

*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

I. Evolution from Universal to Precision Screening

CASE-PASS Demonstration: Adjusting for Self-Selection and Lead-Time Bias in Screening Evaluation

Dr. Chiu-Wen Su

National Taiwan University



IACCS 2026

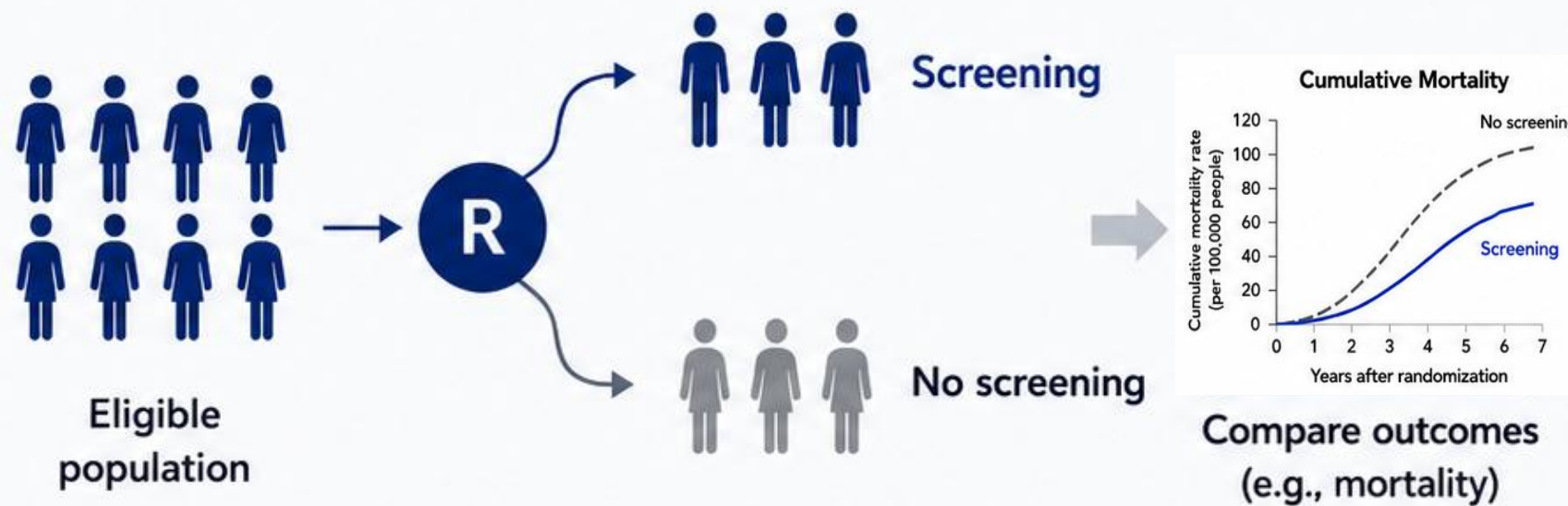
Real-world Screening Evaluation



Traditional Evaluation of Screening Effectiveness

Randomized Controlled Trials (RCTs) have been the gold standard for evaluating screening.

Randomized Controlled Trials (RCTs)



Examples of Screening Evaluated with RCTs



Mammography
(breast cancer)



FOBT / FIT
(colorectal cancer)



Pap smear
(cervical cancer)

Strengths



- ✓ Randomization minimizes selection bias
- ✓ Gold standard evidence for mortality reduction

Limitations



- ⚠ Difficult after nationwide implementation
- ⚠ Not suitable for ongoing program evaluation

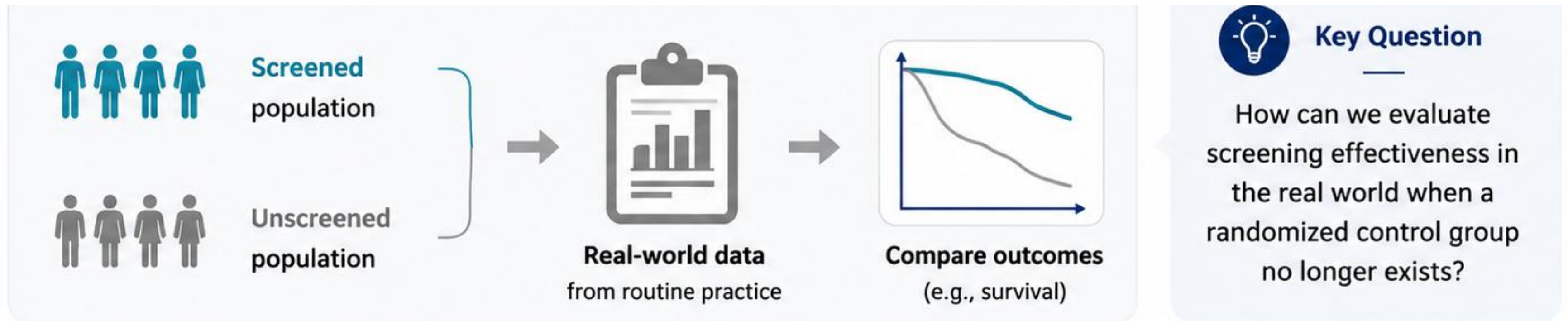


RCTs establish efficacy under controlled conditions.

Question: How do we evaluate screening after implementation?

But What Happens After Implementation?

Once screening programs are implemented nationwide, we rely on observational data from routine practice.



Limitations: Traditional Approaches Are Subject to Major Biases



Lead-time Bias

Earlier diagnosis looks better, even if time of death is unchanged.



Length Bias

Slow-growing tumors are more likely to be detected by screening.



Self-selection Bias

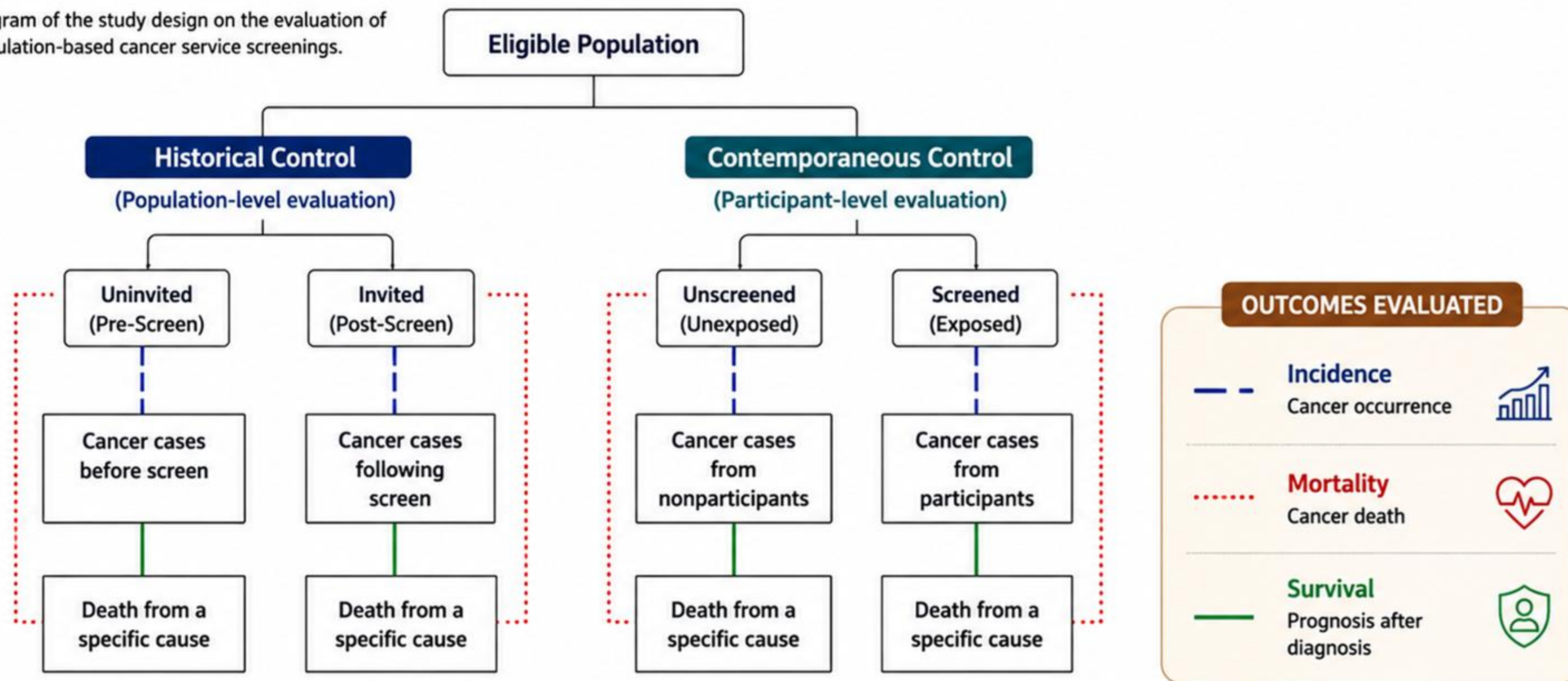
People who participate in screening are healthier and have different risk profiles.

CASE-PASS Study Design Architecture

Two complementary approaches for evaluating real-world screening effectiveness

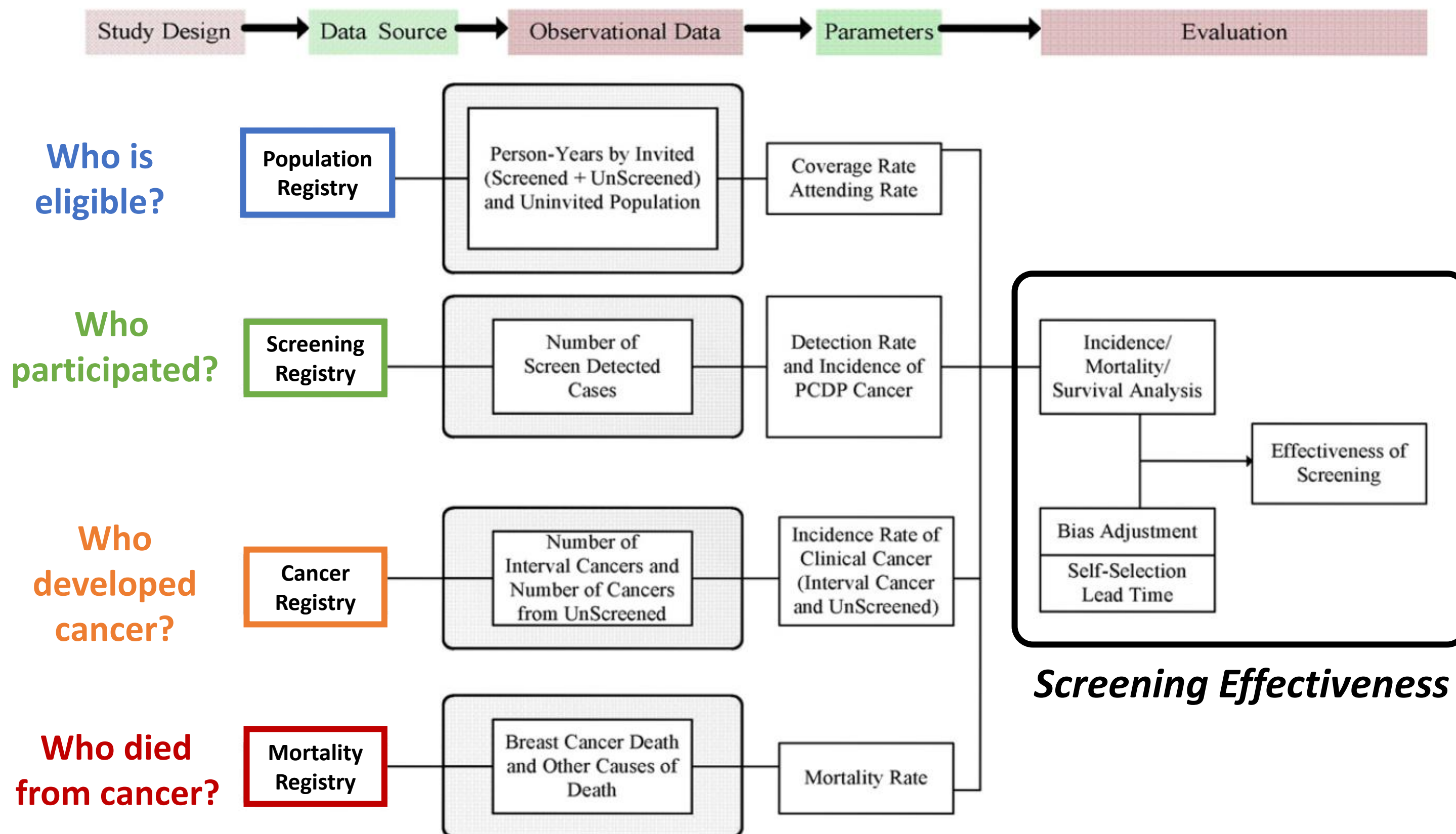


Diagram of the study design on the evaluation of population-based cancer service screenings.



Chen LS, Yen AM, Duffy SW, Tabar L, Lin WC, Chen HH. Ann Epidemiol. 2010 Oct;20(10):786-96.

Data Required for the Evaluation of Cancer Service Screenings



Chen LS, Yen AM, Duffy SW, Tabar L, Lin WC, Chen HH. *Ann Epidemiol.* 2010 Oct;20(10):786-96.

CASE-PASS Analytical Platform – How Selection Bias is Adjusted

Historical Control Design

Calculate the 20-year cumulative breast cancer mortality among women aged 40–69 years during the prescreening period (1960–1981) and the screening period (1982–2003)

Step 1. Construct Exposed and Unexposed Person-Years

1 Parameter Setting

Data Set

Essp.Casepass

Attributes

DDIAG
DETECT
DIAAGE
DOFDEATH
EXPOSE
MDIAG
MOFDEATH
NECD
SD
YDIAG
YOFDEATH
stime

Location

X City

Prescreen period

1960 ~ 1981

Screen period

1982 ~ 2003

Age range

40 ~ 69

Save Model

Load Model

Clear

View Log

View Output

Exit

Prescreen period

Screen period

Year	Population in year	Person-years exposed to screening	Total person-years exposed for each period	Total person-years unexposed for each period
1960	5,116	0	4,737 (First period)	108,574 (First period)
1961	5,116	0		
1962	5,116	0		
⋮	⋮	⋮		
1978	5,105	0		
1979	5,069	469		
1980	5,040	1,613		
1981	5,009	2,655		
1982	4,987	2,792	88,672 (Second period)	23,365 (Second period)
⋮	⋮	⋮		
2002	5,225	4,290		
2003	5,245	4,419		

CASE-PASS Analytical Platform – How Selection Bias is Adjusted

Historical Control Design

Calculate the 20-year cumulative breast cancer mortality among women aged 40–69 years during the prescreening period (1960–1981) and the screening period (1982–2003)

3.1 Mortality

Data Set: Essp.Casepass

Attributes:

- DDIAG
- DETECT
- DIAAGE
- DOFDEATH
- EXPOSE**
- MDIAG
- MOFDEATH
- NBCD
- SD
- YDIAG
- YOFDEATH
- stme

Load Pre-definal Model: []

Year of diagnosis: X > YDIAG

Year of death: X > YOFDEATH

Age of diagnosis: X > DIAAGE

Die of breast cancer: X > NBCD

Exposure of screening: X > EXPOSE

Save Model: []

Load Model: []

Run Poisson Rrg Excel Clear View Log View Output Exit

Step 2. Estimate Mortality in Each Period

VIEWTABLE: Essp.Casepass

	YDIAG	YOFDEATH	DIAAGE	NBCD	EXPOSE
1	1961	1984	068	0	0
2	1962	1982	068	0	0
3	1963	1968	069	1	0
4	1963	1968	068	1	0
5	1964	1968	067	1	0
6	1966	1974	069	1	0
7	1963	1964	066	1	0
8	1962	1964	065	1	0
9	1964	1968	066	0	0
10	1964	1965	065	1	0
11	1964	1975	065	1	0
12	1965	1967	066	1	0
13	1961	1968	062	1	0
14	1961	1961	061	1	0
15	1961	1967	061	1	0
16	1968	1983	067	1	0
17	1965	1965	064	1	0

CASE-PASS Analytical Platform – How Selection Bias is Adjusted

Historical Control Design

Calculate the 20-year cumulative breast cancer mortality among women aged 40–69 years during the prescreening period (1960–1981) and the screening period (1982–2003)

3.1 Mortality

Data Set: Esp.Casepass

Attributes: DDIAG, DETECT, DIAAGE, DOFDEATH, EXPOSE, MDIAG, MOFDEATH, NBED, SD, YDIAG, YOFDEATH, stime

Load Pre-definal Model: []

Year of diagnosis: X > YDIAG

Year of death: X > YOFDEATH

Age of diagnosis: X > DIAAGE

Die of breast cancer: X > NBED

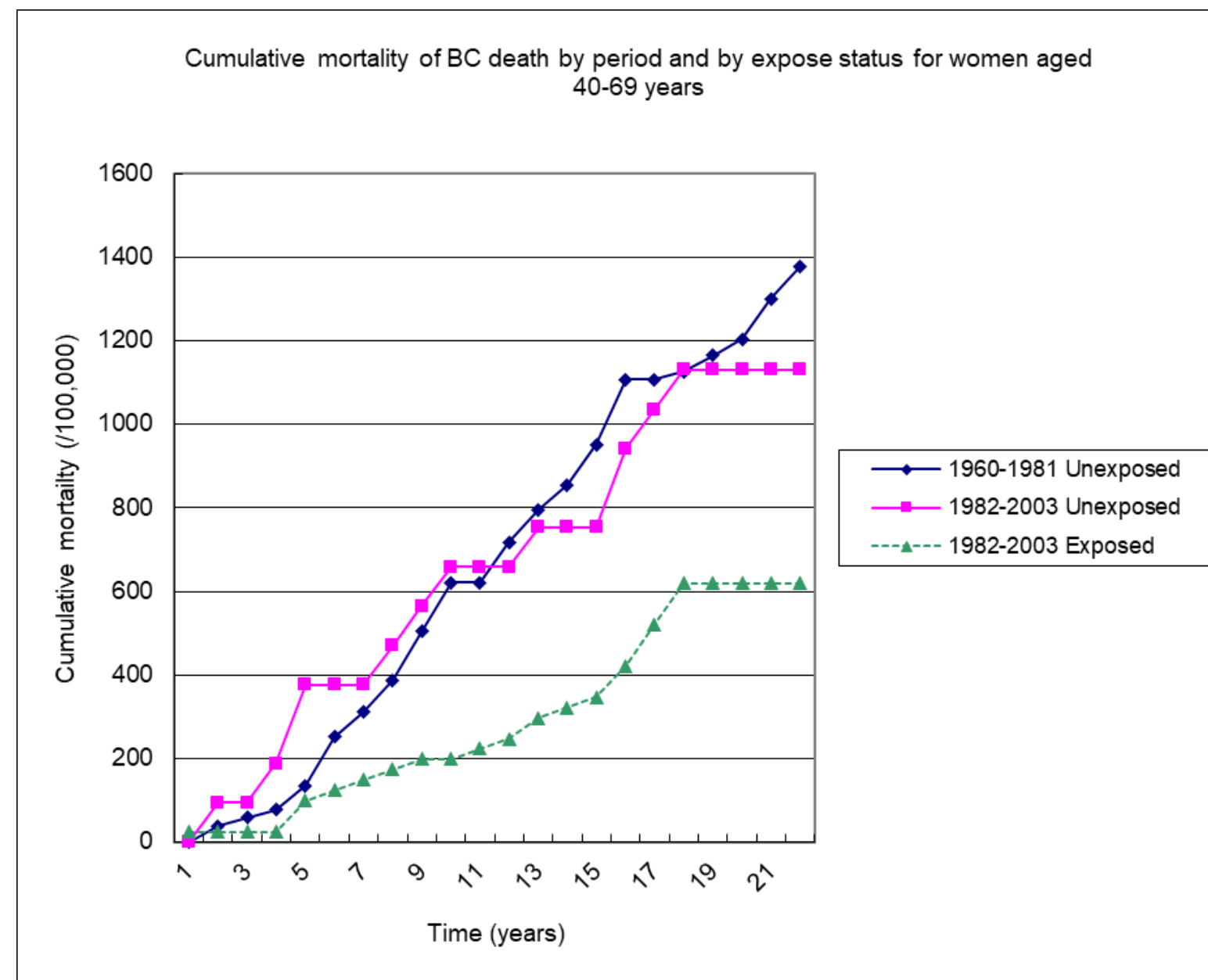
Exposure of screening: X > EXPOSE

Save Model: []

Load Model: []

Run Poisson Rrg Excel Clear View Log View Output Exit

Step 2. Estimate Mortality in Each Period



CASE-PASS Analytical Platform – How Selection Bias is Adjusted

Historical Control Design

Calculate the 20-year cumulative breast cancer mortality among women aged 40–69 years during the prescreening period (1960–1981) and the screening period (1982–2003)

1 RR_1 : Screened vs. Uninvited (Baseline)

$$RR_1 = \text{Mortality}_{\text{Screened}} / \text{Mortality}_{\text{Uninvited}}$$

2 RR_2 : Nonparticipants vs. Uninvited (Baseline)

$$RR_2 = \text{Mortality}_{\text{Nonparticipants}} / \text{Mortality}_{\text{Uninvited}}$$

3 AR : Attendance Rate

$$AR = \text{Number Attending} / \text{Number Invited}$$

CASE-PASS Adjusted Relative Risk

$$RR_{\text{adj}} = RR_1 \times AR + RR_2 \times (1 - AR)$$

This adjusts for self-selection bias.

C. Selection Bias Adjustment Output

3.3 Bias Adjustment

Selection Bias and Lead-time Bias Adjustment

Crude RR from Table 6

1 RR_1 : $\frac{P(\text{Death} | S)}{P(\text{Death} | \bar{I})} = 0.40338$ SE: 0.14532 (95% CI: 0.30339 0.53631)

From meta-analysis of randomized trials

2 RR_2 : $\frac{P(\text{Death} | \bar{S})}{P(\text{Death} | \bar{I})} = 1.49$ SE: 0.2

From meta-analysis of randomized trials

Attendance rate = 0.87741

Corrected for lead time = Yes (1.05357)

Results

Intention to Treat

Adjusted RR = 0.55554 (95% CI: 0.44468 0.69403)

Noncompliance Adjustment in Service Screening

Adjusted RR = 0.45622 (95% CI: 0.26117 0.79694)

Run Clear View Log View Output Exit

Lead-time Bias in Cancer Screening

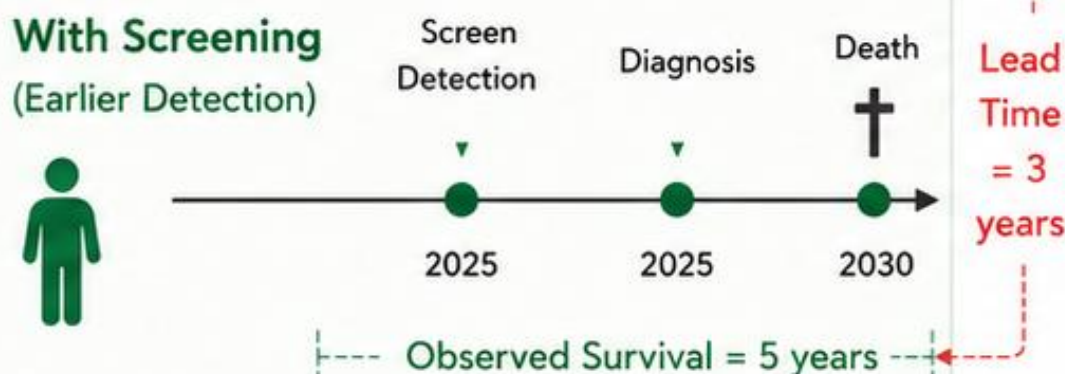
Screening detects cancer earlier without changing the time of death, creating an apparent survival advantage.

What is Lead-time Bias?

Without Screening (Clinical Diagnosis)



With Screening (Earlier Detection)



Survival appears longer with screening, but the time of death is unchanged. The extra survival time is the lead time.

How Lead-time Bias is Adjusted (CASE-PASS)

Observed Survival (From Screening Detection)

$$\text{True Survival Benefit (what screening actually provides)} + \text{Lead Time (earlier diagnosis)} = \text{Observed Survival Time}$$



Adjusted Survival (Removes Lead Time)

$$\text{Observed Survival Time} - \text{Lead Time (earlier diagnosis)} = \text{Adjusted Survival Time (true benefit)}$$

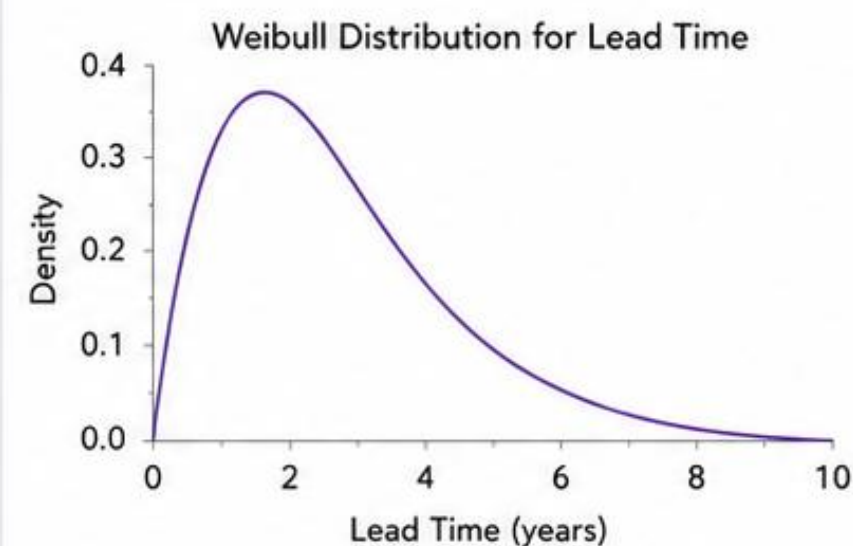
Illustrative Example

$$\text{Observed Survival (with screening)} = 5 \text{ years} - \text{Lead Time } 3 \text{ years} = \text{Adjusted Survival (true benefit)} = 2 \text{ years}$$

How CASE-PASS Estimates Lead Time

Assume Lead Time Follows a Weibull Distribution

$$T_{lead} \sim \text{Weibull}(\lambda, k)$$



Lead time is estimated from observed data using a Weibull distribution.



The expected lead time is subtracted from the observed survival time to obtain lead-time-adjusted survival.

CASE-PASS Analytical Platform – Lead-time Bias in Cancer Screening

3.2 Survival (Lead Time Adjustment)

Load Pre-definal Model

Data Set: Essp.Casepass

Attributes: DDIAG, DETECT, DIAAGE, DOFDEATH, EXPOSE, MDIAG, MOFDEATH, NBCD, SD, YDIAG, YOFDEATH, stime

Death: NBCD

Survival Time: stime

Detection Mode: SD

Mean Lead Time: .

Time Interval of Life Table: 1

Max Follow-up Time: .

Distribution: Weibull

Competing Risk: No

Save Model

Load Model

(1) Adjusted Survival

(2) Survival Curves

(3) Hazard Ratios

Clear View Log View Output Exit

Calculate breast cancer survival time among women aged 40–69 years by screening detected and clinical detected.

VIEWTABLE: Essp.Casepass

	YDIAG	YOFDEATH	DIAAGE	DETECT	NBCD	EXPOSE	stime	SD
1	1961	1984	068	15	0	0	23.05270363	0
2	1962	1982	068	15	0	0	20.85147159	0
3	1963	1968	069	15	1	0	5.037645448	0
4	1963	1968	068	15	1	0	5.593429158	0
5	1964	1968	067	15	1	0	4.052854209	0
6	1966	1974	069	15	1	0	8.788501027	0
7	1963	1964	066	15	1	0	1.152635181	0
8	1962	1964	065	15	1	0	2.833675565	0
9	1964	1968	066	15	0	0	23.84941821	0
10	1964	1965	065	15	1	0	0.980150582	0
11	1964	1975	065	15	1	0	11.06913073	0
12	1965	1967	066	15	1	0	2.496919918	0
13	1961	1968	062	15	1	0	7.529089665	0
14	1961	1961	061	15	1	0	0.501026694	0
15	1961	1967	061	15	1	0	5.976728268	0
16	1968	1983	067	15	1	0	14.88295688	0
17	1965	1965	064	15	1	0	0.186173854	0

CASE-PASS Analytical Platform – Lead-time Bias in Cancer Screening

3.2 Survival (Lead Time Adjustment)

Load Pre-definal Model

Data Set

Attributes

- DDIAG
- DETECT
- DIAAGE
- DOFDEATH
- EXPOSE
- MDIAG
- MOFDEATH
- NBCD
- SD**
- YDIAG
- YOFDEATH
- stime

Death

Survival Time

Detection Mode

Mean Lead Time

Time Interval of Life Table

Max Follow-up Time

Distribution

Competing Risk

Save Model

Load Model

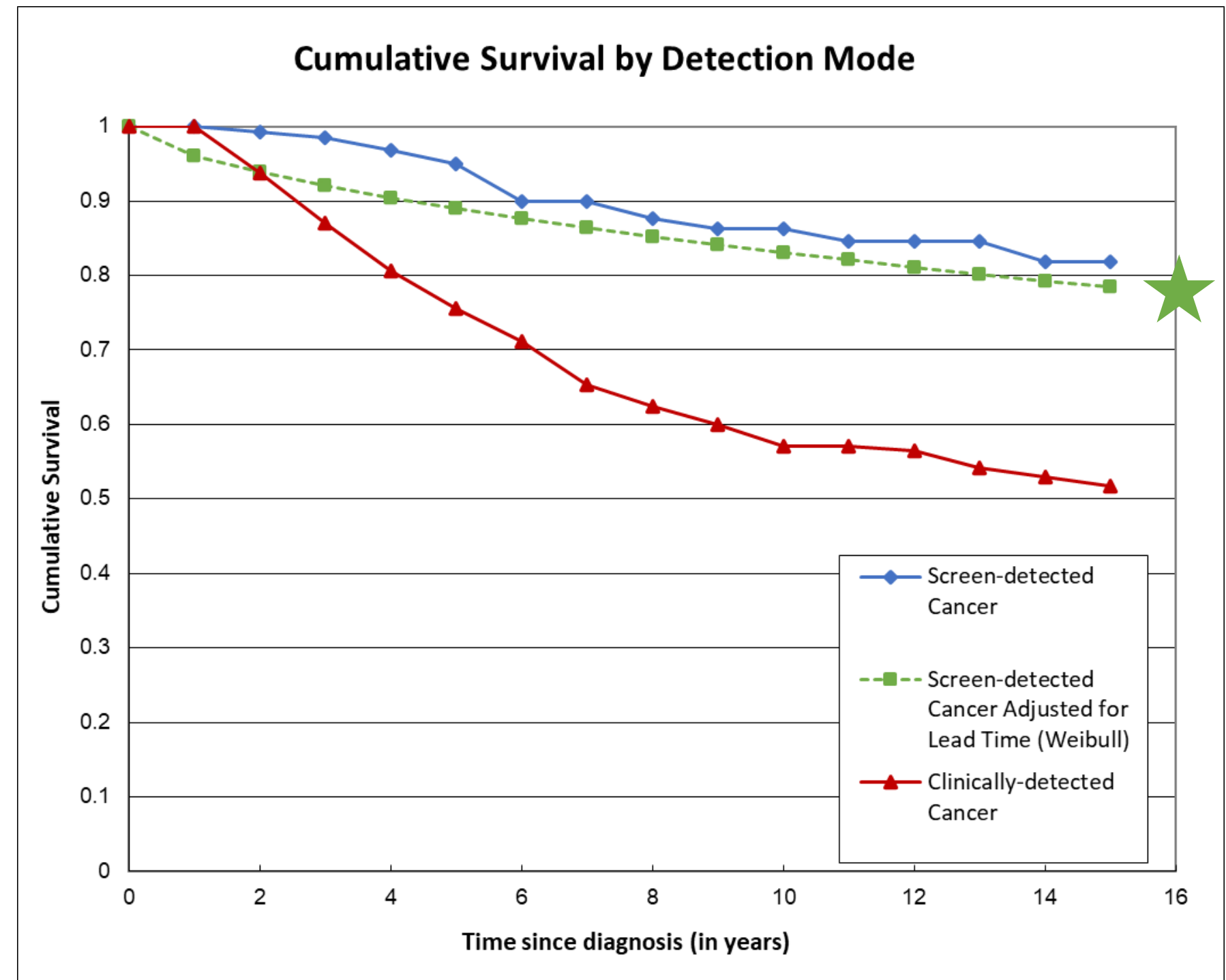
(1) Adjusted Survival

(2) Survival Curves

(3) Hazard Ratios

Clear View Log View Output Exit

Calculate breast cancer survival time among women aged 40–69 years by screening detected and clinical detected.



Chen LS, Yen AM, Duffy SW, Tabar L, Lin WC, Chen HH. Ann Epidemiol. 2010 Oct;20(10):786-96.

Thank you for your attention



IACCS 2026

*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

II. Artificial Intelligence and Machine Learning for Precision Screening

Machine Learning Classifiers for Risk Prediction: *from Conventional Algorithms to Deep Learning*

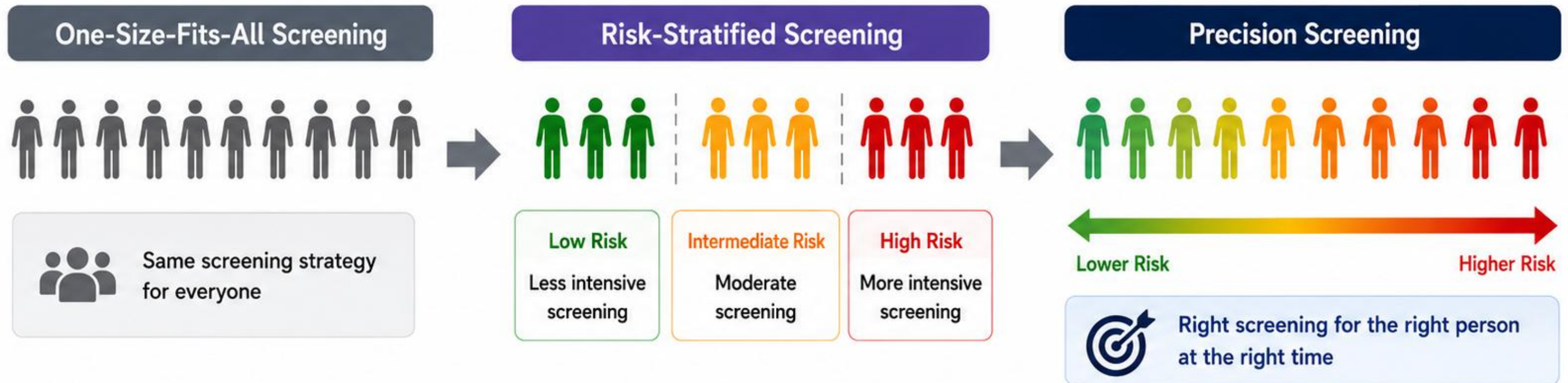
Dr. Chiu-Wen Su

National Taiwan University



IACCS 2026

From One-Size-Fits-All to Precision Screening



How do we identify
the right person
for the right screening
at the right time?



Right Person
Who is at higher risk?



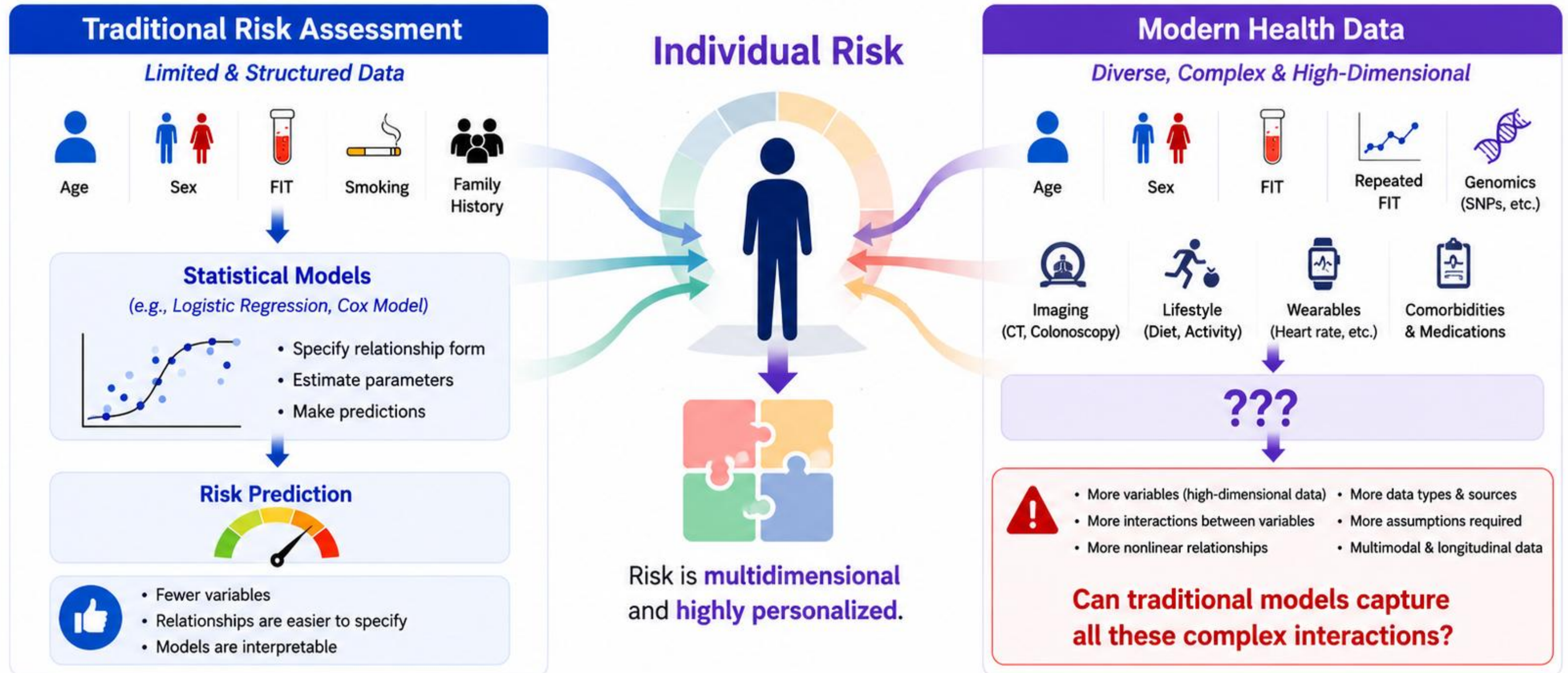
Right Screening
What type and intensity
of screening?



Right Time
When should
screening be performed?

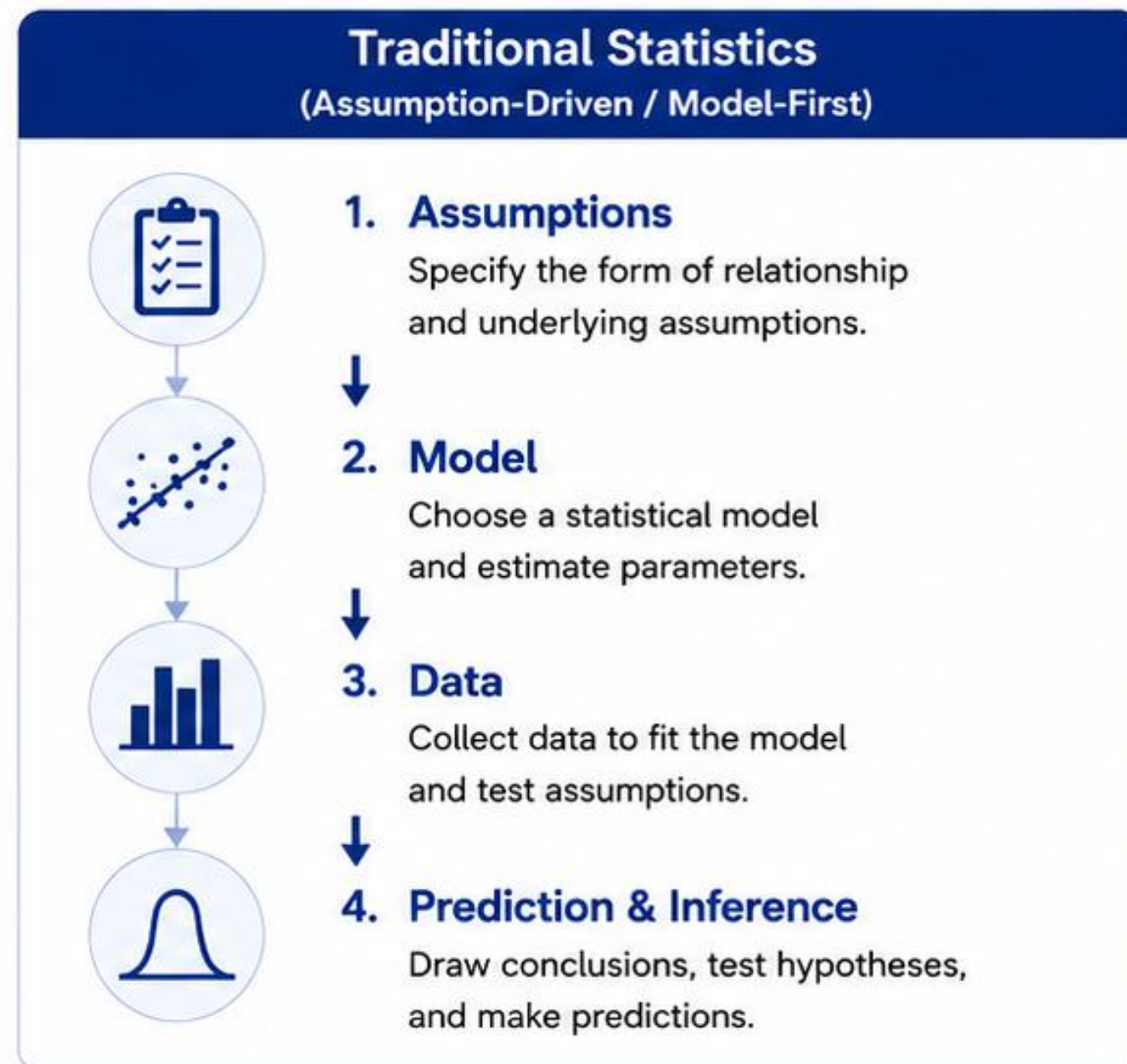
Why Is Individual Risk So Difficult to Predict?

Individual risk depends on many interacting factors.

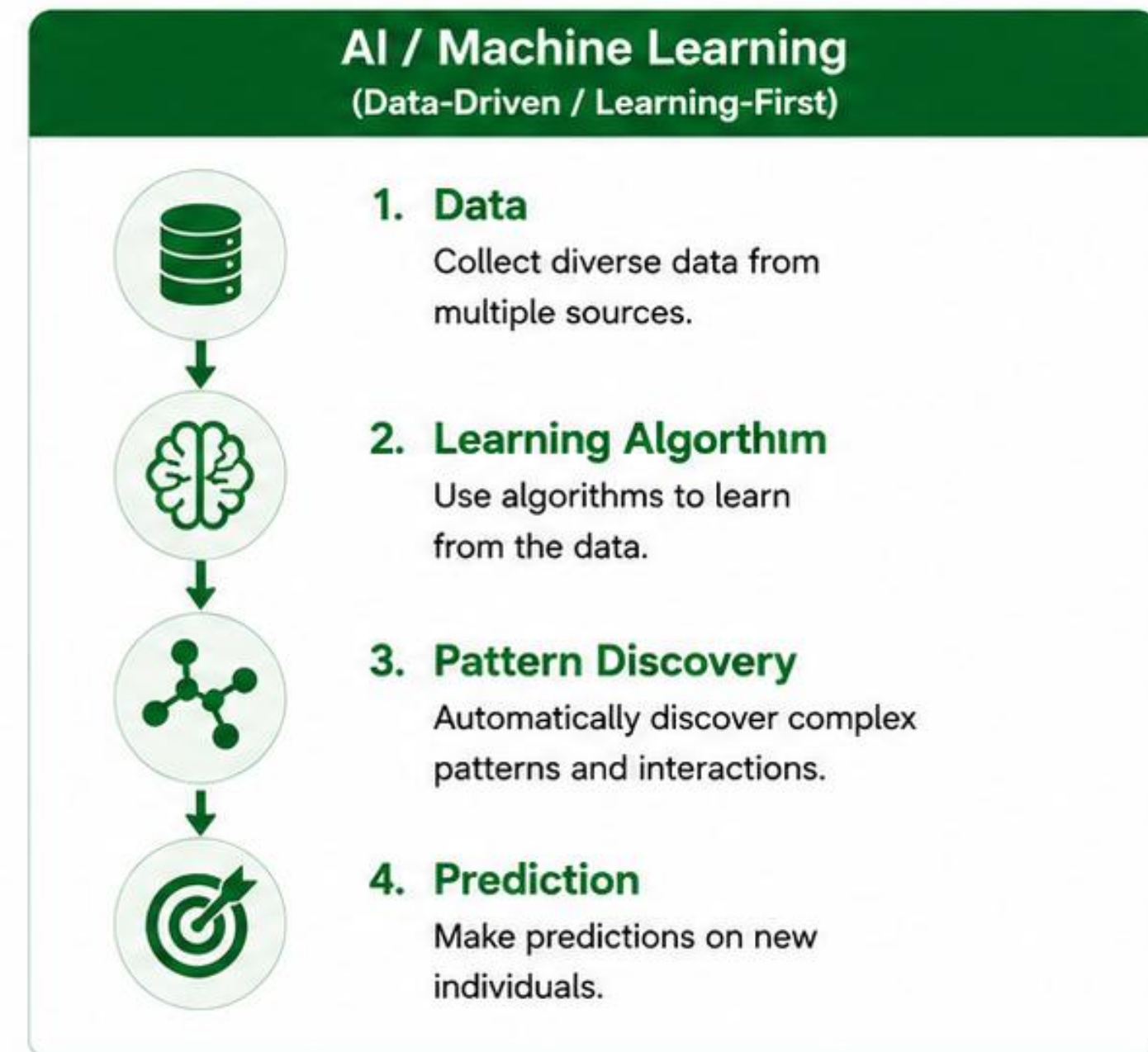


How Can We Learn from Complex Data?

Traditional statistics and AI take different approaches.



VS.



What *Assumptions* should we make?



What patterns can we learn from the *data*?



Three Major Types of Machine Learning

Different learning paradigms for extracting knowledge from data



Supervised Learning

Learning from labeled outcomes

Input Data → Known Outcome

Typical Tasks



Classification

- Cancer vs. No Cancer
- FIT Positive vs. Negative



Regression

- Individual Risk Prediction
- Probability Estimation

Common Algorithms

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM)
- Neural Network



Unsupervised Learning

Learning from unlabeled data

Typical Tasks



Clustering

- Patient Stratification
- Risk Group Identification
- Phenotype Discovery

Common Algorithms

- K-Means
- Hierarchical Clustering
- Principal Component Analysis (PCA)



Reinforcement Learning

Learning through interaction and feedback

Typical Tasks



Sequential Decision Making

- Adaptive Screening Strategies
- Personalized Surveillance
- Dynamic Clinical Decision Support

Learning Cycle



Action



Environment



Reward



Update Policy



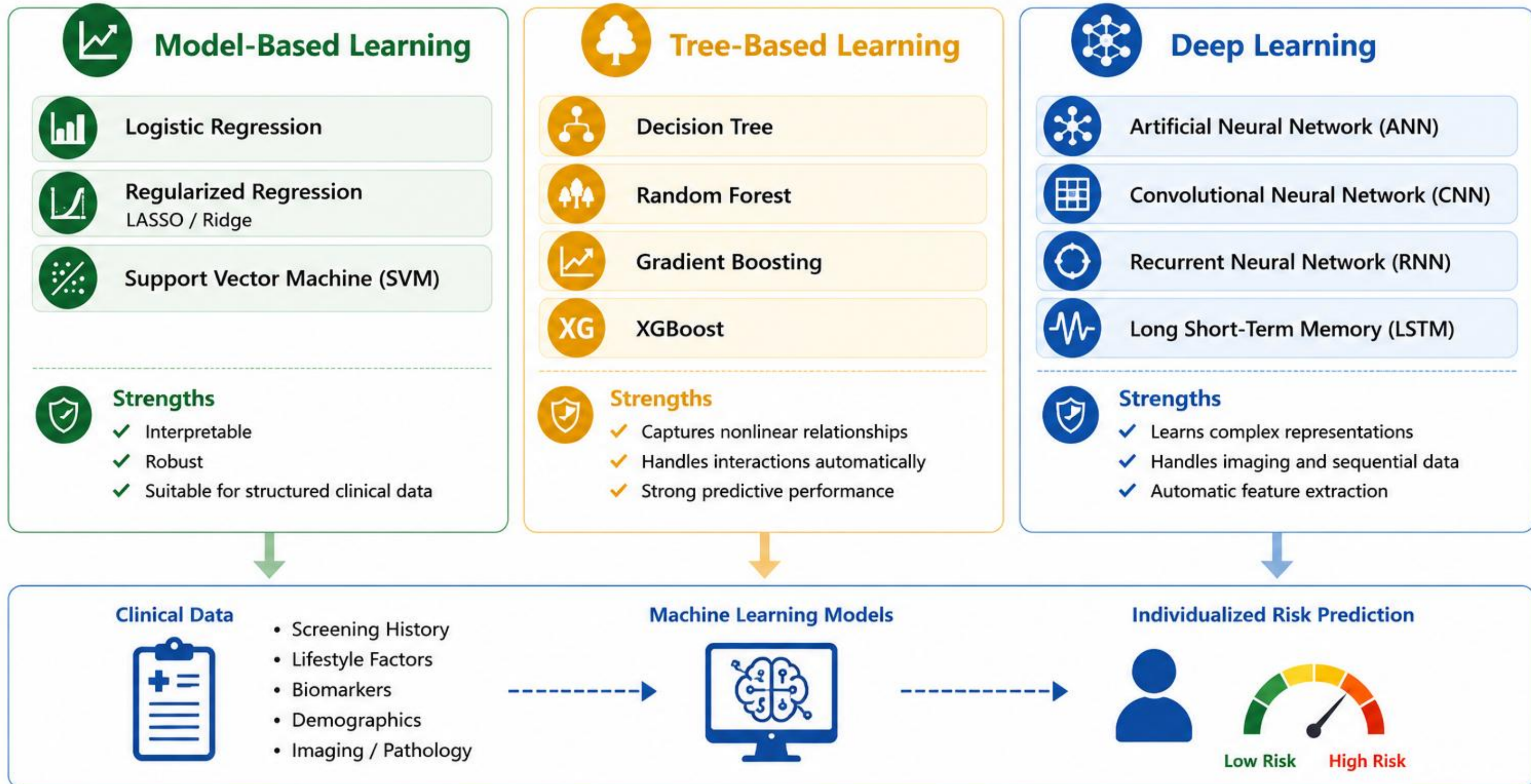
Supervised Learning

Learning from labeled outcomes

Input Data → **Known Outcome**

Machine Learning Models for Precision Screening

Commonly used algorithms for cancer risk prediction



Decision Tree: Learning Through Sequential Questions

The idea is simple: ask the right questions to split the data into more homogeneous groups.

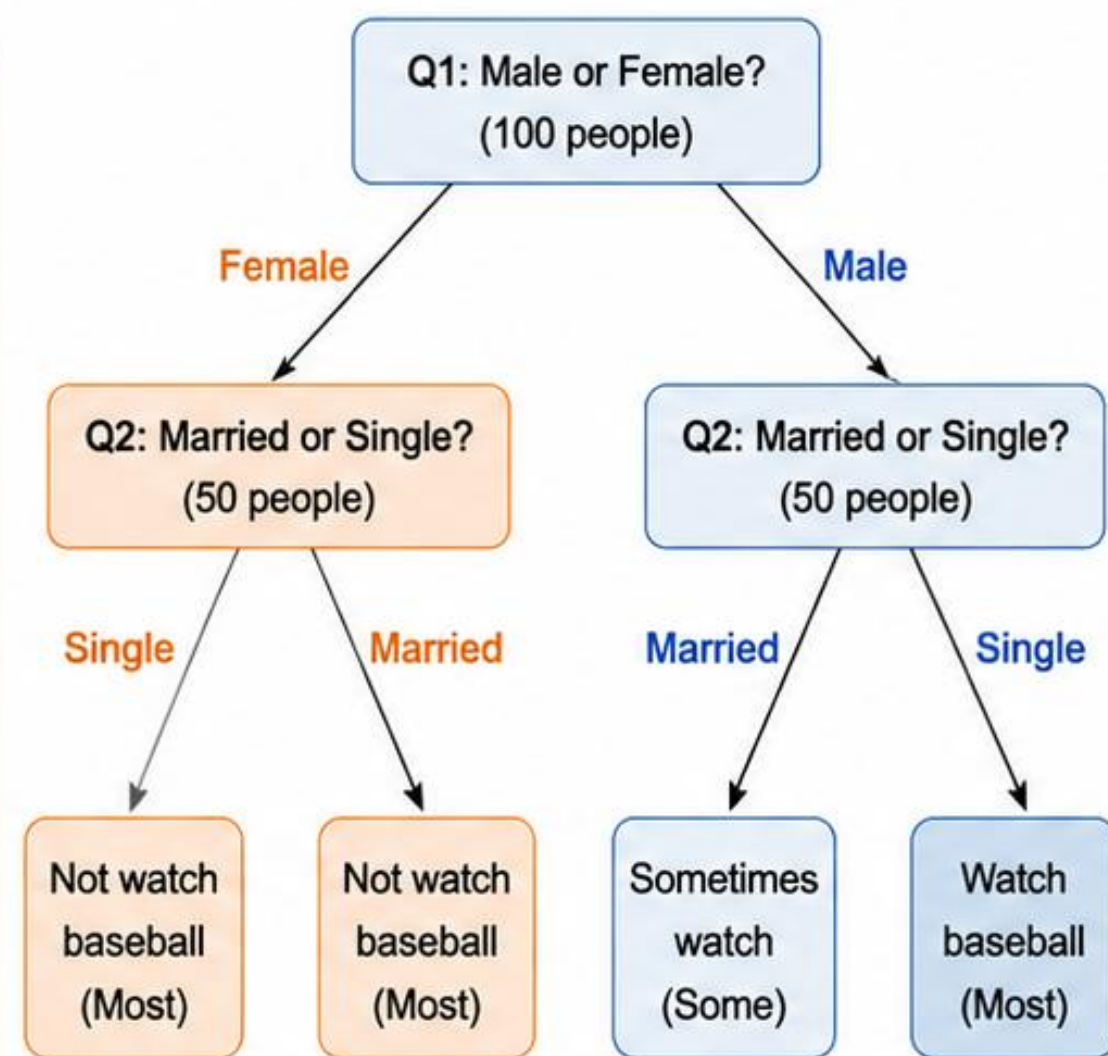
Example 1: Who is most likely to watch baseball?



We survey 100 people

Target: Watch baseball? (Yes / No)

- ✓ Male / Female
- ✓ Married / Single



★ Conclusion: **Single males** are most likely to watch baseball.

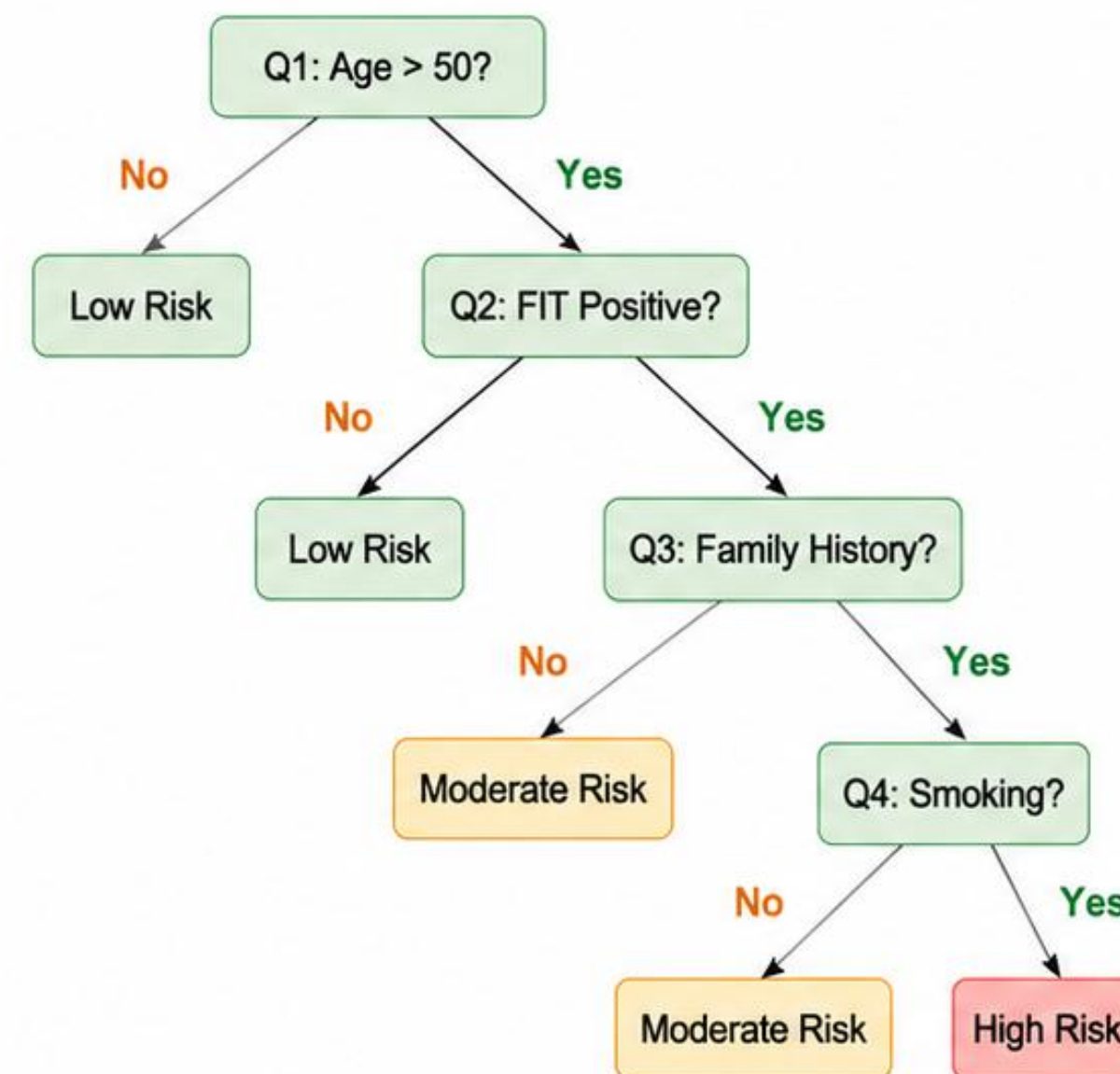
Example 2: Who is most likely to develop colorectal cancer?



In precision screening,
we want to find:

**Who is at high risk
for colorectal cancer?**

- ✓ Age
- ✓ FIT Result
- ✓ Family History
- ✓ Smoking
- ✓ ...



★ The model classifies individuals into **Low** / **Moderate** / **High Risk** groups.

How Does a Decision Tree Choose the Best Split?

Selecting the question that best separates the data



Example: Should we first ask “Male or Female?” or “Married or Single?”

→ The decision tree selects the question that **makes the data purer** (more homogeneous).

What is Impurity?

Impurity measures how **mixed** the data are.

Mixed data (uncertain)



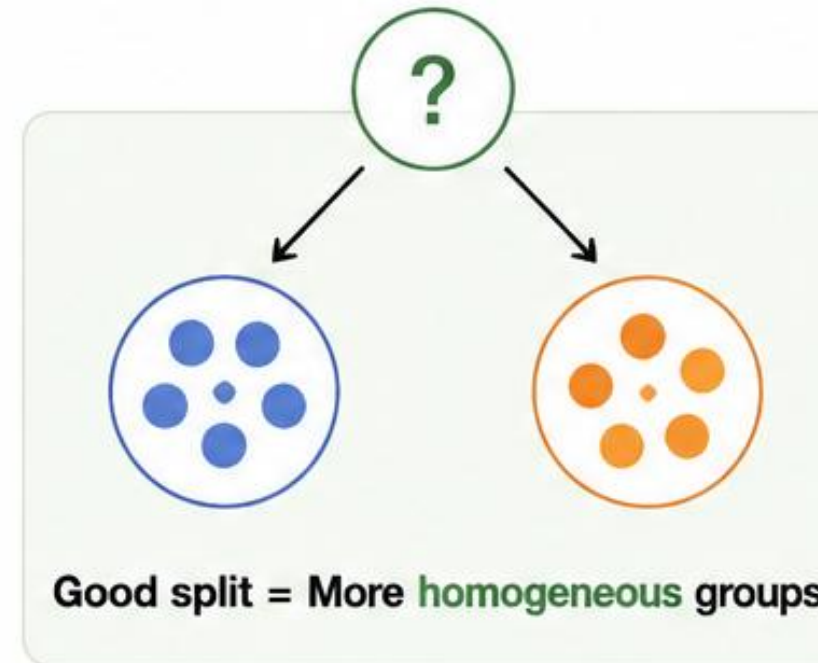
→ High impurity

Well-separated data (more certain)



→ Low impurity

The Principle



Decision Tree selects the split that produces the **largest reduction in impurity**.

How is Impurity Measured?

Decision Trees use mathematical criteria to measure impurity.



Entropy

Measures uncertainty

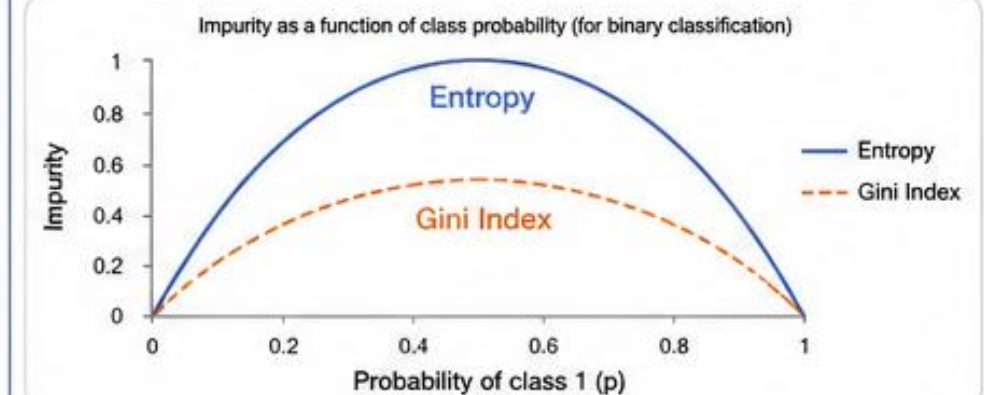
$$Entropy = - \sum_{i=1}^n p_i \log_2 p_i$$



Gini Index

Measures impurity

$$Gini = 1 - \sum_{i=1}^n p_i^2$$

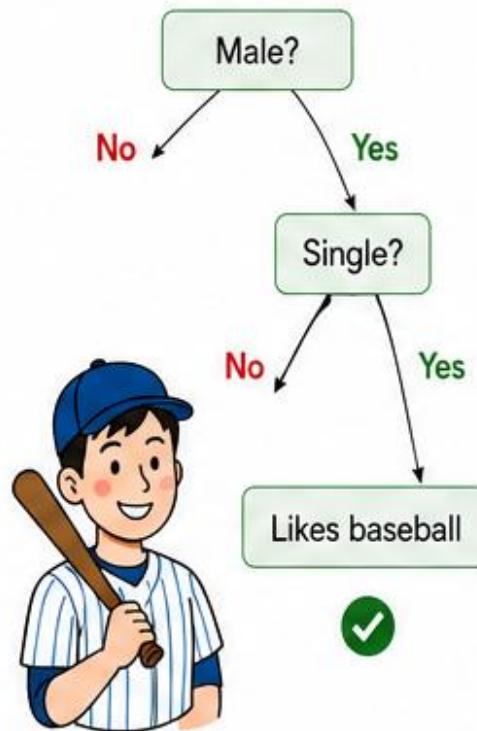


Both metrics are 0 when data are perfectly pure, and highest when data are maximally mixed.

Overfitting

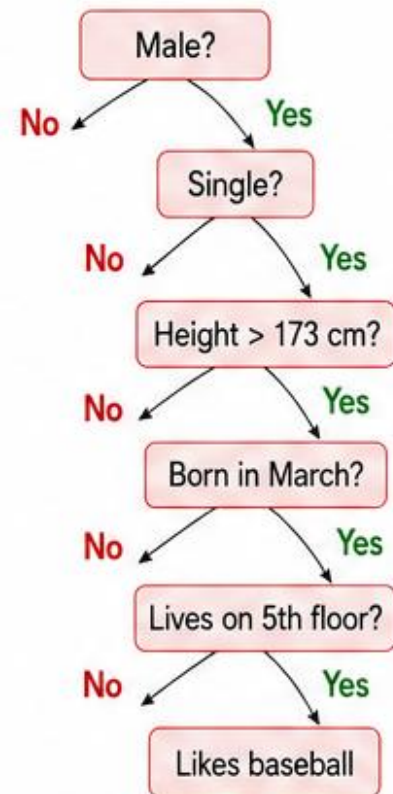
Analogy: Who likes baseball?

Good Fit (Generalizes Well)



- ✓ Simple rules capture the true pattern. Works on new people.

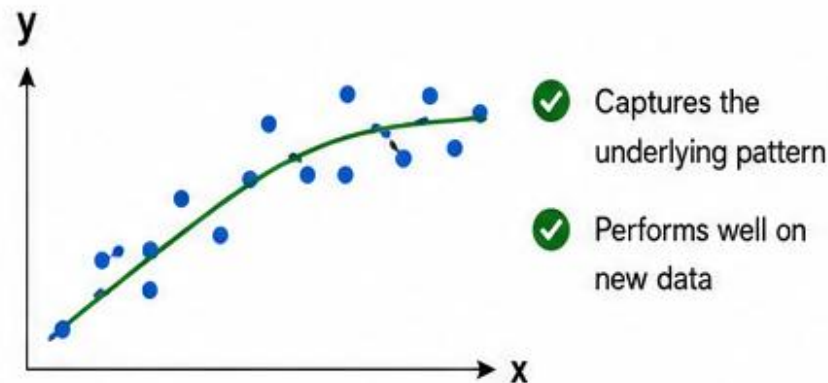
Overfit (Does Not Generalize)



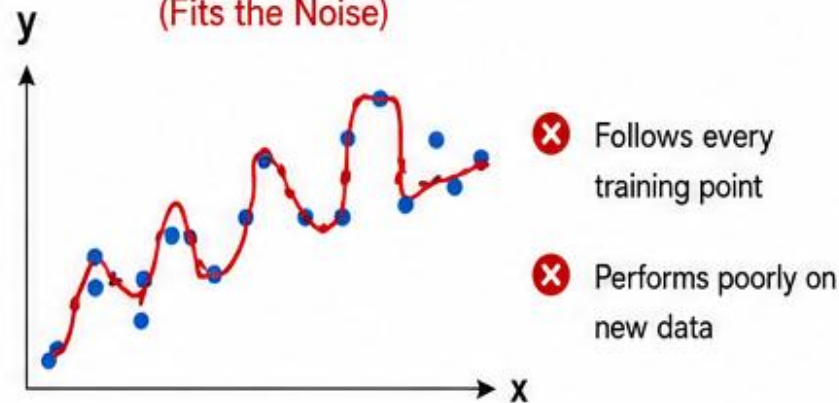
- ✗ Too many rules fit the noise. Fails on new people.

Good Fit vs. Overfit

Good Fit (Captures the Trend)



Overfit (Fits the Noise)



Example in Cancer Risk Prediction

Good Model



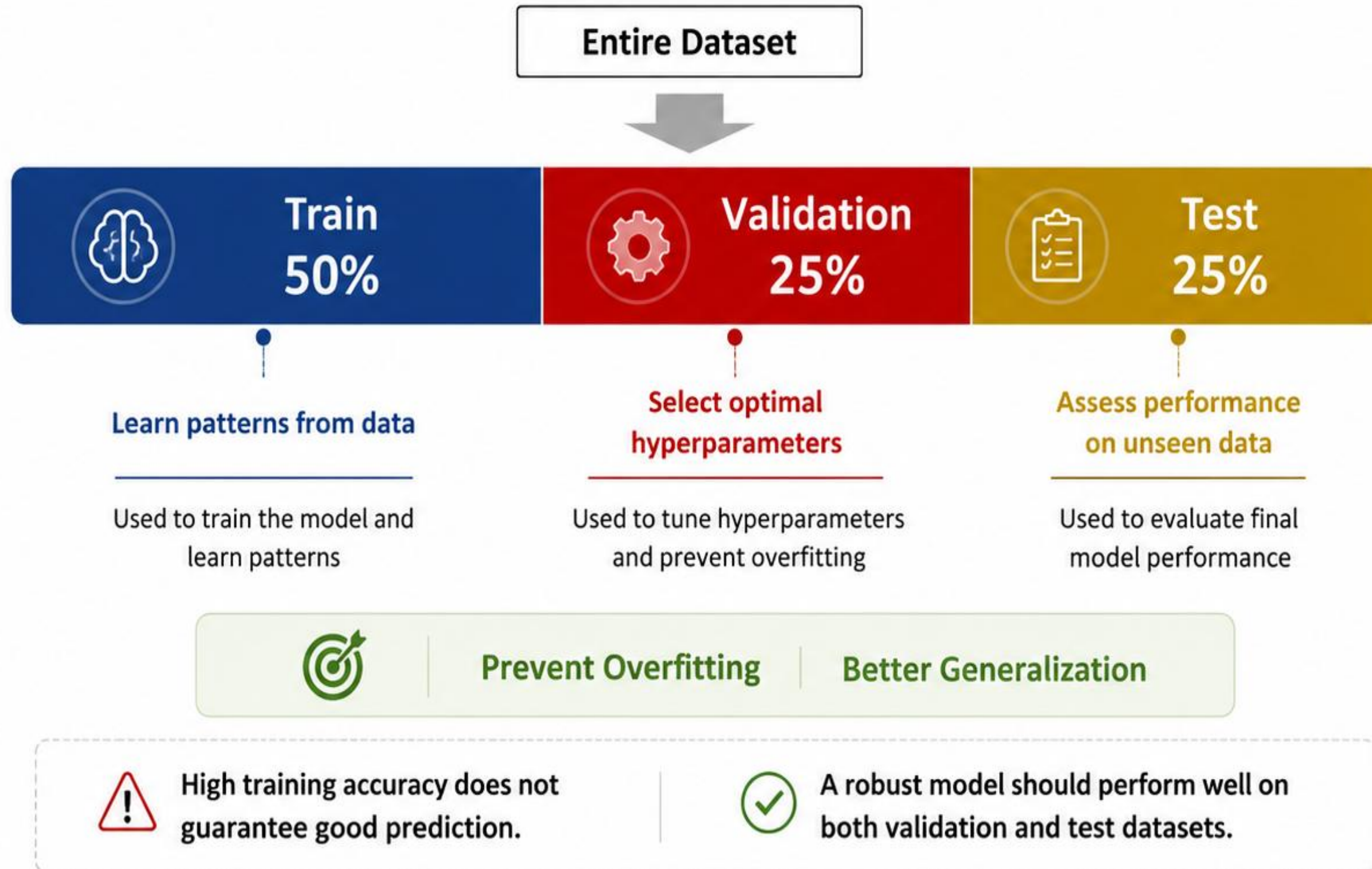
Overfit Model



Key Message

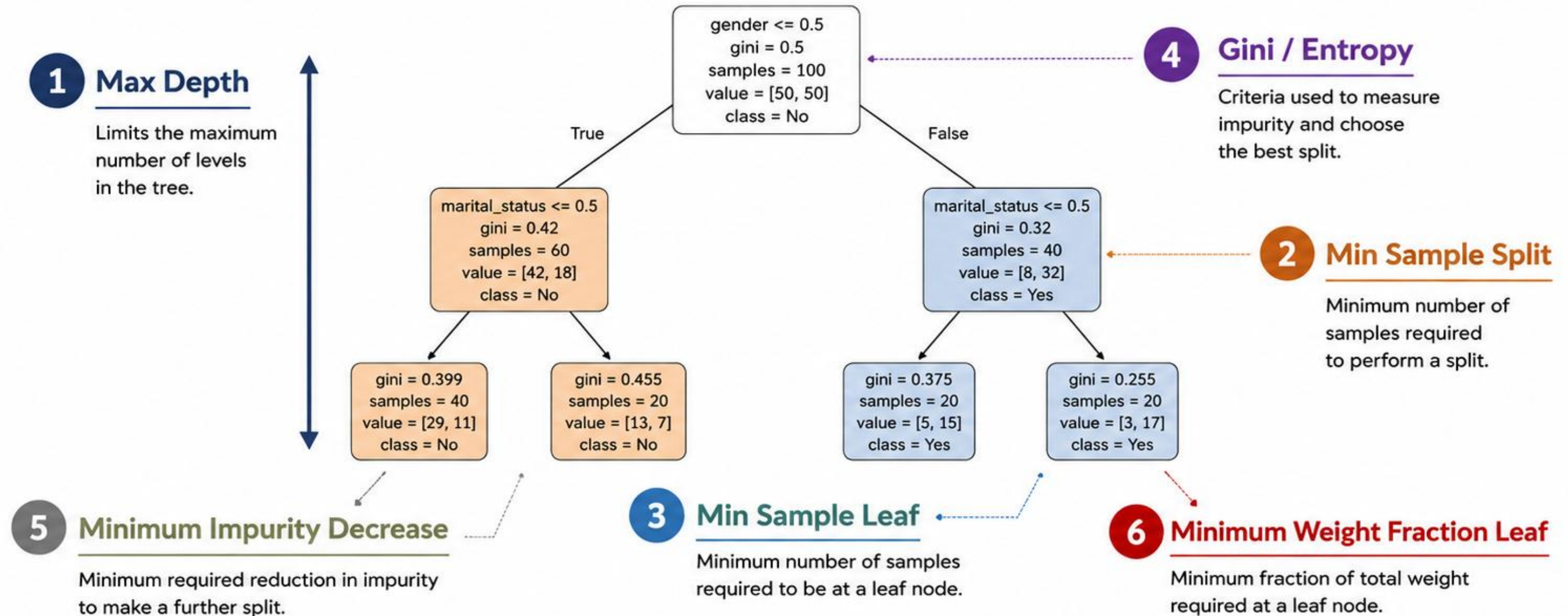
Overfitting happens when a model is too complex and learns the noise. It looks great on **training data**, but performs poorly on **new, unseen data**.

Train, Validation and Test Sets



Hyperparameters That Control Tree Complexity

Preventing Overfitting in Decision Trees

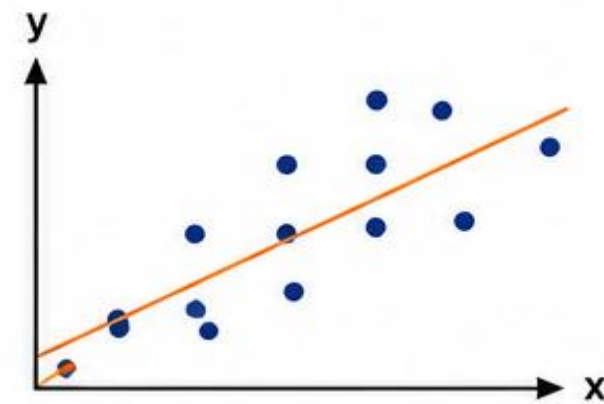


These hyperparameters control the growth of the tree, helping to **prevent overfitting** and improve generalization.

Bias–Variance Tradeoff

Finding the right balance for the best generalization

Logistic Regression



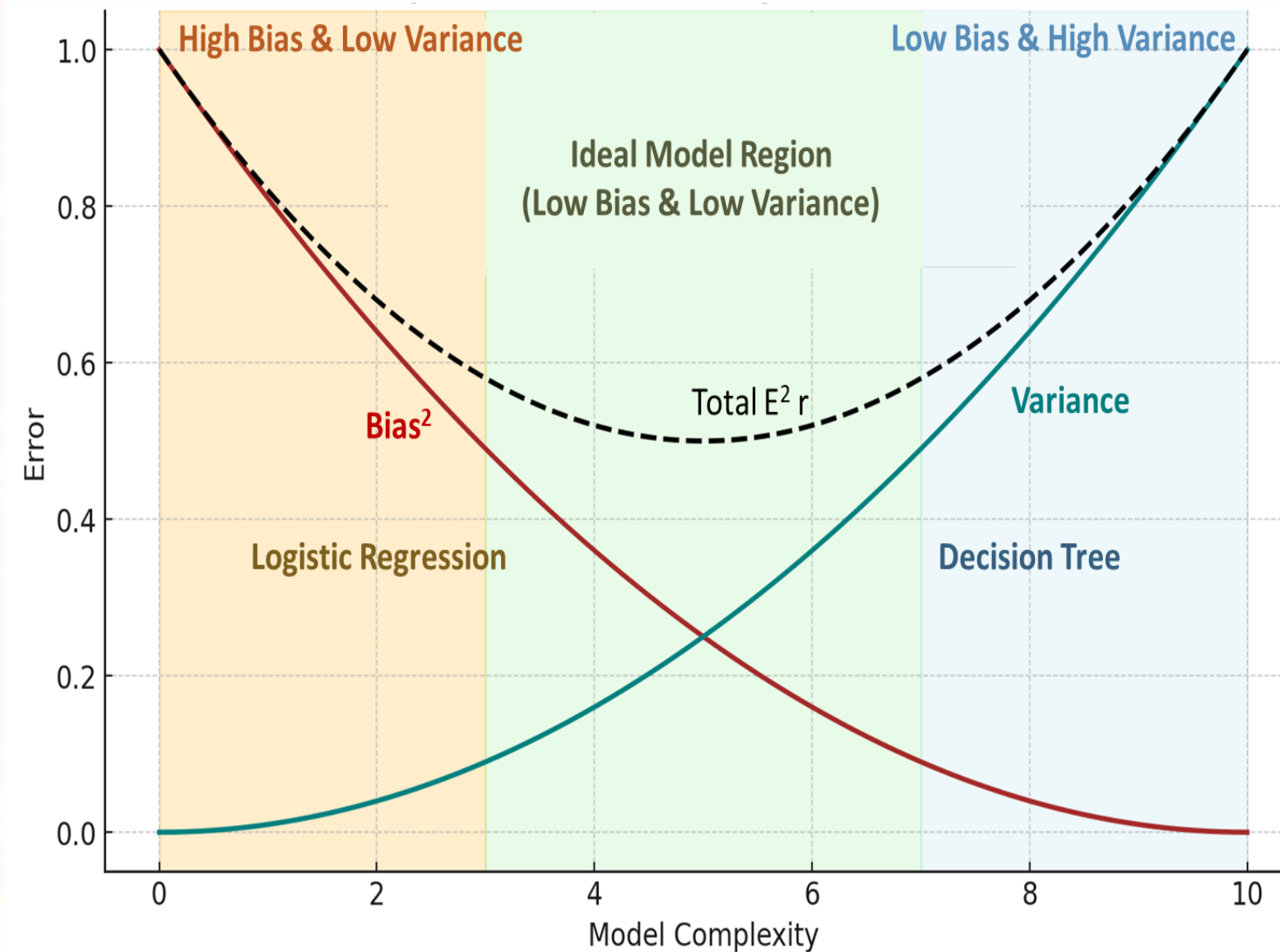
- ✓ High Bias (Underfits)
- ✓ Low Variance (Stable)



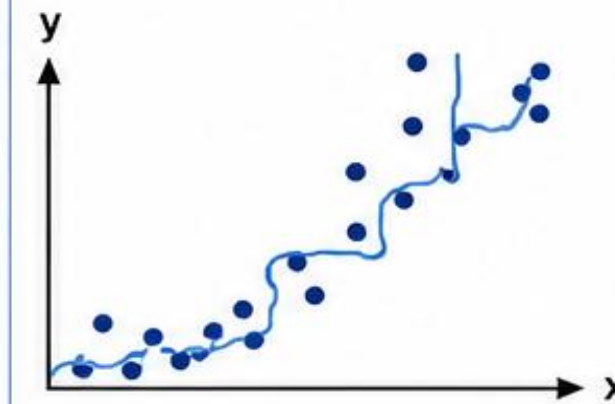
Too simple →
Misses complex patterns

The model is too rigid to capture the true relationship.

Bias–Variance Tradeoff



Decision Tree



- ✓ Low Bias (Captures patterns)
- ✓ High Variance (Unstable)



Too complex →
Overfits the training data

The model learns the noise instead of the true pattern.



Key Takeaway

Overfitting happens when a model is too complex and learns the noise.

We need to balance **Bias** and **Variance** to achieve the best performance on **new, unseen data**.



Solution:

Use ensemble methods such as **Random Forest!**

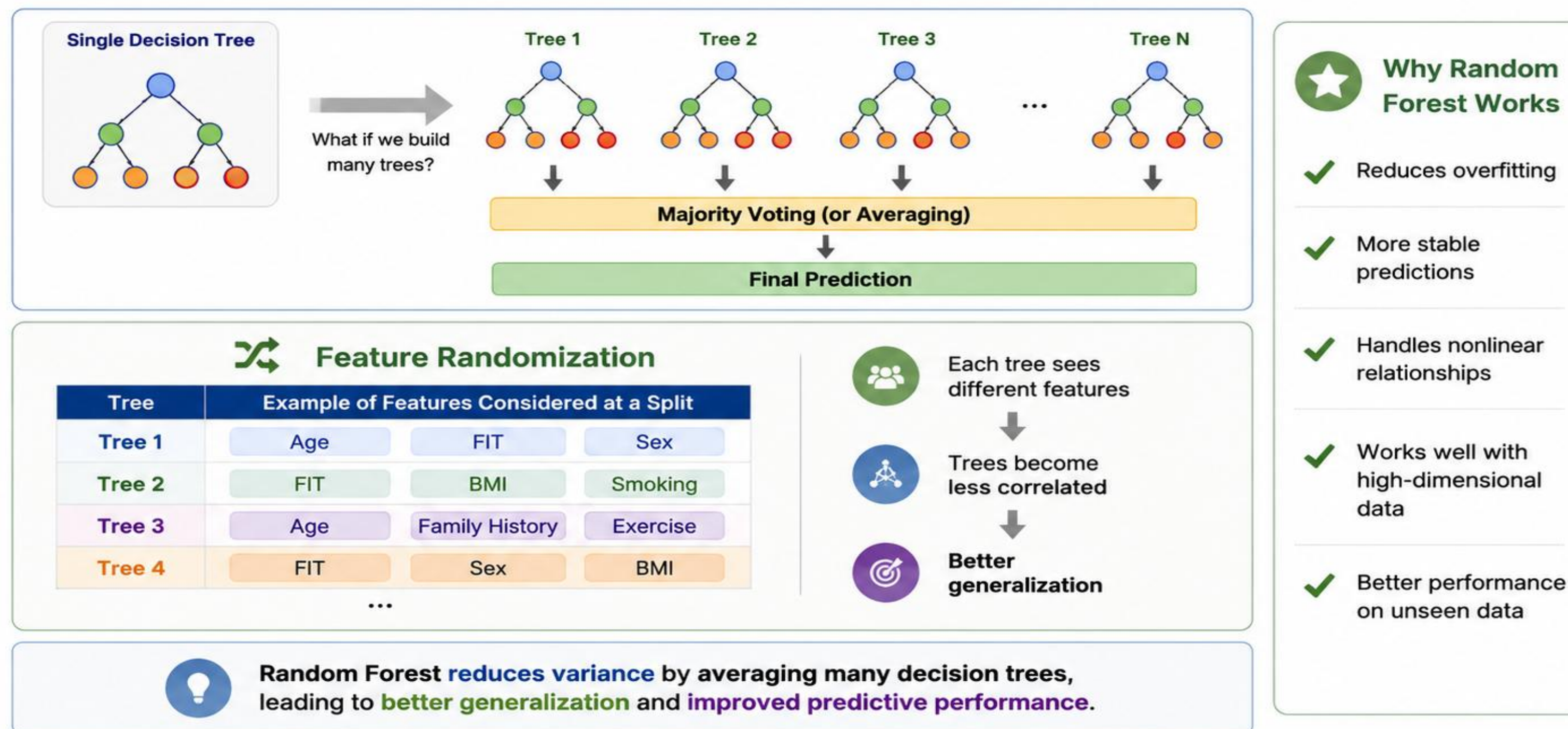
From a Single Decision Tree to a Random Forest

How Random Forest Reduces Overfitting

1 Bootstrap Sampling
Different trees see different subsets of data

2 Feature Randomization
Different trees use different predictors

3 Majority Voting
Combine predictions from all trees



How Do We Measure Model Performance?

From Confusion Matrix to AUC

1 Confusion Matrix

		Actual	
		1 (Positive)	0 (Negative)
Predicted	1 (Positive)	TP (True Positive)	FP (False Positive)
	0 (Negative)	FN (False Negative)	TN (True Negative)



$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



$$Precision = \frac{TP}{TP + FP}$$

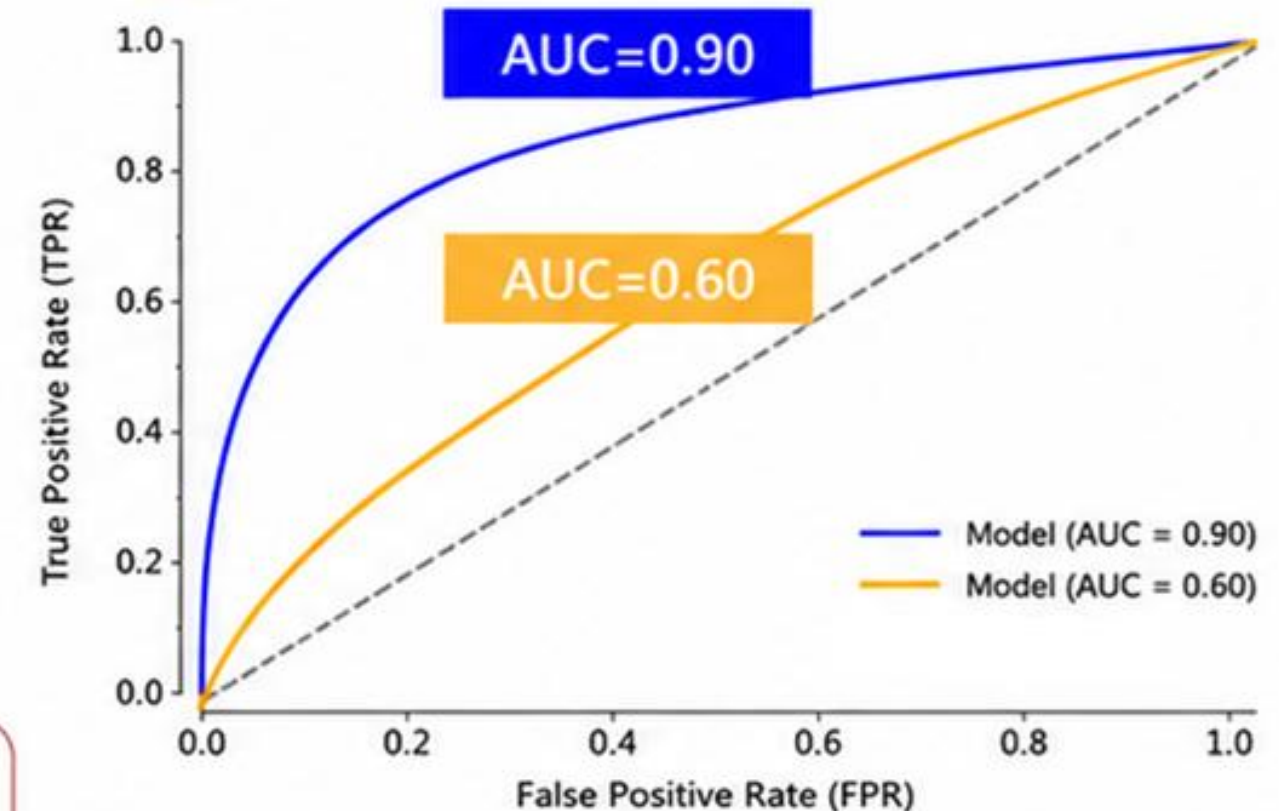


$$Recall (Sensitivity) = \frac{TP}{TP + FN}$$



$$F1 \text{ score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

2 ROC Curve and AUC



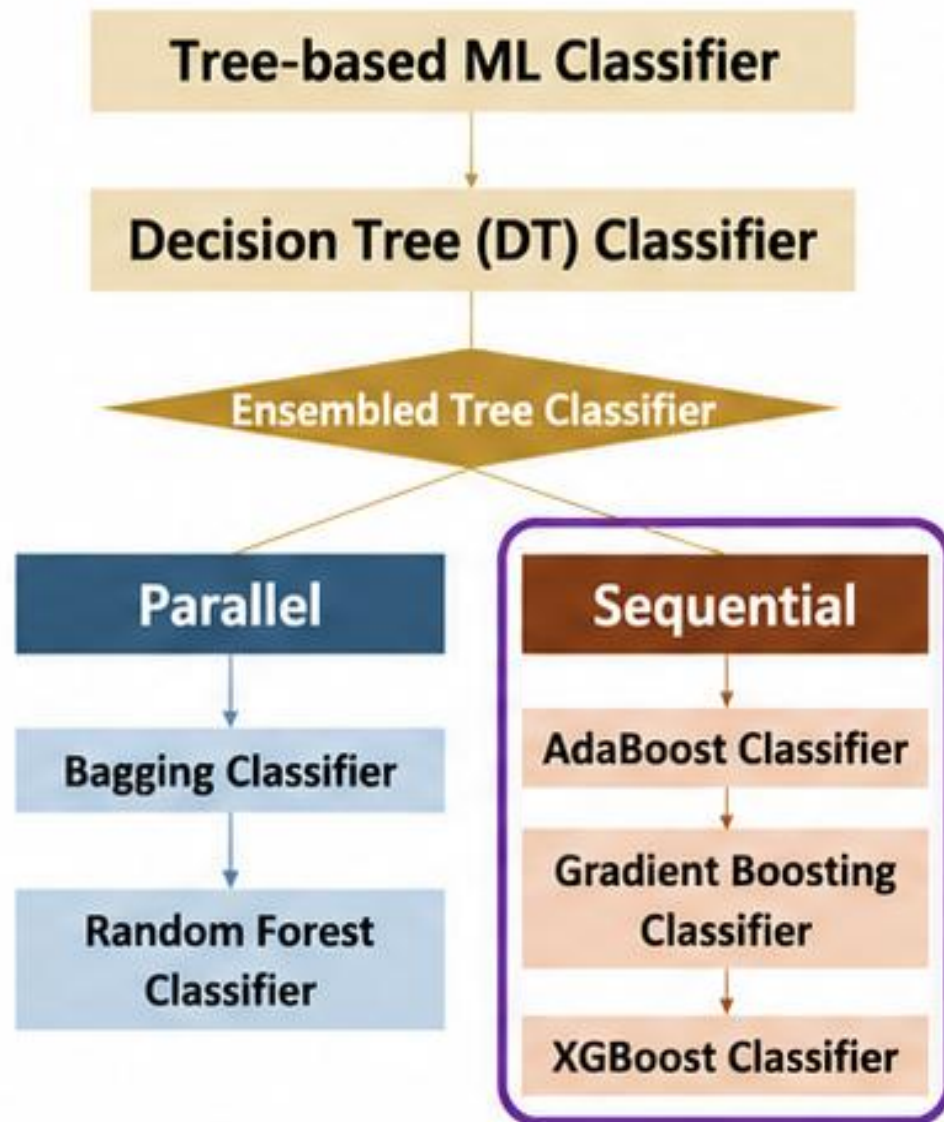
$$G\text{-mean} = \sqrt{Recall \times Precision}$$



$$P\text{-mean} = \sqrt{Recall \times Specificity}$$

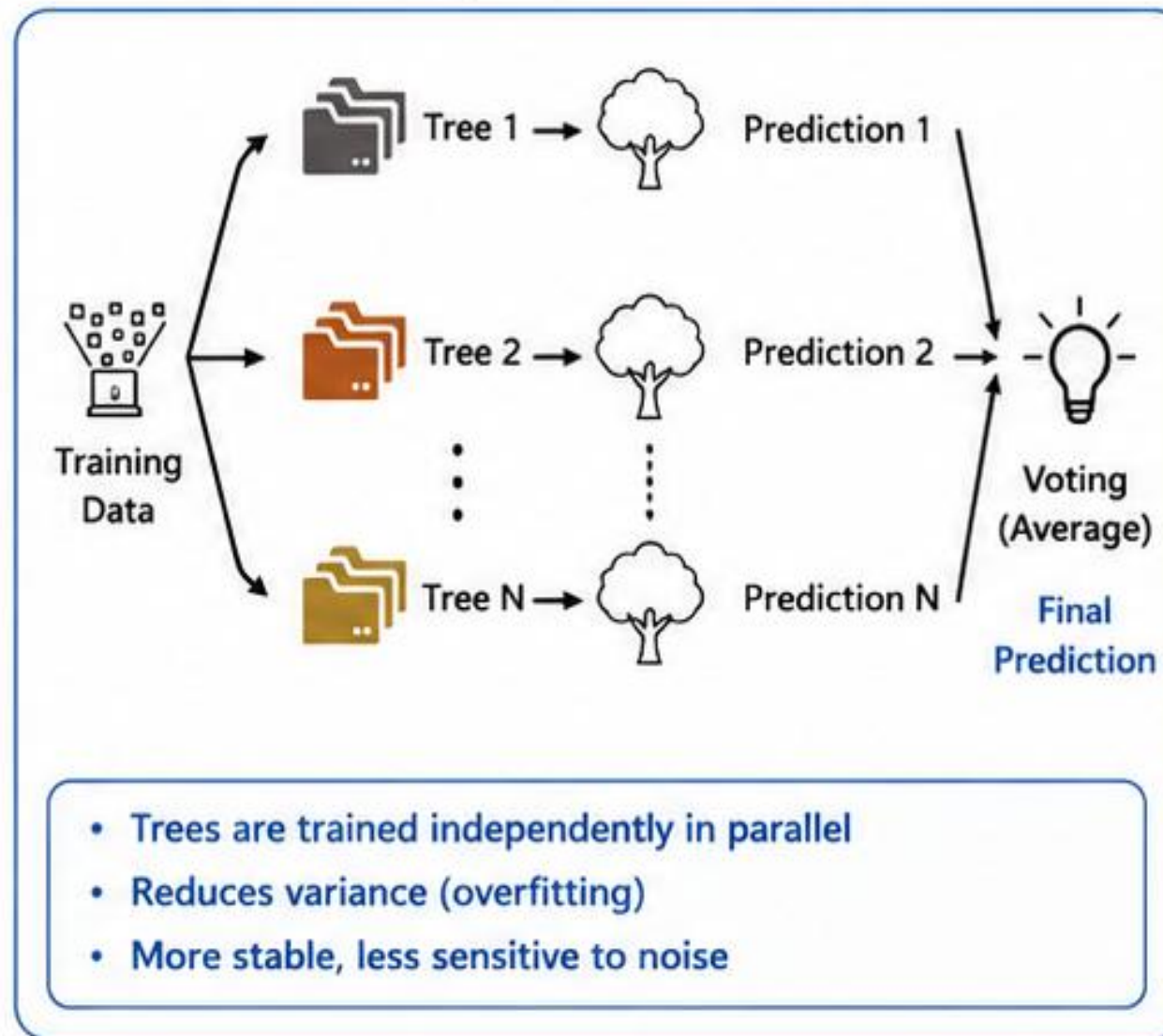
From Random Forest to XGBoost

Ensemble Learning Approaches



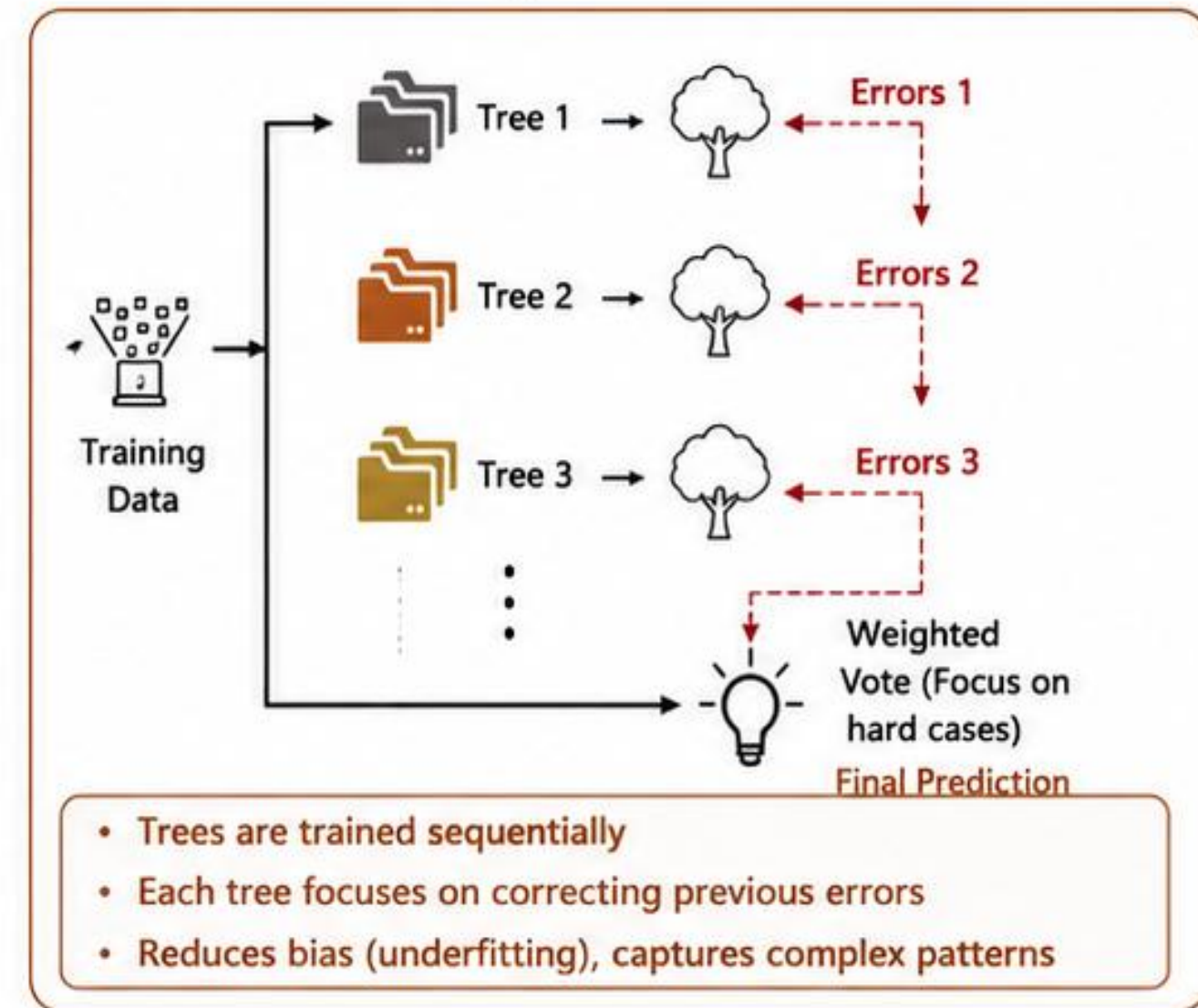
Bagging (Parallel)

Example: Random Forest



Boosting (Sequential)

Example: XGBoost

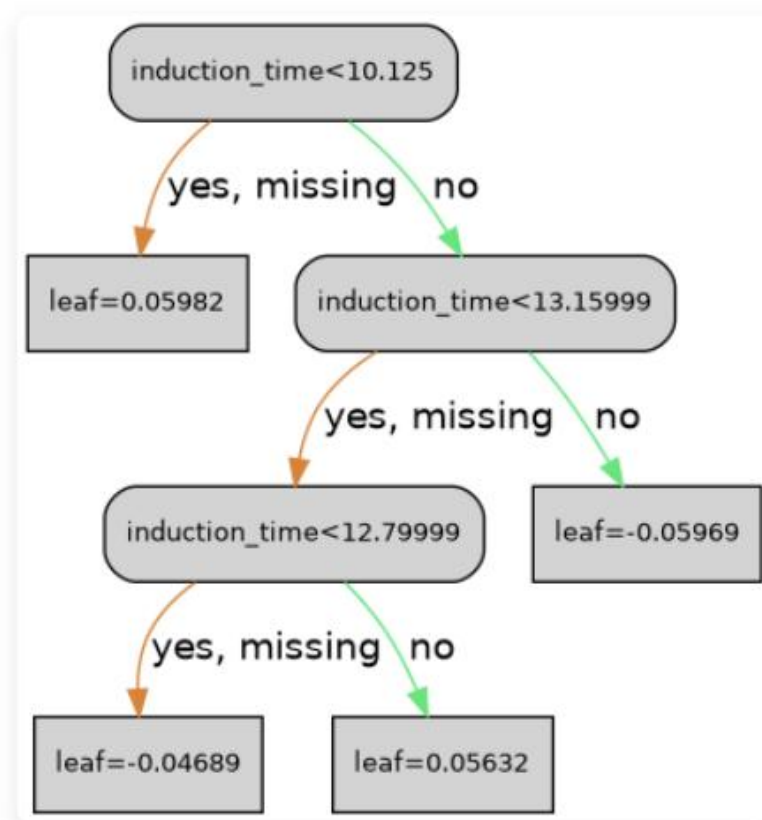


How Does XGBoost Learn?

Each tree learns from the errors of previous trees

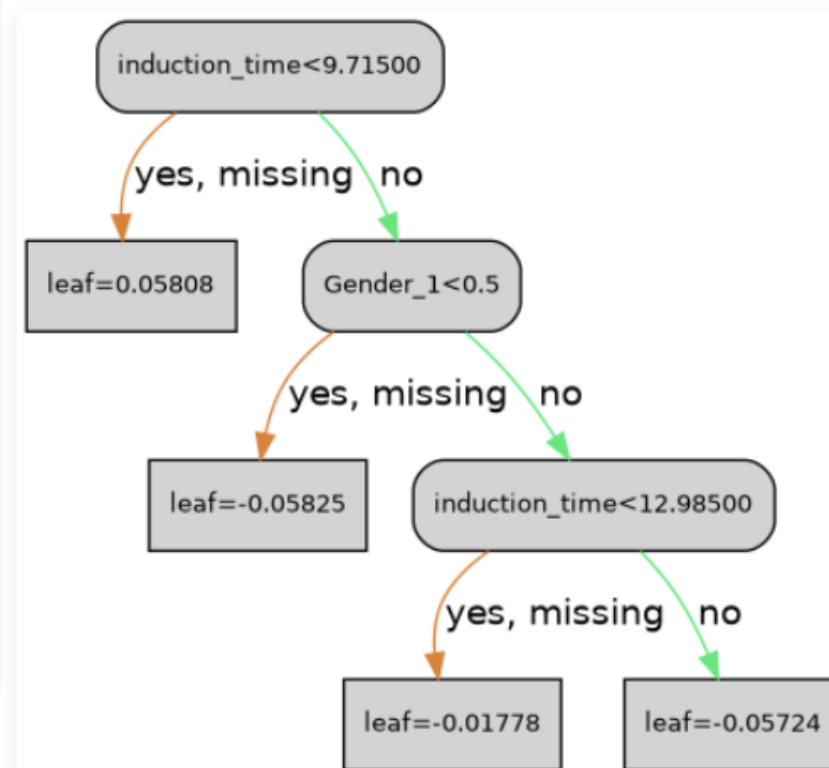
Identify Prediction Errors

First Tree



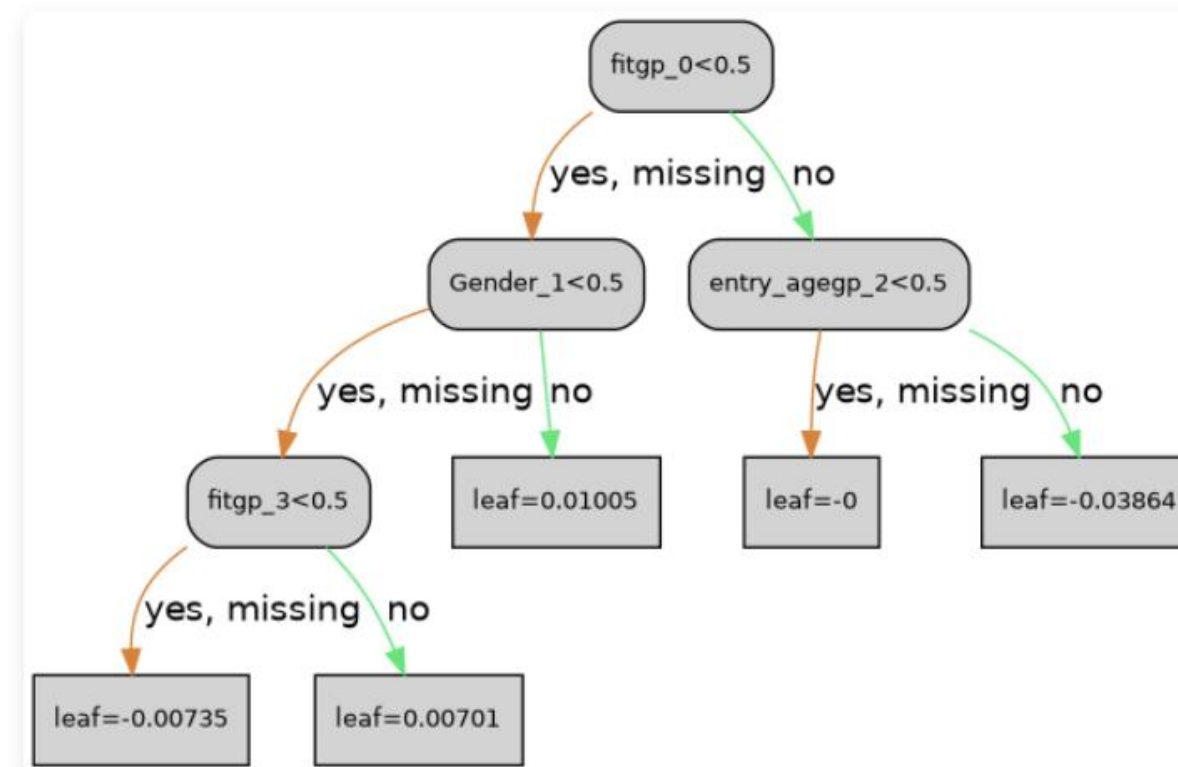
Focus on Difficult Cases

Second Tree



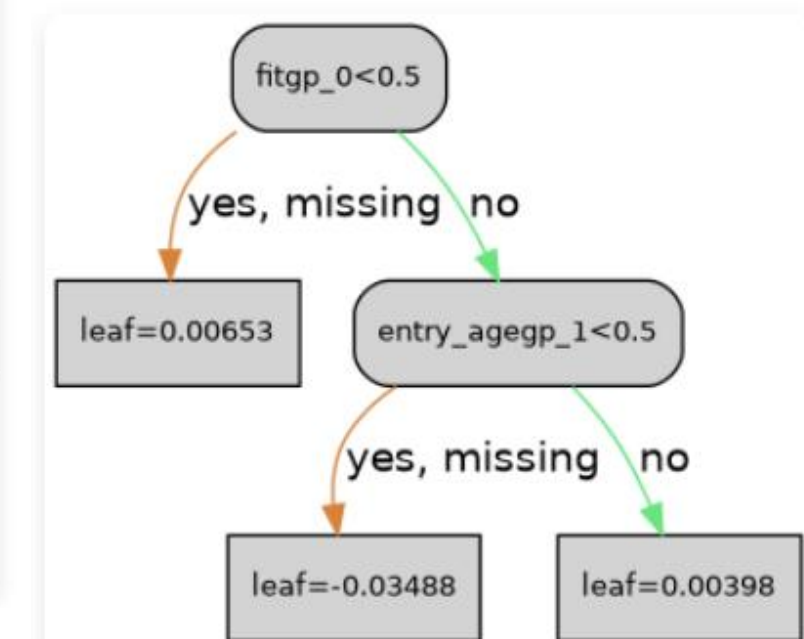
Correct Remaining Errors

Penultimate Tree



Improved Prediction

last Tree



- Each new tree is trained on the **residual errors** from previous trees.
- The model gradually improves prediction by focusing on difficult cases.

Support Vector Machine (SVM)

Classifying CRC Risk Using BMI and FIT Concentration

Screening Cohort (100 people)



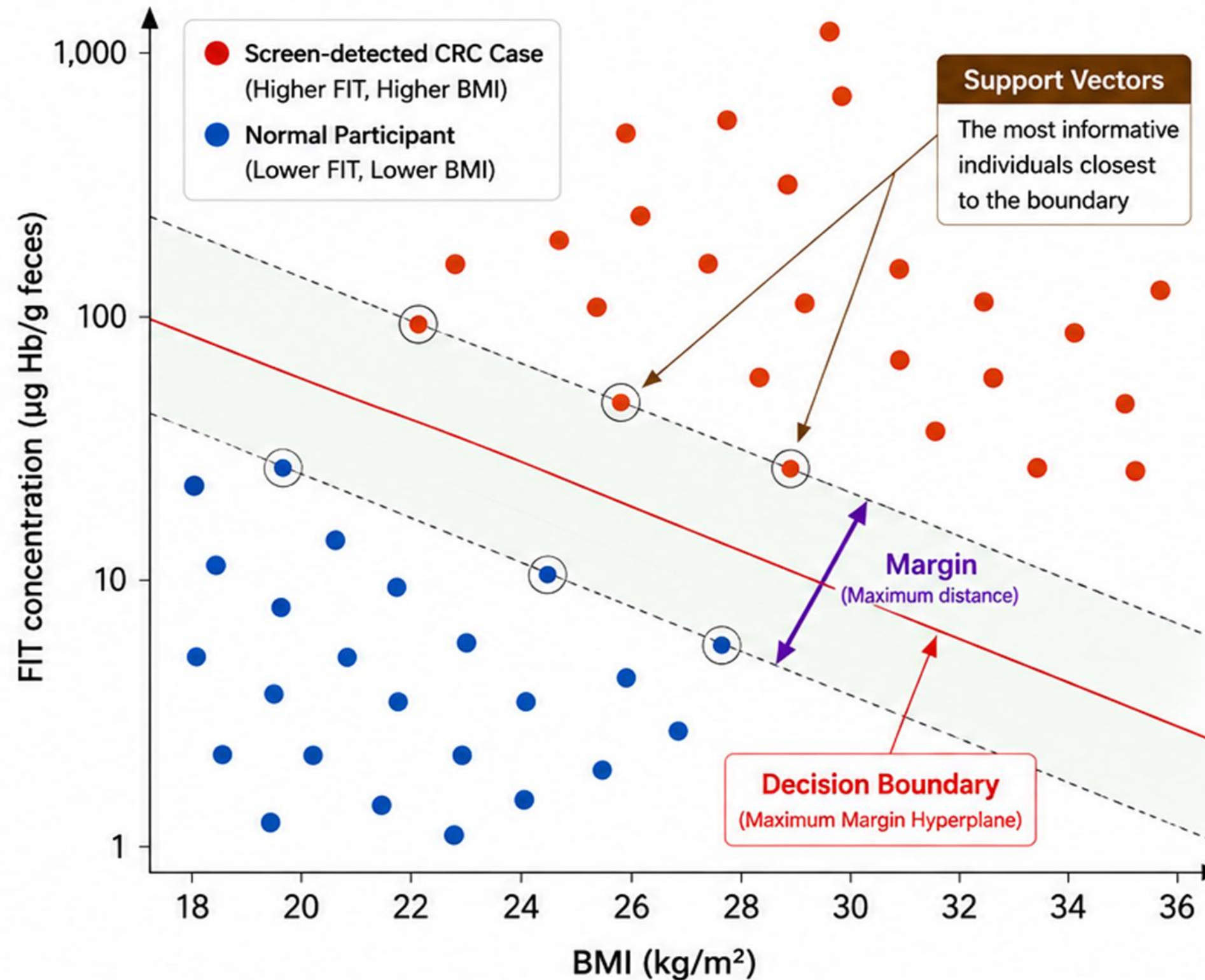
We survey 100 participants in CRC screening program

Features (Continuous Variables)

- X-axis: BMI (kg/m^2)
- Y-axis: FIT concentration ($\mu\text{g Hb/g feces}$, log scale)

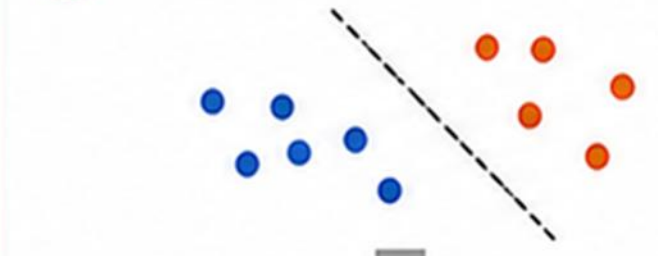
Target

- Classify individuals into:
- Screen-detected CRC Case (Higher FIT, Higher BMI)
 - Normal Participant (Lower FIT, Lower BMI)

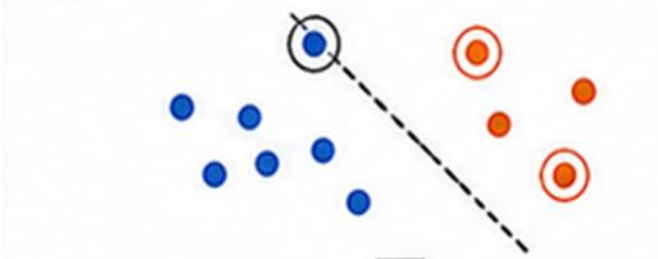


SVM Key Idea

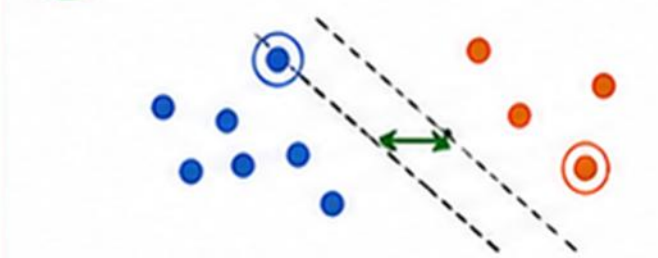
1 Find a decision boundary



2 Identify support vectors



3 Maximize the margin

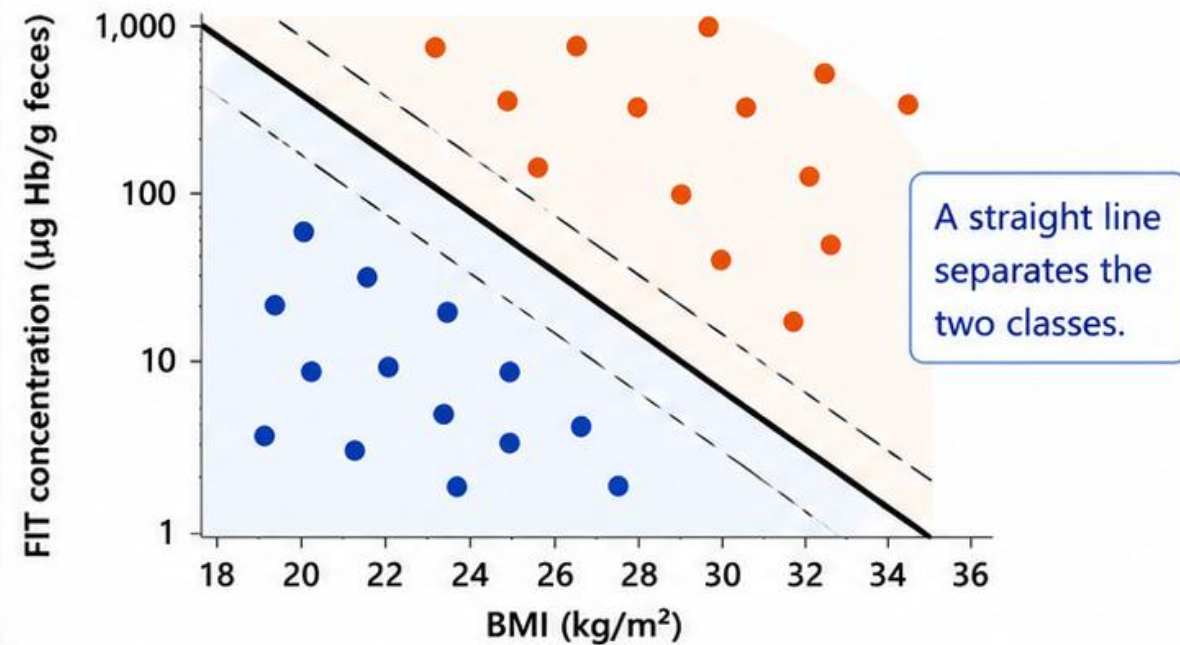


SVM Kernel Functions

Handling Non-linear Relationships in CRC Risk Prediction

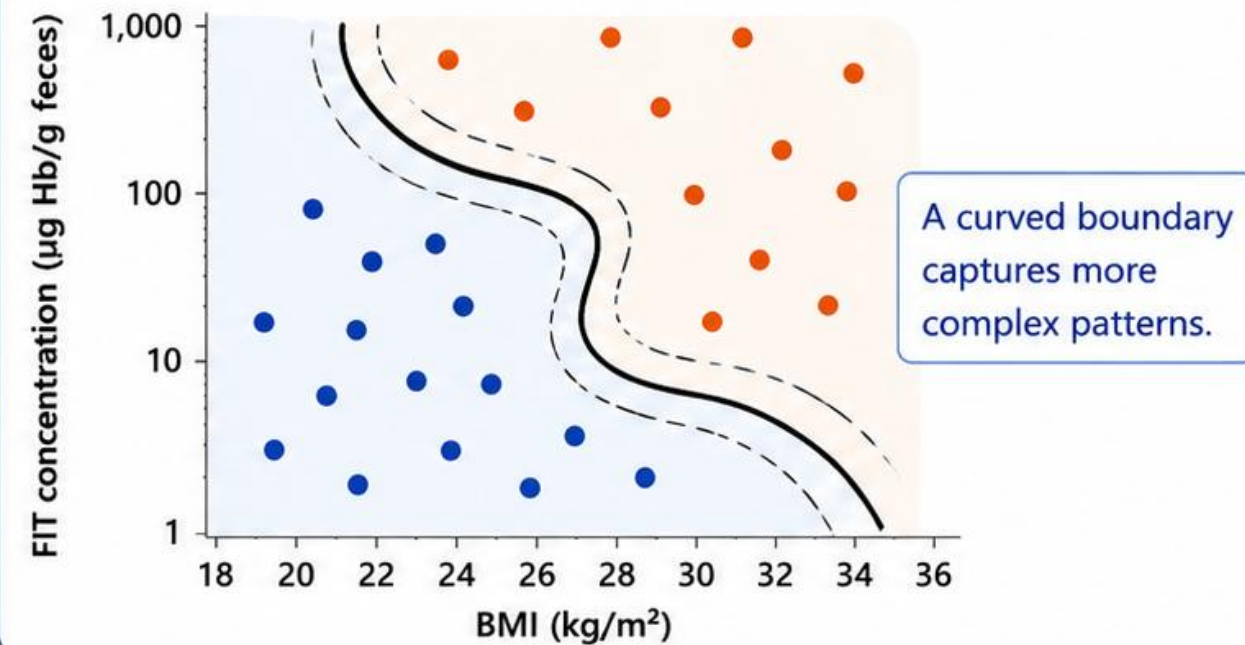
1 Linear Kernel (Linear SVM)

Simplest, fastest



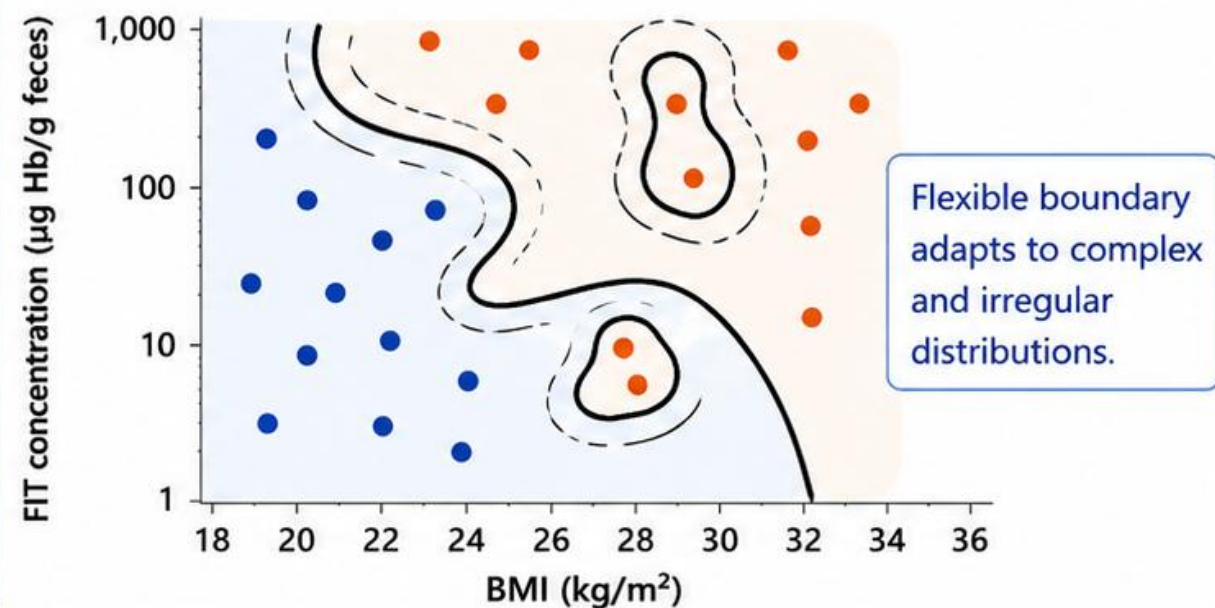
2 Polynomial Kernel

Captures interactions



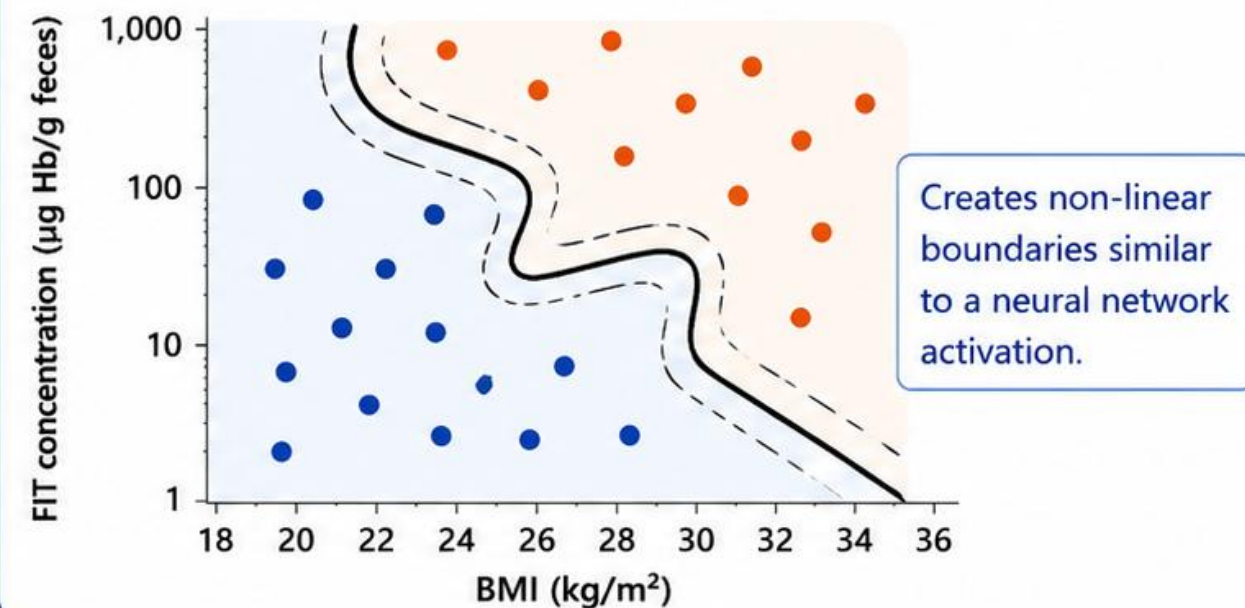
3 RBF Kernel (Gaussian)

Most flexible, widely used



4 Sigmoid Kernel

Similar to neural networks



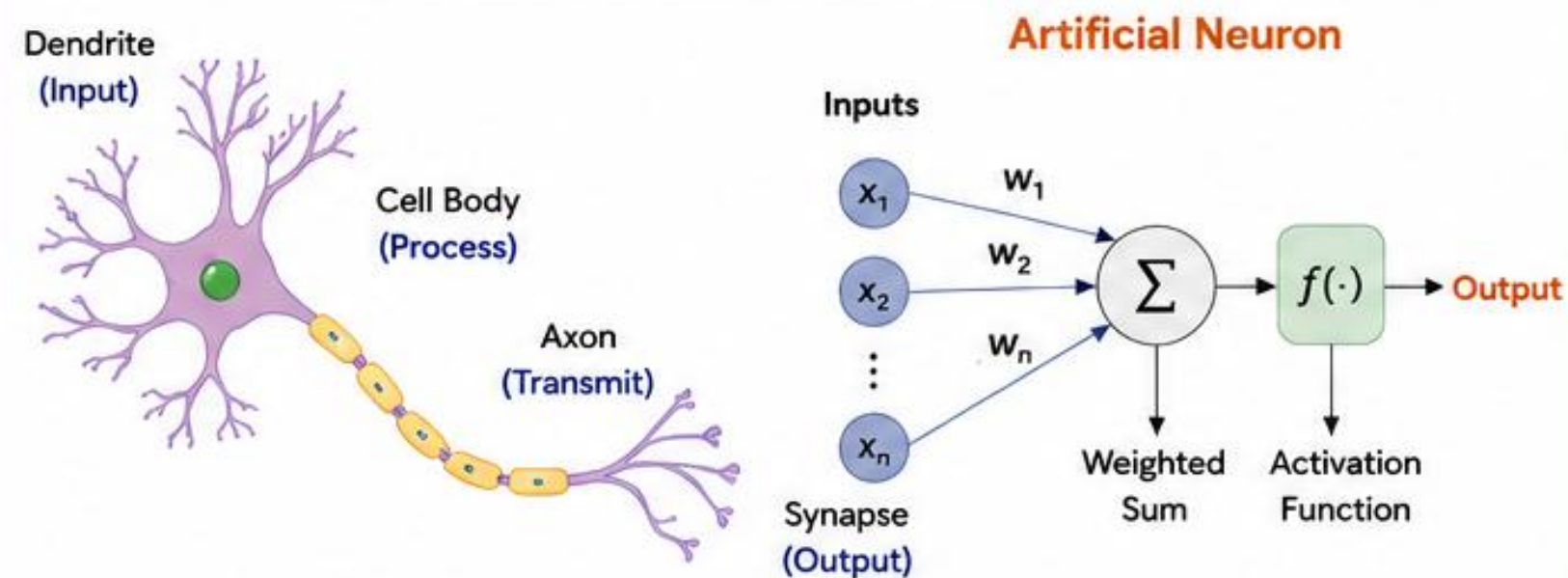
Artificial Neural Networks (ANN)

Learning Complex Patterns from Screening Data



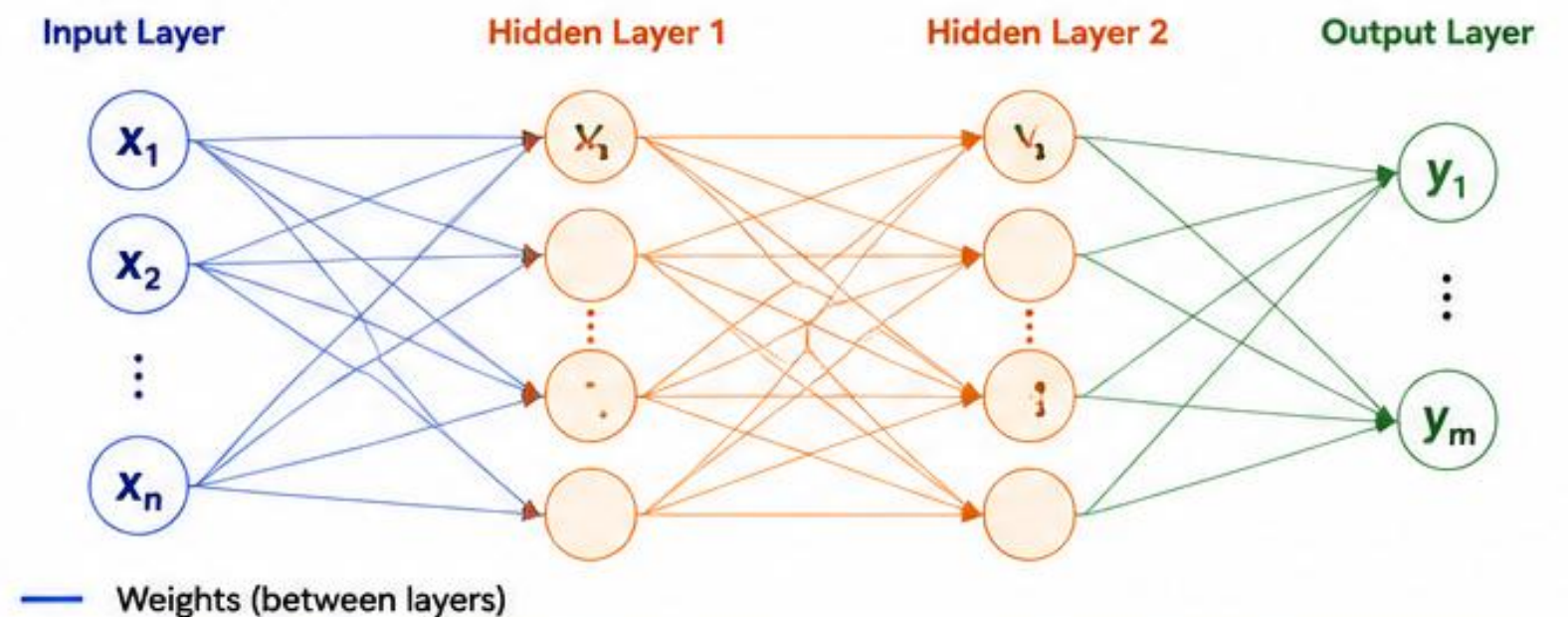
ANNs are inspired by the human brain and can **automatically learn complex, non-linear relationships** among many variables to **improve prediction accuracy**.

1. Inspired by the Human Brain



Many inputs are combined, processed, and passed through an activation function to produce an output.

2. How ANN Works (Forward Propagation)



Each neuron receives inputs, applies weights and an activation function, and passes the result to the next layer.

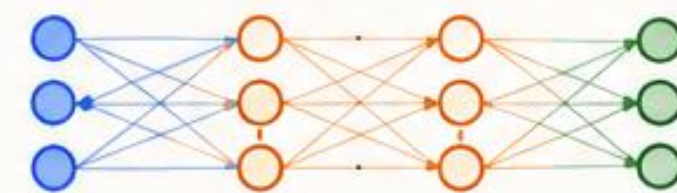
3. How ANN Works for CRC Risk Prediction

1. Input (Screening & Clinical Variables)



Many variables describe each individual.

2. Neural Network (Learn Hidden Patterns)



Automatically learns complex, non-linear relationships and hidden disease patterns.

3. Output (Prediction)

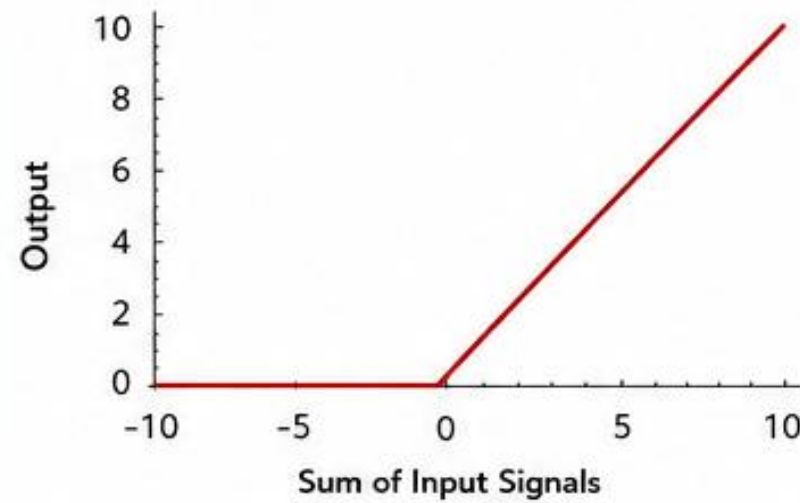


Activation Function in ANN



Activation function introduces **non-linearity** into ANN, enabling the network to learn **complex patterns** from data.

1 ReLU (Rectified Linear Unit)



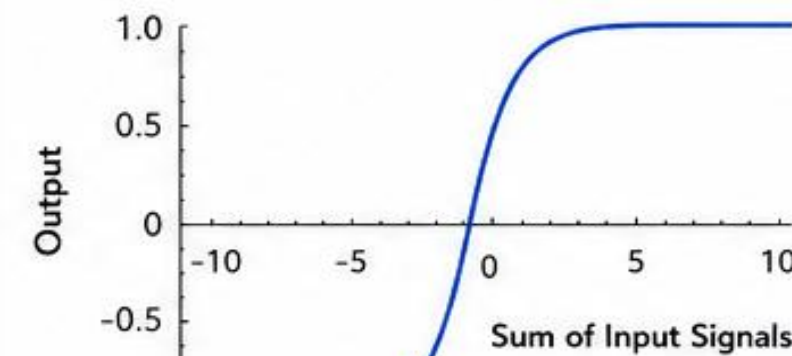
$$f(x) = \max(0, x)$$

- Outputs 0 for negative values and x for positive values.
- Simple and computationally efficient.
- Most widely used in hidden layers.



Fast training, reduces vanishing gradient problem.

2 Tanh (Hyperbolic Tangent)



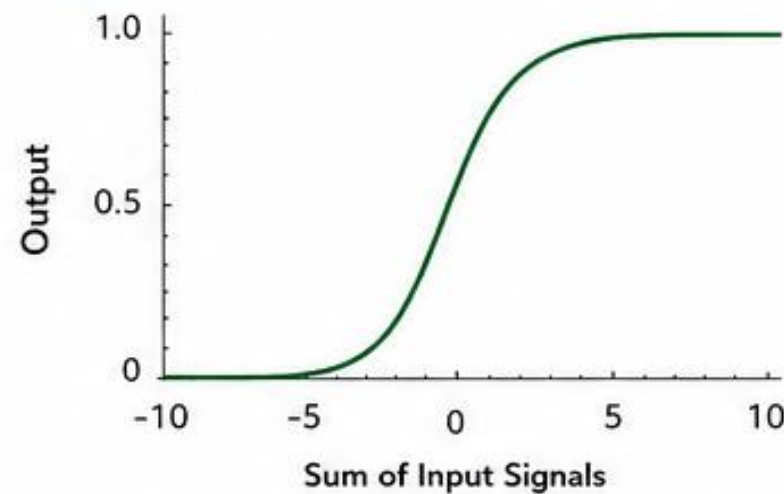
$$f(x) = \tanh(x)$$

- Outputs values between -1 and 1.
- Zero-centered output (mean close to 0).
- Better than sigmoid for hidden layers.



Stronger gradient than sigmoid, improves learning.

3 Sigmoid (Logistic)



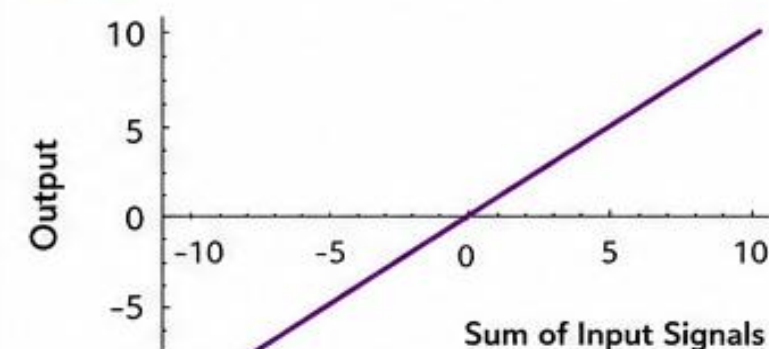
$$f(x) = \frac{1}{1 + e^{-x}}$$

- Outputs values between 0 and 1.
- Commonly used in output layer for binary classification.
- Not ideal for hidden layers (vanishing gradient).



Good for probability output (e.g., 0 to 1).

4 Linear (Identity)



$$f(x) = x$$

- Outputs the input directly.
- Used in regression output layer.
- Not suitable for hidden layers (no non-linearity).



Useful for continuous output problems.

ANN Demo: Predicting Breast Cancer Death Risk

Evolution of Prognostic Factors in Breast Cancer

1st Generation
Traditional Clinical Factors
 TNM staging
 T (size), N (nodes), M (metastasis)

2nd Generation
Molecular Biomarkers
 ER, PR, HER2, Ki67, etc.

3rd Generation
Imaging Biomarkers & AI-discovered Patterns
 Long-term outcomes by imaging features

Size	Grade	Nodes	Extent	ER	PR	Basal like	LVI	nbcd	Mammography appearance	HER2	Detection mode	Focality
26.0	1	0	0	1	1	0	0	0	2	0	1	1
15.0	2	1	1	1	0	0	0	0	5	0	1	2
20.0	2	0	0	1	1	0	0	0	5	0	1	1
14.0	2	0	0	1	0	0	0	0	1	1	1	1
14.0	2	0	1	1	1	0	1	0	2	0	1	2
5.0	2	9	0	1	0	0	0	0	2	0	1	1
22.0	2	0	0	1	0	0	0	0	1	0	1	1
2.0	3	0	0	0	0	0	0	0	4	1	1	3

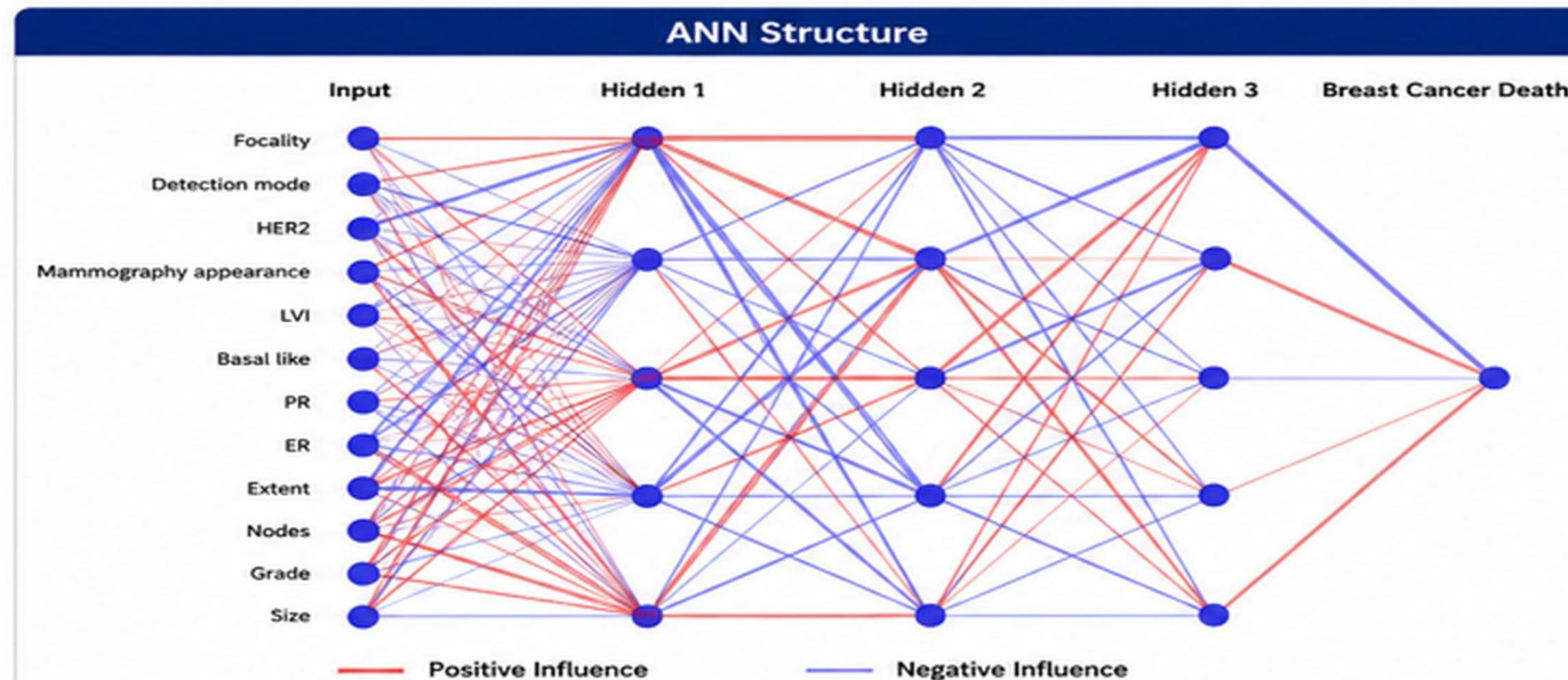
Data source: Prof. László Tabár, Sweden.

ANN Demo: Predicting Breast Cancer Death Risk

From Input Variables → Neural Network → Prediction



This demo shows how an ANN **learns from clinical and image features** to **predict breast cancer death risk**.



How It Works



Input Variables
Clinical & imaging features are entered.



ANN Learning
The network learns complex patterns in the data.



Hidden Layers
Capture non-linear relationships among features.



Output Prediction
The model predicts the probability of breast cancer death.

Output



Breast Cancer Death Risk
0.76 (High Risk)
Probability (0 to 1)

But What If the Input Is an Image?

1. Structured Data

(User Original Data)

Age	Size	Grade	Nodes	Extent	ER	PR	Basal like	LVI	nbc	Mammography appearance	Diagnose
26.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0	0	2	69
15.0	2.0	1.0	1.0	1.0	0.0	0.0	0.0	0	0	5	69
20.0	2.0	0.0	0.0	1.0	1.0	0.0	0.0	0	0	5	69
14.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	0	0	1	69
14.0	2.0	0.0	1.0	1.0	1.0	0.0	1.0	0	0	2	69
5.0	2.0	NaN	0.0	1.0	0.0	0.0	0.0	0	0	2	68
22.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	0	0	1	68



31 input features



ANN



Structured Variables

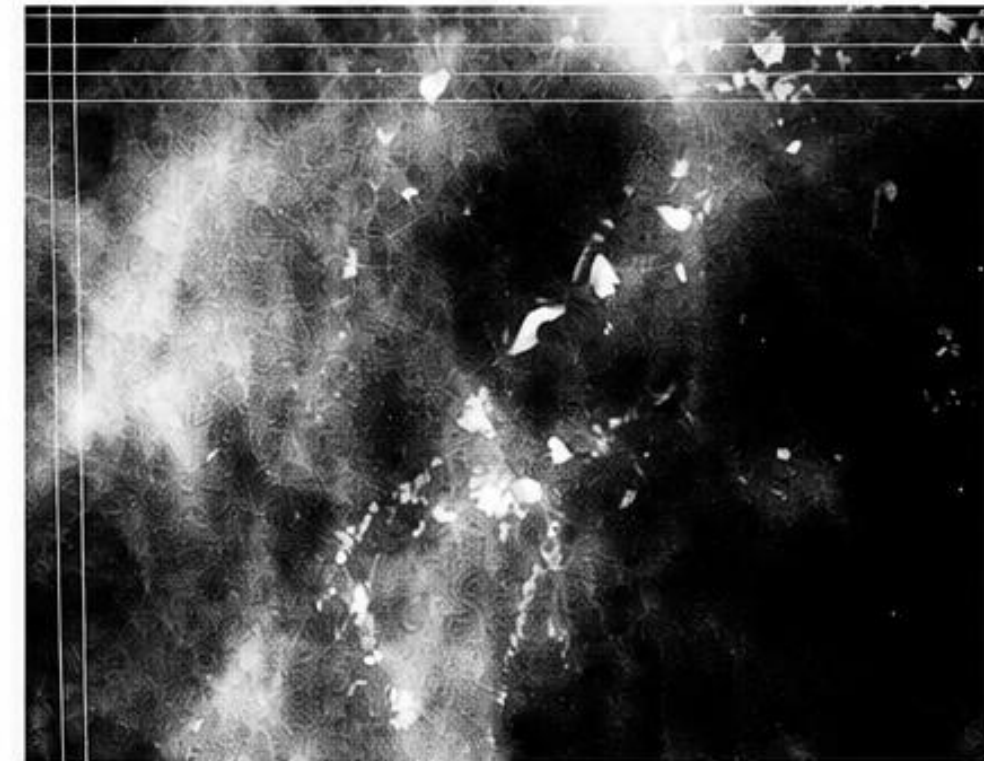
→ ANN



Images

→ CNN

2. Mammography Image



256 × 256 pixels
=
65,536 input features



CNN



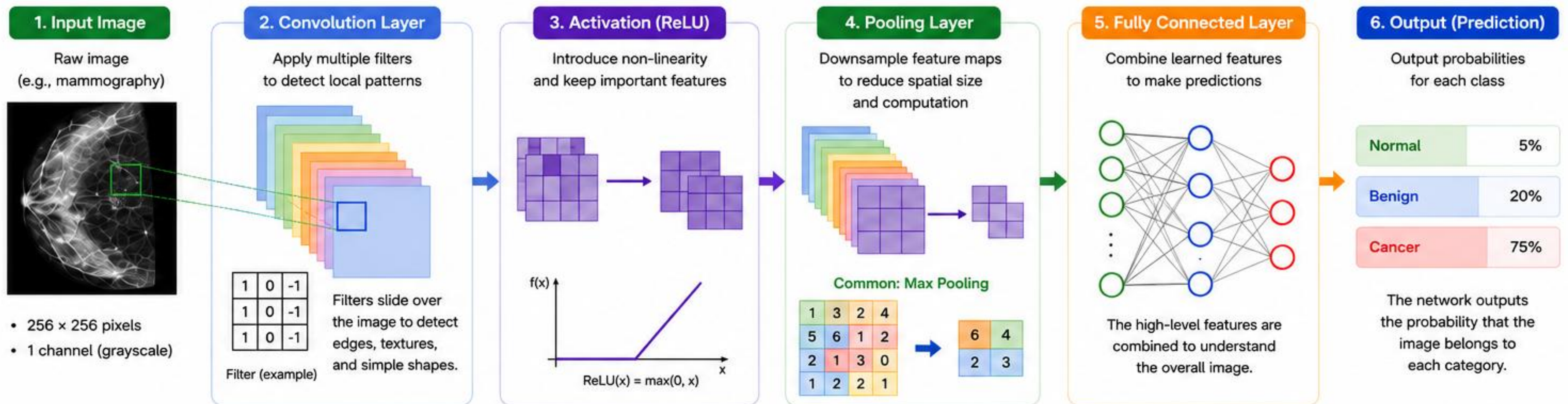
Too many parameters



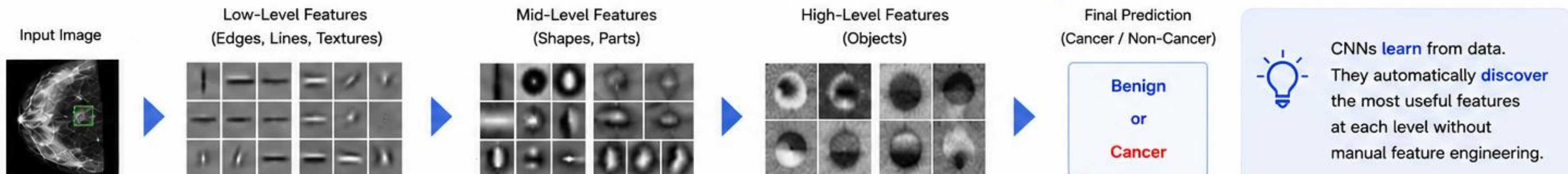
Loss of spatial information

How CNN Learns from Images

From pixels to edges, shapes, and clinically meaningful patterns



How CNN Learns: Hierarchical Feature Learning



LSTM: Learning Long-Term Dependencies

An enhanced neural network designed to learn from sequential data over time.

Why do we need LSTM?



Screening data are collected **repeatedly** over time.

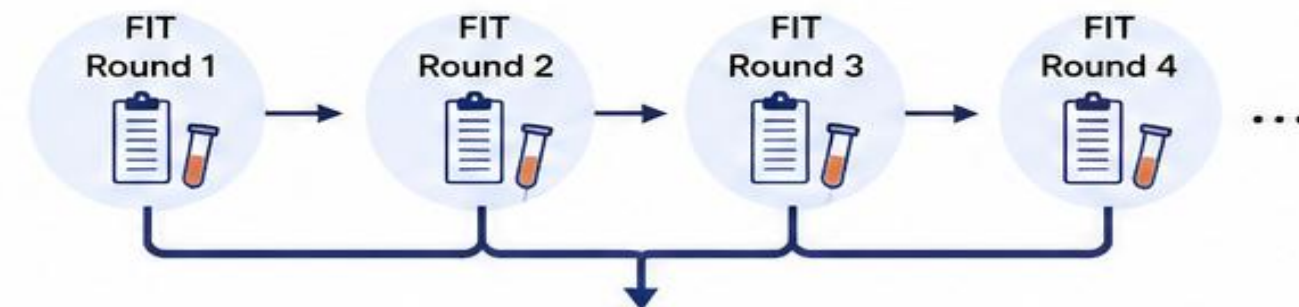


A single FIT result provides only a **snapshot**.



LSTM uses a **Memory Cell** to preserve important information across multiple time points.

Example: Longitudinal FIT Results



LSTM learns patterns across the **entire** screening history to predict future risk.



LSTM selectively remembers important information from multiple screening rounds and ignores irrelevant information, enabling **dynamic and personalized** colorectal cancer **risk prediction**.


**Personalized
Risk Prediction**

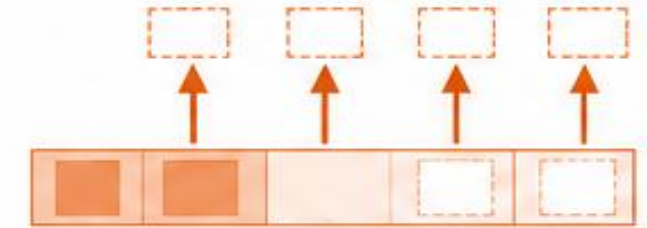
How does LSTM work?

LSTM uses a Memory Cell and three gates to make smart decisions.



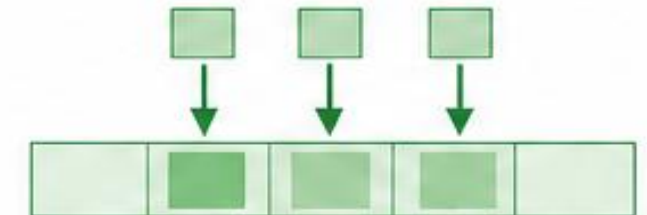
1. Forget Gate (f_t)

Decides what information is no longer important.



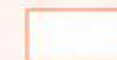
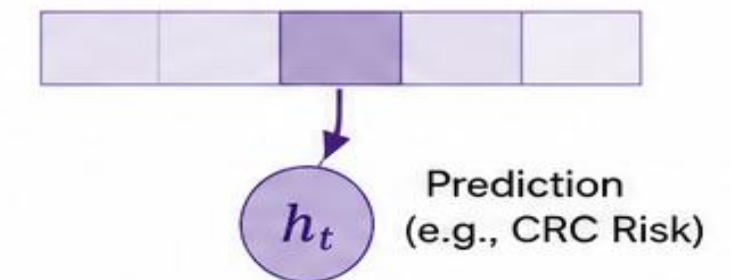
2. Input Gate (i_t)

Decides what new information should be remembered.



3. Output Gate (o_t)

Decides what information should be used for prediction.



Less important information



Important information

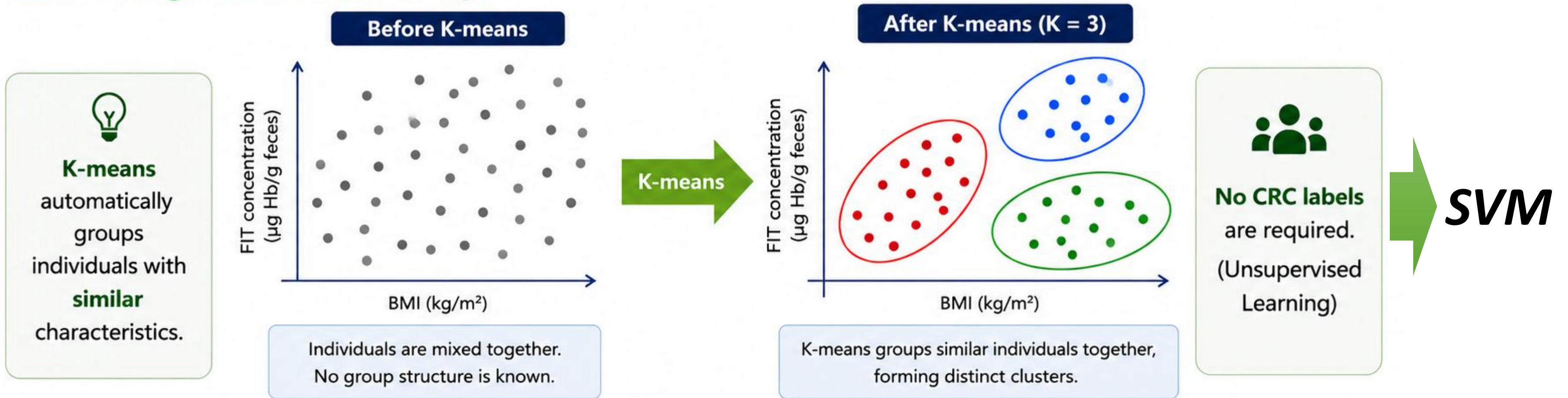


Unsupervised Learning

Learning from unlabeled data

K-means Clustering

Discovering Hidden Risk Groups



Example: Clustering by BMI and FIT Concentration (K = 3)

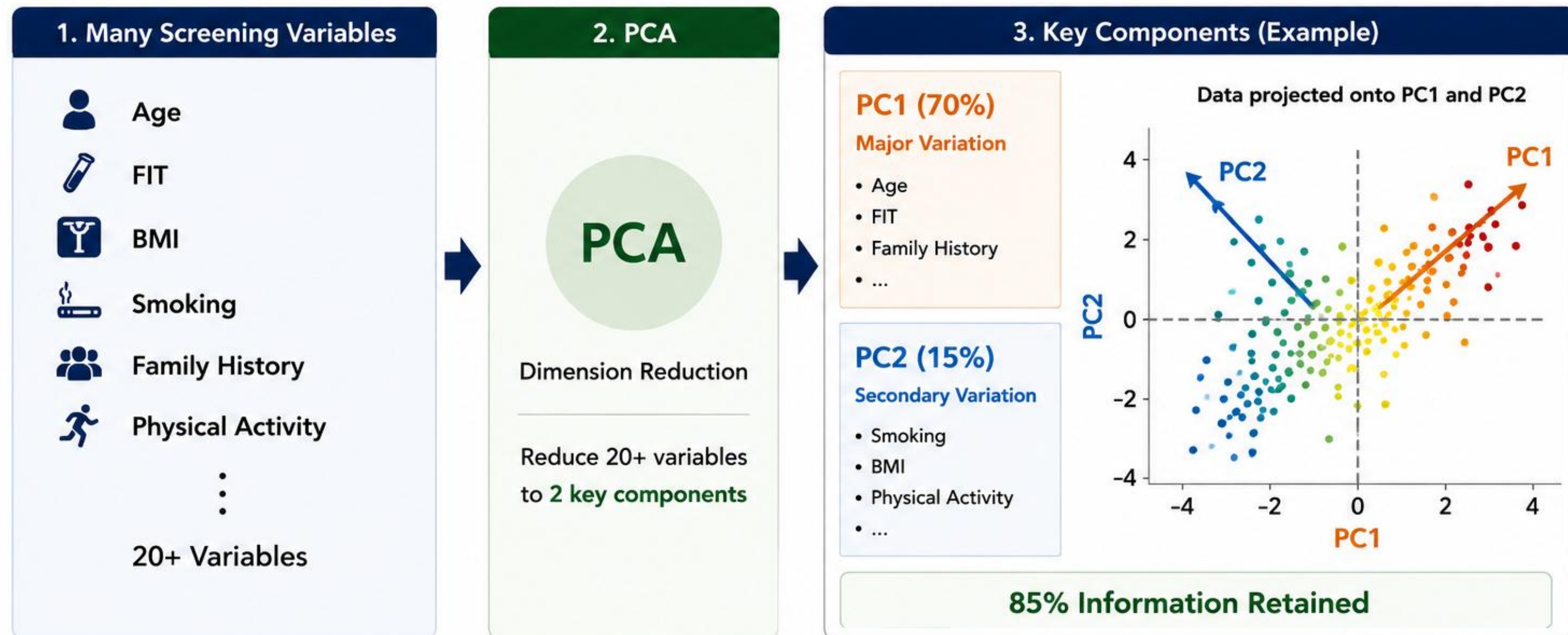


Principal Component Analysis (PCA)

Reduce Many Variables into Key Components



PCA transforms **correlated variables** into a small set of **uncorrelated components**, preserving most of the **important information** in the data.



Take-home Message

PCA reduces complex screening data into a **few key components** while **preserving most of the information**.

*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

II. Artificial Intelligence and Machine Learning for Precision Screening

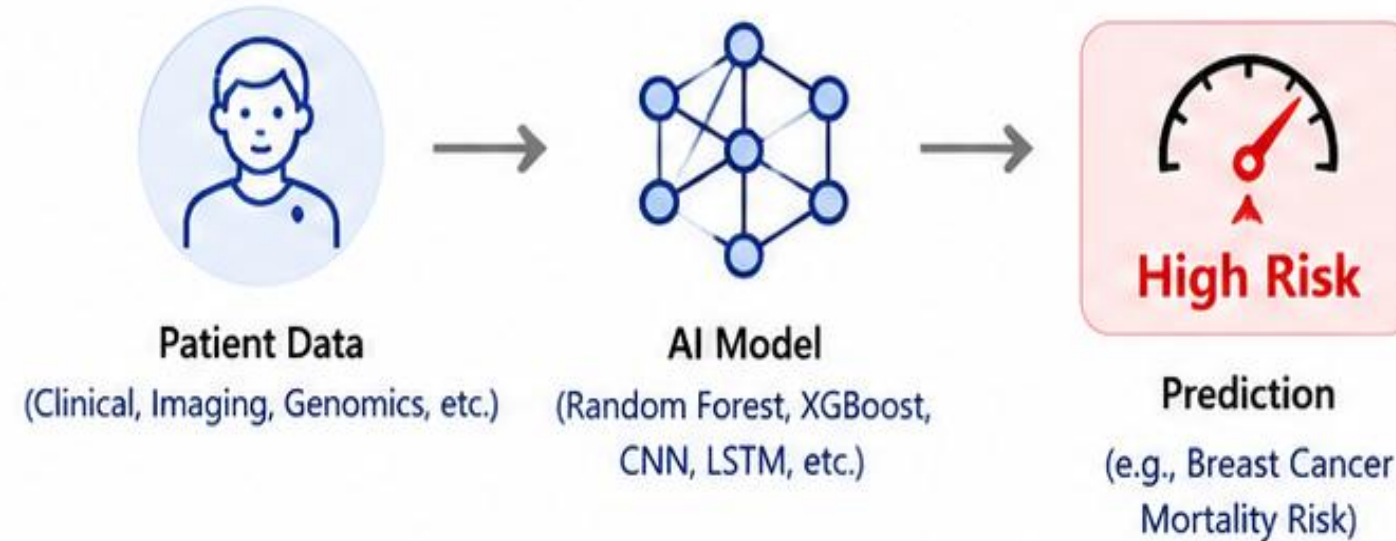
Explainable Artificial Intelligence (XAI) for Personalized Risk Assessment and Prediction



Explainable AI (XAI) for Personalized Risk Assessment and Prediction

*Accurate predictions are not enough. We need to understand **WHY**.*

1. AI Models Can Make Accurate Predictions



2. The Black Box Problem



- Why was this patient predicted as high risk?
- Which factors contributed the most?

3. The Need for Explainable AI (XAI)

-  **Build Trust**
Clinicians can understand and trust the model.
-  **Support Clinical Decisions**
Understand the key factors behind each prediction.
-  **Empower Patients**
Provide clear explanations to support shared decisions.



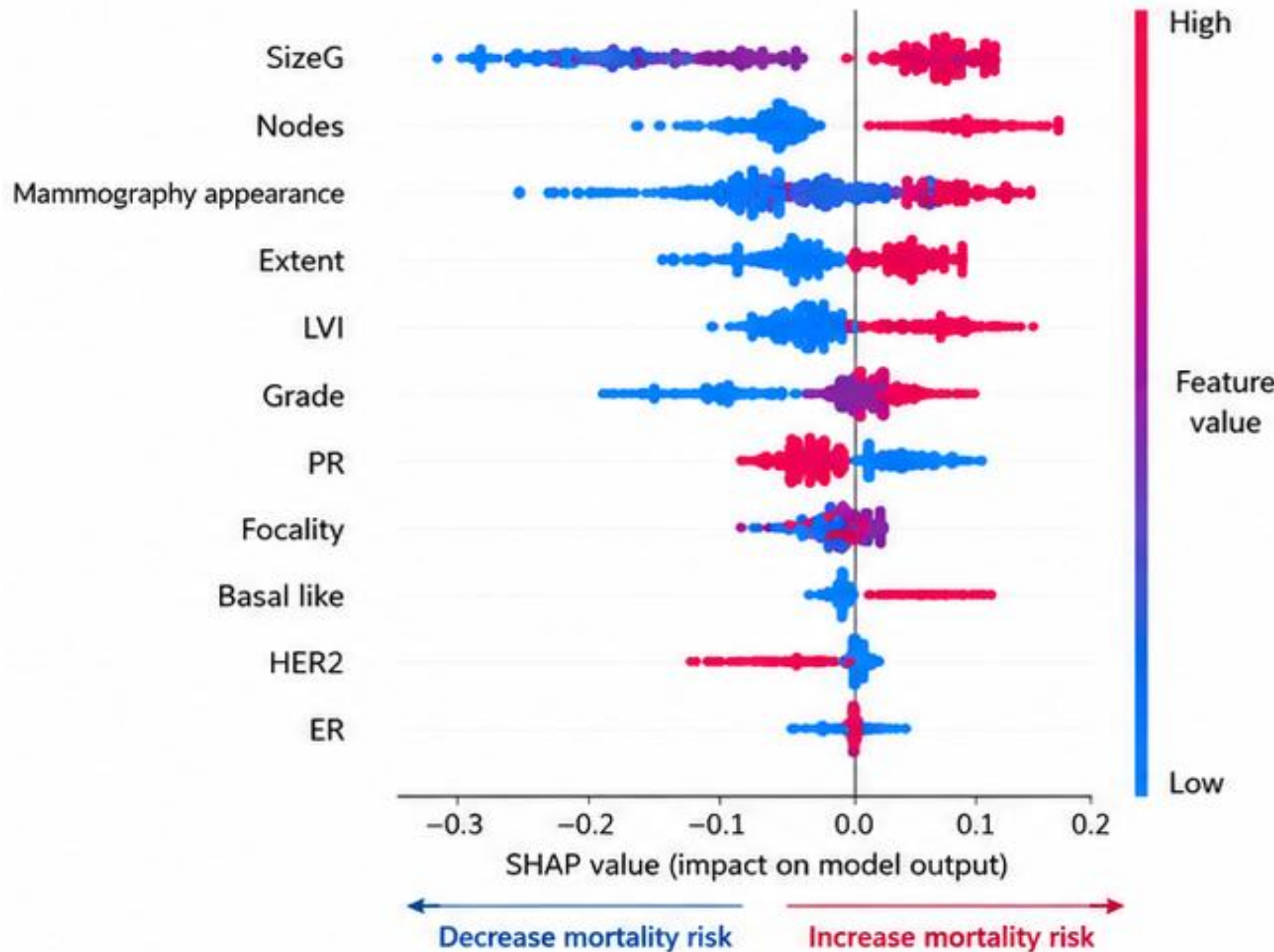
XAI helps us answer:
“Why was this patient predicted as high risk?”



Understanding Model Predictions with SHAP

Opening the Black Box of Random Forest and XGBoost

SHAP Summary Plot (Feature Impact on Breast Cancer Death Prediction)



How to Read This Plot

- 1 Importance Ranking**
Features are ranked by their overall impact on the model. Top features (top rows) are the most important.
- 2 Color (Feature Value)**
Red = High value
Blue = Low value
- 3 Direction (Impact)**
Right (positive SHAP value):
Increases mortality risk
Left (negative SHAP value):
Decreases mortality risk

Understanding Model Predictions with SHAP

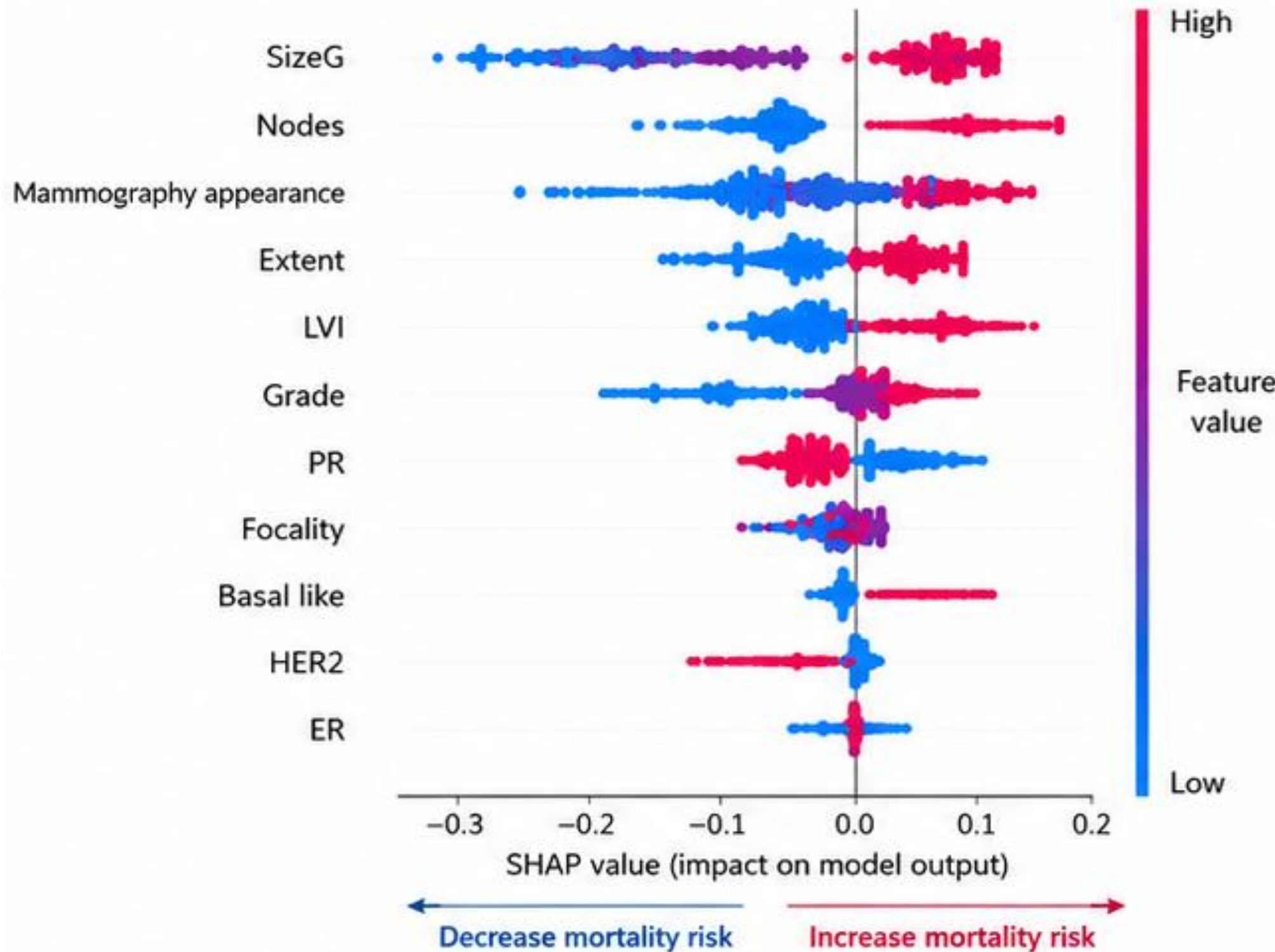
Opening the Black Box of Random Forest and XGBoost



Suppose a patient has a **larger tumor size, positive lymph nodes, and HER2-negative status.**

What would the predicted risk of breast cancer mortality be?

SHAP Summary Plot (Feature Impact on Breast Cancer Death Prediction)



How to Read This Plot

- 1 Importance Ranking**
Features are ranked by their overall impact on the model. Top features (top rows) are the most important.
- 2 Color (Feature Value)**
Red = High value
Blue = Low value
- 3 Direction (Impact)**
Right (positive SHAP value): **Increases** mortality risk
Left (negative SHAP value): **Decreases** mortality risk

Understanding Model Predictions with SHAP

Opening the Black Box of Random Forest and XGBoost

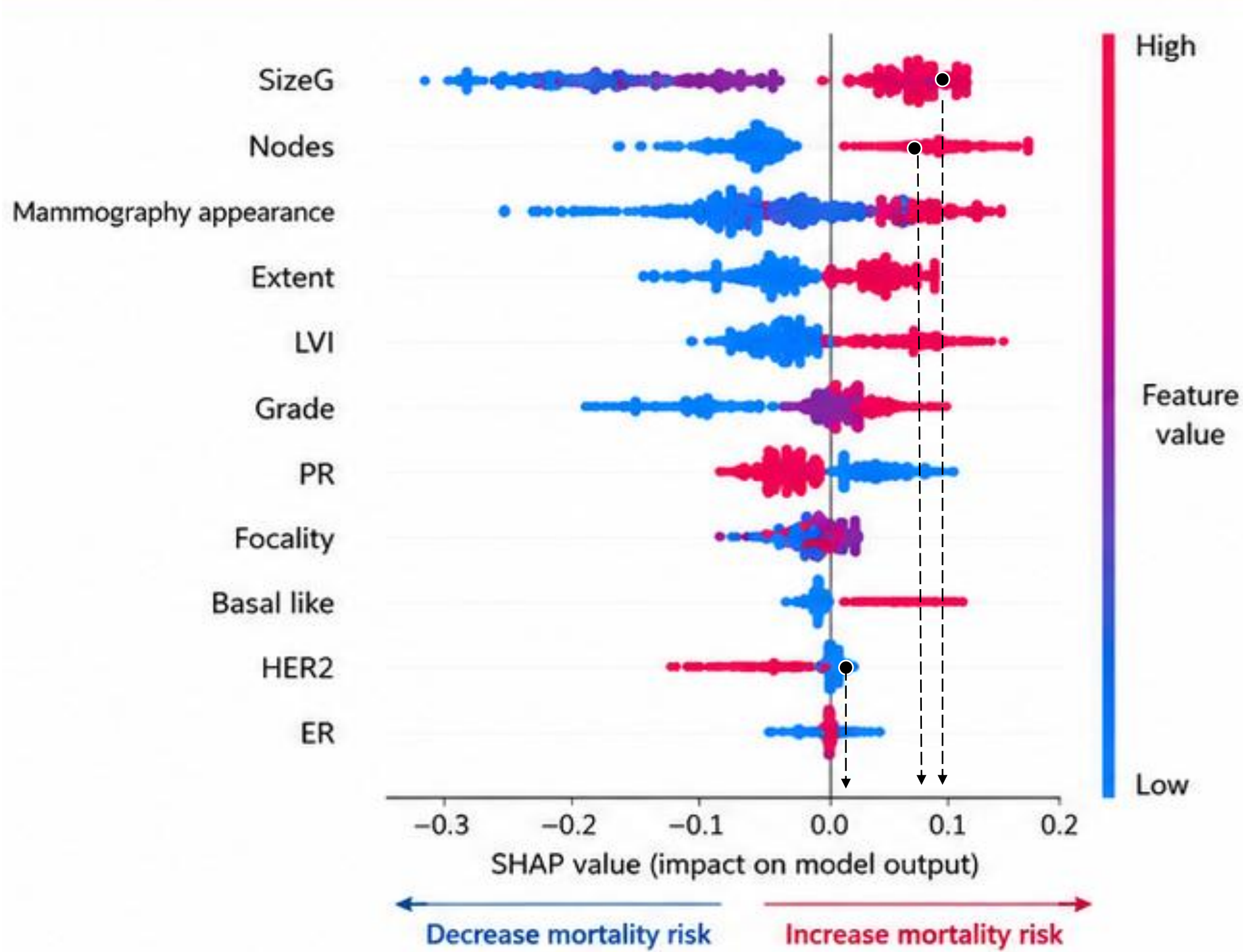


Suppose a patient has a **larger tumor size, positive lymph nodes, and HER2-negative status.**

What would the predicted risk of breast cancer mortality be?



SHAP Summary Plot (Feature Impact on Breast Cancer Death Prediction)



How to Read This Plot

1 Importance Ranking

Features are ranked by their overall impact on the model. Top features (top rows) are the most important.

2 Color (Feature Value)

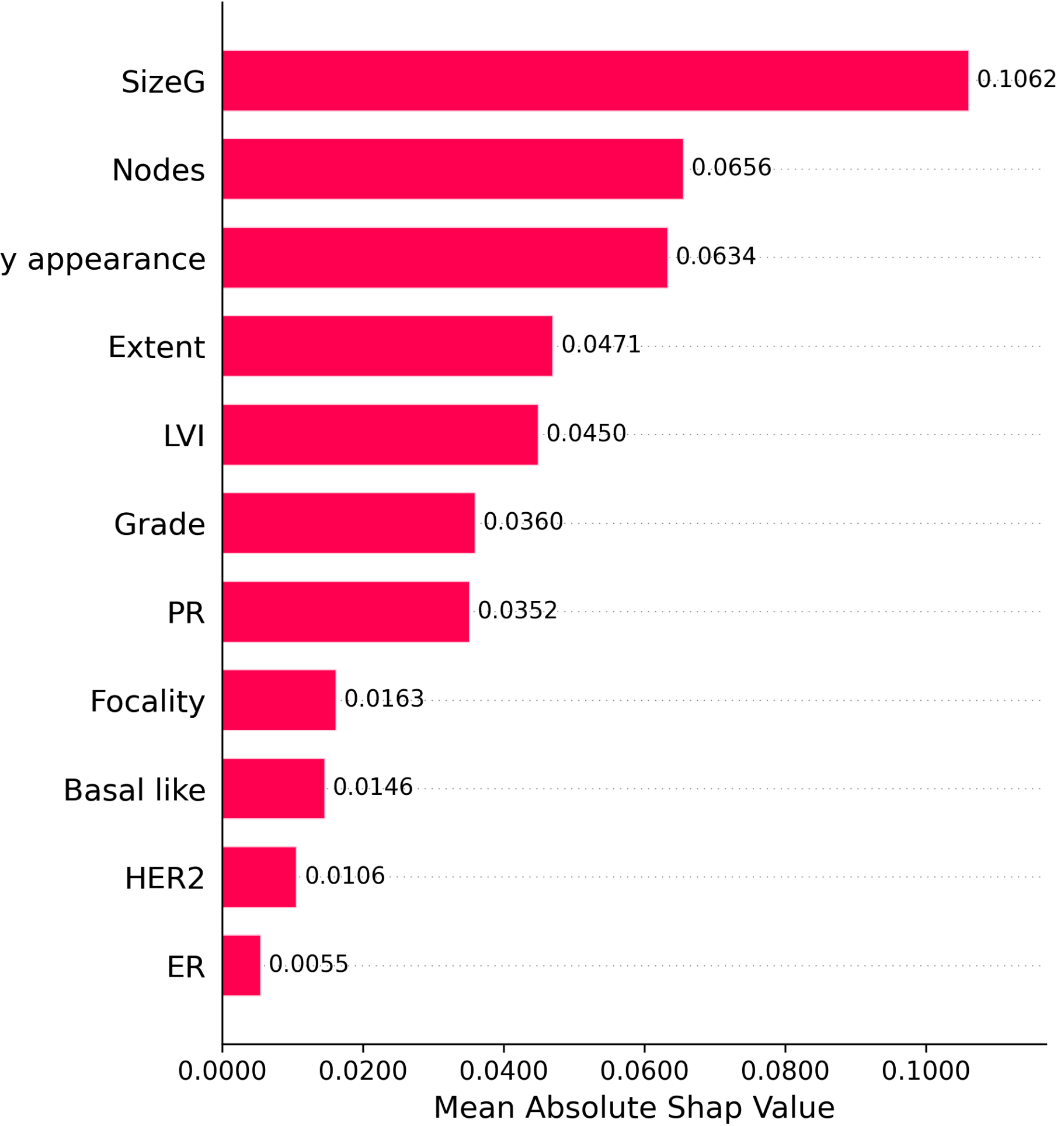
Red = High value
Blue = Low value

3 Direction (Impact)

Right (positive SHAP value):
Increases mortality risk
Left (negative SHAP value):
Decreases mortality risk

Global SHAP Importance Plot

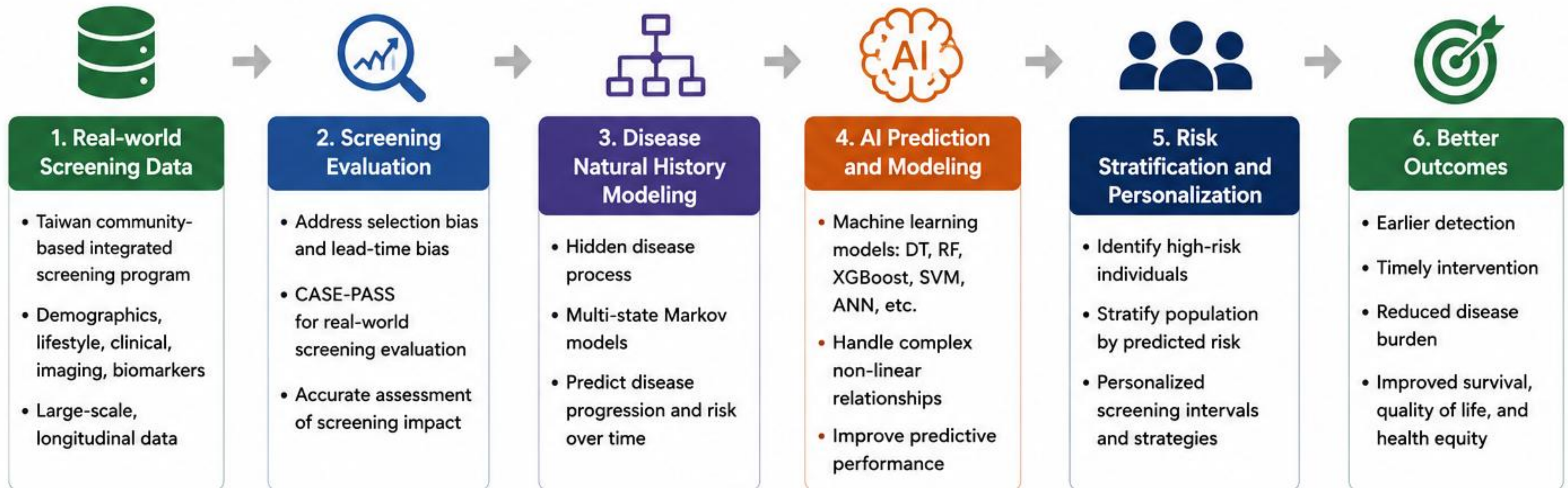
Features



AI-empowered Precision Screening

From Population Screening to Personalized Prevention

A **Data-driven** • **Model-informed** • **AI-powered** Framework for Better Outcomes



*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

II. Artificial Intelligence and Machine Learning for Precision Screening

Practice: Machine Learning Approaches for Differentiating Screen-Detected and Clinically Detected Cases

Dr. Chiu-Wen Su
National Taiwan University



IACCS 2026

Real-World Data :

Taiwan Community-based Integrated Screening Program



Dynamic data

Characteristics

*Age, Gender, Education,
Family History*

Dietary Habits

*Meat/Vegetable
/Bean & Egg
/Milk/ Coffee/ Fruit*

Life Style

*Smoking/
Betel quid chewing/
Alcohol Screening/
Awareness/ Health check*

Metabolic syndrome factors

*Waist/ TG/ HDL/ Blood
Pressure/ Glucose*

Biomarkers

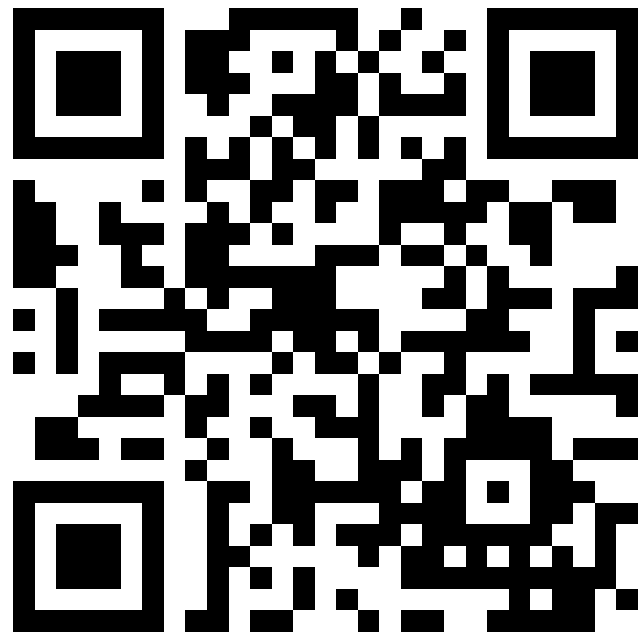
*Fecal hemoglobin
concentration (FIT)/ Uric
Acid/ Creatinine*

Who is likely be detected from CRC screening?

▼ Colorectal Cancer Dataset

	Serial_no	Fitgp	Gender	entry_agegp	Education	CRC_familyhistory	Smoke	Alcohol	Betel	Exercise	Vegetable
0	1	6	1	2	3	0	0	0	0	1	2
1	2	1	2	3	1	0	0	0	0	1	2
2	3	4	1	2	2	1	2	0	0	0	1
3	4	0	2	3	1	0	0	0	0	0	1
4	5	0	2	4	1	0	0	0	0	1	1
5	6	5	2	1	1	0	0	0	0	0	2
6	7	1	2	1	3	0	0	0	0	1	3
7	8	0	1	1	1	0	2	0	0	1	2
8	9	3	1	3	3	0	0	0	0	1	3

Who is likely be detected from CRC screening?



Select Multistate Outcomes

Normal → Adenoma → Cancer

Screen Detected
Interval Cancer

Sample Fraction of Mass Screening Dataset

Number of Features per Split (Integer >= 10)

Number of Features-Related Pathways (Integer >= 10)

Max Depth of Pathways (Integer >= 1 or None)

Min Screenes for Node Split Minimum Screenes to split a node (Integer >= 2)

Min Screenes for Node Leaf Minimum Screenes in a leaf node (Integer >= 1)

Cross-Validation Number of folds for cross-validation.

Correlation
Reduces correlation among trees

Strength
Controls the strength of individual trees

Validation
Helps prevent overfitting and ensure robustness

Random Forest Classification

Random Forest Model for Screen-detected CRC

Identifying individuals most likely to be detected through CRC screening



Outcome

Screen-detected CRC
vs Normal

Continuous

- ☐ Age
- ☒ Induction time

Categorical

- ☒ f-Hb level
- ☒ Gender
- ☒ Age group

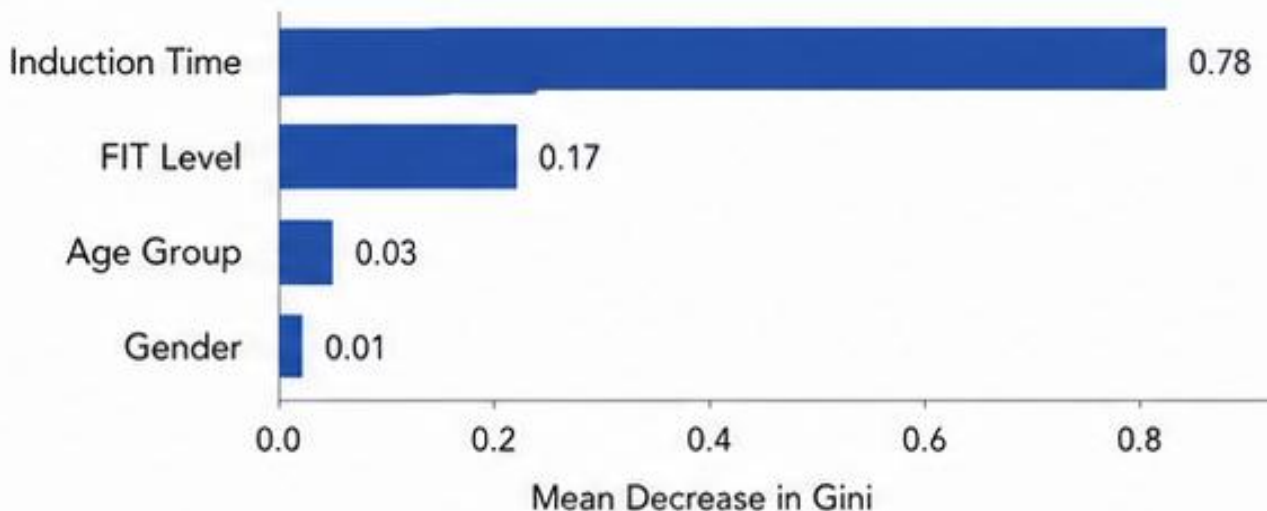


Model Hyperparameters

Max Depth: 4
Min Samples Split: 5
Min Samples Leaf: 2
Criterion: Gini
Number of Trees: 100



Top Predictors (Variable Importance)

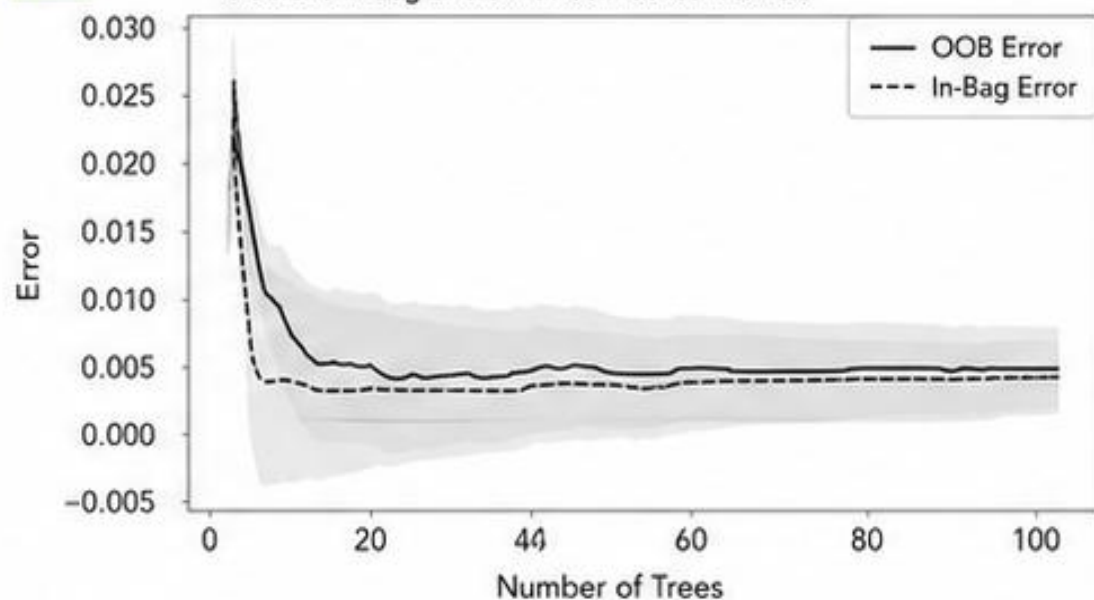


→ FIT level and screening history (induction time)
were the strongest predictors of screen-detected CRC.



Model Stability

OOB and In-Bag Errors with Confidence Intervals



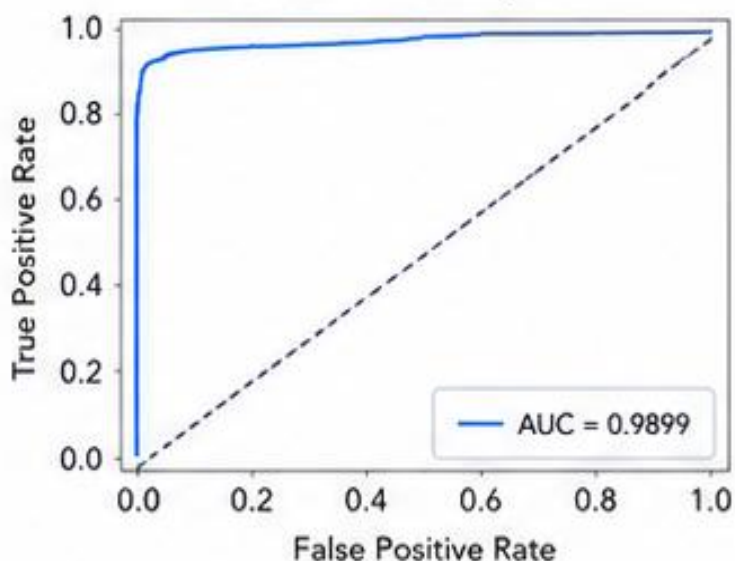
→ OOB error stabilized after ~20 trees.

Model Performance

Training Set

Metric	Value
Precision	0.998
Recall (Sensitivity)	0.996
F1-score	0.996
Specificity	0.996
AUC (ROC)	0.9899

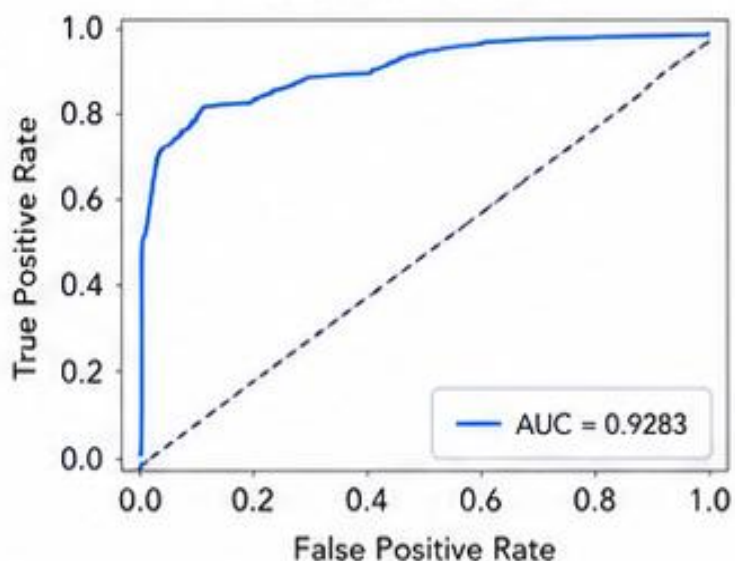
ROC Curve – Training Data



Test Set

Metric	Value
Precision	0.897
Recall (Sensitivity)	0.788
F1-score	0.825
Specificity	0.788
AUC (ROC)	0.9283

ROC Curve – Test Data



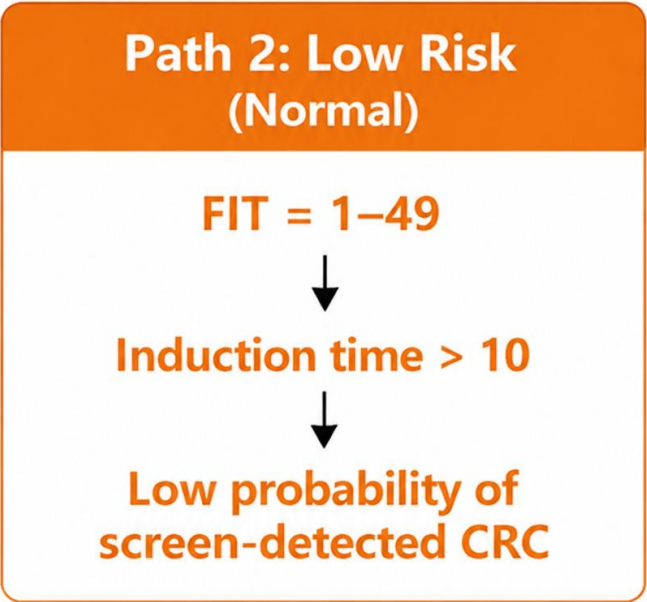
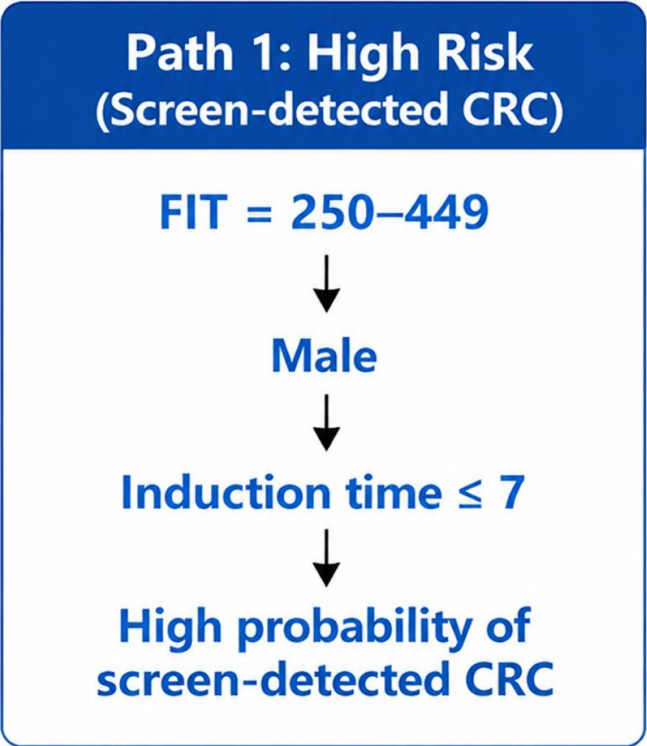
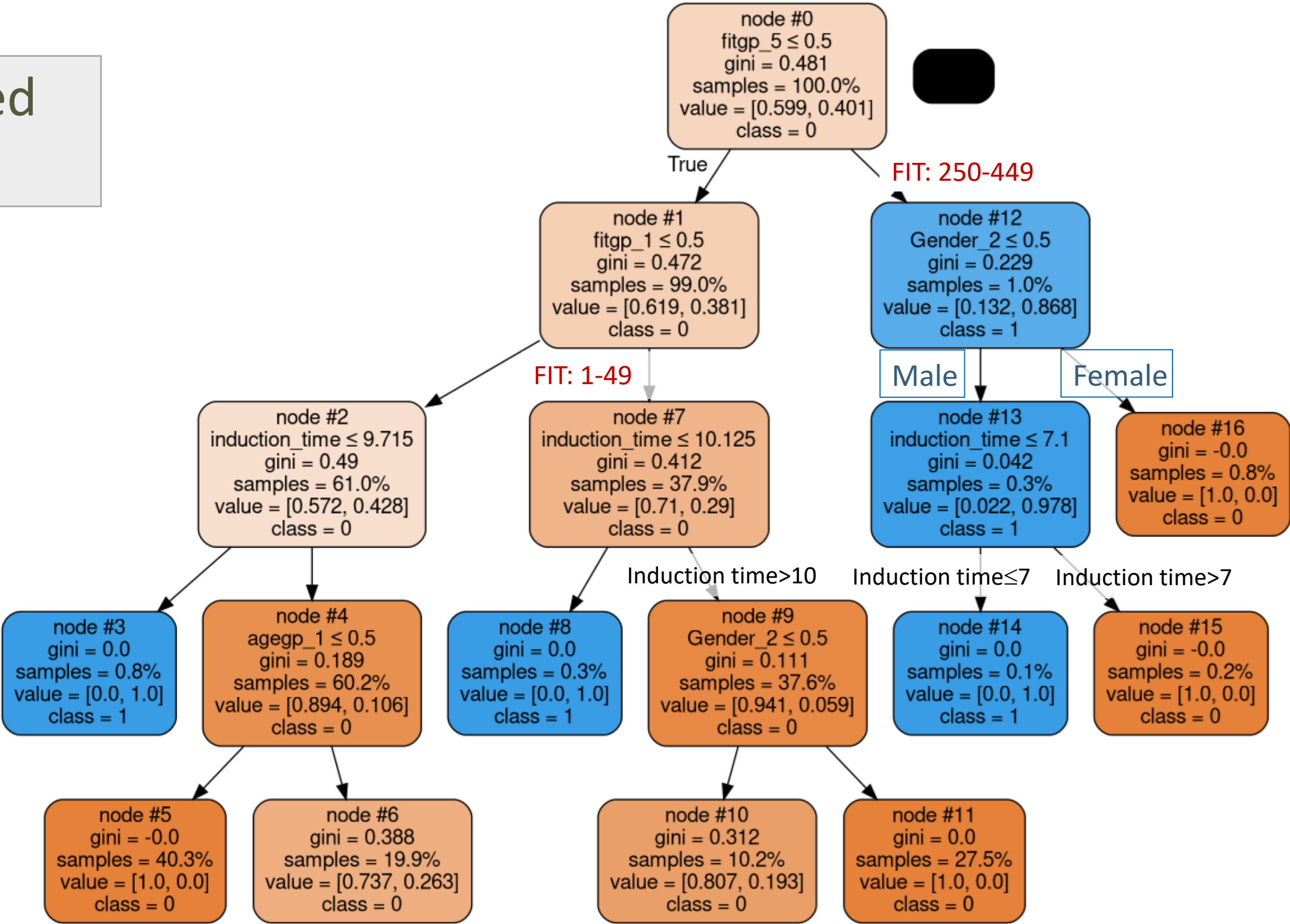
How Does the Model Make Decisions?

Example: One Decision Tree from the Random Forest

Screen-detected
vs Normal

FITGP

- 0: Undetected
- 1: 1-49
- 2: 50-99
- 3: 100-149
- 4: 150-249
- 5: 250-449**
- 6: >450



RF1. Which AUC should we use to estimate real-world performance?

A. Training AUC = 0.9899

B. Test AUC = 0.9283



RF2. If the training AUC is 0.98 and the testing AUC is 0.72, does this indicate overfitting?

A. Yes

B. No



Who is likely be detected from CRC screening?



Artificial Neural Network Analysis

Epochs:

Batch Size:

Hidden Layers:

Neurons per Layer:

Activation Function: Select the activation function to use.

Learning Rate: Default is 0.001

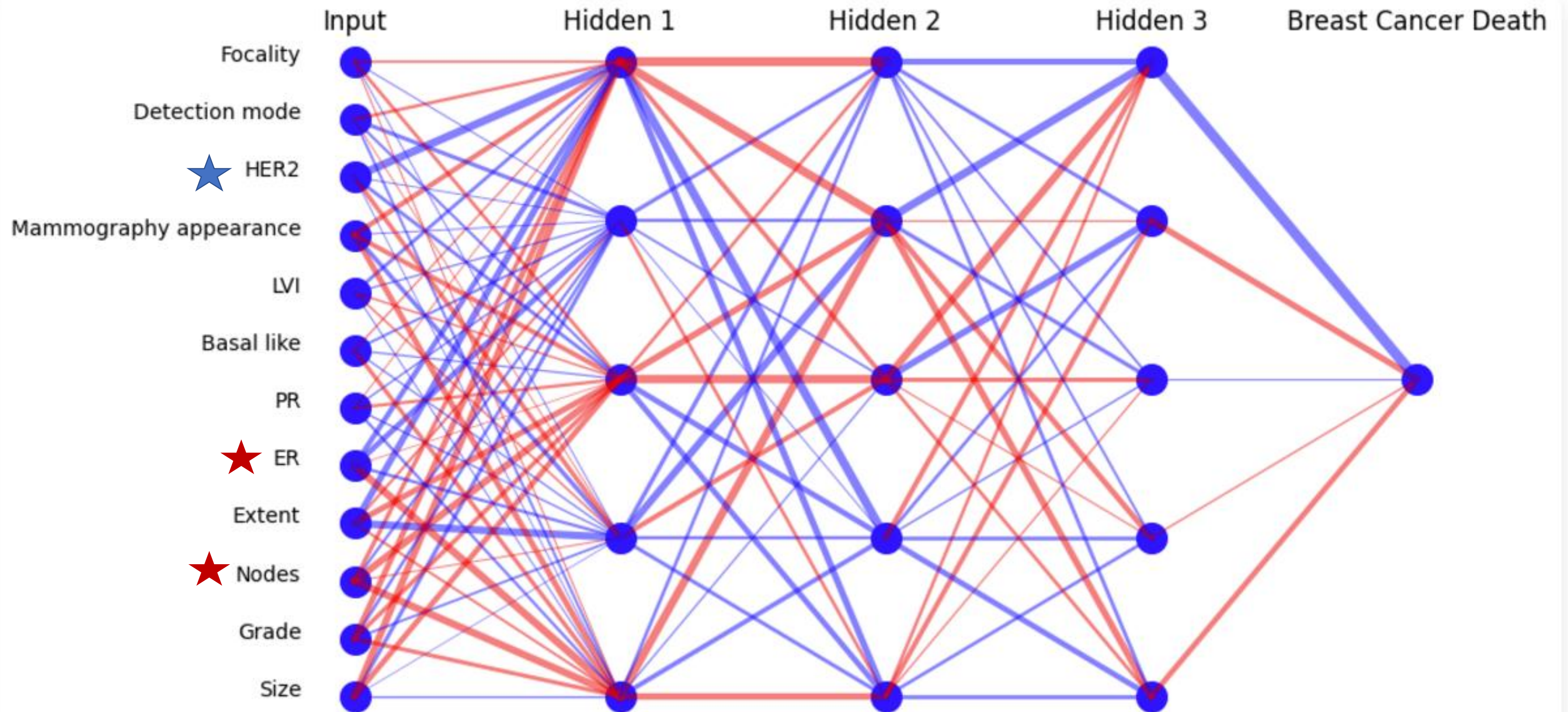
Random Seed: Seed value for random number generators to ensure reproducibility.

Run

Clear Value

Artificial Neural Network Analysis

ANN Structure



ANN-1. What is the predicted probability of breast cancer under the following conditions (ANN)? *Please enter a 3-digit number.*

size	Grade	Nodes	Extent	ER	PR	Basal like	LVI	Mammography appearance	HER2	Detection mode	Focality	BCD	Predicted BCD	Predicted Probabilities Death	ID
60	Grade 3	Positive	Positive	Positive	Negative	Negative	Positive	Architecture distortion	Negative	Interval cancer	Diffuse	1			520



ANN-1. What is the predicted probability of breast cancer under the following conditions (ANN)? *Please enter a 3-digit number.*

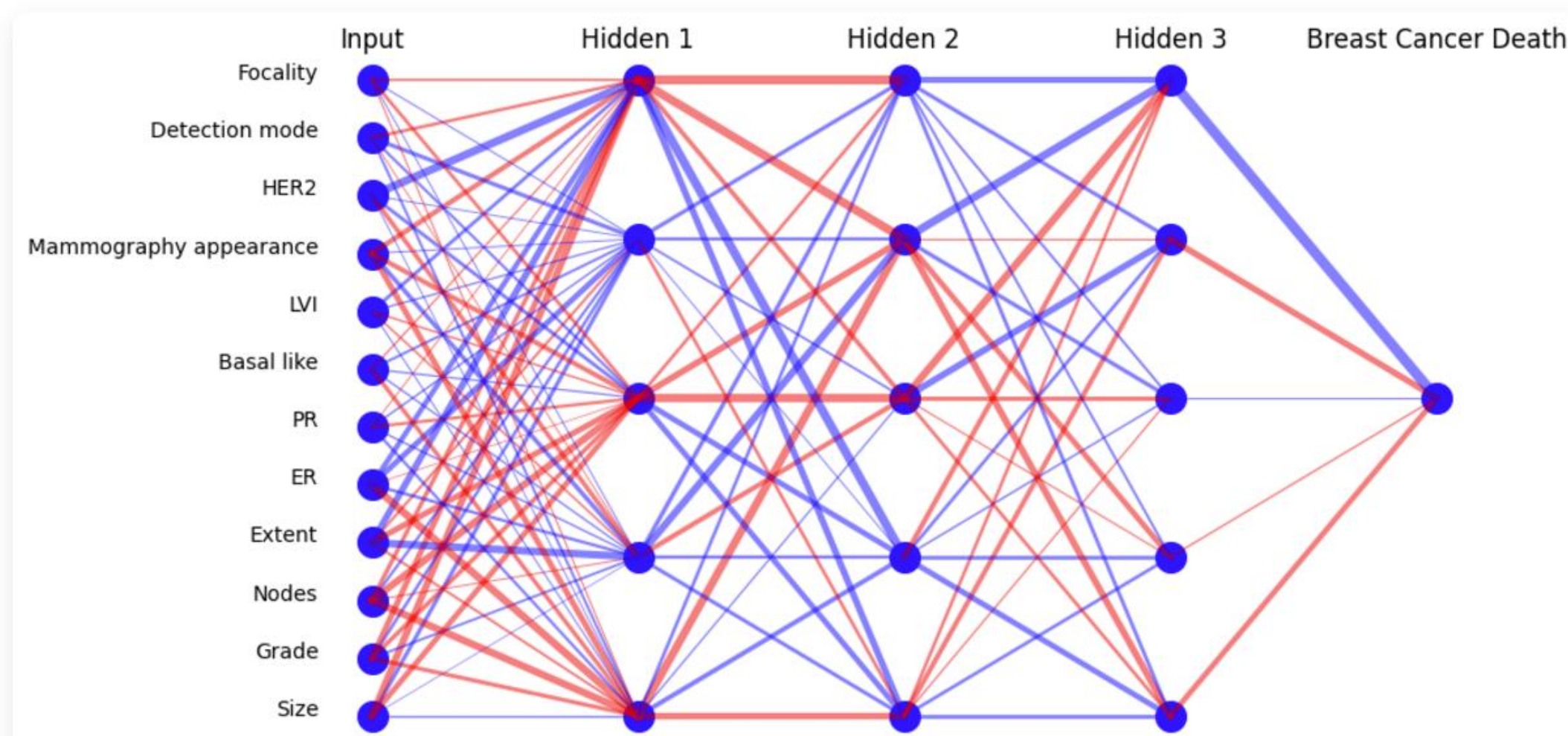
size	Grade	Nodes	Extent	ER	PR	Basal like	LVI	Mammography appearance	HER2	Detection mode	Focality	BCD	Predicted BCD	Predicted Probabilities Death	ID
60	Grade 3	Positive	Positive	Positive	Negative	Negative	Positive	Architecture distortion	Negative	Interval cancer	Diffuse	1	1	0.929	520

ANN-2. How many hidden layers does this ANN have in this example?

A. 3

B. 5

ANN Structure



*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

III. Digital Twin AI for Precision Screening

M-state demonstration



*AI-empowered Precision Disease Screening of Population-based
Organized Service Program*

III. Digital Twin AI for Precision Screening

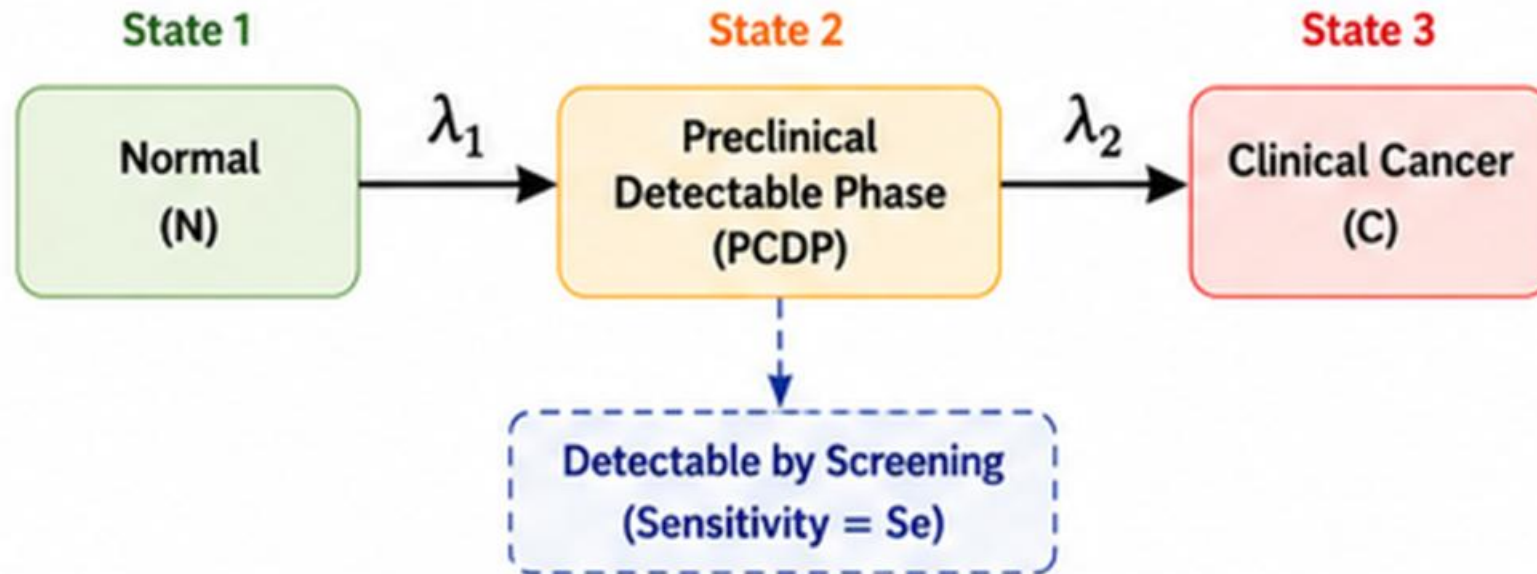
M-state demonstration



From population-average progression to individualized disease trajectories

Three-state model without covariates (Homogeneous)

Assumes all individuals share the same disease progression process.



1. Intensity Matrix

$$Q = \begin{pmatrix} -\lambda_1(t) & \lambda_1(t) & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 \end{pmatrix}$$

Defines transition rates between disease states.

2. Transition Probability Matrix

$$P(t) = \begin{pmatrix} P_{11}(t) & P_{12}(t) & P_{13}(t) \\ 0 & P_{22}(t) & P_{23}(t) \\ 0 & 0 & P_{33}(t) \end{pmatrix}$$

Describes the probability of moving between states over time.

3. Likelihood Function

$$L = \prod_{j=1}^3 \prod_{m=1}^{n_j} P_{1j}^+(t_m)^{\delta_{mj}}$$

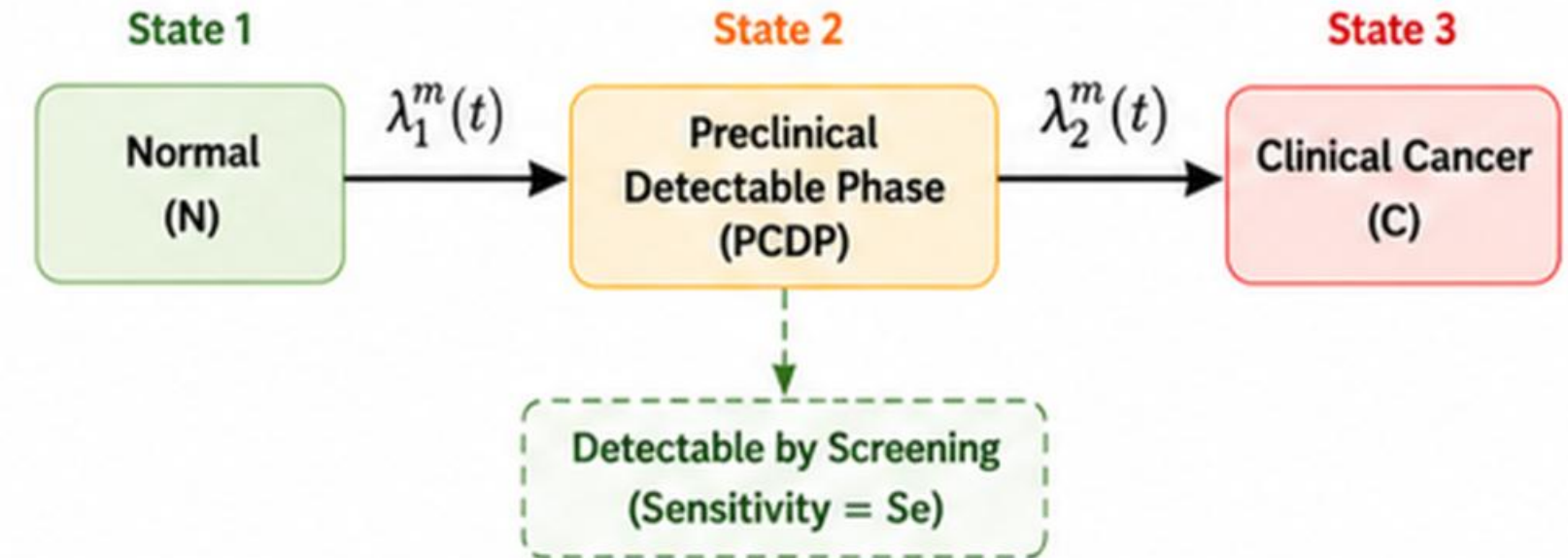
Used to estimate model parameters from observed screening data.

Key parameters (same for everyone):

λ_1 (disease onset rate), λ_2 (disease progression rate), Se (screening sensitivity)

Three-state model with covariates (Non-homogeneous)

Transition rates depend on individual characteristics.



Individual Covariates

- Gender
- Age
- Smoking status
- Family history
- Genetic factors
- Other risk factors
- ...

Covariate-dependent transition rates

$$\lambda_1^m(t) \text{ Disease onset rate for individual } m$$

$$\lambda_2^m(t) \text{ Disease progression rate for individual } m$$

Different disease trajectories across individuals

Example (Gender)

$$\lambda_1^m(t) = \lambda_{10} \exp(\beta_{10} \text{ Gender}) \gamma_1 t^{\gamma_1 - 1}$$

$$\lambda_2^m = \lambda_2 \exp(\beta_2 \text{ Gender})$$

(Similar forms for other covariates)

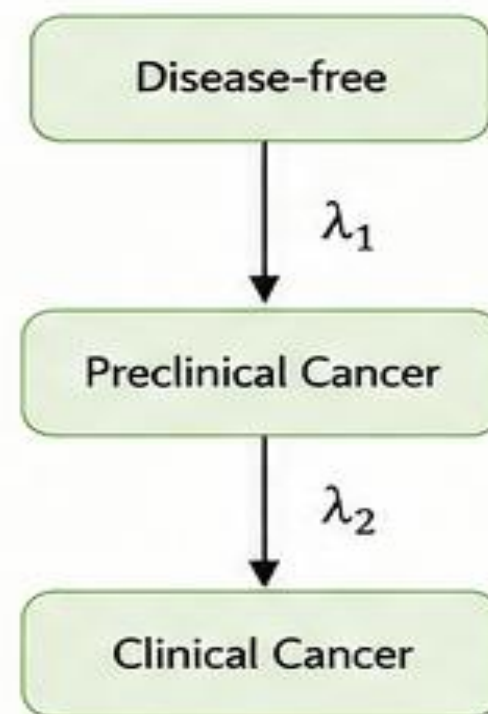
m : individual index ($m = 1, 2, \dots, M$)

Estimation of Natural History Parameters of Breast Cancer

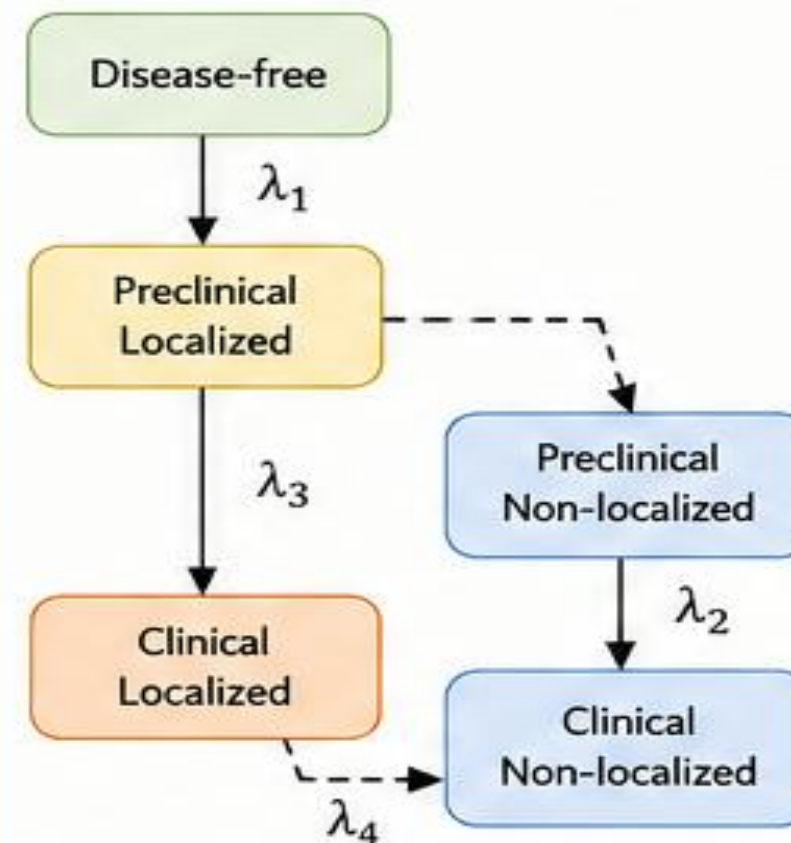
Based on Non-randomized Organized Screening Data

Multi-State Markov Models

Three-State Model



Five-State Model (by stage)



Parameters Estimated

- ✓ Transition probability ($\lambda_1, \lambda_2, \lambda_3, \lambda_4$)
- ✓ Mean sojourn time (MST)
- ✓ Screening sensitivity

Simulating Screening Strategies

Model inputs



Screening Interval

- Annual
- Every 2 years
- Every 3 years



Screening Sensitivity

- 68%
- 80%
- 90%



Attendance Rate

- 30%
- 60%
- 90%

Model outputs



Advanced Breast Cancer Reduction



Mortality Reduction

Multi-state Disease Prediction System

