

## Article

# Can AI Automatically Assess Scan Quality of Hip Ultrasound?

Abhilash Rakkunedeth Hareendranathan <sup>1,\*</sup>, Myles Mabee <sup>2</sup>, Baljot S. Chahal <sup>1</sup> , Sukhdeep K. Dulai <sup>1</sup> and Jacob L. Jaremko <sup>1</sup> 

<sup>1</sup> Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, AB T6G 2B7, Canada; baljot@ualberta.ca (B.S.C.); sdulai@alberta.ca (S.K.D.); jjaremko@ualberta.ca (J.L.J.)

<sup>2</sup> College of Medicine, University of Saskatchewan, 107 Wiggins Rd, Saskatoon, SK S7N 5E5, Canada; mmabee@ualberta.ca

\* Correspondence: hareendr@ualberta.ca

**Featured Application:** Our AI-based approach flags low-quality ultrasound hip images as inadequate for diagnosis. This would help sonographers to collect high-quality hip scans suitable for early diagnosis of DDH.

**Abstract:** Ultrasound images can reliably detect Developmental Dysplasia of the Hip (DDH) during early infancy. Accuracy of diagnosis depends on the scan quality, which is subjectively assessed by the sonographer during ultrasound examination. Such assessment is prone to errors and often results in poor-quality scans not being reported, risking misdiagnosis. In this paper, we propose an Artificial Intelligence (AI) technique for automatically determining scan quality. We trained a Convolutional Neural Network (CNN) to categorize 3D Ultrasound (3DUS) hip scans as ‘adequate’ or ‘inadequate’ for diagnosis. We evaluated the performance of this AI technique on two datasets—Dataset 1 (DS1) consisting of 2187 3DUS images in which each image was assessed by one reader for scan quality on a scale of 1 (lowest quality) to 5 (optimal quality) and Dataset 2 (DS2) consisting of 107 3DUS images evaluated semi-quantitatively by four readers using a 10-point scoring system. As a binary classifier (adequate/inadequate), the AI technique gave highly accurate predictions on both datasets (DS1 accuracy = 96% and DS2 accuracy = 91%) and showed high agreement with expert readings in terms of Intraclass Correlation Coefficient (ICC) and Cohen’s kappa coefficient (K). Using our AI-based approach as a screening tool during ultrasound scanning or postprocessing would ensure high scan quality and lead to more reliable ultrasound hip examination in infants.

**Keywords:** hip; developmental dysplasia of the hip; 3D ultrasound; scan quality assessment; deep learning; convolutional neural networks



**Citation:** Hareendranathan, A.R.; Mabee, M.; Chahal, B.S.; Dulai, S.K.; Jaremko, J.L. Can AI Automatically Assess Scan Quality of Hip Ultrasound? *Appl. Sci.* **2022**, *12*, 4072. <https://doi.org/10.3390/app12084072>

Academic Editors: Yujin Lim and Hideyuki Takahashi

Received: 15 March 2022

Accepted: 12 April 2022

Published: 18 April 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Developmental Dysplasia of the Hip (DDH) is common in infants, present in one to three per 1000 live births [1] with large variations among ethnic groups [2]. Undiagnosed DDH is a major risk factor for early hip Osteoarthritis (OA) [2,3] which is associated with a huge economic burden [4,5]. It is also seen as the underlying cause for more than 1/3 of hip replacements in adults who are less than 60 years of age [6]. If diagnosed in early infancy (<3 months), DDH can be treated by simple techniques such as bracing (e.g., Pavlik harness) [7]. Despite obvious advantages of early diagnosis [8,9], hip screening programs for infants are not common in most countries, due in large part to the high variability in scan assessment [10].

Clinically, DDH is diagnosed by physical examination including Barlow and Ortolani maneuvers which have poor sensitivity beyond the neonatal period, resulting in missed cases, particularly of mild DDH or fixed dislocations [11–13]. Compared to physical examination, ultrasound imaging is more reliable and more sensitive [11–13] and is ideal for screening programs as it is safe, easily portable and inexpensive.

During hip examination, 2D ultrasound images are usually acquired and interpreted based on the Graf technique [14] which measures the alpha angle between the ilium and acetabular roof. Alpha angles greater than 60 degrees are considered normal and those less than 43 degrees are considered severely dysplastic [14]. This technique relies heavily on obtaining a single 2D image in the Graf plane containing all landmarks such as ilium, labrum, os ischium and femoral head (Figure 1A). Acquiring such images requires many hours of training, since slight variations in probe orientation can result in suboptimal images. Previous studies have shown that deviation from the Graf plane by novice users can result in incorrect diagnosis in two-thirds of neonates and half of infants evaluated [15]. These variations can be reduced to some extent by using 3D Ultrasound (3DUS). Since 3DUS covers a larger area, it is more likely to contain the Graf plane, thereby making it more reliable than 2DUS, especially when scanning is performed by novice sonographers [16].



**Figure 1.** Examples of ultrasound hip images of different scan quality. (A) Example of a high-quality image in which all landmarks such as ilium, os ischium, labrum and femoral head are clearly visible. The alpha angle is measured as the inner angle between the iliac line and acetabular roof. (B) Moderate-quality scan containing all landmarks, but with the blurring of the os ischium and blurring and slight tilt of the iliac line. (C) Poor-quality image with none of the landmarks visible.

However, in both 2DUS and 3DUS, reliability of the ultrasound examination depends to a large extent on scan quality. In current clinical practice, sonographers manually assess scan quality based on visibility of key landmarks. This approach is prone to high inter-observer variance. As a result, poor-quality images can be presented to radiologists, risking misdiagnosis. Examples of images of varying scan quality are shown in Figure 1. In earlier work [17], a semi-quantitative scoring technique that evaluates individual imaging landmarks (such as ilium, labrum, femoral head and os ischium) and artifacts (such as movement and other artifacts) is proposed. While such approaches make scan quality assessment less subjective, there can still be variability in the manual assessment of individual features.

Automatic assessment of scan quality could address these issues and provide the sonographer with a reliable and non-subjective assessment of image adequacy at the time of scanning. Automatic assessment of ultrasound is challenging due to various factors such as the presence of spurious image effects that resemble anatomical structures, blurred image boundaries, mirror artifacts and shadowing artifacts. The 3DUS images (as well as 2D sweeps) also contain artifacts resulting from patient movement or hand movement at the time of scanning. As a result, commonly used image processing techniques such as template-matching [18,19] and feature extraction [20,21] cannot be directly applied to ultrasound.

#### *Related Work on Deep Learning in Hip Ultrasound*

With the recent success of Convolutional Neural Networks (CNNs) in medical image processing, data-driven approaches have been proposed for automatic image segmentation and classifications in hip ultrasound. Hareendranathan et al. [22] developed a CNN for segmenting the acetabulum that used multi-scale super pixels as inputs to an AlexNet model.

The 3D convolutional networks (3DconvNet) such as C3D and I3D have been used for video classification. The C3D network uses a homogeneous architecture with  $3 \times 3 \times 3$  convolution kernels in all layers and gave 52.8% accuracy on the UCF101 dataset [23]. The Inflated 3D ConvNet (I3D) used a new two-stream approach to combine 2D and 3D convolutions. With pretraining on Kinetics, the I3D model gave accuracies of 80.9% on the HMDB-51 dataset and 98.0% on the UCF-101 dataset [24]. Similarly, 2D segmentation models such as the 3D U-Net [25], 3D V-Net [26] and the 3D Deeply Supervised network [27] have also been used for image segmentation in videos. Extending these models to 3DUS volumes of the hip is non-trivial, mainly due to the lack of large datasets.

Generally, deep learning models are trained on hip ultrasound images acquired by well-trained sonographers in research settings. Such images contain all necessary imaging landmarks such as straight and horizontal acetabulum and round femoral head. The scan quality of the images used in these datasets was assessed subjectively by the sonographer. In this paper, we develop a deep learning technique to automatically assess the image quality of hip ultrasound images and categorize it as adequate vs. inadequate for DDH diagnosis. Similar deep learning techniques have been developed using CNNs [28,29] and Recurrent Neural Networks (RNNs) [30] for 2DUS scan quality assessment. Along these lines, we propose a new 3D CNN model to predict the adequacy of hip scans for DDH diagnosis in 3D ultrasound images and validate it on two large clinical datasets.

This is the first study to validate an automatic hip quality assessment technique on a large dataset, of 2187 hip scans performed by 3DUS (Dataset 1). Any systematic error in prediction is also examined by evaluating the algorithm on subgroups of images based on age {0–3 months, >3 months} and sex {Male, Female}. On a separate dataset of 107 images (Dataset 2), we conducted a multi-reader study with four readers who semi-quantitatively evaluated scan quality using a 10-point scoring system [17] and compared it to predictions from our AI-based approach.

## 2. Materials and Methods

We recruited infants who had been referred for ultrasound examination based on clinical suspicion of DDH (due to risk factors such as hip laxity, asymmetrical skin creases, breech position, female sex, positive family history of DDH, first-born infants and ethnicity). In this institutional health research ethics board-approved study, we added 3DUS scanning to the routine 2DUS scanning protocol and obtained written informed parental consent from each participant. Since DDH can be unilateral or bilateral, we included all dysplastic hips separately but, in cases where both hips were normal, we only included one hip per subject in the study. Each hip was categorized as normal or abnormal based on the assessment of the consulting radiologist.

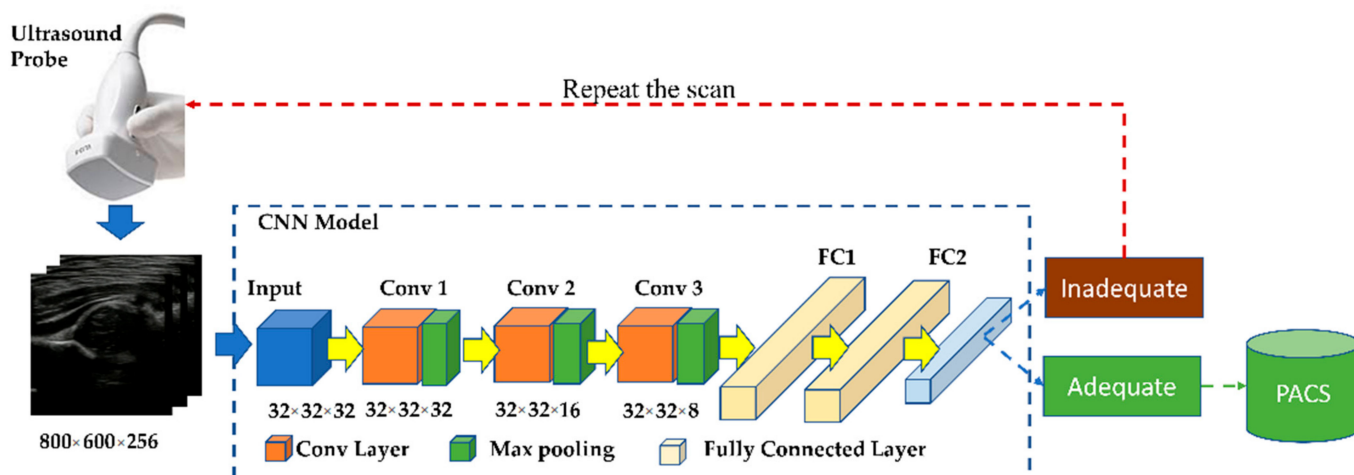
### 2.1. Ultrasound Scanning

Routine ultrasound examination for both hips was performed per American College of Radiology recommendations [31]. In addition to 2D scans, we also acquired coronal 3DUS images of both hips using a Philips iU22 scanner with a Philips 13VL5 (Philips Healthcare, Andover, MA, USA) linear array transducer (having a center frequency of 13 MHz). Each 3DUS image consisted of 3.2 s sweeps (sweep angle  $+/-$  15 degrees, 256 slices), with the head of the transducer positioned near the greater trochanter of the infant. Sweeps were acquired such that the central slice of the 3DUS volume roughly approximated the Graf plane. Each slice was  $411 \times 193$  pixels (with each pixel measuring  $0.11 \text{ mm} \times 0.20 \text{ mm}$ ) and a slice thickness of 0.13 mm.

### 2.2. Model Development

We used a 3D CNN consisting of convolutional and fully connected layers to predict scan quality. The optimal architecture for the network (as shown in Figure 2) was dynamically determined based on accuracy of classification on the validation set. Like the C3D model, we used a homogenous kernel size of  $3 \times 3 \times 3$  in all layers. To minimize

the memory requirement in training, we resized the original  $800 \times 600 \times 256$  image to a  $32 \times 32 \times 32$  tensor. Since the input tensor is already much smaller compared to the original image, we applied max pooling only in the slice dimension, i.e., a pooling ratio of  $1 \times 1 \times 2$ . Hence, the height and width of each slice was maintained at  $32 \times 32$  pixels. We used a stride size of 1 in all convolutional layers and used 16, 32 and 64 kernels in the convolutional layers conv1, conv2 and conv3, respectively. The model was trained for 100 epochs to minimize a categorical cross entropy loss optimized using rmsprop at learning rate  $lr = 0.001$ . To prevent the model from overfitting, we used 20% dropout in the fully connected layers.



**Figure 2.** Overview of the proposed technique showing the CNN model consisting of 3 convolutional layers and 2 Fully Connected (FC) layers. Each convolutional layer consists of a convolution layer, ReLU activation and softmax (applied only in the slice dimension).

Labels for training the network were obtained from the manual scoring on 1548 other 3DUS scans excluded from the test set by one reader (MM with 4 years of experience in hip ultrasound image analysis). Factors such as the straightness of the ilium, visibility of labrum, os ischium and femoral head and presence of movement artifacts were considered while scoring the images. These labeled images were divided into the training (70%, 1083 images, 651 high-quality images, 432 low-quality images) and validation set (30%, 465 images, 300 high-quality images, 165 low-quality images).

In our dataset, we had a relatively low number of poor-quality images. This is typical of ultrasound scanning taking place in research settings where scans are acquired by well-trained sonographers. We tried to address this issue by assigning a higher weight (0.65 vs. 0.45) for misclassified low-quality scans.

### 2.3. Model Evaluation

We tested the AI technique on two datasets—(1) Dataset 1 (DS1): A large dataset of 2187 3D hip ultrasound images obtained from 508 infants (each 3D image from a patient scan was analyzed separately for image quality) in which one reader (MM), a radiology resident with 4 years of expertise in hip dysplasia ultrasound, scored images from 1–5, and (2) Dataset 2 (DS2): A separate set of 107 3D images from 101 subjects in which 4 readers evaluated the quality more systematically using a previously published 10-point scoring system [17] as summarized in Table 1.

**Table 1.** Method of scoring used to evaluate Datasets 1 and Dataset 2. #+, #- indicate the number of adequate-quality and inadequate-quality images in each dataset.

Dataset	# Images, (#+, #-)	# Readers	Method	Scan Adequacy Criteria
Dataset 1	2187 (1568, 619)	1	Holistic score: Low quality: 1 High quality: 5	>3
Dataset 2	107 (75, 32)	4	10-point scoring <sup>1</sup> : High quality: 10 Straight and horizontal ilium: 2 Clearly visible Os ischium: 2 Visible labrum: 1 Round femoral head: 1 No motion artifacts: 2 No other image artifacts: 2	>6

<sup>1</sup> Details on the 10-point scoring system can be found in [17].

#### 2.4. Statistics

The readers included one expert (reader 4) who is our lead radiologist, with fellowship training in pediatric and musculoskeletal radiology and 13 years of experience (JJ) and three non-expert readers (reader 1, BC, a radiology resident with 2 years' experience in hip ultrasound, reader 2, EO, a graduate student in radiology with 1 year experience in hip ultrasound, reader 3, AH, a research associate with 5 years of experience in ultrasound hip image analysis). For binary classification, scans with a score of above 3/5 in DS1 and above 6/10 in DS2 were considered 'adequate' quality.

Accuracy, sensitivity, specificity, Negative Predictive Value (NPV), Positive Predictive Value (PPV) and Area Under the ROC Curve (AUC) were calculated for each image with the manual categorization treated as ground truth. Descriptive statistics of agreement of AI prediction with manual readings are also reported in terms of ICC (3,k) and kappa. We evaluated the Accuracy, Sensitivity (SN), Specificity (SP), and Positive and Negative Predictive Value (PPV, NPV) of the AI network as a diagnostic test for image quality in which a true-positive result was an image manually defined as inadequate (i.e., a test that flags poor-quality scans). Calculations were performed using our inhouse software developed in Python 3.6 using sklearn and pingouin libraries.

### 3. Results

As seen in Table 2, within DS1 (2187 3DUS images from 508 infants) the AI technique correctly predicted 596/619 images as inadequate quality and 1501/1568 as adequate quality. In DS2 (107 3DUS images), the AI technique gave the highest accuracy, sensitivity, specificity, NPV and PPV when compared to three non-expert readers. We also tried to address the data imbalance (651 high-quality images vs. 432 low-quality images) by assigning a higher weight to the underrepresented class (i.e., low-quality images). However, this gave a lower AUC of 0.89.

In order to check for any systematic error in predictions, we divided the images into subgroups based on age and sex and evaluated all parameters within the subgroups as summarized in Table 3. Accuracies were similar in all subgroups, indicating no systematic error in prediction.

Agreement of AI-based predictions with manual scoring was quantified using ICC and Cohen's kappa as shown in Table 4. We used kappa to measure the inter-observer agreement between AI and the expert reader in Dataset (DS) 1. The DS was evaluated by three non-expert readers and one expert (JJ). Compared to the non-expert readers, AI showed higher agreement (0.77 vs. 0.72) with the expert. The difference in kappa scores



between AI and non-expert human readers was statistically significant ( $p < 0.05$ ) for readers 2 and 3.

**Table 2.** Accuracy, sensitivity, specificity, Negative Predictive Value (NPV), Positive Predictive Value (PPV) and AUC of our AI-based approach. AUC represents the area under the Receiver Operating Characteristic (ROC) curve.

Dataset	Reader	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
Dataset 1	AI	0.96	1.0	0.87	0.94	1.0	0.93
Dataset 2	AI	0.91	0.9	0.93	0.97	0.76	0.91
	Reader 1	0.89	0.9	0.85	0.95	0.74	0.88
	Reader 2	0.7	0.59	1.0	1.0	0.46	0.8
	Reader 3	0.75	0.72	0.81	0.92	0.5	0.77

**Table 3.** Comparison of accuracy, sensitivity, specificity, PPV, NPV and AUC in subgroups of data categorized based on age and sex. Note that there was no systematic bias in any subgroup.

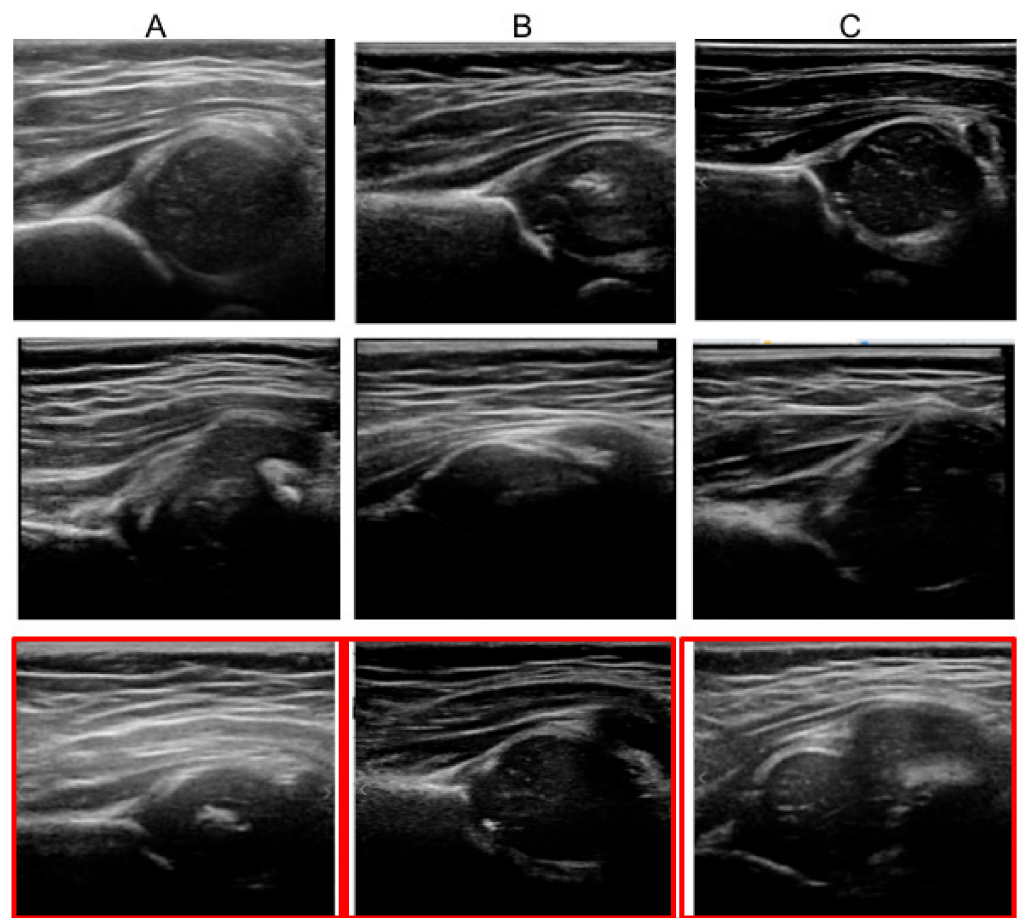
Variable	Subgroup	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC
Age	0–3 months	0.97	1.0	0.84	0.96	1.0	0.92
	>3 months	0.93	1.0	0.89	0.83	1.0	0.94
Sex	Male	0.96	1.0	0.87	0.94	1.0	0.94
	Female	0.96	1.0	0.86	0.94	1.0	0.93

**Table 4.** Agreement of AI predictions with scores provided by expert and non-expert readers. The 95% Confidence Intervals (CIs) of each value is provided in square brackets.

Dataset	Readers	Ground Truth	Kappa	ICC
DS1	AI	Expert	0.90 [0.97, 0.82]	0.95 [0.94, 0.95]
DS2	AI	Expert	0.77 [0.85, 0.68]	0.88 [0.84, 0.91]
	Reader 1	Expert	0.72 [0.79, 0.65]	
	Reader 2	Expert	0.43 [0.50, 0.35]	
	Reader 3	Expert	0.44 [0.50, 0.38]	

Overall agreement of AI with manual reading was quantified using ICC considering AI-based prediction as the 2nd reader in DS1 and as the 5th reader in DS2. In both cases, agreement among readers was high (DS1 = 0.95; DS2 = 0.88).

Some examples of images of varying quality that were evaluated by AI are shown in Figure 3. All images in row 1 are high-quality images with all necessary landmarks that the CNN correctly identified as adequate. Similarly, images in row 2 are images that are correctly categorized as inadequate by AI. Images in row 3 are moderate-quality images in which there exists variability in manual assessment. The expert radiologist scored two of these images (A, B) as adequate (with score 7) and one image (C) as inadequate (with score 6). AI classified images A and B as inadequate and C as adequate. This row of images highlights the inevitable subjectivity of assessment of intermediate-quality scans.



**Figure 3.** Examples of images analyzed by AI. Row 1: Images of high quality that were correctly identified AI. Row 2: Low-quality images correctly categorized as inadequate by AI. Row 3: Cases where AI failed are indicated with a red box. Images (A,B) were categorized as inadequate by AI and two human readers but were categorized as adequate by the expert reader (these images represent false positives in classification). Similarly, image (C) was classified as adequate by AI and 2 human readers, but inadequate by the expert reader (false negative).

#### 4. Discussion

We developed a 3D CNN for automatic ultrasound scan quality assessment and validated it on two datasets, one large (>2000 scans) and one as part of a multi-reader exercise (four human readers, 107 scans). This is the first study of automatic ultrasound scan quality assessment on a large dataset with wide representation from infants of different age groups and sex. On the large data set, our CNN was 96% accurate (with 100% sensitivity and 87% specificity) when compared to expert manual assessment of quality. In the multi-reader study, the AI technique actually performed slightly better than each of the three non-expert human readers in agreeing with the 4th expert human reader on diagnostic quality as rated in a 10-point scoring system [17]. Based on the commonly used interpretation of ICC (where values less than 0.5 indicate poor, 0.5–0.75 moderate, 0.75–0.9 good and greater than 0.90 excellent [32]), the CNN showed excellent reliability.

The 3D CNN approach flags low-quality images containing imaging artifacts that are commonly seen in 3DUS (as well as 2D sweeps). These artifacts usually occur due to patient movement, hand movement and ultrasound shadowing. Although we do not specifically segment these artifacts, since our technique considers temporal and spatial information around each pixel, images with high occurrence of image artifacts are categorized as inadequate.

Since 3DUS scans consist of a large number (~250) of slices, manual assessment of scan quality is tedious. As a result, anatomical landmarks crucial to DDH assessment

could be missed in 3DUS. Our automatic assessment technique is fast (average execution time  $\sim 2$  s/image on the Compute Canada Cedar cluster with an NVIDIA V100 GPU) and accurate and it can be used to provide feedback on scan quality to the sonographer in real time during hip examination.

This can also be used as a preprocessing step for automatic interpretation of hip ultrasound. Automatic interpretation usually involves calculation of alpha angle and/or predicting the probability of DDH from ultrasound images. Since these techniques [33–35] generally rely on imaging landmarks such as ilium, acetabulum, os ischium and femoral head, their accuracy depends to a large extent on adequate visualization of these features. Our technique could flag inadequate images upfront and thereby improve the accuracy of these automatic hip interpretation systems. Since the focus of this paper is to apply a CNN model on the new use case of hip image quality assessment, we have not compared it to other 3D networks such as C3D, I3D or 3D U-Net.

One limitation of our study is the variability in ground truth as it is based on manual assessment. Images in Dataset 1 were holistically scored from 1–5 by a manual reader. There could be variability in these scores if the same images were to be scored by a second reader. We address this issue to some extent in Dataset 2 which was semi-quantitatively assessed using the scoring system described in [17]. Using the scoring system, the human reader makes a series of (mostly binary) decisions based on landmarks found in the image. Although this reduces the variability, there are inevitably some variations in manual assessment of these individual features, most relevant in intermediate-quality images. For example, as shown Figure 2 row 3, two images were assessed as adequate quality by two human readers but with a low score of 7 points. These images were assessed as inadequate by AI with a probability of 0.42 and 0.4. Such images represent edge cases where there exists ambiguity in ground truth assessment.

Images used in our study were acquired in a research environment by experienced sonographers in dedicated sessions which represent the idealized scenario. Hence, most images were of high quality. For example, in Dataset 2 more than 90% of the images have clearly visible labrum and femoral head. As future work, we plan a multi-center study on 2D sweep images acquired from tertiary centers and small clinics using handheld pocket-sized devices. In this study, we expect to acquire more low-quality scans that can be used to further validate the utility of our AI-based approach. Demonstrating validity of the AI technique for assessing scan quality in these settings is critical to the feasibility of using 2DUS for population screening of DDH.

## 5. Conclusions

We developed a new AI technique for automatic interpretation of ultrasound scan quality and validated it on two datasets of 3D ultrasound images. Our approach was fast, highly reliable and showed agreement with an expert reader that was equivalent to or better than that of three other human readers. This new AI approach can be used to provide feedback on scan quality to sonographers during or immediately after ultrasound examination. It can also be used as a preprocessing step in interpretation systems to filter low-quality images and improve the reliability of hip ultrasound examinations.

**Author Contributions:** Conceptualization, J.L.J. and A.R.H.; methodology, A.R.H., B.S.C., M.M. and J.L.J.; software, A.R.H.; validation, A.R.H., S.K.D. and J.L.J.; formal analysis, A.R.H.; investigation, A.R.H. and J.L.J.; resources, J.L.J.; data curation, M.M. and B.S.C.; writing—original draft preparation, A.R.H.; writing—review and editing, A.R.H., S.K.D. and J.L.J.; visualization, A.R.H.; supervision, J.L.J.; project administration, J.L.J.; funding acquisition, J.L.J. All authors have read and agreed to the published version of the manuscript.



**Funding:** This research was funded by the Women and Children’s Health Research Institute (WCHRI) and the Resource Allocation Panel (RAP) funding by Alberta Machine Intelligence Institute (AMII).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of University of Alberta.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors thank the Women and Children’s Health Research Institute (WCHRI) for the research funding that supported this work. We acknowledge the support of Compute Canada for providing GPU instances that were used to train and test the AI models developed in this project.

**Conflicts of Interest:** Jacob Jaremko is a co-founder of MEDO.ai Inc.—a company that develops AI-based solutions for ultrasound. Other authors declare no conflict of interest. The funders (WCHRI) had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Furnes, O.; Lie, S.A.; Espehaug, B.; Vollset, S.E.; Engesaeter, L.B.; Havelin, L.I. Hip Disease and the Prognosis of Total Hip Replacements. *J. Bone Jt. Surg. Br. Vol.* **2001**, *83-B*, 579. [[CrossRef](#)]
2. Loder, R.T.; Skopelja, E.N. The Epidemiology and Demographics of Hip Dysplasia. *ISRN Orthop.* **2011**, *2011*, 238607. [[CrossRef](#)] [[PubMed](#)]
3. Jacobsen, S.; Sonne-Holm, S. Hip Dysplasia: A Significant Risk Factor for the Development of Hip Osteoarthritis. A Cross-Sectional Survey. *Rheumatology* **2005**, *44*, 211–218. [[CrossRef](#)] [[PubMed](#)]
4. Bitton, R. The Economic Burden of Osteoarthritis. *Am. J. Manag. Care* **2009**, *15*, S230–S235.
5. Loza, E.; Lopez-Gomez, J.M.; Abasolo, L.; Maese, J.; Carmona, L.; Batlle-Gualda, E. Artrocad Study Group Economic Burden of Knee and Hip Osteoarthritis in Spain. *Arthritis Rheum.* **2009**, *61*, 158–165. [[CrossRef](#)]
6. Price, C.T.; Ramo, B.A. Prevention of Hip Dysplasia in Children and Adults. *Orthop. Clin. N. Am.* **2012**, *43*, 269–279. [[CrossRef](#)]
7. Atalar, H.; Sayli, U.; Yavuz, O.Y.; Uras, I.; Dogruel, H. Indicators of Successful Use of the Pavlik Harness in Infants with Developmental Dysplasia of the Hip. *Int. Orthop.* **2007**, *31*, 145–150. [[CrossRef](#)]
8. Buonsenso, D.; Curatola, A.; Lazzareschi, I.; Panza, G.; Morello, R.; Marrocco, R.; Valentini, P.; Cota, F.; Rendeli, C. Developmental Dysplasia of the Hip: Real World Data from a Retrospective Analysis to Evaluate the Effectiveness of Universal Screening. *J. Ultrasound* **2020**, *24*, 403–410. [[CrossRef](#)]
9. Buonsenso, D.; Menzella, N.; Morello, R.; Valentini, P. Indirect Effects of COVID-19 on Child Health Care: Delayed Diagnosis of Developmental Dysplasia of the Hip. *J. Ultrasound* **2020**, *23*, 443–444. [[CrossRef](#)]
10. Shorter, D.; Hong, T.; Osborn, D.A. Cochrane Review: Screening Programmes for Developmental Dysplasia of the Hip in Newborn Infants. *Evid. Based Child Health* **2013**, *8*, 11–54. [[CrossRef](#)]
11. Dezateux, C.; Rosendahl, K. Developmental Dysplasia of the Hip. *Lancet* **2007**, *369*, 1541–1552. [[CrossRef](#)]
12. Bache, C.E.; Clegg, J.; Herron, M. Risk Factors for Developmental Dysplasia of the Hip: Ultrasonographic Findings in the Neonatal Period. *J. Pediatric Orthop. B* **2002**, *11*, 212–218.
13. Clarke, N.M.; Clegg, J.; Al-Chalabi, A.N. Ultrasound Screening of Hips at Risk for CDH. Failure to Reduce the Incidence of Late Cases. *J. Bone Jt. Surg. Br.* **1989**, *71*, 9–12. [[CrossRef](#)] [[PubMed](#)]
14. Graf, R. Fundamentals of Sonographic Diagnosis of Infant Hip Dysplasia. *J. Pediatr. Orthop.* **1984**, *4*, 735–740. [[CrossRef](#)]
15. Jaremko, J.L.; Mabee, M.; Swami, V.G.; Jamieson, L.; Chow, K.; Thompson, R.B. 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* **2014**, *273*, 870–878. [[CrossRef](#)]
16. Mostofi, E.; Chahal, B.; Zonoobi, D.; Hareendranathan, A.; Roshandeh, K.P.; Dulai, S.K.; Jaremko, J.L. Reliability of 2D and 3D Ultrasound for Infant Hip Dysplasia in the Hands of Novice Users. *Eur. Radiol.* **2019**, *29*, 1489–1495. [[CrossRef](#)]
17. Hareendranathan, A.R.; Chahal, B.; Ghasseminia, S.; Zonoobi, D.; Jaremko, J.L. Impact of Scan Quality on AI Assessment of Hip Dysplasia Ultrasound. *J. Ultrasound* **2021**, *5*, 1–9. [[CrossRef](#)]
18. Kwitt, R.; Vasconcelos, N.; Razzaque, S.; Aylward, S. Localizing Target Structures in Ultrasound Video—A Phantom Study. *Med. Image Anal.* **2013**, *17*, 712–722. [[CrossRef](#)]
19. Ni, D.; Yang, X.; Chen, X.; Chin, C.-T.; Chen, S.; Heng, P.A.; Li, S.; Qin, J.; Wang, T. Standard Plane Localization in Ultrasound by Radial Component Model and Selective Search. *Ultrasound Med. Biol.* **2014**, *40*, 2728–2742. [[CrossRef](#)]
20. Rahmatullah, B.; Papageorghiou, A.T.; Noble, J.A. Integration of Local and Global Features for Anatomical Object Detection in Ultrasound. *Med. Image Comput. Comput. Assist. Interv.* **2012**, *15*, 402–409.
21. Maraci, M.A.; Napolitano, R.; Papageorghiou, A.; Noble, J.A. Searching for Structures of Interest in an Ultrasound Video Sequence. In *Machine Learning in Medical Imaging*; Springer International Publishing: Cham, Switzerland, 2014; pp. 133–140.

22. Hareendranathan, A.R.; Zonoobi, D.; Mabee, M.; Cobzas, D.; Punithakumar, K.; Noga, M.; Jaremko, J.L. Toward Automatic Diagnosis of Hip Dysplasia from 2D Ultrasound. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 18–21 April 2017; pp. 982–985.
23. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features with 3d Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
24. Carreira, J.; Zisserman, A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, CA, USA, 21–26 July 2017; pp. 6299–6308.
25. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016, Athens, Greece, 17–21 October 2016; Springer International Publishing: Athens, Greece, 2016; pp. 424–432.
26. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
27. Dou, Q.; Yu, L.; Chen, H.; Jin, Y.; Yang, X.; Qin, J.; Heng, P.-A. 3D Deeply Supervised Network for Automated Segmentation of Volumetric Medical Images. *Med. Image Anal.* **2017**, *41*, 40–54. [[CrossRef](#)] [[PubMed](#)]
28. Paserin, O.; Mulpuri, K.; Cooper, A.; Hodgson, A.J.; Abugharbieh, R. Automatic Near Real-Time Evaluation of 3D Ultrasound Scan Adequacy for Developmental Dysplasia of the Hip. In Proceedings of the Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures, Québec City, QC, Canada, 14 September 2017; Springer International Publishing: Québec City, QC, Canada, 2017; pp. 124–132.
29. Chen, H.; Wu, L.; Dou, Q.; Qin, J.; Li, S.; Cheng, J.-Z.; Ni, D.; Heng, P.-A. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework. *IEEE Trans. Cybern.* **2017**, *47*, 1576–1586. [[CrossRef](#)] [[PubMed](#)]
30. 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 Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Granada, Spain, 16–20 September 2018; Springer International Publishing: Granada, Spain, 2018; pp. 365–373.
31. Harcke, H.T.; Paltiel, H.; Rosenberg, H.K.; Barr, L.L.; Ruzal-Shapiro, C.; Wolfson, B.J.; Paushter, D.M.; Angtuaco, T.L.; Ackerman, S.; Crino, J.; et al. AIUM Practice Guideline for the Performance of an Ultrasound Examination for Detection and Assessment of Developmental Dysplasia of the Hip. *J. Ultrasound Med.* **2009**, *28*, 114–119.
32. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)] [[PubMed](#)]
33. Hareendranathan, A.R.; Mabee, M.; Punithakumar, K.; Noga, M.; Jaremko, J.L. A Technique for Semiautomatic Segmentation of Echogenic Structures in 3D Ultrasound, Applied to Infant Hip Dysplasia. *Int. J. Comput. Assist. Radiol. Surg.* **2016**, *11*, 31–42. [[CrossRef](#)]
34. Hareendranathan, A.R.; Mabee, M.; Punithakumar, K.; Noga, M.; Jaremko, J.L. Toward Automated Classification of Acetabular Shape in Ultrasound for Diagnosis of DDH: Contour Alpha Angle and the Rounding Index. *Comput. Methods Programs Biomed.* **2016**, *129*, 89–98. [[CrossRef](#)]
35. Quader, N.; Hodgson, A.; Abugharbieh, R. Confidence Weighted Local Phase Features for Robust Bone Surface Segmentation in Ultrasound. In *Workshop on Clinical Image-Based Procedures*; Springer: Boston, MA, USA, 2014; pp. 76–83.