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

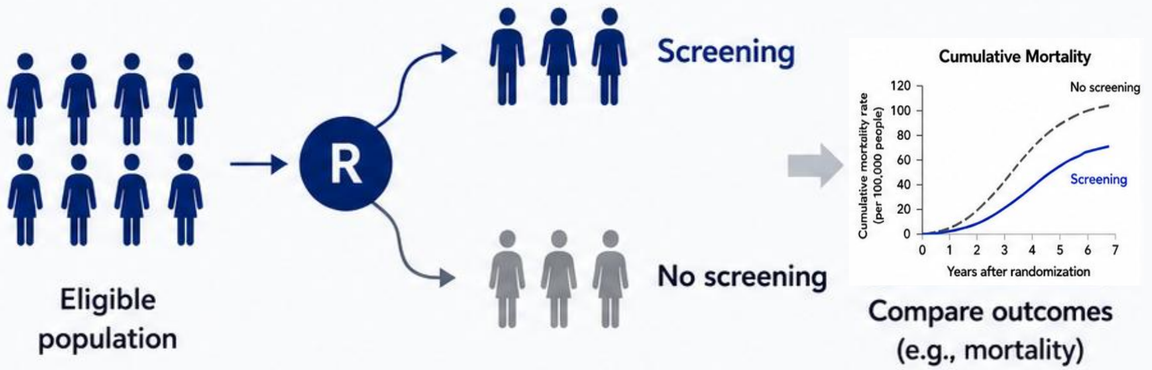
Dr. Chiu-Wen Su
National Taiwan University

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)



Compare outcomes (e.g., mortality)

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.

Key Question

How can we evaluate screening effectiveness in the real world when a randomized control group no longer exists?

Limitations: Traditional Approaches Are Subject to Major Biases

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

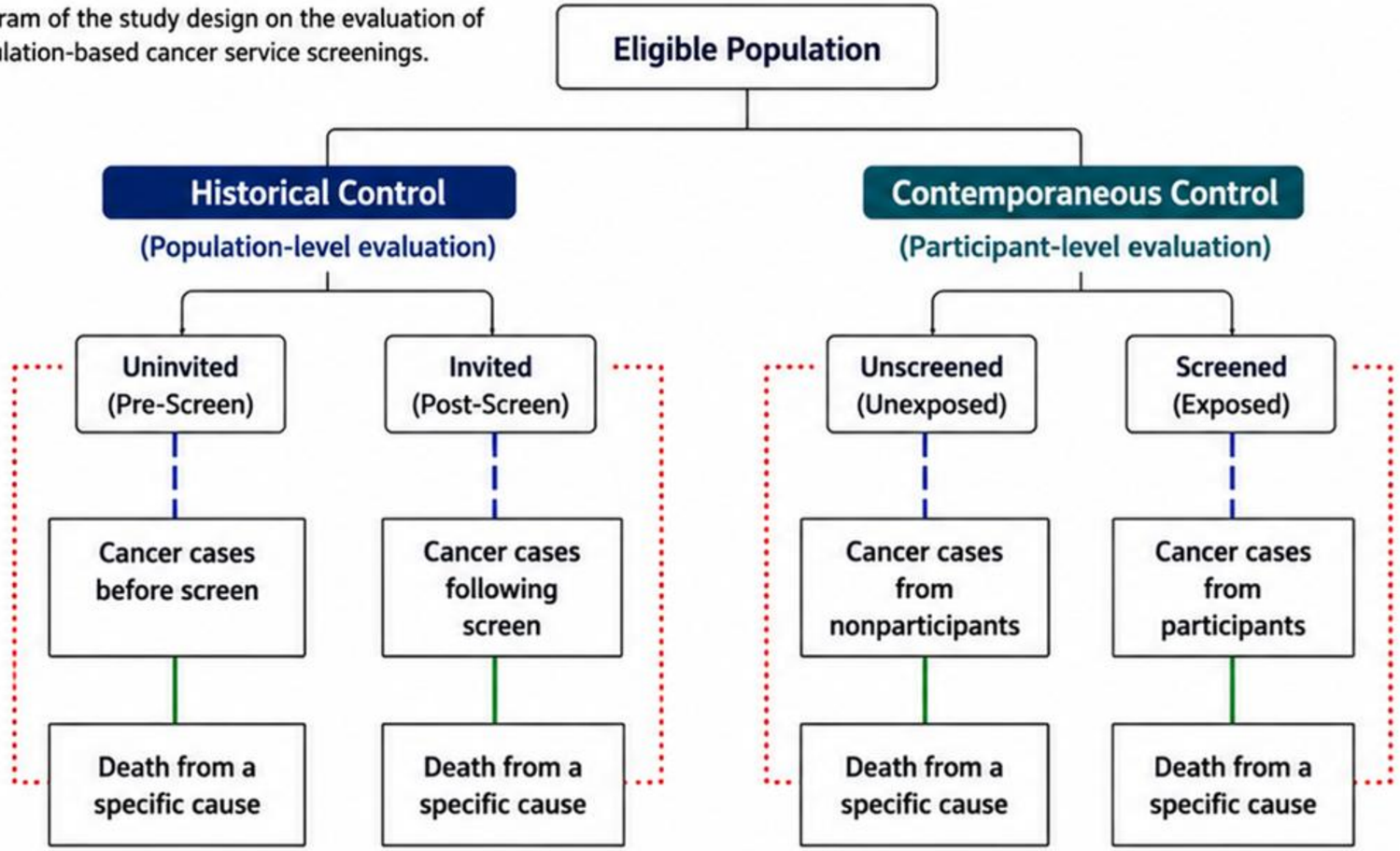
Length Bias
Slow-growing tumors are more likely to be detected by screening.

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.



OUTCOMES EVALUATED

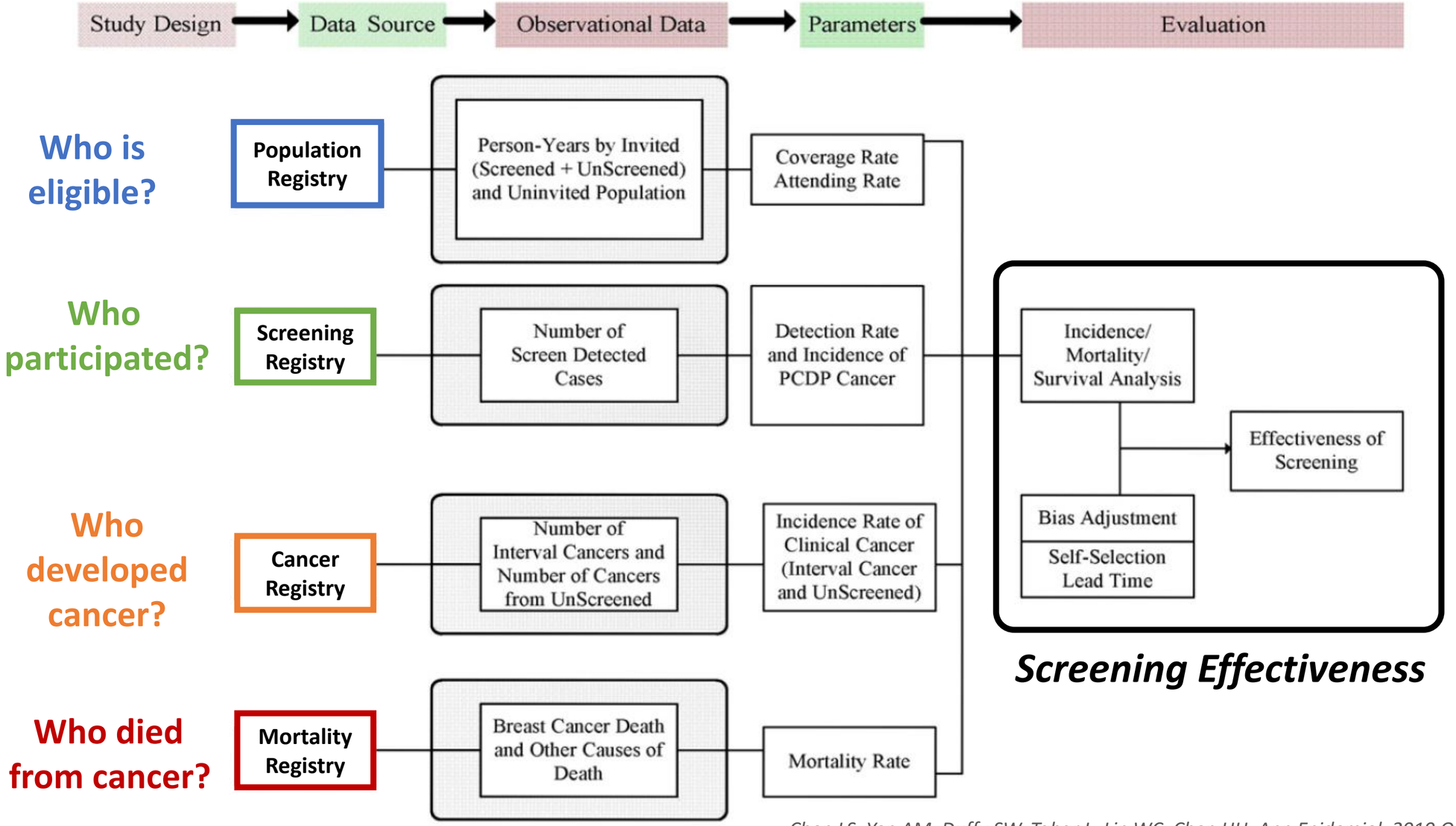
Incidence
Cancer occurrence

Mortality
Cancer death

Survival
Prognosis after diagnosis

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

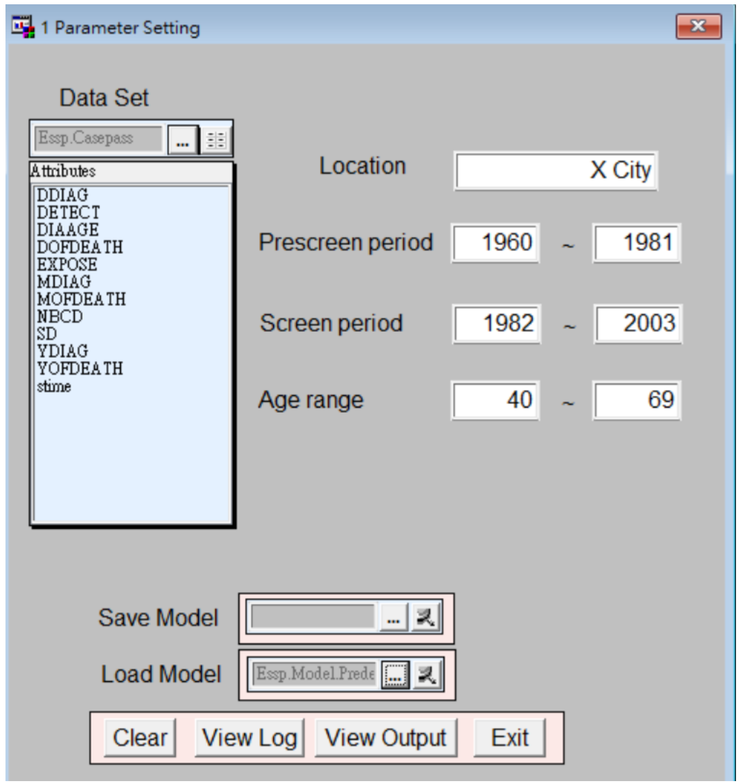


Chen LS, Yen AM, Duffy SW, Tabar L, Lin WC, Chen HH. Ann Epidemiol. 2010 Oct;20(10):786-96.

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)



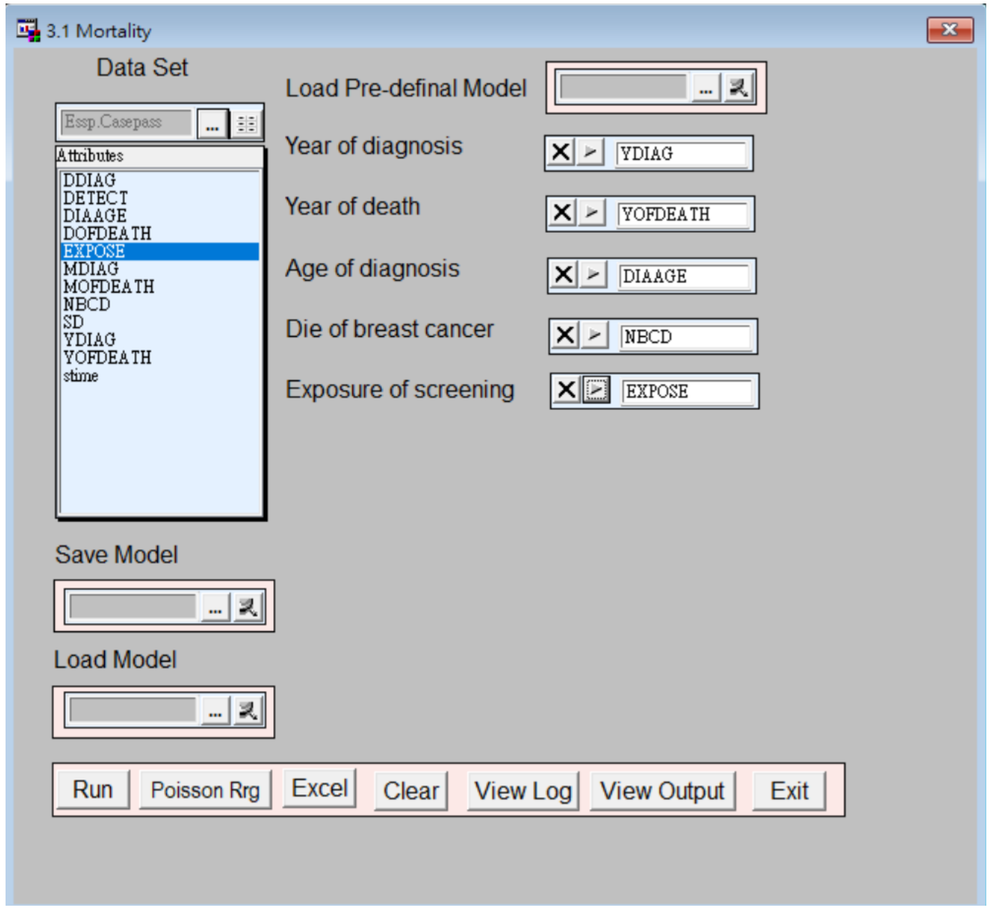
Step 1. Construct Exposed and Unexposed Person-Years

	Year	Population in year	Person-years exposed to screening	Total person-years exposed for each period	Total person-years unexposed for each period
Prescreen 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		
Screen period	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)



Step 2. Estimate Mortality in Each Period

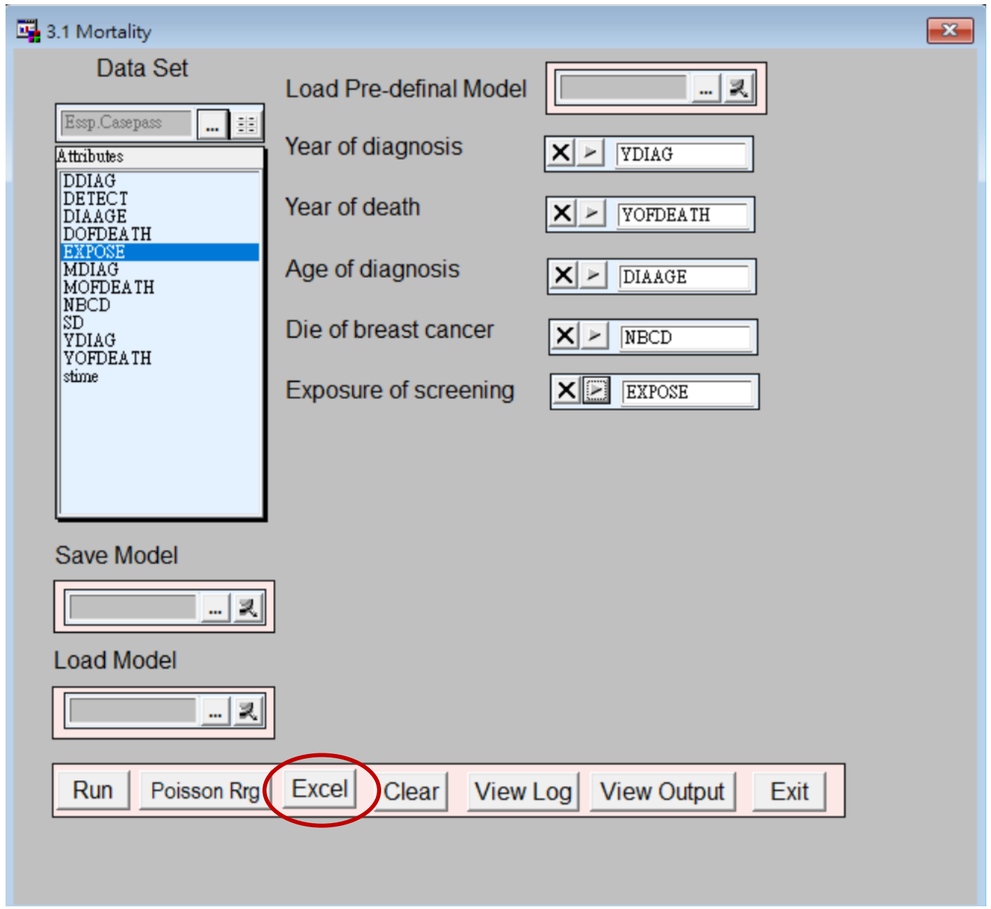
VIEWTABLE: Essp.Casepass

	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

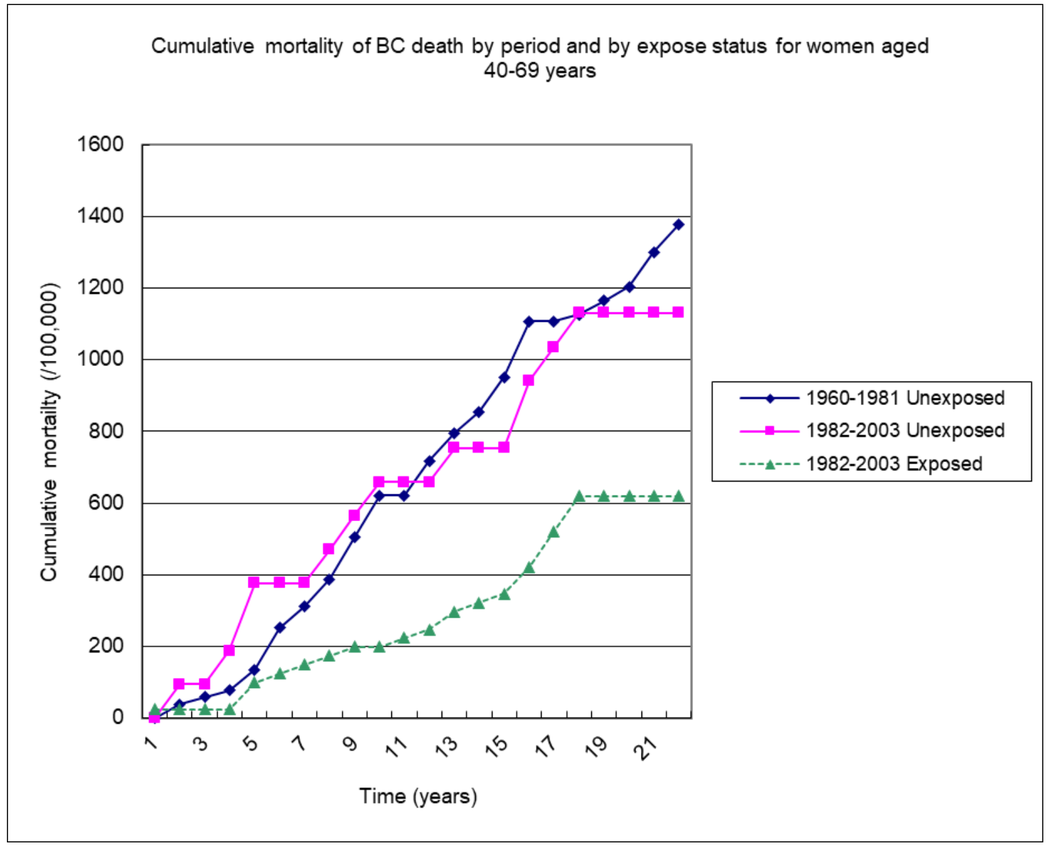
CASE-PASS Analytical Platform – How Selection Bias is Adjusted

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)



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 = $Mortality_{Screened}$ / $Mortality_{Uninvited}$

2. RR_2 : Nonparticipants vs. Uninvited (Baseline)

   RR_2 = $Mortality_{Nonparticipants}$ / $Mortality_{Uninvited}$

3. AR : Attendance Rate

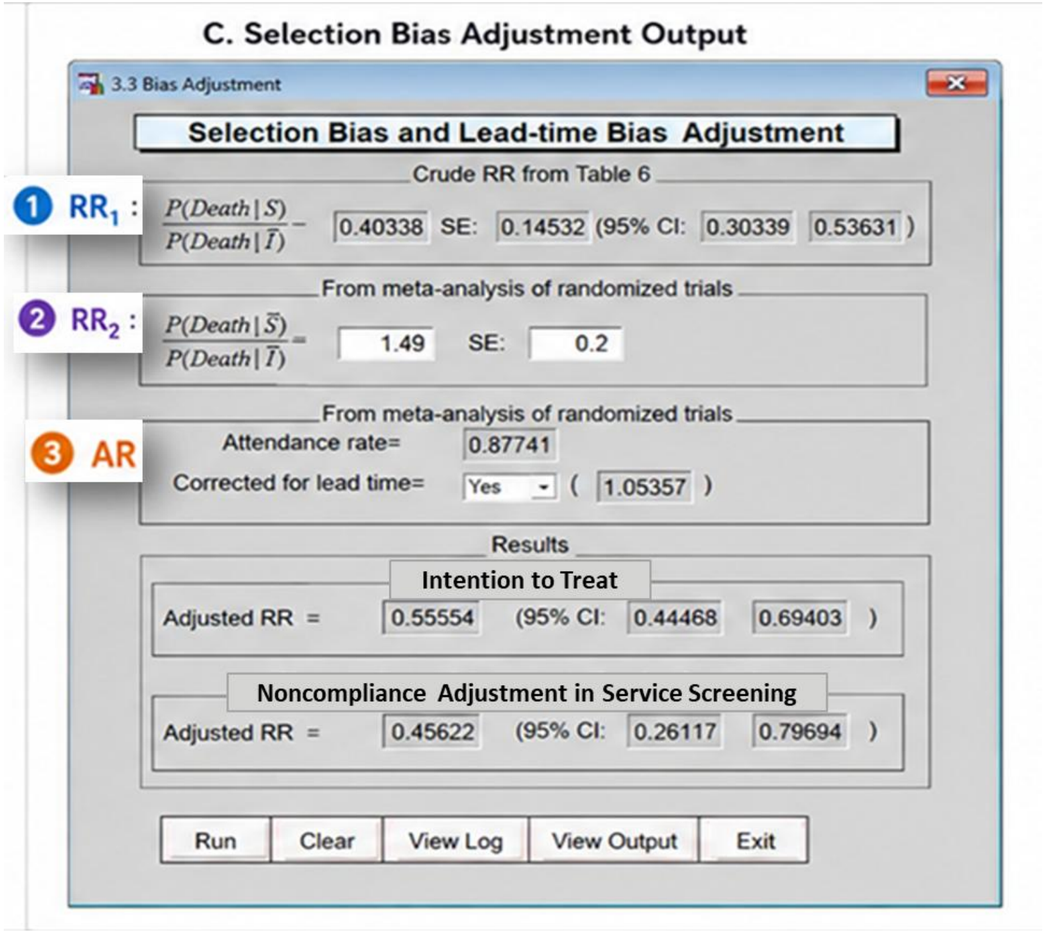
   AR = Number Attending / Number Invited

CASE-PASS Adjusted Relative Risk

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

This adjusts for self-selection bias.

C. Selection Bias Adjustment Output



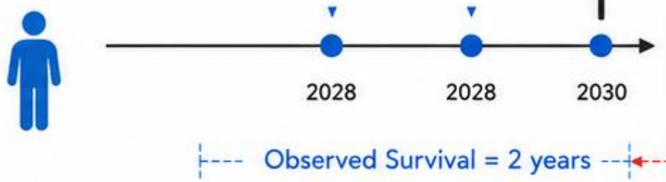
Chen LS, Yen AM, Duffy SW, Tabar L, Lin WC, Chen HH. Ann Epidemiol. 2010 Oct;20(10):786-96.

Lead-time Bias in Cancer Screening

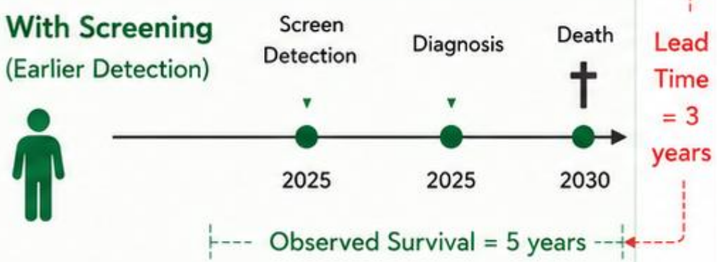
Screening detects cancer earlier without changing the time of death, creating an apparent survival advantage.

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)

True Survival Benefit (what screening actually provides) + Lead Time (earlier diagnosis) = Observed Survival Time

Adjusted Survival *(Removes Lead Time)*

Observed Survival Time − Lead Time (earlier diagnosis) = Adjusted Survival Time (true benefit)

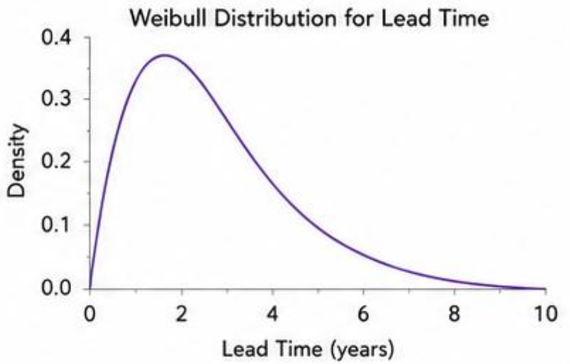
Illustrative Example

Observed Survival (with screening) = 5 years − Lead Time 3 years = Adjusted Survival (true benefit) = 2 years

How CASE-PASS Estimates Lead Time

Assume Lead Time Follows a Weibull Distribution

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

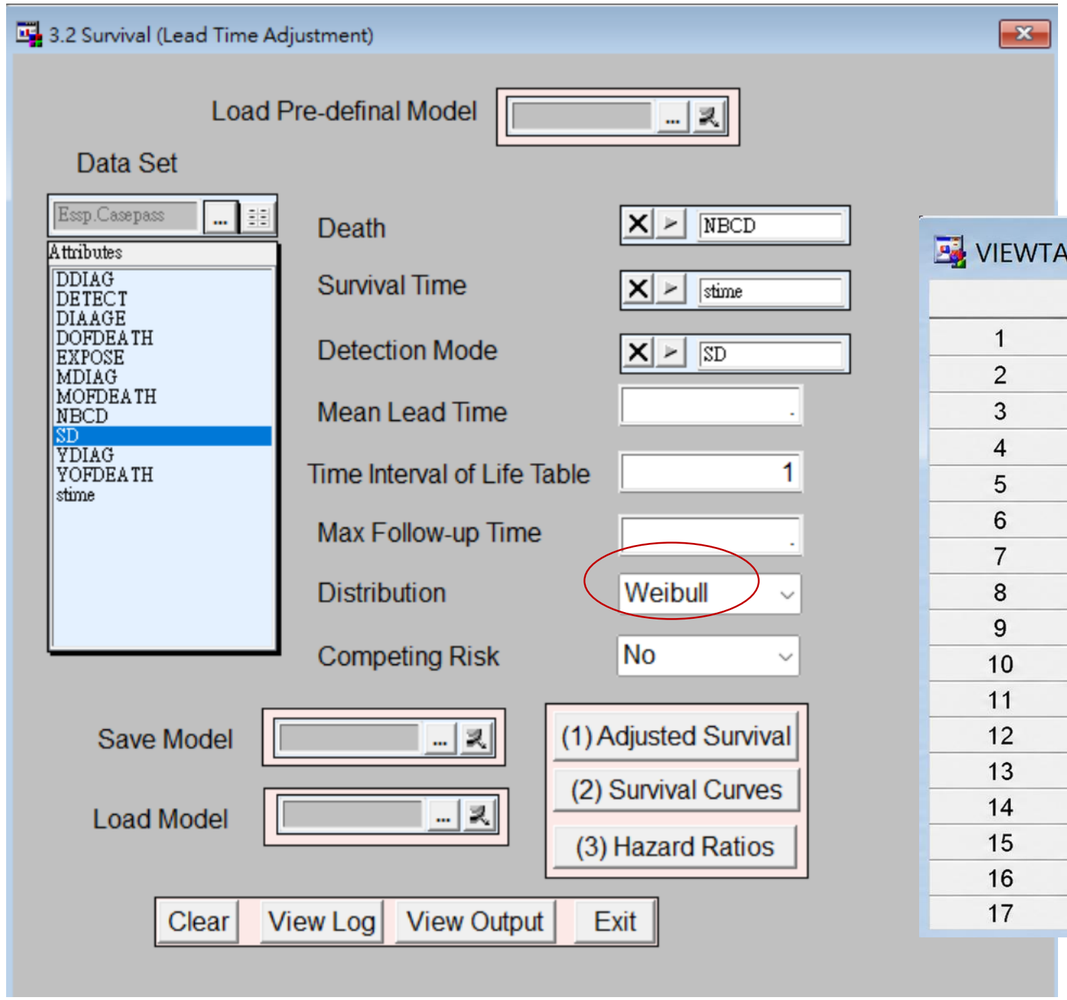


Lead time is estimated from observed data using a Weibull distribution.

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

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

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



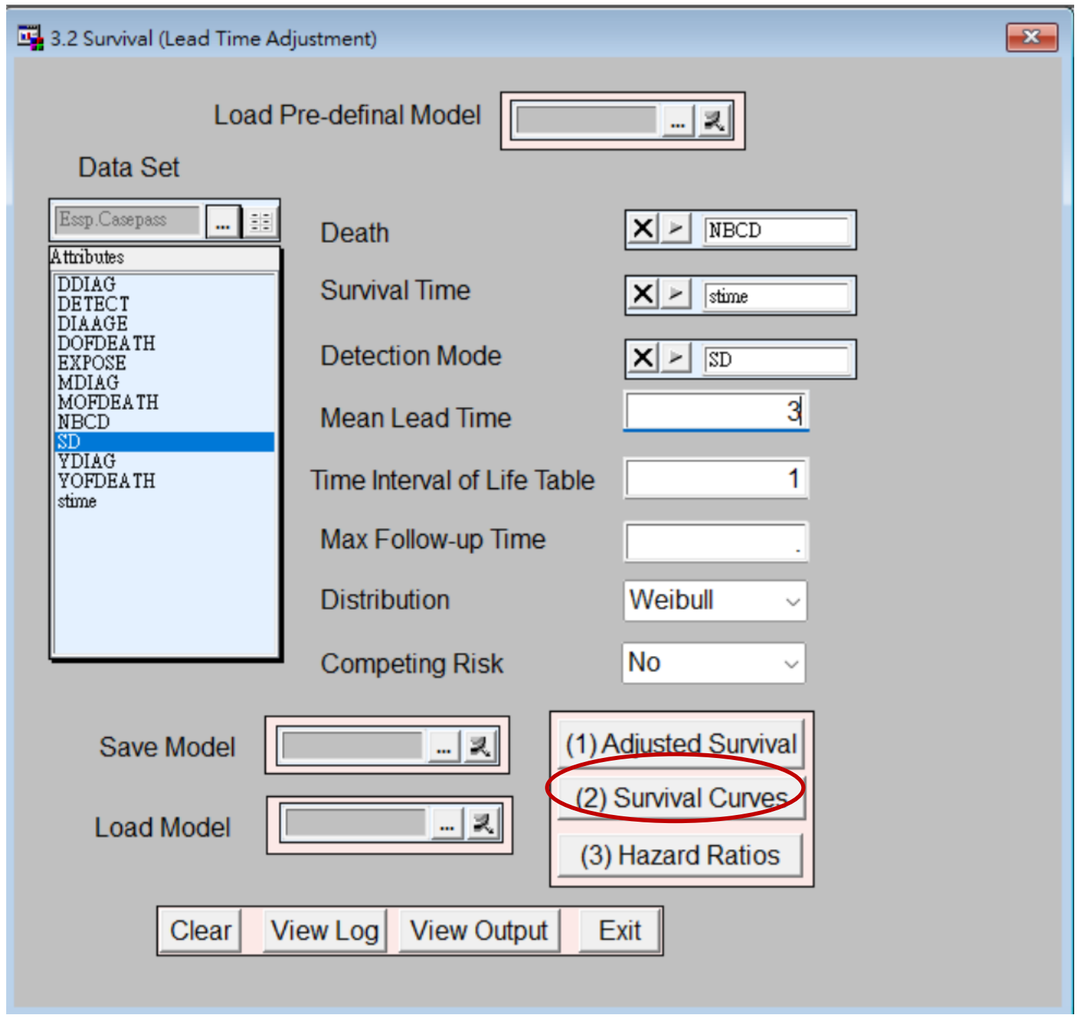
Calculate breast cancer survival time among women aged 40–69 years by screening detected and clinical detected.

VIEWTABLE: Essp.Casepass

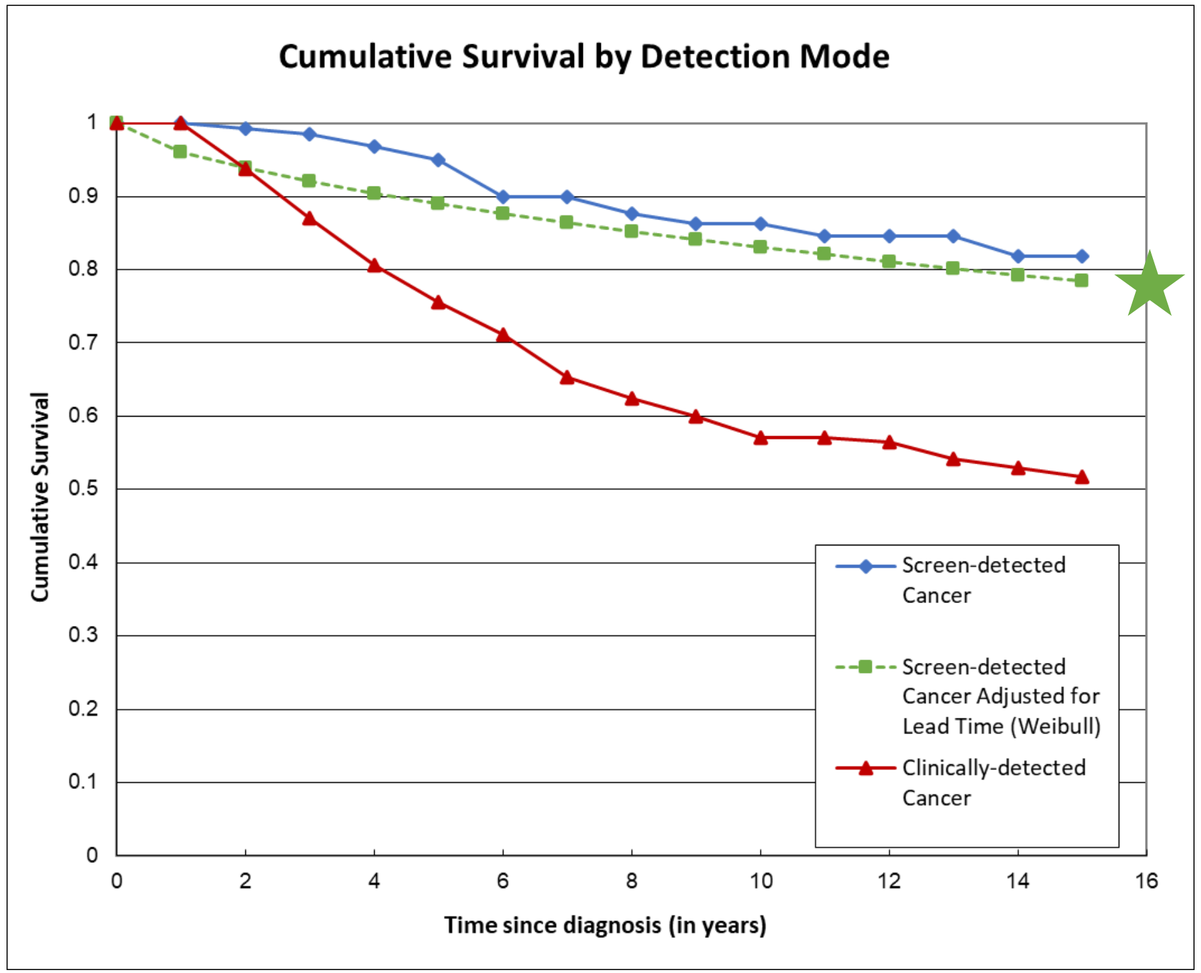
	YDIAG	YOFDEATH	DIAAGE	DETECT	NBCD	EXPOSE	stime	SD
1	1961	1984	068	15	0	0	23.05270363	0
2	1962	1982	068	15	0	0	20.85147159	0
3	1963	1968	069	15	1	0	5.037645448	0
4	1963	1968	068	15	1	0	5.593429158	0
5	1964	1968	067	15	1	0	4.052854209	0
6	1966	1974	069	15	1	0	8.788501027	0
7	1963	1964	066	15	1	0	1.152635181	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

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

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



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

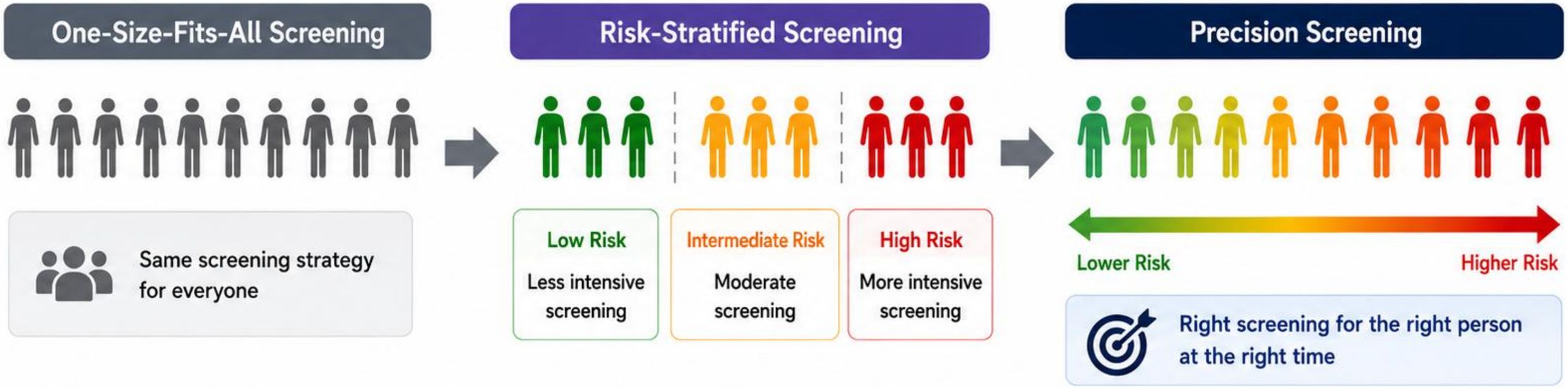
AI-empowered Precision Disease Screening of Population-based Organized Service Program

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

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.

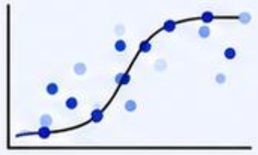
Traditional Risk Assessment

Limited & Structured Data

Age | Sex | FIT | Smoking | Family History

Statistical Models

(e.g., Logistic Regression, Cox Model)



- Specify relationship form
- Estimate parameters
- Make predictions

Risk Prediction

- Fewer variables
- Relationships are easier to specify
- Models are interpretable

Individual Risk

Risk is **multidimensional** and **highly personalized**.

Modern Health Data

Diverse, Complex & High-Dimensional

Age | Sex | FIT | Repeated FIT | Genomics (SNPs, etc.)

Imaging (CT, Colonoscopy) | Lifestyle (Diet, Activity) | Wearables (Heart rate, etc.) | Comorbidities & Medications

???

- More variables (high-dimensional data)
- More interactions between variables
- More nonlinear relationships
- More data types & sources
- More assumptions required
- Multimodal & longitudinal data

Can traditional models capture all these complex interactions?

How Can We Learn from Complex Data?

Traditional statistics and AI take different approaches.

Traditional Statistics
(Assumption-Driven / Model-First)

1. **Assumptions**
   Specify the form of relationship and underlying assumptions.
2. **Model**
   Choose a statistical model and estimate parameters.
3. **Data**
   Collect data to fit the model and test assumptions.
4. **Prediction & Inference**
   Draw conclusions, test hypotheses, and make predictions.

VS.

AI / Machine Learning
(Data-Driven / Learning-First)

1. **Data**
   Collect diverse data from multiple sources.
2. **Learning Algorthim**
   Use algorithms to learn from the data.
3. **Pattern Discovery**
   Automatically discover complex patterns and interactions.
4. **Prediction**
   Make predictions on new individuals.

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

Model-Based Learning

- Logistic Regression
- Regularized Regression
  LASSO / Ridge
- Support Vector Machine (SVM)

Strengths

- ✓ Interpretable
- ✓ Robust
- ✓ Suitable for structured clinical data

Tree-Based Learning

- Decision Tree
- Random Forest
- Gradient Boosting
- XGBoost

Strengths

- ✓ Captures nonlinear relationships
- ✓ Handles interactions automatically
- ✓ Strong predictive performance

Deep Learning

- Artificial Neural Network (ANN)
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)

Strengths

- ✓ Learns complex representations
- ✓ Handles imaging and sequential data
- ✓ Automatic feature extraction

Clinical Data

- Screening History
- Lifestyle Factors
- Biomarkers
- Demographics
- Imaging / Pathology

Machine Learning Models

Individualized Risk Prediction

Low Risk — High Risk

Decision Tree: Learning Through Sequential Questions

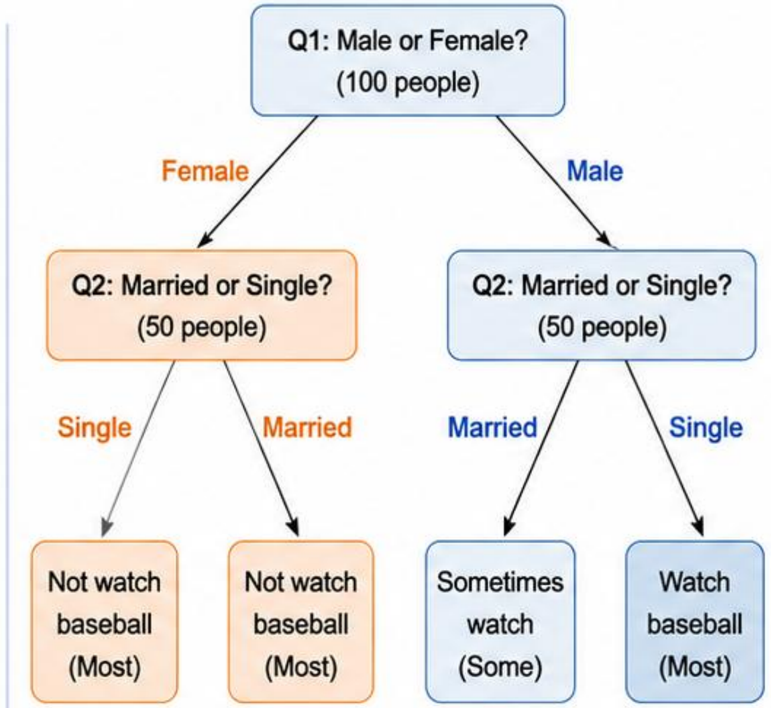
The idea is simple: ask the right questions to split the data into more homogeneous groups.

Example 1: Who is most likely to watch baseball?

We survey 100 people

Target: Watch baseball? (Yes / No)

- ✓ Male / Female
- ✓ Married / Single



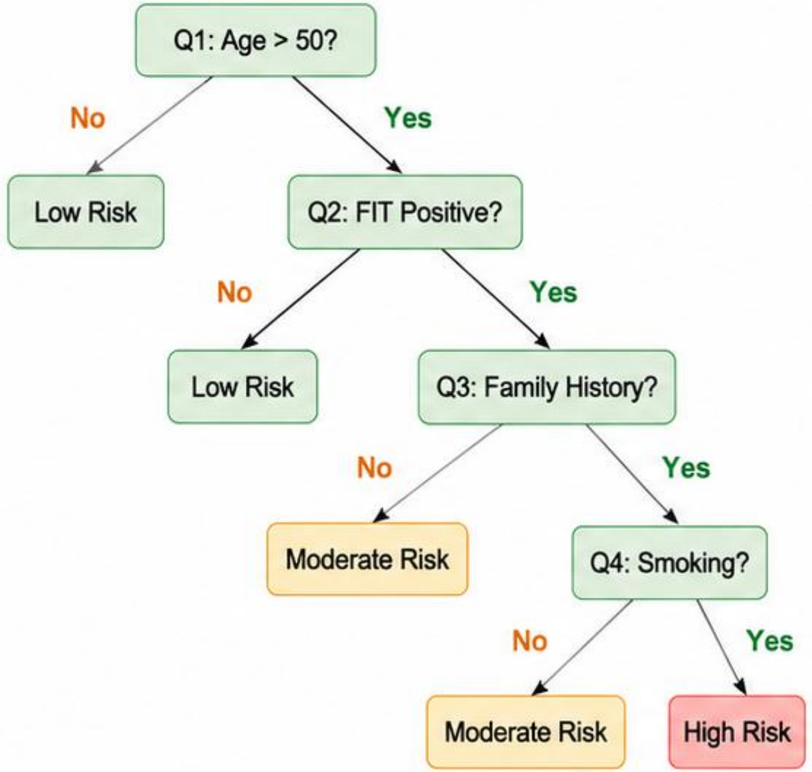
Conclusion: Single males are most likely to watch baseball.

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.

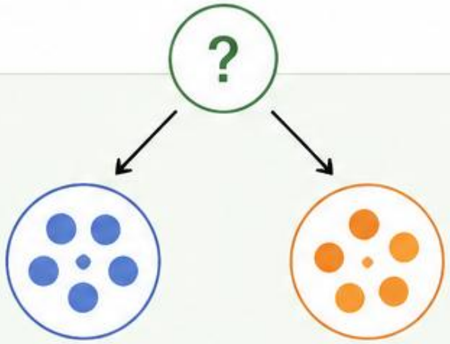
Mixed data (uncertain)

→ High impurity

Well-separated data (more certain)

→ Low impurity

The Principle



Good split = More homogeneous groups

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

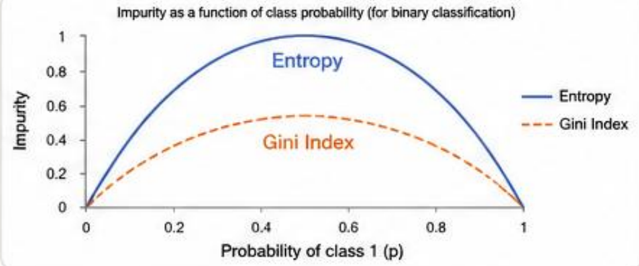
Measures uncertainty

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

Gini Index

Measures impurity

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

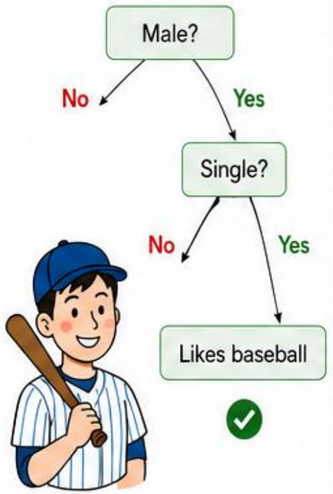


Both metrics are 0 when data are perfectly pure, and highest when data are maximally mixed.

Overfitting

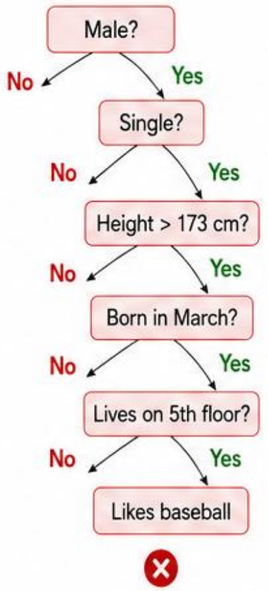
Analogy: Who likes baseball?

Good Fit (Generalizes Well)



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

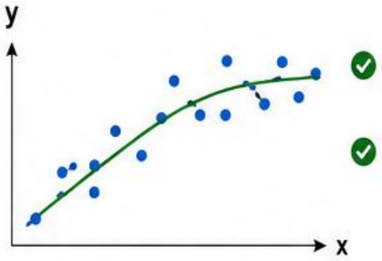
Overfit (Does Not Generalize)



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

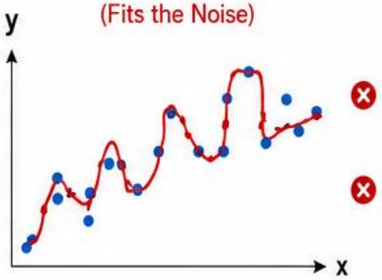
Good Fit vs. Overfit

Good Fit
(Captures the Trend)



- Captures the underlying pattern
- Performs well on new data

Overfit
(Fits the Noise)



- Follows every training point
- Performs poorly on new data

Example in Cancer Risk Prediction

Good Model

- Age
- FIT Result
- Family History
- ...

→ Predicts true cancer risk

Overfit Model

- Age
- FIT Result
- Family History
- ZIP code
- Hospital ID
- Visit date
- Random variation
- ...

→ Fits training data too well

Training AUC **0.99**

External Validation AUC **0.65**

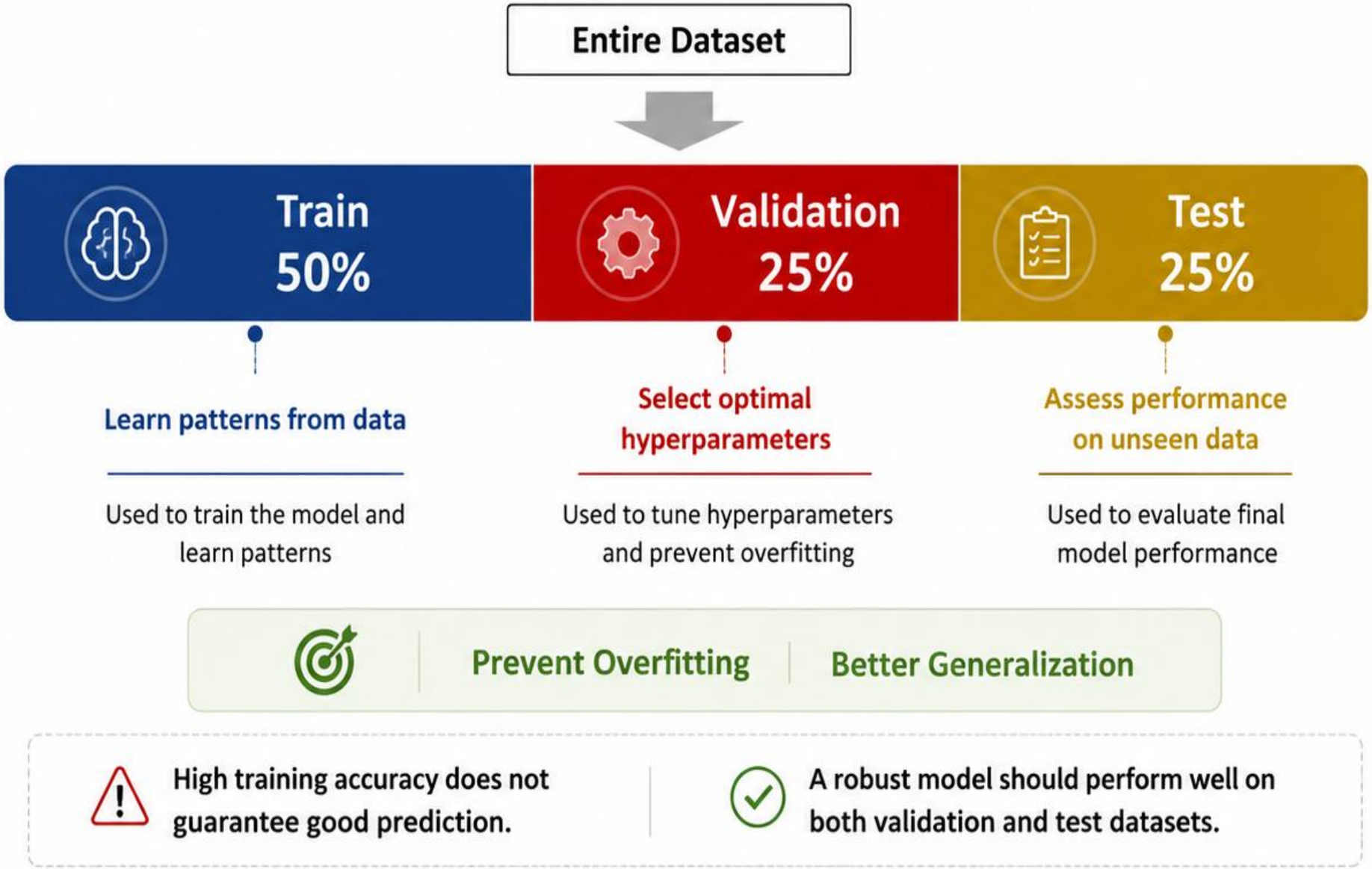
⚠ Poor generalization!

Key Message

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

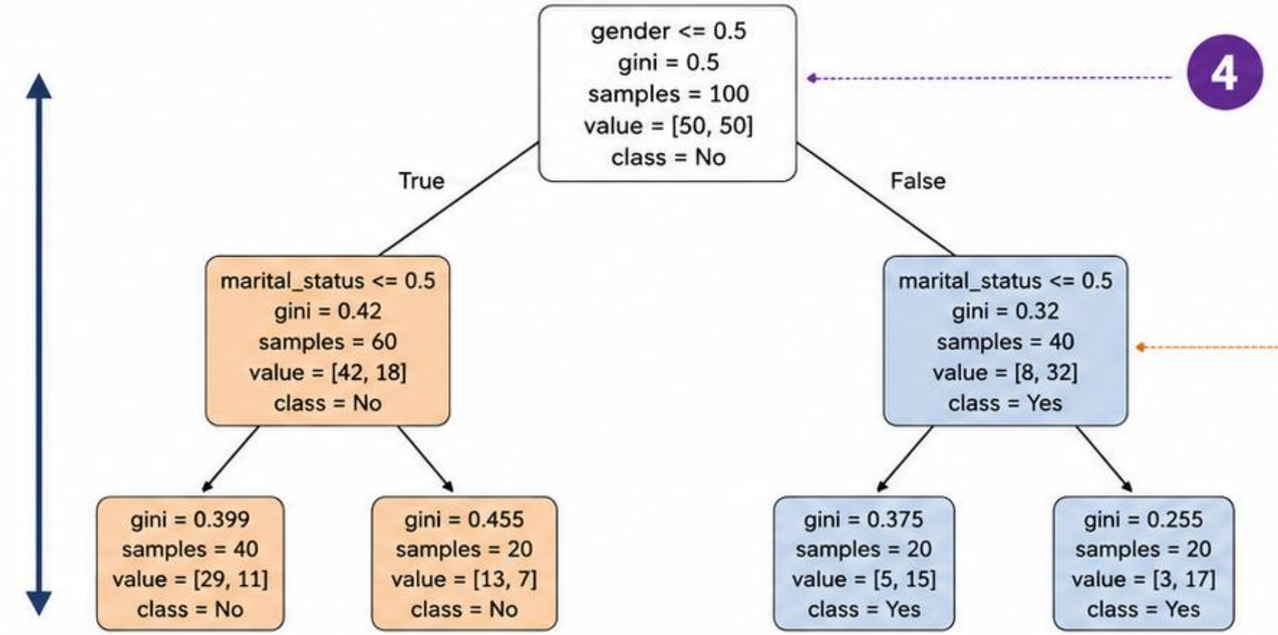
It looks great on training data, but performs poorly on new, unseen data.

Train, Validation and Test Sets



Hyperparameters That Control Tree Complexity

Preventing Overfitting in Decision Trees



1 Max Depth

Limits the maximum number of levels in the tree.

4 Gini / Entropy

Criteria used to measure impurity and choose the best split.

2 Min Sample Split

Minimum number of samples required to perform a split.

5 Minimum Impurity Decrease

Minimum required reduction in impurity to make a further split.

3 Min Sample Leaf

Minimum number of samples required to be at a leaf node.

6 Minimum Weight Fraction Leaf

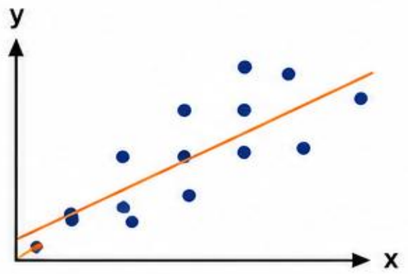
Minimum fraction of total weight required at a leaf node.

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

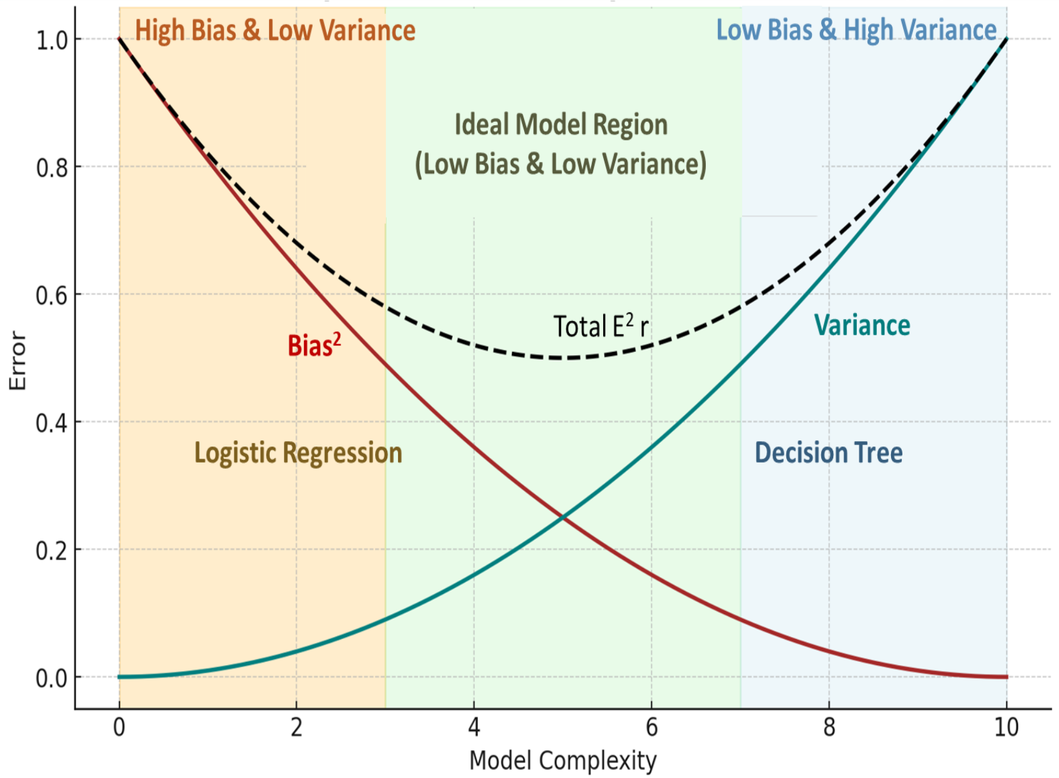


- High Bias (Underfits)
- Low Variance (Stable)

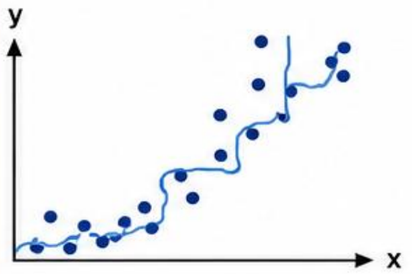
Too simple → **Misses complex patterns**

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**!

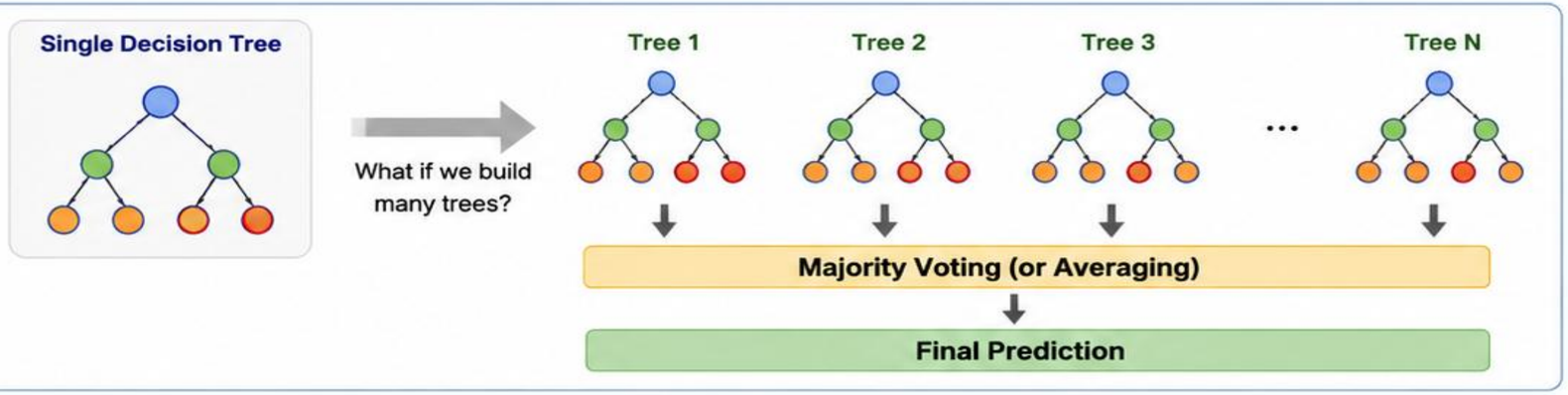
From a Single Decision Tree to a Random Forest

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



Feature Randomization

Tree	Example of Features Considered at a Split		
Tree 1	Age	FIT	Sex
Tree 2	FIT	BMI	Smoking
Tree 3	Age	Family History	Exercise
Tree 4	FIT	Sex	BMI
...			

Each tree sees different features

↓

Trees become less correlated

↓

Better generalization

Random Forest reduces variance by averaging many decision trees, leading to better generalization and improved predictive performance.

Why Random Forest Works

- ✓ Reduces overfitting
- ✓ More stable predictions
- ✓ Handles nonlinear relationships
- ✓ Works well with high-dimensional data
- ✓ Better performance on unseen data

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)

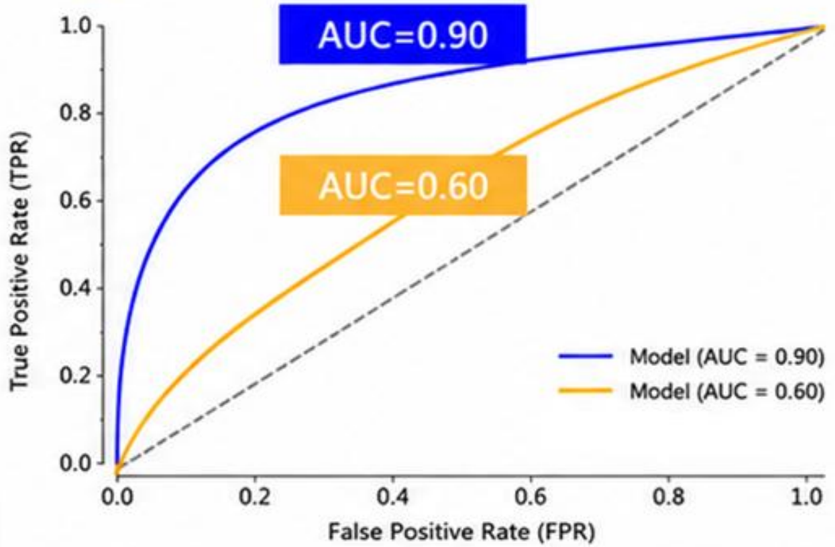
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

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

2 ROC Curve and AUC

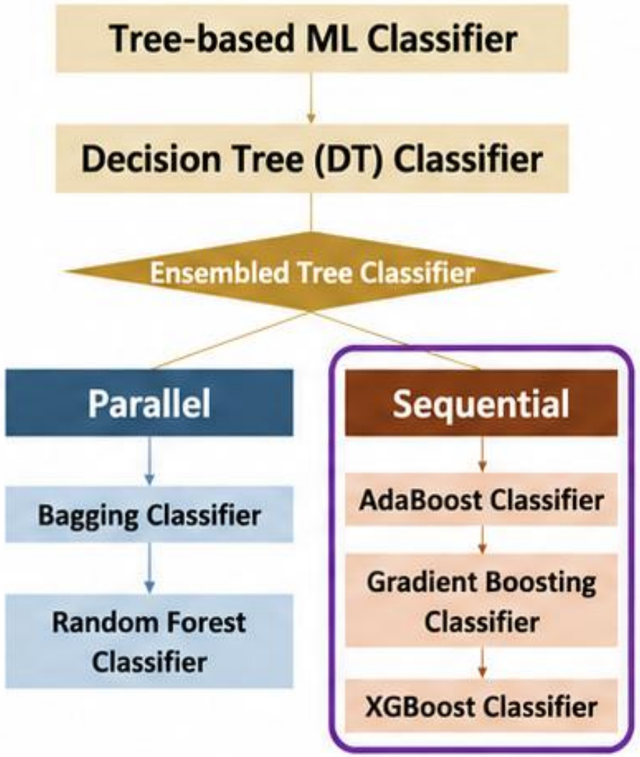


$$G\text{-}mean = \sqrt{Recall \times Precision}$$

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

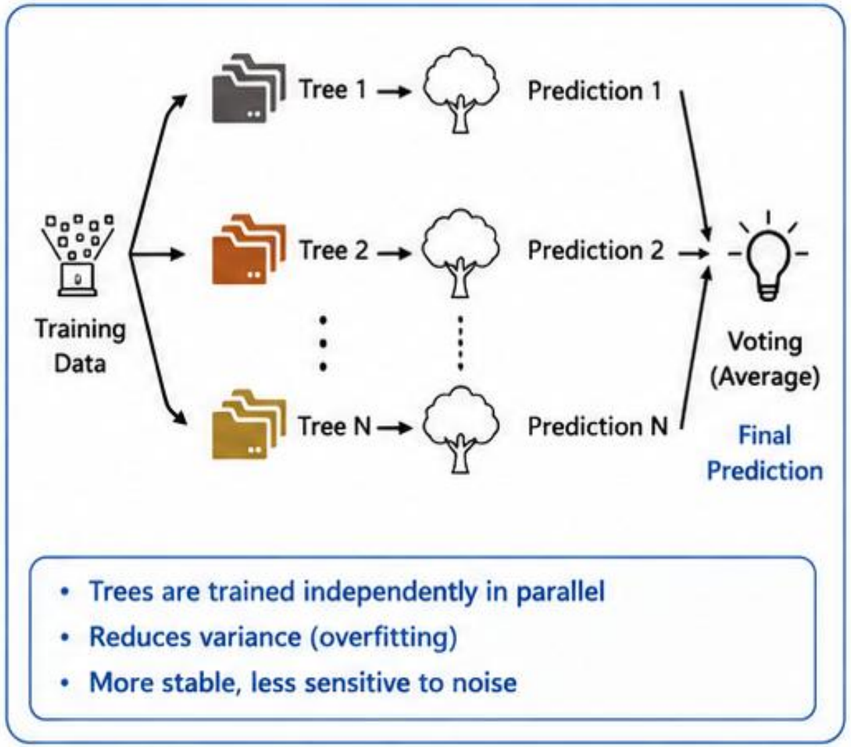
From Random Forest to XGBoost

Ensemble Learning Approaches



Bagging (Parallel)

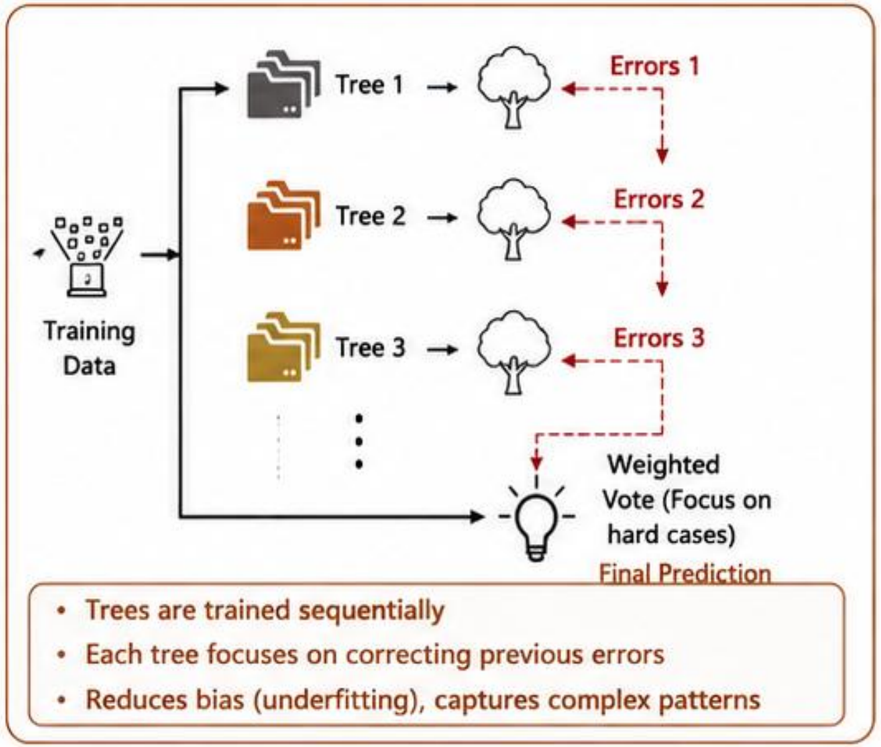
Example: Random Forest



- Trees are trained independently in parallel
- Reduces variance (overfitting)
- More stable, less sensitive to noise

Boosting (Sequential)

Example: XGBoost



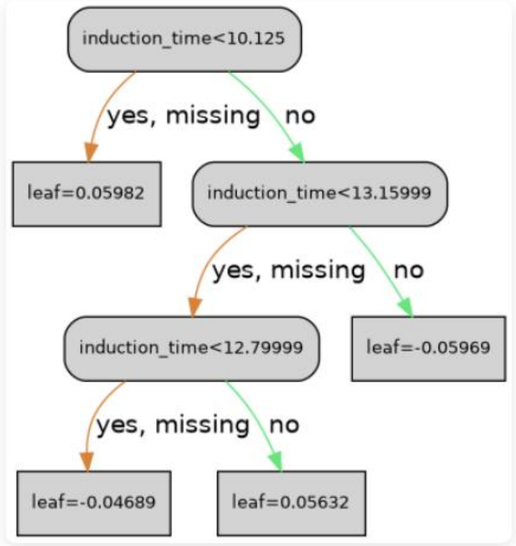
- Trees are trained sequentially
- Each tree focuses on correcting previous errors
- Reduces bias (underfitting), captures complex patterns

How Does XGBoost Learn?

Each tree learns from the errors of previous trees

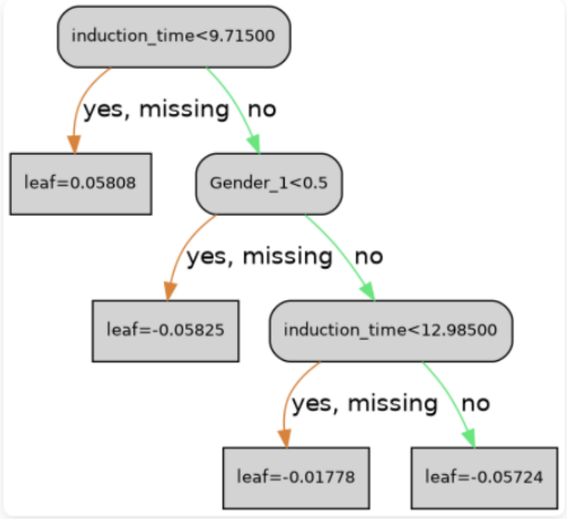
Identify Prediction Errors

First Tree



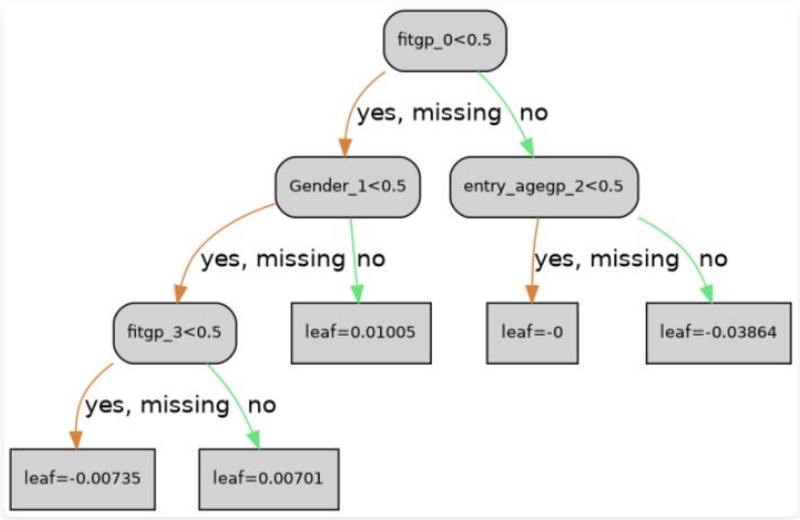
Focus on Difficult Cases

Second Tree



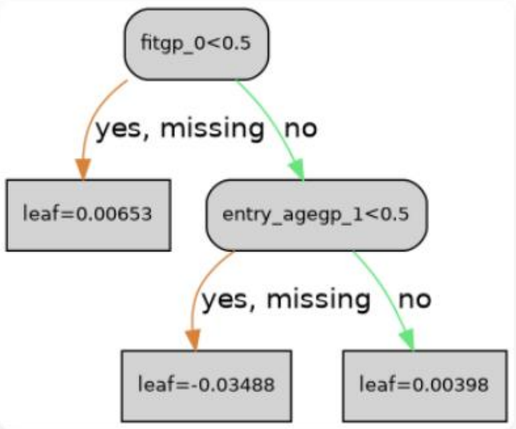
Correct Remaining Errors

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

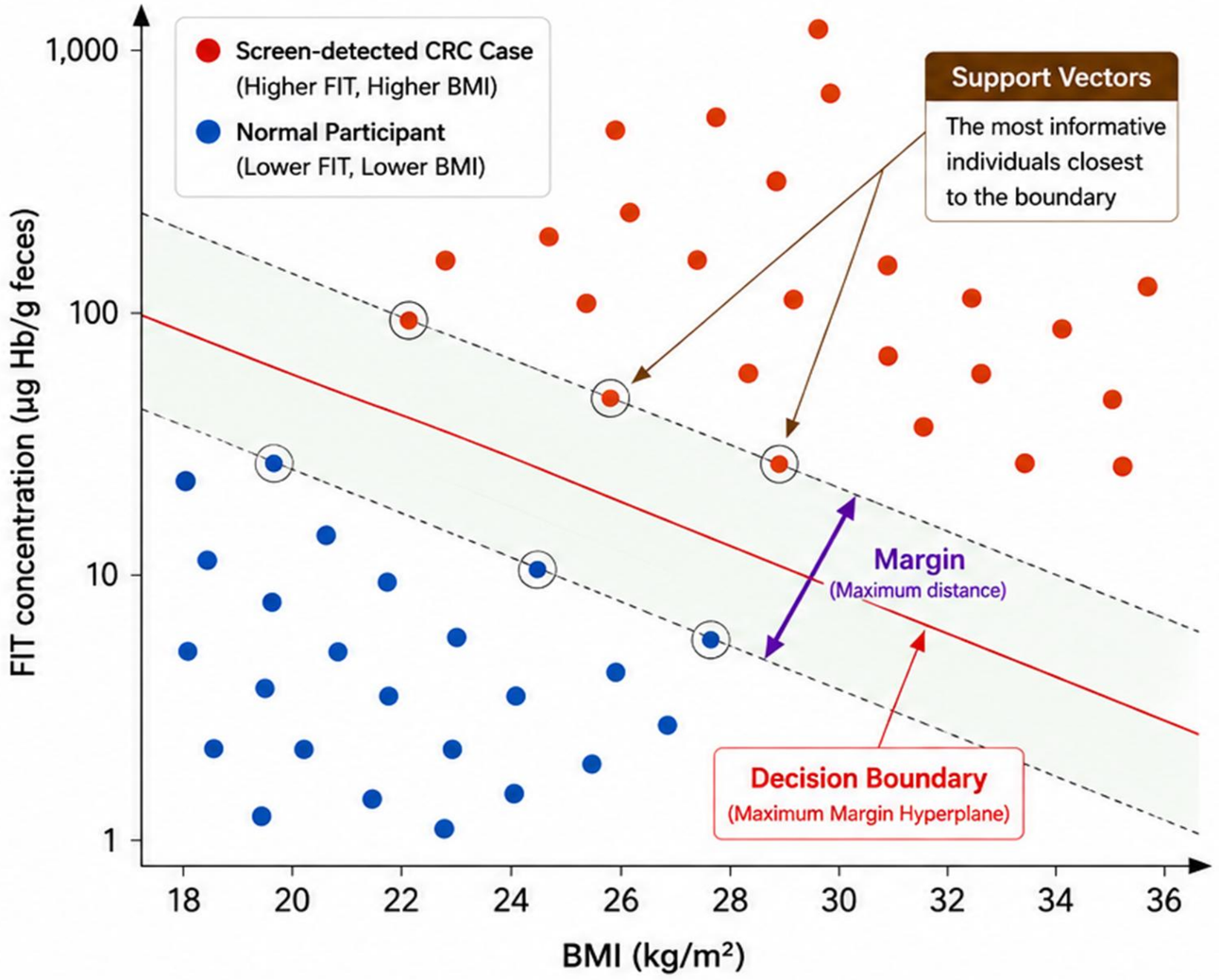
Features (Continuous Variables)

- X-axis: BMI (kg/m²)
- Y-axis: FIT concentration (µg Hb/g feces, log scale)

Target

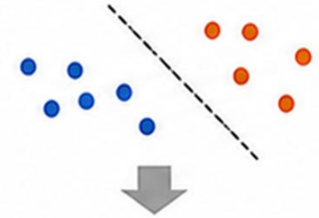
Classify individuals into:

- Screen-detected CRC Case (Higher FIT, Higher BMI)
- Normal Participant (Lower FIT, Lower BMI)

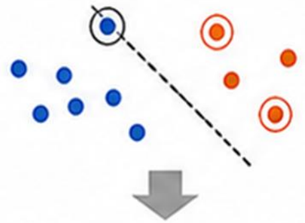


SVM Key Idea

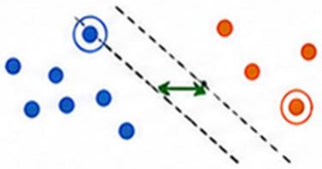
1. Find a decision boundary



2. Identify support vectors



3. Maximize the margin

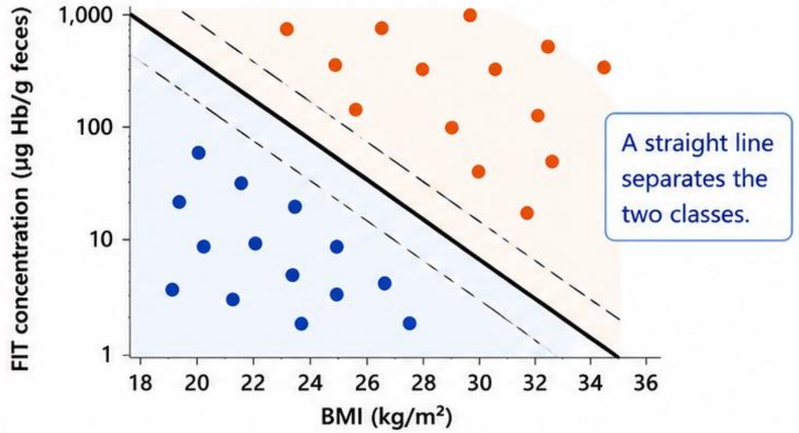


SVM Kernel Functions

Handling Non-linear Relationships in CRC Risk Prediction

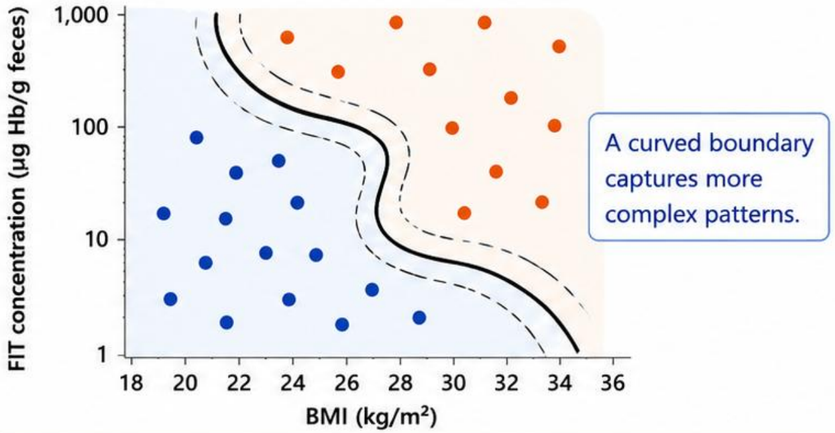
1 Linear Kernel (Linear SVM)

Simplest, fastest



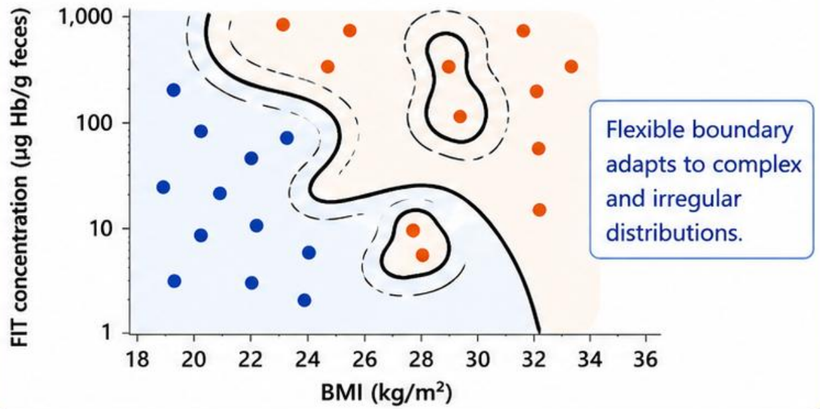
2 Polynomial Kernel

Captures interactions



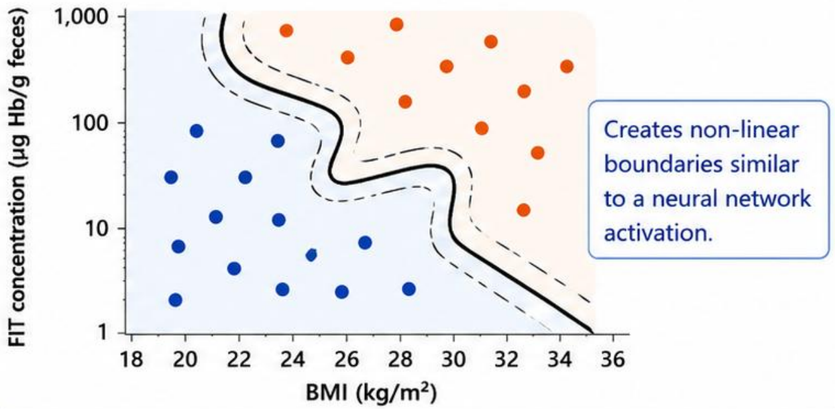
3 RBF Kernel (Gaussian)

Most flexible, widely used



4 Sigmoid Kernel

Similar to neural networks

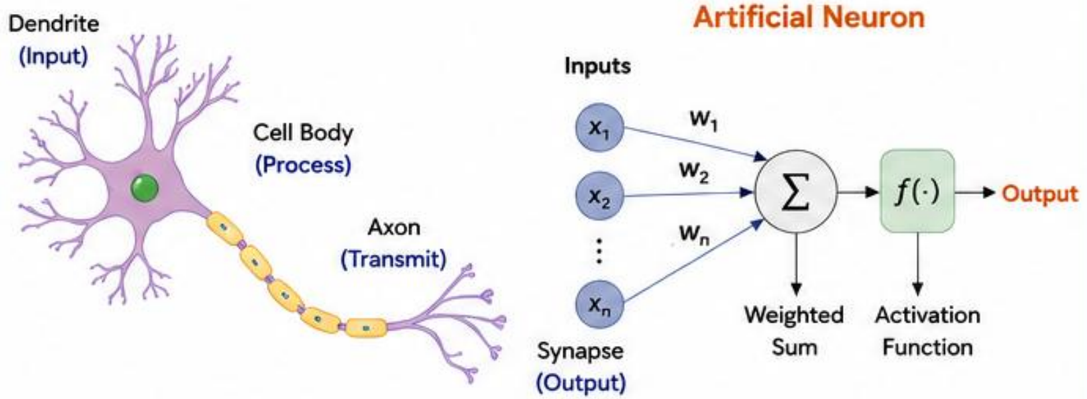


Artificial Neural Networks (ANN)

Learning Complex Patterns from Screening Data

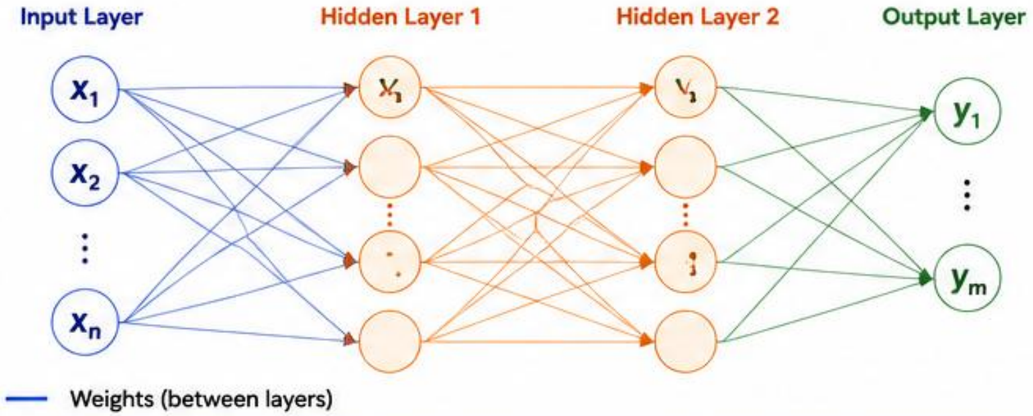
ANNs are inspired by the human brain and can **automatically learn complex, non-linear relationships** among many variables to **improve prediction accuracy**.

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)

Age · FIT · BMI · Smoking · Family History · Physical Activity · ...

Many variables describe each individual.

2. Neural Network (Learn Hidden Patterns)

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

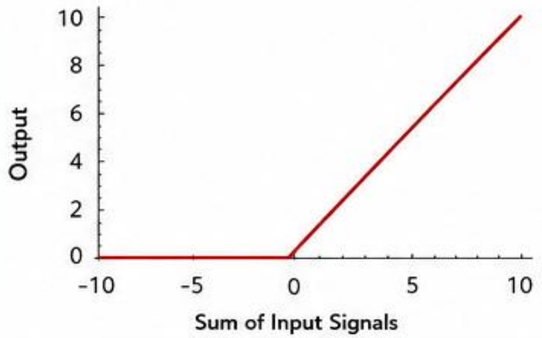
3. Output (Prediction)

- CRC Risk (High / Low)
- Advanced Adenoma Risk
- Personalized Screening Recommendation

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)

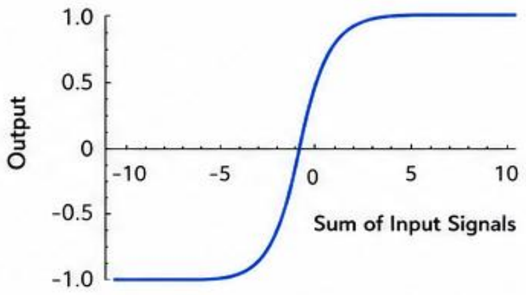


$$f(x) = max(0, x)$$

- 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)

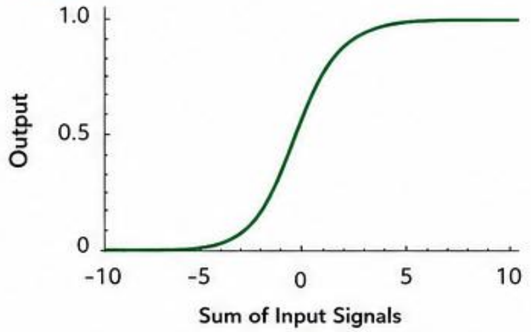


$$f(x) = \tanh(x)$$

- 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)

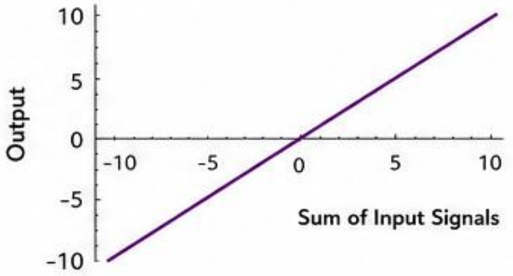


$$f(x) = \frac{1}{1+e^{-x}}$$

- 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)

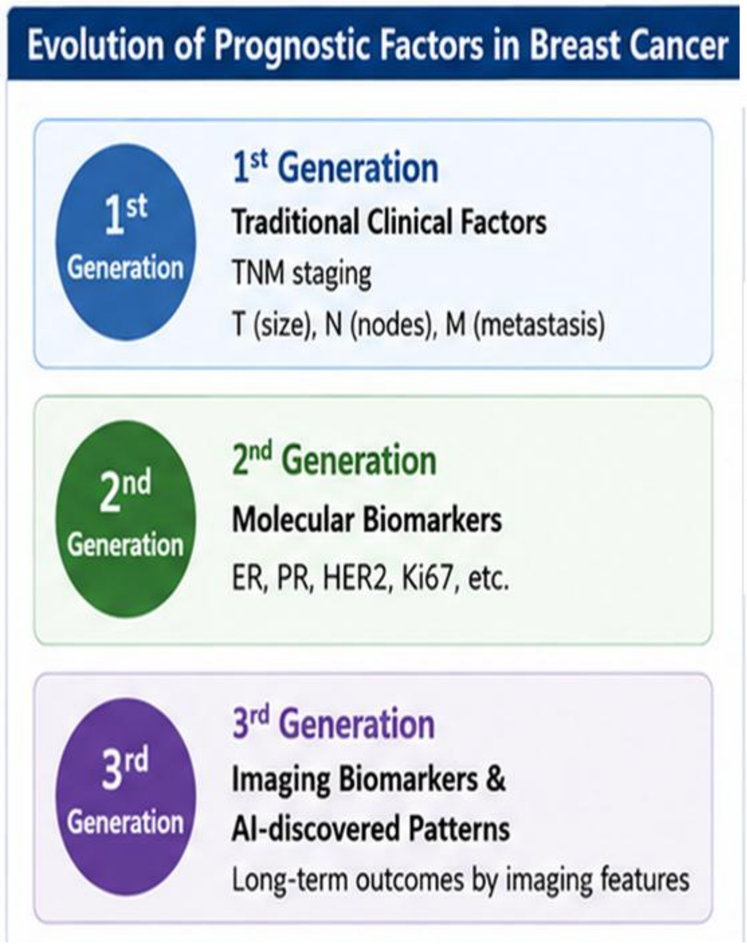


$$f(x) = x$$

- 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



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

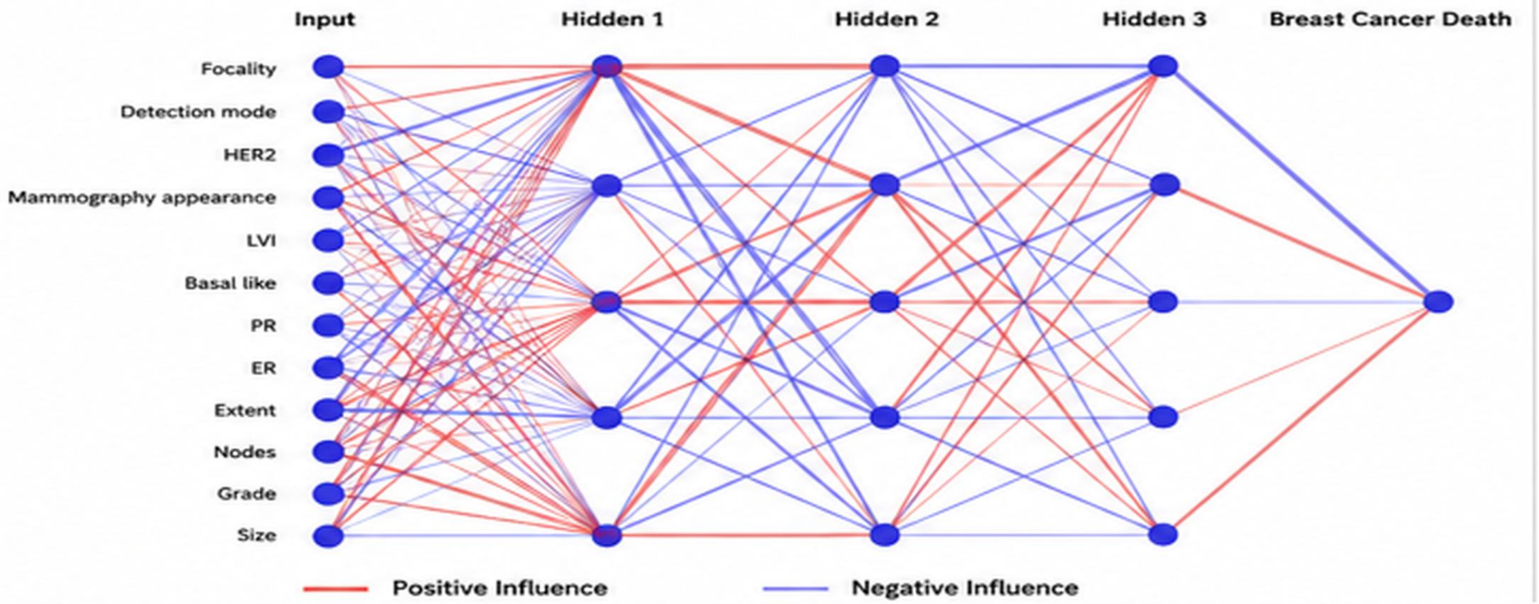
Data source: Prof. László Tabár, Sweden.

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**.

ANN Structure



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	nbcd	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

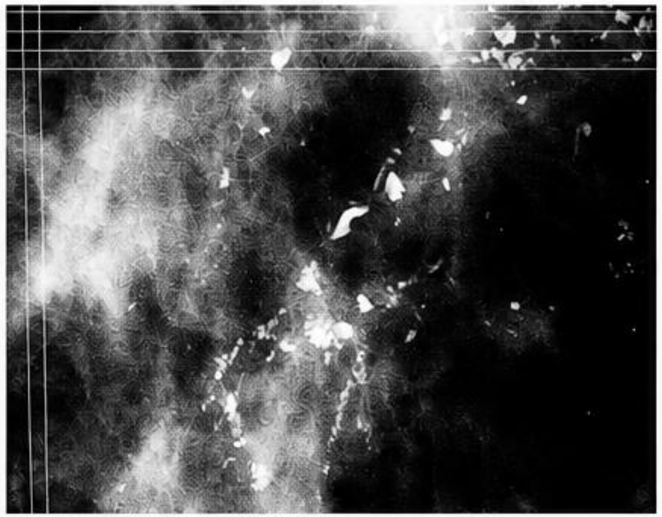
31 input features

ANN

Structured Variables → ANN

Images → CNN

2. Mammography Image



256 × 256 pixels = 65,536 input features

CNN

Too many parameters

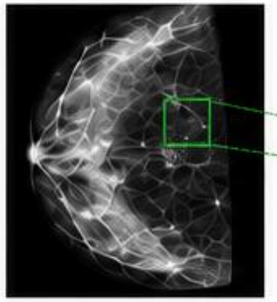
Loss of spatial information

How CNN Learns from Images

From pixels to edges, shapes, and clinically meaningful patterns

1. Input Image

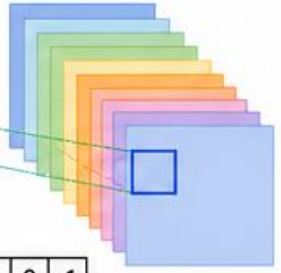
Raw image
(e.g., mammography)



- 256 × 256 pixels
- 1 channel (grayscale)

2. Convolution Layer

Apply multiple filters
to detect local patterns



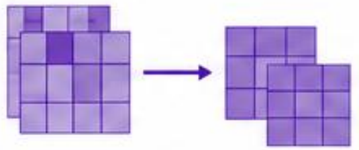
1	0	-1
1	0	-1
1	0	-1

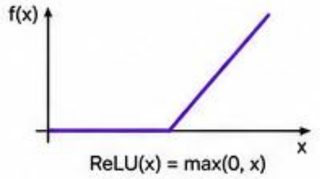
Filter (example)

Filters slide over the image to detect edges, textures, and simple shapes.

3. Activation (ReLU)

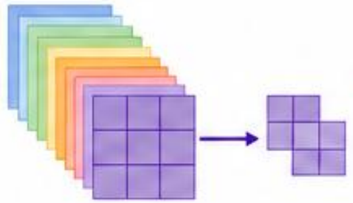
Introduce non-linearity
and keep important features





4. Pooling Layer

Downsample feature maps
to reduce spatial size
and computation



Common: Max Pooling

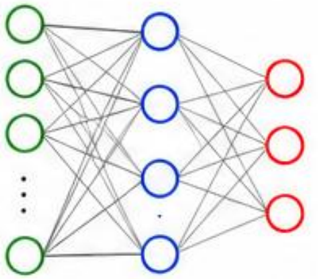
1	3	2	4
5	6	1	2
2	1	3	0
1	2	2	1

→

6	4
2	3

5. Fully Connected Layer

Combine learned features
to make predictions



The high-level features are combined to understand the overall image.

6. Output (Prediction)

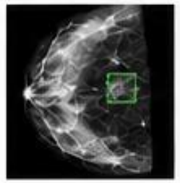
Output probabilities
for each class

Normal	5%
Benign	20%
Cancer	75%

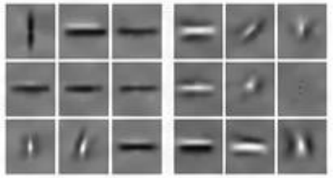
The network outputs the probability that the image belongs to each category.

How CNN Learns: Hierarchical Feature Learning

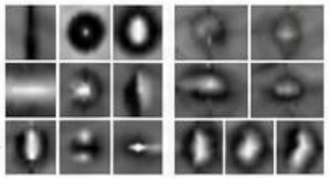
Input Image



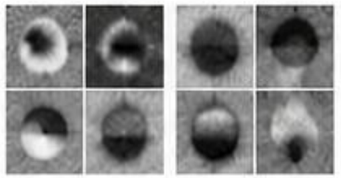
Low-Level Features
(Edges, Lines, Textures)



Mid-Level Features
(Shapes, Parts)



High-Level Features
(Objects)



Final Prediction
(Cancer / Non-Cancer)

Benign
or
Cancer

CNNs **learn** from data. They automatically **discover** the most useful features at each level without manual feature engineering.

LSTM: Learning Long-Term Dependencies

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

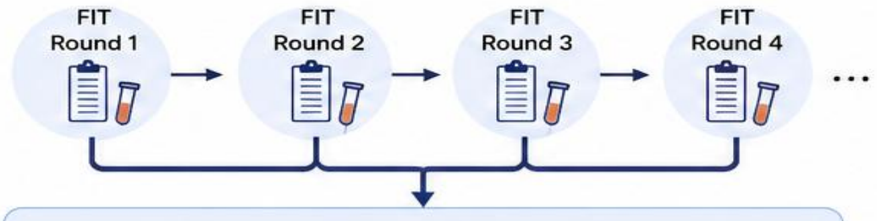
Why do we need LSTM?

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



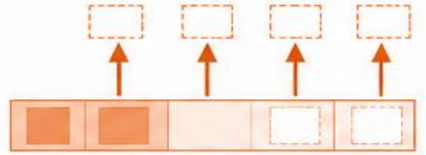
LSTM learns patterns across the **entire screening history** to predict future risk.

How does LSTM work?

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

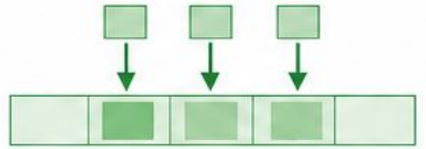
1. Forget Gate (f_t)

Decides what information is no longer important.



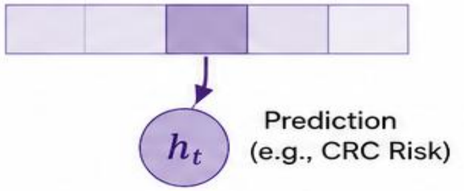
2. Input Gate (i_t)

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

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

Personalized Risk Prediction

Unsupervised Learning

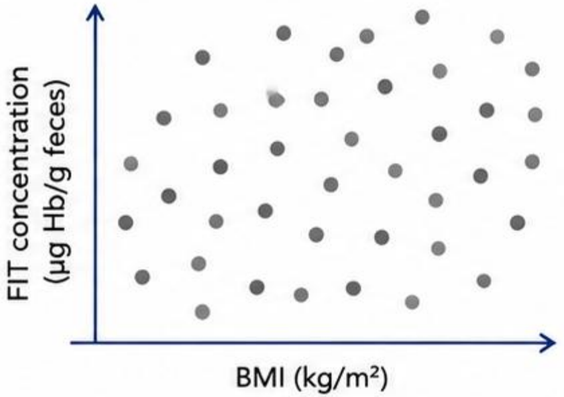
Learning from unlabeled data

K-means Clustering

Discovering Hidden Risk Groups

K-means automatically groups individuals with **similar** characteristics.

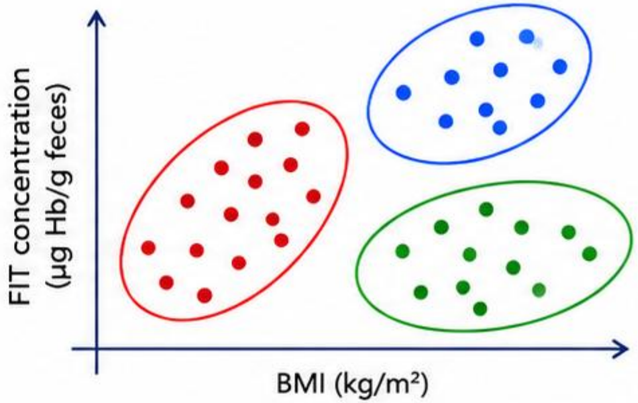
Before K-means



Individuals are mixed together.
No group structure is known.

K-means

After K-means (K = 3)



K-means groups similar individuals together, forming distinct clusters.

No CRC labels are required.
(Unsupervised Learning)

SVM

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

1 Cluster 1

Lower BMI
+ Lower FIT

Low-risk group

2 Cluster 2

Moderate BMI
+ Moderate FIT

Intermediate-risk group

3 Cluster 3

Higher BMI
+ Higher FIT

High-risk group

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.

1. Many Screening Variables

- Age
- FIT
- BMI
- Smoking
- Family History
- Physical Activity
- ⋮

20+ Variables

2. PCA

PCA

Dimension Reduction

Reduce 20+ variables to **2 key components**

3. Key Components (Example)

PC1 (70%)

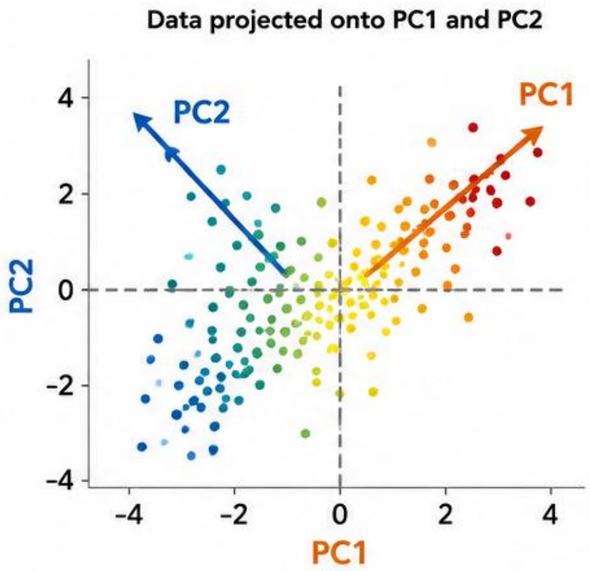
Major Variation

- Age
- FIT
- Family History
- ...

PC2 (15%)

Secondary Variation

- Smoking
- BMI
- Physical Activity
- ...



85% Information Retained

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

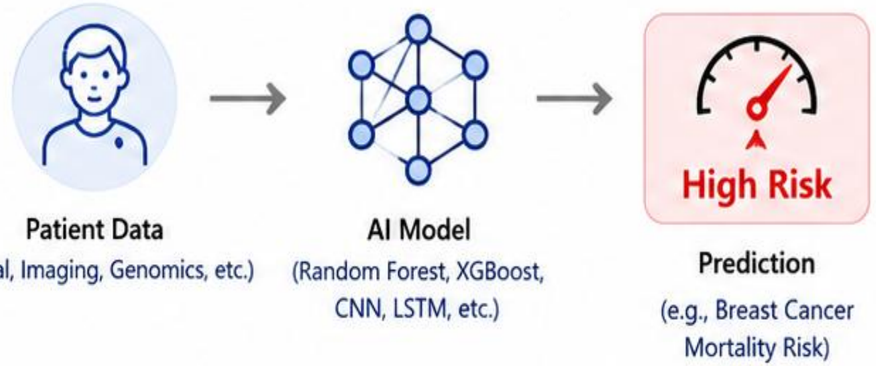
II. Artificial Intelligence and Machine Learning for Precision Screening

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



Patient Data
(Clinical, Imaging, Genomics, etc.)

AI Model
(Random Forest, XGBoost, CNN, LSTM, etc.)

Prediction
(e.g., Breast Cancer Mortality Risk)

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.

Open the Black Box
Understand how AI makes predictions.

XAI helps us answer:

" Why was this patient predicted as high risk?"

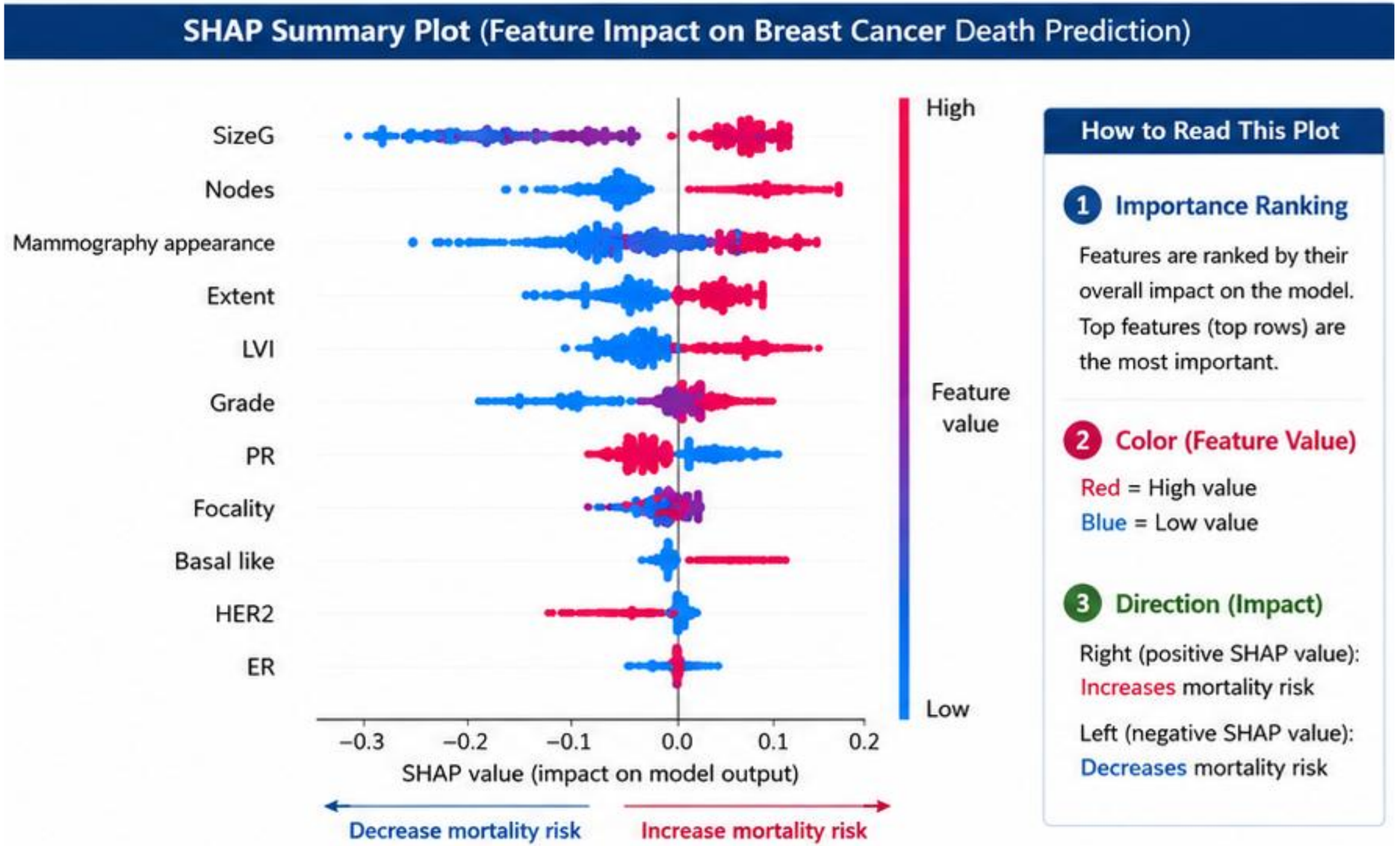
SHAP

One powerful approach:

SHAP
(SHapley Additive exPlanations)

Understanding Model Predictions with SHAP

Opening the Black Box of Random Forest and XGBoost



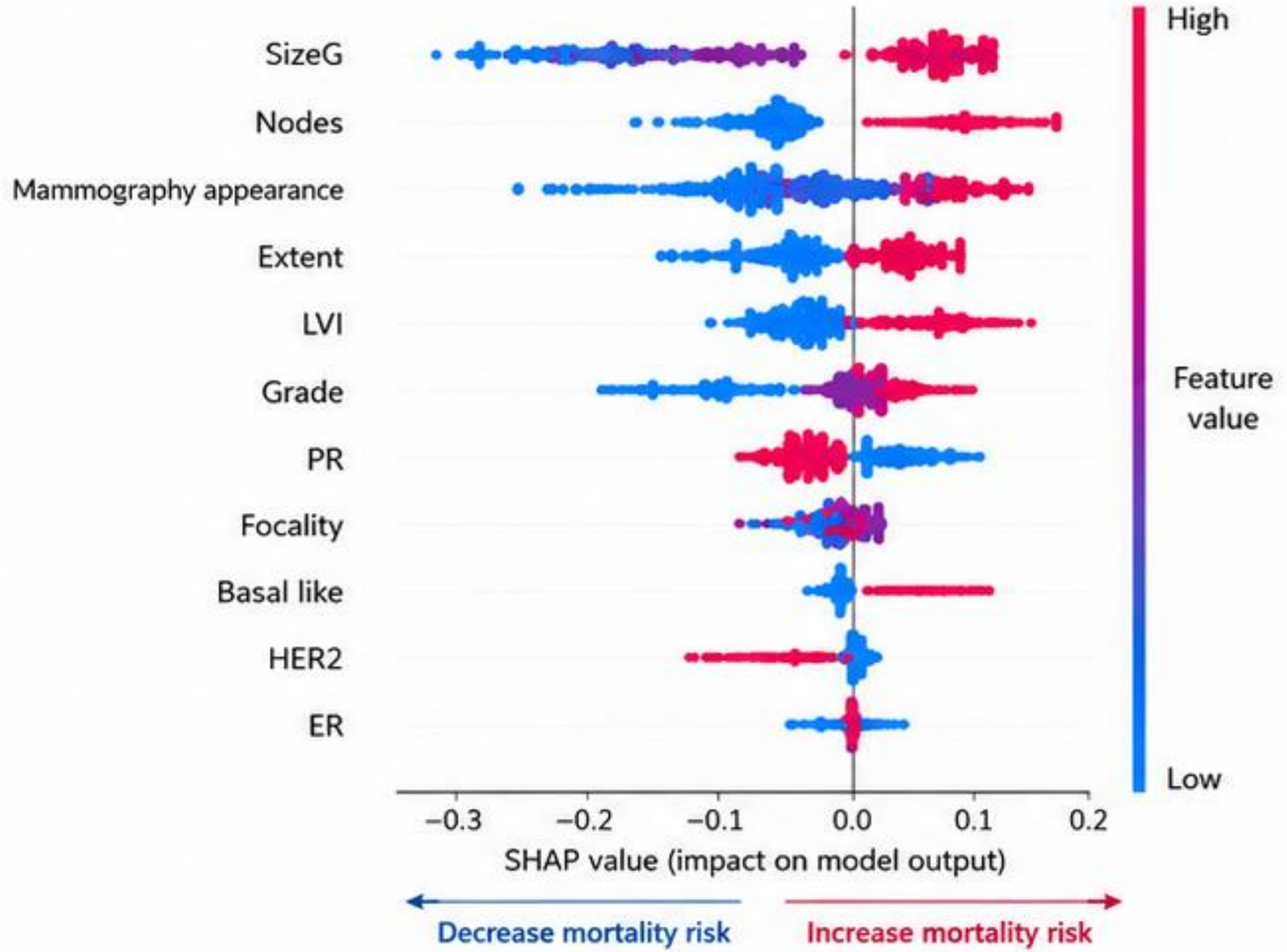
Understanding Model Predictions with SHAP

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

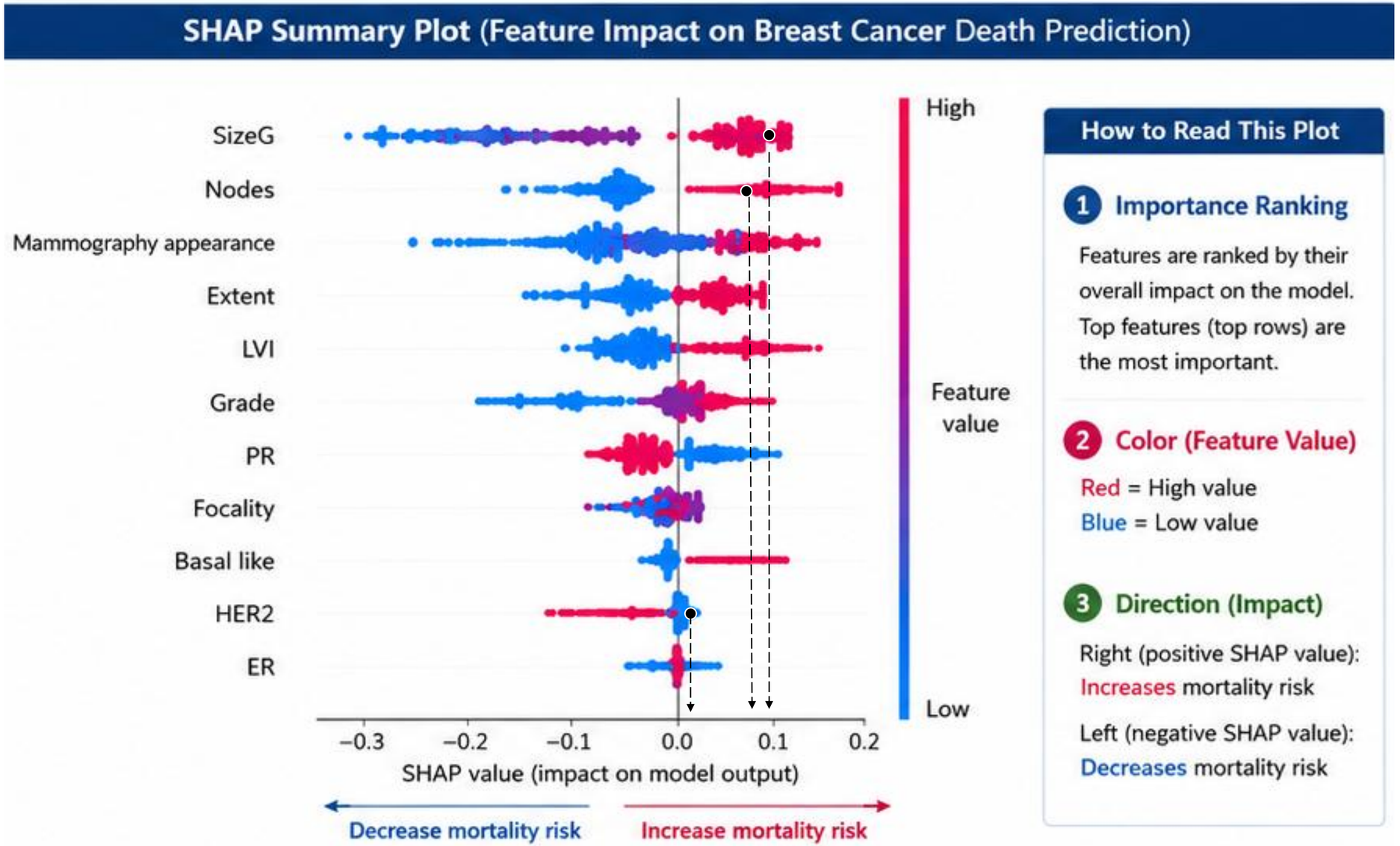
Left (negative SHAP value): Decreases mortality risk

Understanding Model Predictions with SHAP

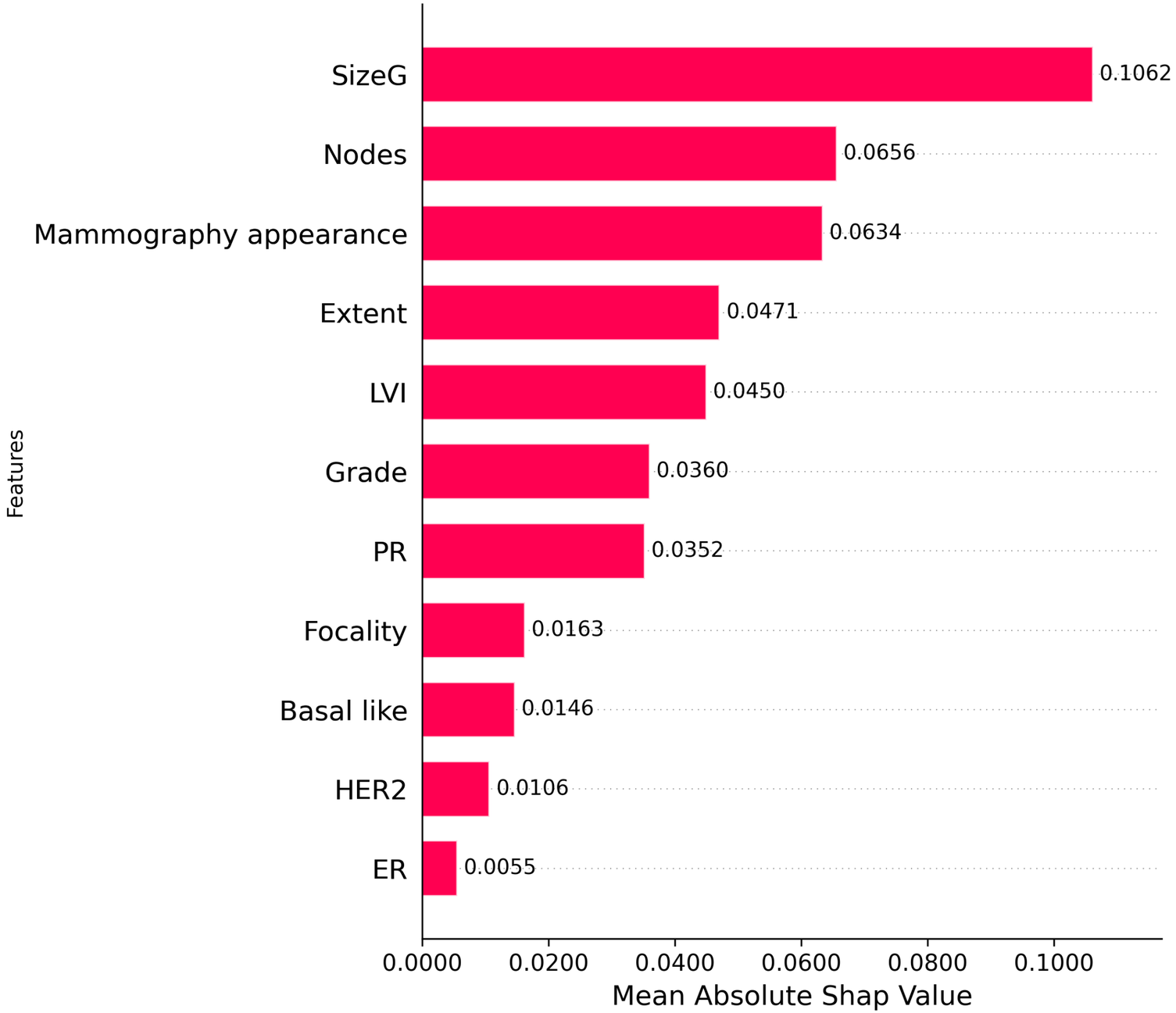
Opening the Black Box of Random Forest and XGBoost

*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?



Global SHAP Importance Plot



AI-empowered Precision Screening

From Population Screening to Personalized Prevention

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

1. Real-world Screening Data

- Taiwan community-based integrated screening program
- Demographics, lifestyle, clinical, imaging, biomarkers
- Large-scale, longitudinal data

2. Screening Evaluation

- Address selection bias and lead-time bias
- CASE-PASS for real-world screening evaluation
- Accurate assessment of screening impact

3. Disease Natural History Modeling

- Hidden disease process
- Multi-state Markov models
- Predict disease progression and risk over time

4. AI Prediction and Modeling

- Machine learning models: DT, RF, XGBoost, SVM, ANN, etc.
- Handle complex non-linear relationships
- Improve predictive performance

5. Risk Stratification and Personalization

- Identify high-risk individuals
- Stratify population by predicted risk
- Personalized screening intervals and strategies

6. Better Outcomes

- Earlier detection
- Timely intervention
- Reduced disease burden
- Improved survival, quality of life, and health equity

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

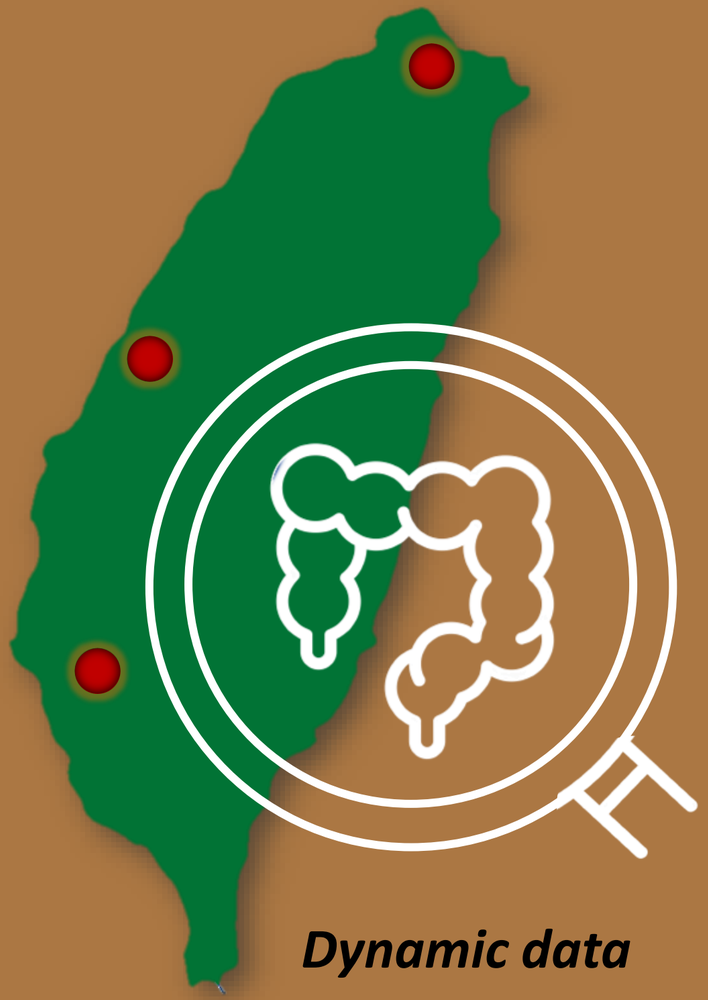
II. Artificial Intelligence and Machine Learning for Precision Screening

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

Dr. Chiu-Wen Su
National Taiwan University

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

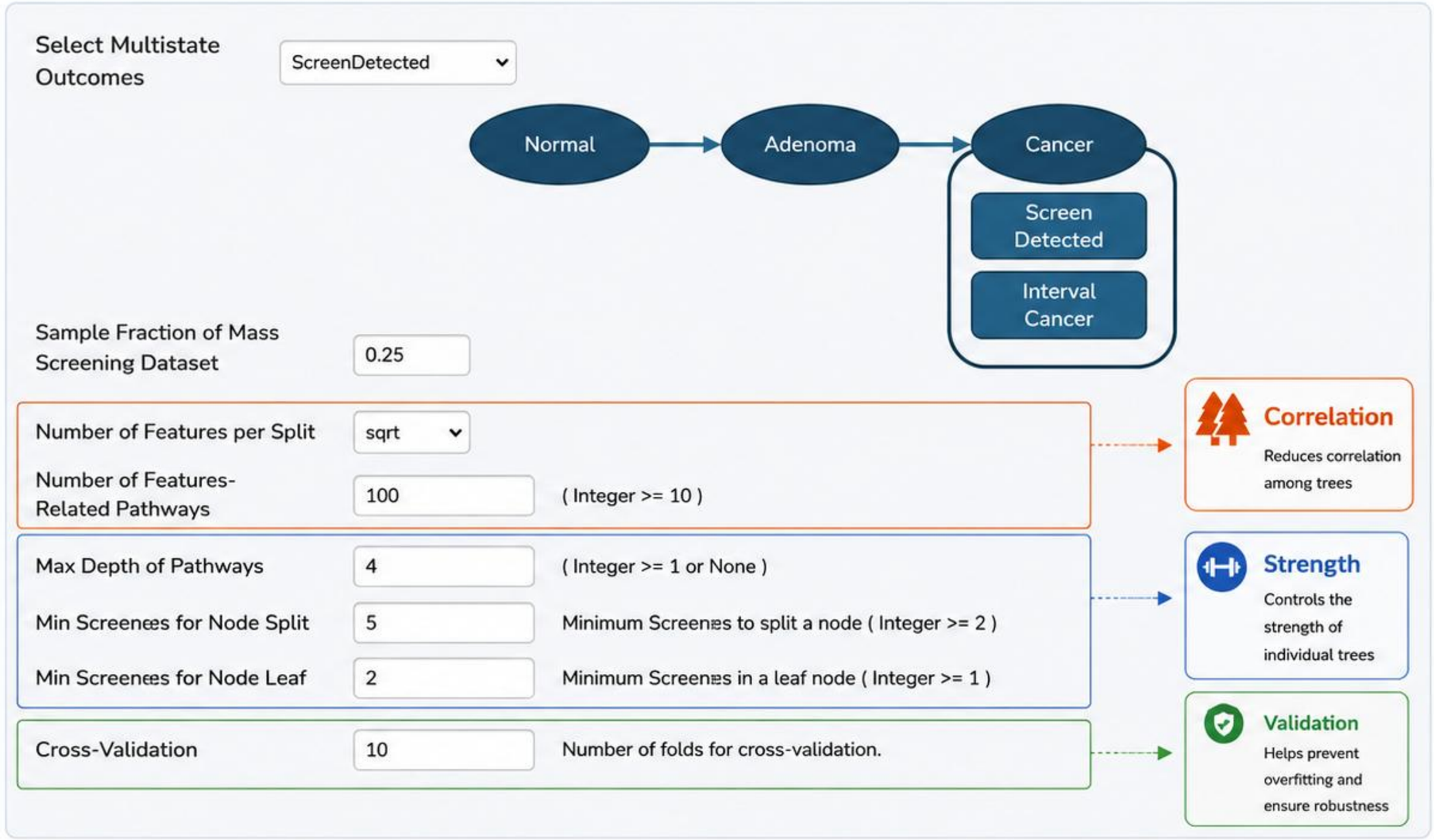
Fecal hemoglobin concentration (FIT)/ Uric Acid/ Creatinine

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?



Random Forest Classification

Random Forest Model for Screen-detected CRC

Identifying individuals most likely to be detected through CRC screening

Outcome

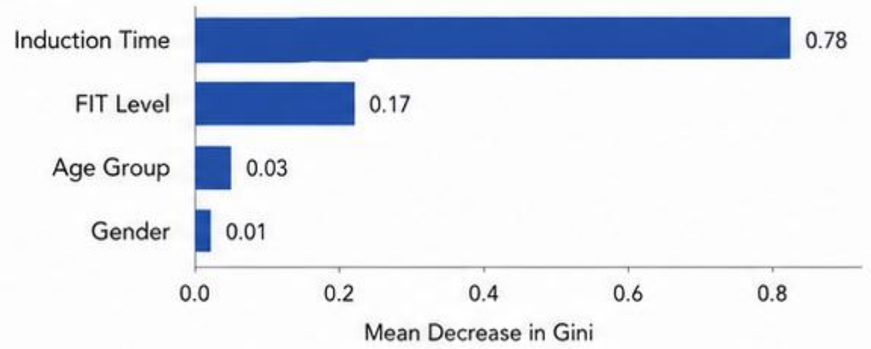
Screen-detected CRC vs Normal

Continuous	Categorical
☐ Age	☑ f-Hb level
☑ Induction time	☑ Gender
	☑ Age group

Model Hyperparameters

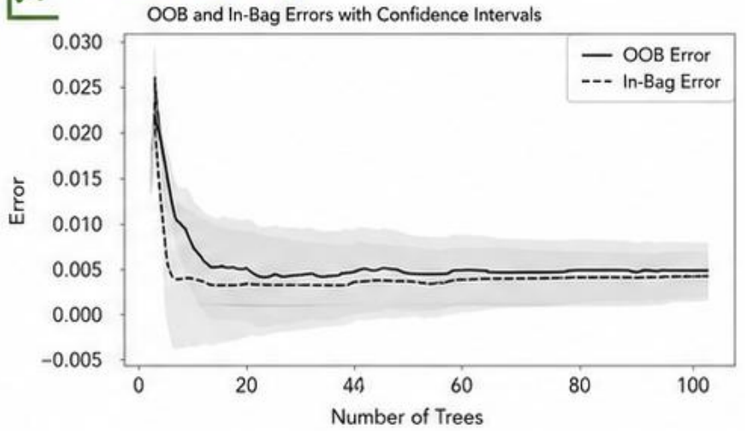
- Max Depth: 4
- Min Samples Split: 5
- Min Samples Leaf: 2
- Criterion: Gini
- Number of Trees: 100

Top Predictors (Variable Importance)



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

Model Stability

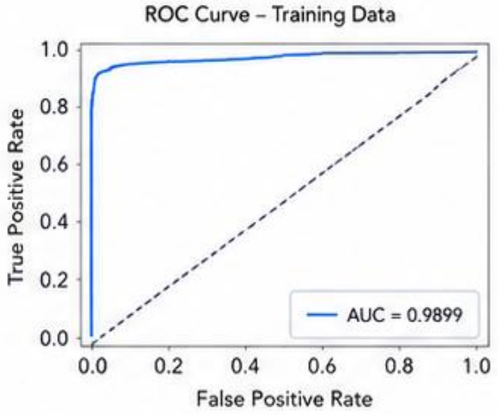


➜ OOB error stabilized after ~20 trees.

Model Performance

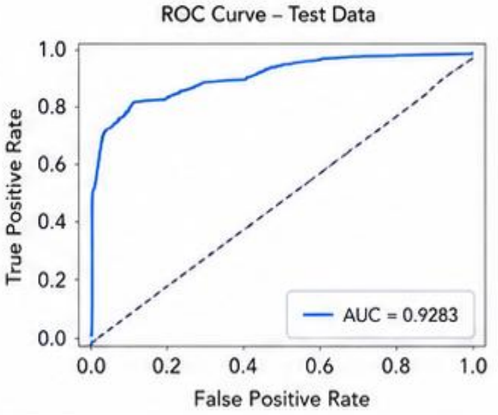
Training Set

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



Test Set

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

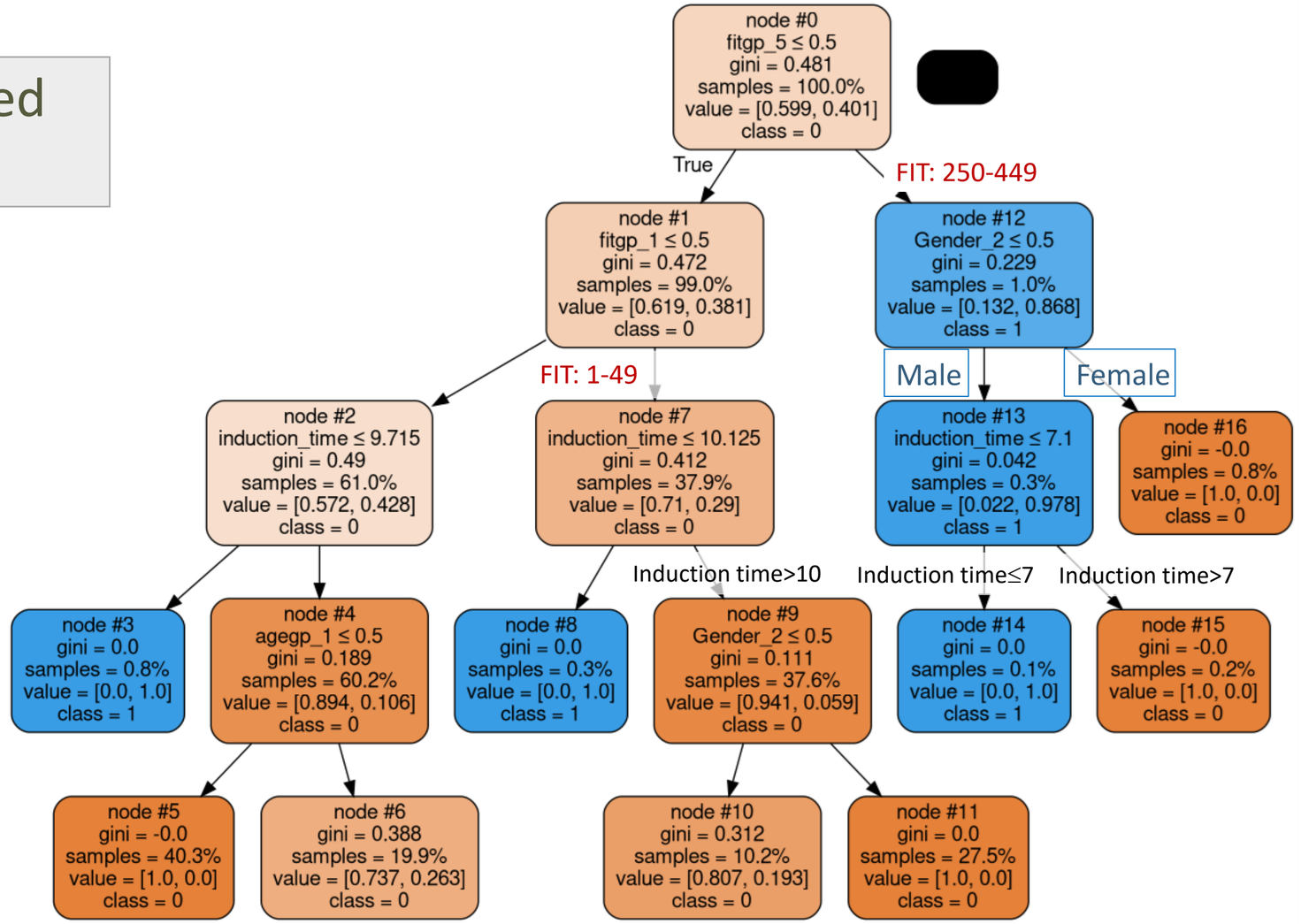


How Does the Model Make Decisions?

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



Path 1: High Risk
(Screen-detected CRC)

FIT = 250–449
↓
Male
↓
Induction time ≤ 7
↓
High probability of screen-detected CRC

Path 2: Low Risk
(Normal)

FIT = 1–49
↓
Induction time > 10
↓
Low probability of screen-detected CRC

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

A. Training AUC = 0.9899

B. Test AUC = 0.9283

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

A. Training AUC = 0.9899

B. Test AUC = 0.9283

Who is likely be detected from CRC screening?

Artificial Neural Network Analysis

Epochs: 200

Batch Size: 64

Hidden Layers: 3

Neurons per Layer: 5

Activation Function: ReLU Select the activation function to use.

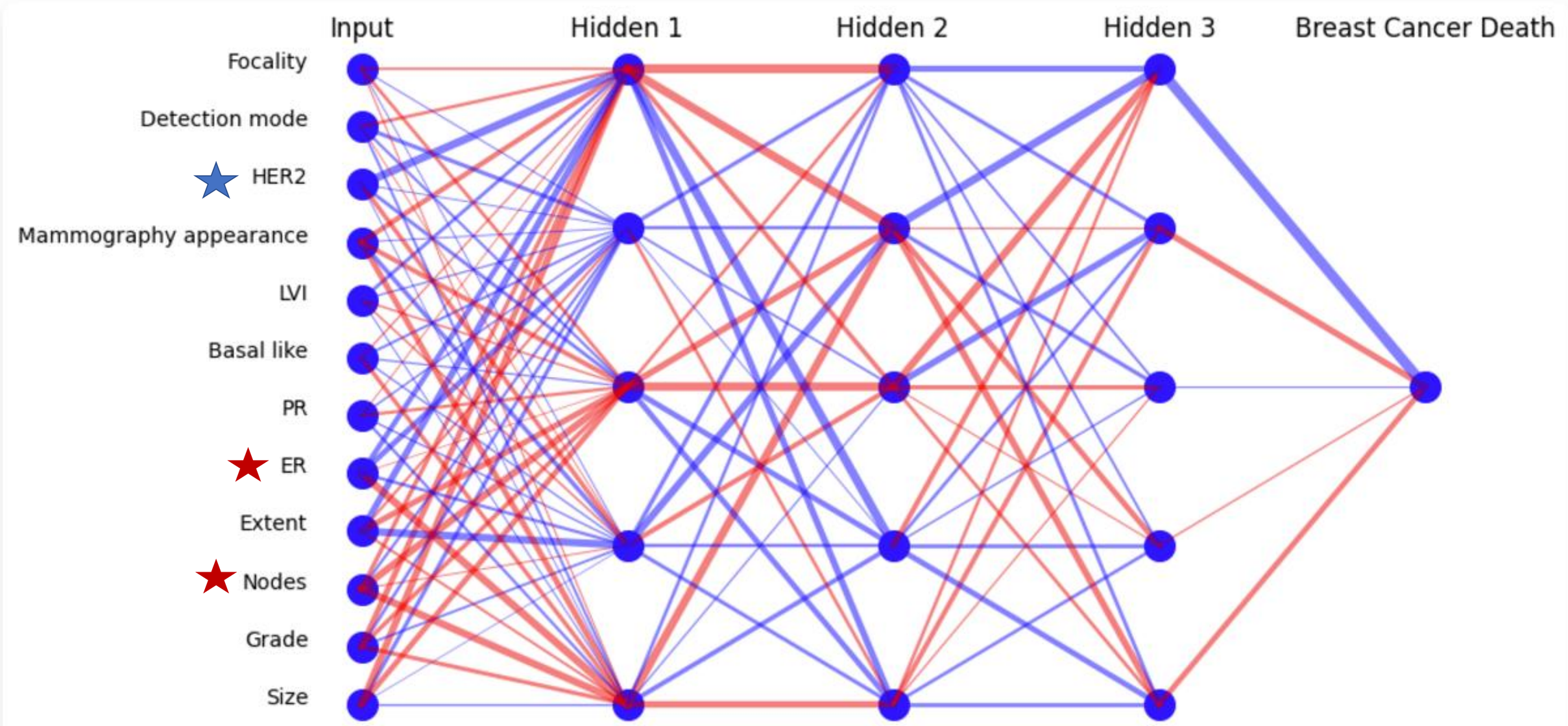
Learning Rate: 0.001 Default is 0.001

Random Seed: 12345 Seed value for random number generators to ensure reproducibility.

Run Clear Value

Artificial Neural Network Analysis

ANN Structure



What is the predicted probability of breast cancer under the following conditions (ANN)?

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

What is the predicted probability of breast cancer under the following conditions (ANN)?

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

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

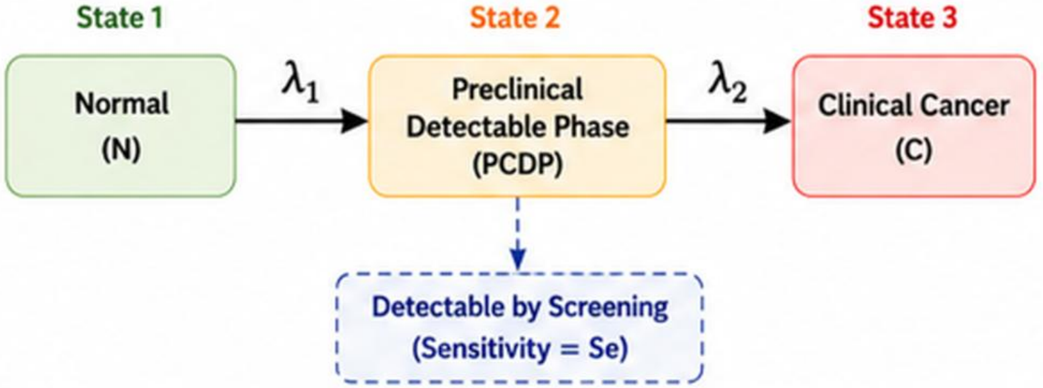
III. Digital Twin AI for Precision Screening

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

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

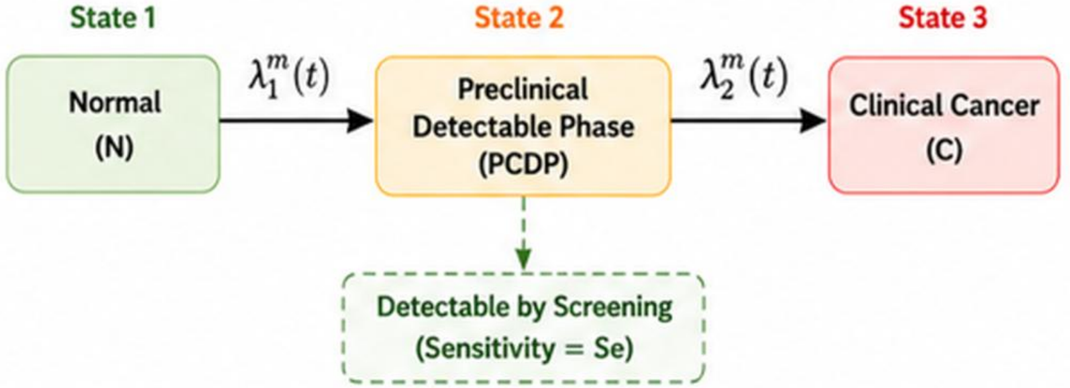
Used to estimate model parameters from observed screening data.

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

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

$\lambda_2^m(t)$ 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 ($m = 1, 2, \ldots, M$)

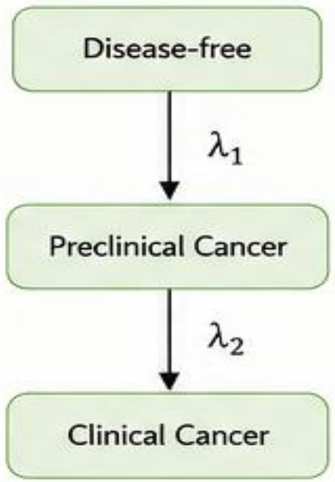
Hui-Min W, Ming-Fang Y, Chen TH. SAS macro program for non-homogeneous Markov process in modeling multi-state disease progression. Comput Methods Programs Biomed. 2004 Aug;75(2):95-105.

Estimation of Natural History Parameters of Breast Cancer

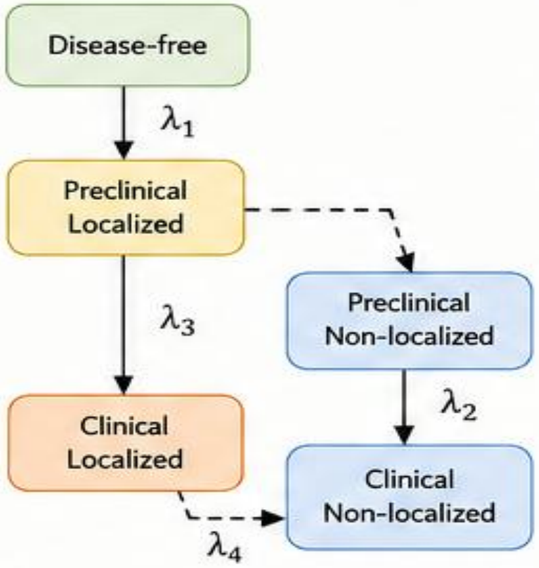
Based on Non-randomized Organized Screening Data

Multi-State Markov Models

Three-State Model



Five-State Model (by stage)



Parameters Estimated

- ✓ Transition probability (λ_1, λ_2, λ_3, λ_4)
- ✓ Mean sojourn time (MST)
- ✓ Screening sensitivity

Simulating Screening Strategies

Model inputs

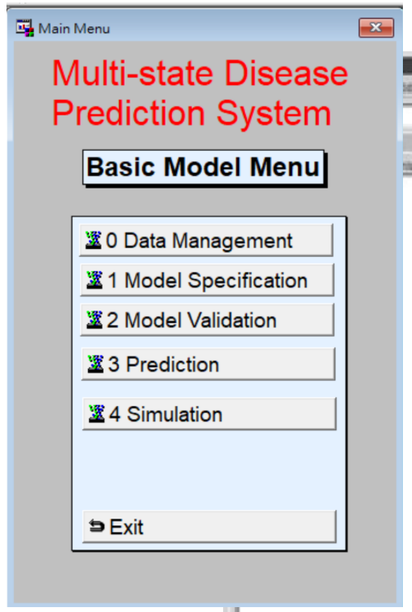
Screening Interval	Screening Sensitivity	Attendance Rate
• Annual	• 68%	• 30%
• Every 2 years	• 80%	• 60%
• Every 3 years	• 90%	• 90%

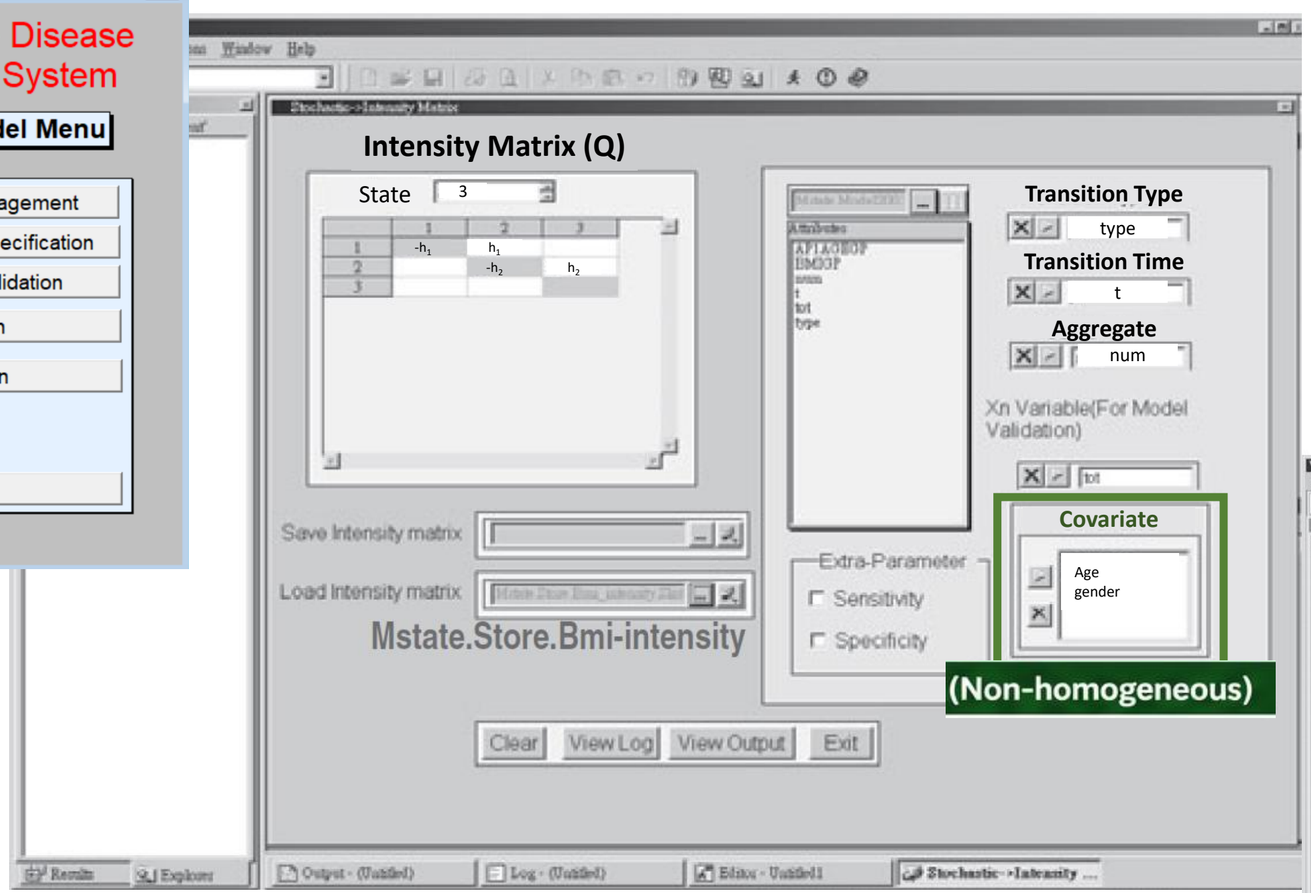
Model outputs

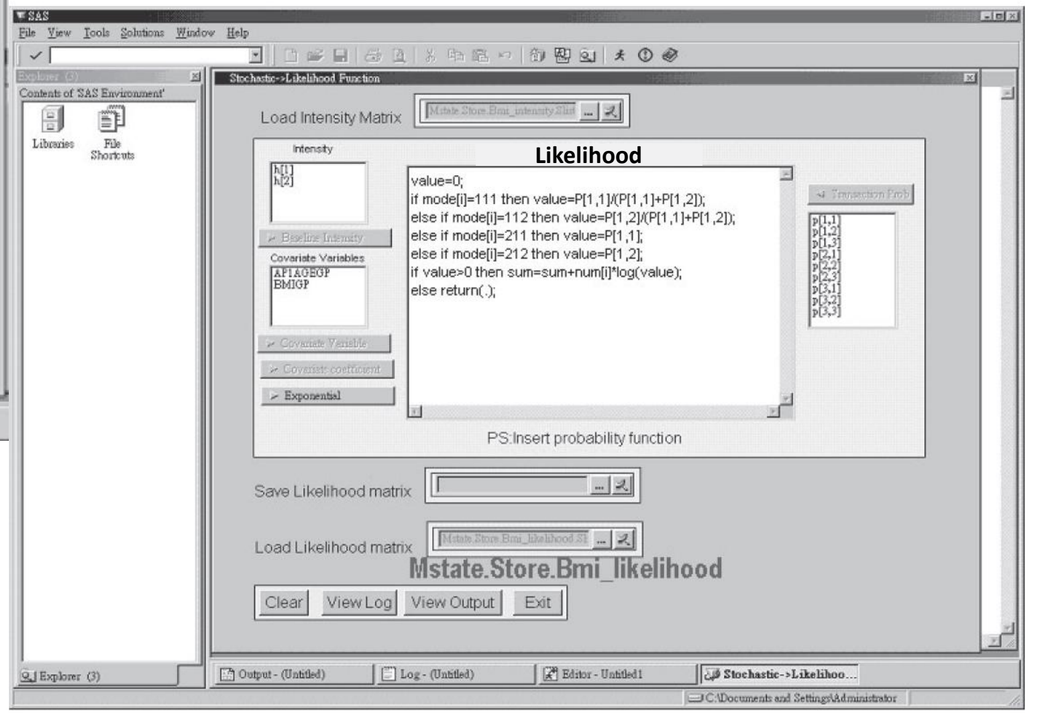
Advanced Breast Cancer Reduction → Mortality Reduction

Wu JC, Hakama M, Anttila A, Yen AM, Malila N, Sarkeala T, Auvinen A, Chiu SY, Chen HH. Estimation of natural history parameters of breast cancer based on non-randomized organized screening data: subsidiary analysis of effects of inter-screening interval, sensitivity, and attendance rate on reduction of advanced cancer. Breast Cancer Res Treat. 2010 Jul;122(2):553-66.

Multi-state Disease Prediction System







Chang, C. M., Lin, W. C., Kuo, H. S., Yen, M. F., & Chen, T. H. H. (2007). Estimation and prediction system for multi-state disease process: Application to analysis of organized screening regime. Journal of Evaluation in Clinical Practice, 13(6), 867-881.