

Supplementary Table 1. Agreement in the held-out internal test cohort: (a) Descriptive statistics; (b) mean absolute error of the model against each reader and the three-reader mean; (c) inter-reader agreement; and (d) intraclass correlation coefficients among the model and the three readers

(a) Descriptive statistics					
	Minimum	Maximum	Mean	SD	n
Model	5.47	33.14	21.12	5.52	105
Reader 1	5.66	34.80	21.56	5.79	105
Reader 2	4.72	34.21	21.49	5.77	105
Reader 3	5.20	34.61	21.57	5.72	105
(b) Model vs. readers					
	MAE (95% CI)		Median	Max	
Model vs. 3-reader mean	0.690° (0.525–0.855)		0.458°	6.11°	
Model vs. Reader 1	0.675° (0.509–0.840)		0.381°	5.96°	
Model vs. Reader 2	0.876° (0.705–1.046)		0.637°	6.07°	
Model vs. Reader 3	0.743° (0.576–0.911)		0.483°	6.32°	
(c) Inter-reader agreement					
	MAE (95% CI)		Bias	95% limits of agreement	
Reader 1 vs. Reader 2	0.591° (0.494–0.688)		+0.072°	–1.444° to +1.589°	
Reader 1 vs. Reader 3	0.362° (0.302–0.423)		–0.011°	–0.949° to +0.927°	
Reader 2 vs. Reader 3	0.486° (0.396–0.575)		–0.083°	–1.392° to +1.225°	
(d) Intraclass correlation coefficients					
	ICC		95% CI		
4 raters (model + 3 readers)	0.9863		0.9808–0.9904		
3 readers	0.9936		0.9912–0.9955		
Model–Reader 1 (pairwise)	0.9813		0.9652–0.9890		
CI, confidence interval; ICC, intraclass correlation coefficient; MAE, mean absolute error; SD, standard deviation.					

Supplementary Table 2. Inter-reader agreement in the held-out test cohort

Comparison	MAE (95% CI)	Bias	95% limits of agreement
Reader 1 vs. Reader 2	0.591° (0.494–0.688)	+0.072°	–1.444° to +1.589°
Reader 1 vs. Reader 3	0.362° (0.302–0.423)	–0.011°	–0.949° to +0.927°
Reader 2 vs. Reader 3	0.486° (0.396–0.575)	–0.083°	–1.392° to +1.225°
MAE, mean absolute error; CI, confidence interval.			

Supplementary Table 3. Mean absolute angular error (degrees) produced by random isotropic displacement of each landmark, 200 draws per radiograph per radius over all 105 held-out radiographs

Displacement	P1	P2	P3	P4	All four
1 px (0.14 mm)	0.039	0.040	0.101	0.101	0.140
2 px (0.28 mm)	0.079	0.079	0.205	0.204	0.279
5 px (0.70 mm)	0.198	0.196	0.508	0.512	0.703
10 px (1.39 mm)	0.396	0.398	1.015	1.017	1.394
20 px (2.79 mm)	0.791	0.791	2.034	2.022	2.799
40 px (5.58 mm)	1.589	1.581	4.063	4.039	5.559

Supplementary Table 4. Comparison of design choices on the held-out cohort (n = 105; the multi-class configuration was evaluated on the 100 radiographs it could measure). Per-channel segmentation thresholds were 0.50/0.30/0.40 for the per-class networks and 0.90/0.05/0.50 for the multi-class network

Configuration	Validation dice (CAL / M5 / CC)	Held-out MAE (°)	Within 2° (%)
Per-class networks + deterministic geometry (deployed)	0.936/0.865/0.807	0.690	92.4
Single multi-class network + deterministic geometry	0.932/0.874/0.802	0.68	96.0
Per-class networks + direct angle regression	—	5.00	20.0
CAL, calcaneus inferior; CC, calcaneocuboid joint; M5, head of 5th metatarsal; MAE, mean absolute error.			

Supplementary Table 5. Post-hoc subgroup analysis of absolute error against the three-reader means on the held-out cohort (n = 105)

Subgroup	n	MAE (°)	Within 2° (%)	Max error (°)
Reference CIA < 20°	44	0.59	95.5	2.53
Reference CIA 20–25°	31	0.57	93.5	2.49
Reference CIA ≥ 25°	30	0.97	86.7	6.11
Calcaneal-spur flag present	19	0.57	94.7	2.49
Calcaneal-spur flag absent	86	0.72	91.9	6.11
Female	83	0.69	91.6	6.11
Male	22	0.67	95.5	2.49
CIA, calcaneal inclination angle; MAE, mean absolute error.				