Computer-Aided Diagnosis: Concepts and Applications

University of Michigan
Medical School

Heang-Ping Chan, Ph.D.
Berkman Sahiner, Ph.D.
Lubomir Hadjiiski, Ph.D.

Department of Radiology
University of Michigan

Outline

- Concepts of CAD
- Current CAD applications
- Computer vision techniques used in CAD
- Issues involved in development of CAD systems
- Future directions

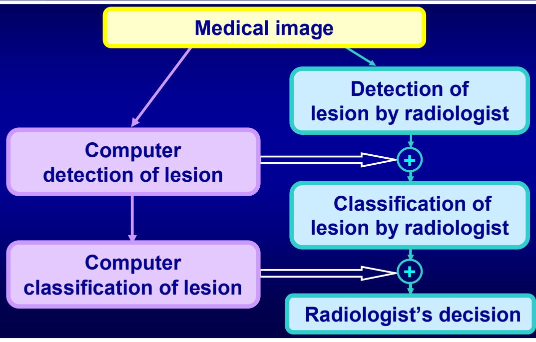
Computer-Aided Diagnosis –

Concepts

Why CAD?

- Interpretation of medical images is never 100% accurate
  - ▹ Detection error
  - ▹ Characterization error
  - ▹ Intra-reader variability
  - ▹ Inter-reader variability

Computer-Aided Diagnosis



CAD

- Provide second opinion to radiologist in the detection and diagnostic processes
- Radiologist is the primary reader and decision maker

CAD Current Applications

CAD Applications

- Breast Cancer
  - Mammography
  - Ultrasound
  - MR breast imaging
  - Multi-modality

CAD Applications

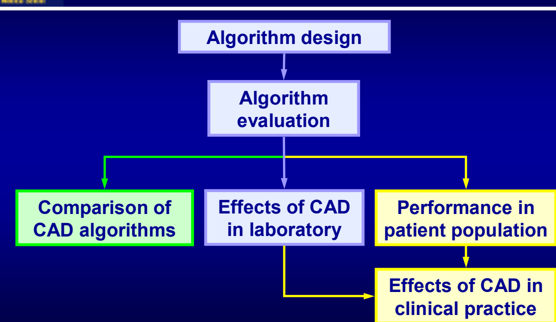
- Lung Cancer
  - Chest radiography
  - Thoracic CT

CAD Applications

- Colon Cancer
  - CT virtual colonography
- Pulmonary embolism
  - CT angiography
- Interstitial disease
  - Chest radiography

Development of CAD System

CAD Algorithm Development Process



CAD System

- Detection
- Characterization
- Combined detection and characterization

General Scheme of Detection System

Preprocessing
↓
Segmentation of Lesion Candidates
↓
Extraction of Features
↓
Feature Selection
↓
Feature Classification
↓
Detected Lesions

General Scheme of Characterization System

Preprocessing
↓
Segmentation of Lesion Candidates
↓
Extraction of Features
↓
Feature Selection
↓
Feature Classification
↓
Likelihood of Malignancy

General Scheme of Combined System

Preprocessing
↓
Segmentation of Lesion Candidates
↓
Extraction of Features
↓
Feature Selection
↓
Feature Classification
↓
Detected Lesions &
Likelihood of Malignancy

Preprocessing

- Trim off regions not of interest
- Suppress structural background, and enhance signal-to-noise ratio
  - Linear spatial filtering
  - Non-linear enhancement
  - Multi-scale wavelet transform
  - etc

Image Segmentation

- Identify objects of interest
  - Contrast
  - Signal-to-noise ratio
  - Gradient field convergence
  - Line convergence – spiculated lesion
  - etc

Object Segmentation

- Extract object from image background
  - Gray level thresholding
  - Region growing
  - Edge detection
  - Active contour model
  - Level set
  - etc

Feature Extraction

- Descriptors of object characteristics
  - Morphological features
    - Shape, size, edge, surface roughness, spiculation, …
  - Gray level features
    - Contrast, density, …
  - Texture features
    - Patterns, homogeneity, gradient, …
  - Wavelet features
  - etc

Feature Selection

- Select effective features for the classification task
- Reduce the dimensionality of feature space
- Reduce the requirement of training sample size

Feature Selection

- Exhaustive search
- Stepwise feature selection
- Genetic algorithm
- Simplex optimization
- etc.

Feature Classification

- Detection
  - ▹ Differentiate true lesions from normal anatomical structures
- Characterization
  - ▹ Differentiate malignant from benign lesions
- Multi-class classification
  - ▹ Differentiate normal tissue, malignant, and benign lesions

Feature Classifiers

- Linear classifiers
- Non-linear classifiers
  - ▹ Quadratic classifier
  - ▹ Bayesian classifier
  - ▹ Neural networks
  - ▹ Support vector machine
  - ▹ etc.

Feature Classifiers

- Other classifiers
  - K-means clustering
  - N nearest neighbors
  - etc.

CAD Algorithm Design

- Application specific
- Designed with training samples

CAD Algorithm Design

- Database
  - Training set
  - Validation set
  - Independent test set

Database for CAD Algorithm Design

Database Collection:

- Quality -
  - Adequate representation of various types of normals and abnormals (of the disease of interest) in patient population
- Quantity -
  - Bias and variance of classifier performance depend on design sample size
  - larger data set ⟶ better training

Sample Size Effect

- Bias of a classifier performance index estimated from a set of design samples, relative to the estimate at infinite sample size:

$$E\{f(N_{train})\} - f(\infty) \cong \frac{1}{N_{train}} + \textit{higher-order terms}$$

Ref. Fukunaga K, *Introduction to Statistical Pattern Recognition*. 2nd Ed. (Academic Press, NY) 1990.

Sample Size Effect on Bias of Classifier Performance

- Monte Carlo simulation study
- Classifier performance index: Area under the ROC curve, A_z
- Dependence of A_z on design sample size
- Comparison of classifiers of different complexity

Classifiers of Different Complexity

Examples:

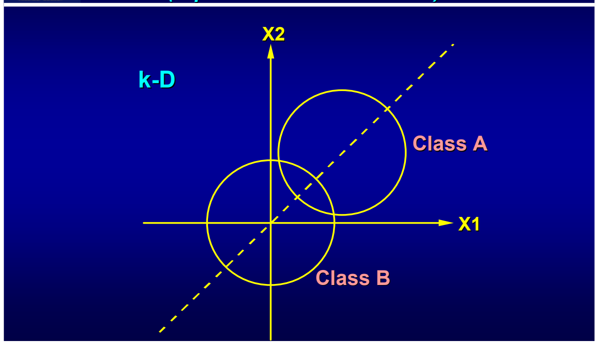
- Linear discriminant classifier
- Quadratic classifier
- Artificial neural network (ANN)

Class Distributions in Feature Space

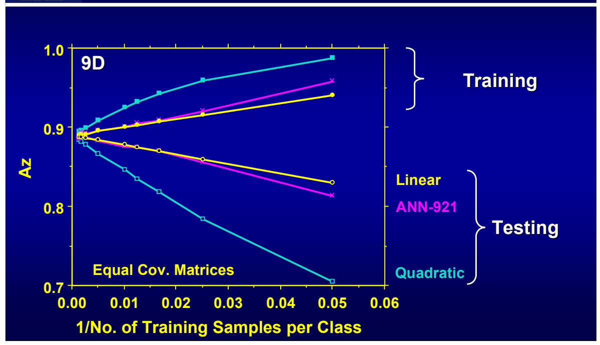
- Two-class classification:
  - ▹ Multivariate normal distributions
    Equal covariance matrices

    Optimal classifier - Linear discriminant
  - ▹ Theoretical optimum at $N_{train} \to \infty$
    $A_z = 0.890$

Class Distributions in Feature Space
(Equal Covariance Matrices)



Comparison of Classifiers

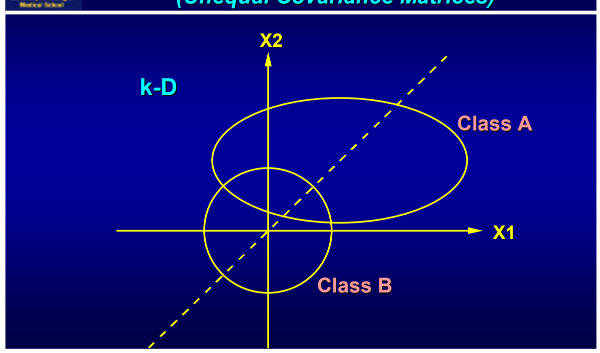


Class Distributions in Feature Space

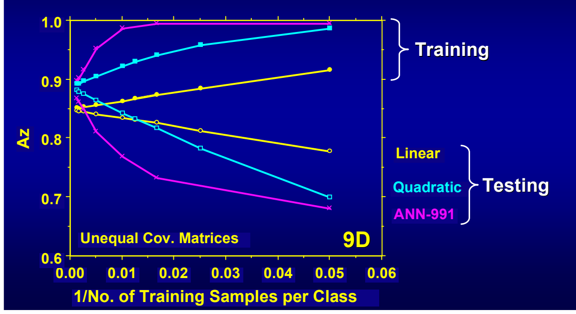
- **Two-class classification:**
  - Multivariate normal distributions
    Unequal covariance matrices

    Optimal classifier - Quadratic discriminant
  - Theoretical optimum at $N_{train} \rightarrow \infty$
    A_z = 0.890

Class Distributions in Feature Space
(Unequal Covariance Matrices)



Comparison of Classifiers

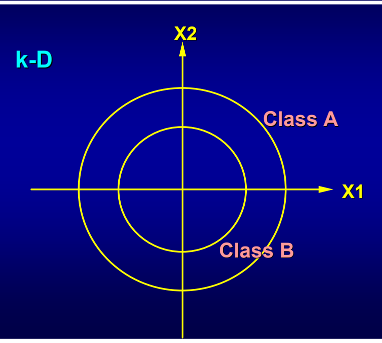


Class Distributions in Feature Space

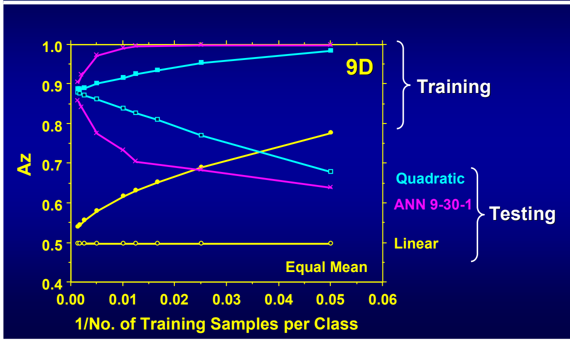
- **Two-class classification:**
  - **Multivariate normal distributions**
    Unequal covariance matrices
    Equal mean

    Optimal classifier - Quadratic discriminant
  - **Theoretical optimum at $N_{train} \rightarrow \infty$**
    A_z = 0.890

Class Distributions in Feature Space
(Equal Mean)



Comparison of Classifiers



Summary - Sample Size Effects

- A classifier trained with a larger sample set has better generalization to unknown samples.
- If design sample set is small, a simpler classifier will provide better generalization.

Summary - Sample Size Effects

- Classification in certain feature spaces requires more complex classifiers such as neural networks or quadratic discriminant.
- When design sample set is small, over-training of neural network will degrade generalizability.

Ref. Chan HP, Sahiner B, Wagner RF, Petrick N. *Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers.* Med Phys 1999; 26: 2654-2668.

Resampling Methods: Evaluation of Classifier Performance with a Limited Data Set

CAD Algorithm Design

- Classifier design is straightforward if available samples with verified truth are unlimited
- Unfortunately, in the real world, especially in medical imaging, collecting data with ground truth is time consuming and costly
- Database is never large enough!

Maximize Trainers

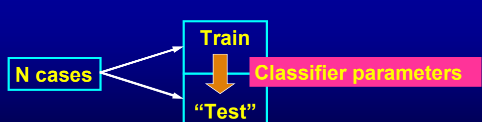
- The larger the number of cases for training, the more accurate the classifier
  - Ideally, use all available cases $x_1, \ldots, x_N$ for training
  - How to evaluate the performance without collecting new cases from the population?

Resampling Methods for CAD Design

- Resampling methods for training and testing CAD algorithm with a limited data set:
  - Resubstitution
  - Hold out
  - Cross validation
  - Leave-one-out
  - Bootstrap
  - etc.

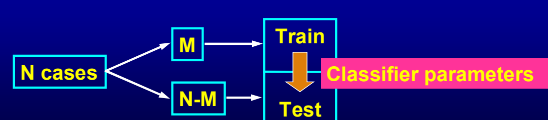
Resampling Scheme 1 - Resubstitution

- Training set = Test set
  - Resubstitution estimate (Consistency Test)
  - Optimistically bias test results



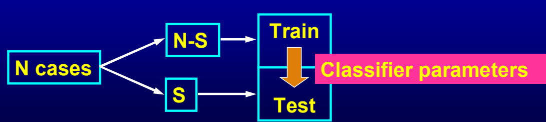
Resampling Scheme 2 - Hold Out

- Partition data set into M trainers and N-M testers
  - Training set independent of test set
  - Training size substantially smaller than N
  - Pessimistically bias test results



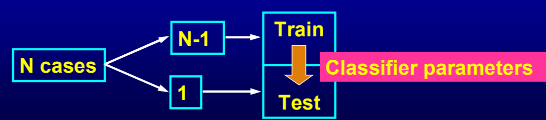
Resampling Scheme 3 - Cross Validation

- **K-fold cross validation**
  - Divide into k subsets of size: S = N / k
  - Train on (k-1) subsets, test on the left-out subset
  - Average results of all k test groups



Resampling Scheme 4: Leave-One-Out

- **Use leave-one-out estimate**
  - Train on N-1 cases
    - Less pessimistic than Hold-Out
  - Test on left-out case, accumulate N test cases
  - Variance?



Resampling Scheme 5 - Bootstrap

- **Bootstrap**
  - Resampling with replacement from available samples
  - Estimate and remove resubstitution bias by using bootstrap samples
    - Ordinary bootstrap
    - 0.632 bootstrap
    - Other variants

Bootstrap Estimate of Resubstitution Bias

- $\underline{x} = (x_1,\ldots,x_N)$, N cases from the population
- Randomly draw a bootstrap sample $\underline{x}^{*b}$
- In the bootstrap world:
  - "Resubstitution" A_z value: $A_z(\underline{x}^{*b}, \underline{x}^{*b})$
  - "Test" A_z value: $A_z(\underline{x}^{*b}, \underline{x})$
  - Bias of resubstitution:

$$w^b = A_z(\underline{x}^{*b}, \underline{x}^{*b}) - A_z(\underline{x}^{*b}, \underline{x})$$

Bootstrap Estimate of Resubstitution Bias

- Repeat for B bootstrap samples
- Average the bias w^b over B replications
- Remove bias from true resubstitution A_z

$$A_z(\underline{x}, \underline{x}) - \frac{1}{B}\sum_{b=1}^{B}[A_z(\underline{x}^{*b}, \underline{x}^{*b}) - A_z(\underline{x}^{*b}, \underline{x})]$$

Resampling Methods

- Bias:
  - Leave-one-out: Pessimistic
  - Bootstrap: Optimistic
- Variance:
  - In general, bootstrap has lower variance than leave-one-out
- Depend on feature space

Evaluation of CAD System Performance and Comparison of CAD Systems

Evaluation of CAD Algorithm Performance

- CAD algorithm performance depends on
  - How well it was trained
  - How it was tested

Testing

- True test performance requires independent test set
- CAD performance depends on properties of test cases
  - Test set should be statistically similar to patient population

Testing

- Comparison of CAD algorithms -
  - use large random samples from patient population
  - Relative performance – use common test set

Validation vs. True Test

- Important issue:

  Even an "independent" test set will turn into a part of the training set if algorithm is optimized based on the test results
  - Optimistic estimate of algorithm performance
- True "test" comparison has to be based on test set unknown to all algorithms

Effects of CAD on Radiologists' Performance – Laboratory Studies

Effects of CAD

- Laboratory observer performance studies
  - ROC methodology
    - *Dorfman-Berbaum-Metz method*
  - ROC with localization
    - LROC - Swensson
    - XFROC, JAFROC - Chakraborty (under development)

Effects of CAD

- Variance component analysis
  - *Beiden-Wagner-Campbell method*
  - Power estimate
  - Sizing of larger trial from pilot study

Effect of CAD

- Findings of laboratory ROC studies
  - CAD improves radiologists' performance in detection and characterization of
    - Breast lesions
    - Lung nodules

CAD in Clinical Practice

Commercial CAD Systems in Mammography

- R2
  - FDA approval in 1998
- CADX (Qualia)
  - FDA approval in Jan. 2002
- ISSI
  - FDA approval in Jan. 2002
- Kodak
  - FDA approvable letter in Nov. 2003

Prospective Study of CAD in Screening Mammography

- Freer *et al.*
  - 19.5% increase in cancer detection
  - 18% increase in call-back rate
- Bandodkar *et al.*
  - 12% increase in cancer detection
  - 12% increase in call-back rate

Prospective Study of CAD in Screening Mammography

- Cupples *et al.*
  - 13% increase in cancer detection
- Helvie *et al.*
  - 10% increase in cancer detection
  - 10% increase in call-back rate

Prospective Study of CAD in Screening Mammography

- Gur *et al.*
  - No statistically significant increase in cancer detection
  - No statistically significant increase in call-back rate

Effects of CAD in Clinical Practice

- CAD system does not need to be more accurate than radiologists to provide useful second opinion
  - Complementary to radiologists is more important

Effects of CAD in Clinical Practice

- CAD system designed as an aid to radiologists should not be misused as a primary reader
  - Loss of advantage of double-reading

Clinical Trial

- Outcome of CAD clinical trial can be affected by
  - Radiologists' vigilance in reading
  - Radiologists' interpretation of CAD marks
  - Study design
- Continued effort to study the effects of CAD in clinical settings needed to resolve differences in outcomes

Future Directions

Future Directions

- **Breast Cancer**
  - Combined mammography/Ultrasound
  - Tomosynthesis
  - Breast CT
  - Dual-energy imaging
  - Dynamic MR breast imaging
  - Risk prediction
  - Prognosis estimation
  - Multi-modality CAD
  - …

Future Directions

- **Similar development of CAD for other diseases**
- **CAD in other image analysis tasks**
  - Optical imaging
  - Molecular imaging
  - etc

Future Directions

- **Beyond the role of second reader**
  - Provide quantitative or other info to radiologists to assist in detection and/or diagnostic decision
  - As a "physician assistant" to screen out negative cases ??