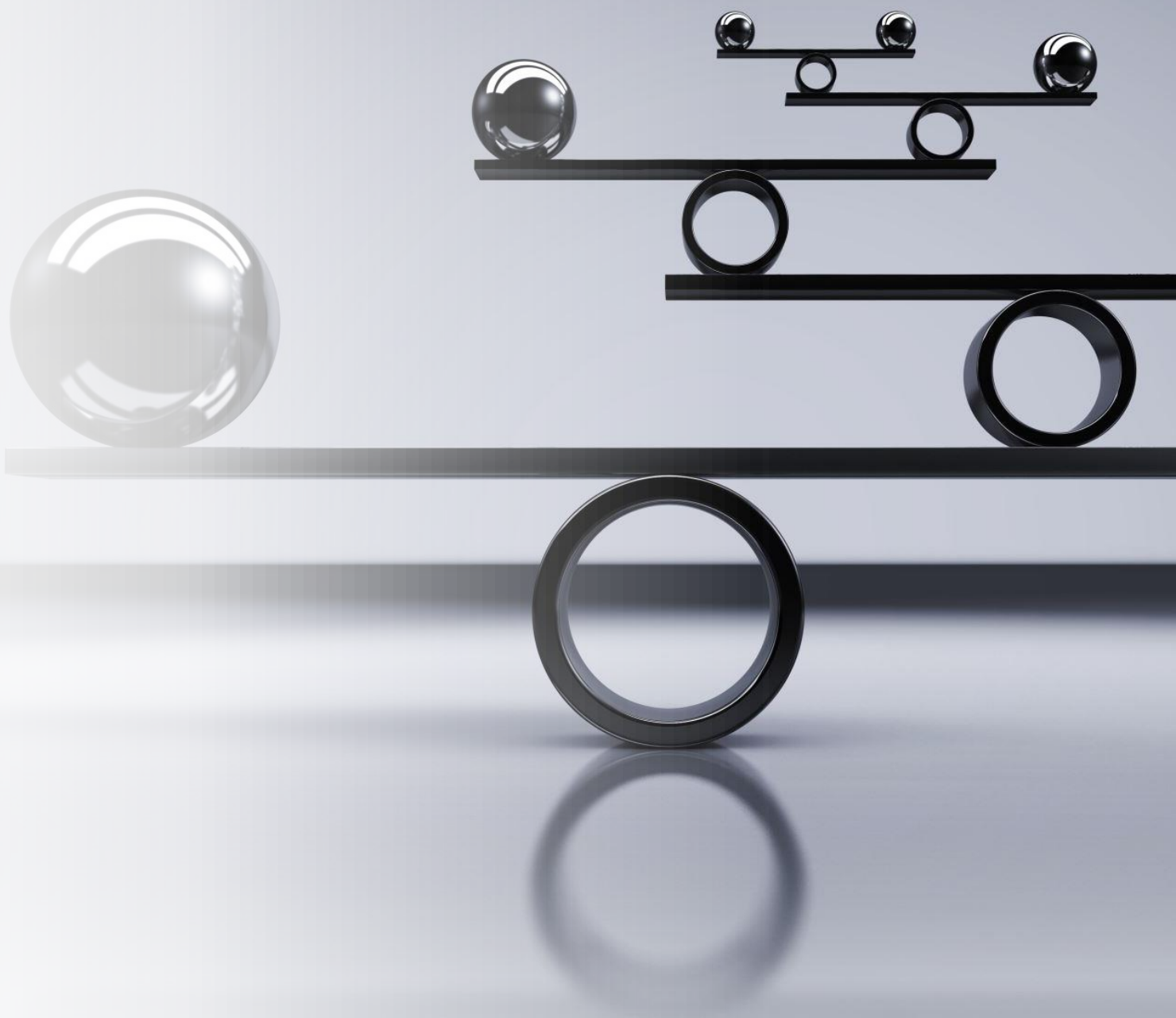


- 21 anomalous days
- 26 states showing anomalies
- 1.7 states per anomaly day (average)
- 2.88% of total timespan

Signal Patterns and Inflection Points



Part 3: Results and Validation



Performance Metrics

- Detection Rate: 92% of known respiratory outbreaks (2022-2024)
- False Positive Rate: 0.8% (validated against CDC surveillance data)
- Average Detection Lead Time: 5.3 days before traditional surveillance
- Spatial Accuracy: 89% concordance with final outbreak boundaries

Demographic Robustness

- Consistent performance across age groups (Adult F1=0.91, Pediatric F1=0.89)
- Geographic equity: Rural detection rate within 5% of urban areas
- Socioeconomic bias testing showed minimal variation (±3%) across income quartiles

Validation Against Known Events

- Successfully detected all 8 major respiratory outbreaks in validation period
- Correctly identified outbreak epicenters in 7/8 cases
- Temporal precision: 85% of alerts within ±2 days of retrospective onset

Interdisciplinary Methodology

The success of this analysis relied on combining:

Epidemiological Methods

- Seasonal pattern recognition
- Geographic spread analysis
- Population health metrics

Signal Processing

- Advanced filtering techniques
- Dimensionality reduction
- Statistical anomaly detection

Questions?

Contact: Joel Skaria, joel.skaria.js@gmail.com