SUPPLEMENTARY MATERIAL 6

The Detailed Statistical Methods

Normally distributed (Kolmogorov-Smirnov test) numerical data are reported as the mean ± standard deviation, otherwise they are expressed as median (range). Differences in sex distribution and slice thickness between the monocentric training and validation set, as well as differences in sex distribution among the validation sets (five multicenter/multiparameter validation sets and one control set) were analyzed using Pearson's chi-squared test. Age was not normally distributed across all datasets, so it is reported as the median (range). The Mann-Whitney U test/Kruskal-Wallis H test was used to compare ages between the monocentric training set and validation set, as well among the validation sets (five multicenter/multiparameter validation sets and one control set), respectively. Differences in the diagnostic performance of Faster R-CNN and YOLOv3 were compared using a paired-samples *t* test. The detection differences for the three categories of rib fractures were analyzed using the Kruskal-Wallis H test and least significant difference post hoc test (rank conversion). The sensitivity and average number of FPs/patient in the CNN model in the whole CT images from 33 patients without merging results are shown using free-response receiver operating characteristic curves, and the 11 points of the structured report and radiologists with and without artificial intelligence (AI) assistance were overlaid. The precision and sensitivity of the structured report and five radiologists without AI assistance were compared using a single sample *t* test, and radiologists with and without AI assistance were compared using the Mann-Whitney U test. In addition, the diagnosis time was analyzed using one-way ANOVA and the paired-samples *t* test was used for self-comparison. A two-tailed p value < 0.05 was statistically significant. Statistical analyses and graphing were performed using R (version 3.5.3, available at https://www.r-project.org/).