

Data Drift Detection

Data Drift: When production data distribution **differs from training data**



Covariate Shift

$P(X)$ changes

Input distribution changes

| *Example: New imaging equipment with different resolution/contrast*



Label Shift

$P(Y)$ changes

Output distribution changes

| *Example: Seasonal disease prevalence changes*



Concept Drift

$P(Y|X)$ changes

Relationship between input and output changes

| *Example: Treatment protocols evolve, outcomes change*



Feature Drift

Individual features change

Specific feature distributions shift

| *Example: Patient demographics shift over time*



Detection Methods

KS Test (Kolmogorov-Smirnov)

Chi-Square Test

PSI (Population Stability Index)

Wasserstein Distance

KL Divergence