KU LEUVEN

Evaluation of fault detection methods in condition monitoring

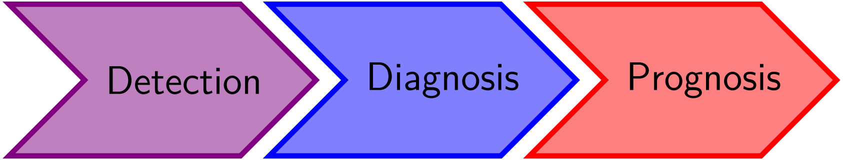
MOIRA Public Technical Course: 2024/06/07

Douw Marx, Konstantinos Gryllias

Outline

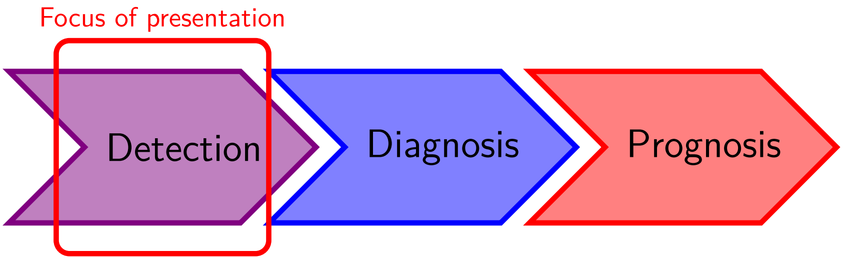
Outline

Placement of detection in condition monitoring



Three stages of the condition monitoring

Placement of detection in condition monitoring



- Fault detection
- Anomaly detection
- Out of distribution detection
- Novelty detection

Motivation for proper fault detection evaluation

Evaluate the following model with test data on the left:

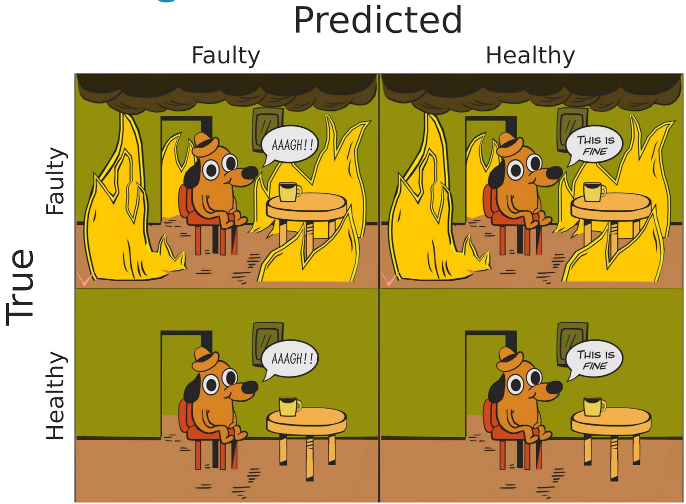
Model: f(apple) = good
(Model that always says an apple is good)

Accuracy:
accuracy = 25/26 = 96% : High accuracy!

Is this an effective measure if the model is performing well?

Outline

What makes for a good fault detection method?



The confusion matrix

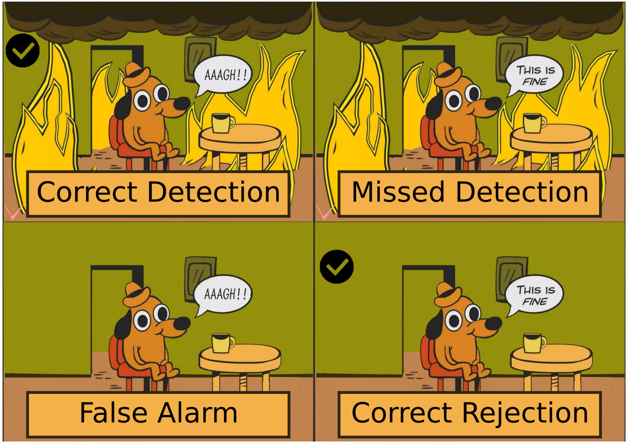
Predicted

Faulty Healthy

True

Faulty

Healthy



Desired Behaviour

The confusion matrix

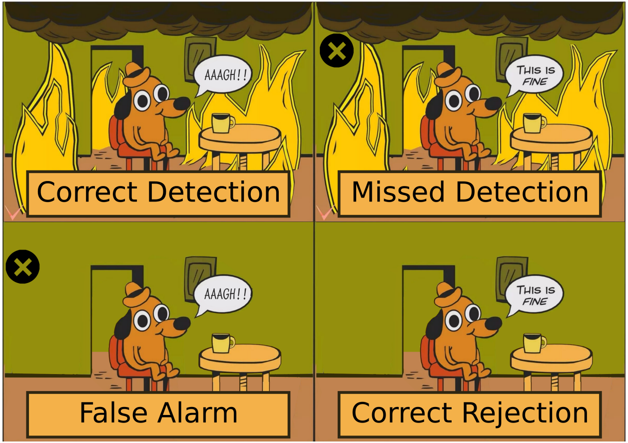
Predicted

Faulty | Healthy

True

Faulty

Healthy



Two types of mistakes can be made

The confusion matrix

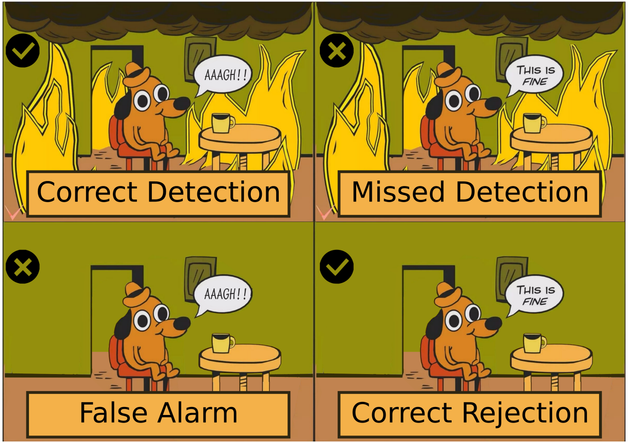
Predicted

Faulty | Healthy

True

Faulty

Healthy



Each quadrant has an associated cost

The confusion matrix

		Predicted	
		Faulty	Not Faulty
Actual	Faulty	TP	FP
	Not Faulty	FN	TN

TP: Correct detection, FP: False alarm,
FN: Missed detection, TN: Correctly identified as not faulty.

- Fault detection metrics should incorporate performance on both normal and faulty data (F1, ROC, PRC etc)

Confusion matrix example

Model: f(apple) = good (Model that always says an apple is good)

		Actual	
		Positive	Negative
Predicted	Positive	True Positive 0	False Positive 0
	Negative	False Negative 1	True Negative 25

Earlier detection is not always better

		Predicted	
		Faulty	Not Faulty
Actual	Faulty	TP	FP
	Not Faulty	FN	TN

TP: Correctly predicted as faulty, FP: Incorrectly predicted as faulty,
FN: Incorrectly predicted as not faulty, TN: Correctly predicted as not faulty.

- Early detection is often mainly concerned with correct detections.

Earlier detection is not always better

		Predicted	
		Faulty	Not Faulty
Actual	Faulty	TP	FP
	Not Faulty	FN	TN

TP: Correctly predicted as faulty, FP: Incorrectly predicted as faulty,
FN: Incorrectly predicted as not faulty, TN: Correctly predicted as not faulty.

- Early detection is often mainly concerned with correct detections.
- The importance of identifying reference samples is often overlooked.

Earlier detection is not always better

		Predicted	
		Faulty	Not Faulty
Actual	Faulty	TP	FP
	Not Faulty	FN	TN

TP: Correctly predicted as faulty, FP: Incorrectly predicted as faulty,
FN: Incorrectly predicted as not faulty, TN: Correctly predicted as not faulty.

- Early detection is often mainly concerned with correct detections.
- The importance of identifying reference samples is often overlooked.
- $\text{total cost} = \text{cost(TP)} \cdot \text{TP} + \text{cost(FN)} \cdot \text{FN} + \text{cost(FP)} \cdot \text{FP} + \text{cost(TN)} \cdot \text{TN}$

Outline

Confusion Matrix Cocktails

"Accuracy on positive class"

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = 1 - \mathrm{FPR}$$

"How often was the model correct when it was betting on the positive class"

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{FDR}$$

negative predictive value (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} = 1 - \mathrm{FOR}$$

miss rate or false negative rate (FNR)

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{P}} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1 - \mathrm{TPR}$$

fall-out or false positive rate (FPR)

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = 1 - \mathrm{TNR}$$

false discovery rate (FDR)

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}} = 1 - \mathrm{PPV}$$

false omission rate (FOR)

$$\mathrm{FOR} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TN}} = 1 - \mathrm{NPV}$$

F1 score for balanced performance

- We want to find a compromise between precision (If model says the data point is positive it is actually positive) and recall/sensitivity (You have a high accuracy on the positive class).
- Measure between 0 and 1

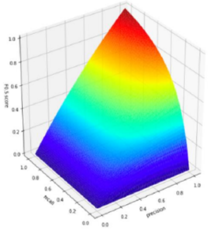
$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

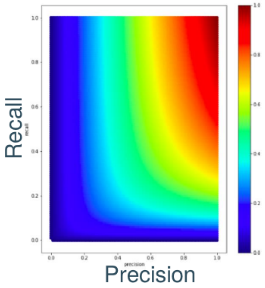
Example

- True positives (TP): 75 correctly identified as faulty machinery.
- False positives (FP): 10 samples incorrectly identified as faulty machinery.
- False negatives (FN): 5 faulty machinery samples incorrectly classified as non-faulty.
- True negatives (TN): 10 samples were correctly identified as non-faulty machinery.

- Precision = TP / (TP + FP) = 75/(75+10) = 0.882
- Recall = TP / (TP + FN) = 75/(75+5) = 0.938

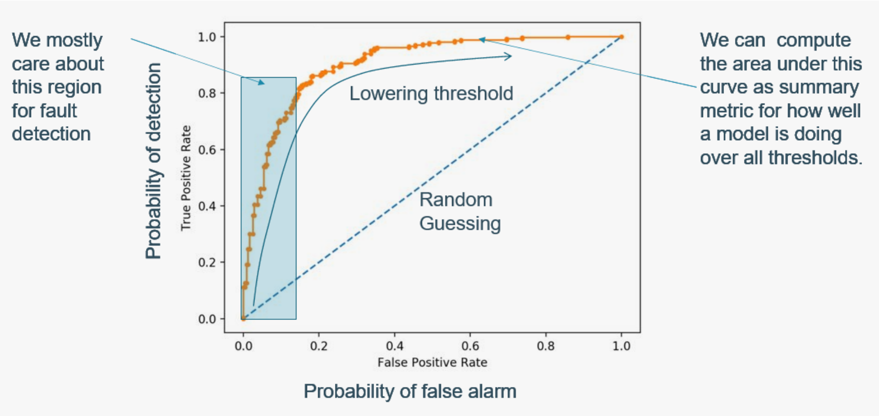
- F1 score = 2 * ((Precision * Recall) / (Precision + Recall)) = 2 * ((0.882 * 0.938) / (0.882 + 0.938)) = 0.909





Outline

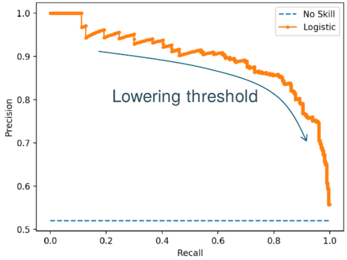
The Receiver Operating Curve (ROC)



The ROC curve shows the trade-off between hit rate (TPR) and false alarm rate (FPR).

Precision Recall Curve (PRC)

- Show the trade-off between precision and recall/sensitivity as the threshold of detection is varied.



Threshold example:

Faulty if RMS vibration above 8G vs 9G vs 10G.

The PRC curve shows the trade-off between precision and recall.

Jason Brownlee, ROC Curves and Precision-Recall Curves for Imbalanced Classification, Machine Learning Mastery, Available from https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-imbalanced-classification/, accessed May 3rd, 2020.

Precision Recall Curve vs ROC Curve

Precision Recall Curve:

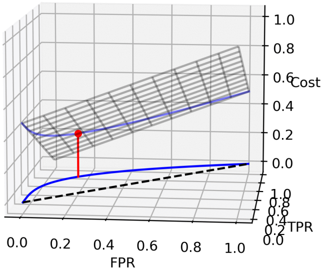
- Better for imbalanced datasets: Does not accounts for TN unlike ROC.
- Less intuitive than ROC

ROC Curve:

- Often easier to interpret than PRC
- Overly optimistic for highly imbalanced datasets

Outline

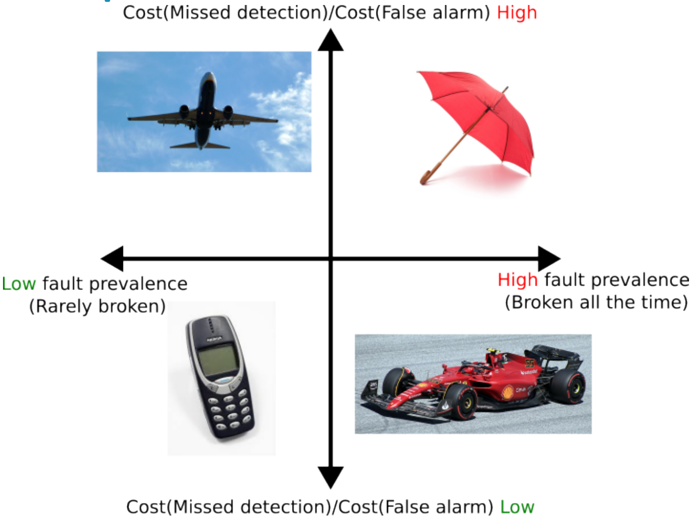
The ROC curve and CBM cost



$$\text{Cost} = \text{TPR} \times C_{TP} + \text{TNR} \times C_{TN} + \text{FPR} \times C_{FP} + \text{FNR} \times C_{FN}$$

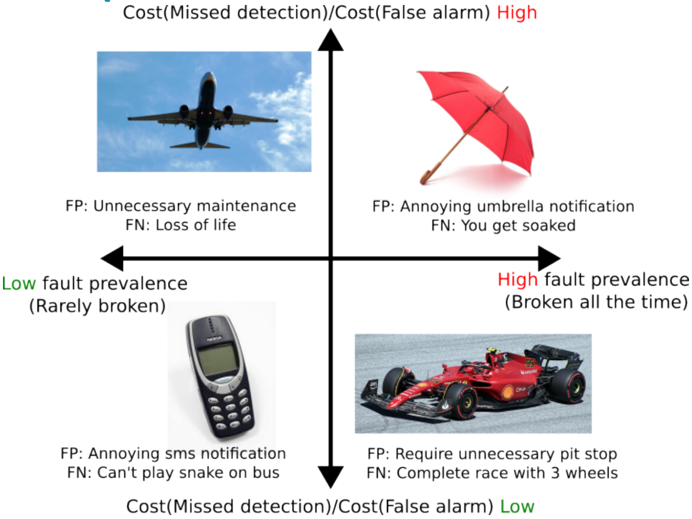
- Assume
  - $C_{TP} = 0$ (No cost for correctly detecting a fault)
  - $C_{TN} = 0$ (No cost for correctly detecting a healthy sample)
- C_{FP} (False alarm cost) and C_{FN} (Missed detection cost) are varied.
- C_X: Accounts for 1) Cost per occurrence 2) Prevalence of the fault
- FNR = 1 - TPR

Cost and prevalence differences for different applications



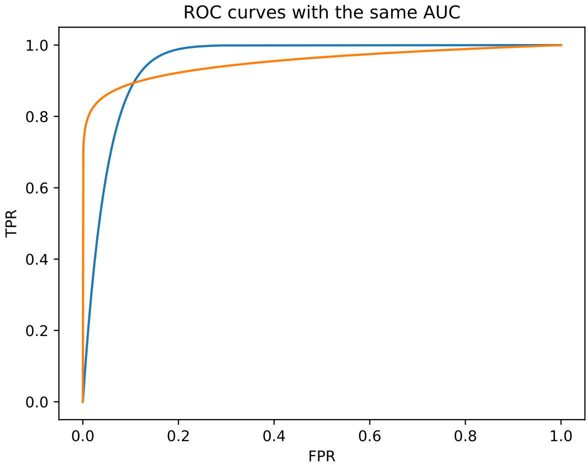
Both the relative cost of false alarms and missed detections and the prevalence of faults should be considered

Cost and prevalence differences for different applications



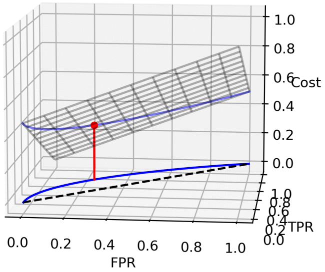
Both the relative cost of false alarms and missed detections and the prevalence of faults should be considered

Models suitable for different applications



Not all ROC curves are created equal. Orange: Preferred for e.g. package inspection, Blue: Preferred for e.g. nuclear power plants.

Goal: Optimise ROC curve given cost and prevalence



- Increase correct detections, without contributing to more false alarms.
- Consider the relative cost of false alarms and missed detections and the prevalence of faults.

Outline

Conclusions

- Evaluation metrics used to evaluate fault detection methods should incorporate performance on both normal and faulty data.
- Evaluation metrics should be designed based on the application and relative cost of false positives and false negatives.

Thank you for your attention
Any comments / critique will be appreciated

douw.marx@kuleuven.be

The authors gratefully acknowledge the European Commission for its support of the Marie Sklodowska Curie program through the ETN MOIRA project (GA 955681).