BJO

■ **HIP**

Artificial intelligence-generated hip radiological measurements are fast and adequate for reliable assessment of hip dysplasia

AN EXTERNAL VALIDATION STUDY

H. Archer,
S. Reine,
A. Alshaikhsalama,
J. Wells,
A. Kohli,
L. Vazquez,
A. Hummer,
M. D. DiFranco,
R. Ljuhar,
Y. Xi,
A. Chhabra

From The University of Texas Southwestern Medical Center, Dallas, Texas, USA.

Aims

Hip dysplasia (HD) leads to premature osteoarthritis. Timely detection and correction of HD has been shown to improve pain, functional status, and hip longevity. Several time-consuming radiological measurements are currently used to confirm HD. An artificial intelligence (AI) software named HIPPO automatically locates anatomical landmarks on anteroposterior pelvis radiographs and performs the needed measurements. The primary aim of this study was to assess the reliability of this tool as compared to multi-reader evaluation in clinically proven cases of adult HD. The secondary aims were to assess the time savings achieved and evaluate inter-reader assessment.

Methods

A consecutive preoperative sample of 130 HD patients (256 hips) was used. This cohort included 82.3% females (n = 107) and 17.7% males (n = 23) with median patient age of 28.6 years (interquartile range (IQR) 22.5 to 37.2). Three trained readers' measurements were compared to AI outputs of lateral centre-edge angle (LCEA), caput-collum-diaphyseal (CCD) angle, pelvic obliquity, Tönnis angle, Sharp's angle, and femoral head coverage. Intraclass correlation coefficients (ICC) and Bland-Altman analyses were obtained.

Results

Among 256 hips with AI outputs, all six hip AI measurements were successfully obtained. The AI-reader correlations were generally good (ICC 0.60 to 0.74) to excellent (ICC > 0.75). There was lower agreement for CCD angle measurement. Most widely used measurements for HD diagnosis (LCEA and Tönnis angle) demonstrated good to excellent inter-method reliability (ICC 0.71 to 0.86 and 0.82 to 0.90, respectively). The median reading time for the three readers and AI was 212 (IQR 197 to 230), 131 (IQR 126 to 147), 734 (IQR 690 to 786), and 41 (IQR 38 to 44) seconds, respectively.

Conclusion

This study showed that AI-based software demonstrated reliable radiological assessment of patients with HD with significant interpretation-related time savings.

Introduction

Hip dysplasia (HD) is a developmental condition where the acetabulum does not sufficiently cover the femoral head. This insufficient coverage places excessive stresses on the acetabular rim and can lead to hip pain,
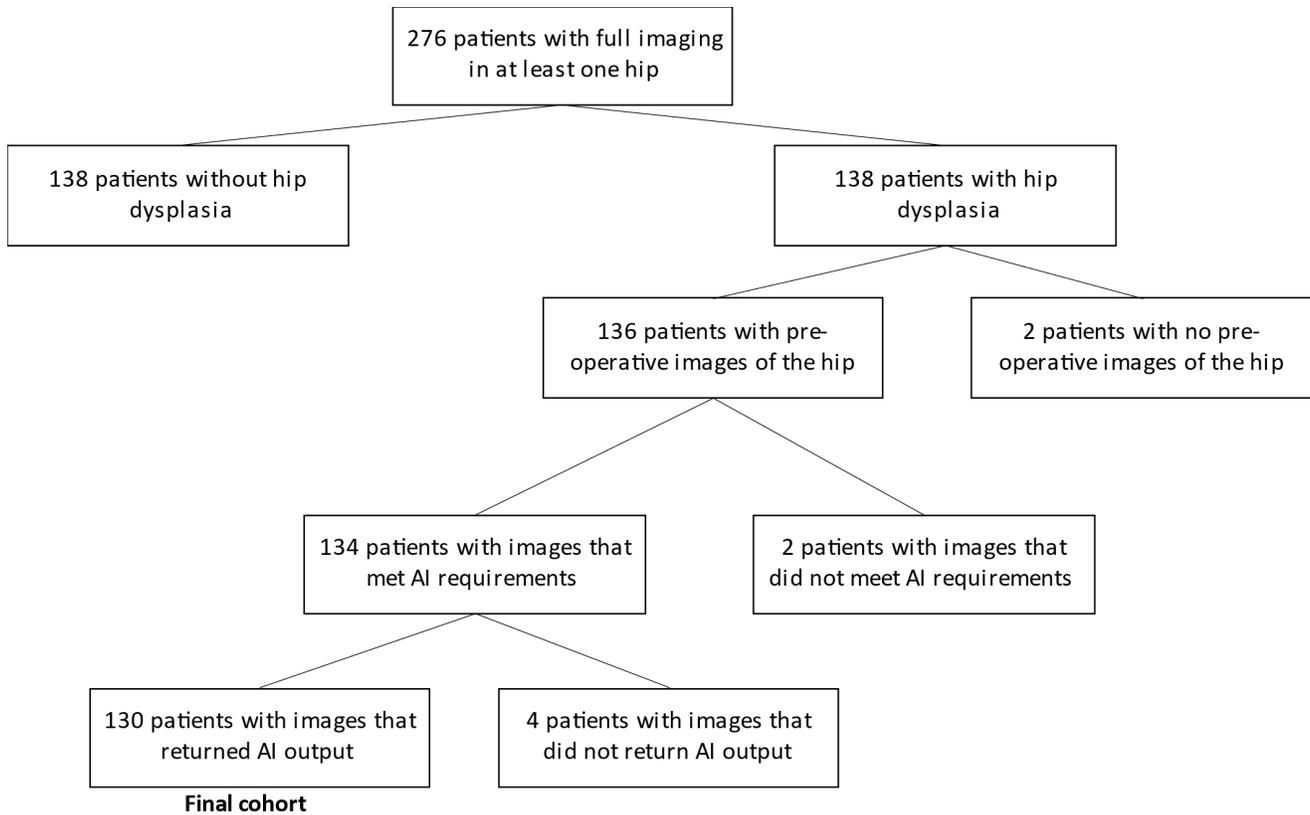
```
              ┌─────────────────────────────────┐
              │  276 patients with full imaging  │
              │       in at least one hip        │
              └─────────────────────────────────┘
             /                                     \
┌───────────────────────────┐        ┌───────────────────────────┐
│  138 patients without hip  │        │   138 patients with hip   │
│         dysplasia          │        │         dysplasia          │
└───────────────────────────┘        └───────────────────────────┘
                                       /                         \
                   ┌───────────────────────────┐   ┌───────────────────────────┐
                   │  136 patients with pre-    │   │  2 patients with no pre-   │
                   │ operative images of the hip│   │operative images of the hip │
                   └───────────────────────────┘   └───────────────────────────┘
                    /                    \
  ┌───────────────────────────┐   ┌───────────────────────────┐
  │  134 patients with images  │   │  2 patients with images    │
  │   that met AI requirements │   │ that did not meet AI       │
  │                            │   │ requirements               │
  └───────────────────────────┘   └───────────────────────────┘
    /                  \
┌───────────────────────┐  ┌───────────────────────┐
│ 130 patients with      │  │ 4 patients with images │
│ images that returned   │  │ that did not return    │
│ AI output              │  │ AI output              │
└───────────────────────┘  └───────────────────────┘
      Final cohort
```

Fig. 1

Flowchart for final study sample containing preoperative hip dysplasia patients with complete radiological imaging who met artificial intelligence (AI) requirements.

apprehension, instability, progressive chondrolabral injury, and premature osteoarthritis.[1,2] HD prevalence ranges from 5.4% to 12.8%, depending on the radiological index applied for the diagnosis.[3] Timely detection and correction of HD has been shown to improve hip pain, joint functional status, and hip longevity.[1,2,4,5]

Several radiological measurements have been used to diagnose HD, especially the lateral centre-edge angle (LCEA) of Wiberg,[6] femoral head coverage, and Tönnis angle.[7] It is controversial as to how some of these angles are defined. For example, the LCEA, which measures acetabular coverage of the femoral head in the coronal plane, is sometimes measured to the most lateral acetabular rim edge instead of the sclerotic lateral sourcil edge, resulting in statistically and clinically significant differences.[6] The differences in measurements can lead to different hip diagnoses, such as HD and femoroacetabular impingement (FAI), leading to different or inadequate treatments. In addition to potential problems with the accuracy of manual readings, performing multiple diagnostic measurements for each patient is time-consuming and requires full attention and diligence for consistency.

Therefore, there is an unmet need for standardized and reproducible radiological measurements of the hip. The primary aim of this study was to assess the agreement between a Conformité Européenne (CE)-certified artificial intelligence (AI)-based algorithm (software) and manual measurements by multiple readers in adult patients with HD. The secondary aims were to assess the time savings achieved and evaluate inter-reader assessment.

Methods

This study received institutional review board approval for retrospective cross-sectional evaluation of a prospectively gathered sample from the institutional hip registry. All patients had provided informed consent for future use of their images in our tertiary care institutional hip preservation practice. All Health Insurance Portability and Accountability Act of 1996 regulations were followed.[8]

Patients. From our hip preservation database, we identified 325 hips from 276 patients with complete radiological imaging from May 2016 to December 2021. The complete radiological imaging consisted of an anteroposterior (AP) pelvis, 45° Dunn, frog-leg lateral, and false-profile views. The inclusion criteria included: ages 14 to 100 years; any sex; complete radiological imaging series; and a reference final diagnosis of HD based on consensus radiological opinions of an independent
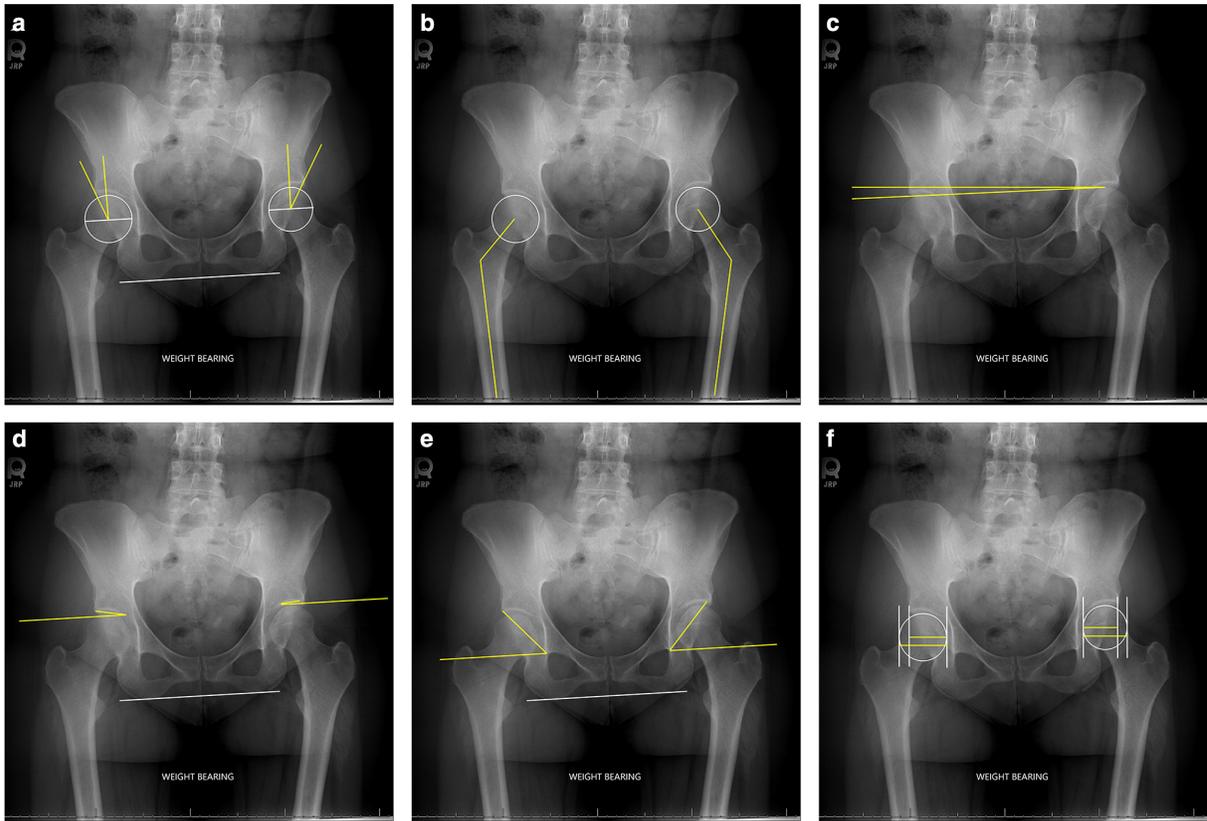
Fig. 2

Landmarks used by manual readers. a) Lateral centre-edge angle. b) Caput-collum-diaphyseal angle. c) Obliquity. d) Tönnis angle. e) Sharp's angle. f) Femoral head coverage.
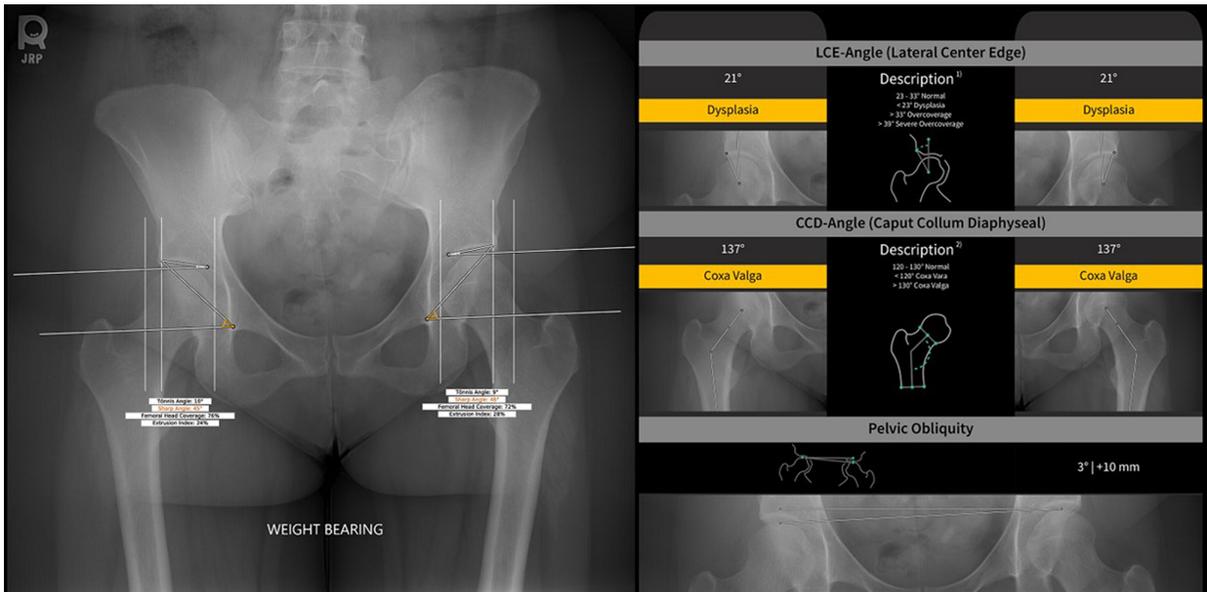

Fig. 3

Example HIPPO output. This figure shows an example of the reports that HIPPO produces.

fellowship-trained musculoskeletal (MSK) radiologist and hip preservation surgeon using the four-view hip series as well as surgical findings of arthroscopy and/or periacetabular osteotomy in the electronic health records. The

Table I. Landmarks used for manual measurements and artificial intelligence-based algorithm on anteroposterior pelvis images.

Measurement	Method
LCEA	The LCEA was measured between a line originating at the centre of the femoral head extending upwards perpendicular to a line connecting the inferior aspects of the ischial tuberosities and a line from the centre of the femoral head to the lateral acetabular sourcil.[6]
CCD	The CCD angle was measured as the angle between the femoral neck axis and the femoral shaft axis.[9]
Obliquity	The pelvic obliquity was measured by drawing an angle between a horizontal line extending from the apex of the femoral head on the side that is higher and a line connecting the apex of each femoral head.[10]
Tönnis angle	The Tönnis angle was measured by drawing an angle between a line parallel to the line connecting the inferior aspect of the ischial tuberosities and a line connecting the inferior and lateral aspects of the acetabular sourcil.[7]
Sharp's angle	Sharp's angle was measured by drawing a line at the level of the lower edge of the acetabular teardrop that is parallel to the line connecting the inferior aspect of the ischial tuberosities and a line connecting the lower edge of the acetabular teardrop and the lateral edge of acetabular sourcil.[12]
Femoral head coverage	The femoral head coverage was calculated by using three vertical lines: one representing the medial aspect of the femoral head, one representing the lateral aspect of the femoral head, and one representing the lateral edge of the acetabular sourcil. The femoral head coverage was represented by the percentage of femoral head covered versus the total horizontal head diameter.[12] The extrusion index was simply the percentage femoral head coverage subtracted from one.

CCD, caput-collum-diaphyseal; LCEA, lateral centre-edge angle.

Table II. Patient characteristics for the final study sample.

Variable	Male	Female	Overall
Patients, n	23	107	130
Median age, yrs (IQR)	23.98 (19.91 to 35.39)	29.32 (21.90 to 36.21)	28.61 (21.81 to 36.22)
Median weight, kg (IQR)	81.00 (72.50 to 94.50)	67.00 (55.00 to 77.50)	71.00 (58.25 to 81.75)
Median height, m (IQR)	1.80 (1.75 to 1.85)	1.63 (1.60 to 1.70)	1.65 (1.60 to 1.73)
Median BMI, kg/m^2 (IQR)	24.00 (22.69 to 29.00)	24.00 (21.00 to 29.50)	24.00 (21.00 to 29.75)

IQR, interquartile range.

exclusion criteria included: lack of complete radiological series; lack of concordant diagnosis among two specialists; hips with prior surgical intervention; avascular necrosis; and hip arthroplasty. The concordant diagnosis of HD by both specialists resulted in 276 hips from 138 patients in the study sample. In addition, two patients had both hips excluded because they did not have immediate preoperative images, five patients had a single hip excluded due to lack of preoperative images, two patients were excluded because they did not meet AI image quality criteria due to inadequate femoral visibility, and four patients had seven hips excluded as the AI did not generate output due to technical failures. It is not clear why the AI did not generate output for these cases. This resulted in a final cohort of 130 patients and 256 hips (Figure 1).

Patient demographic data including age, sex, and BMI were also extracted from the electronic health records.

Imaging parameters. All scans were performed using the standing (weightbearing) AP pelvis view, which allows visualization of both hips. The tube-to-film distance was 120 cm using 80 to 90 kilovoltage peak (kVp) and 20 to 30 milli-ampere-second (mAs) depending upon the size of the patient. For the AI algorithm to work, at least 1.5 times the femur's width must be visible below the most distal point of the lesser trochanter as per the vendor specifications. Four hips did not meet the image quality criteria from the vendor due to inadequate femoral visibility.

AI algorithm. A vendor-provided deep-learning-based software (HIPPO; ImageBiopsy Lab, Austria) automatically locates anatomical landmarks on AP pelvis radiographs and performs the six measurements including LCEA, caput-collum-diaphyseal (CCD) angle (also known as the femoral neck-shaft angle),[9] pelvic obliquity,[10] Tönnis angle,[11] Sharp's angle,[12] and femoral head coverage (Table I, Figure 2, Figure 3).[12] The software returns an error if specific DICOM metadata are not present or incorrectly specified, or if the image cropping prevents reliable measurements. All images were transferred via a secure research picture archiving and communication system (IPACS; Philips, the Netherlands) server to the vendor for evaluation.

Manual measurements. The three readers for manual measurements in the study were trained medical students (HA, SR, AA). After the senior MSK radiologist (AC) instructed the readers on how to properly measure, each reader practised measurements on ten images and, in addition, compiled images demonstrating the landmarks used for all measurements on an additional ten cases. The radiologist re-evaluated the landmarks each reader used and provided feedback for appropriate use of landmarks.

Following this process, the LCEA, CCD angle, pelvic obliquity, Tönnis angle, femoral head coverage, and Sharp's angle were measured by each reader using IPACS with a built-in measurement tool. Each reader measured all values on all of the patients in the study independently and was blinded to the AI measurements. Each reader also used a stopwatch to record the time spent obtaining all measurements for each patient, from the time images

Table III. Intraclass correlation coefficient and 95% confidence intervals for the reader agreement analysis.

Variable	Reader 1 vs Reader 2 ICC (95% CI)		Reader 1 vs Reader 3 ICC (95% CI)		Reader 2 vs Reader 3 ICC (95% CI)	
	Left hip	Right hip	Left hip	Right hip	Left hip	Right hip
LCEA	0.87 (0.81 to 0.91)	0.88 (0.74 to 0.94)	0.76 (0.67 to 0.82)	0.81 (0.68 to 0.88)	0.82 (0.69 to 0.89)	0.84 (0.78 to 0.88)
Tönnis angle	0.74 (0.30 to 0.88)*	0.86 (0.71 to 0.93)	0.80 (0.64 to 0.88)	0.84 (0.43 to 0.93)	0.84 (0.77 to 0.89)	0.89 (0.84 to 0.92)
Sharp's angle	0.89 (0.81 to 0.93)	0.90 (0.86 to 0.93)	0.86 (0.76 to 0.91)	0.84 (0.78 to 0.89)	0.84 (0.78 to 0.89)	0.83 (0.76 to 0.88)
CCD angle	0.61 (0.00 to 0.87)*	0.63 (0.00 to 0.86)*	0.63 (0.28 to 0.80)*	0.54 (0.40 to 0.66)	0.69 (0.56 to 0.78)	0.59 (0.42 to 0.71)
Femoral head coverage	0.84 (0.78 to 0.89)	0.84 (0.77 to 0.89)	0.77 (0.59 to 0.86)	0.80 (0.52 to 0.90)	0.81 (0.71 to 0.87)	0.81 (0.71 to 0.87)
Pelvic obliquity	0.83 (0.66 to 0.90)		0.80 (0.61 to 0.89)		0.93 (0.91 to 0.95)	

*Wide confidence intervals were observed.
CCD, caput-collum-diaphyseal; CI, confidence interval; ICC, intraclass correlation coefficient; LCEA, lateral centre-edge angle.

Table IV. Bland-Altman bias values with lower and upper limits of agreement for the reader agreement analysis.

Variable	Reader 1 vs Reader 2 (LOA)		Reader 1 vs Reader 3 (LOA)		Reader 2 vs Reader 3 (LOA)	
	Left hip	Right hip	Left hip	Right hip	Left hip	Right hip
LCEA,°	-1.2 (-8.5 to 6.2)	2.0 (-4.7 to 8.7)	1.0 (-9.5 to 11.6)	2.0 (-7.0 to 11.0)	2.2 (-7.1 to 11.5)	0 (-9.2 to 9.2)
Tönnis angle,°	-2.8 (-9.1 to 3.6)*	-1.9 (-8.2 to 4.4)	-1.7 (-8.2 to 4.8)	-2.7 (-8.7 to 3.3)	1.1 (-5.0 to 7.1)	-0.9 (-7.3 to 5.6)
Sharp's angle,°	-0.8 (-4.5 to 2.8)	0.2 (-3.8 to 4.1)	-0.9 (-5.3 to 3.4)	-0.5 (-5.3 to 4.4)	-0.1 (-5.1 to 4.9)	-0.6 (-5.6 to 4.4)
CCD angle,°	5.5 (0.4 to 10.7)*	4.0 (-2.1 to 10.0)*	3.8 (-6.5 to 14.1)*	1.7 (-9.5 to 13.0)	-1.7 (-12.1 to 8.7)	-2.2 (-12.7 to 8.2)
Femoral head coverage, %	1.0 (-7.5 to 9.4)	1.2 (-7.0 to 9.5)	2.7 (-7.6 to 13.1)	3.1 (-5.5 to 11.6)	1.8 (-8.0 to 11.6)	1.8 (-7.8 to 11.5)
Pelvic obliquity,°	-0.3 (-1.6 to 0.9)		-0.4 (-1.7 to 0.9)		0.0 (-0.8 to 0.7)	

*Large bias was observed.
CCD, caput-collum-diaphyseal; LCEA, lateral centre-edge angle; LOA, limit of agreement.

Table V. Intraclass correlation coefficient and 95% confidence intervals for the manual reader vs HIPPO agreement analyses, excluding HIPPO measurements generated from a single outlying patient.

Variable	Reader 1 vs HIPPO ICC (95% CI)		Reader 2 vs HIPPO ICC (95% CI)		Reader 3 vs HIPPO ICC (95% CI)	
	Left hip	Right hip	Left hip	Right hip	Left hip	Right hip
LCEA	0.85 (0.78 to 0.89)	0.84 (0.78 to 0.88)	0.86 (0.81 to 0.90)	0.78 (0.64 to 0.86)	0.75 (0.61 to 0.83)	0.71 (0.57 to 0.81)
Tönnis angle	0.82 (0.69 to 0.89)	0.82 (0.38 to 0.92)*	0.83 (0.72 to 0.90)	0.87 (0.81 to 0.91)	0.84 (0.78 to 0.88)	0.90 (0.86 to 0.93)
Sharp's angle	0.86 (0.81 to 0.90)	0.83 (0.77 to 0.88)	0.80 (0.60 to 0.89)	0.81 (0.73 to 0.86)	0.80 (0.57 to 0.89)	0.74 (0.63 to 0.82)
CCD angle	0.73 (0.06 to 0.90)*	0.75 (0.62 to 0.84)	0.79 (0.60 to 0.88)	0.72 (0.38 to 0.86)*	0.65 (0.54 to 0.74)	0.62 (0.50 to 0.72)
Femoral head coverage	0.73 (0.07 to 0.90)*	0.68 (0.00 to 0.88)*	0.67 (0.00 to 0.88)*	0.64 (0.00 to 0.88)*	0.61 (0.00 to 0.85)*	0.5 (0.00 to 0.79)*
Pelvic obliquity	0.83 (0.59 to 0.91)		0.95 (0.93 to 0.96)		0.98 (0.97 to 0.98)	

*Wide confidence intervals were observed.
CCD, caput-collum-diaphyseal; CI, confidence interval; ICC, intraclass correlation coefficient; LCEA, lateral centre-edge angle.

were loaded on PACS until the recording of all measurements on Excel (Microsoft, USA).

Statistical analysis. Patient demographic variables were summarized by median and interquartile range (IQR) if continuous, and by counts if categorical. In addition, the mean and standard error (SE) were reported for each of the seven measurement variables by each of the four readers (three readers and one AI algorithm) in the study.

Two separate agreement analyses were conducted through the calculation of intraclass correlation coefficients (ICC).[13] The first assessed pairwise inter-reader reliability among the three readers by estimating ICC values from a single-rating, absolute-agreement, two-way random-effects model. The second assessed inter-reader reliability between each of the readers and the HIPPO algorithm by estimating ICC values from a single-rating, absolute agreement, two-way mixed-effects model. In both analyses, 95% confidence intervals (CIs)

were reported for the ICC estimates. Benchmarks used for interpretation of the ICC estimates were: 0.00 to 0.40 poor; 0.40 to 0.59 fair; 0.60 to 0.74 good; and 0.75 to 1.00 excellent.[14]

Bland-Altman analyses were also conducted between all readers to supplement the ICC analysis results.[15] The estimated bias between reader measurements was reported along with the lower and upper limits of agreement. The estimated limits of agreement provide a reference interval within which most differences between measurements by the two readers are expected to occur.

To calculate the percentage time reduction offered by the HIPPO algorithm for a given patient, a linear mixed model was fit with log-transformed time as the dependent variable and a four-level categorical variable indicating the reader (three readers and one AI algorithm) as the independent variable. Random intercepts were included for each patient hip. Linear contrasts were

Table VI. Bland-Altman bias values with limits of agreement for the reader vs HIPPO agreement analyses, excluding HIPPO measurements generated from a single outlying patient.

Variable	Reader 1 vs HIPPO (LOA)		Reader 2 vs HIPPO (LOA)		Reader 3 vs HIPPO (LOA)	
	Left hip	Right hip	Left hip	Right hip	Left hip	Right hip
LCEA, °	-1.1 (-8.0 to 5.8)	0.0 (-7.5 to 7.5)	0.1 (-7.6 to 7.8)	-2.0 (-10.5 to 6.6)	-2.1 (-12.1 to 7.9)	-2.0 (-11.8 to 7.8)
Tönnis angle, °	-1.4 (-7.3 to 4.4)	-2.6 (-8.2 to 3.1)*	1.3 (-4.4 to 7.1)	-0.8 (-6.9 to 5.3)	0.3 (-5.9 to 6.5)	0.1 (-5.5 to 5.7)
Sharp's angle, °	0.5 (-3.6 to 4.6)	0.5 (-4.0 to 4.9)	1.4 (-3.2 to 5.9)	0.3 (-4.5 to 5.1)	1.5 (-3.2 to 6.1)	0.9 (-4.6 to 6.4)
CCD angle, °	3.6 (-2.4 to 9.6)*	1.5 (-5.5 to 8.4)	-1.9 (-8.5 to 4.7)	-2.5 (-9.2 to 4.2)*	-0.2 (-11.9 to 11.6)	-0.3 (-11.2 to 10.6)
Femoral head coverage, %	-4.8 (-13 to 3.3)*	-5.2 (-12.8 to 2.4)*	-5.8 (-14.4 to 2.8)*	-6.4 (-13.6 to 0.8)*	-7.6 (-17.2 to 2.1)*	-8.2 (-18.5 to 2.1)*
Pelvic obliquity	-0.4 (-1.6 to 0.8)		-0.1 (-0.9 to 0.8)		0.0 (-0.6 to 0.5)	

*Analyses that resulted in wide intraclass correlation coefficient confidence intervals.
CCD, caput-collum-diaphyseal; LCEA, lateral centre-edge angle; LOA, limit of agreement.

Table VII. Reader measurements.

Reader	LCEA, °	CCD, °	Obliquity, °	Tönnis grade	Sharp's angle, °	Femoral head coverage, %	Extrusion index, %
Reader 1 mean (range)	18.1 (-9.4 to 34.4)	137.5 (124.3 to 168.5)	1.3 (0.0 to 7.2)	10.9 (-4.8 to 36.9)	43.5 (33.2 to 60.3)	71.1 (46.7 to 88.5)	28.9 (11.5 to 53.3)
Reader 2 mean (range)	17.7 (-14.7 to 39.7)	132.7 (119.7 to 155.3)	1.6 (0.0 to 9.4)	13.3 (-2.6 to 44.3)	43.9 (33.5 to 61.5)	70.0 (41.4 to 87.8)	30 (12.2 to 58.6)
Reader 3 mean (range)	16.6 (-12.1 to 38.2)	134.7 (115.6 to 160.6)	1.7 (0.0 to 9.5)	13.2 (-1.3 to 38.8)	44.3 (34.2 to 59.5)	68.2 (37.8 to 89.3)	31.8 (10.7 to 62.2)
AI mean (range)*	19.3 (-2.7 to 34.2)	134.9 (84.3 to 180.0)	1.7 (0.0 to 9.3)	12.5 (-2.5 to 28.9)	43.0 (32.7 to 55.5)	76.2 (53.6 to 100.8)	23.8 (-0.8 to 46.4)

*One extreme outlier was not included in the HIPPO calculations due to erroneous landmark placement.
AI, artificial intelligence; CCD, caput-collum-diaphyseal; LCEA, lateral centre-edge angle.

estimated and exponentiated to calculate the percentage time reduction produced by the HIPPO algorithm for a given patient relative to each of the three readers. Three statisticians (LV, AH, YX) were involved in the discussion of methods and assisted with the statistical analysis.

Agreement analyses were performed in R (R Core Team; R Foundation for Statistical Computing, Austria) using the irr and BlandAltmanLeh packages. The timing mixed model analysis was performed in the SAS v. 9.4 Mixed Procedure (SAS Institute, USA).

Results

Patients. An orthopaedic surgeon (JW) classified the hips according to the Tönnis grade.[11] The median Tönnis grade was 0 with the majority 204 hips (79.7%) having Tönnis grade 0, 51 hips (19.9%) with Tönnis grade 1, and one hip (0.4%) with Tönnis grade 2. Further patient characteristics are described in Table II. During AI algorithm (HIPPO) implementation, seven hips could not be processed by the AI algorithm due to technical issues.

Reader measurements. The mean measurements of the three readers and HIPPO are presented in Table VII.

Inter-reader analysis. ICC estimates across each pairwise reader analysis demonstrated fair to excellent agreement (Table III). Wide 95% CIs were observed in three of the measurements of the Reader 1 versus Reader 2 analysis: the left hip Tönnis angle, the left hip CCD angle, and the right hip CCD angle. In addition, wide intervals were observed for the left hip CCD angle measurement in

the Reader 1 versus Reader 3 analysis. The corresponding Bland-Altman results in Table IV indicated that these four analyses exhibited larger bias when compared to the other reader analysis within the same variable. Overall, Reader 1 recorded larger CCD angles and smaller Tönnis angles relative to Readers 2 and 3.

Reader-AI analysis. ICC estimates for Sharp's angle, CCD angle, and pelvic obliquity all demonstrated good to excellent agreement across each reader-AI analysis. However, the AI algorithm severely miscalculated several measurements for one of the 256 hips in the study by erroneously placing the lateral acetabular marker on the femur. The miscalculations led to the presence of an extreme outlier which impacted the ICC analyses for the right hip measurements of LCEA, Tönnis angle, and femoral head coverage.

Sensitivity analysis excluding one outlier. Table V and Table VI display the ICC estimates and 95% CIs of the analyses excluding the hip with the outlying HIPPO measurements, respectively. In the absence of the outlying hip, all ICC estimates demonstrated fair to excellent agreement across all reader-HIPPO analysis. However, some analyses, as highlighted in the tables, still resulted in wide 95% CIs after removing the outlier. Inspection of the corresponding Bland-Altman results indicated that the observed ICC variability was again associated with large bias when compared to the other reader analysis within the same variable. In particular, the AI algorithm

generated systematically larger femoral head coverage measurements than each of the three readers.

Time savings. The median reading time for the three readers and AI was 212 seconds, 131 seconds, 734 seconds, and 41 seconds, respectively. For a given patient, the AI algorithm performed reads a mean 80.4% (79.7% to 81.1%), 70.1% (69.1% to 71.1%), and 94.4% (94.2% to 94.6%) faster than Reader 1, Reader 2, and Reader 3, respectively.

Discussion

To the authors' knowledge, there is no commercially available software that performs all of these fully automated HD measurements. The AI algorithm (HIPPO) has been validated in Europe and is CE-certified. It was used on an independent sample from USA and the external validation was successfully performed confirming reliable assessment of HD.

Excluding one severe outlier that significantly influenced the measurements, HIPPO-reader correlations were generally in the good to excellent range. This AI method was more reliable for LCEA, Tönnis grade, and Sharp's angle. Integration of this AI system could provide preliminary measurements to the physicians and direction for more thorough assessment for HD, especially in places without access to board-certified radiologists or orthopaedic surgeons to conduct the measurements. We used a large sample of proven cases of HD from our practice with standardized imaging and believe that this model will perform well in other settings if AP pelvis imaging is obtained with adequate inclusion of the proximal femur.

The HIPPO AI system performed reads between 70.1% and 80.4% faster than manual readers. Because comparable measurements were obtained between AI and the manual readers, implementing this AI-based model can produce rapid, consistent, and standardized measurements that may aid in the timely diagnosis of HD. In addition, the measurements can be imported in the electronic reports for future reference and longitudinal data collection.

Moreover, the AI system has the potential for significant cost savings. Based on time spent on HD measurements and the average orthopaedic surgeon and radiologist salaries as per the 2021 Doximity Physician Compensation Report,[16] for an average orthopaedic surgeon, each AI read would cost about $4.18 of the orthopaedic surgeon's time whereas the manual read would cost $36.59. For an average radiologist, each AI read would cost about $3.27 of the radiologist's time whereas the manual read would cost about $28.66. There are also non-financial costs, such as stress or fatigue from reading and inconsistent measurements. These radiographs are also very commonly obtained. At a large tertiary care hospital system like ours, there are tens of thousands of hip radiographs performed every year. Given high frequency of such radiographs, automated and consistent measurements that could make it to the electronic health record would be useful akin to echocardiogram-like measurements for a heart study.

This study was focused on preoperative HD patients with a reference standard diagnosis. It is possible that the AI software may perform better or worse on different patient populations, such as patients with FAI or normal hip anatomy. Although patients included in the study had a final diagnosis based on radiological assessment of two specialists, medical students (rather than attending physicians) performed all manual measurements in the study. However, this was intentional and meant to replicate more generalizability of the results and wider use in settings where trained radiologists are not available to perform the measurements. It is also possible that the time savings may be distorted by the latency of the image storage system. There was much wider variation in manual reading times than the automated reading times because of lag in the imaging system. Accurate time savings from AI, in part, are also dependent upon the consistency of the internet connection. In addition, this study included mostly Tönnis grade 0 and 1 patients as such patients are commonly referred for hip preservation, so this work represents a proof of concept for lower grades of hip degeneration. Higher Tönnis grades will be a subject of a future study.

As this study has shown how quickly and reliably this AI method can perform radiological measurements on HD patients, future studies could compare the AI measurements with patient-reported outcome measures, clinical symptoms of pain, functional hip scores, or intraoperative findings of labrum and cartilage damage. To conclude, this study demonstrated that the AI-based trained software contributed to significant time savings in reliable radiological assessment of patients with HD.

Take home message

- Most widely used measurements for hip dysplasia diagnosis (lateral centre-edge angle and Tönnis angle) demonstrated good to excellent inter-method reliability between the trained readers and artifical intelligence (AI)-based algorithm.
- Substantial time savings (to the order of 70% to 94%) were observed for hip radiological measurements per patient for all readers by using AI algorithm.

Twitter

Follow J. Wells @joelwellsmd
Follow A. Kohli @ajaykohlimd
Follow A. Hummer @ImageBiopsyLab
Follow A. Chhabra @AChhabraMD

References

1. **Morvan J**, **Bouttier R**, **Mazieres B**, **et al**. Relationship between hip dysplasia, pain, and osteoarthritis in a cohort of patients with hip symptoms. *J Rheumatol.* 2013;40(9):1583–1589.
2. **Gala L**, **Clohisy JC**, **Beaulé PE**. Hip dysplasia in the young adult. *J Bone Joint Surg Am.* 2016;98-A(1):63–73.

3. **Jacobsen S**, **Sonne-Holm S**. Hip dysplasia: a significant risk factor for the development of hip osteoarthritis. a cross-sectional survey. *Rheumatology (Oxford)*. 2005;44(2):211–218.

4. **Wenger DR**, **Bomar JD**. Human hip dysplasia: evolution of current treatment concepts. *J Orthop Sci*. 2003;8(2):264–271.

5. **Wells J**, **Millis M**, **Kim Y-J**, **Bulat E**, **Miller P**, **Matheney T**. Survivorship of the Bernese periacetabular osteotomy: what factors are associated with long-term failure? *Clin Orthop Relat Res*. 2017;475(2):396–405.

6. **Hanson JA**, **Kapron AL**, **Swenson KM**, **Maak TG**, **Peters CL**, **Aoki SK**. Discrepancies in measuring acetabular coverage: revisiting the anterior and lateral center edge angles. *J Hip Preserv Surg*. 2015;2(3):280–286.

7. **Clohisy JC**, **Carlisle JC**, **Beaulé PE**, **et al**. A systematic approach to the plain radiographic evaluation of the young adult hip. *J Bone Joint Surg Am*. 2008;90-A(Suppl 4):47–66.

8. Health Insurance Portability and Accountability Act of 1996 (HIPAA). Centers for Disease Control and Prevention. 2022. https://www.cdc.gov/phlp/publications/topic/hipaa.html (date last accessed 17 October 2022).

9. **Isaac B**, **Vettivel S**, **Prasad R**, **Jeyaseelan L**, **Chandi G**. Prediction of the femoral neck-shaft angle from the length of the femoral neck. *Clin Anat*. 1997;10(5):318–323.

10. **Giles LG**, **Taylor JR**. Low-back pain associated with leg length inequality. *Spine (Phila Pa 1976)*. 1981;6(5):510–521.

11. **Tönnis D**. *Congenital Dysplasia and Dislocation of the Hip in Children and Adults*. Berlin, Germany: Springer-Verlag, 1987.

12. **Tannast M**, **Hanke MS**, **Zheng G**, **Steppacher SD**, **Siebenrock KA**. What are the radiographic reference values for acetabular under- and overcoverage? *Clin Orthop Relat Res*. 2015;473(4):1234–1246.

13. **McGraw KO**, **Wong SP**. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1(1):30–46.

14. **Cicchetti DV**. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994;6(4):284–290.

15. **Bland JM**, **Altman DG**. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160.

16. **No authors listed**. 2021 Physician Compensation Report. Doximity. 2021. https://c8y.doxcdn.com/image/upload/v1/Press%20Blog/Research%20Reports/Doximity-Compensation-Report-2021.pdf (date last accessed 17 October 2022).

Author information:
- H. Archer, BSA, BBA, Medical Student
- S. Reine, BS, Medical Student
- A. Alshaikhsalama, BS, Medical Student
- J. Wells, MD, MPH, Orthopaedic Surgeon, Assistant Professor of Orthopaedic Surgery
- A. Kohli, MD, Musculoskeletal Radiologist, Assistant Professor of Radiology
- L. Vazquez, BS, PhD Student
- Y. Xi, PhD, Statistician
- A. Chhabra, MD, Chief of Musculoskeletal Radiology, Medical Director of Musculoskeletal Radiology
  The University of Texas Southwestern Medical Center, Dallas, Texas, USA.
- A. Hummer, PhD, Senior Clinical Research Specialist
- M. D. DiFranco, PhD, Chief Scientific Officer, Head of Clinical Research
- R. Ljuhar, PhD, Chief Executive Officer
  Image Biopsy Lab, Vienna, Vienna, Austria.

Author contributions:
- H. Archer: Conceptualization, Methodology, Project administration, Supervision, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing.
- S. Reine: Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Writing – review & editing.
- A. Alshaikhsalama: Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Writing – review & editing.
- J. Wells: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.
- A. Kohli: Conceptualization, Project administration, Resources, Supervision, Writing – review & editing.
- L. Vazquez: Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing.
- A. Hummer: Resources, Software.
- M. D. DiFranco: Resources, Software.
- R. Ljuhar: Resources, Software.
- Y. Xi: Methodology, Project administration, Supervision, Data curation, Formal analysis, Writing – review & editing.
- A. Chhabra: Conceptualization, Methodology, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.