



Automated Dynamic 3D Ultrasound Assessment of Developmental Dysplasia of the Infant Hip

Olivia Paserin¹(✉), Kishore Mulpuri², Anthony Cooper², Antony J. Hodgson³,
and Rafeef Garbi¹

¹ Department of Electrical and Computer Engineering,
University of British Columbia, Vancouver, Canada
opaserin@ece.ubc.ca

² Department of Orthopaedic Surgery, British Columbia Children's Hospital,
Vancouver, Canada

³ Department of Mechanical Engineering,
University of British Columbia, Vancouver, Canada

Abstract. Dynamic two-dimensional sonography of the infant hip is a commonly used procedure for developmental dysplasia of the hip (DDH) screening by many clinicians. It however has been found to be unreliable with some studies reporting associated misdiagnosis rates of up to 29%. Aiming to improve reliability of diagnosis and to help in standardizing diagnosis across different raters and health-centers, we present a preliminary automated method for assessing hip instability using three-dimensional (3D) dynamic ultrasound (US). To quantify hip assessment, we propose the use of femoral head coverage variability (ΔFHC_{3D}) within US volumes collected during a dynamic scan which uses phase symmetry features to approximate the vertical cortex of the ilium and a random forest classifier to identify the approximate location of the femoral head. We measure the change in FHC_{3D} across US scans of the hip acquired under posterior stress vs. rest as maneuvered during a 3D dynamic assessment. Our findings on 38 hips from 19 infants scanned by one orthopedic surgeon and two radiology technicians suggests the proposed ΔFHC_{3D} may provide a good degree of repeatability with an average test-retest intraclass correlation measure of 0.70 (95% confidence interval: 0.35 to 0.87, $F(21, 21) = 7.738$, $p < 0.001$). This suggests that our 3D dynamic dysplasia metric may prove valuable in improving reliability in diagnosing hip laxity due to DDH, which may lead to a more standardized DDH assessment with better diagnostic accuracy. The long-term significance of this approach to evaluating dynamic assessments may lie in increasing early diagnostic sensitivity in order to prevent dysplasia remaining undetected prior to manifesting itself in early adulthood joint disease.

Keywords: Pediatric · Ultrasound · Hip · Bone imaging
Developmental dysplasia of the hip · DDH · Dynamic assessment
Repeatability

1 Introduction

Performing dynamic sonography on the infant hip is a routine part of screening for developmental dysplasia of the hip (DDH) in many clinical settings [1]. However, such screening has been shown to be unreliable. For example, in a study on a cohort of 266 infants by Imrie et al. [2]., dynamic assessment was associated with misdiagnosis rates of 29% where infants who had been screened as healthy were later found to display sufficient signs of dysplasia to require treatment. In a dynamic assessment of an infant's hip, clinicians apply stress to the adducted hip, in a posterior direction to provoke dislocation, and observe the resulting joint movement with ultrasound (US) [3]. Barlow (dislocation) and Ortolani (reduction) maneuvers have become the basis for the clinical classification of hip abnormality [3]. Resulting observations are currently described qualitatively using terms such as normal, lax, dislocatable, reducible and not reducible, and are not based on measured quantities. Evaluating the hip's stability dynamically is crucial because in order to avoid DDH, the development of the neonatal cartilaginous acetabulum must occur around "*a properly seated femoral head*" [4]. In a complementary manner, Graf's method [5] is the standardized method for assessing acetabular morphology of the infant hip using US-based angle measurements to estimate the depth of the acetabular socket during a static assessment. However, characterizing acetabular morphology with Graf's technique alone does not evaluate or screen for loose ligaments supporting the hip - a factor in hip dysplasia. As such, dynamic assessment is recommended as a routine part of every infant hip clinical exam [3]. In a recent study by Alamdaran et al. [6]), 100% of hips with dysplastic morphology (mild and severe) were unstable in dynamic analysis while 9% of unstable hips had normal morphology in static evaluation, supporting the need for dynamic assessments to be performed on every hip that appears morphologically normal as they may reduce the number of missed DDH cases [7].

In a recent systematic review, Charlton et al. [8] examined dynamic US screening for hip instability in the first six weeks after birth and found current best practices for such early screening techniques to be still divergent between different institutions in terms of clinical scanning protocols, namely the most appropriate scanning plane and position, diagnostic metrics, patient age to scan, and followup procedures, used internationally. They in fact identified nearly 20 early dynamic US screening techniques present in the literature each with different imaging and measurement protocols. To the best of our knowledge, all previous dynamic assessment studies employed two-dimensional (2D) ultrasonography. However, it has been recently shown that using three-dimensional (3D) US can markedly improve the reliability of dysplasia metric measurements during static assessment of an infant's hip compared to 2D US as volumetric scans can capture the entire hip joint and are less prone to probe orientation errors compared to 2D scans [9,10]. Further, none of the previous dynamic studies explored automating the assessment via computational image analysis despite inter-assessor variability likely accounting for much of the poor reproducibility of dynamic assessments and the related rates of misdiagnosis.

Our main objective in this work is to develop and evaluate an automated quantitative method for assessing hip instability volumetrically in a dynamic examination in order to improve reliability of diagnosis and reduce misdiagnosis rates. To enable this, we developed an approach to automatically calculate femoral head coverage (FHC_{3D}) from volumetric US scans, a ratio describing how much of the femoral head sits in the acetabular cup of the hip joint [11]. While the intra-rater variability of automatic FHC_{3D} measurements in static assessments was found to be 5.4% and the inter-rater variability to be 6.1% [11], the variability of the measurement during dynamic assessment it is not yet known, nor is the range of differences in FHC_{3D} during dynamic evaluations across normal and dysplastic hips. In an initial feasibility study [12] we recently used our automatic FHC_{3D} measurement technique to perform a dynamic assessment on a single patient, but this protocol did not include a repeated measurement. In this paper, therefore, we report on a clinical study evaluating the test-retest repeatability of an automatic technique for estimating *change* in FHC_{3D} during dynamic assessments in a larger and more clinically representative cohort of patients. We estimate hip joint laxity by quantifying the change in FHC_{3D} observed during a dynamic assessment as the hip is posteriorly stressed by a clinician. With stress applied to a stable joint, we expect femoral head coverage to vary minimally while an unstable hip would show larger changes in the measurement during distraction.

2 Materials and Method

2.1 Data Acquisition and Experimental Setup

In this study, one pediatric orthopedic surgeon and two technologists from the radiology department at British Columbia Children’s Hospital participated in collecting B-mode 3D US images of 38 infant hips from 19 patients. All patients had been referred to the orthopedic clinic due to DDH risk factors and/or clinical suspicion for DDH in one or both hips. Patient inclusion criteria included being between the ages of 0–4 months and attending an appointment for suspected or confirmed DDH. The principal exclusion criterion was that subjects had not received a diagnosis of a genetic syndrome, since patients with genetic syndromes often have abnormal hips due to non DDH-related conditions and we wanted to eliminate that as a potential confounder. Parents were informed of our research goals and protocols, as well as their right to withdraw their consent at any time during the imaging procedure without affecting their child’s clinical care. Age at scan, sex, born by caesarian section, breech birth position, and birth order patient demographics were recorded and are summarized in Table 1. 79% of the patients were female. Interestingly, the majority (68%) of referred patients were first born and had no familial history of DDH.

3D US volumes were obtained as part of routine clinical care under appropriate institutional review board approval. Data collection for our study coincided with patients’ regular clinic visit and increased each appointment duration by approximately five minutes. 3D US volumes were collected using a SonixTouch

Table 1. Summary demographics of our 19 patient cohort, including developmental dysplasia of the hip risk factors.

Patient demographics in our clinical study	
Age	0 to 4 months
Sex	14 female; 5 male
Affected hip	8 left; 5 right; 6 bilateral
Familial history of DDH	3 yes; 13 no; 3 unknown
First born child	13 yes; 6 no
Breech presentation	11 yes; 8 no
Caesarian section	11 yes; 8 no

Q+ scanner (BK ultrasound, Analogic Inc., Peabody, MA, USA) with a 4DL14-5/38 linear 4D transducer set at 7 MHz. The probe was held laterally and positioned in the coronal plane with the infant laying on their side with their hip flexed. Each acquired volume comprised 245 slices with an axial resolution of 0.17 mm and an in-plane resolution of 256×256 pixels corresponding to a physical slice dimension of 38×38 mm.

To investigate test-retest repeatability, each hip examination involved two dynamic assessments performed by one rater (out of the three participating sonographers). The pediatric orthopaedic surgeon, first radiology technologist, and second radiology technologist performed the assessment on twelve, three, and four of the 19 patients, respectively. Each dynamic assessment involved acquiring two 3D US volumes - one with and one without stress applied to the joint in an effort to observe maximal displacement (i.e. we acquired four 3D US volumes for each hip and eight 3D US volumes in total for each patient). We did not evaluate inter-rater repeatability due to the additional time that would have been required from attending clinical staff and families for each patient visit.

2.2 Identifying Adequate Volumes

In order to evaluate test-retest repeatability, we required all four volumes (test neutral, test stressed, retest neutral, retest stressed) from each hip to be adequate for interpretation as in [13]. Volumes were independently classified as adequate vs. inadequate using the deep learning-based classifier presented in [13] found to perform with 82% accuracy compared with an expert radiologist’s labels as ground truth. The classifier was comprised of five convolutional layers to extract hierarchical features from a scan, followed by a recurrent, long short-term memory layer to capture the spatial relationship of their responses.

2.3 Femoral Head and Ilium Segmentation

To automatically estimate FHC_{3D} in each US volume, we used the method proposed by [11]. Before segmenting anatomical structures, we first extracted

a 3D bone boundary of the hip joint using a rotation-invariant local symmetry feature, structured phase symmetry, which extracts sheet-like hyperechoic responses in the volume including bone boundaries, cartilage boundaries, and soft tissue interfaces [14]. The bone boundaries were isolated using attenuation-based post-processing.

From the extracted bone boundaries, we next identified a planar approximation to the vertical cortex of the ilium using geometric priors and a M-estimator sample consensus (MSAC) algorithm [15]. Next, we extracted a voxel-wise probability map characterizing the likelihood of a voxel belonging to the femoral head, a hypoechoic spherical structure localized with a trained random forest classifier.

Once both the ilium and femoral head were segmented, we then used both these structures to calculate FHC_{3D} as the ratio of femoral head portion medial to plane of the ilium as illustrated in Fig. 1. Finally, we estimate the joint laxity by computing $\Delta FHC_{3D} = FHC_{neutral} - FHC_{stressed}$ where $FHC_{neutral}$ was measured from the volume in which no stress was applied and $FHC_{stressed}$ was measured from the volume in which the clinician applied stress in a direction posterior to the hip joint.

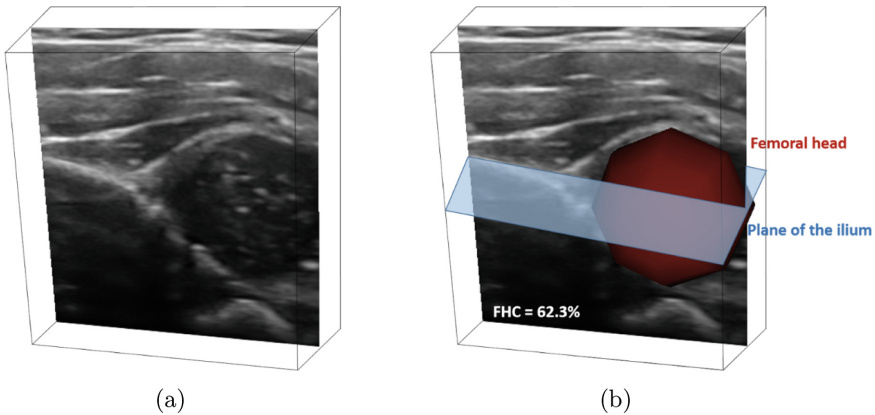


Fig. 1. (a) Three-dimensional visualization of the raw ultrasound (US) data with one coronal slice from the volume displayed. (b) Overlay of the example US volume and its automatically extracted femoral head and planar ilium.

2.4 Quantifying Joint Laxity

Our proposed FHC_{3D} ratio is again illustrated in Fig. 2; FHC_{3D} was calculated as the ratio of the femoral head portion medial to the plane of the ilium. Examples of both unstable and stable hips are shown. With stress applied to a stable hip joint, we expect FHC_{3D} to vary minimally. On the other hand, an unstable hip would show large changes in FHC_{3D} as the femoral head does not sit well in the acetabular socket. Hence, we used ΔFHC_{3D} to quantify joint stability, where $\Delta FHC_{3D} = FHC_{neutral} - FHC_{stressed}$.

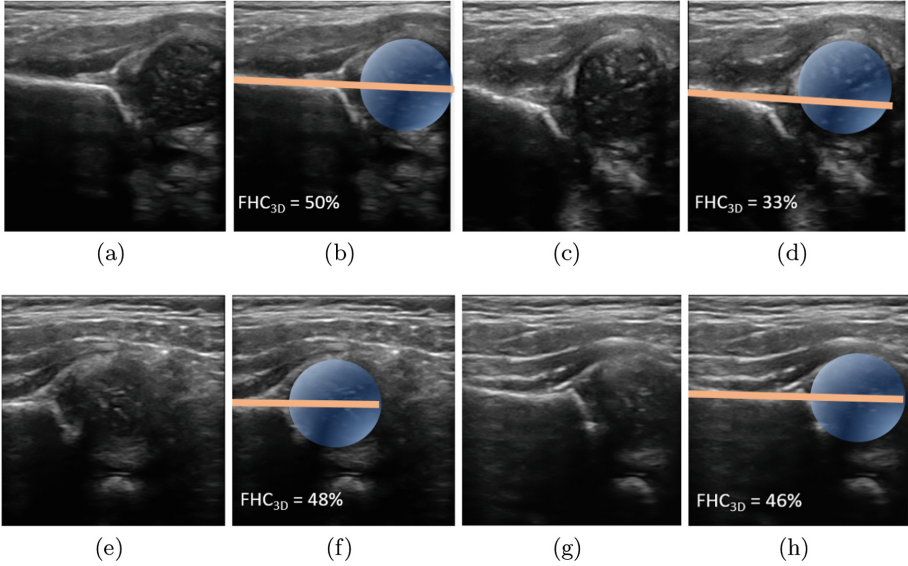


Fig. 2. Qualitative results. (a)–(d) Hip demonstrating 17% change in FHC_{3D} . (a) Raw ultrasound (US) volume with hip at rest. (b) Overlay of the femoral head and planar ilium segmentations. (c) Raw US volume with the hip stressed posteriorly. (d) Overlay of the femoral head and planar ilium segmentations. (e)–(h) Hip demonstrating 2% change in FHC_{3D} . (e) Raw US volume with hip at rest. (f) Overlay of the femoral head and planar ilium segmentations. (g) Raw US volume with the hip stressed posteriorly. (h) Overlay of the femoral head and planar ilium segmentations.

3 Results and Discussion

Our resulting FHC_{3D} and α_{3D} from all recorded repeated dynamic assessments are plotted in Figs. 3 and 4. In post-clinical visit analysis, half of our dynamic assessment recordings were found to have had one or more of the four volumes classified as inadequate, which demonstrates that, despite the increased viewing volumes that 3D scan provide, it is nonetheless still difficult for clinicians to reliably acquire high quality 3D US volumes. We found that most inadequate volumes were acquired during sessions where a new sonographer was collecting the data, which suggests that additional training in operating the 3D US probe may be required. This left us with seventeen test-retest assessments that had all four required dynamic volumes successfully classified as adequate; these were therefore included in the statistical analysis below.

High test-retest repeatability is a necessary requirement for demonstrating the utility of a quantified dynamic assessment diagnostic tool. Using the well-established intra-class correlation coefficient (ICC) performance measures [16], we quantify the reliability of the dynamic assessment done with 3D US. A good degree of reliability was found between the measurements. The test-retest

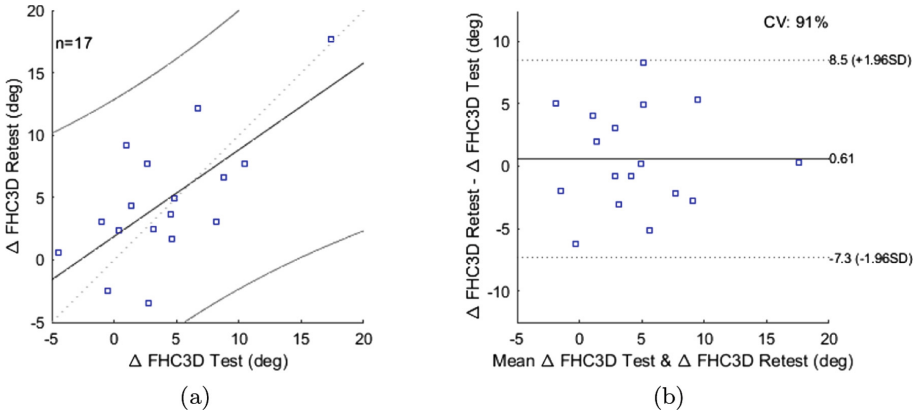


Fig. 3. FHC_{3D} measurements and Bland-Altman plot (ΔFHC_{3D} repeatability). (a) Scatter plot of test-retest FHC_{3D} measurements. Every dot represents two measurements acquired by one rater for one hip. Solid line shows line of best fit and dotted line shows 1–1 line. Curved lines show fit line confidence intervals. (b) Bland-Altman plot for test-retest FHC_{3D} measurements. The solid line indicates the mean difference ($M = 0.61$), dashed lines mark mean difference ± 1.96 standard deviations (SDs). $SD = 4.05$. CV is coefficient of variation (SD of mean values as a percentage).

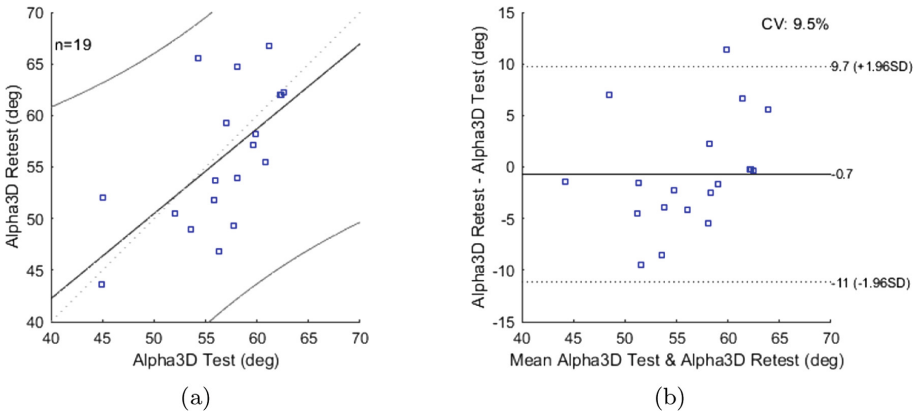


Fig. 4. α_{3D} measurements and Bland-Altman plot (α_{3D} repeatability). (a) Scatter plot of test-retest α_{3D} measurements. Every point represents two measurements acquired by one rater for one hip. Solid line shows line of best fit and dotted line shows 1–1 line. Curved lines show fit line confidence intervals. (b) Bland-Altman plot for test-retest α_{3D} measurements. The solid line indicates the mean difference ($M = -0.70$), dashed lines mark mean difference ± 1.96 standard deviations (SDs). $SD = 5.33$. CV is coefficient of variation (SD of mean values as a percentage).

ICC measure of 0.70 (95% confidence interval: 0.35 to 0.87, $F(21, 21) = 7.738$, $p < 0.001$). ICC estimates and their 95% confident intervals were calculated using MATLAB (Mathworks Inc., Natick, MA, USA) based on a single measurement, absolute-agreement, 2-way mixed-effects model. As shown in Fig. 3, mean difference of ΔFHC_{3D} measurements was 0.61 with standard deviation (SD) 4.05. This suggests that the proposed metric and technique likely have sufficient resolution and repeatability to quantify differences in laxity between stable and mildly unstable hips, since the observed changes in ΔFHC_{3D} range up to about 18% in this cohort. Due to the unbalanced number of volumes recorded by each observer, we have not included their results separately, however this comparison analysis will be performed and reported in future work once we acquire a larger dataset.

Additionally, we automatically computed the α_{3D} measurements on each hip and measured a test-retest ICC measure of 0.61 (95% confidence interval: 0.22 to 0.83, ($F(18, 18) = 4.05$, $p < 0.01$) for α_{3D} . This calculation was also based on a single measurement, absolute-agreement, 2-way mixed-effects model. As shown in Fig. 4, mean difference of repeated α_{3D} measurements was 0.70 with SD 5.33. We found that the most unstable hip, as determined by ΔFHC_{3D} with a change of 17% during the dynamic assessment, corresponded to the most dysplastic hip, as determined by α_{3D} with an angle of less than 40° , demonstrating agreement between the two diagnostic metrics (ΔFHC_{3D} and α_{3D}) in that case. In future work, we plan to collect more US-recorded dynamic assessments in order to determine a reliable range of ΔFHC_{3D} for both stable and dysplastic infant hips.

Computation Complexity. The process of segmenting one US volume and extracting FHC_{3D} took approximately 100 seconds when run on a Intel (R) Xeon(R) 3.70 GHz CPU computer with 8 GB RAM. All processes were executed using MATLAB 2018a. Current practice has a sonographer process the images post-acquisition, so this computation time is not a significant barrier to implementation as it is not necessary to deliver the head coverage metric in near-real time. Nonetheless, although not critical for clinical use, we plan to work towards optimizing our code with graphics processing unit (GPU) parallel programming to significantly reduce this computation time. Volume adequacy classification was performed in one second per volume, a time suitable for clinical workflow, although this was not implemented in near-real time in this study, but was performed post-facto. This time was achieved on a Intel(R) Core(TM) i7-7800X 3.50 GHz CPU, with a NVIDIA TITAN Xp GPU and 64 GB of RAM.

4 Conclusions

We presented an automatic 3D dysplasia metric, ΔFHC_{3D} , to characterize DDH from 3D US images of the neonatal hip through a 3D dynamic assessment procedure. In previous studies [9], α_{3D} was reported to have an intra-rater reliability of 2.2° ($p < 0.01$) and an inter-rater reliability of 2.35° ($p < 0.01$). We suspect that

the observed increase in variability in our study was due to the increased amount of movement introduced during dynamic assessment, as the values reported in [9] were acquired during static assessments only. Mean difference of ΔFHC_{3D} measurements was 0.61 with SD 4.05, with ΔFHC_{3D} values ranging from 0 to 17%. Using the proposed ΔFHC_{3D} we achieved a good degree of reliability. This suggests that this 3D dynamic dysplasia metric could be valuable in improving the reliability in diagnosing hip laxity due to DDH, which may lead to a more standardized DDH assessment with better diagnostic accuracy.

Acknowledgements. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X GPU used for this research.

References

1. Bracken, J., Ditchfield, M.: Ultrasonography in developmental dysplasia of the hip: what have we learned? *Pediatr. Radiol.* **42**(12), 1418–1431 (2012). <https://doi.org/10.1007/s00247-012-2429-8>
2. Imrie, M., Scott, V., Stearns, P., Bastrom, T., Mubarak, S.: Is ultrasound screening for DDH in babies born breech sufficient? *J. Child. Orthop.* **4**(1), 3–8 (2010). <https://doi.org/10.1007/s11832-009-0217-2>
3. Harcke, H., Grissom, L.: Performing dynamic sonography of the infant hip. *AJR Am. J. Roentgenol.* **155**(4), 837–844 (1990). <https://doi.org/10.2214/ajr.155.4.2119119>
4. Gomes, H., Ouedraogo, T., Avisse, C., Lallemand, A., Bakhache, P.: Neonatal hip: from anatomy to cost-effective sonography. *Eur. Radiol.* **8**(6), 1030–1039 (1998). <https://doi.org/10.1007/s003300050510>
5. Graf, R.: New possibilities for the diagnosis of congenital hip joint dislocation by ultrasonography. *J. Pediatr. Orthop.* **3**(3), 354–359 (1983)
6. Alamdaran, S., Kazemi, S., Parsa, A., Moghadam, M., Feyzi, A., Mardani, R.: Assessment of diagnostic value of single view dynamic technique in diagnosis of developmental dysplasia of hip: a comparison with static and dynamic ultrasound techniques. *Arch. Bone Joint Surg.* **4**(4), 371–375 (2016)
7. Koşar, P., Ergun, E., Unlübay, D., Koşar, U.: Comparison of morphologic and dynamic US methods in examination of the newborn hip. *Diagn. Interv. Radiol.* **15**(4), 284–289 (2009). <https://doi.org/10.4261/1305-3825.DIR.2557-09.2>
8. Charlton, S., Schoo, A., Walters, L.: Early dynamic ultrasound for neonatal hip instability: implications for rural Australia. *BMC Pediatr.* **17**, 82 (2017). <https://doi.org/10.1186/s12887-017-0830-z>
9. Quader, N., Hodgson, A., Mulpuri, K., Cooper, A., Abugharbieh, R.: Towards reliable automatic characterization of neonatal hip dysplasia from 3D ultrasound images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016*. LNCS, vol. 9900, pp. 602–609. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-46720-7.70>
10. Jaremko, J., Mabee, M., Swami, V., Jamieson, L., Chow, K., Thompson, R.: Potential for change in US diagnosis of hip dysplasia solely caused by changes in probe orientation: patterns of alpha-angle variation revealed by using three-dimensional US. *Radiology* **273**(3), 870–878 (2014). <https://doi.org/10.1148/radiol.14140451>

11. Quader, N., Hodgson, A.J., Mulpuri, K., Cooper, A., Abugharbieh, R.: A 3D femoral head coverage metric for enhanced reliability in diagnosing hip dysplasia. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 100–107. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_12
12. Paserin, O., Mulpuri, K., Cooper, A., Abugharbieh, R., Hodgson, A.: Improving 3D ultrasound scan adequacy classification using a three-slice convolutional neural network architecture. In: Zhan, W., Rodriguez Y Baena, F. (eds.) Proceedings of Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery - CAOS 2018, EPiC Series in Health Sciences, vol. 2, pp. 152–156 (2018). <https://doi.org/10.29007/2tct>
13. Paserin, O., Mulpuri, K., Cooper, A., Hodgson, A.J., Garbi, R.: Real time RNN based 3D ultrasound scan adequacy for developmental dysplasia of the hip. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 365–373. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_42
14. Quader, N., Hodgson, A., Abugharbieh, R.: Confidence weighted local phase features for robust bone surface segmentation in ultrasound. In: Linguraru, M.G., et al. (eds.) CLIP 2014. LNCS, vol. 8680, pp. 76–83. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13909-8_10
15. Torr, P., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.* **78**(1), 138–156 (2000). <https://doi.org/10.1006/cviu.1999.0832>
16. McGraw, K., Wong, S.: Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1**(1), 30–46 (1996). <https://doi.org/10.1037/1082-989X.1.1.30>