SUPPLEMENTARY MATERIAL 8

The Detailed Performance of Individual Radiologists in the Reader Test

The detailed performance of six radiologists including conventional diagnosis without AI assistance and AI-assisted diagnosis are shown in Tables 1 and 2.

Table 1. Performance of Conventional Diagnosis by 5 Radiologists on Validation Set 1 (n = 33)

Indicators	Radiologists				
	1	2	3	4	5
TPs					
Fresh fractures	33 (25–39)	33 (26–38)	26 (18–32)	34 (27–39)	37 (29–42)
Healing fractures	32 (23–39)	25 (17–32)	36 (28–42)	41 (32–48)	38 (30–44)
Old fractures	21 (13–28)	27 (18–34)	29 (21–36)	31 (23–38)	36 (27–43)
FNs					
Fresh fractures	12 (3–22)	12 (5–19)	19 (11–28)	11 (3–19)	8 (1–15)
Healing fractures	24 (14–36)	31 (22–39)	20 (12–30)	15 (5–25)	18 (9–27)
Old fractures	33 (19–47)	27 (16–38)	25 (13–35)	23 (13–32)	18 (7–27)
FPs					
Fresh fractures	10 (2–21)	2 (0–9)	4 (0–12)	5 (0–14)	4 (0–13)
Healing fractures	24 (12–35)	0 (0–5)	2 (0–10)	9 (1–21)	4 (0–13)
Old fractures	23 (11–40)	18 (8–30)	10 (4–22)	6 (0–16)	9 (3–20)
Precision					
Fresh fractures	33/43 = 0.767 (0.571–0.950)	33/35 = 0.943 (0.750–1.000)	26/30 = 0.867 (0.625–1.000)	34/39 = 0.872 (0.659–1.000)	37/41= 0.902 (0.714–1.000)
Healing fractures	32/56 = 0.571 (0.397–0.766)	25/25 = 1.000 (0.783–1.000)	36/38 = 0.947 (0.744–1.000)	41/50 = 0.820 (0.638–0.977)	38/42 = 0.905 (0.723–1.000)
Old fractures	21/44 = 0.477 (0.259–0.694)	27/45 = 0.600 (0.383–0.800)	29/39 = 0.744 (0.512–0.895)	31/37 = 0.838 (0.610–1.000)	36/45 = 0.800 (0.592–0.933)
Sensitivity					
Fresh fractures	33/45 = 0.733 (0.532–0.929)	33/45 = 0.733 (0.578–0.881)	26/45 = 0.578 (0.400–0.744)	34/45 = 0.756 (0.596–0.927)	37/45 = 0.823 (0.667–0.977)
Healing fractures	32/56 = 0.571 (0.397–0.731)	25/56 = 0.446 (0.304–0.593)	36/56 = 0.643 (0.483–0.778)	41/56 = 0.732 (0.561–0.904)	38/56 = 0.679 (0.526–0.824)
Old fractures	21/54 = 0.389 (0.236–0.587)	27/54 = 0.500 (0.333–0.660)	29/54 = 0.537 (0.382–0.729)	31/54 = 0.574 (0.418–0.745)	36/54 = 0.667 (0.509–0.854)
F1-score					
Fresh fractures	1.124/1.500 = 0.749 (0.625–0.830)	1.382/1.676 = 0.825 (0.712–0.894)	1.002/1.445 = 0.693 (0.537–0.790)	1.318/1.628 = 0.810 (0.701–0.876)	1.485/1.725 = 0.861 (0.744–0.923)
Healing fractures	0.652/1.142 = 0.571 (0.447–0.656)	0.892/1.446 = 0.619 (0.466–0.727)	1.218/1.590 = 0.766 (0.651–0.840)	1.200/1.552 = 0.773 (0.660–0.850)	1.229/1.584 = 0.776 (0.667–0.846)
Old fractures	0.371/0.866 = 0.428 (0.289–0.533)	0.600/1.100 = 0.545 (0.400–0.642)	0.799/1.281 = 0.624 (0.494–0.720)	0.962/1.412 = 0.681 (0.554–0.776)	1.067/1.467 = 0.727 (0.600–0.811)

Corresponding 95% confidence intervals, shown inside parentheses, were estimated by using bootstrapping with 1000 bootstraps and randomly sampled at lesions level.

Table 2. Performance of AI-Assisted Diagnosis by 5 Radiologists on Validation Set 1 (n = 33)

Indicators	Radiologists				
	1	2	3	4	5
TPs					
Fresh fractures	42 (36–46)	41 (34–45)	41 (33–46)	42 (36–45)	42 (36–46)
Healing fractures	50 (42–55)	51 (43–56)	49 (42–53)	54 (46–59)	52 (46–55)
Old fractures	41 (33–46)	37 (29–42)	41 (33–47)	37 (29–43)	47 (39–52)
FNs					
Fresh fractures	3 (0–8)	4 (0–12)	4 (0–11)	3 (0–9)	3 (0–8)
Healing fractures	6 (1–15)	5 (0–12)	7 (2–14)	2 (0–7)	4 (0–11)
Old fractures	13 (3–22)	17 (9–25)	13 (6–21)	17 (8–26)	7 (1–14)
FPs					
Fresh fractures	4 (1–11)	6 (1–15)	9 (2–18)	1 (0–6)	6 (1–14)
Healing fractures	7 (0–16)	6 (1–14)	1 (0–7)	10 (4–19)	1 (0–6)
Old fractures	3 (0–13)	2 (0–10)	3 (0–12)	3 (0–11)	5 (0–14)
Precision					
Fresh fractures	42/46 = 0.913 (0.771–0.979)	41/47 = 0.872 (0.694–0.978)	41/50 = 0.820 (0.647–0.958)	42/43 = 0.977 (0.861–1.000)	42/48 = 0.875 (0.726–0.977)
Healing fractures	50/57 = 0.877 (0.729–1.000)	51/57 = 0.895 (0.754–0.982)	49/50 = 0.980 (0.863–1.000)	54/64 = 0.844 (0.708–0.937)	52/53 = 0.981 (0.885–1.000)
Old fractures	41/44 = 0.932 (0.729–1.000)	37/39 = 0.949 (0.769–1.000)	41/44 = 0.932 (0.745–1.000)	37/40 = 0.925 (0.744–1.000)	47/52 = 0.904 (0.750–1.000)
Sensitivity					
Fresh fractures	42/45 = 0.933 (0.818–1.000)	41/45 = 0.911 (0.750–1.000)	41/45 = 0.911 (0.766–1.000)	42/45 = 0.933 (0.800–1.000)	42/45 = 0.933 (0.826–1.000)
Healing fractures	50/56 = 0.892 (0.737–0.982)	51/56 = 0.911 (0.800–1.000)	49/56 = 0.875 (0.750–0.964)	54/56 = 0.964 (0.868–1.000)	52/56 = 0.928 (0.807–1.000)
Old fractures	41/54 = 0.759 (0.607–0.932)	37/54 = 0.685 (0.546–0.824)	41/54 = 0.759 (0.611–0.887)	37/54 = 0.685 (0.544–0.843)	47/54 = 0.870 (0.741–0.981)
F1-score					
Fresh fractures	1.704/1.846 = 0.923 (0.847–0.968)	1.589/1.783 = 0.891 (0.800–0.938)	1.494/1.731 = 0.863 (0.759–0.920)	1.823/1.910 = 0.954 (0.878–0.989)	1.632/1.808 = 0.903 (0.828–0.949)
Healing fractures	1.565/1.769 = 0.885 (0.800–0.932)	1.630/1.860 = 0.876 (0.819–0.949)	1.715/1.855 = 0.925 (0.849–0.964)	1.627/1.808 = 0.865 (0.821–0.944)	1.821/1.961 = 0.929 (0.893–0.982)
Old fractures	1.415/1.691 = 0.837 (0.733–0.893)	1.300/1.634 = 0.796 (0.682–0.857)	1.415/1.691 = 0.837 (0.733–0.904)	1.267/1.610 = 0.787 (0.674–0.860)	1.573/1.774 = 0.887 (0.796–0.937)

Corresponding 95% confidence intervals, shown inside parentheses, were estimated by using bootstrapping with 1000 bootstraps and randomly sampled at lesions level.