

Artificial Intelligence for Patient Data and Drug Safety: Insights for Safer Care

Dr Sakinat Folorunso

Associate Professor of AI Systems and FAIR Data

Olabisi Onabanjo University

**Evidence-Based Medicine Research &
Pharmacovigilance Training**

9/09/2025



Objectives

Understand	Understand AI and ML fundamentals in healthcare
Identify	Identify types and sources of patient data
Apply	Apply basic AI techniques
Interpret	Interpret AI outputs
Address	Address ethical and privacy concerns

Presentation Outline



Introduction to AI, machine learning (ML), and deep learning (DL)



Overview of AI applications in healthcare



Types of patient data:
structured vs unstructured



Data quality, bias, and
preprocessing in healthcare
datasets

AI, ML, and DL:

Definitions and Analogies



Healthcare
Analogy



Artificial Intelligence (AI)

AI refers to computer systems designed to mimic human intelligence, capable of performing tasks like interpreting data, making decisions, and solving problems.

Healthcare Analogy

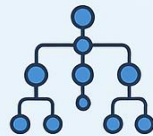


A seasoned doctor rapidly diagnoses and treats illnesses using extensive clinical experience.

Pharmacovigilance



An experienced safety officer swiftly recognizes potential adverse drug reactions (ADRs) by evaluating multiple patient reports and drug safety data.



Machine
Learning (ML)



Machine Learning (ML)

ML is a subset of AI that involves algorithms learning from historical data, improving predictions and decisions without explicit programming.

Healthcare Analogy



A medical resident enhances their diagnostic skills by systematically reviewing and learning from numerous patient

Pharmacovigilance



A trainee pharmacist learns to identify medication risks by analyzing multiple patient medication records and recognizing patterns of drug-related issues.



Deep
Learning (DL)

Deep Learning (DL)

DL is a specialized form of ML using multi-layered neural networks to detect intricate patterns and insights from complex datasets automatically.

Healthcare Analogist



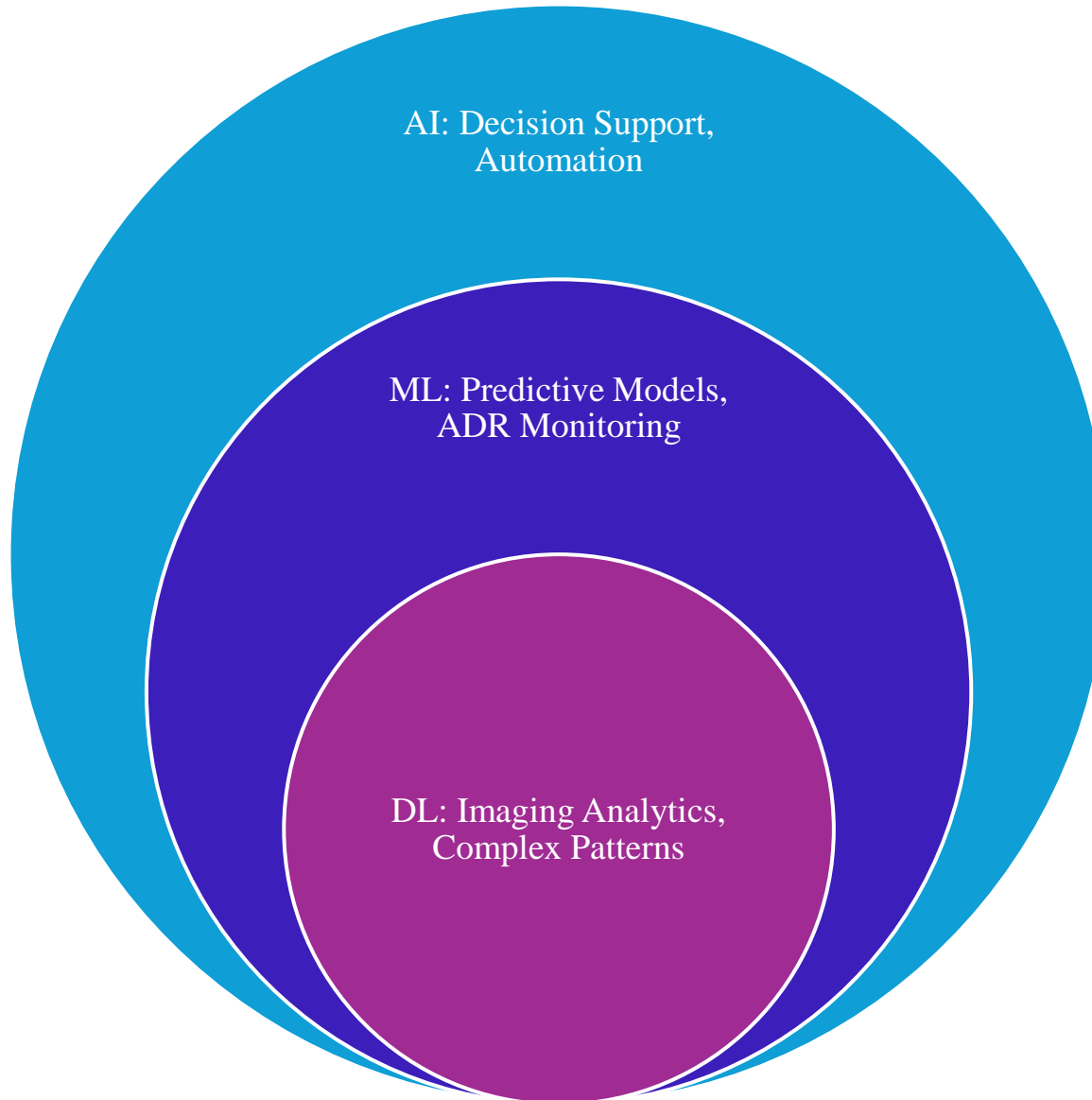
A specialist radiologist interpreting complex

Pharmacovigilance



An expert toxicologist closely examining complex,

Relationship between AI, ML, and DL



AI Applications in Healthcare



Diagnostic imaging



Predictive analytics
(patient outcomes,
readmissions)



Pharmacovigilance
using NLP



AI supports clinicians
with data-driven
decision-making.



Enables faster, more
targeted diagnostics
and interventions.



Moves healthcare from
reactive to proactive
care.



Think of AI as a clinical
assistant trained on
millions of cases.

Definitions and Examples of Structured and Unstructured Data

Structure Patient Data

- a. Organized into rows and columns (e.g., databases, spreadsheets).
- b. Easily searchable, analyzable, and standardized.
- c. Examples: EHR tables, lab results, vitals, and medication records.
- d. Dataset Example: MIMIC-III database (ICU patient data).

Unstructure Patient Data

- a. Free-form data without a pre-defined structure.
- b. Challenging to analyze directly without NLP, imaging, or audio processing.
- c. Examples: Physician notes, radiology images, audio dictations.
- d. Dataset Example: Sample clinical notes or discharge summaries from MIMIC-III.

Structured vs. Unstructured Data

Structured Data

Highly organized information stored in predefined formats like tables or databases

Blood pressure
= 120/80 mmHg

- Laboratory test results
HbA1c = 6.2%
- Medication lists
and dosages

Unstructured Data

Information not stored in a predefined data model or format

“Patient complains of chest tightness and fatigue...”



Radiology images



Voice recordings



Examples of both structured and unstructured patient data types,



Structured Data (Tabular Format)

Patient ID	Temp (°C)	BP (mmHg)	Heart Rate (bpm)	Diagnosis Code	Medication	Lab Result
001	37.2	120/80	82	I10	Atenolol	HbA1c = 6.4
002	38.0	135/85	90	J20	Azithromycin	WBC = 11.2
003	36.6	115/70	75	E11	Metformin	Glucose = 140

Unstructured Data (Free Text Narrative)

- **Example 1 – Physician Note:**
“Patient reports fatigue, shortness of breath, and mild fever. Vitals stable. Suspect early-stage infection. Plan: start empirical antibiotics and monitor for 48 hours.”
- **Example 2 – Adverse Drug Reaction Report (Patient):**
“Since I started the new blood pressure medicine, I’ve been feeling dizzy in the mornings. Sometimes I get a rash on my arms, too.”
- **Example 3 – Audio Transcription (Consultation Summary):**
“Patient says they’re tolerating Metformin but occasionally feel nauseated after meals. No signs of hypoglycemia reported.”

Issue	Why It Matters in Pharmacovigilance / Medicine	Typical Example	Practical Fixes & Mitigation
Missing values	Spontaneous safety reports (ICSRs) or EHR extracts may lack dose, onset date, or lab confirmation—hampering signal detection and case assessment.	12 % of Individual Case Safety Reports omit time-to-onset , making disproportionality analysis less reliable.	<ul style="list-style-type: none"> • Follow-up queries to reporters / clinicians. • Domain-aware imputation (e.g., impute onset date from median latency in similar cases). • Flag “critical missing fields” to rout cases for manual review before database lock.
Noise & outliers	Implausible doses, negative ages, or duplicate MedDRA codes inflate or mask drug–event signals.	Dose recorded as “50 000 mg” instead of “50 mg” triggers a false safety alert.	<ul style="list-style-type: none"> • Range rules & units checks at point of data entry. • Automated outlier detection (IQR/Z-score) followed by pharmacovigilance-expert review. • Maintain a drug-specific plausibility table (e.g., maximum daily oral dose).
Duplicate records	The same ADR case can flow into EV/FAERS from multiple reporters; duplicates over-count events and bias disproportionality metrics.	A hospital EHR export and a spontaneous report both describe the identical anaphylaxis case—counted twice in signal analysis.	<ul style="list-style-type: none"> • Deterministic / probabilistic de-duplication using patient initials + age + event date + drug. • Apply WHO UMC vigiMatch or similar duplicate-detection algorithms.
Bias <ul style="list-style-type: none"> – Selection – Measurement – Demographic 	Skewed case mix distorts risk estimates and ML model performance.	<p>Selection: 80 % of ADR reports come from tertiary hospitals—community reactions are under-captured.</p> <p>Measurement: A new e-reporting app asks for “severity” on a 1–3 scale, and an older paper form uses SOC codes to shift the severity distribution.</p> <p>Demographic: Under-reporting in paediatrics yields sparse safety data for children.</p>	<ul style="list-style-type: none"> • Post-stratification weighting or inverse-probability weighting when analysing signals. • Harmonise data-collection instruments; apply calibration factors if devices/scales change. • Targeted awareness campaigns to boost reporting in under-represented groups (e.g., paediatrics, geriatrics).

Common dataset issues with examples and fixes from real datasets

Data Preprocessing Techniques in Medicine & Pharmacovigilance

Preprocessing transforms raw, inconsistent data into clean, reliable input for meaningful analysis and AI models.

1. Data Cleaning 🧹

Goal: Remove errors, fill gaps, and correct inconsistencies.

Task	Example (Pharmacovigilance/Medicine)	Why It Matters
Missing value handling	Fill in missing age in ICSR with median from similar reports	Prevents model distortion or analysis errors
Outlier removal	Remove Metformin dose = 50,000 mg	Prevents false signal detection or clinical misinterpretation
Duplicate detection	Same ADR case reported by hospital and pharmacy	Avoids overcounting in signal detection
Inconsistent units	Normalize "2 g" vs "2000 mg" to same unit	Ensures uniformity in dose-response analysis

PowerBI (Business Intelligence)

□ What is Power BI? – One Slide Overview

Facilitator: Dr. Sakinat Folorunso

□ Key Features

- Connect to data (Excel, CSV, SQL, cloud)
- Interactive visuals (charts, KPIs, maps)
- Data cleaning with Power Query
- AI features: Key Influencers, forecasting
- Build dashboards & share insights

□ Versions

- Power BI Desktop (Free, Windows app)
- Power BI Service (Pro/Premium for sharing)
- Power BI Mobile (view dashboards)

⌘ Uses in Healthcare

- Analyze patient data & outcomes
- Monitor Adverse Drug Reactions (ADRs)
- KPIs: Readmission %, Mortality %, LOS
- Share dashboards for decision support

□ *Power BI turns raw data into clear, interactive stories for better decision-making.*

PowerBI (Business Intelligence)

□ Power BI Desktop Quick Start - One Slide

Facilitator: Dr. Sakinat Folorunso

1□ Download

- Microsoft Store OR
- <https://aka.ms/pbiSingleInstaller>

2□ Install & Open

- ⚙ Setup wizard
- ▶ Start Menu → Power BI Desktop

3□ First-Time Setup

- Sign in OR 'Skip for now'

4□ Load Data

- Home → Get Data → Excel/CSV → Load

5□ Build Visuals

- Pick visual (bar, line, KPI, map)
- Drag fields to Axis/Values/Legend

6□ Create Report Pages

- Multiple visuals per page
- Add slicers (filters)

7□ Save & Share

- Save as .pbix
- ▲ Publish to Power BI Service

Why Power BI Works for Healthcare & Pharmacovigilance

Low learning curve.

Handles structured patient data

Built-in AI/ML features

Visualization-first

Integration

Simple AI Features in Power BI



Key Influencers Visual

Automatically shows which factors (e.g., age, dosage, comorbidities) influence outcomes like adverse drug reactions.



Decomposition Tree

Lets users explore data hierarchically (e.g., ADR by drug class → hospital → age group).



Forecasting & Trend Analysis

Time series forecasting for patient admission rates, adverse drug reports over months.



Q&A (Natural Language Queries)

Participants can *type questions* like “Which drug had the highest number of side effects in 2024?” and Power BI generates a visual.



Quick Insights

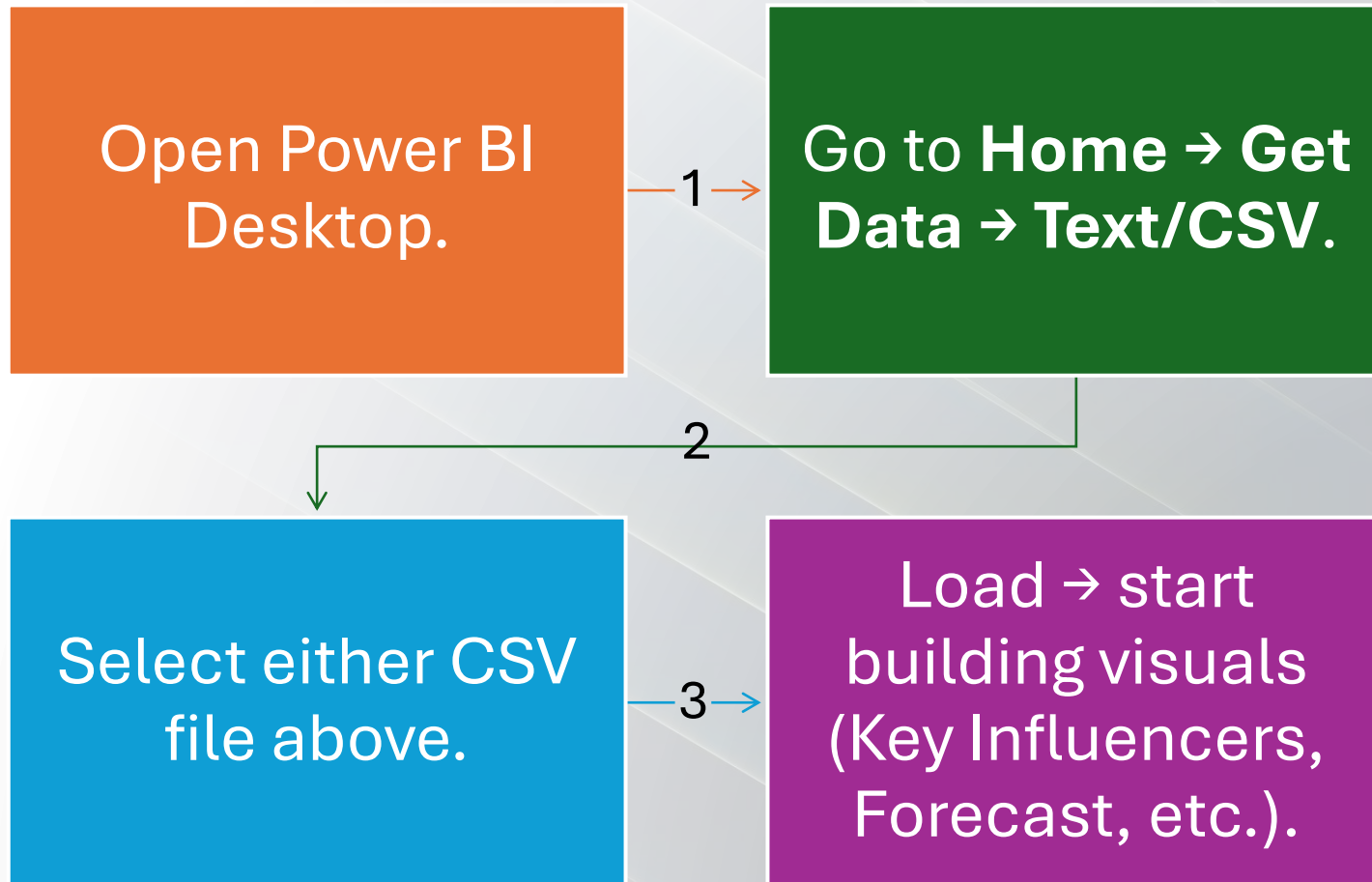
Power BI can scan datasets and highlight anomalies, trends, and correlations automatically.



Fast Power BI Demo (step-by-step)



How to use **Power BI**:



What we'll build in Power BI

3 visuals:

Key Influencers (risk drivers)

Trend forecast

Decomposition Tree

Key Influencers (Risk drivers) – Pharmacovigilance

1. Load *pharmacovigilance_adr_reports.csv*.
2. Insert **Key Influencers** visual.
3. *Analyze*: seriousness (treat "Serious" vs "Non-serious" as outcome).
4. *Explain by*: age, dosage_mg, drug_class, event_type, sex, onset_days, causality_score, concomitant_meds_count, state.
5. Talking point: dose, age ≥65, and specific drug classes emerge as top drivers → *clinical risk flags* for pharmacovigilance.

Key influencers Top segments

What influences seriousness to be Serious

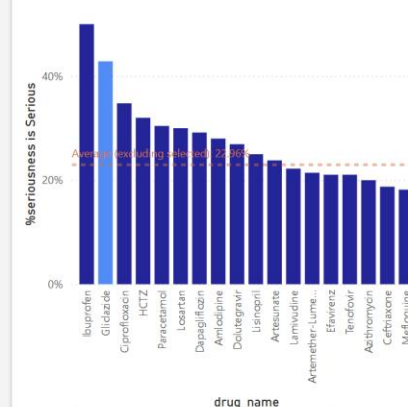
When...

...the likelihood of seriousness being Serious increases by

drug_name is Glucilazide

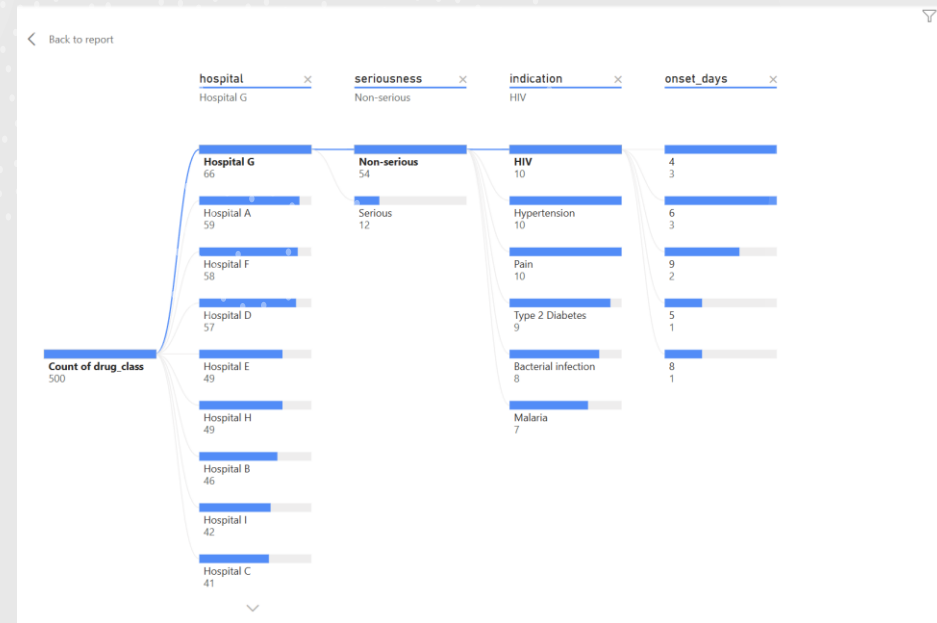
1.87x

← seriousness is more likely to be Serious when drug_name is Glucilazide than otherwise (on average).



☐ Only show values that are influencers

Decomposition Tree (Drill-downs) – Pharmacovigilance

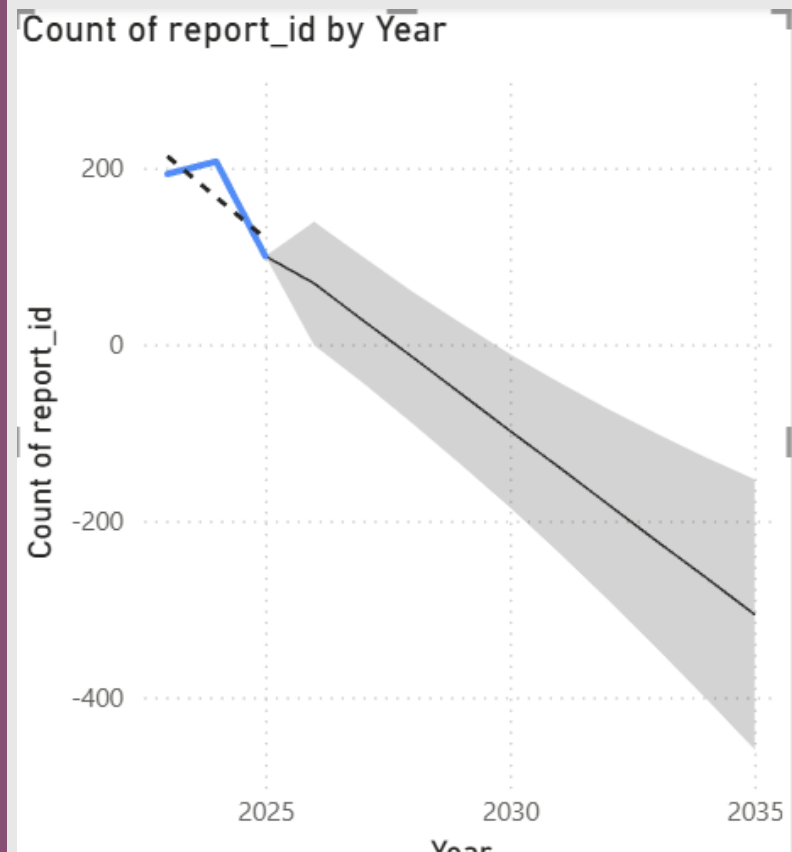


1. Add Decomposition Tree.
2. Root: `count of report_id`.
3. Drill by: `drug_class` → `hospital` → `age` → `seriousness`.
4. Talking point: quickly locate clusters of serious ADRs by hospital/drug class.

Forecasting (Trends)

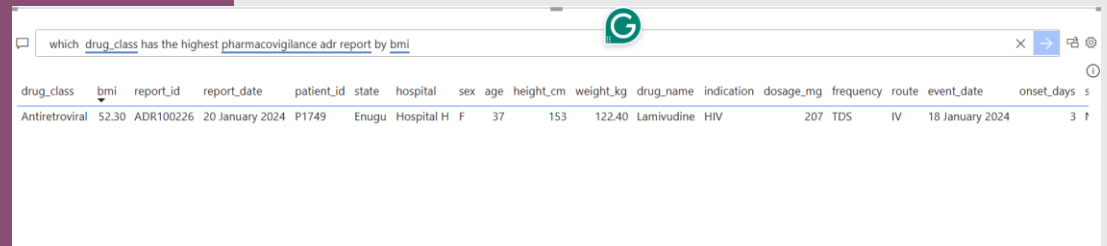
— Pharmacovigilance

1. Create a line chart: `report_date` (Axis) vs `count of report_id` (Values).
2. Turn on **Analytics** → **Forecast** (e.g., 6–12 steps).
3. Talking point: show **seasonal spikes** or **post-policy changes** in reports.



Q&A (Natural language) – Both datasets

- Enable Q&A and ask:
 - "Which drug class has the highest number of serious events in 2024?"
 - "Show average length_of_stay_days by diagnosis in 2025."

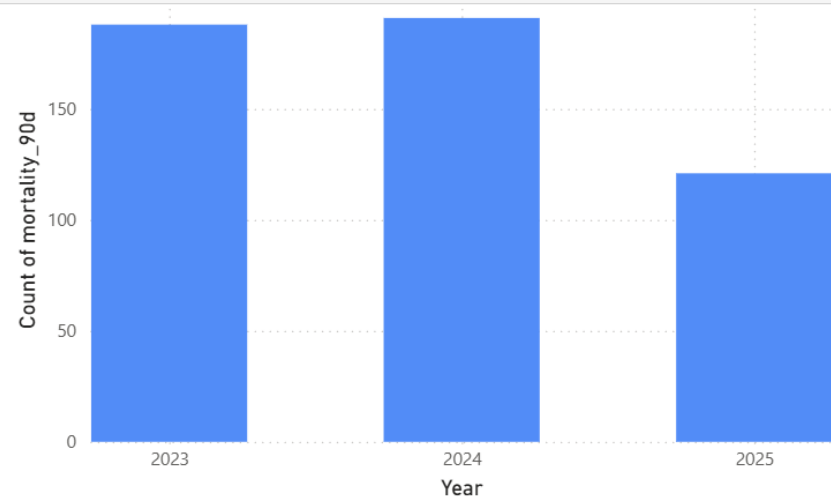


The screenshot shows a web application interface. At the top, there is a search bar with the text "which drug_class has the highest pharmacovigilance adr report by bmi". Below the search bar, there is a table with the following columns: drug_class, bmi, report_id, report_date, patient_id, state, hospital, sex, age, height_cm, weight_kg, drug_name, indication, dosage_mg, frequency, route, event_date, onset_days, and s. The table contains one row of data for the drug class "Antiretroviral".

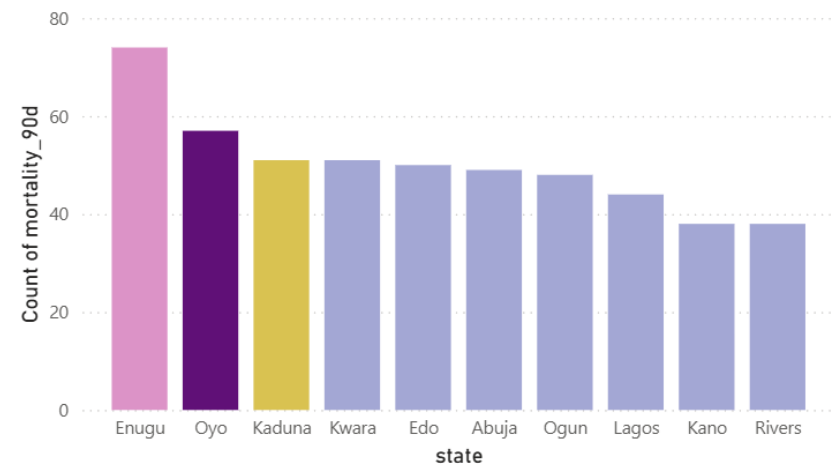
drug_class	bmi	report_id	report_date	patient_id	state	hospital	sex	age	height_cm	weight_kg	drug_name	indication	dosage_mg	frequency	route	event_date	onset_days	s
Antiretroviral	52.30	ADR100226	20 January 2024	P1749	Enugu	Hospital H	F	37	153	122.40	Lamivudine	HIV	207	TDS	IV	18 January 2024	3	f

Outcomes Dashboard – Medicine

1. Load *medicine_patient_outcomes.csv*.
2. Cards: average `length_of_stay_days`, `% readmission_30d`, `% mortality_90d`.
3. Bar chart: `diagnosis` vs `readmission_30d` rate.
4. Key Influencers: **Analyze** `readmission_30d` → **Explain by** `age`, `lab_abnormalities_count`, `creatinine_mg_dl`, `ALT_U_L`, `hemoglobin_g_dl`, `dosage_intensity`, `length_of_stay_days`, `diagnosis`.
5. Talking point: **lab abnormalities + age** drive readmissions → targets for care pathways.



Count of mortality_90d by state




Simple Cohort Filter (Slicers) – Both


- Add slicers for `state`, `hospital`, `sex`, `age` (numeric range) to simulate **site-level** pharmacovigilance or ward-level audit.
-

Training Demo Flow with Power BI


Dataset Import – Load a CSV of patient vitals + drug history + adverse reactions.




Data Cleaning – Show missing value handling and transformations (link to data quality/bias lecture).



Visualization – Build basic charts (adverse events by drug, hospital, age).



AI Insights – Apply *Key Influencers* + *Q&A* + *Forecasting*.



Interpretation – Discuss how clinicians can use these insights for pharmacovigilance decisions.

Clinical Interpretation

Bullet point example: “High dosage_mg + age $\geq 65 \rightarrow$ increased ‘Serious’ ADR. Consider dose review and monitoring.”

Ethical Issues in AI for Pharmacovigilance & Healthcare



Informed Consent

Patients must be fully aware of how their medical data is used, especially in AI research.



Bias and Fairness

AI systems should avoid favoring certain groups or others, ensuring equitable outcomes.



Interpretability and Transparency

AI predictions must be understandable and justifiable to clinicians and regulators.



NDPR & WHO Alignment

AI use must align with Nigeria Data Protection Regulation (NDPR) and WHO guidelines

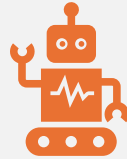
Ethics & Privacy

“De-identify, aggregate; follow NDPR/HIPAA; restrict direct identifiers.”

Data Quality & Bias

Check missing labs
class imbalance (few fatal
cases)
site reporting bias (some
hospitals under-report)

Closing & Takeaways



AI is a *tool*, not a replacement for medical expertise.



Patient data must be used responsibly.



Evidence-based practice + AI = stronger pharmacovigilance