annalise.ai

# Diagnostic Accuracy

tells us how well a test, such as an AI model, can discriminate between two certain conditions (e.g. healthy and diseased)

Measures of diagnostic accuracy: Sensitivity (Se) and Specificity (Sp), Positive and negative predictive values (PPV, NPV), Area under the ROC curve (AUC)

## Sensitivity & Specificity

**Sensitivity** refers to the ability of an AI model to correctly identify positive cases, i.e. **true positives**

**Specificity** refers to the ability of the model to correctly identify negative cases, i.e. **true negatives**

High sensitivity is important to ensure that the model does not miss positive cases, while high specificity is important to minimize false positives.

The **optimal balance** between sensitivity and specificity depends on the **clinical consequences of false positives vs. false negatives.**

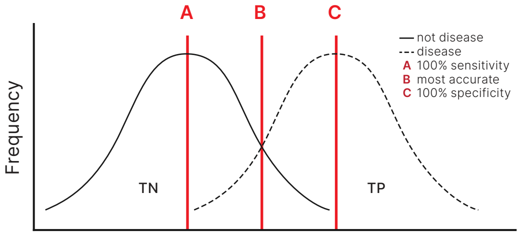
The operating threshold which determines the sensitivity and specificity of individual findings within Annalise CXR can be adjusted. If you think certain findings are being overcalled, please consult the Annalise team.

With Annalise CXR, some findings are **deliberately set with a greater sensitivity threshold**, such as:

Pneumothorax, Lung Nodules, Pleural effusions

This is to ensure these findings are not missed in clinical practice. However, this may result in some **false positive findings or "overcalls".**

### Sensitivity vs Specificity




## Positive & Negative Predictive Values (PPV & NPV)

**PPV** refers to the proportion of positive test results that are true positives, i.e., **the probability that a person with a positive test result actually has the condition.**

**NPV** refers to the proportion of negative test results that are true negatives, i.e., **the probability that a person with a negative test result does not have the condition.**

Unlike sensitivity and specifi city which are intrinsic properties of the test, PPV and NPV depend on the **prevalence of the condition in the population being tested.**

When the prevalence is low, even a highly sensitive and specific test can have a low PPV due to a higher proportion of false positives.

## Area Under the Curve (AUC) & Receiving Operator Characteristics (ROC)

**ROC Curve:** The ROC curve is a graphical representation of the performance of a medical AI model.

It plots the True Positive Rate (TPR) on the y-axis against the False Positive Rate (FPR) on the x-axis at different threshold settings.

For each of the 124 findings detected by Annalise CXR, a specific threshold value must balance the sensitivity and specificity of accurately detecting the finding.

This is known as the Operating Point Threshold.

The ROC curve shows the **trade-off between sensitivity (TPR) and specificity (1-FPR) as** the decision threshold is varied

AUC is a single value that measures the overall performance of a medical AI model.

AUC values range from 0.5 (no discrimination) to 1.0 (perfect discrimination)

AUC can be interpreted as the **probability that the model will correctly distinguish between positive and negative cases**, which is a key consideration for medical applications

Some findings have a greater sensitive treshold to ensure these findings are not missed eg pneumothorax. This may result in some false positive though.

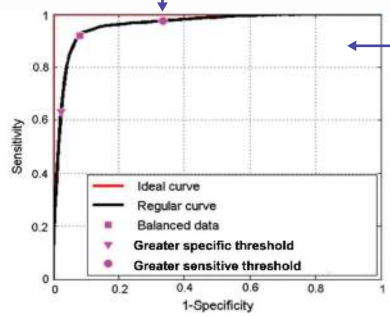



**Fig. 1** A typical representation of ROC curves. The *red curve* represents the ideal curve. The *black curve* shows an example of regular ROC curve (not ideal). The *square*, *triangular* and *circle* magenta markers indicate the results of maximizing the overall accuracy of data with different imbalance level


**Reference:**
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636062/
https://www.nature.com/articles/s41746-020-0253-3
https://fastdatascience.com/measuring-the-accuracy-of-ai-for-healthcare/
https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5