**Supplemental Digital Content 3 Table S1.** Performance of the deep learning-based automatic detection algorithm: error from the ground truth in ETT tip and carina detection and ETT-carina distance of each model during internal 4-fold cross-validation (1,842 images/1,842 patients in total).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metrica | Modelb | Median (IQR), mm | ≤5 mm | ≤10 mm | ≤15 mm |
| ETT tip detection | 1st (n=462) | 2.6 (1.6-4.8) | 77.1% (356) | 92.9% (429) | 96.5% (446) |
| 2nd (n=461) | 2.8 (1.6-5.1) | 73.8% (340) | 91.8% (423) | 95.7% (441) |
| 3rd (n=458) | 2.9 (1.5-5.3) | 72.7% (333) | 92.4% (423) | 95.9% (439) |
| 4th (n=461) | 2.8 (1.5-4.6) | 76.8% (354) | 92.8% (428) | 97.4% (449) |
|  |
| Carina detection | 1st (n=462) | 3.6 (2.1-5.9) | 67.5% (312) | 90.0% (416) | 95.0% (439) |
| 2nd (n=461) | 3.4 (2.2-5.3) | 70.9% (327) | 93.1% (429) | 96.1% (443) |
| 3rd (n=458) | 3.5 (1.9-5.5) | 69.7% (319) | 91.3% (418) | 94.8% (434) |
| 4th (n=461) | 3.8 (2.2-5.6) | 67.2% (310) | 91.8% (423) | 96.3% (444) |
|  |
| ETT-carina distance | 1st (n=462) | 4.0 (1.7-7.2) | 59.3% (274) | 84.4% (390) | 91.1% (421) |
| 2nd (n=461) | 4.0 (1.9-7.4) | 60.0% (278) | 83.3% (384) | 92.6% (427) |
| 3rd (n=458) | 3.8 (1.7-7.0) | 59.8% (274) | 83.4% (382) | 92.6% (424) |
| 4th (n=461) | 3.8 (1.7-6.9) | 62.0% (286) | 85.7% (395) | 94.8% (437) |

ETT, endotracheal tube; IQR, interquartile range.

a Data are expressed as percentage (number) unless otherwise indicated.

b In 4-fold cross-validation, a single fold is retained as the validation dataset for testing the model, and the remaining 3 folds are used as training datasets. For example, the 1st model was trained using the 2nd, 3rd and 4th folds and was tested using the 1st fold. The architecture of the algorithm was consistent during the training process of the 4 models.