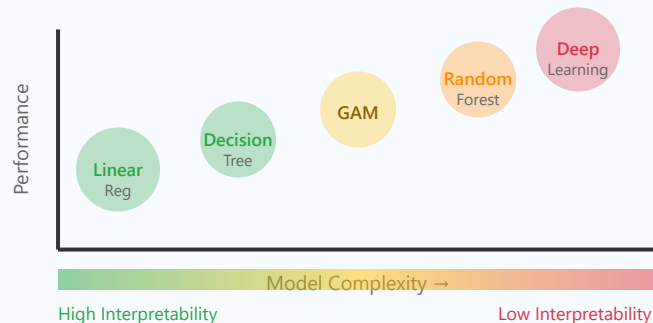# Interpretable ML for Clinical Adoption

⚠️ **Black Box Problem:** Clinicians won't trust models they can't understand

Glass Box Models
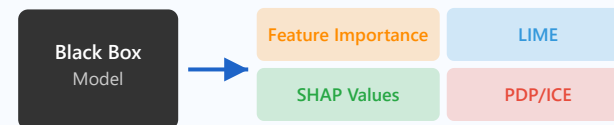
Interpretability-Performance Trade-off

Performance

Deep Learning
Random Forest
GAM
Decision Tree
Linear Reg

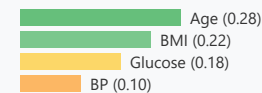Model Complexity →

High Interpretability · Low Interpretability

- Linear/Logistic Regression
- Decision Trees
- GAMs (Generalized Additive Models)
- Rule-based systems

Inherently interpretable

Post-hoc Explanations

Making Black Boxes Transparent

Black Box Model → Feature Importance | LIME | SHAP Values | PDP/ICE

Example: Feature Importance

Age (0.28)
BMI (0.22)
Glucose (0.18)
BP (0.10)

- Feature Importance (RF, XGBoost)
- Partial Dependence Plots
- LIME (Local Interpretable)
- SHAP Values

Explain black boxes

Clinical Acceptance Requires: Model explanation + Clinical validation + Physician trust + Regulatory approval