**Supplemental Digital Content 4 Table S2.** 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 external validation (216 images/216 patients in total).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metrica | Modelb | Median (IQR), mm | ≤5 mm | ≤10 mm | ≤15 mm |
| ETT tip detection | 1st (n=216) | 3.0 (1.8-5.2) | 73.6% (159) | 91.2% (197) | 95.8% (207) |
| 2nd (n=216) | 2.8 (1.7-5.0) | 73.6% (159) | 90.7% (196) | 94.4% (204) |
| 3rd (n=216) | 3.1 (1.7-5.8) | 70.8% (153) | 88.4% (191) | 93.5% (202) |
| 4th (n=216) | 3.1 (1.7-5.3) | 72.2% (156) | 91.2% (197) | 97.2% (210) |
|  |
| Carina detection | 1st (n=216) | 3.4 (1.9-5.5) | 71.3% (154) | 88.0% (190) | 95.4% (206) |
| 2nd (n=216) | 3.5 (1.9-5.9) | 67.1% (145) | 90.3% (195) | 95.4% (206) |
| 3rd (n=216) | 3.6 (2.1-5.8) | 66.7% (144) | 90.7% (196) | 97.2% (210) |
| 4th (n=216) | 3.5 (2.3-6.1) | 66.2% (143) | 88.0% (190) | 95.8% (207) |
|  |
| ETT-carina distance | 1st (n=216) | 4.1 (1.8-8.0) | 58.8% (127) | 82.9% (179) | 92.6% (200) |
| 2nd (n=216) | 4.2 (1.8-7.7) | 57.9% (125) | 80.6% (174) | 90.7% (196) |
| 3rd (n=216) | 4.3 (1.4-8.3) | 54.6% (118) | 82.9% (179) | 92.6% (200) |
| 4th (n=216) | 3.8 (1.7-7.2) | 59.3% (128) | 86.6% (187) | 94.4% (204) |

ETT, endotracheal tube; IQR, interquartile range.

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

b The model numbers are identical in 4-fold cross-validation (Supplemental Digital Content 3 Table S1) and external validation.