

**Supplementary 3.****Average percent systematic difference using pairwise approach**

The pairwise average percent systematic difference  $\delta_{ij}$  was calculated for Bland-Altman analysis. The measure is based on the average difference in measurement between any pair of the  $i$ -th and  $j$ -th radiologists for the  $k$ -th cases as follows.

$$\delta_{ij} = \frac{2}{K} \sum_{k=1}^K \frac{M_{ik} - M_{jk}}{M_{ik} + M_{jk}}$$

Here,  $K$  is the number of cases (in our study  $K = 10$ ), and  $M_{ik}$  is a measurement value of the  $i$ -th radiologist for the  $k$ -th case.

**Bland-Altman outlier scores with standard and fixed-limit**

The standard Bland-Altman outlier scores  $Y_{2SD}$  is reliant on the percentage of pairwise measurement difference less than 2 standard deviations. Similarly, the Bland-Altman scores  $Y_{20\%}$  with 20% fixed limit is reliant on the percentage of pairwise measurement difference less than 20% and calculated as follows:

$$Y_{2SD} = \frac{(N-2)!}{N!} \sum_{i=1}^N \sum_{j=1}^N 1 \left( \frac{|M_{ik} - M_{jk}|}{M_{ik} + M_{jk}} < 2SD \right)$$

$$Y_{20\%} = \frac{(N-2)!}{N!} \sum_{i=1}^N \sum_{j=1}^N 1 \left( \frac{|M_{ik} - M_{jk}|}{M_{ik} + M_{jk}} < 0.2 \right)$$

Here,  $1(A)$  is an indicator function whose value is 1 if  $A$  is true and 0 otherwise.  $N$  represents the number of radiologists. The fixed-limit Bland-Altman outlier scores were based on the percentage of pairs where a pair of radiologists reviewed the same CT image set and resulted in measurements that differ by less than 20%.