

Supplementary Material

Calculations for Performance metrics

Accuracy shows the proportion of correct predictions out of all predictions and is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} ;$$

Precision, given by

$$\text{Precision} = \frac{TP}{TP + FP},$$

measures the proportion of predicted positive cases that are truly positive;

Recall (or Sensitivity) indicates the ability of the model to correctly identify positive cases and is defined as

$$\text{Recall} = \frac{TP}{TP + FN} ;$$

Specificity measures how effectively the model recognizes true negative cases and is computed as

$$\text{Specificity} = \frac{TN}{TN + FP} ;$$

F1-score is the harmonic mean of precision and recall, i.e. the trade-off between them. It is useful when both FP and FN are costly. F1-score is defined as

$$\text{F1-score} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} ;$$

Balanced Accuracy is particularly suitable in healthcare datasets, where positive and negative outcomes may be unevenly distributed. It is given by

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} ;$$

Area Under the Receiver Operating Characteristic Curve (AUC) quantifies the model's ability to discriminate between classes across all thresholds. The AUC of 1 indicates a perfect discrimination, while 0.5 indicates no better than random guessing.

Abbreviations:

TP – True Positive, correctly predicted positive cases

TN – True Negative, correctly predicted negative cases

FP – False Positive, negative cases incorrectly classified as positive

FN – False Negative, positive cases incorrectly classified as negative

AUC – Area Under the ROC Curve; measures discriminatory ability across thresholds