 Received 06-12-22; Accepted 12-12-22; Published 15-12-22

Deep Learning-Generated Radiographic Hip Dysplasia Parameters: Relationship to Postoperative Patient-Reported Outcome Measures

Seth Reine seth.reine@utsouthwestern.edu
Holden Archer holden.archer@utsouthwestern.edu
Ahmed Alshaikhsalama ahmed.alshaikhsalama@utsouthwestern.edu
Joel E Wells joel.wells@utsouthwestern.edu
Ajay Kohli ajay.kohli@utsouthwestern.edu
Department of Orthopedic Surgery,
University of Texas Southwestern Medical
Center, Dallas, TX,
USA

Louis Vazquez louis.vazquez@utsouthwestern.edu
Department of Radiology,
University of Texas Southwestern Medical Center, Dallas, TX,
USA.

Allan Hummer a.hummer@imagebiopsy.com
Matthew D. Difranco m.difranco@imagebiopsy.com
Richard Ljuhar r.ljuhar@imagebiopsy.com
ImageBiopsy Labs, Vienna,
Austria

Yin Xi yin.xi@utsouthwestern.edu
Department of Radiology,
University of Texas Southwestern Medical Center, Dallas, TX,
USA

Avneesh Chhabra avneesh.chhabra@utsouthwestern.edu
MD, MBA, Department of Orthopedic Surgery,
University of Texas Southwestern Medical Center, Dallas, TX,
USA.
Medical advisor to ImageBiopsy Labs, Vienna,
Austria.
Adjunct faculty- Johns Hopkins, MD, Walton Center of Neurosciences,
Liverpool, UK, and University of Dallas,
Richardson, Tx

Corresponding Author: Avneesh Chhabra

Abstract

Background: Hip dysplasia (HD) causes accelerated osteoarthrosis of the acetabulum and is diagnosed through radiographic evaluation. An artificial intelligence (AI) program capable

Citation: Seth Reine, et al. Deep Learning-Generated Radiographic Hip Dysplasia Parameters: Relationship to Postoperative Patient-Reported Outcome Measures. Advances in Artificial Intelligence and Machine Learning 2022;2(4):37.

of measuring the necessary anatomical landmarks relevant to HD could reduce resource utilization, increase standardized HD screenings, and form HD outcome models. The study's aim was to evaluate the relationship between AI measurements of dysplastic hips on initial presentation and changes in patient-reported outcome measures following surgical intervention for HD.

Methods: One hundred nine patients with HD and planned surgical intervention obtained preoperative anterior-posterior pelvic radiographs which were measured by the HIPPO AI for lateral center edge angle, Tönnis angle, Sharp angle, Caput-Collum-Diaphyseal angle, femoral coverage, femoral extrusion, and pelvic obliquity. Patients completed a preoperative survey containing the 12-Item Short Form, EuroQol Visual Analog Scale (EQVAS), International Hip Outcome Tool (iHOT-12), Harris Hip Score, and Visual Analog Pain Scales. Patients were recommended to follow up at four months and one year to complete the same survey. Changes in outcome measures were evaluated with paired t-tests for each follow-up interval. Partial Spearman Rank-order correlations were performed between radiographic measures and changes in outcome measures at each follow-up interval controlling for age, BMI, and follow-up time.

Results: Patients had significant improvement in all outcome measures at four months (N=46,p-values<0.05) and one year (N=49,p-values<0.001), except one-year EQVAS (p-value=0.090). Significant positive correlation of moderate strength existed between the Sharp angle and iHOT-12 at four months postoperatively (r_s=0.472,p-value=0.044). No other significant correlations were found at either follow-up interval between HIPPO measures and outcome measures.

Conclusion: Correlations between deep learning radiographic measurements of dysplastic hips and improvements in postoperative outcomes as evaluated by outcome measures lacked any significant relationships in this study. Physicians treating HD patients can augment care with AI tools but outcomes are likely more multi-factorial and require multi-disciplinary patient care.

Keywords: Deep learning, Hip dysplasia, Outcomes, Artificial intelligence, Periacetabular osteotomy

1. INTRODUCTION

Acetabular dysplasia (HD) is characterized by insufficient coverage of the femoral head by the acetabulum due to a reduced size or exaggerated vertical orientation [1]. The resultant altered biomechanics from repeated acetabular edge loading can accelerate the hyaline cartilage loss and development of premature osteoarthrosis [1, 2]. The incidence of developmental HD varies widely among different ethnic populations with an approximate overall incidence of one per 1000 births [3]. Factors positively associated with HD include female gender, breech position in utero, and first-degree family history [3]. Sequelae of HD include articular cartilage damage, labral tear, osteoarthritis, hip instability, and subluxation with associated pain and functional limitations [1, 2, 4]. Other clinical manifestations in adults include leg length discrepancy, femoral head impingement, Trendelenburg gait, limp, altered pelvic tilt, and limited hip abduction [5, 6]. Treatment of

HD ranges from conservative management to arthroscopic labral repair paired with periacetabular osteotomy (PAO), femoral derotational osteotomy, or total hip arthroplasty (THA) depending upon the pathologic severity and stage of presentation [7-11].

Diagnosis of HD in skeletally mature individuals depends on radiographic measurement of anatomical landmarks. The most used and validated measure of acetabular under coverage is the lateral center-edge angle (LCEA) on the anteroposterior (AP) pelvis view ranging from $<25°$ to $<20°$ with lower LCEA measurements associated with earlier presentation of HD symptoms [6, 12-14]. Additional parameters used during pre-operative planning include the Tönnis angle, Sharp angle, femoral neck-shaft angle/Caput-Collum-Diaphyseal angle (CCD), femoral head extrusion index, acetabular crossover sign, and femoral head obliquity, etc.[12]. MRI and CT imaging are used in pre-operative planning to evaluate the labrum, hyaline cartilage, and soft tissue pathology [6]. A reader must synthesize all available clinical and imaging information to accurately diagnose hip dysplasia, which can be difficult in borderline cases [6]. Individual readers may also measure the same landmarks differently depending on training, familiarity, and variations in radiographic quality, and image viewing software. These factors render timely initiating management of HD both resource and expertise intensive with substantial room for inconsistency.

The clinical manifestations of HD as perceived by the patients themselves are optimally assessed using patient-reported outcome measures (PROM). PROMs elucidate a subjective, patient-centered view of quality of life, physical function, mental health, activity of daily living limitations, and pain through readily distributable surveys. Thus, patients can share consistent information in a standardized manner which updates through time for follow-up of disease progression and recovery. PROMs also allow for shared decision-making with physicians when deliberating conservative versus surgical treatment of HD. Several PROMs including, the Harris Hip Score (HHS), Visual Analog Scale for Pain (VAS), International Hip Outcome Tool (iHOT-12), EuroQol Visual Analog Scale (EQVAS), and 12-Item Short Form survey (SF-12) are commonly used in any hip preservation practice for patient follow-ups [15-18].

Standardization of radiographic HD evaluation provides a platform for generalizable investigation into the relationships between imaging, surgical decisions, and longitudinal changes in PROMs postoperatively. One approach to image measurement standardization involves the implementation of artificial intelligence (AI) image reading software and AI generated consistent radiographic measurements could be correlated against the differences in PROMs between the preoperative and postoperative states. This study is the first investigation of the AI measures' relationship with differences in pre- and post-operative outcomes in the HD population. We hypothesized that preoperative AI hip measures of increasing HD severity will correlate with greater absolute improvement in PROMs postoperatively.

This was a retrospective cross-sectional analysis of a prospectively collected hip preservation database at our tertiary care institute. Institutional review board approval was in place as a part of an ongoing institutional prospective longitudinal cohort study of HD outcomes. Additional approval was obtained for AI analysis of the radiographic images.

2. PATIENTS AND METHODS

2.1 Study Design and Patient Population

The patients were recruited from 1,222 individuals who presented for hip related issues between September 2016 and December 2021 at an urban medical center. A musculoskeletal fellowship trained radiologist and hip preservation expert orthopedic surgeon independently screened these patients using clinical presentation and radiographs to identify 138 mutually agreed upon HD patients who subsequently received surgical interventions. All 138 patients consented to study participation and completed an initial preoperative survey either at pre-operative clinical visit or through the virtual the REDCap® online platform, which included HHS, iHOT-12, SF-12, EQVAS, and VAS pain. All patients were recommended to postoperatively follow up for at least one year. Therefore, we defined short-term follow-up as within 2 weeks of the four-month mark after surgery and long-term follow-up as greater than one year. All analyses were done for the differences in PROMs between the pre-operative baseline and each respective follow-up periods. Follow-up surveys were completed post-operatively at subsequent clinical visits or through digital distribution. Included cases in this study had ages 14-50 years, all genders and possessed a complete preoperative survey, surgical intervention for HD, and a follow-up survey at either the four-month interval and/or greater than one-year interval postoperatively. Cases were excluded if the patients had undergone prior hip related surgery, only followed up outside of the described intervals, or if the individual questionnaires were left incomplete. Patient inclusion flowchart is displayed in FIGURE 1.

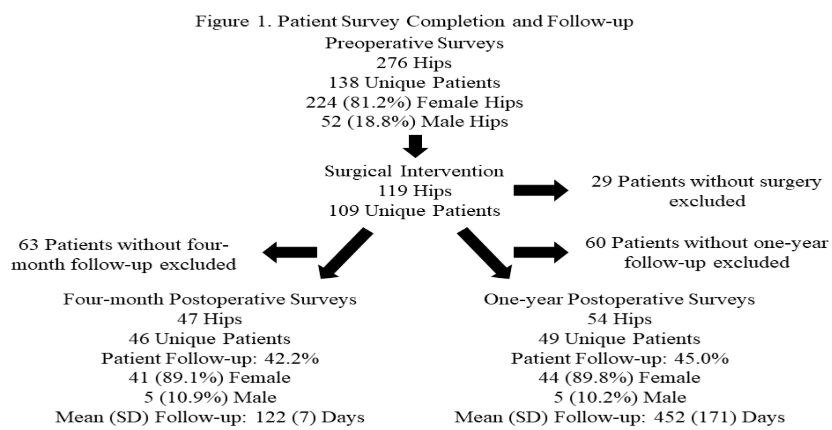

Figure 1: Patient Recruitment and Survey Completion.

2.2 Artificial Intelligence Evaluation

ImageBiopsy Labs © Vienna, Austria, has developed a machine learning AI system, named HIPPO, capable of measuring commonly used metrics to diagnose hip dysplasia in an AP pelvic radiograph [19]. HIPPO has already been validated against multiple human readers on dysplastic hips with a good to excellent intraclass correlation coefficient (manuscript under review). AP pelvis radiographs from the 138 pre-operative patients were evaluated by HIPPO with auto-generated outputs including LCEA, Tönnis angle, Sharp angle, CCD angle, femoral coverage, femoral extrusion, and femoral obliquity. The measurements were used for correlation analysis (FIGURE 2).

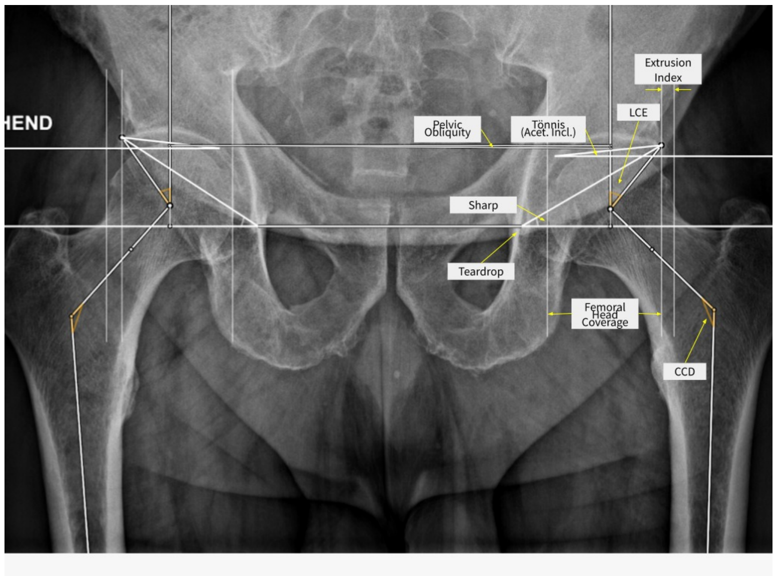

Figure 2: Typical HIPPO (AI) Output of a Hip Dysplasia Patient.

2.3 Statistical Analysis

Descriptive statistics for the hip dysplasia cohort were reported as mean and standard deviation for continuous metrics and count and frequency for categorical metrics. Means and standard deviations of changes in PROM values between preoperative and four months and greater than one-year post-operative follow-up were calculated for SF-12, EQVAS, iHOT-12, HHS, and VAS Pain scales with subsequent dependent Student's t-tests performed between time points. Bilateral surgical cases were handled on a per-patient basis with HIPPO measures used from the most dysplastic hip defined as the smallest LCEA. Correlations between initial presentation AI measures and changes in available PROMs between preoperative baseline and four-month and greater than one-year follow-up were performed using partial Spearman's rank order correlation controlling for body mass index (BMI), age, and follow-up time. Correlation p-values were corrected with the False Discovery Rate (FDR) method to account for multiple comparisons with the final significance alpha threshold set at 0.05 where applicable. All analyses were performed with Python 3.8 statsmodel package v0.13.2.

3. RESULTS

3.1 Patients

Surgical intervention for HD occurred in 119 hips from 109 unique patients out of the original 276 individual hips measured by HIPPO. Preoperative surveys were completed for 103 patients, and the same survey was completed postoperatively by 46 patients and 49 patients at the four-month and greater than one-year postoperative follow-up, respectively (FIGURE 1). Descriptive analysis is reported in TABLE. 1.

Table 1: Descriptive Analysis of HD Cohort.

	4-Month Follow-up	**1-Year Follow-up**
N	46	49
Unique Hips	47	54
Female	41 (89.1%)	44 (89.8%)
Male	5 (10.9%)	5 (10.2%)
Age (years)	29.2 (8.1)	31.2 (9.1)
BMI (kg/m^2)	24.8 (5.5)	25.8 (5.5)
Follow-up (days)	122 (7)	452 (171)
Periacetabular Osteotomy	38 (80.9%)	43 (79.6%)
Total Hip Arthroplasty	3 (6.4%)	3 (5.6%)
Femoral Derotation	1 (2.1%)	0
Arthroscopy	4 (8.5%)	8 (14.8%)
Surgical Hip Dislocation	1 (2.1%)	0

3.2 Change in Patient-Reported Outcome Measures pre- to post-operative

Paired samples Students' t-test found statistically significant improvements four months after surgery for all PROMs (SF-12 p=0.004, EQVAS p=0.026, other PROMs p-values<0.001). Significant improvements were also found at least one year after surgery for all PROMs (p-values<0.001), except EQVAS (p-value=0.090) (TABLE 2).

Table 2: Postoperative Change in PROMs at 4 months and 1 year for HD Patients with Paired t-test [Mean (SD).]

Outcome Measure	Delta Mean 4 Months N=46	p-value	Delta Mean 1 Year N=49	p-value
SF-12	5.7 (10.9)	0.004	10.2 (13.1)	<0.001
EQVAS	8.3 (22.1)	0.026	3.9 (15.6)	0.091
iHOT-12	20.7 (29.5)	<0.001	29.3 (27.7)	<0.001
HHS	12.9 (20.3)	<0.001	25.1 (19.6)	<0.001
VAS Best	-1.8 (1.9)	<0.001	-1.5 (1.9)	<0.001
VAS Worst	-3.1 (2.8)	<0.001	-3.7 (2.7)	<0.001
VAS Now	-2.9 (2.7)	<0.001	-2.8 (2.5)	<0.001
VAS Average	-2.9 (2.7)	<0.001	-3.0 (2.6)	<0.001

3.3 Correlation of HIPPO Radiographic Measures With Change In PROMs

There was a significant positive correlation of moderate strength between the Sharp angle and iHOT-12 at four months postoperatively (r_s=0.472, p-value=0.044) (TABLE 3). No other significant correlations were found at either follow-up interval between HIPPO measures and PROMs (TABLE 3, TABLE 4). Scatterplots of radiographic measures versus PROMs are displayed in FIGURE 3 and FIGURE 4. Complete partial Spearman rank order correlation analysis can be found in supplemental materials.

Table 3: Radiographic Measure to PROMs Four-month Correlation Matrix (*,**,*** indicating <0.05,<0.01, <0.001 after FDR, respectively).

PROM	LCEA	Tönnis	Sharp	CCD	Fem	Extrusion	Obliquity
SF-12	-	-	-	-	-	-	-
EQVAS	-	-	-	-	-	-	-
iHOT-12	-	-	*	-	-	-	-
HHS	-	-	-	-	-	-	-
VAS Best	-	-	-	-	-	-	-
VAS Worst	-	-	-	-	-	-	-
VAS Now	-	-	-	-	-	-	-
VAS Average	-	-	-	-	-	-	-

Table 4: Radiographic Measure to PROMs One-year Correlation Matrix (*,**,*** indicating <0.05,<0.01, <0.001 after FDR, respectively).

PROM	LCEA	Tönnis	Sharp	CCD	Fem	Extrusion	Obliquity
SF-12	-	-	-	-	-	-	-
EQVAS	-	-	-	-	-	-	-
iHOT-12	-	-	-	-	-	-	-
HHS	-	-	-	-	-	-	-
VAS Best	-	-	-	-	-	-	-
VAS Worst	-	-	-	-	-	-	-
VAS Now	-	-	-	-	-	-	-
VAS Average	-	-	-	-	-	-	-

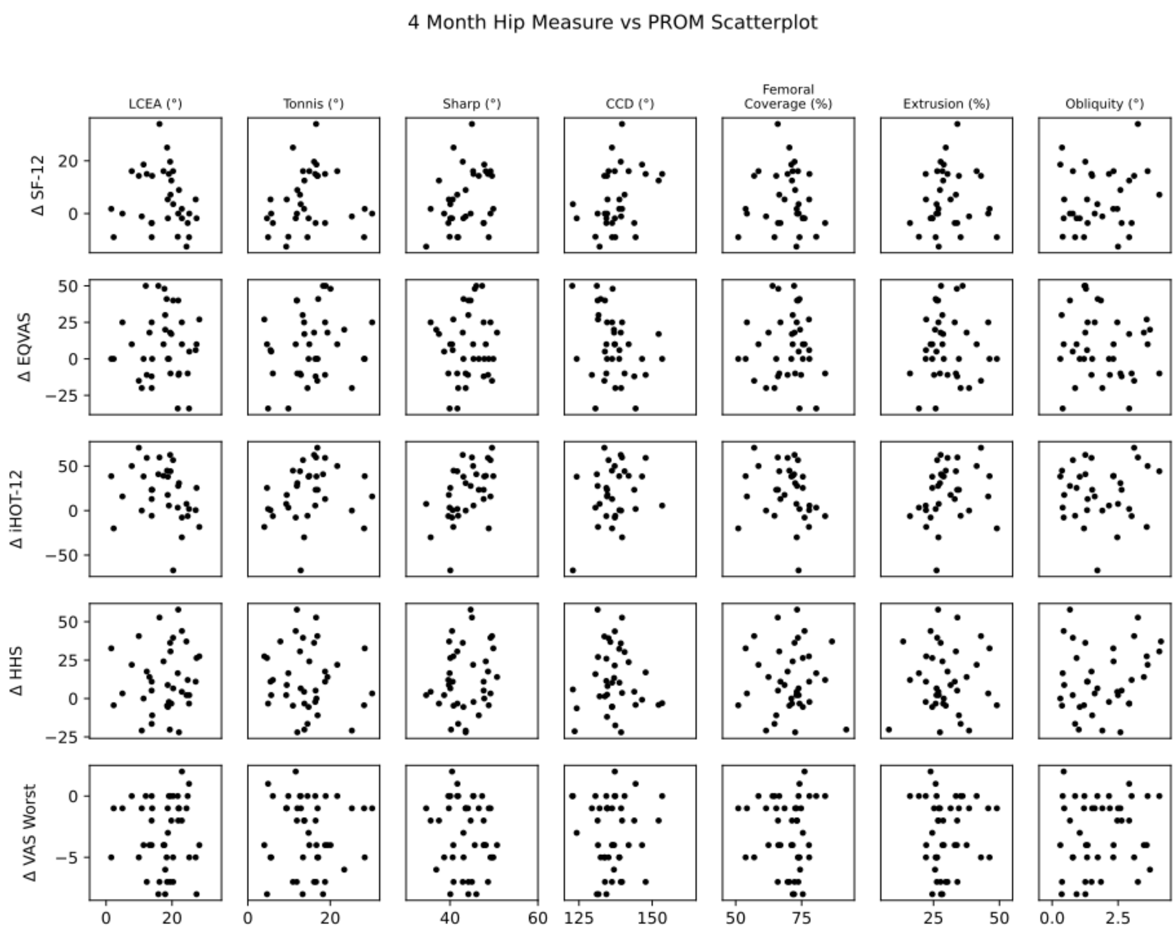

Figure 3: HIPPO Radiographic Measures and Patient Reported Outcome Measures (PROM) at Four Months.

Figure 4: HIPPO Radiographic Measures and Patient Reported Outcome Measures (PROM) at One Year.

4. DISCUSSION

The applications of artificial intelligence in medical imaging workflows have grown to encompass many medical specialties including several orthopedic conditions [20-22]. These orthopedic applications of deep learning software often focus on image segmentation and landmark-related measurements, and some secondary applications of deep learning have been explored for risk stratification of hip dislocations [22, 23]. The use of deep learning relating to HD has largely been limited to diagnostic endeavors without exploration of the impact of auto-generated radiographic measurements on patient outcome measures [24, 25]. Therefore, this study aimed to investigate the relationship between AI-driven deep learning generated measurements in HD patients and their postoperative outcomes.

As expected, the patients experienced statistically significant improvements in PROMs at both four months and greater than one year, post-operatively. Patients had minimal clinically important difference (MCID) at both follow-up intervals for SF-12 (MCID 4.3) and VAS Pain Scales (MCID 1.4) [15, 26]. Patients exceeded the MCID for iHOT-12 (MCID 23) and HHS (MCID 18) after one year but not at four months postoperatively suggesting that patients continue to experience improvements in their condition that have not yet manifested in short term follow-up [27, 28]. However, the patients did not obtain the MCID of 24 for EQVAS at any follow-up interval [29]. It suggests that surgical HD candidates may attain improvements in outcomes postoperatively below their threshold of notice.

Overall, there was a paucity of significant correlations between the radiographic parameters measured by the deep learning HIPPO system and changes in postoperative patient-reported outcome measures at any follow-up interval. It might be due to good generalized post-operative improvements among most PROMs with lack of substantial variation among different patients. The lack of correlation between the radiographic severity and symptom severity also suggests that other factors, such as the age of intervention, sex, family history, muscle strength, ligamentous laxity, functional limitations, mental health, activity level, and physical activity preferences must be considered alongside imaging to holistically characterize HD patients [5, 13, 14, 30]. The diagnosis of HD traditionally relies upon radiographic measurements, such as the LCEA which presents an opportunity for deep learning program use in screening, initial evaluation, and surgical planning to augment a physician's time efficiency. However, single-view radiographic measurements either performed by AI or manually lack a relationship to patient outcomes but do relate to the age of onset of symptoms [13, 14, 31]. Future deep learning systems trained on radiographic measures in addition to other aforementioned patient factors may produce decision models capable of drawing meaningful relationships to surgical outcomes in such patients.

This study possesses some limitations. The primary limitation was a relatively small initial cohort size and follow-up response rate of less than 50% at both intervals. A larger sample size would improve the study power and make the detection of smaller correlation effects more apparent. However, we included all consecutive cases in our practice and all patients were diagnosed as HD by two experts of different disciplines. The cohort is also skewed to encompass more female patients, which mirrors the general HD population, but future gender-specific investigations may reveal diverging results. The patients presenting for this cohort were subjectively symptomatic enough to seek surgical interventions, whereas the inclusion of non-symptomatic HD may better account for self-selection biases. Lastly, we utilized preoperative measurements as our initial study, not post-operative measurements.

The application of deep learning radiographic measurements in dysplastic hips likely has value in repetitive, standardized, and time-consuming tasks; however, the subjective variation in PROMs precludes the use of radiographs to predict patient symptomology. Future investigation of deep learning models accounting for radiographs in addition to other imaging modalities and subjective patient factors to predict postoperative outcomes in HD patients may have more value at the expense of overfitting to niche use. Not to mention, further investigation will be performed looking at changes in radiographic parameters pre and post operatively utilizing AI software.

5. CONCLUSION

To summarize, correlations between deep learning radiographic measurements of dysplastic hips and improvements in postoperative outcomes as evaluated by PROMs did not have any significant relationships in this study. Physicians treating HD patients can augment care with AI tools but continue to prioritize holistic approaches to multi-disciplinary patient care.

6. LIST OF ABBREVIATIONS

HD: Hip Dysplasia; AI: Artificial Intelligence; PROM: Patient Reported Outcome Measure; MCID: Minimal Clinically Important Difference; THA: Total Hip Arthroplasty; LCEA: Lateral Center Edge Angle; AP: anterior-posterior; CCD: caput-collum-diaphyseal; HHS: Harris Hip Score; VAS: Visual Analog Scale for Pain, iHOT-12: International Hip Outcome Tool; EQVAS: EuroQol Visual Analog Scale; SF-12: 12-Item Short Form survey, BMI: Body Mass Index, FDR: False Discovery Rate, PAO: Periacetabular Osteotomy

7. DECLARATIONS

Ethics approval and consent to participate

Institutional review board approval was obtained for longitudinal collection of clinical data related to hip pathology. All patients provided informed consent for study participation. The ethics committee was from the University of Texas Southwestern Medical Center.

8. CONSENT FOR PUBLICATION

The radiograph provided in FIGURE 2 comes from a deidentified patient who provided informed consent for participation and use.

9. AVAILABILITY OF DATA AND MATERIALS

The deep learning software used is proprietary to ImageBiopsy Labs and is not currently available for journal submission. The outcome data is available upon request.

10. COMPETING INTERESTS

AC: Consultant: ICON Medical and TREACE Medical Concepts Inc., Book Royalties: Jaypee, Wolters, Speaker: Siemens, Medical advisor: ImageBiopsy Inc.

MD, AH: Employee: IB Lab GmbH.

RL: Employee: IB Lab

JW: Consultant: Ethicon

Funding

Not applicable

11. AUTHORS' CONTRIBUTIONS

SR assisted in study design, data acquisition, data analysis, result interpretation, manuscript drafting, and manuscript revision. HA assisted in data acquisition and manuscript revision. AA assisted in data acquisition and manuscript revision. JW assisted with study design, data acquisition, and manuscript revision. AK assisted in study design. YX assisted in study design, data analysis, and manuscript revision. AC assisted with study design, data analysis, and manuscript revision. All authors read and approved the final manuscript and accept responsibility for their contributions.

12. ACKNOWLEDGEMENT

Allan Hummer: Deep learning system technical support
Matthew D. DiFranco: Deep learning system access
Richard Ljuhar: Deep learning system access
Ajay Kohli- Research grant from Imagebiopsy Lab, inc.
Avneesh Chhabra- Research grant from Imagebiopsy Lab, inc.

13. SUPPLEMENTAL MATERIALS

Number of Tables : 8.
Tables from TABLE a - TABLE h.

References

[1] Wiberg G. Studies on Dysplastic Acetabula and Congenital Subluxation of the Hip Joint With Special Reference to the Complication of Osteoarthritis. Stockholm. Norstedt: Exp. 1939.

[2] Gala L, Clohisy JC, Beaulé PE. Hip Dysplasia in the Young Adult. J Bone Joint Surg Am. 2016;98:63-73.

[3] Loder RT, Skopelja EN. The Epidemiology and Demographics of Hip Dysplasia. ISRN Orthop. 2011;2011:238607.

[4] Cooperman D. What Is the Evidence to Support Acetabular Dysplasia as a Cause of Osteoarthritis? J Pediatr Orthop. 2013;33:S2-S7.

[5] Nunley RM, Prather H, Hunt D, Schoenecker PL, Clohisy JC. Clinical Presentation of Symptomatic Acetabular Dysplasia in Skeletally Mature Patients. J Bone Joint Surg American;2011;93:17-21.

[6] Kraeutler MJ, Safran MR, Scillia AJ, Ayeni OR, Garabekyan T, et al. A Contemporary Look at the Evaluation and Treatment of Adult Borderline and Frank Hip Dysplasia. Am J Sports Med. 2020;48:2314-2323.

[7] Buly RL, Sosa BR, Poultsides LA, Caldwell E, Rozbruch SR. Femoral Derotation Osteotomy in Adults for Version Abnormalities. J Am Acad Orthop Surg. 2018;26:e416-e425.

[8] Armand M, Lepistö J, Tallroth K, Elias J, Chao E. Outcome of Periacetabular Osteotomy: Joint Contact Pressure Calculation Using Standing AP Radiographs, 12 Patients Followed for Average 2 Years. Acta Orthop. 2005;76:303-13.

[9] Kamath AF. Bernese Periacetabular Osteotomy for Hip Dysplasia: Surgical Technique and Indications. World J Orthop. 2016;7:280-6.

[10] Learmonth ID, Young C, Rorabeck C. The Operation of the Century: Total Hip Replacement. Lancet. 2007;370:1508-1519.

[11] Kelley SP, Feeney MM, Maddock CL, Murnaghan ML, Bradley CS. Expert-Based Consensus on the Principles of Pavlik Harness Management of Developmental Dysplasia of the Hip. JB JS Open Access. 2019;4:e0054.

[12] Clohisy JC, Carlisle JC, Beaulé PE, Kim YJ, Trousdale RT, et al. A Systematic Approach to the Plain Radiographic Evaluation of the Young Adult Hip. J Bone Joint Surg American;2008;90:47-66.

[13] Matheney T, Zaltz I, Kim YJ, Schoenecker P, Millis M, et al. Activity Level and Severity of Dysplasia Predict Age at Bernese Periacetabular Osteotomy for Symptomatic Hip Dysplasia. J Bone Joint Surg Am. 2016;98:665-671.

[14] Okpara S, Nakonezny P, Wells J. Do Psychological Factors or Radiographic Severity Play a Role in the Age of Onset in Symptomatic Developmental Dysplasia of Hip and Femoroacetabular Impingement Syndrome? BMC Musculoskelet Disord. 2019;20:412.

[15] Martin RL, Kivlan BR, Christoforetti JJ, Wolff AB, Nho SJ, et al. Minimal Clinically Important Difference and Substantial Clinical Benefit Values for a Pain Visual Analog Scale After Hip Arthroscopy. Arthroscopy. 2019;35:2064-2069.

[16] Griffin DR, Parsons N, Mohtadi NG, Safran MR. A Short Version of the International Hip Outcome Tool (iHOT-12) For Use in Routine Clinical Practice. Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc North Am Int Arthrosc Assoc. 2012;28:611-616.

[17] Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, et al. Patient-Reported Outcome Measures in Arthroplasty Registries. Acta Orthop. 2016;87:3-8.

[18] Nilsdotter A, Bremander A. Measures of Hip Function and Symptoms: Harris Hip Score (HHS), Hip Disability and osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. Arthritis Care Res. 2011;63:S200-S207.

[19] Stotter C, Hummer A, Difranco M, Nehrer S. Fully Automated Radiographic Measurements of the Pelvis. In: ICRS, editor. Society ICRJP. Berlin, Germany: ICRS. 2022.

[20] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, et al. A Survey on Deep Learning in Medical Image Analysis. Med Image Anal. 2017;42:60-88.

[21] Karnuta JM, Haeberle HS, Luu BC, Roth AL, Molloy RM, et al. Artificial Intelligence to Identify Arthroplasty Implants From Radiographs of the Hip. J Arthroplasty. 2021;36:S290-S294.e1.

[22] Rouzrokh P, Ramazanian T, Wyles CC, Philbrick KA, Cai JC, et al. Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty From Postoperative Radiographs. J Arthroplasty. 2021;36:2197-2203.e3.

[23] Yu JS, Yu SM, Erdal BS, Demirer M, Gupta V, et al. Detection and Localisation of Hip Fractures on Anteroposterior Radiographs With Artificial Intelligence: Proof of Concept. Clin Radiol. 2020;75:237.e1-237.e9.

[24] Park HS, Jeon K, Cho YJ, Kim SW, Lee SB, et al. Diagnostic Performance of a New Convolutional Neural Network Algorithm for Detecting Developmental Dysplasia of the Hip on Anteroposterior Radiographs. Korean J Radiol. 2021;22:612-623.

[25] Zhang SC, Sun J, Liu CB, Fang JH, Xie HT, et al. Clinical Application of Artificial Intelligence-Assisted Diagnosis Using Anteroposterior Pelvic Radiographs in Children With Developmental Dysplasia of the Hip. Bone Joint J. 2020;102-B:1574-1581.

[26] Canfield M, Savoy L, Cote MP, Halawi MJ. Patient-Reported Outcome Measures in Total Joint Arthroplasty: Defining the Optimal Collection Window. Arthroplast Today. 2020;6:62-67.

[27] Martin RL, Kivlan BR, Christoforetti JJ, Wolff AB, Nho SJ, et al. Minimal Clinically Important Difference and Substantial Clinical Benefit Values for the 12-Item International Hip Outcome Tool. Arthroscopy. 2019;35:411-416.

[28] Chahal J, Thiel GSV, Mather RC, Lee S, Salata MJ, Nho SJ. The Minimal Clinical Important Difference (MCId) And Patient Acceptable Symptomatic State (PASS) For the Modified Harris Hip Score and Hip Outcome Score Among Patients Undergoing Surgical Treatment for Femoroacetabular Impingement. Orthop J Sports Med. 2014;2:2325967114S00105.

[29] Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal Clinically Important Improvement (MCII) And Patient-Acceptable Symptom State (PASS) In Total Hip Arthroplasty (THA) Patients 1 Year Postoperatively. Acta Orthop. 2014;85:39-48.

[30] Kraeutler MJ, Garabekyan T, Pascual-Garrido C, Mei-Dan O. Hip Instability: A Review of Hip Dysplasia and Other Contributing Factors. Muscles Ligaments Tendons J. 2016;6:343-353.

[31] Takegami Y, Seki T, Osawa Y, Kusano T, Ishiguro N. The Relationship Between Radiographic Findings and Patient-Reported Outcomes in Adult Hip Dysplasia Patients: A Hospital Cross-Sectional Study. J Orthop Sci. 2020;25:606-611.

Supplemental Materials

Table a: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in SF-12 ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

SF-12	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	-0.25 [-0.55, 0.10]	0.161	n/a	-0.01 [-0.31, 0.29]	0.954	n/a
Tönnis	0.25 [-0.11, 0.55]	0.167	n/a	-0.06 [-0.36, 0.24]	0.694	n/a
Sharp	0.31 [-0.04, 0.59]	0.084	n/a	0.06 [-0.24, 0.36]	0.679	n/a
CCD	0.28 [-0.08, 0.57]	0.125	n/a	0.11 [-0.20, 0.40]	0.481	n/a
Coverage	-0.19 [-0.50, 0.17]	0.305	n/a	-0.02 [-0.32, 0.28]	0.899	n/a
Extrusion	0.19 [-0.17, 0.50]	0.305	n/a	0.02 [-0.28, 0.32]	0.899	n/a
Obliquity	0.12 [-0.24, 0.45]	0.513	n/a	0.09 [-0.22, 0.38]	0.585	n/a

Table b: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in EQVAS ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

EQVAS	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	0.11 [-0.22, 0.43]	0.513	n/a	0.18 [-0.12, 0.44]	0.239	n/a
Tönnis	0.12 [-0.21, 0.43]	0.472	n/a	-0.03 [-0.32, 0.26]	0.829	n/a
Sharp	0.01 [-0.32, 0.34]	0.952	n/a	-0.001 [-0.29, 0.29]	0.992	n/a
CCD	-0.25 [-0.54, 0.08]	0.139	n/a	0.02 [-0.27, 0.31]	0.903	n/a
Coverage	0.10 [-0.24, 0.41]	0.563	n/a	0.20 [-0.09, 0.47]	0.175	n/a
Extrusion	-0.10 [-0.41, 0.24]	0.563	n/a	-0.20 [-0.47, 0.09]	0.175	n/a
Obliquity	-0.20 [-0.50, 0.14]	0.240	n/a	0.06 [-0.23, 0.34]	0.693	n/a

Table c: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in iHOT-12 ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

iHOT-12	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	-0.35 [-0.63, -0.01]	0.047	0.099	0.15 [-0.16, 0.42]	0.343	0.657
Tönnis	0.29 [-0.07, 0.58]	0.109	0.152	-0.19 [-0.46, 0.11]	0.219	0.657
Sharp	0.47 [0.15, 0.70]	0.006	0.044	-0.07 [-0.36, 0.23]	0.657	0.657
CCD	0.13 [-0.23, 0.46]	0.475	0.554	0.08 [-0.23, 0.36]	0.623	0.657
Coverage	-0.34 [-0.62, 0.01]	0.057	0.099	0.10 [-0.2, 0.39]	0.509	0.657
Extrusion	0.34 [-0.01, 0.62]	0.057	0.099	-0.10 [-0.39, 0.2]	0.509	0.657
Obliquity	0.05 [-0.30, 0.39]	0.788	0.788	0.39 [0.11, 0.62]	0.008	0.057

Table d: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in HHS ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

HHS	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	0.08 [-0.26, 0.40]	0.643	n/a	0.15 [-0.16, 0.43]	0.341	n/a
Tönnis	-0.08 [-0.40, 0.25]	0.628	n/a	-0.11 [-0.39, 0.19]	0.477	n/a
Sharp	0.13 [-0.21, 0.44]	0.452	n/a	-0.02 [-0.32, 0.28]	0.879	n/a
CCD	0.04 [-0.29, 0.36]	0.823	n/a	-0.07 [-0.36, 0.23]	0.652	n/a
Coverage	0.02 [-0.31, 0.34]	0.917	n/a	0.14 [-0.17, 0.42]	0.379	n/a
Extrusion	-0.02 [-0.34, 0.31]	0.917	n/a	-0.14 [-0.42, 0.17]	0.379	n/a
Obliquity	0.19 [-0.15, 0.49]	0.276	n/a	0.24 [-0.06, 0.5]	0.113	n/a

Table e: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in VAS Pain Best ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

VAS Best	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	0.14 [-0.17, 0.43]	0.363	0.756	0.04 [-0.24, 0.32]	0.769	n/a
Tönnis	-0.12 [-0.41, 0.19]	0.432	0.756	-0.07 [-0.34, 0.22]	0.643	n/a
Sharp	-0.22 [-0.5, 0.08]	0.148	0.520	-0.14 [-0.4, 0.15]	0.343	n/a
CCD	0.33 [0.03, 0.58]	0.031	0.220	-0.01 [-0.29, 0.27]	0.941	n/a
Coverage	0.01 [-0.29, 0.31]	0.942	0.942	0.0009 [-0.28, 0.28]	0.994	n/a
Extrusion	-0.01 [-0.31, 0.29]	0.942	0.942	-0.0009 [-0.28, 0.28]	0.994	n/a
Obliquity	0.06 [-0.24, 0.36]	0.685	0.942	-0.22 [-0.47, 0.06]	0.125	n/a

Table f: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in VAS Pain Worst ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

VAS Worst	**4 months**			**Over 1 year**		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	0.07 [-0.24, 0.37]	0.665	n/a	-0.07 [-0.34, 0.22]	0.644	n/a
Tönnis	-0.08 [-0.38, 0.23]	0.603	n/a	0.09 [-0.2, 0.36]	0.545	n/a
Sharp	-0.15 [-0.43, 0.16]	0.353	n/a	0.03 [-0.25, 0.31]	0.826	n/a
CCD	-0.004 [-0.31, 0.30]	0.980	n/a	0.03 [-0.25, 0.3]	0.848	n/a
Coverage	0.02 [-0.29, 0.32]	0.900	n/a	-0.08 [-0.35, 0.2]	0.567	n/a
Extrusion	-0.02 [-0.32, 0.29]	0.900	n/a	0.08 [-0.2, 0.35]	0.567	n/a
Obliquity	0.15 [-0.16, 0.43]	0.354	n/a	-0.04 [-0.31, 0.25]	0.806	n/a

Table g: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in VAS Pain Now ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

VAS Now	4 months			Over 1 year		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	-0.01 [-0.31, 0.29]	0.941	n/a	-0.16 [-0.42, 0.13]	0.277	n/a
Tönnis	-0.02 [-0.32, 0.29]	0.918	n/a	0.18 [-0.11, 0.44]	0.219	n/a
Sharp	-0.15 [-0.43, 0.17]	0.359	n/a	-0.01 [-0.29, 0.27]	0.958	n/a
CCD	0.28 [-0.02, 0.54]	0.069	n/a	0.07 [-0.21, 0.35]	0.610	n/a
Coverage	-0.07 [-0.36, 0.24]	0.671	n/a	-0.17 [-0.43, 0.12]	0.254	n/a
Extrusion	0.07 [-0.24, 0.36]	0.671	n/a	0.17 [-0.12, 0.43]	0.254	n/a
Obliquity	0.19 [-0.13, 0.46]	0.240	n/a	-0.20 [-0.45, 0.09]	0.176	n/a

Table h: Partial Spearman's Rank-order correlation between HIPPO Measures and Change in VAS Pain Average ([95% Confidence Interval]; r_s: Spearman Coefficient; FDR: False Discovery Rate).

VAS Average	4 months			Over 1 year		
	rs	p-value	FDR	rs	p-value	FDR
LCEA	0.12 [-0.19, 0.41]	0.464	n/a	-0.09 [-0.36, 0.2]	0.556	n/a
Tönnis	-0.17 [-0.45, 0.14]	0.287	n/a	0.09 [-0.19, 0.36]	0.536	n/a
Sharp	-0.26 [-0.52, 0.05]	0.097	n/a	-0.07 [-0.34, 0.21]	0.626	n/a
CCD	0.16 [-0.15, 0.44]	0.308	n/a	0.08 [-0.2, 0.35]	0.578	n/a
Coverage	0.11 [-0.20, 0.40]	0.503	n/a	-0.10 [-0.37, 0.18]	0.477	n/a
Extrusion	-0.11 [-0.40, 0.20]	0.503	n/a	0.10 [-0.18, 0.37]	0.477	n/a
Obliquity	0.25 [-0.05, 0.52]	0.103	n/a	-0.09 [-0.36, 0.19]	0.532	n/a