Measures of Diagnostics Accuracy

Dept of. E & EC Engg.
IIT Kharagpur

Dr. Sudipta Mukhopadhyay
(smukho@ece.iitkgp.ernet.in)

Outline

- True Positive, (TPF) and True Negative, (TNF)
- False Negative, (FNF) and False Positive, (FPF)
- Relationships among them
- Classification Matrix or Cost Matrix
- Positive Predictive Value & Negative Predictive Value
- Receiver Operating Characteristics (ROC)
- Free Response Receiver Operating Characteristics (FROC)
- McNemar's test of symmetry

Sensitivity -Specificity

Let us consider a screening test to detect the presence or absence of any disease (*binary decision*)

A: the event that a subject has particular pathology

N: the event that a subject does not have particular pathology

P(A): The prior probabilities represent the fractions of subjects with the disease in the test population

P(N): Prior probability represents fraction of the normal subjects in the test population

T^+ : group tested positive

T^- : group tested negative

Sensitivity -Specificity

- True positive (TP) or "hit" is a situation when the test is positive for a subject with disease. *True positive fraction* (TPF) or *Sensitivity* (S^+) is given by $P(T^+/A)$ or

$$S^+ = \frac{\text{number of TP decisions}}{\text{number of subjects with disease}}$$

- True negative (TN) is a situation when the test is negative for a subject without disease. *True negative fraction* (TNF) or *Specificity* (S^-) is given by $P(T^-/N)$ or

$$S^- = \frac{\text{number of TN decisions}}{\text{number of subjects without disease}}$$

Sensitivity -Specificity

- False negative (FN) or "miss" is a situation when the test is negative for a subject with disease. False negative fraction (FNF) is given by $P(\bar{T}/A)$ or

$$P\left(T^{-}/A\right)=\frac{\text{number of FN decisions}}{\text{number of subjects with disease}}$$

- False positive(FP) or a false alarm is a situation when the test is positive for a subject without disease. False positive fraction (FPF) is given by P(T+/N) or

$$P\left(T^{+}/N\right)=\frac{\text{number of FP decisions}}{\text{number of subjects without disease}}$$

Sensitivity –Specificity (Relationship)

- FNF+TPF=1
- FPF+TNF=1
- Specificity S^{-}=1-FPF=TNF
- Sensitivity S^{+}=1-FNF=TPF
- Accuracy of the screening test = $S^{+}P(A)+S^{-}P(N)$

where P(A) is the fraction of the study population that actually has the disease (that is the prevalence of the disease) and P(N) is the fraction of the study population that is actually free of the disease

Sensitivity –Specificity (Relationship)

Classification/Confusion matrix

	Predicted group	
Actual group	Normal	Abnormal
Normal	S^{--}=TNF	FPF
Abnormal	FNF	S^{+}=TPF

Cost Matrix

	Predicted group	
Actual group	Normal	Abnormal
Normal	C_N	C_{FP}
Abnormal	C_{FN}	C_A

Loss for misclassification is $L=FPF \times C_{FP} + FNF \times C_{FN}$

Total cost of screening is

$CS = TPF \times C_A + TNF \times C_N + FPF \times C_{FP} + FNF \times C_{FN}$

where C_A: cost of the test when a TP is found

C_N : cost of conducting the test and arriving at a TN decision

C_{FP}: cost of FP result

C_{FN}: cost of FN result

Positive Predictive Value & Negative Predictive Value

The efficiency of the test may also be indicated by predictive values

- Positive predictive value (PPV)=100*TP/(TP+FP)
  Represent % of the cases labeled as '+' by the test that are actually '+'

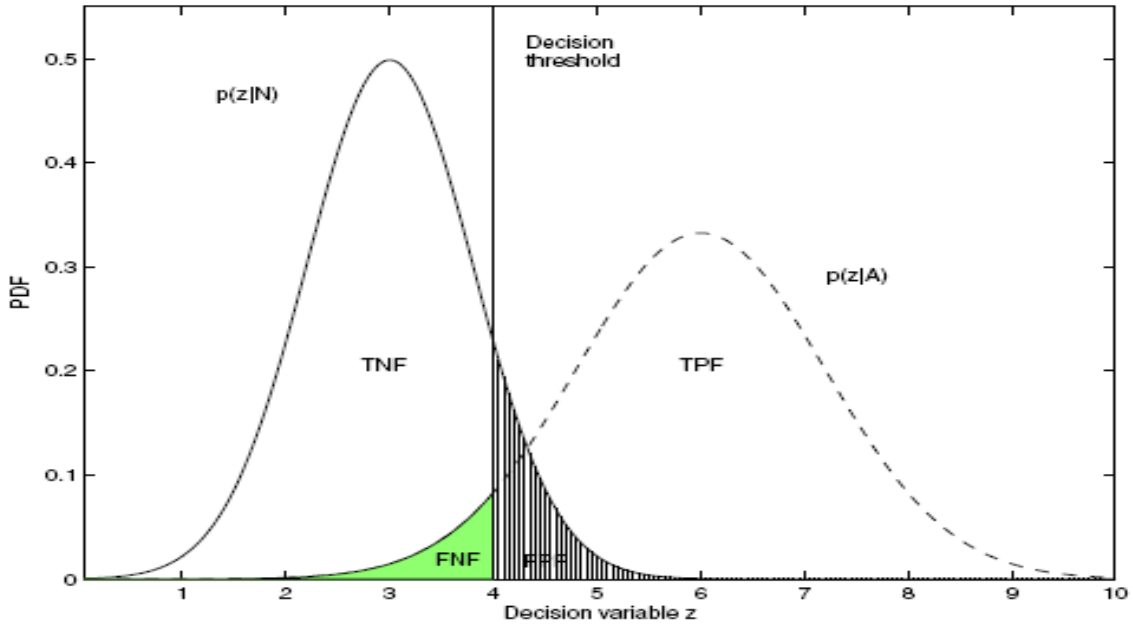
- Negative predictive value (NPV)=100*TN/(TN+FN)
  Represent % of the cases labeled as '-' by the test that are actually '-'

Receiver Operating Characteristics

- Correct classification of patterns as percentages provide limited indications of the accuracy of a diagnostic method.
- The provision of a separate correct classification rate for each category, such as sensitivity and specificity, can facilitate improved analysis.
- However, these measures do not indicate the dependence of the results upon the decision threshold.
- It is desirable to have a screening or diagnostic test that is both highly sensitive and highly specific. In reality, however, such a test is usually not achievable.
- The relationship between sensitivity and specificity is illustrated by the receiver operating characteristics (ROC) curve which facilitates improved analysis of the classification accuracy of a diagnostic method.

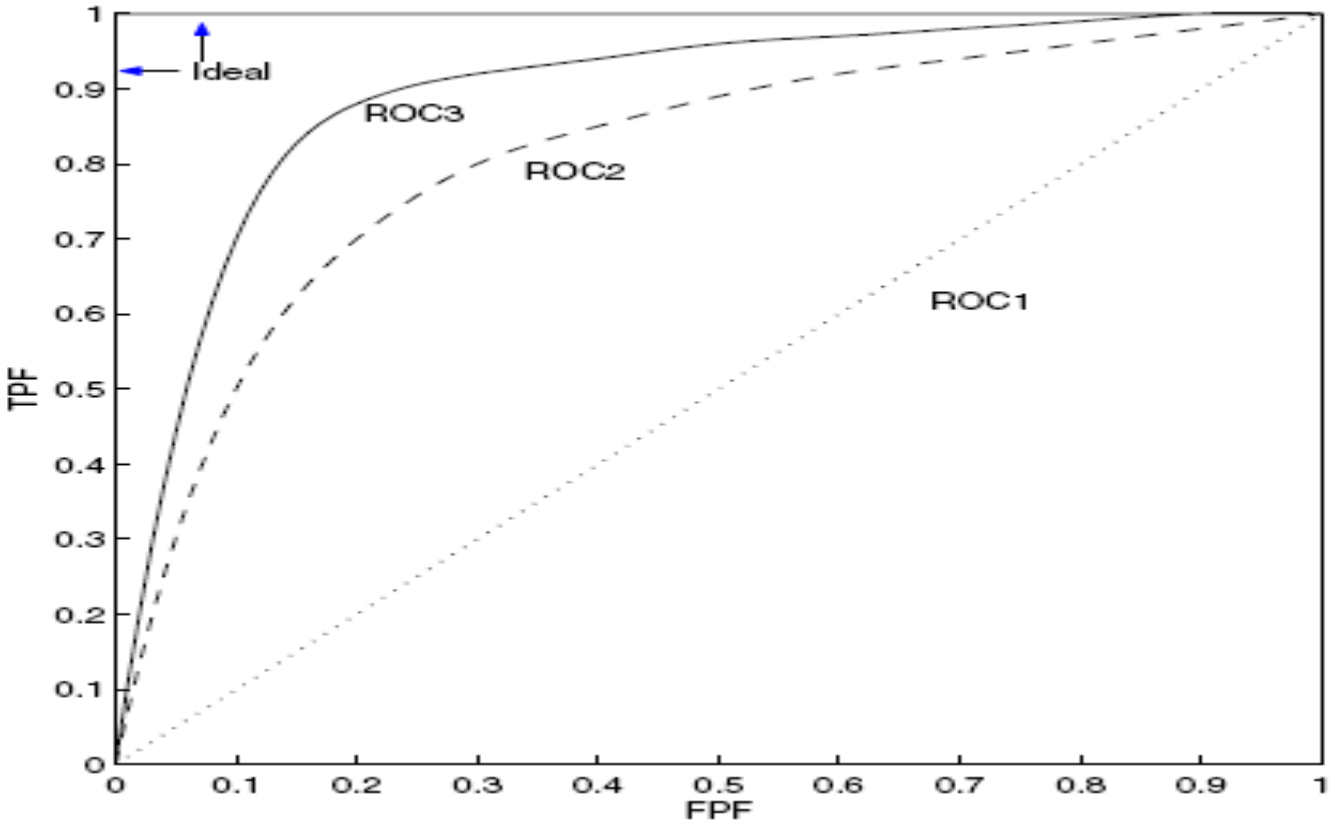
Receiver Operating Characteristics

- For a given diagnostic test with the decision variable z, we have predetermined state-conditional PDFs of the decision variable z for actually negative indicated as p(z/N) and for actually positive indicated as p(z/A). [Overlapping]
- The user or operator needs to determine a decision threshold (indicated by the vertical line) so as to strike a compromise between sensitivity and specificity.
- Lowering the decision threshold will increase TPF at the cost of increased FPF.



State conditional PDFs of a diagnostic decision variable

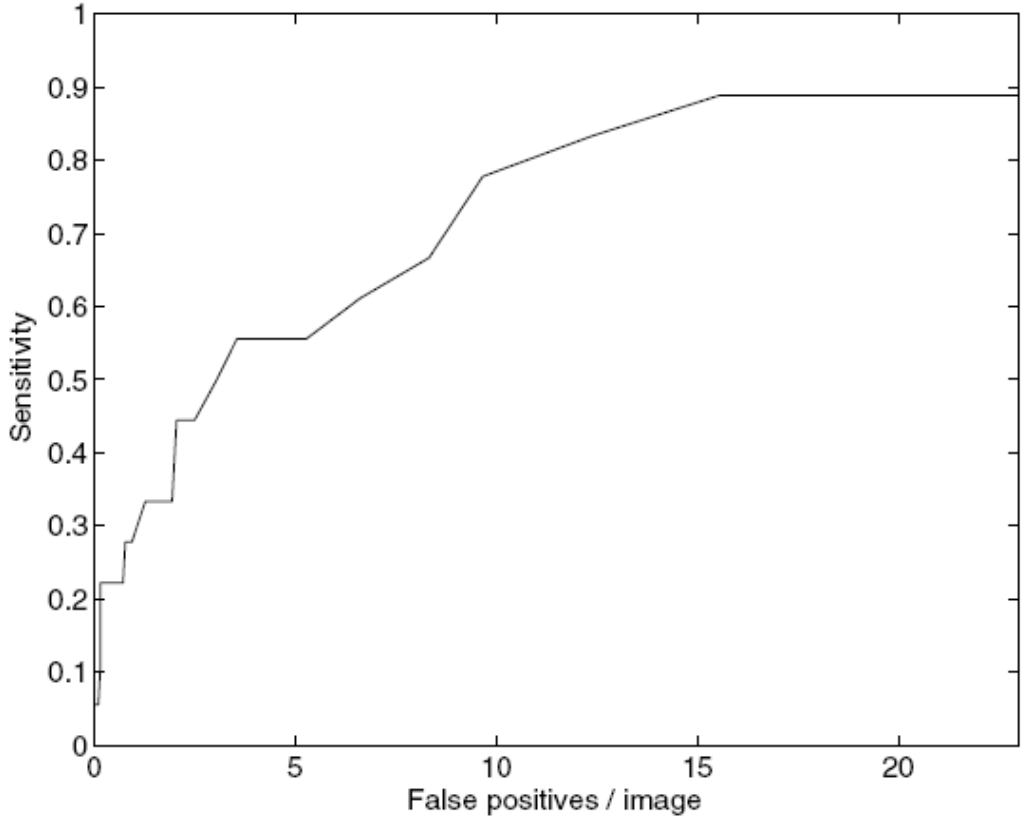
Receiver Operating Characteristics



Receiver Operating characteristics (ROC) curve

Free Response ROC (FROC)

- Another form of ROC analysis known as the free response ROC or FROC where the sensitivity is plotted against the number of false positives per image.



McNemar's test of symmetry

- McNemar's test is based on the construction of contingency tables that compare the results of two classification methods.
- Rows represent the outcomes of one methods used as the reference (gold standard).
- Columns represent outcomes of the other method, which is a new method (To be evaluated).
- The variables a, b, c, d, e, f, g, h, and i denote the counts in each cell, and the numbers in parentheses denote the cell number.
- The variables C1,C2, and C3 denote the total numbers of counts in the corresponding columns; R1, R2, and R3 denote the total numbers of counts in the corresponding rows.
- The total number of cases in the true category represented by the table is N = C1+C2+C3=R1+R2+R3

	Method B			
Method A	Normal	Intermediate	Abnormal	Total
Normal	a (1)	b (2)	c (3)	R1
Intermediate	d (4)	e (5)	f (6)	R2
Abnormal	g (7)	h (8)	i (9)	R3
Total	C1	C2	C3	N

Contingency Table for McNemar's test of asymmetry

McNemar's test of symmetry

- Each cell in a contingency table represents a paired outcome. (Example)
- The row totals R1, R2, and R3 and the column totals C1, C2, and C3 may be used to determine the sensitivity and specificity of the two methods.
- High values along the main diagonal a, e, i of a contingency table indicate no significant change in diagnostic performance with Method B as compared to Method A.
- In contingency table for truly abnormal cases a high value in the upper, right portion (cell number 3) will indicate an improvement in diagnosis (higher sensitivity) with Method B as compared to Method A.
- In evaluating a contingency table for truly normal cases, Method B will have a higher specificity than Method A if a large value is found in cell 7.

	Method B			
Method A	Normal	Intermediate	Abnormal	**Total**
Normal	a (1)	b (2)	c (3)	R1
Intermediate	d (4)	e (5)	f (6)	R2
Abnormal	g (7)	h (8)	i (9)	R3
Total	C1	C2	C3	N

Contingency Table for McNemar's test of asymmetry

References

- Rangayyan, RM: Biomedical Image Analysis. CRC, Boca Raton, FL (2005) Page 1132-1140
- ROCKIT,

  http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm