

Automatic identification of clinical landmarks and calculation of Graf angles in newborn hip screening: A pilot study

Singh Abhinav, McCouat James, Perry Daniel, Hemmadi Sandeep, Voiculescu Irina

Background: The UK screening programme has not improved early detection of developmental dysplasia of the hip (DDH) since 1986. Late diagnosis leads to multiple operations and persistent pain. Artificial Intelligence (AI) assisted diagnostics represents the potential for a universal ultrasound screening programme with improved accuracy and objectivity.

Methods: This study utilised an anonymised dataset obtained by experienced radiologists during routine NHS practice. The dataset contained 190 2D hip ultrasound images (normal= 71, dysplastic= 66, dislocated= 53) from 100 multi-ethnic babies (aged 4-12 weeks). Two clinical experts provided the five anatomical landmarks representing base of ilium (1st and 2nd points), turning point (3rd), lower limb (4th) and labrum (5th). Corresponding Graf angles were used to classify screening diagnosis - normal (Graf 1/2a) or abnormal (Graf 2b,2c,D,3,4). We designed a convolutional neural network (modified U-Net, 94:47:49 training:validation:testing data split) to detect anatomical landmarks for automatically calculating Graf angles and providing a screening diagnosis.

Results: Within the test dataset (n=49), the five landmarks were identified in all images with a mean pixel error of 7.1. When comparing Graf angles between ground truth and AI-derived landmarks the mean difference for alpha was 2.8 degrees (range 0-7.6; interclass correlation coefficient (ICC) 0.88) and beta was 5.4 (range 0.3-16; ICC 0.87) degrees. Automated Graf angle diagnosis agreed with clinical experts in 85.7% (n=42) of the images with a sensitivity of 90.6% and specificity of 76%.

Conclusions: Our AI successfully identifies key landmarks with minimal pixel error. It automatically calculates Graf angles, exceeding the existing state of the art AI performance and provides a diagnosis which agrees with expert clinicians on 85.7% of images.

Implications: This method will improve the accuracy of DDH screening programme, whilst reducing clinician fatigue and improving resource allocation. Objective data will also facilitate high-quality outcomes research.

Disclosure: None.