

## **Can artificial intelligence outperform clinicians in interpreting newborn hip screening ultrasounds? Pilot study with external validation results**

Singh Abhinav, Clement Allison, Wynell-Mayow William, Perry Daniel, Eastwood Deborah, Voiculescu Irina

**Background:** The UK screening programme has not improved early detection of developmental dysplasia of the hip (DDH) since 1986. Artificial Intelligence (AI) methods show reasonable performance in calculating Graf angles and diagnostic class from predicted anatomical landmarks. This study evaluated AI's performance after training it on additional diagnostic data alongside labelled images.

**Methods:** Our study utilised two routine anonymised datasets containing hip ultrasound 2D images from babies aged 4-12 weeks. Dataset A (DA) had 516 images from 397 babies from one hospital and dataset B (DB) had 50 images from 34 babies from a different hospital. Two clinicians (X and Y) labelled the five anatomical landmarks which represent the ilium (1<sup>st</sup> and 2<sup>nd</sup> points), turning point (3<sup>rd</sup>), lower limb (4<sup>th</sup>) and labrum (5<sup>th</sup>). Labelled images were quality-checked by two different blinded senior clinicians before inclusion. The AI model was trained with both landmarks and weighted diagnostic data. Its output was the alpha angle and Graf class (centred or decentred). The performance between clinicians (X vs Y) and AI vs clinician X were measured.

**Results:** For the DA test dataset (n=77), the best-performing model predicted alpha to within 2 degrees of the clinician more frequently (81.9%) than when clinicians (X vs Y) were compared to each other (33.2%). This performance was maintained on the smaller (n=50) external DB dataset (93.7%). The AI's prediction of the Graf class was also superior to the clinician on both DA (90.2% sensitivity, 66.7% specificity) and DB (93% sensitivity, 80% specificity).

**Conclusions:** Our AI model outperformed clinicians in generating the alpha angles and Graf class on both datasets. Additional diagnostic data may allow the model to better understand the relationship between anatomical landmarks.

**Implications:** When translated into the clinical setting, this method will improve the accuracy and objectivity of DDH screening.

**Disclosure:** None.