Confusion Matrix

Dr. Muhammad Madeem Majeed

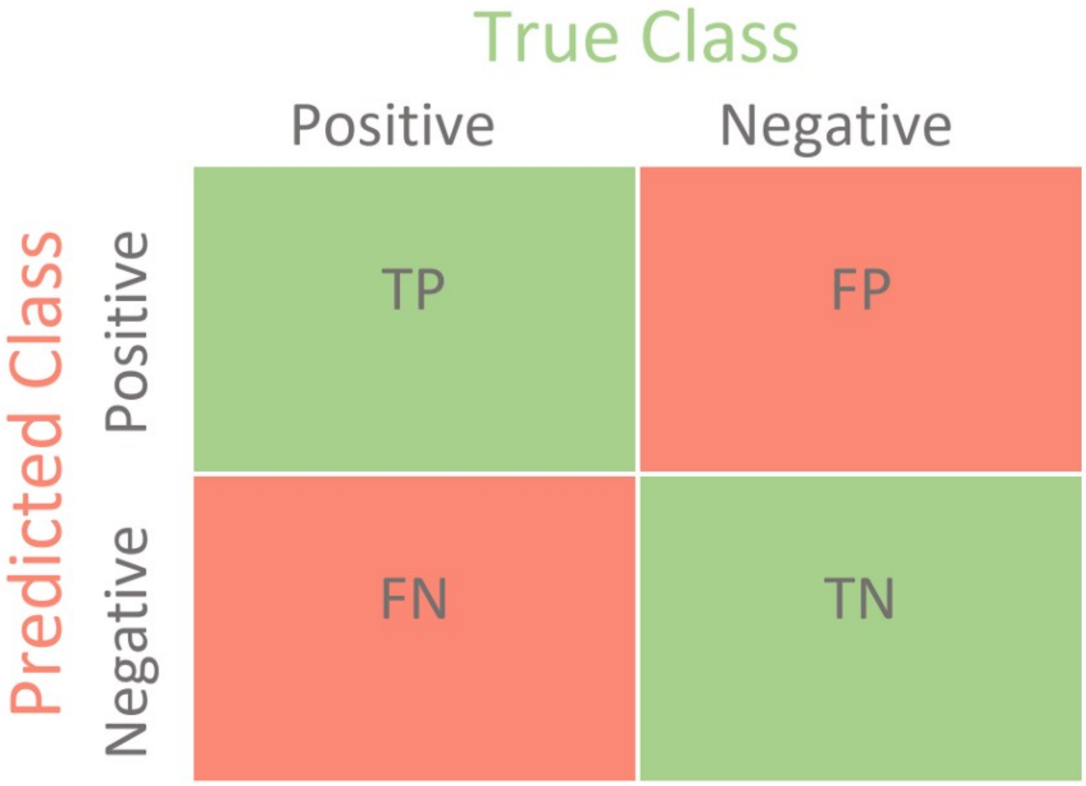
Confusion Matrix

- Confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning.
- Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa – both variants are found in the literature.
- The name stems from the fact that it makes it easy to see whether the system is confusing two classes (i.e. commonly mislabeling one as another).

Confusion Matrix

- It is a special kind of contingency table, with two dimensions ("actual" and "predicted"), and identical sets of "classes" in both dimensions (each combination of dimension and class is a variable in the contingency table).

Confusion Matrix



Confusion Matrix: Example

- Given a sample of 12 pictures, 8 of cats and 4 of dogs, where cats belong to class 1 and dogs belong to class 0,

  actual = [1,1,1,1,1,1,1,1,0,0,0,0],

- assume that a classifier that distinguishes between cats and dogs is trained, and we take the 12 pictures and run them through the classifier. The classifier makes 9 accurate predictions and misses 3: 2 cats wrongly predicted as dogs (first 2 predictions) and 1 dog wrongly predicted as a cat (last prediction).

  prediction = [0,0,1,1,1,1,1,1,0,0,0,1]

- With these two labeled sets (actual and predictions), we can create a confusion matrix that will summarize the results of testing the classifier:

Confusion Matrix: Example

- In this confusion matrix, of the 8 cat pictures, the system judged that 2 were dogs, and of the 4 dog pictures, it predicted that 1 were cats.
- All correct predictions are located in the diagonal of the table (highlighted in bold), so it is easy to visually inspect the table for prediction errors, as values outside the diagonal will represent them.

Predicted class / Actual class	Cat	Dog
Cat	**6**	2
Dog	1	**3**

Confusion Matrix: Example

Predicted class / Actual class	P	N
P	**TP**	FN
N	FP	**TN**

Predicted class / Actual class	Cat	Non-cat
Cat	**6 true positives**	2 false negatives
Non-cat	1 false positive	**3 true negatives**

Terminologies used

- condition positive (P)
  the number of real positive cases in the data
- condition negative (N)
  the number of real negative cases in the data
- true positive (TP)
  eqv. with hit
- true negative (TN)
  eqv. with correct rejection
- false positive (FP)
  eqv. with false alarm, type I error or underestimation
- false negative (FN)
  eqv. with miss, type II error or overestimation

Sensitivity and Specificity

- In a diagnostic test, sensitivity is a measure of how well a test can identify true positives and specificity is a measure of how well a test can identify true negatives.
- For all testing, both diagnostic and screening, there is usually a trade-off between sensitivity and specificity, such that higher sensitivities will mean lower specificities and vice versa.

Sensitivity and Specificity

- Consider the example of a medical test for diagnosing a condition.
- Sensitivity refers to the test's ability to correctly detect ill patients who do have the condition.
- In the example of a medical test used to identify a condition, the sensitivity (sometimes also named the detection rate in a clinical setting) of the test is the proportion of people who test positive for the disease among those who have the disease.

Sensitivity and Specificity

- Mathematically, this can be expressed as:

$$\begin{aligned}\text{sensitivity} &= \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \\ &= \frac{\text{number of true positives}}{\text{total number of sick individuals in population}} \\ &= \text{probability of a positive test given that the patient has the disease}\end{aligned}$$

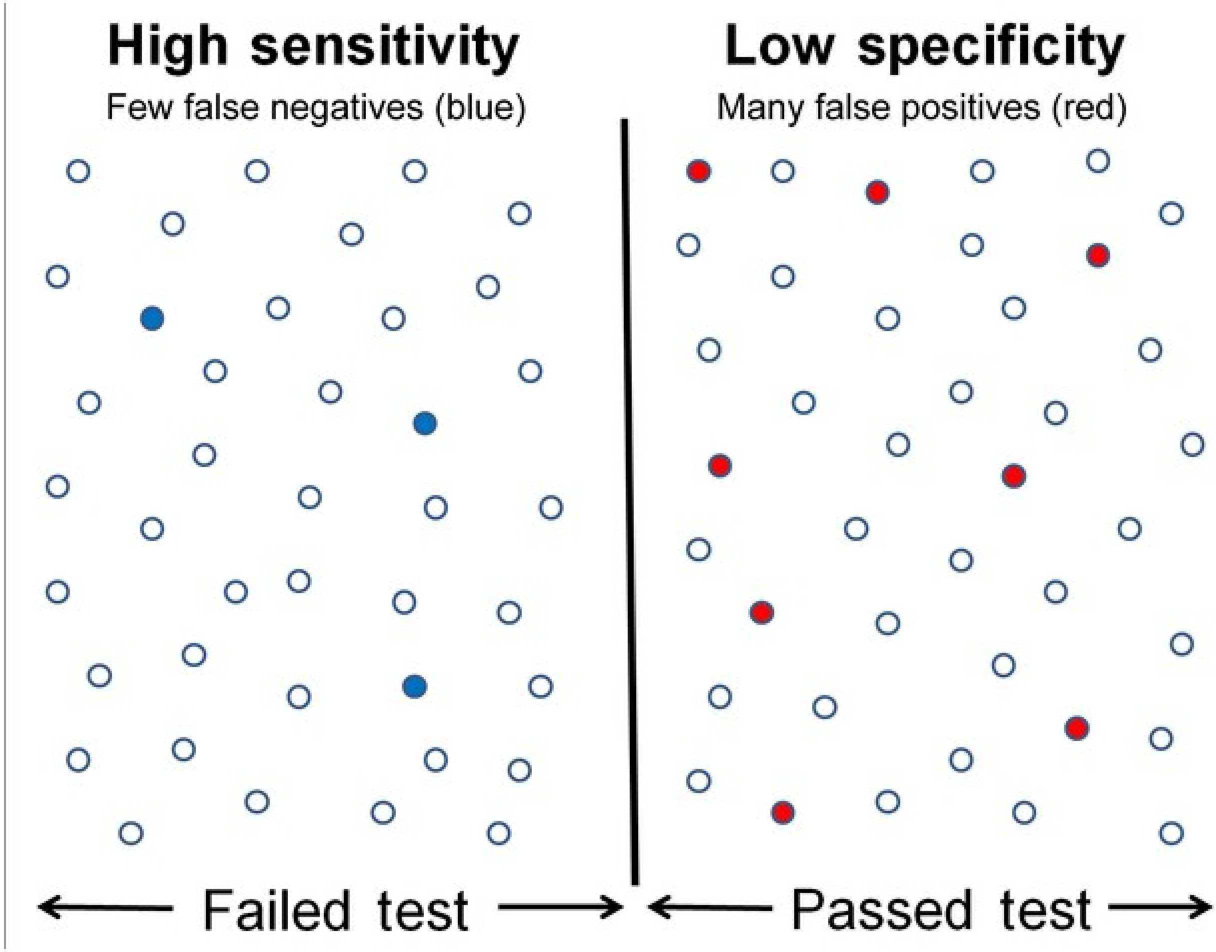
- A negative result in a test with high sensitivity is useful for ruling out disease

Sensitivity and Specificity

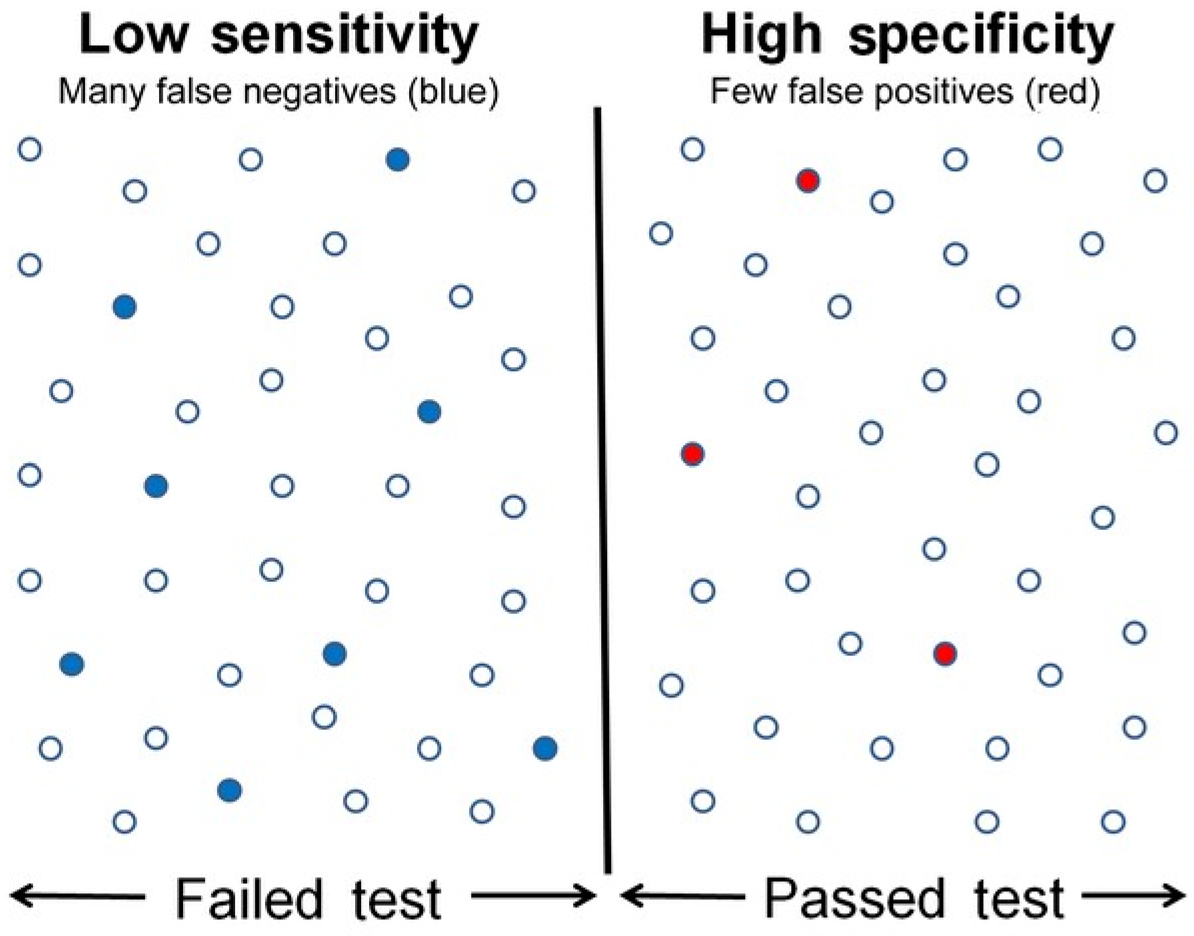
- Consider the example of a medical test for diagnosing a disease. Specificity relates to the test's ability to correctly reject healthy patients without a condition.
- Specificity of a test is the proportion of who truly do not have the condition who test negative for the condition.
- Mathematically, this can also be written as:

$$\begin{aligned}\text{specificity} &= \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}\\ &= \frac{\text{number of true negatives}}{\text{total number of well individuals in population}}\\ &= \text{probability of a negative test given that the patient is well}\end{aligned}$$

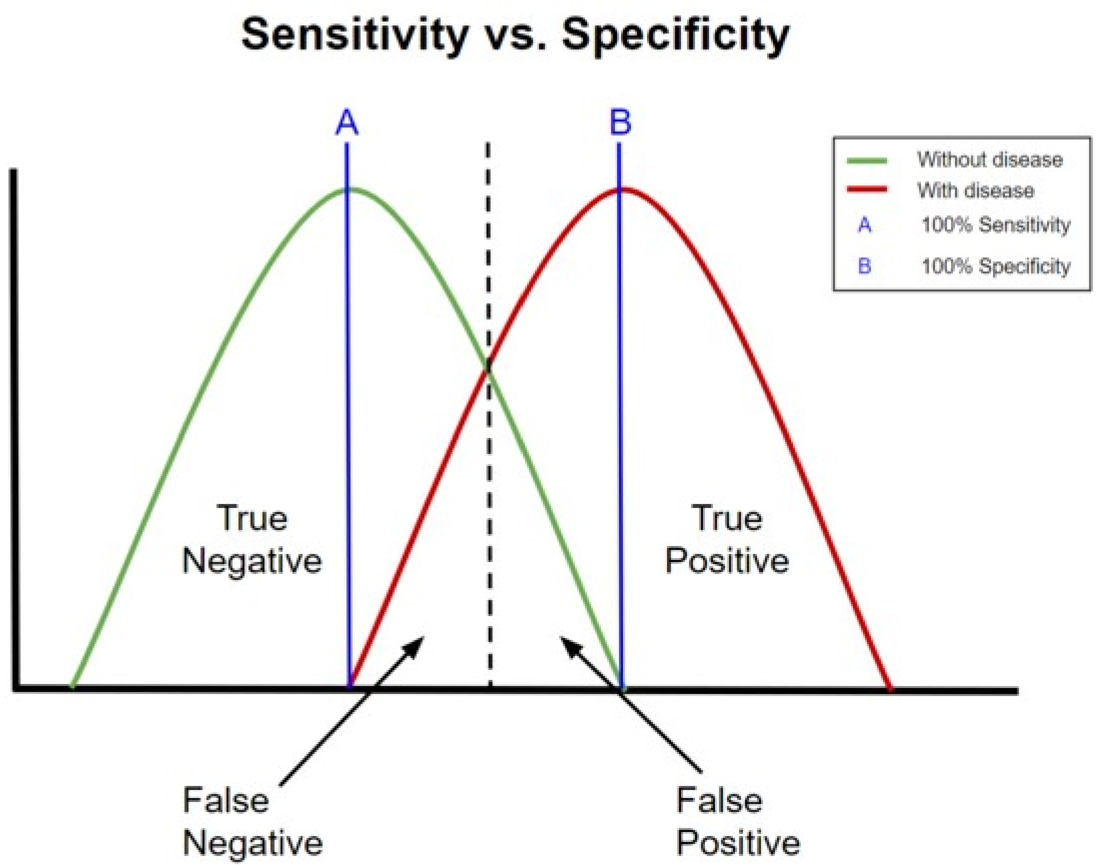
Sensitivity and Specificity



Sensitivity and Specificity



Sensitivity and Specificity



Positive and Negative predictive values

- The positive and negative predictive values (PPV and NPV respectively) are the proportions of positive and negative results in statistics and diagnostic tests that are true positive and true negative results, respectively.
- The PPV and NPV describe the performance of a diagnostic test or other statistical measure.
- A high result can be interpreted as indicating the accuracy of such a statistic.

Precision

- The positive predictive value (PPV), also called precision, is defined as,

$$\mathrm{PPV} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false positives}} = \frac{\text{Number of true positives}}{\text{Number of positive calls}}$$

Negative Predictive Value

- The negative predictive value is defined as:

$$\mathrm{NPV} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{Number of false negatives}} = \frac{\text{Number of true negatives}}{\text{Number of negative calls}}$$

Errors

- Tabularised relations between truth/falseness of the null hypothesis and outcomes of the test

actual predicted reject don't reject
True a fp 1-a

Table of error types		**Null hypothesis (H_0) is**	
		True	**False**
Decision about null hypothesis (H_0)	**Don't reject**	Correct inference (true negative) (probability = $1-\alpha$)	Type II error (false negative) (probability = β)
	Reject	Type I error (false positive) (probability = α)	Correct inference (true positive) (probability = $1-\beta$)

Accuracy

- Accuracy is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition.
- That is, the accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

Balanced Accuracy

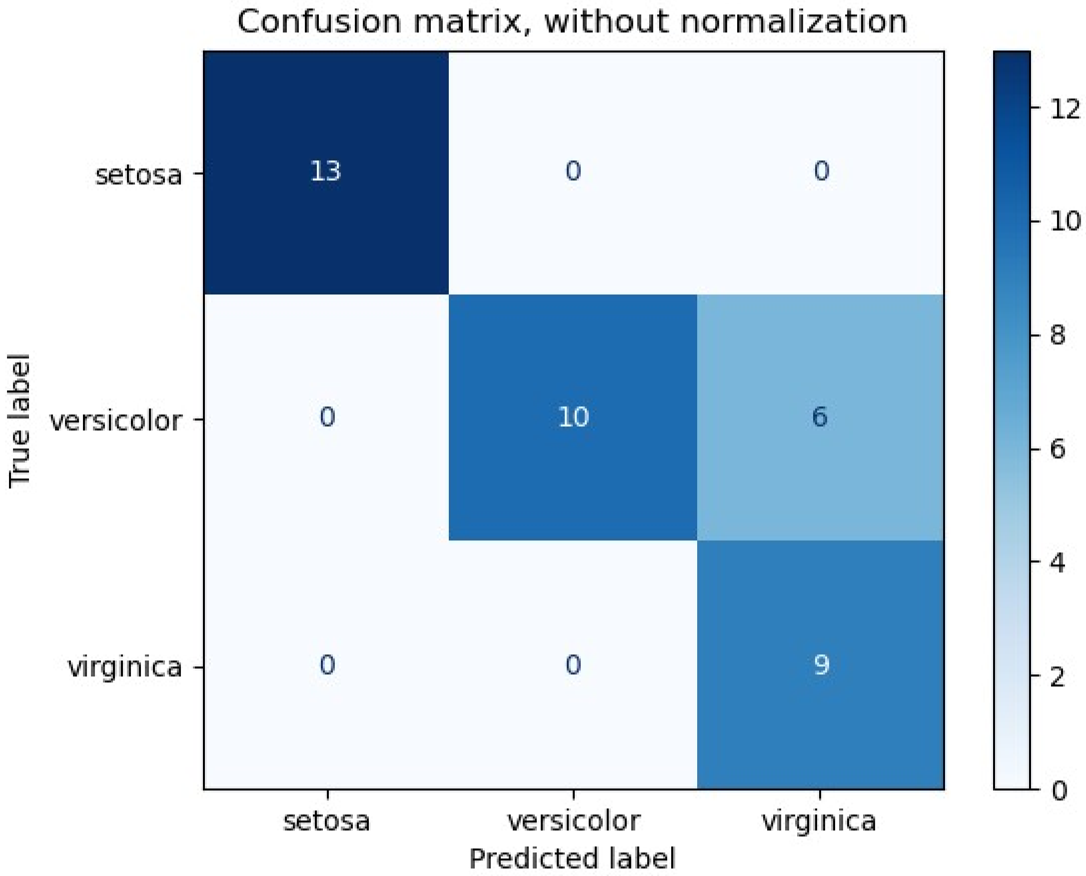
$$\mathrm{BA} = \frac{TPR + TNR}{2}$$

F1 Score

- In statistical analysis of binary classification, the F-score or F-measure is a measure of a test's accuracy.
- It is calculated from the precision and recall of the test, where the precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive.

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})} .$$

Confusion Matrix for multiple classes



Thank you