Automatic Report Generation from models evaluation results

Model ID: {{ Model }}

Dataset ID: {{ DatasetID }}

Training Results date: {{ TrainingDate }}

|  |  |
| --- | --- |
| **Accuracy** | **Risk Score** |
| **{{ accuracy }} %** | **{{ riskscore }}** |

Contents

[1. Dataset details 3](#_Toc141708749)

[1.1. Data source 3](#_Toc141708750)

[1.2. Dataset split 3](#_Toc141708751)

[2. Confusion matrix 3](#_Toc141708752)

[3. Recall, Precision and F1 performance metrics 3](#_Toc141708753)

[4. Misclassification cost 4](#_Toc141708754)

[5. Classification confidence for correct predictions 4](#_Toc141708755)

[6. Classification confidence for incorrect predictions 4](#_Toc141708756)

[7. Correctly classified image examples 5](#_Toc141708757)

[8. Incorrectly classified image examples 5](#_Toc141708758)

List of figures

[Figure 1: Number of sample images in the training and testing dataset per class 3](#_Toc141191567)

[Figure 1: Number of sample images in the training and testing dataset per class 3](#_Toc141191568)

[Figure 1: Confusion Matrix 4](#_Toc141191569)

[Figure 1: Recall, Precision and F1 4](#_Toc141191570)

[Figure 5: Risk score due to incorrect classification 4](#_Toc141191571)

[Figure 6: Confidence levels for correct classifications 5](#_Toc141191572)

[Figure 7: Confidence levels for Incorrect classification 5](#_Toc141191573)

List of tables

[Table 1: Data location and number of sample images 3](#_Toc141274827)

[Table 2: Number of image samples 3](#_Toc141274828)

[Table 2: PFMEA Matrix provided by Rolls-Royce 4](#_Toc141274829)

[Table 3: Misclassification cost per class 4](#_Toc141274830)

## Dataset details

## Data source

Table 1: Image dataset location

|  |  |
| --- | --- |
| Data | Source path |
| Training | {{ DataPathTrain }} |
| Testing | {{ DataPathTest }} |

This table list the location of the training and testing dataset. The locations and names are in line with the naming and data structure.

## Dataset split

Table 2: Number of image samples

|  |  |
| --- | --- |
| Total number of image samples in the training dataset | **{{ NumTrainImages }}** |
| Total number of image samples in the testing dataset | **{{ NumTestImages }}** |

**Training and testing dataset distribution:** The graph below illustrates the number of image samples per class in the training and testing dataset.

{{ DatasetSplit1 }}

Figure : Number of sample images in the training and testing dataset per class

**Percentage and count of training and testing dataset distribution per class:** The graph below illustrates the dataset distribution on each training and testing dataset as a percentage of the total samples:

{{ DatasetSplit2 }}

Figure : Number of sample images in the training and testing dataset per class

## Confusion matrix

{{ ConfusionMatrix }}

Figure : Confusion Matrix

A confusion matrix shows how many defects were classified correctly and incorrectly.

The columns represent the predicted class labels, while the rows represent the actual class labels. Each cell contains the number of instances that belong to that particular category. The diagonal cells show the number of correct predictions, while the off-diagonal cells show the number of incorrect predictions. From this matrix, we can calculate other metrics such as accuracy, precision, recall and F1 score.

## Recall, Precision and F1 performance metrics

{{ rpf1 }}

Figure : Recall, Precision and F1

**Recall**: Provides the proportion of positive instances of a particular class that were correctly classified out of all defects of that class in the holdout dataset.

**Precision**: Provides the proportion of correctly classified instances of a particular class out of all positive instances the model claims belong to that class from the holdout dataset.

**F1 score:** This is the harmonic mean of the precision and recall meant to assign equal weight to both as a measure of model performance.

**Accuracy:** This is the total number of correct predictions using the test dataset divided by the total number of test samples.

**Macro Average:** This is the average for each metric (Recall, Precision and F1)

**Weighted Average:** This average considers the number of samples per class that have been used when testing the prediction of the model.

## Misclassification cost

Table 2: Misclassification cost matrix

{{ CostMatrix}}

The misclassification cost matrix represents the weighted misclassification risk score table. The diagonal cells show a zero value as they relate to correct predictions, while the off-diagonal cells show the weighted value of the risk associated for the incorrect prediction.

Table 3: Misclassification cost

{{ MisclassificationCostPerClass }}

Figure : Risk score due to incorrect classification

The misclassification cost graph shows the result of applying the weighted risk score from the PFMEA matrix to the incorrect number of predictions per class. The higher the value it means that the model has. The graph also includes the total risk score (sum of all risk scores per class).

## Classification confidence for correct predictions

{{ CCconfidence }}

Figure : Confidence levels for correct classifications

The graph above illustrates the confidence levels (probability) for the correct prediction. Ideally the values should be closer to one.

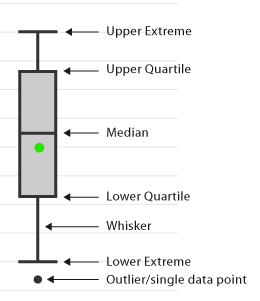
## Classification confidence for incorrect predictions

{{ ICconfidence }}

Figure : Confidence levels for Incorrect classification

The figure above shows the confidence levels for the incorrect predictions. Ideally these values should be closer to 0.5. This is because the model shouldn’t be very confident when making an incorrect prediction.

Boxplot interpretability shown on in the two figures above.



The box plot represents the probability distribution for each class, where 25% of the probability results is between the lower extreme and the lower quartile. 50% of the probability are between the lower extreme and the Median and 75% between the lower extreme and the Upper quartile. The plot also includes a green dot that represents the average confidence. There is also black dots located outside the lower and upper extreme representing the outliers.

## Correctly classified image examples

This section will show samples of correctly classified images (it needs implementing – we need folders containing these results)

## Incorrectly classified image examples

This section will show incorrectly classified images (it needs implementing - we need folders containing these results)