

Impact of artificial intelligence assistance on clinician performance in diagnosing developmental dysplasia of the hip

Introduction: Artificial intelligence (AI) models have shown reasonable ability in identifying developmental dysplasia of the hip (DDH). This study evaluates the impact of AI-assisted image interpretation on the diagnostic performance of 'expert' and 'non-expert' clinicians.

Methods: A multi-reader, multi-case study was conducted with 10 readers (5 consultants and 5 registrars). The study included 70 static 2D hip ultrasound scans from 70 unique patients (age range 0.86-19.86 weeks, mean age 8.6 weeks, 55.7% female and 65.7% left). There were 35 normal (Graf 1) and 35 abnormal (Graf 2 n=25; Graf 3/4 n=10) images. The reference standards were the Graf alpha angle and femoral head coverage (FHC) derived from anatomical points (landmarks) checked by two senior experts. The clinicians placed landmarks without (Phase1) and with AI assistance (Phase2). Their diagnostic performance (sensitivity/specificity), confidence rating (1-5) and time taken were reported.

Results: For the Graf method, the mean sensitivity/specificity values for 'experts' without vs with AI were 0.75/0.93 vs 0.71/0.96, respectively. For 'non-experts', this was 0.75/0.90 vs 0.70/0.95. For FHC%, the sensitivity/specificity values for 'experts' (0.97/0.92) and 'non-experts' (0.98/0.93) were unchanged between phases. AI assistance significantly increased the rate that all clinicians reported the highest level of confidence (5) in their annotations. The mean time taken (hh:mm) by 'experts' without vs with AI was 04:49 vs 05:07, for 'non-experts' it was 05:04 vs 04:52.

Conclusions: The FHC demonstrated better overall sensitivity as a screening method. AI assistance improved the diagnostic confidence of all clinicians without reducing their performance and made 'non-experts' faster.