Accella AI

How to Measure AI Performance in Quality Control

A Confusion Matrix Brings Clarity!

Q "How do I know how well the AI model performs and how many mistakes it makes?"

A A confusion matrix allows you to assess the performance of AI models by providing insights into model accuracy and the types of errors it makes.

DEFINITIONS

Positives: the class you're trying to detect, in QC, somewhat unintuitively, the positives are defective products

Negatives: the default or non-target class, in QC those a the non-defective, good products

Positives and negatives can be both correctly or incorrectly identified resulting in true and false positives and negatives

	Actually Positive	Actually Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

EXAMPLE

A small manufacturer makes 200 widgets with the following confusion matrix:

	Actually Positive	Actually Negative
Predicted Positive	40 (TP)	20 (FP)
Predicted Negative	10 (FN)	130 (TN)

4 CONFUSION MATRIX PERFORMANCE METRICS

ACCURACY

Proportion of correctly classified items (both defective and non-defective) out of the total items

$$\frac{TP\,(40) + TN\,(130)}{200} = 0.85 \text{ or } 85\%$$

The model correctly classified 85% of the products

Accuracy shows the overall proportion of correct predictions (defect and non-defect)

PRECISION

Proportion of true positive predictions out of all positive predictions (TP and FP).

$$\frac{TP\,(40)}{TP\,(40) + FP\,(20)} = 0.67 \text{ or } 67\%$$

The model correctly identified positive/defective itm in 67% of the cases

Precision shows how often a model is correct when predicting the target class

SENSITIVITY (RECALL)

Proportion of actual defective items that are correctly identified as defective.

$$\frac{TP\,(40)}{TP\,(40) + FN\,(10)} = 0.8 \text{ or } 80\%$$

The model correctly identifies 88.9% of all the positive/ defective products

Sensitivity measures how well AI detects all actual defects. Higher recall = fewer missed defects

SPECIFICITY

Proportion of actual non-defective items that are correctly identified as non-defective.

$$\frac{TN\,(130)}{TN\,(130) + FP(20)} = 0.87 \text{ or } 87\%$$

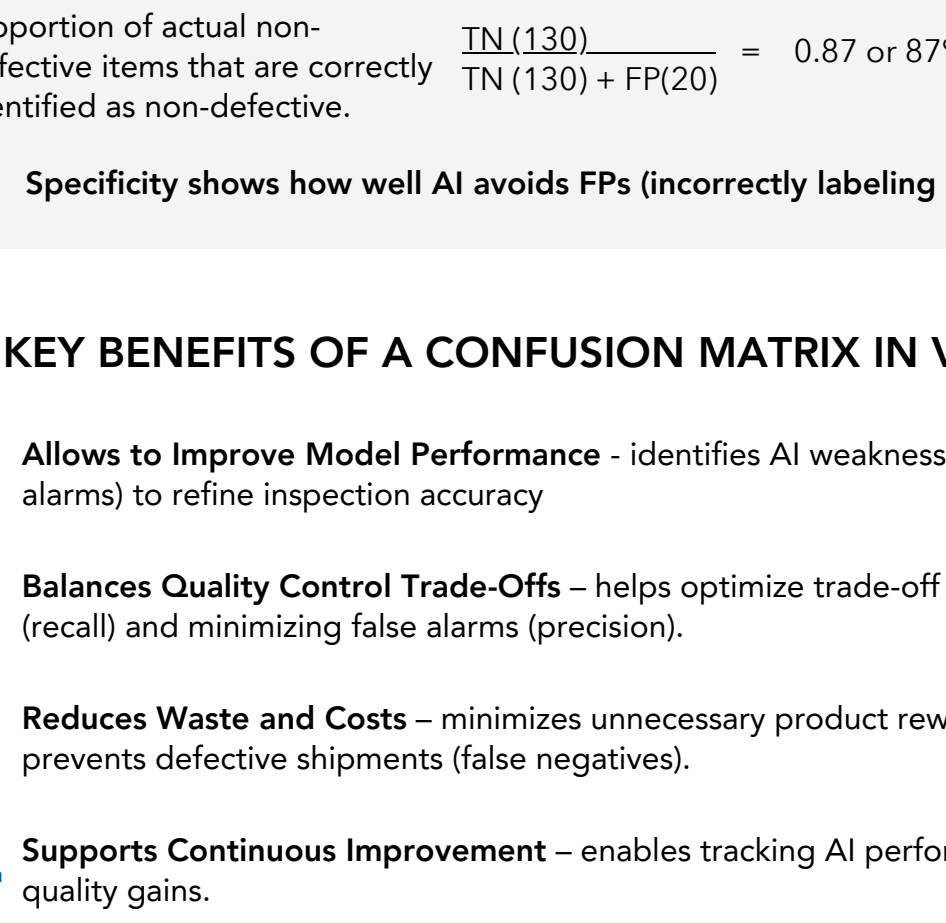
The model correctly identifies 87% of the good products as good

Specificity shows how well AI avoids FPs (incorrectly labeling good products as defective)

7 KEY BENEFITS OF A CONFUSION MATRIX IN VISUAL INSPECTION

1. **Allows to Improve Model Performance** - identifies AI weaknesses (missed defects or false alarms) to refine inspection accuracy
2. **Balances Quality Control Trade-Offs** – helps optimize trade-off between catching all defects (recall) and minimizing false alarms (precision).
3. **Reduces Waste and Costs** – minimizes unnecessary product rework (false positives) and prevents defective shipments (false negatives).
4. **Supports Continuous Improvement** – enables tracking AI performance over time for long-term quality gains.
5. **Helps with Threshold Adjustments** – allows manufacturers to tweak sensitivity based on defect severity and production needs.
6. **Ensures Regulatory and Quality Compliance** – provides data-backed insights for audits and industry quality standards.
7. **Enhances Trust in AI Systems** – transparency in AI decision-making improves acceptance by quality teams.

USE CASE: A REAL-LIFE CONFUSION MATRIX



EXPLANATION

True Positives (TP1, TP2) – Correctly identified defects 1 and 2

True Negatives (TN) – Correctly identified OK products

False Positives (FP1, FP2) – OK product wrongly labeled as defect 1 or 2

False Negatives (FN1, FN2) – defect 1 or 2 wrongly classified as OK product

Defect Misclassification (M12, M21) – Defect 1 mistaken for defect 2 and vice versa

Real life is more complicated than a 2 x 2 confusion matrix. This matrix shows two different defect categories and the OK product and all possible permutations of classification errors that can be made.

In this case the accuracy of the algorithm is 94%, precision overall is 99%, sensitivity is 100% and specificity 98%.
As an additional metrics we can calculate the misclassification rate (M12 + M21/all) as 5.5%

CONTACT

Tina Baumgartner, VP of Business Development
tina@accellagroup.com
www.accella.ai