Table 11 presents the definitions and detailed descriptions of the five performance measures we mentioned in Section 4.2.4.

Table 11. A summary of the studied measures.

| **Measure** | **Definition** | **Description** |
|---|---|---|
| AUC | The area under the receiver operator characteristic curve. | The receiver operator characteristic Curve plots the true positive rate against the false positive rate across all the thresholds measure of the mean squared difference between the predicted probability assigned to a module and the actual probability. |
| Recall | $\text{Recall} = \frac{TP}{(TP+FN)}$ | The percentage of defective modules that are correctly classified as defective. |
| F1 | $F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}$ | The harmonic mean of precision and recall, providing a balance between the two, and is particularly useful when the class distribution is uneven. |
| Brier Score | $\text{Brier} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$ | A measure for assessing the accuracy of probabilistic predictions. |
| PF | $pf = \frac{FP}{TN+FP}$ | The ratio of incorrect positive predictions to the total number of actual negative instances. |