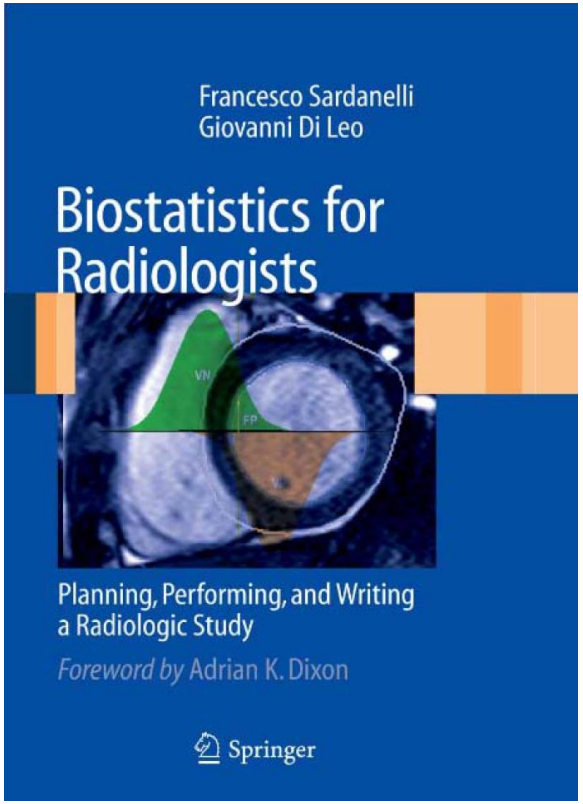
BIOSIGNAL PROCESSING

Topic 2: Automated ECG Arrhythmia Diagnosis

Prof. Yasser Mostafa Kadah – www.k-space.org

Recommended Reference

- *Biostatistics for Radiologists: Planning, Performing, and Writing a Radiologic Study*, by Francesco Sardanelli and Giovanni Di Leo, Springer, 2009.



Evidence-Based Medicine (EBM)

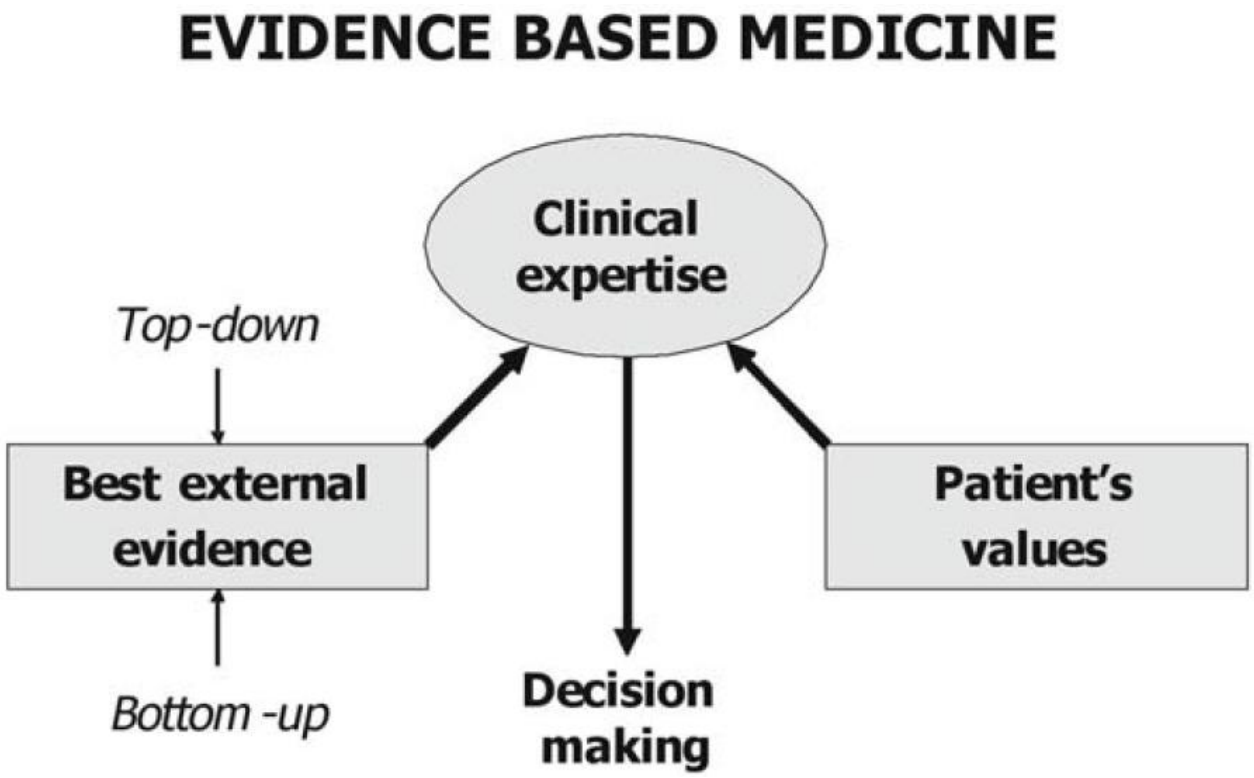
- Clinical practice should be based on the critical evaluation of the results obtained from medical scientific research
- This notion of a clinical practice based on the results (the evidence) given by the research has engendered the discipline of evidence-based medicine (EBM)
  - Also referred to as evidence-based healthcare or evidence-based practice
- In this context the term *evidence* is more closely associated with the concepts of proof, demonstration, or testability than simply with visibility or clarity
- In fact, the general meaning of the new discipline suggests a clinical practice no longer based on bequeathed knowledge, on opinions, impressions, and perceptions, but on demonstrable proofs

Evidence-Based Medicine (EBM)

- Definitions of EBM:
  - "The systematic application of the best evidence to evaluate the available options and decision making in clinical management and policy settings"
  - "Integrating clinical expertise with the best available external clinical evidence from research"
- EBM is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients
- EBM is the use of mathematical estimates of the risk, of benefit and harm, derived from high-quality research on population samples, to inform clinical decision making in the diagnosis, investigation or management of individual patients

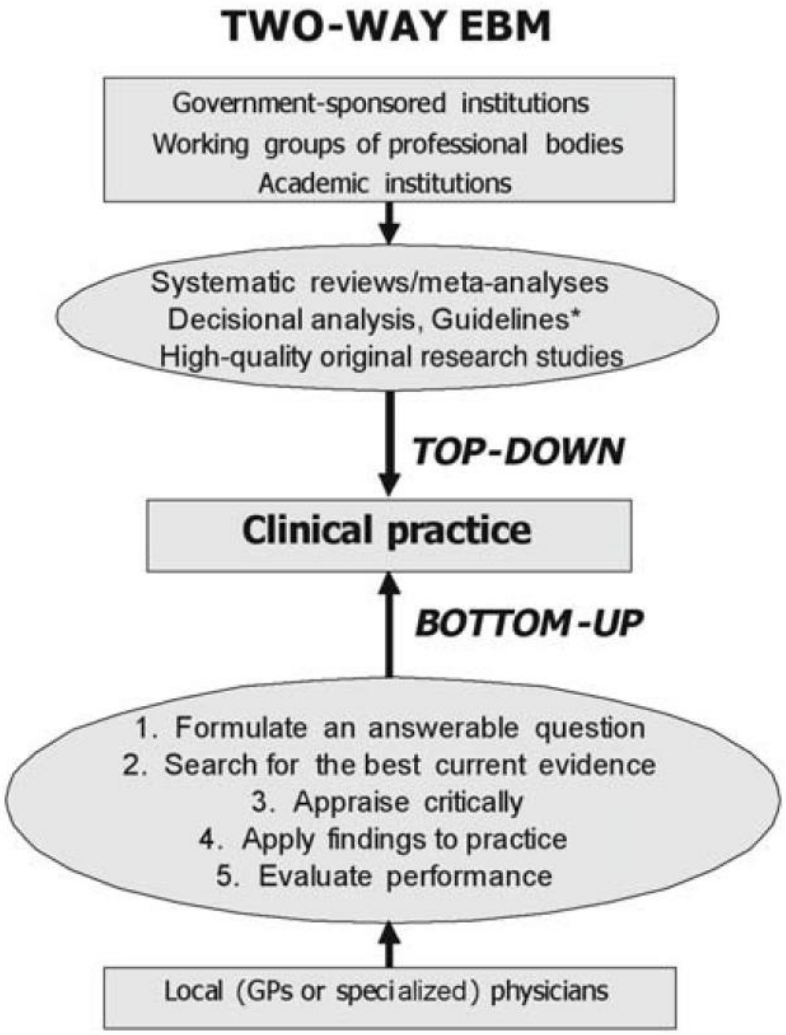
Evidence-Based Medicine (EBM)

- EBM is not only the combination of current best available external evidence and individual clinical expertise. A third factor must be included in EBM: *patient's values and choice*



Evidence-Based Medicine (EBM)

- Two approaches:
  - Top-Down
  - Bottom-Up
- General EBM aim: *"to maximize the quality and quantity of life for the individual patient"*



EBM Limitations

- Judged as unproven
- Very time-consuming (and therefore expensive)
- Narrowing the research agenda and patients' options
- Threatening professional autonomy and clinical freedom
- Large clinical areas have not been sufficiently explored by studies according to EBM criteria
  - Real patients can be totally different from those described in literature
  - Clinical trials not directly applicable

Diagnostic Performance Definitions

- The *performance* of a diagnostic examination can be basically considered as its degree of accuracy, namely its ability to find the subjects affected with a given disease as positive and the subjects not affected with same disease as negative
- The indices which in different ways measure this performance are defined *measures of diagnostic performance* and the studies aimed at measuring the diagnostic performance of an examination or, more often, at comparing the diagnostic performance of two or more examinations, are defined *studies of diagnostic performance*.

Results of an Examination Compared to Reference Standard

- To evaluate the performance of a diagnostic examination, we need to compare its results to a reference standard
  - "Gold standard"
- Typical example: to verify each result of a diagnostic examination for a sample of n patients with pathology report
- Suppose that both doctor and pathologist are required to give a dichotomous judgment (yes/no) about malignancy:
  - True positive
  - False positive
  - True negative
  - False negative

Two-by-Two Contingency Table

		Reference standard	
		Positive	Negative
Radiologic examination	Positive	True positives (TP)	False positives (FP)
	Negative	False negatives (FN)	True negatives (TN)

		Reference standard		
		Affected	Nonaffected	Total
Radiologic examination	Positive	True positives (TP)	False positives (FP)	All positives (TP + FP)
	Negative	False negatives (FN)	True negatives (TN)	All negatives (FN + TN)
Total		All affected (TP + FN)	All nonaffected (FP + TN)	Grand total (TP + FP + FN + TN)

Terminology

- Different terms: *cases, lesions, findings, patients*, and *subjects*
- Consider the study subjects as *patients* when they present with symptoms or signs for a disease
- Name the asymptomatic persons enrolled in population screening program only as *subjects*
- *Statistical Unit* to be considered
  - Patient, organ, segment, or lesion
- Avoid the term *case* in a scientific context
  - Ambiguous because it can be used for both patients and lesions

Measures of Diagnostic Performance

Index	Definition	Formula	Dependence on disease prevalence
1. Sensitivity (or TP rate)	Ability to identify the presence of disease	TP/(TP+FN)	No
2. Specificity (or TN rate)	Ability to identify the absence of disease	TN/(TN+FP)	No
3. Positive predictive value (PPV)	Reliability of the positive result	TP/(TP+FP)	Yes
4. Negative predictive value (NPV)	Reliability of the negative result	TN/(TN+FN)	Yes
5. Overall accuracy	Global reliability	(TP+TN)/(TP+TN+FP+FN)	Yes

Measures of Diagnostic Performance

Index	Definition	Formula	Dependence on disease prevalence
6. FN rate	Proportion between FN and all affected	FN/(FN+TP) = (1 – Sensitivity)	No
7. FP rate	Proportion between FP and all nonaffected	FP/(FP+TN) = (1 – Specificity)	No
8. Positive likelihood ratio	Increase in disease probability when the result is positive	Sensitivity/(1 – Specificity)	No
9. Negative likelihood ratio	Decrease in disease probability when the result is negative	(1 – Sensitivity)/Specificity	No

Note that *disease prevalence* is equal to (TP+FN) / (TP+TN+FP+FN), being the ratio between the number of subjects affected by the disease and the grand total of sample of subjects under investigation.

Sensitivity

- Sensitivity: the ability to identify the presence of a disease
- Example [SARDANELLI ET AL, 2004]: Sensitivity of mammography and dynamic contrast enhanced MRI for the detection of malignant lesions in patients candidate for mastectomy. The authors investigate 99 breasts in 90 candidates for unilateral (n = 81) or bilateral (n = 9) mastectomy. The reference standard, i.e. the pathology exam of the whole excised breast, establishes the presence of 188 malignant lesions. Mammography has 124 true positives and 64 false negatives, MR imaging 152 true positives and 36 false negatives.
  - Sensitivity is 124/(124+64) = 0.66 (66%) for mammography and 152/(152+36) = 0.809 (80.9%) for MRI.
  - The FN rate is 0.340 (34.0%) and 0.191 (19.1%), respectively.
  - Note that the statistical unit is the lesion and not the patient or the breast

Specificity

- *Specificity*: the ability to identify the absence of a disease
- Example [SOBUE ET AL, 2002]: Low-dose CT screening for lung cancer: Of a total of 1611 asymptomatic subjects who undergo the first screening event, 186 are found to be positive and are further studied with high-resolution scanning; 21 of these undergo biopsy. Thirteen subjects are found to be affected by lung cancer. There are no interval cancers (cancers detected between the first and the second screening event). As a result there are 1425 true negatives (=1611-186) and 173 false positives (=186-13)
  - Specificity is 1425/(1425+173)= 0.892 = 89.2%.
  - In this series only one possible lesion is considered for each subject. Lesion and subject are coincident as a statistical unit.

Notes on Different Measures

- Sensitivity and specificity: answers to pretest questions
  - If the patient is affected by the disease, what is the probability that the examination produces a positive result (sensitivity)?
  - If the patient is not affected by the disease, what is the probability that the examination produces a negative result (specificity)?
- Differentiation between *sensitivity* and *specificity* as answers to pre-examination questions and *predictive values* as answers to post-examination questions
- Sensitivity and specificity do not depend on disease prevalence
  - *Prevalence* indicates proportion between number of subjects affected by disease and total number of subjects of an entire population for a defined time interval
  - *Incidence* indicates the number of subjects newly diagnosed as affected by the disease during a defined time interval

Notes on Different Measures

- Optimal situation in clinical practice is when a single diagnostic examination is available with levels of sensitivity or specificity high enough to produce conclusive decision-making
- An examination is SNOUT when its negative result excludes the possibility of the presence of the disease
  - When a test has a very high Sensitivity, a Negative result rules OUT the diagnosis
- An examination is SPIN when its positive result definitely confirms the presence of the disease
  - When a test has a very high Specificity, a positive result rules IN the diagnosis
- In most situations, a certain degree of certainty can be reached with a single diagnostic examination but not a definitive conclusion
  - More than one examination is generally needed

Predictive Values

- Indicate the reliability of positive or negative result and answer questions posed after having performed the examination
  - If the result of the examination is positive, what is the probability that the patient really is affected by the disease (positive predictive value)?
  - If the result of the examination is negative, what is the probability that the patient is really not affected by the disease (negative predictive value)?
- Predictive values depend on disease prevalence
  - Positive predictive value is directly related to disease prevalence
  - Negative predictive value is inversely related to disease prevalence
- Reliability of reports also depends on patient selection by the referring physicians

Predictive Values

- A disease can affect a patient with different levels of severity (or stage) and the probability of a positive result of an examination increases with the level of severity.
  - Level of severity should be lower in subjects in whom the disease is diagnosed with periodic screening than that found in symptomatic subjects in whom the disease is diagnosed in clinical practice
- In this way we observe a direct influence on sensitivity and specificity: they are higher in symptomatic subjects than in asymptomatic subjects in whom the disease is more likely in an early stage

Overall Accuracy

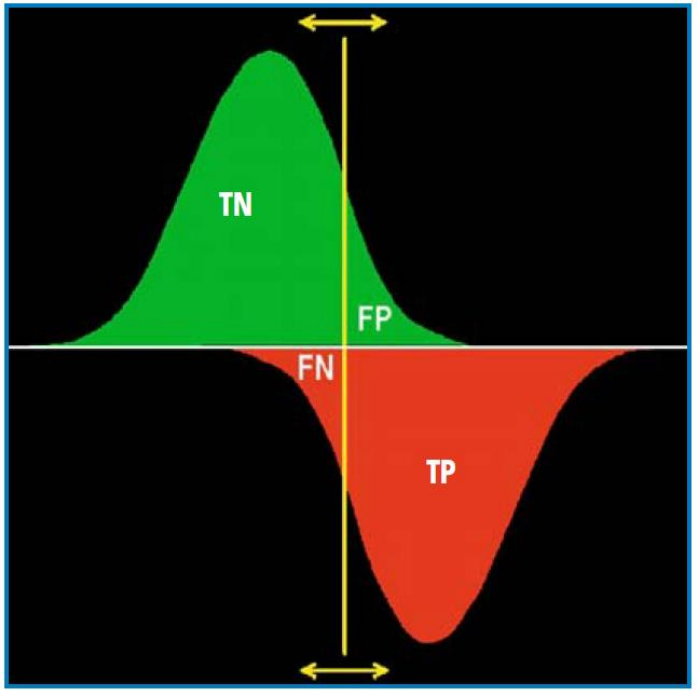
- Ability to correctly identify the presence and the absence of a disease
- It answers the question: what is the probability of a correct result?
  - Somewhat like a global index of diagnostic performance, but its linear distribution ranges between the sensitivity value and the specificity value.
  - It approaches the higher of the two with increasing disease prevalence and approaches the lower of the two with decreasing disease prevalence.
- In practice, it is a kind of "mean" between sensitivity and specificity which is weighted for disease prevalence
  - Dependence on disease prevalence is the feature shared with the predictive values

Notes

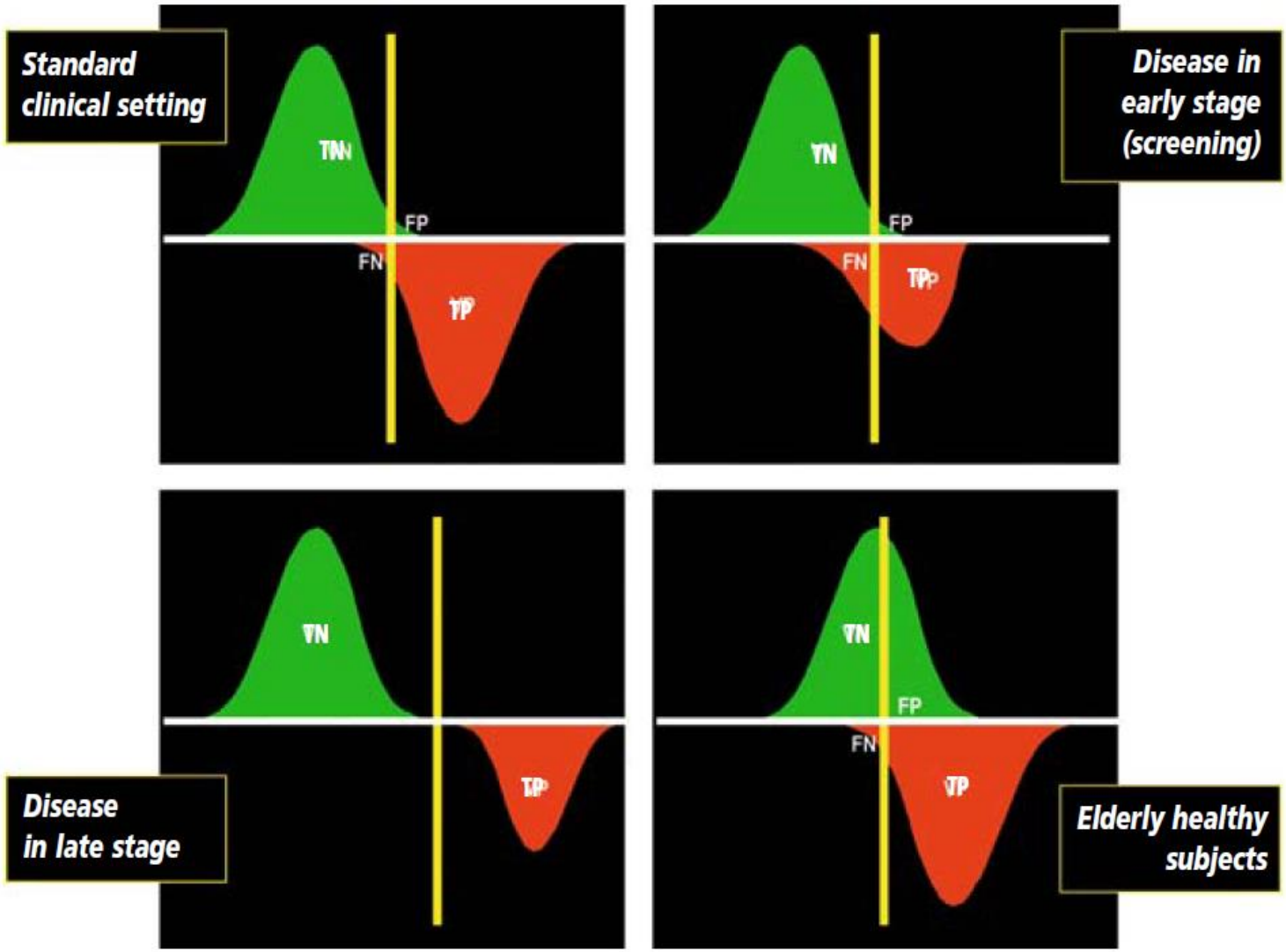
- Sensitivity and specificity may appear to be properties intrinsic to the examination and independent of the disease we would like to confirm or to exclude, which is not the case
  - Always relate measures of diagnostic performance to defined disease
- *Clinical Diagnosis* vs. *Screening*
  - Clinical diagnosis (symptomatic subjects): try to use examinations with a high sensitivity, even in the presence of a relatively low specificity
  - Screening (asymptomatic subjects): try to use examinations with a high specificity, also accepting a trade-off for sensitivity
  - While in clinical diagnosis, major priority is to diagnose symptomatic disease (possibly advanced stage), in screening, diagnosis of asymptomatic disease must be balanced by need of limited amount of useless diagnostic work-up in screened population

Thresholds and Cutoff

- Both the doctor and the pathologist are required to give a dichotomous judgment (yes/no) about the malignancy of the lesion
  - Problem is related to the threshold we choose for our diagnostic decision, i.e. the cutoff
  - Above the cutoff a signal sign is considered predictive of a disease
  - If we lower the cutoff, we gain in sensitivity and lose in specificity
  - If we raise the cutoff, we gain in specificity and lose in sensitivity
- The cutoff could be optimized by choosing the level which minimizes total errors (sum of false negatives and false positives)

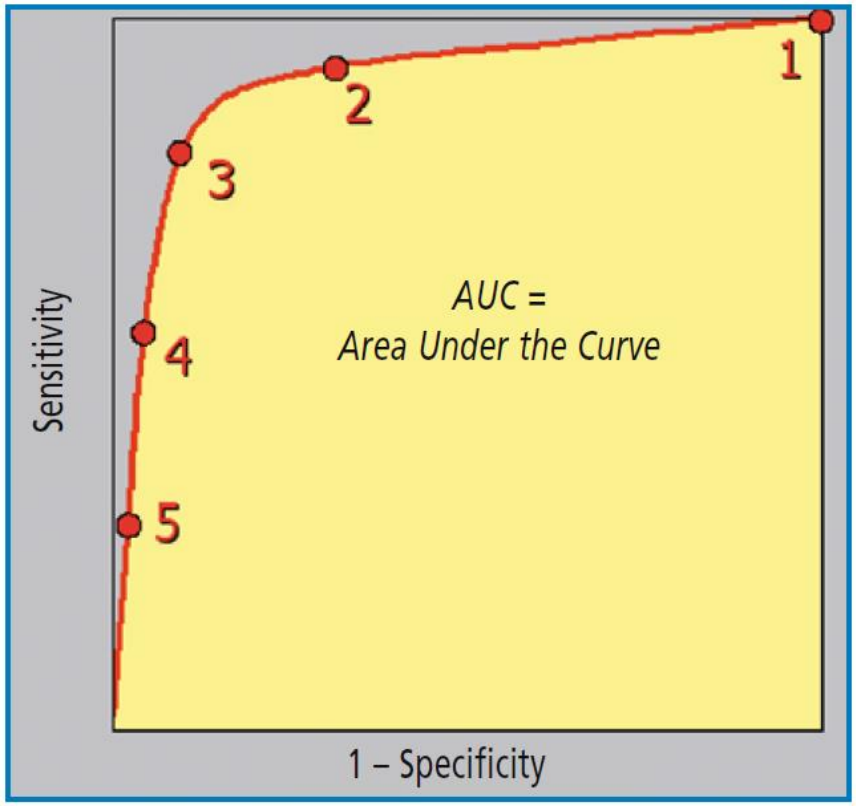


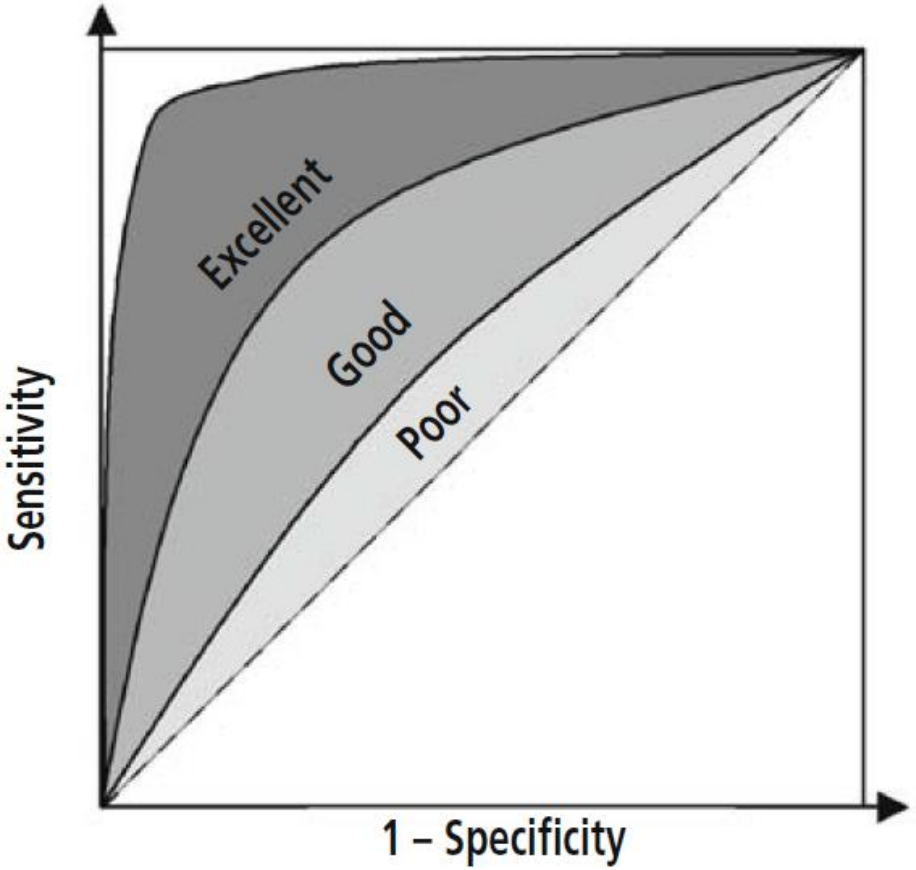
Role of Disease Spectrum



Receiver Operator Characteristic (ROC) Curve

- Sensitivity is graphed on the y-axis and (1– Specificity) on the x-axis using different cutoffs





Computer Aided Diagnosis (CAD)

- Computer-aided diagnosis (CAD) is defined as a diagnosis made by a doctor who uses the output of a computer analysis of the signals when making his/her interpretation
- Computer-aided detection (CADe)
  - Identify and mark suspicious segments in signal
  - Goal of CADe is to help doctor avoid missing a problem
  - Used with “Screening”
- Computer-aided diagnosis (CADx)
  - CADx help doctor decide the type of an existing problem
  - Report the likelihood that the signal exhibits a life-threatening disease
  - Used with “Classification”

Goals of CAD

- Improving doctors' performance
  - Better accuracy
- Reduce intra- and inter-variability of doctors
- It is also hoped that CAD can improve doctors' productivity
  - Faster diagnosis

General Methodology

- Steps that are essential in all CAD systems are shown

CADe as Pre-screen

- As the performance of CADe schemes has improved, the number of false detections has decreased while the sensitivity has improved or remained the same
- The performance of clinical CADe systems is now at the point where it may be possible to use CADe as a pre-screener of screening mammograms.
  - That is, before the cases are reviewed by a doctor, a CADe scheme is run and cases without any detections are considered normal and not read by the doctor
  - Goal here is to reduce the doctors' workload
  - CADe scheme will pre-screen out "easy" cases
  - Problem: doctors may become more exhausted, as every case will be "difficult"

Concurrent Reading with CADe

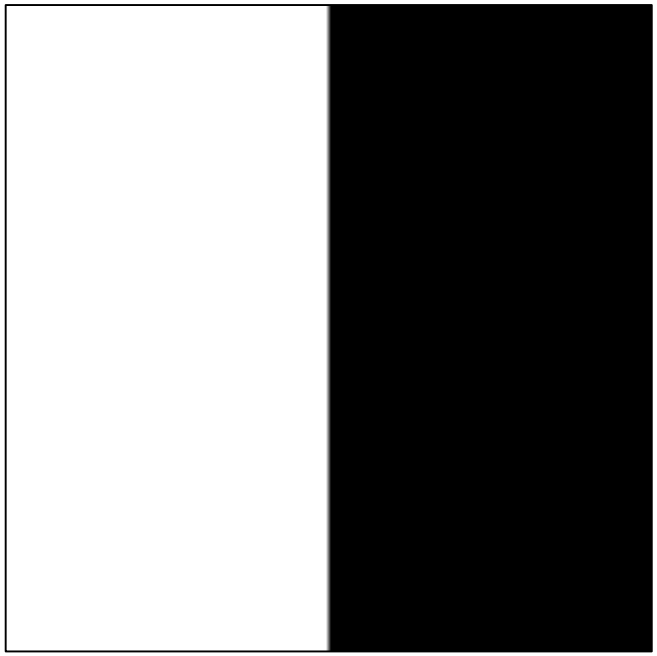
- Currently, CADe is used as a second reader, where the doctor fi rst reads the signals without CADe and then re-reads with the knowledge of the CADe findings
- An alternative paradigm is to have CADe as the first reader and have the doctor verify (i.e., accept or reject) each CADe prompt
  - For this to be effective, the sensitivity of the CADe system must be very high
  - For mass detection, which is a very difficult task, sensitivity is not high enough
  - For clustered microcalcification detection, sensitivity is approximately 98%
- At this level of performance, it is conceivable that CADe could be the first reader, which is sometimes referred to as *concurrent reading*
  - CADe finds clustered microcalcifications at very high rate

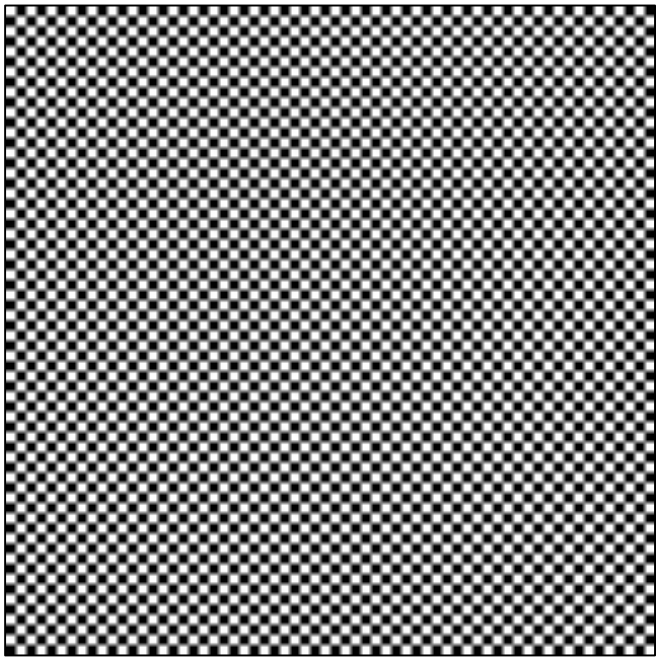
Feature Extraction

- Features are quantitative measures of texture that describe salient characteristics in the signal
- Spatial domain features
  - First order statistical or Histogram-based features
  - Higher order statistical features (HOSA)
- Transform domain features
  - Fourier descriptors
  - Wavelet features

First Order Statistical Features

- Depend only on pixel values and independent of spatial distribution of pixels
  - Example: images below have same first order features





First Order Statistical Features

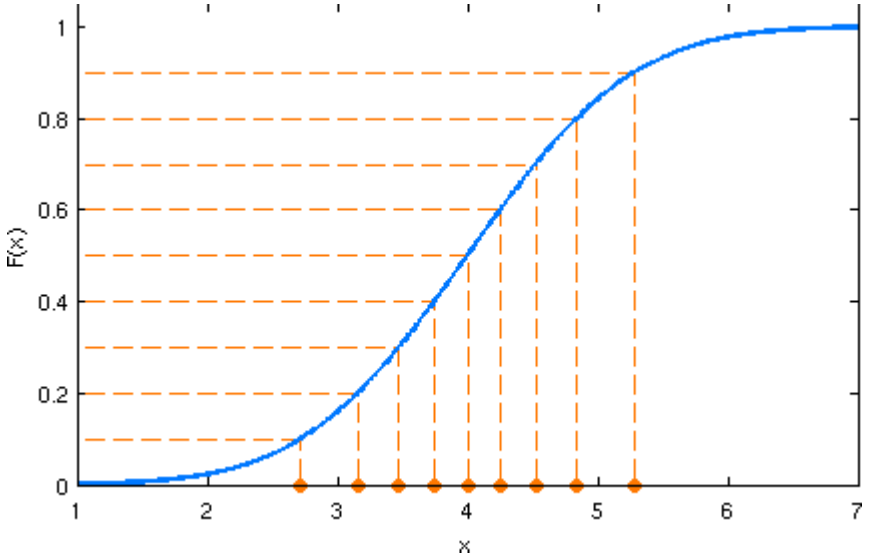
- Example features:

Moment	Expression	Measure of texture
Mean	$m = \sum_{i=0}^{L-1} z_i \; p(z_i)$	A measure of average intensity
Standard deviation	$\sigma = \sqrt{\mu_2 \; (z)} = \sqrt{\sigma^2}$	A measure of average contrast
Smoothness	$R = 1 - 1/(1+\sigma^2)$	Measures the relative smoothness of the intensity in a region.
Third Moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \; p(z_i)$	Measures the skew ness of a histogram
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$	Measures the uniformity. of intensity in the histogram
Entropy	$e = -\sum_{i=0}^{L-1} p(z_i) \; log_2 \; p(z_i)$	A measure of randomness

First Order Statistical Features

- Quantiles (or percentiles) are points taken at regular intervals from the cumulative distribution function (CDF) of a random variable
  - Dividing ordered data into essentially equal-sized data subsets is the motivation for q-quantiles; quantiles are the data values marking the boundaries between consecutive subsets
  - Common to use 0.1, 0.2, …, 0.9 as features
  - 0.5-quantile is the median

Matlab function "quantile(data,p)":
Returns quantile of the values in "data" for the cumulative probability or probabilities "p" in the interval [0,1]



Higher Order Statistical Features

- Depend on pixel values and their temporal inter-relationships.
- Example: a tabulation of how often different combinations of signal values (amplitudes) occur in a specific signal duration
  - Constructed as a matrix by observing pairs of point signal intensities with time *T* from each other and incrementing the matrix position corresponding to the amplitudes of both points
  - Different realization depending on distance

Higher Order Statistical Features

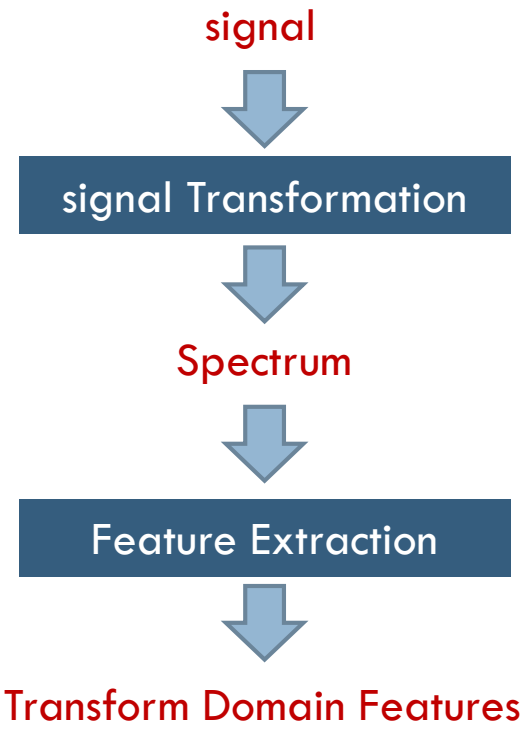
Descriptor	Explanation	Formula
Maximum probability	Measures the strongest response of C. The range of values is [0,1]	$\max_{(i,j)}(P_{ij})$ (2.5)
Correlation	A measure of how correlated a pixel is to its neighbour over the entire image. Range of values is 1 to -1, corresponding to perfect positive and perfect negative correlations. This measure is not defined if either standard deviation is zero.	$\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(i-m_r)(j-m_c)P_{ij}}{\sigma_r\sigma_c}$ $\sigma_r \neq 0; \sigma_c \neq 0$ (2.6)
Contrast	A measure of intensity contrast between a pixel and its neighbour over the entire image. The range of values is 0 (when C is constant) to $(k-1)^2$	$\sum_{i=1}^{k}\sum_{j=1}^{k}(i-j)^2 P_{ij}$ (2.7)

Higher Order Statistical Features

Uniformity (also called Energy)	A measure of uniformity in the range [0,1]. Uniformity is 1 for a contrast image.	$\sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}^2$ (2.8)
Homogeneity	Measures the spatial closeness of the distribution of elements in C to the diagonal. The range of values is [0,1], with the maximum being achieved when C is a diagonal matrix.	$\sum_{i=1}^{k}\sum_{i=1}^{k} \frac{P_{ij}}{1+\lvert i-j\rvert}$ (2.9)
Entropy	Measures the randomness of the elements of C. The entropy is 0 when all P_{ij}'s are 0 and is maximum when all P_{ij}'s are equal.	$-\sum_{i=1}^{k}\sum_{i=1}^{k} P_{ij}\log_2 P_{ij}$ (2.10)

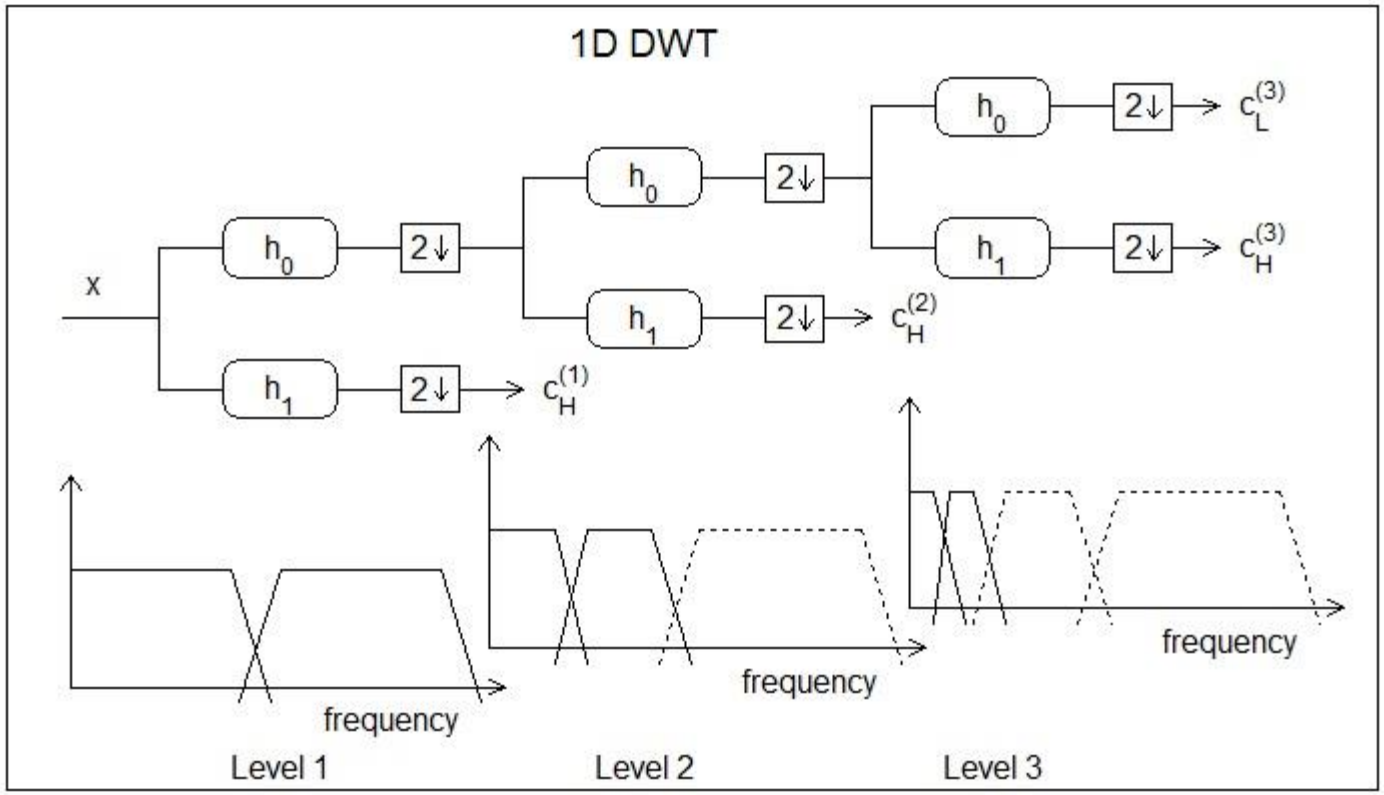
Transform Domain Features

- Features computed AFTER transformation to another domain
  - Discrete Fourier transform
  - Discrete wavelet transform



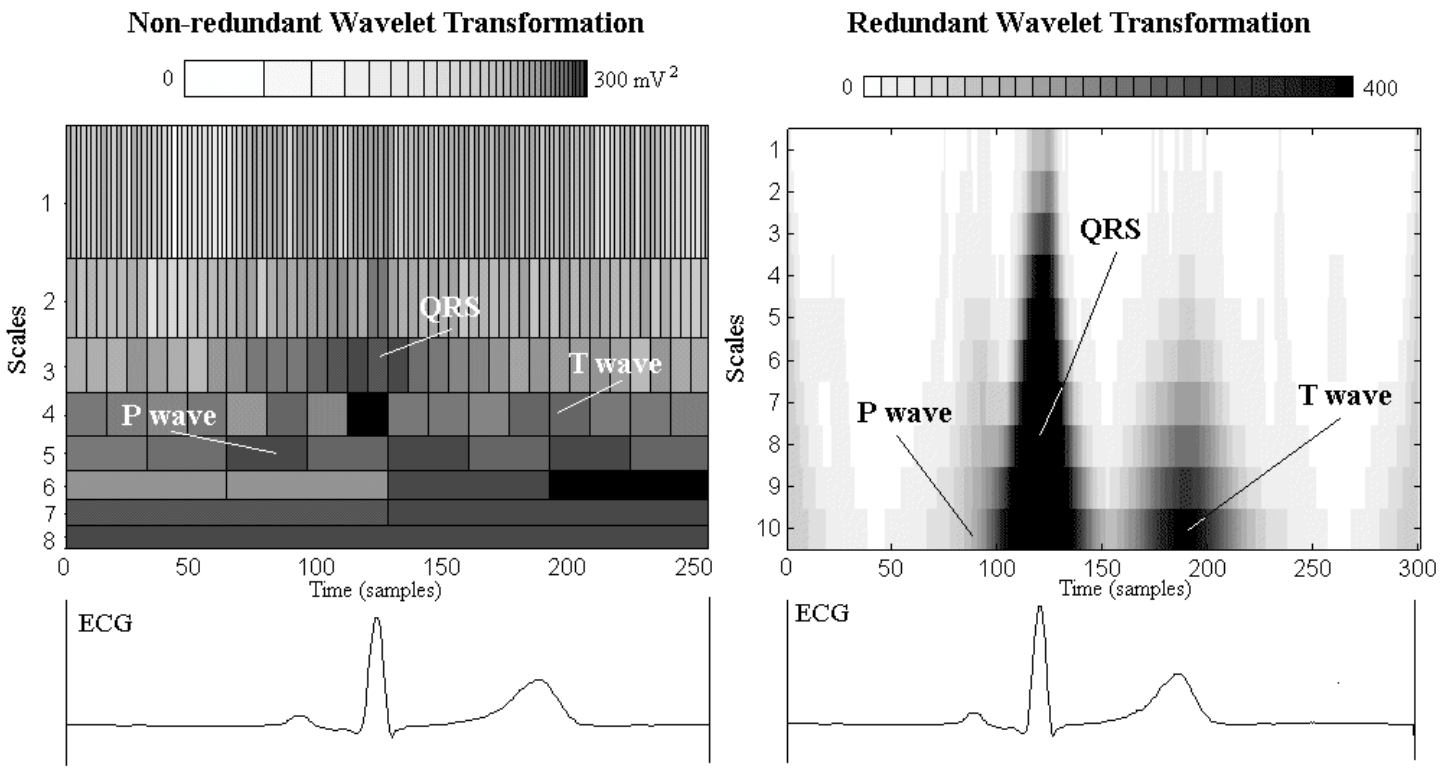
Transform Domain Features

- Example: Wavelet decomposition
  - Selection of basic wavelet type and size and number of levels



Transform Domain Features

- Fourier and Wavelet decomposition of ECG signal
  - Different details in different scales
  - Consider as multiple new signals and do feature extraction



Final Feature Extraction Notes

- Newer features based on signal modeling
  - Markov random model
  - ARMA model
  - Fractal models
- Next step: feature selection
  - Not all features are correlated with disease
  - Must include only relevant features to avoid misclassification
  - Normalization of feature values is necessary preprocessing step before training classifiers

Feature Selection

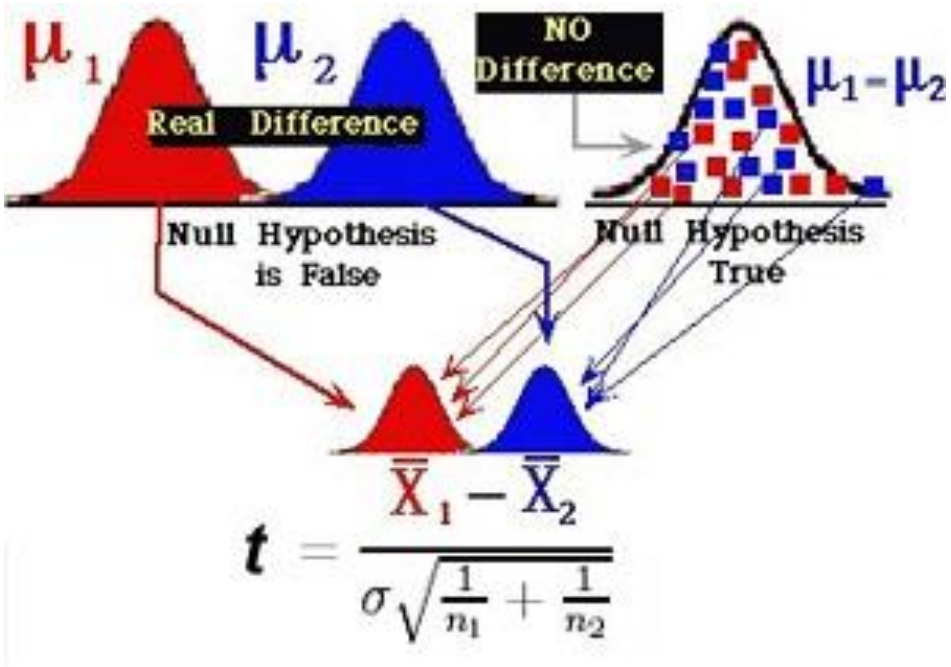
- Feature extraction and selection in pattern recognition are based on finding mathematical methods for reducing dimensionality of pattern Representation
- A lower-dimensional representation based on independent pattern descriptors is important
  - Plays crucial role in determining separating properties of pattern classes
- The choice of features, attributes, or measurements to be included has an important influence on:
  - (1) accuracy of classification
  - (2) time needed for classification
  - (3) number of examples needed for learning
  - (4) cost of performing classification

Feature Selection Goal

- Features extracted from signals need not represent significant information to diagnosis
  - May describe aspects of no relevance to the pathology of interest
  - May vary a lot with acquisition settings (pose, processing, etc.)
- Several problems should be mitigated in feature selection
  - Features that do not correlate with pathology
  - Features that are not independent
- Building classifiers with features that are not properly selected will cause problems in the training phase and will not yield the best overall classification accuracy

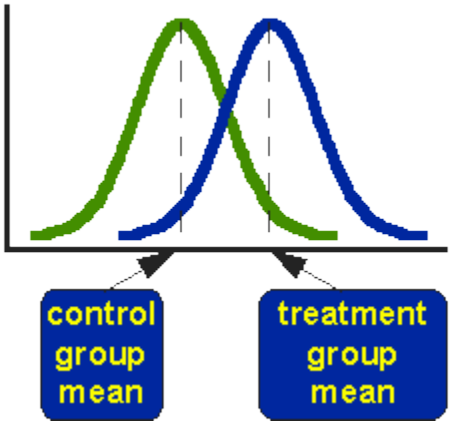
Statistical Significance of Features

- Idea: if the feature changes consistently with pathology, then the hypothesis of a statistically significant difference between the set of values for normal and abnormal cases will be true
  - Inferential Statistical tests like Student t-test should detect a difference



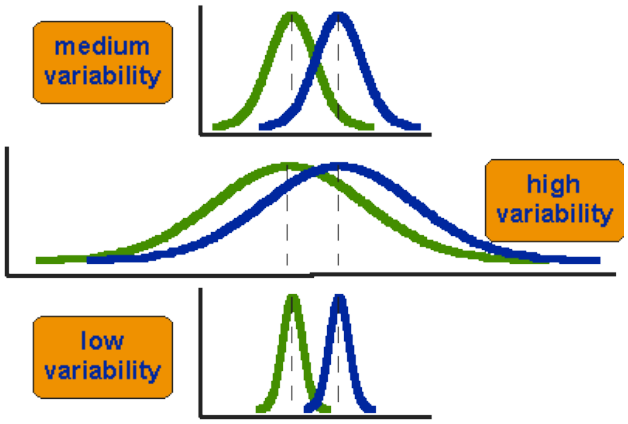
Statistical Significance of Features

- Student t-test steps:
  - Consider a particular feature of interest
  - Divide the values into two sets for normal and abnormal cases
  - Compute the mean and standard deviation for both sets
  - Use the t-test to compute the p-value of the null hypothesis that both sets do not have a statistically significant difference
  - The feature is suitable if the p-value is small (e.g., 0.05, 0.01, etc.)



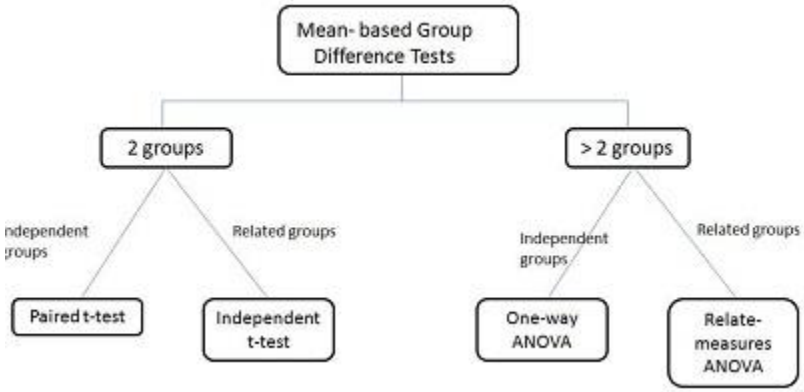
Statistical Significance of Features

- Important to keep in mind that large difference in value does not mean statistical significance
  - Data dispersion is a key factor



- General form: multiple groups
  - Diagnosis not detection
  - More general inferential statistics

- Nonparametric methods
  - Kolmogorov-Smirnov test
  - Wilcoxon signed rank test



Assessment of Feature Independence

- Some features may end up being dependent
  - Example: feature computed as a constant factor of another
  - Only one of them should be included as the input to classification stage
- Several methods can help identify such dependence
  - Pearson's linear correlation coefficient
  - Principal component analysis

Pearson's linear correlation coefficient

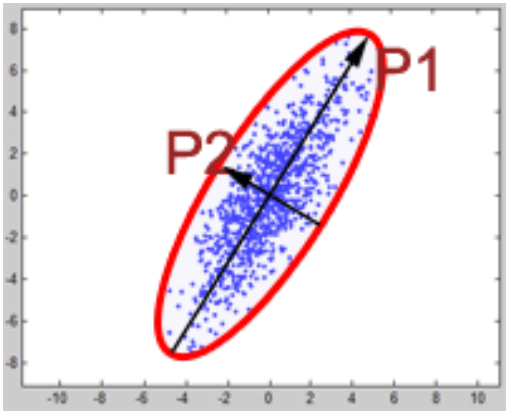
- Computes the correlation between a given pair of features
- Computing formula:

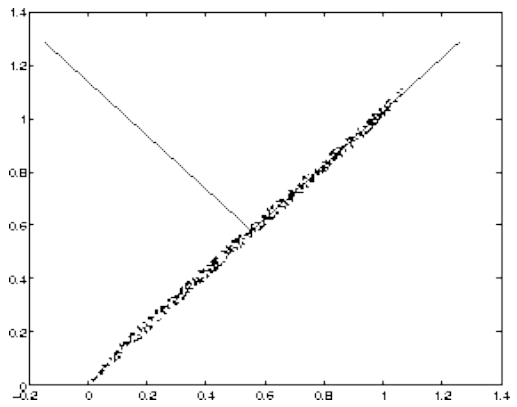
$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

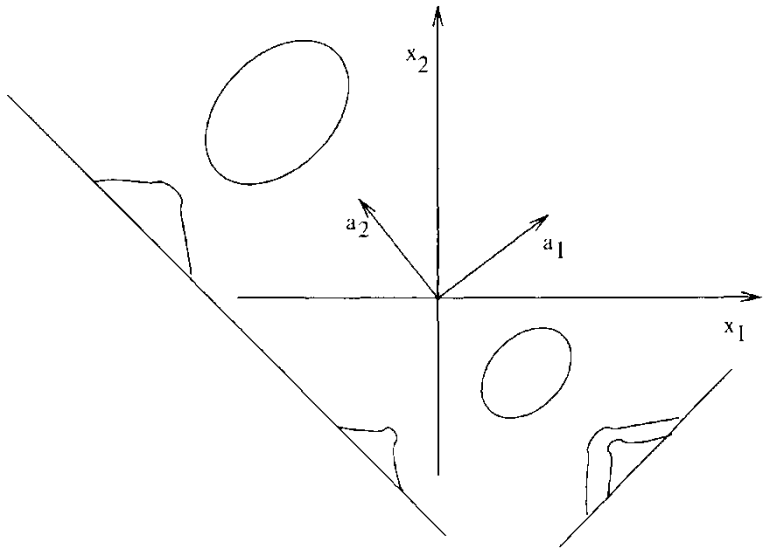
- The value of r lies between -1 and 1, inclusive
  - Value of 1 is called "complete positive correlation": straight line y = c x
  - Value of -1 is called "complete negative correlation": straight line y= -c x
  - Value of r near zero indicates that the variables x and y are *uncorrelated*

Principal Component Analysis

- Computes the eigenvalue decomposition of the matrix of features
  - Rank of matrix = number of independent features
  - Directions of principal components may have different performance in classification







Retrospective Assessment of Features

- Retrospective: evaluation after seeing the classification results based on these features
- Basic idea: use for classification and then choose the features that produce the best results
  - Exhaustive search
  - Branch and Bound Algorithm
  - Max-Min Feature Selection
  - Sequential Forward and Sequential Backward Selection
  - Fisher's Linear Discriminant

Exhaustive Search

- Let y with $y = [y_1, y_2, \ldots, y_D]$ be a pattern vector, exhaustive search selects the **d** best features out of the maximal available features D as to minimize the classification error
- The resulting number of total combinations is:

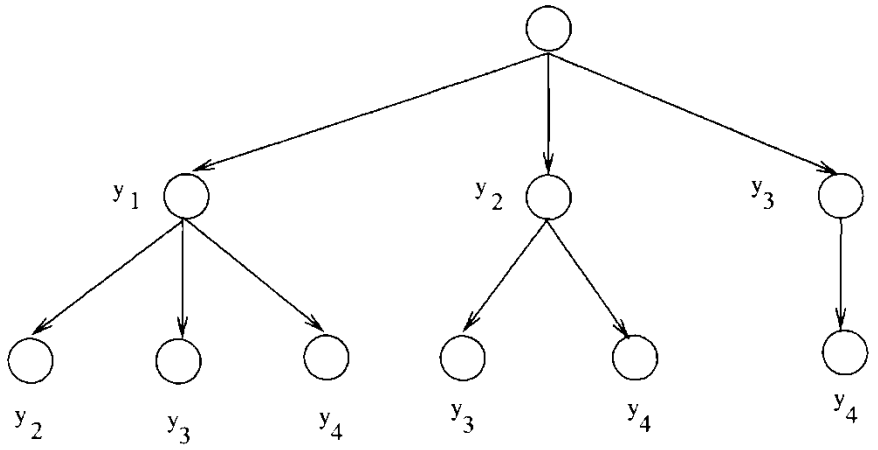
$$\binom{D}{d} = \frac{D!}{(D-d)!d!}$$

- Main advantage: guaranteed best answer
- Main disadvantage of exhaustive search is that the total number of combinations increases exponentially with the dimension of the feature vector

Branch and Bound Algorithm

- Solves the problem of huge computational cost associated with the exhaustive search
- New technique to determine the optimal feature set without explicit evaluation of all possible combinations of d features
- This technique is applicable when the separability criterion is monotonic: $\text{If} \quad \chi_1 \subset \chi_2 \subset \cdots \chi_D, \quad \text{then} \quad J(\chi_1) \leq J(\chi_2) \leq \cdots J(\chi_D)$

- Combinatorial optimization
- Efficient computation



Classification: Assumptions

- There is a visible difference between normal and abnormal signals
  - If you cannot see consistent differences, you cannot program a CAD system to see them
  - Training set with normal and abnormal cases of interest
- Features selected describe such difference effectively
  - Irrelevant features will only confuse and misguide the CAD system
- Normal and abnormal cases form distinct clusters that are somewhat apart according to a distance measure
  - Cases from the same pathology are represented by points in feature space that with smaller distance separation than with cases from other pathologies
  - "Intra-cluster" distance is significantly smaller than "Inter-cluster" distance

Classification: Model

- Parametric classification
  - Assumed a certain distribution for data clusters (e.g., Gaussian)
  - Estimates model parameters from the data
  - Uses this a priori information to design the classification method and estimate its parameters
  - Good if model is correct but bad if not (difficult to know ahead)
- Non-Parametric classification
  - Does not assume or impose any model on the data
  - "Model-Free" or "Data-Dependent"

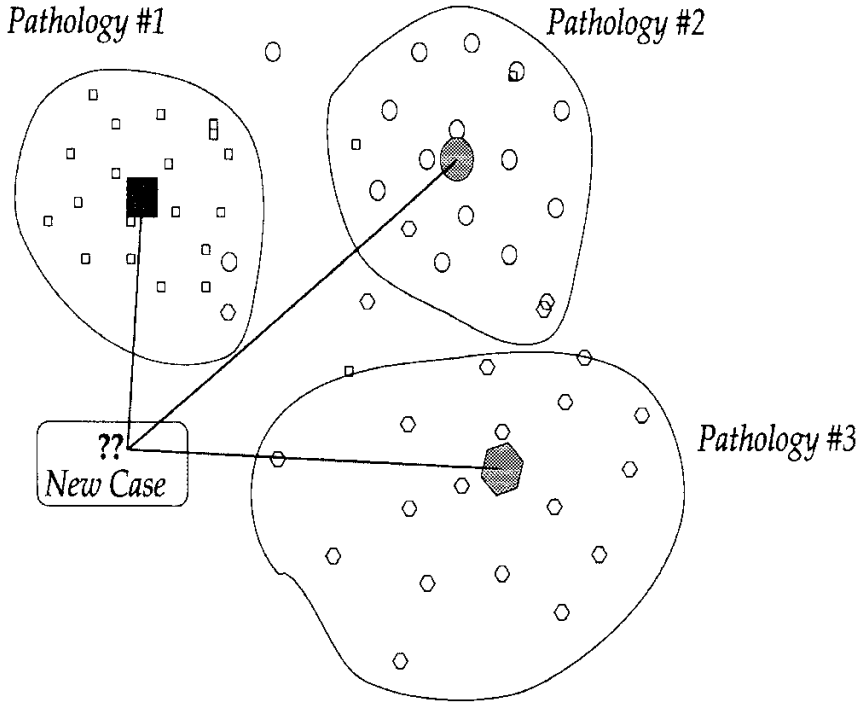
Classification: Learning

- Supervised classifiers
  - Classifiers are trained with data samples of known labels
  - Number of clusters known a priori and cannot be changed
  - Relies on “trainer” to provide the correct information

- Unsupervised classifiers
  - Classifiers are trained with data samples with unknown labels
  - Discover underlying clusters according to particular criteria
  - Interesting to identify new clusters representing sub-classes within large normal or pathological classes

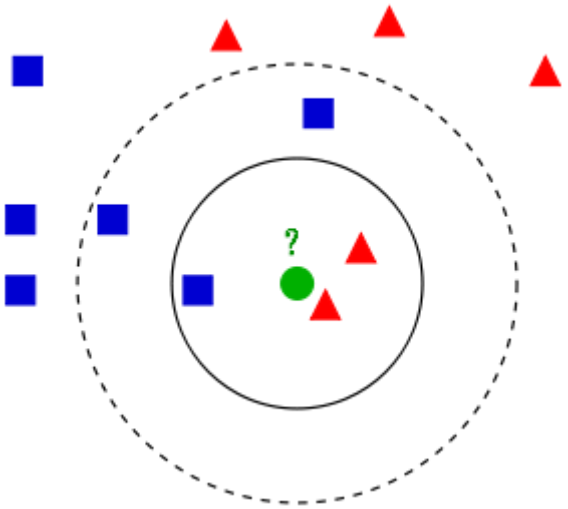
Minimum Distance Classifier

- Idea: Compute the distance between unknown case and the center of known clusters and assign the case to the closest cluster
  - How to compute cluster centers
  - Distance computation
  - Simple but not good in many cases



K-Voting Nearest Neighbor

- Assumption: an unknown case is likely to belong to the class of its neighbors in space
- Idea: Find the K neighbors of the unknown case and make a majority vote among them and assign the case to the pathology of the majority
  - Select K in such a way to avoid "ties": e.g., odd number for binary classification
  - Special case: K=1 – Nearest Neighbor Classifier
  - Matlab function "knnclassify"

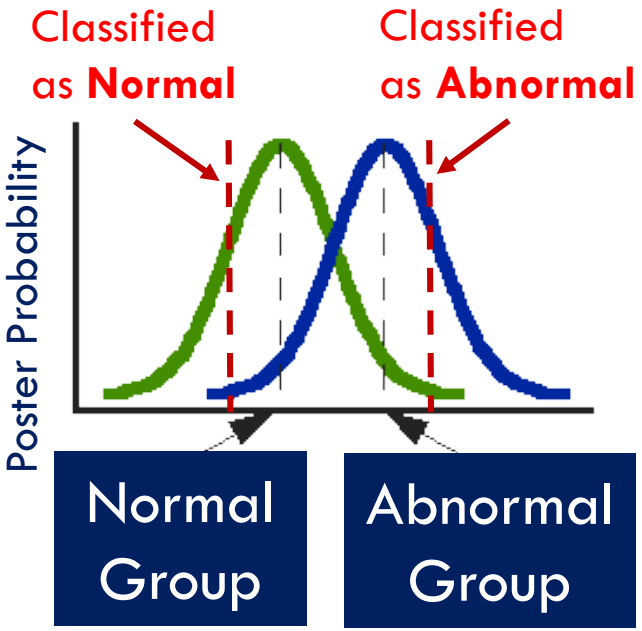


Naïve Bayesian Classifer

- Idea: Assign the unknown case to the class of largest posterior probability

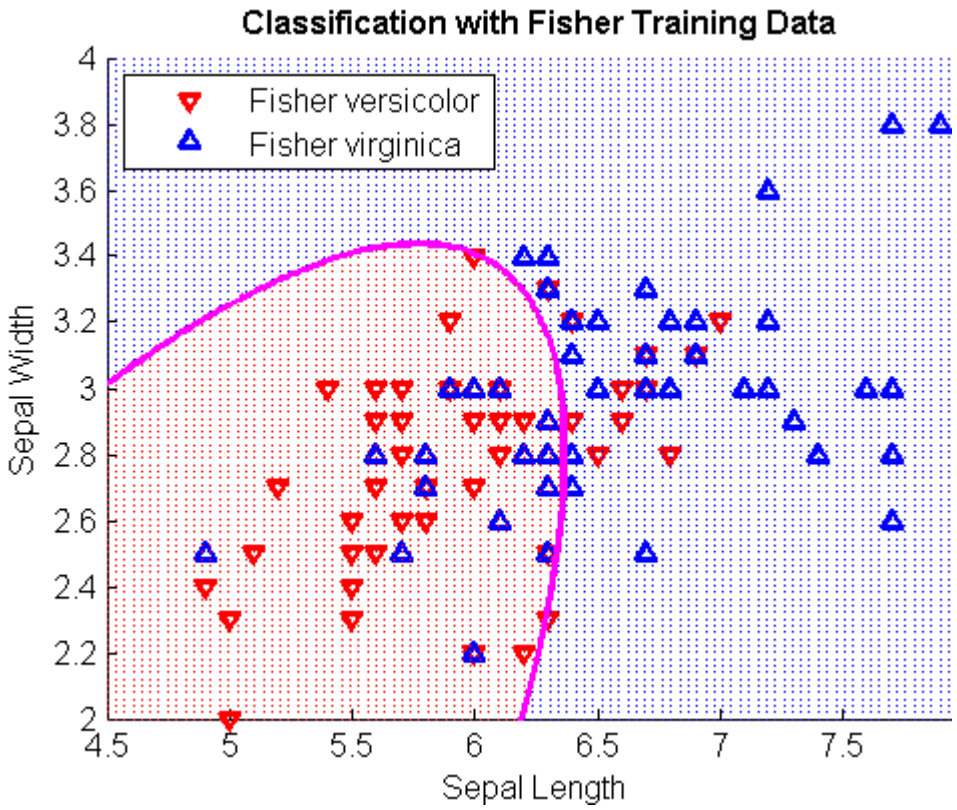
$$P(y \mid x) = \frac{P(x \mid y)P(y)}{P(x)}$$

  - P(y) is the *a priori* probability of y
  - P(x | y) is the likelihood function;
  - P(x) is the marginal probability
  - P(y | x) is the *a posteriori* probability



Discriminant Analysis

- Idea: Find the best separation surface between clusters based
- Generalization of Naïve Bayesian classifier which is a linear version of it
  - Linear: LDA
  - Quadratic: QDA
  - Matlab function "classify"



Things to Watch for

- Feature normalization is important and can improve training and classification results
- Independent training and testing data is a must
  - Half and half for example
  - Leave-one-out cross validation can be used for small data sets

Assignments

- Prepare a list of past papers on the topic of automated ECG arrhythmia diagnosis Download and select one of them to and summarize its idea in less than 500 words. Comment on their use of diagnostic performance measures to describe the performance of their technique.

Information Source	Best For:	The Information:	Watch For:
Books	• Comprehensive information • Background and historical information • Bibliography of sources	• Historical context • Broad overviews • broader audience	• Dated information • Content level • Bias or slant (author)
Popular/Special Interest Magazine	• Current information • Shorter, easy to understand • Photos and illustrations	• Long-form stories. • Discuss impact on society • Offers perspectives • General audience	• Authors not experts • May lack depth • Sources not always cited • Editorial bias
Professional/Trade Magazines	• Specialized information • Current information • Some bibliographies	• Long articles or reports • Context and analysis • Professional audience	• Vary between short, easy to lengthy and specialized • Sources not always cited
Scholarly/Academic Journals	• Depth • Written by experts • Charts and graphs • Recent research • Bibliographies of sources	• Often theoretical • Peer-reviewed • Often narrow focus • Scholars, researchers, professionals and students	• Specialized Terminology and depth • Dated information
Newspapers	• Daily information • Local information	• Quotes from experts, government, witnesses, etc. • General audience	•Authors usually not experts
Web Sites	• Government information • Varied points of view on a topic	•Is primarily provided through resources like Internet news sites when	•Credibility and accuracy cannot be assured (check for author credentials

Lab Notebook Documentation