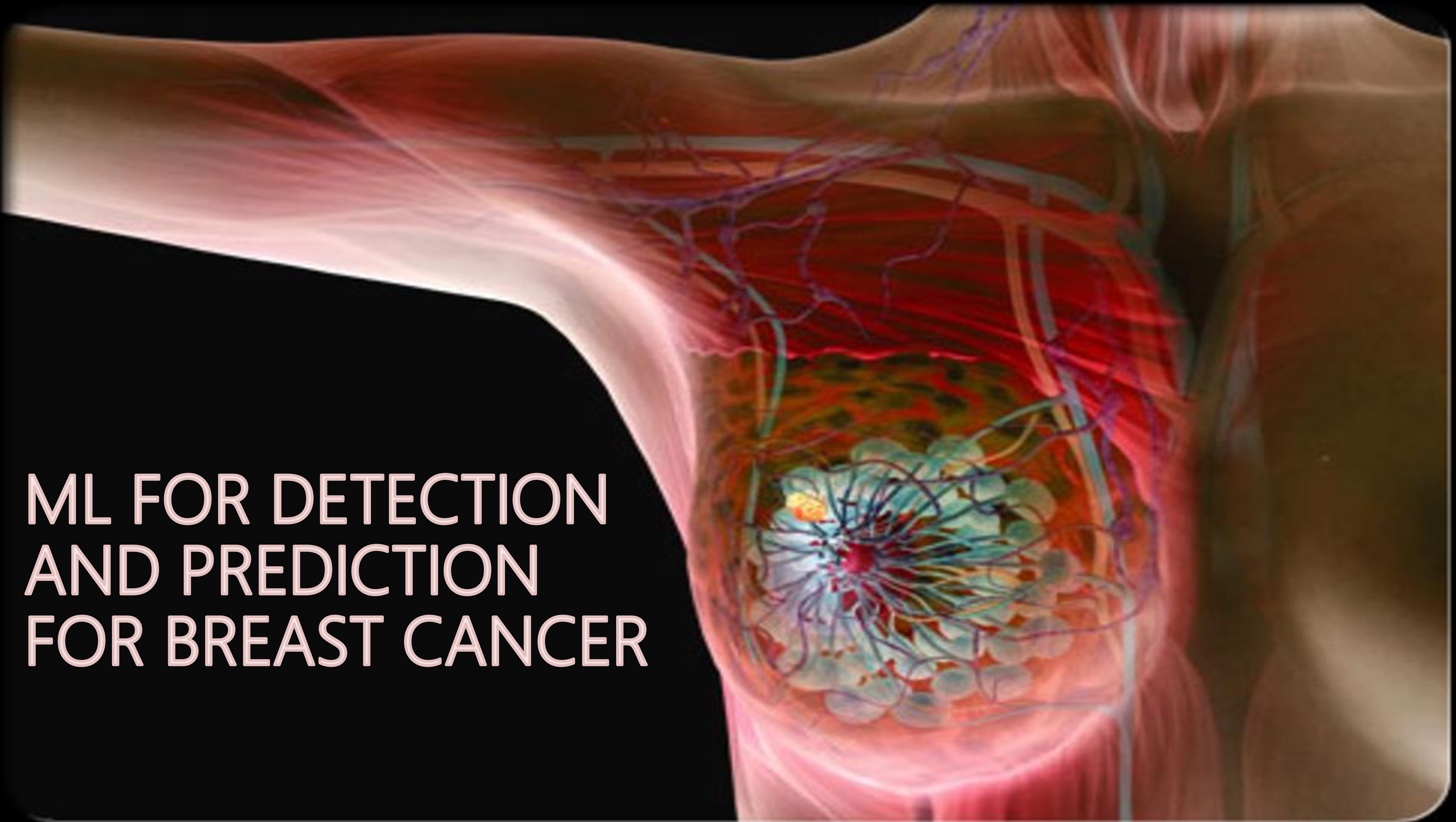


ML FOR DETECTION AND PREDICTION FOR BREAST CANCER

Breast Cancer Receptors

- Hormone receptor (HR)
- HR+ means that tumor cells have receptors for estrogen or progesterone. Certain therapies can target HR+ tumors.

- Human epidermal growth factor receptor (HER2)
- HER2+ means that tumor cells overexpress the protein HER2 which is associated with aggressive breast cancers. Certain therapies can target HER2+ tumors.

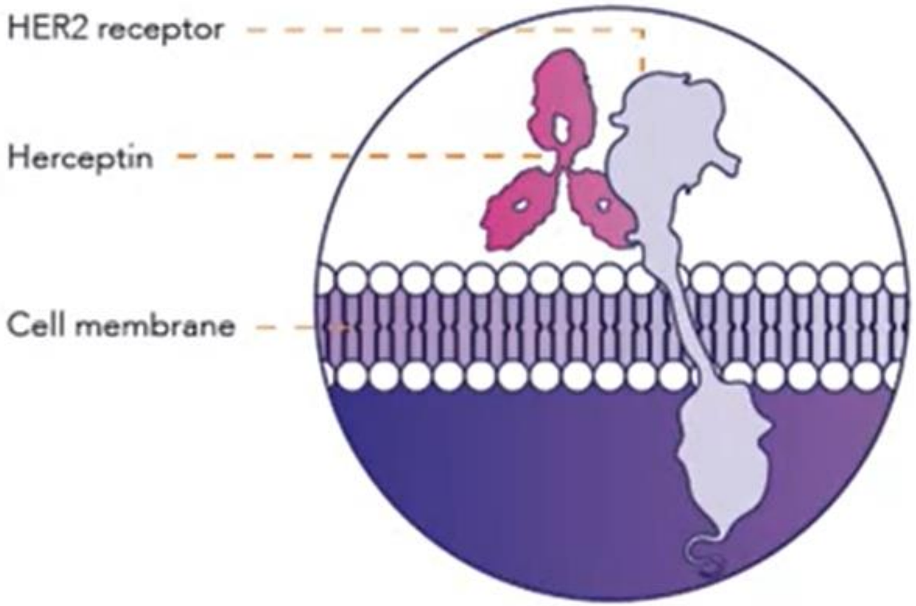
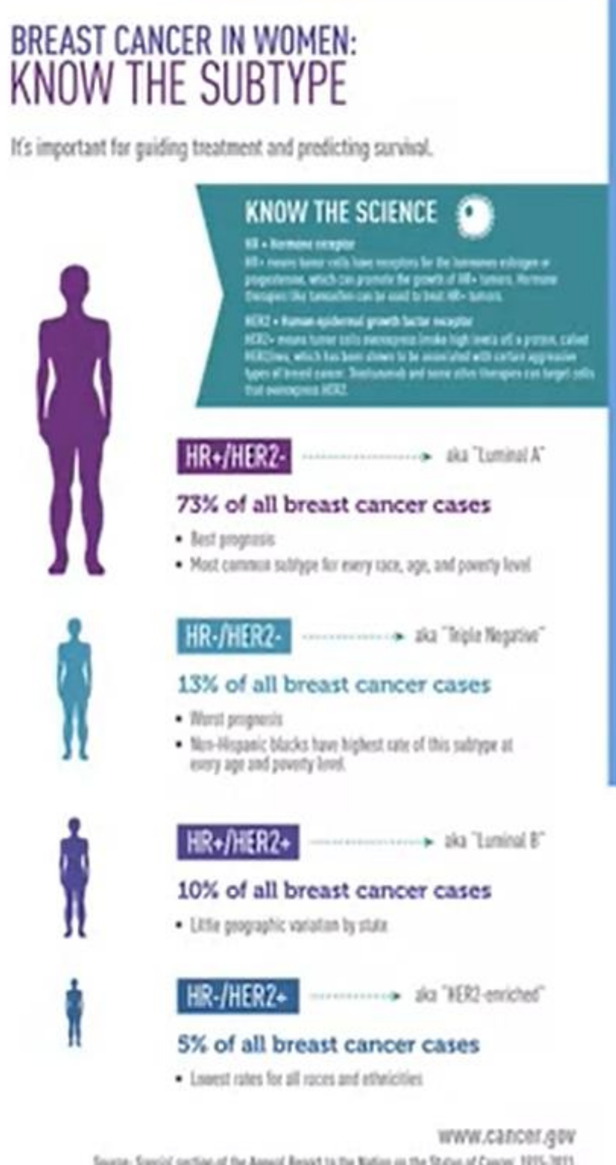


Image courtesy of Herceptin.com

Breast Cancer Subtypes

- Luminal A = HR+/HER2-
  73% of breast cancers
  Best prognosis
- Triple negative = HR-/HER2-
  13% of breast cancers
  Worst prognosis
- Luminal B = HR+/HER2+
  10% of breast cancers
- HER2 enriched = HR-/HER2+
  5% of breast cancers



Relevance to Algorithm Development

- Quantitative Feature effectiveness measure
  - False positive rate and false negative rate
  - How well the feature distinguish two classes
- Facilitate feature selection process
  - Improve detection performance
  - Reduce run-time complexity

Machine Learning Processing Paradigm

- **Sensor**: phenomenology
- **Image Acquisition**: sampling, coding, compression
- **Data Conditioning**: contrast enhancement, object segmentation
- **Feature Extraction**: attributes representing various object types
- **Modeling**: distribution of the feature data
- **Classification**: binary, multi-class

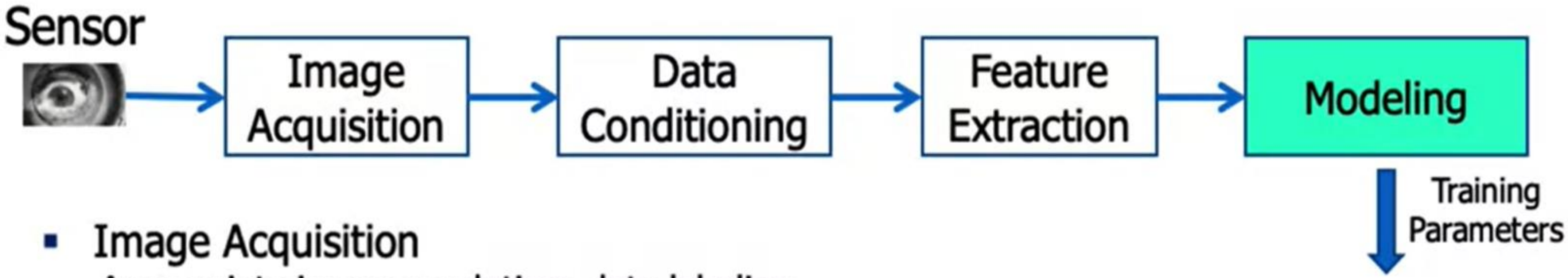
Feature Extraction

- Perform calculations of the attributes of the segmented object
  - Size, shape, texture, etc.
- Features represent the characteristics of the object
  - Effective features to distinguish various object types
- Features needs to be stable
  - Insensitive to slight spatial translation and rotation

Data Conditioning

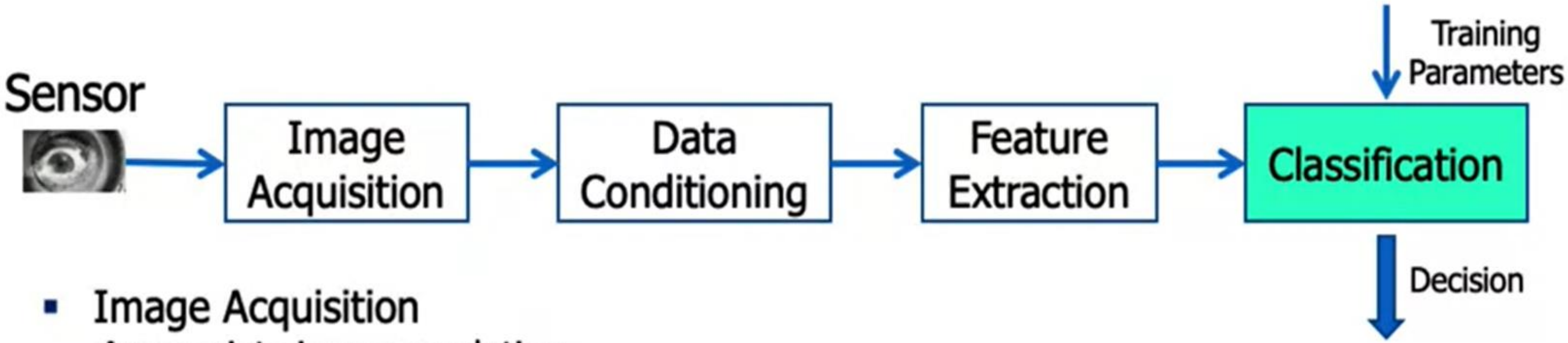
- Speckle (noise) filtering
  - Remove any system noise for further processing
- Object Segmentation
  - Separate object of interest from the background clutter

Algorithm Training



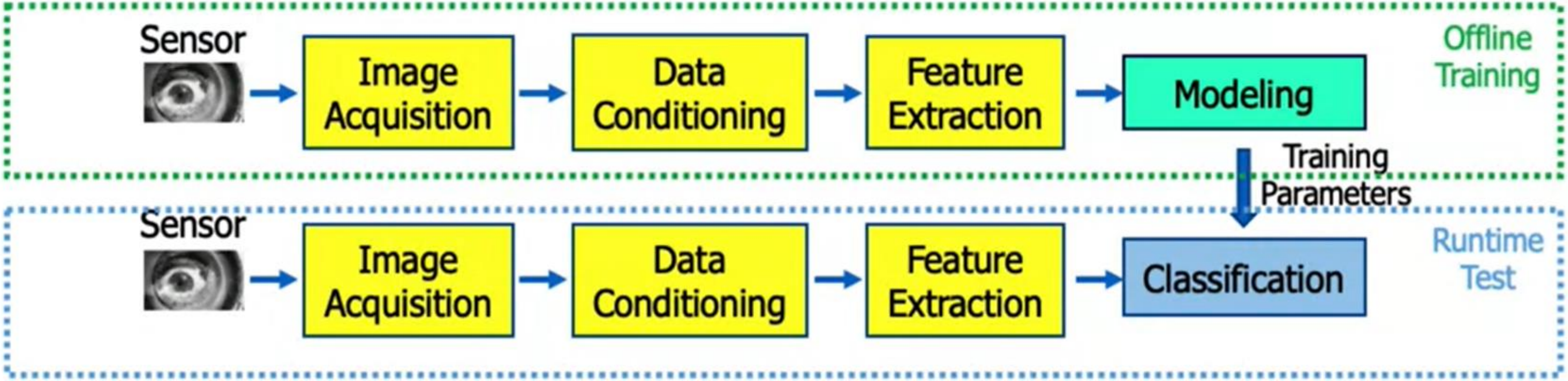
- **Image Acquisition**
  Appropriate image resolution; data labeling
- **Data Conditioning**
  Object relevant information, e.g., object segmentation
- **Feature Extraction**
  A multi-dimensional feature space
- **Modeling**
  Training parameters

Algorithm Test



- Image Acquisition
  Appropriate image resolution
- Data Conditioning
  Object relevant information, e.g., object segmentation
- Feature Extraction
  A multi-dimensional feature space
- Classification
  Training parameters is an input; decision can be either binary or multi-class

Training and Test - Revisit



- The processes, highlighted in yellow, are identical
- Input data are different
  - Training – data labeled with classification types
  - Test – data are unknown, not labeled
- Time requirements
  - Offline training is less critical than run-time test (in general).

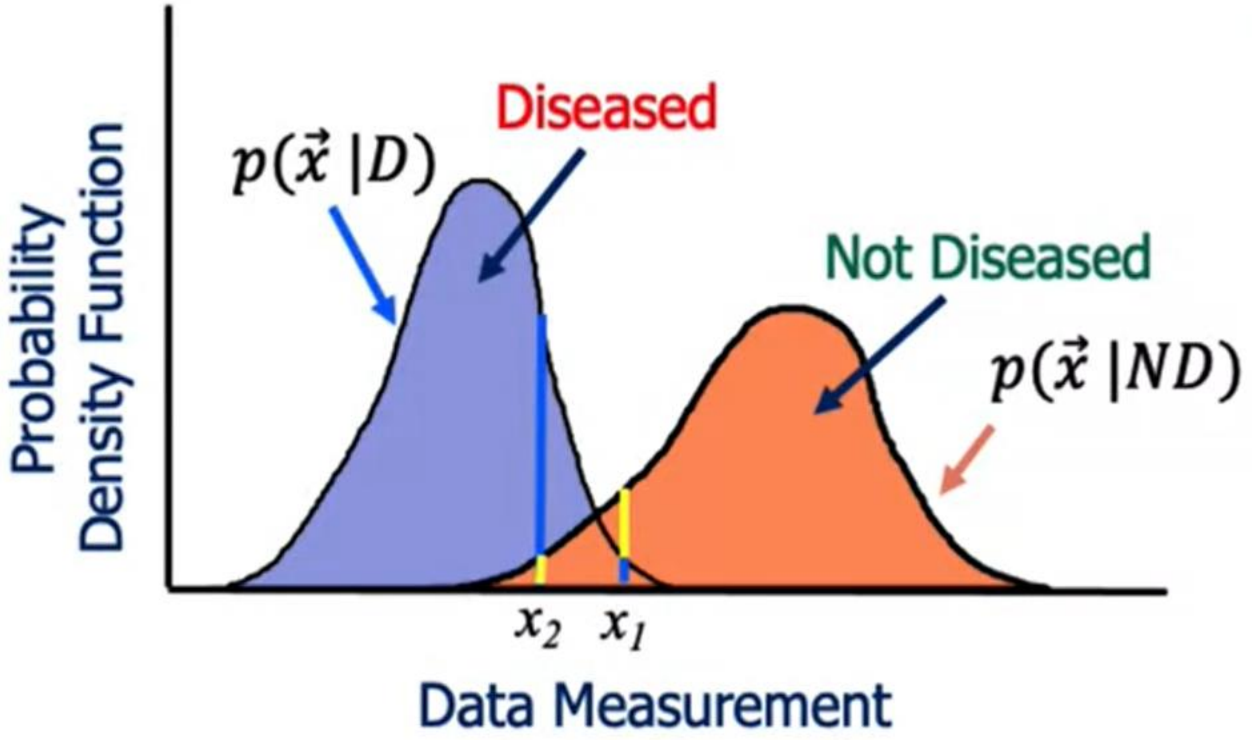
Classification - Bayes' Theorem

$$P(D|\vec{x}) = \frac{P(\vec{x}|D)\ P(D)}{P(\vec{x})} = \frac{P(\vec{x}|D)\ P(D)}{\boxed{P(\vec{x}|D)P(D) + P(\vec{x}|ND)P(ND)}}$$

$$P(ND|\vec{x}) = \frac{P(\vec{x}|ND)\ P(ND)}{P(\vec{x})} = \frac{P(\vec{x}|ND)\ P(ND)}{P(\vec{x}|D)P(D) + P(\vec{x}|ND)P(ND)}$$

- D: Diseased; ND: Not Diseased
- $\vec{x}$: Data measurement
- $P(D|\vec{x})$: *a posteriori* probability of D, given a measurement of $\vec{x}$
- $P(\vec{x}|D)$: Probability of a measurement of $\vec{x}$, given classification type D
- $P(\vec{x}|ND)$: Probability of a measurement of $\vec{x}$, given classification type ND
- $P(D)$: *a priori* probability of Classification type D
- $P(ND)$: *a priori* probability of Classification type ND

Data Distribution & Decision



- Based on the data distributions and Bayes' theorem

$P(D|\vec{x}_2) > P(ND|\vec{x}_2)$ Sensitivity

$P(D|\vec{x}_1) < P(ND|\vec{x}_1)$ Specificity

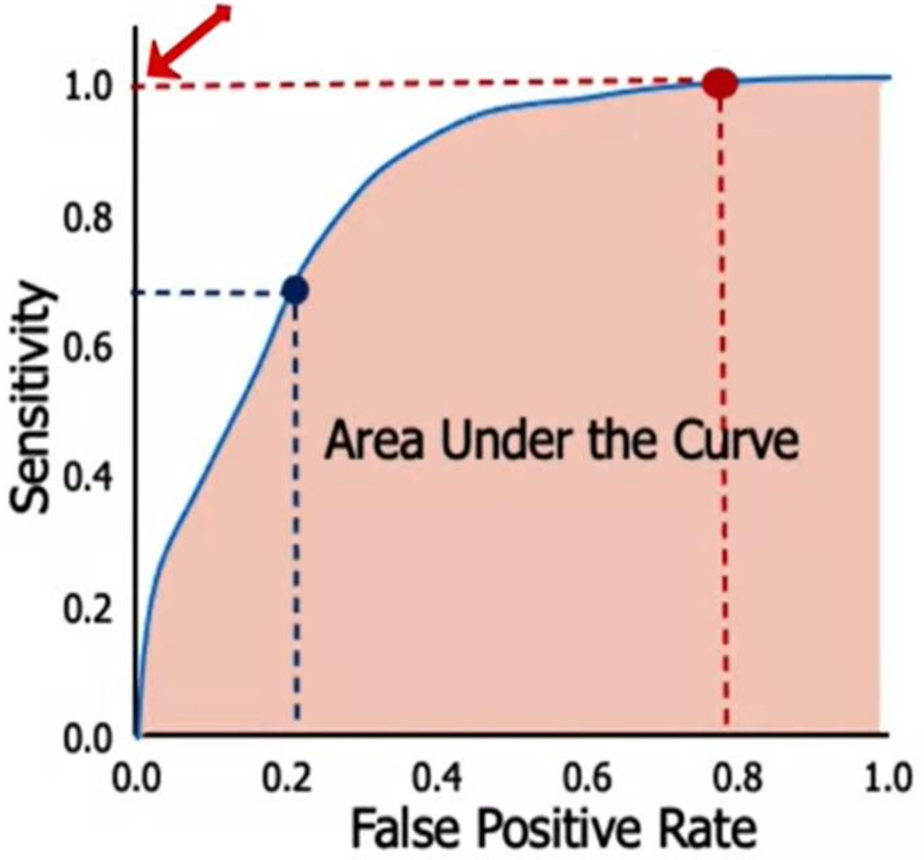
Confusion Matrix

- Table layout to visualize the algorithm performance
  - Row: instances of predicted class
  - Column: instances of actual class
- Sensitivity: 2350/2382 = 98.66%
- Specificity: 2063/2103 = 98.10%
- False positive rate: 40/2103 = 1.90%
- False negative rate: 32/2382 = 1.34%
- Accuracy: (2350+2063)/4485 = 98.39%

Prediction \ Truth	D	ND	Total
D	2350	40	2390
ND	32	2063	2095
Total	2382	2103	4485

ROC Curve

- Employed for a binary classifier, i.e., D or ND
- Plot the sensitivity against the false positive rate (FPR) at various threshold settings.
- Relate to cost/benefit analysis of diagnostic decision making
- Area under the curve (AUC) – a single metric



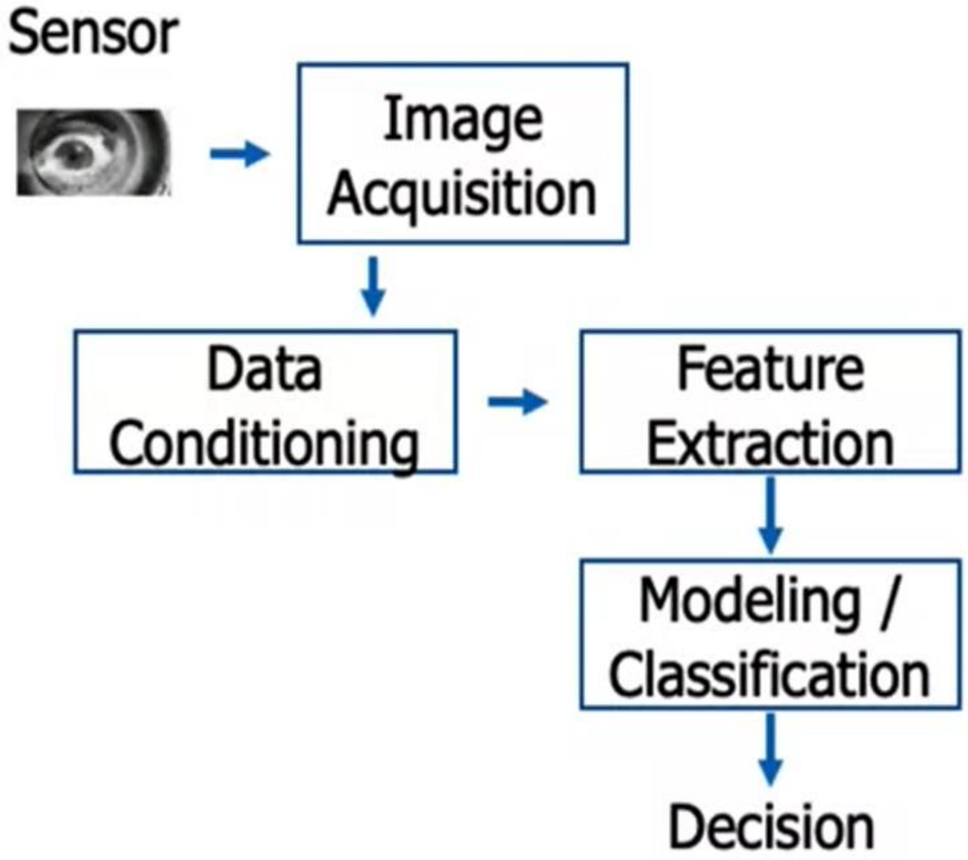
Cross-Validation

- Gain statistical confidence of the performance results
  - More indicative to the performance estimate on a blind test data
  - Minimize the effect due to inadvertent bias occurred in training and test data selection
  - The processing paradigm of each fold is the same
- Not to design the optimal modeling approach
  - Exclusively relevant to performance assessment

Approach

- n-fold cross-validation
  - Randomly partitioned into n equal-sized subsets.
  - One subset is used as the validation / test data, and the remaining $n - 1$ subsets are used as training data.
  - The cross-validation process is then repeated n times
  - The size of the subset determines the confidence of the training model and test results
  - $n = 10$ commonly used
- Collect statistics of Confusion matrices, ROC curves, AUC's, etc. of multiple runs
- Other folds can be used to accommodate sample data volume

Summary



- Six key elements of Machine Learning Processing
  - Modeling for algorithm training
  - Classification for algorithm test
- Input labeled data for training and unknown data for test
- Identical processes in both algorithm training and test
  - Image Acquisition, Data Conditioning, and Feature Extraction

THANK YOU