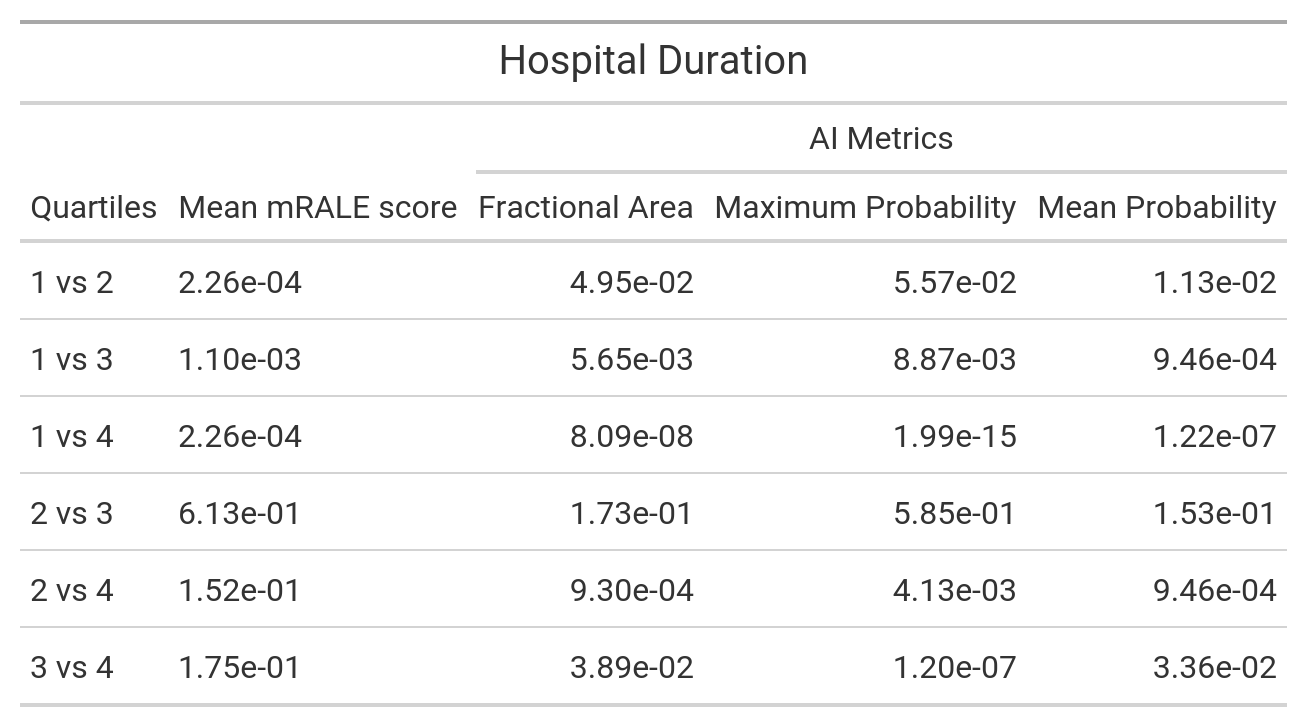
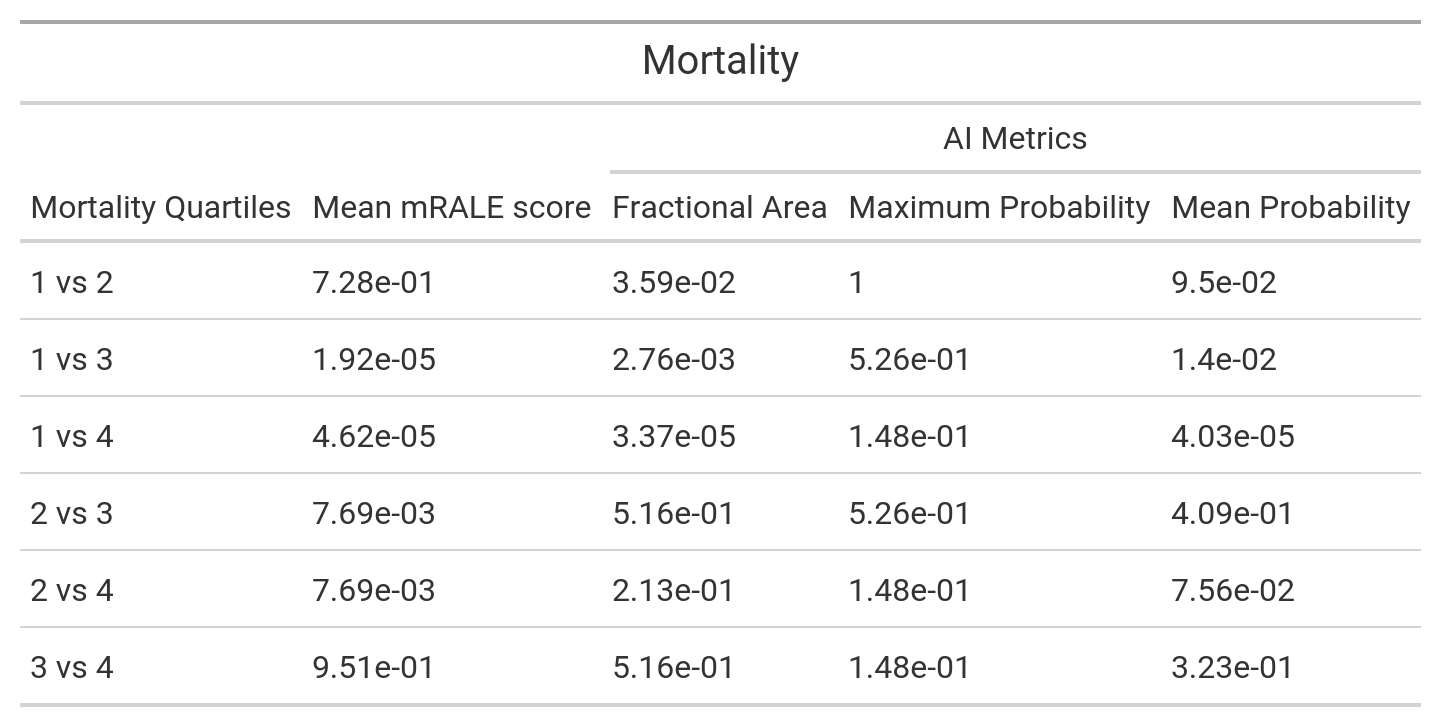
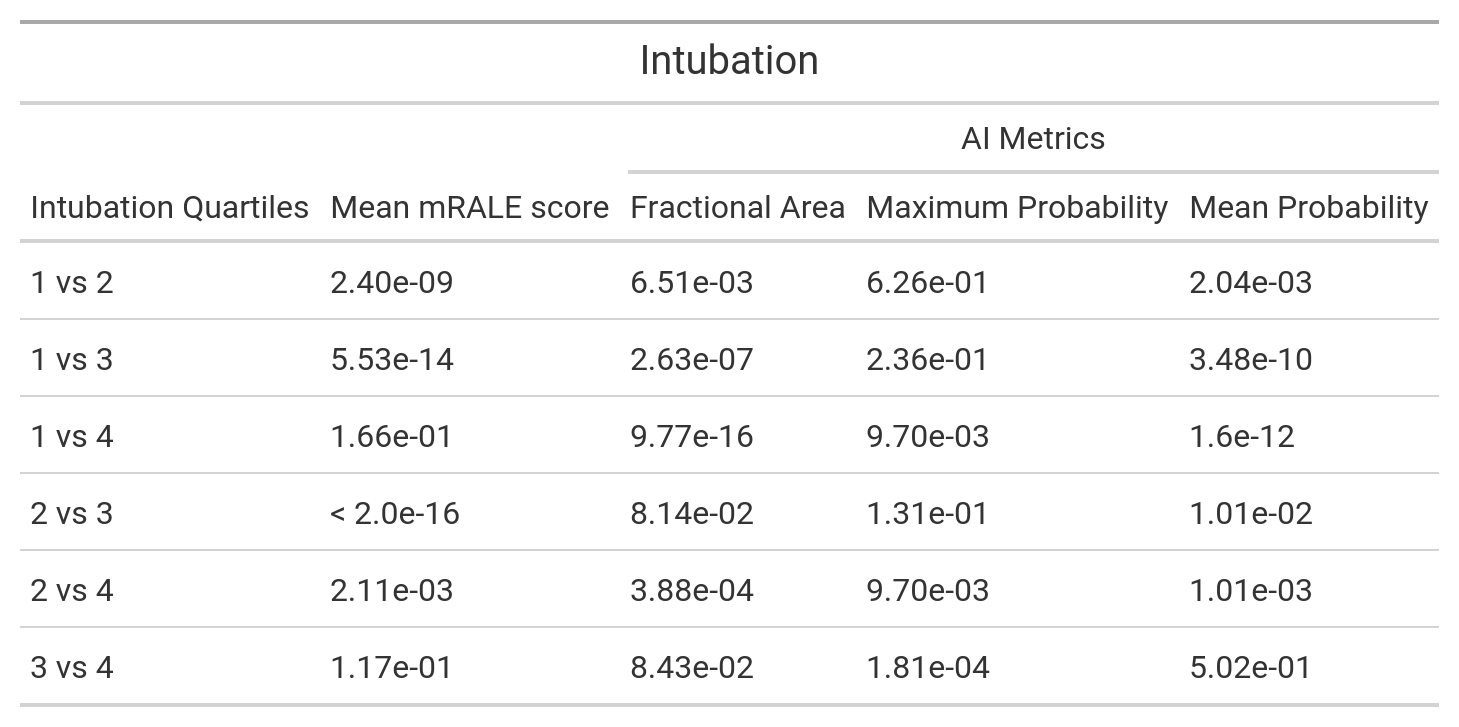
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Loss Function** | **Internal Data Random Split**   **Method** | **External Cases (neg/pos)** | **AUC Improvement** | **Dice Improvement** |
| **30x Pixel-Weighted MSE** | **1** | **1200/600** | 0.0 85 | 0.185 |
| **Kullback-Liebler Divergence** | **1** | **1200/600** | 0. 085 | 0.025 |
| **Inverse Frequency-Weighted MSE** | **1** | **1200/600** | 0.083 | 0.209 |
| **Inverse Frequency-Weighted MSE** | **2** | **2400/600** | 0. 083 | 0.188 |
| **Inverse Frequency-Weighted MSE** | **2** | **1200/600** | 0. 082 | 0.162 |
| **Inverse Frequency-Weighted MSE** | **1** | **2400/1200** | 0. 081 | 0.149 |
| **Kullback-Liebler Divergence** | **2** | **1200/600** | 0. 081 | 0.107 |
| **Binary Focal Loss** | **1** | **1200/600** | 0. 081 | -0.092 |
| **30x Pixel Weighted MSE** | **2** | **1200/600** | 0. 081 | 0.153 |
| **30x Pixel Weighted MSE** | **1** | **2400/600** | 0. 080 | 0.165 |

***Supplemental Table 1. Hyperparameter search and selection of the optimal updated CNN.***The top 10 candidate convolutional neural networks (CNNs) from 102 total CNNs were decided by the amount of area under the curve (AUC) improvement from the internal validation data set. The randomization method (“Internal Data Random Split Method”) to split the internal data into training/validation, choice of loss function, and the number of negative and positive cases in the external data set that were used in each epoch of training were included in the hyperparameter search. A CNN that increased the weighting of positive pixels in small foci of pneumonia, “30x pixel-weighted mean squared error (MSE)” was selected as the “*updated CNN.”* This *updated CNN* maximized both AUC and Dice overlap between the CNN inference and ground truth annotation.

**Supplemental Table 2: Pairwise cooccurrences between mRALE and AI severity score quartiles.** Log-rank pairwise cooccurrence comparison p-values with Benjamini-Hochberg correction between severity quartiles for intubation, hospital duration, and mortality for each severity score (mRALE and AI), computed from the *updated CNN*. Corrected p-values < 0.05 indicate a significant difference between severity quartile survival curves (Figure 8B) for a given metric.