ACCSCIENCE PUBLISHING

Surgical Technology International

RESEARCH ARTICLE

Agreement between artificial intelligence-assisted and clinician-derived proximal femoral anatomical measurements on computed tomography

Styliani Stergiadou[1]*, Antonios Koutalos[2], Grégoire Micicoi[3], Athanasios Batagiannis[4], Michael Klontzas[5,6], Konstantinos Malizos[2], and Aristeidis Zibis[1]

[1]Department of Anatomy, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece

[2]Department of Orthopaedic Surgery and Musculoskeletal Trauma, General University Hospital of Larissa, Larissa, Greece

[3]University Institute for Locomotion and Sports, Pasteur 2 Hospital, Université Côte d'Azur, Nice, France

[4]Arthroniq SMPC, Athens, Greece

[5]Department of Medical Imaging, University Hospital of Heraklion, Heraklion, Crete, Greece

[6]Artificial Intelligence and Translational Imaging (ATI) Lab, Department of Radiology, School of Medicine, University of Crete, Heraklion, Crete, Greece

***Corresponding author:**
Styliani Stergiadou
(stellastergiadou@gmail.com)

Citation: Stergiadou S, Koutalos A, Micicoi G, *et al*. Agreement between artificial intelligence-assisted and clinician-derived proximal femoral anatomical measurements on computed tomography. *Surg Technol Int*.
doi: 10.36922/STI026380043

Received: September 14, 2026

Revised: November 3, 2026

Accepted: November 4, 2026

Published online: November 27, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Accurate preoperative planning for total hip arthroplasty (THA) requires precise identification of proximal femoral anatomical landmarks, which remains observer-dependent. Although machine learning software can automate landmark detection, comparisons with clinician variability are scarce. This study evaluated whether automated landmark extraction on computed tomography (CT) scans falls within the range of human interobserver variability. Forty-three CT scans from patients undergoing preoperative planning for primary THA were analyzed. Three-dimensional femoral models were reconstructed in the CHIDOS software, which automatically segmented bone structures and extracted anatomical landmarks. Automated measurements were compared with manual measurements independently performed by two orthopedic consultants. Agreement was assessed using Euclidean deviations, intraclass correlation coefficients (ICC), Bland–Altman analysis, and paired Wilcoxon signed-rank tests, supplemented by an exploratory post-hoc equivalence analysis using two one-sided t-tests (TOST). Mean Euclidean deviations for landmark localization ranged from 2.79 ± 1.47 mm to 5.01 ± 1.73 mm, and angular deviations ranged from 1.33 ± 0.97° to 2.38 ± 1.75°. Agreement for angular measurements was good to excellent (ICC: 0.89–0.98). Artificial intelligence (AI)–clinician deviations were lower for femoral anteversion ($p < 0.001$) and neck-shaft angle ($p = 0.004$) relative to Clinician 2, higher for femoral head center ($p < 0.001$) and lesser trochanter ($p = 0.008$), and lower for femoral neck isthmus ($p = 0.006$) relative to Clinician 1. The 95% confidence intervals of the mean deviation relative to interobserver variability lay within literature-reported margins (±2° angular, ±3 mm linear) for all five endpoints. Post-hoc TOST analysis supported this observation ($p < 0.001$ for all endpoints), but did not establish clinical equivalence or non-inferiority. These preliminary findings

support further evaluation of AI-assisted landmark extraction as a reproducible first-pass step in CT-based THA planning workflows; effects on implant selection, operative performance, and clinical outcomes remain to be demonstrated.

Keywords: Hip arthroplasty; Bone segmentation; Landmark detection; Machine learning; Femoral anatomy; Three-dimensional preoperative planning

1. Introduction

Total hip arthroplasty (THA) is one of the most common and successful orthopedic surgical procedures, with more than one million surgeries performed annually worldwide.[1] As the demand for THA continues to rise due to an aging population and expanding indications, accurate preoperative planning has become increasingly important. Optimal outcomes depend on precise implant selection, component positioning, and restoration of native hip biomechanics. Misplacement of components and abnormal postoperative biomechanics may lead to complications including dislocation, leg length discrepancy, impingement, and aseptic loosening, ultimately increasing the risk of revision surgery. Accurate preoperative measurements of femoral and acetabular anatomy are therefore crucial to reduce operative time, costs, and the risk of selecting an inappropriate implant.

Accurate determination of the proximal femoral geometry, in particular the femoral anteversion angle and the neck-shaft angle, is a central element of this planning process. These parameters define the rotational and angular orientation of the femoral component relative to the native bone, and their misestimation has been associated with abnormal postoperative limb kinematics, reduced range of motion, impingement, and dislocation. Computed tomography (CT) provides a detailed three-dimensional (3D) representation of the proximal femur and pelvis, and CT-based assessment of femoral morphology has been described as a useful complement to conventional radiographic evaluation, particularly in anatomically complex cases.[2] Despite these advantages, the identification of the anatomical landmarks from which these parameters are derived remains a manual, observer-dependent step in most current planning workflows.

Conventional preoperative planning relies on two-dimensional (2D) templating using plain radiographs.[2] While widely used and accessible, 2D templating is associated with limited accuracy and substantial interobserver variability. A recent systematic review and meta-analysis comparing 2D and 3D planning in cementless hip arthroplasty demonstrated that 3D planning offers superior precision for both the cup (96.92% vs. 87.14%) and the stem (94.72% vs. 86.28%) compared with 2D planning.[3] These limitations have driven the development of CT-based 3D planning platforms.

Computed tomography-based 3D planning is emerging as a superior alternative for THA, supported by several commercial platforms that integrate image segmentation, anatomical landmark extraction, implant templating, and surgical simulation. By using these software packages, the correct prediction rates for femoral stem and acetabular cup sizes range between 34%–100% and 41%–100%, respectively.[4] Examples include mediCAD® 3D Hip (mediCAD Hectec GmbH), hipEOS (EOS Imaging), Mako Total Hip (Stryker Orthopaedics), MyHip,[5] OrthoGrid Hip AI (Zimmer Biomet), ZedHip (LEXI Co., Ltd.), and Hip-Op.[6] However, many of these software packages require technical expertise, and others require manual detection of the anatomical points of interest. A complete 3D preoperative planning aims to choose the correct implant while also calculating parameters such as range of motion and leg length correction.

The reliability of these planning workflows ultimately depends on the first step of the pipeline, namely the detection of the anatomical landmarks on which all downstream measurements are based. Landmark coordinates are used not only to calculate version angles and component orientation but also to drive the generation of patient-specific implant geometries and surgical guides, in which inaccuracies are directly transferred to the manufactured artifact. Reviews of hip implant design and of patient-specific instrumentation emphasize that the accuracy of the underlying anatomical model determines the clinical value of any personalization strategy.[1,4] Reproducible, automated landmark detection is therefore a prerequisite for patient-specific approaches that can be applied safely and at scale.

The integration of artificial intelligence (AI) into orthopedic surgery has accelerated in recent years, with applications spanning diagnostics, preoperative planning, intra-operative guidance, and postoperative care. A systematic review by Eskandar[7] evaluated the current

applications and limitations of AI in orthopedic surgery, finding that AI-based imaging tools demonstrated high diagnostic accuracy with sensitivities up to 98.2% and area under the curve (AUC) values exceeding 0.95 in fracture detection and musculoskeletal anomaly identification. In pre-surgical planning, AI-driven 3D modeling improved implant conformity (acetabular cup: 90.9% vs. 72.2%; femoral stem: 87.3% vs. 66.7%) and enhanced surgical risk prediction (AUC > 0.85 for complications).[7] A separate systematic review by Khaje Mozafari *et al.*[8] further highlighted the growing role of AI in preoperative planning of THA, while emphasizing the need for further research and clinical validation.

A critical prerequisite for accurate 3D THA planning is reliable segmentation of bone structures. Deep learning approaches have shown considerable promise in this domain. A clinical validation study of segmentation algorithms underlying an automatic surgery planning system for THA demonstrated high accuracy across subgroups including gender, age, CT manufacturer, and slice thickness.[9] A 3D multi-task network for automatic segmentation of CT images featuring hip osteoarthritis achieved high accuracy in segmenting peri-articular bone of the hip joint, enabling accurate reconstruction required for preoperative planning.[10] More recently, automatic multi-class segmentation and landmark-based osteotomy planning for hip arthroplasty have been developed, integrating segmentation and surgical planning into a single framework.[11] These results demonstrate that deep learning-based segmentation can significantly enhance accuracy and reproducibility in THA planning workflows.

Beyond segmentation, automated landmark detection has emerged as a critical area of research. Unstructured CT–High-Resolution Net was specifically designed to address challenges posed by irregular patient postures or occluded anatomical markers, integrating a spatial relationship fusion module and an uncertainty estimation module to achieve over 60% improvement across multiple metrics in unstructured data.[12] A two-stage multi-task deep learning framework for simultaneous pelvic bone segmentation and landmark detection from CT images has also been developed, improving accuracy especially for diseased cases.[13] For femoral version assessment, a fully automated method for femoral neck axis line measurement in an anatomical coordinate system has been validated against manual measurement, demonstrating automation, accuracy, and avoidance of subjective effects.[14] Schmaranzer *et al.*[15] developed a fully automated 3D measurement of femoral version using CT-based automated segmentation and compared it with standard 2D measurements in symptomatic patients, showing the feasibility of automated 3D-based assessment. Xiao *et al.*[16] proposed an automated method for femoral neck anteversion assessment from CT scans, which achieved excellent reproducibility (ICC up to 0.96) and strong agreement with manual techniques.

Clinical validation studies of AI-assisted planning systems have reported promising results. Chen *et al.*[17] validated an AI-assisted preoperative planning system for THA in implant selection and anatomical assessment, while Wu *et al.*[18] reported accuracy of AI-assisted 3D preoperative planning compared with conventional approaches. A 2025 study by Chan *et al.*[19] developed a deep learning model for automatic implant positioning in THA and calculated the mean root-square error between the model and four trained raters, finding values ranging from 3.19 mm to 5.37 mm for the identification of the greater and lesser trochanter. Another study utilizing the AI hip preoperative planning system in developmental dysplasia of the hip reported accuracy within ±0 size of 84.1% for the AI hip preoperative planning system versus 64.0% for 2D templating ($p < 0.05$), and within ±1 size of 95.1% versus 81.1% ($p < 0.05$).[20] Machine learning has also been applied to predict functional outcomes after THA. A systematic review by Clement *et al.*[21] assessed the reliability of machine learning techniques to predict functional outcomes of THA, identifying seven studies comprising 44,121 patients. The best-performing models were for health-related quality of life outcomes, with an AUC of more than 84%.

Despite these advances, several gaps remain in the literature. Rostamian *et al.*[22] presented a deep learning multi-view 3D landmark detection system for lower limbs with malalignment, yielding an average positional error of 2.05 mm and angular error under 1°. Berger *et al.*[23] found that artificial neural networks and statistical shape models achieved accuracy comparable to expert observers, while geometric approaches demonstrated higher error rates. Jang *et al.*[24] documented deviations of 3 to 5 mm in identifying the femoral head center by a deep learning algorithm. Automated pipelines have also been developed to measure linear and angular hip and knee parameters after CT, further supporting the feasibility of automated assessment.[25] However, most prior studies have focused primarily on implant prediction accuracy, segmentation performance, or clinical outcomes rather than direct variability compared with natural human observer variability. Moreover, there is no universal threshold defining an "acceptable" error for all THA planning parameters.

The present study was designed to address this gap by directly comparing the variability of an automated landmark detection system with the interobserver

variability of experienced clinicians under the same 3D conditions. Using a commercially available CT-based planning platform, we quantified the agreement between automated and manual measurements of the proximal femoral landmarks and angular parameters that are most relevant to femoral component positioning, and assessed whether the automated variability remained within the range observed between human observers.

Specifically, we investigated the agreement between AI-assisted automated proximal femoral landmark extraction using the CHIDOS software and manual CT-based measurements performed by orthopedic clinicians. We hypothesized that the software would demonstrate agreement comparable to human interobserver variability.

2. Methods

2.1. Study design and population

This retrospective study included CT datasets obtained from patients undergoing preoperative planning for THA (Figure 1). The study was conducted in accordance with the ethical standards of the Ethical Review Board of the University of Thessaly, Larissa (Acceptance Protocol Number: 407) and the principles of the Declaration of Helsinki, and all patients provided written informed consent for the use of anonymized imaging data for research purposes. Inclusion criteria consisted of patients undergoing primary THA with available preoperative CT imaging, while exclusion criteria comprised developmental hip dysplasia, post-traumatic arthritis, severe imaging artifacts, prior proximal femoral implants, and revision cases. The study population included patients with degenerative hip osteoarthritis undergoing elective THA planning who met the inclusion criteria consecutively between January 2025 and September 2025. Of 50 eligible scans screened, 43 m*et all* inclusion criteria and were included; excluded patients were those with developmental hip dysplasia (n = 4), post-traumatic arthritis (n = 2), and a prior proximal femoral implant (nail) (n = 1) (Figure 2). The patients' age range was 58–76 years. Therefore, the evaluated anatomy represents clinically relevant arthritis rather than healthy volunteer anatomy. Case selection, data anonymization, and data export were performed retrospectively from the institutional preoperative planning database, and all data were derived from imaging acquired for routine clinical care.

2.2. Imaging protocol

Preoperative lower-extremity CT scans were acquired using a multidetector CT scanner (Ingenuity CT; Philips Healthcare, The Netherlands). Imaging acquisition was performed using a standardized institutional protocol with a tube voltage of 120 kV, tube current-time product of 366 mAs, and slice thickness of 0.625 mm. Images were reconstructed with a reconstruction diameter of 500 mm using a standard bone reconstruction protocol for 3D anatomical assessment. The thin slice thickness was selected to support accurate 3D reconstruction and the precise localization of the anatomical landmarks on which the subsequent measurements depend.

2.3. CHIDOS software

The CHIDOS software v. 1.00.00 (Arthroniq, Greece) supports the uploading of CT scans, through which the pelvic and femoral anatomy are automatically reconstructed in 3D phantoms. A customized U-Net architecture first performs a segmentation to isolate the bone structures. The anatomical landmarks are also automatically extracted using a convolutional neural network and highlighted in the graphical user interface, allowing the surgeon to continue preoperative planning after the verification of the measurements. Subsequently, a dedicated landmark localization network identifies critical anatomical landmarks, establishing a robust 3D coordinate system. This software can also proceed to the automated design of

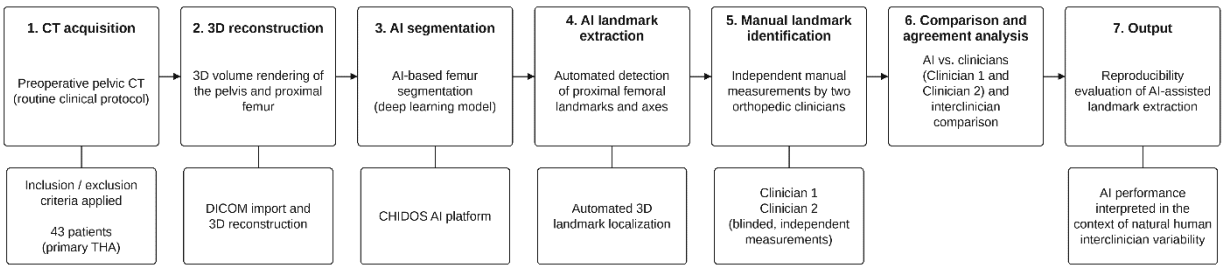

Figure 1. Flow diagram of study design
Abbreviations: 3D: Three-dimensional; AI: Artificial intelligence; CT: Computed tomography; THA: Total hip arthroplasty.

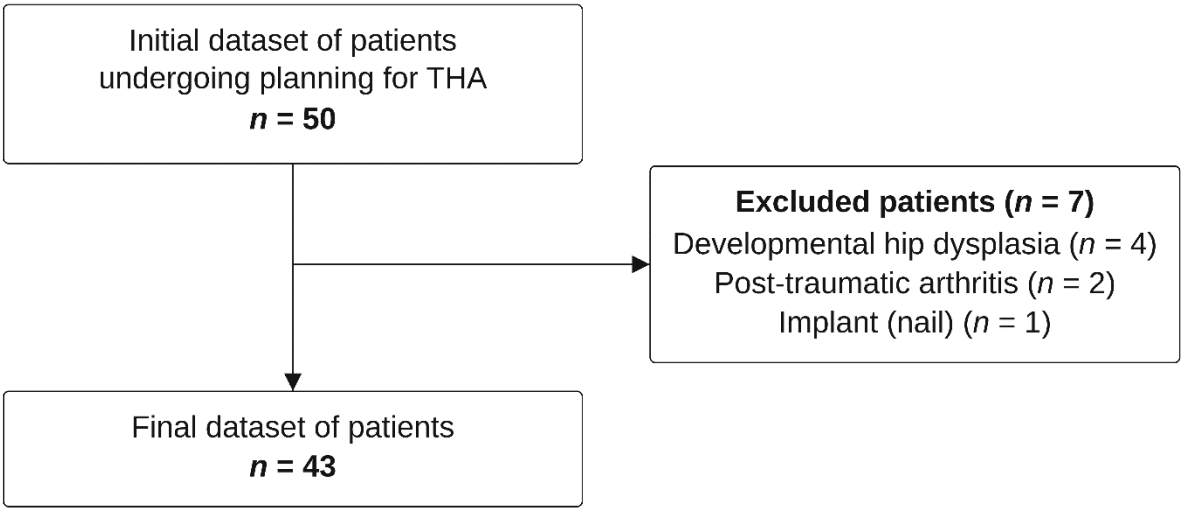

Figure 2. Flow diagram of patient dataset
Abbreviation: THA: Total hip arthroplasty.

a personalized femoral stem accompanied by customized surgical tools for implantation. Figure 3 depicts anatomical landmarks of an example case as they were detected by the CHIDOS software on the reconstructed 3D geometry of the proximal femur. Importantly, the CT datasets included in this study were not used during algorithm training or internal validation, and no model tuning occurred after exposure to this test cohort; therefore, they represent an independent test cohort for evaluation of the CHIDOS software platform. The evaluated model was locked before the evaluation: no retraining, fine-tuning, or parameter adjustment occurred after the test cohort was assembled. The development datasets comprised 400 CT scans from the New Mexico Decedent Image Database,[26] annotated by a research and development engineer with review by a senior engineer, using The Medical Imaging Interaction Toolkit (v. 2024.12).[27] Both the automated and the manual measurements were derived from the identical 3D reconstructions, so that the comparison was not confounded by differences in the underlying geometry. For each case, the complete automated workflow (segmentation, landmark extraction, and computation of the anatomical characteristics) was executed in its default form, without manual correction of the segmentation boundaries.

2.4. Anatomical characteristics of interest

The CT scans were uploaded to the CHIDOS software, and all computed values were collected in a dataset. The computed anatomical characteristics represent parameters that were taken into consideration during THA planning. The femoral anteversion angle and the neck-shaft angle were computed, while the coordinates in the 3D space for the femoral neck isthmus, the lesser trochanter and the femoral head center were also detected. The anatomical definitions used for both manual and automated measurements are summarized in Table 1. Manual measurements were independently performed in a randomized order by two fellowship-trained arthroplasty consultants. Both observers were blinded to the automated measurements and to each other's measurements. The manual measurements were performed between October 2025 and December 2025. 3D anatomical assessment and manual landmark localization were performed using the FreeCAD software (version 1.0.0; FreeCAD https://www.freecad.org) (Figure 4).

For completeness, the measured characteristics are defined as follows:

(i) The femoral head center was determined as the center of the largest sphere that fits within the native femoral head.
(ii) The femoral neck isthmus was defined as the center of the smallest cross-sectional slice perpendicular to the neck axis.
(iii) The lesser trochanter was localized at the midpoint of the prominence on the lesser trochanter surface.
(iv) The femoral anteversion angle was computed as the angle between the femoral neck axis and a line connecting the centers of the medial and lateral

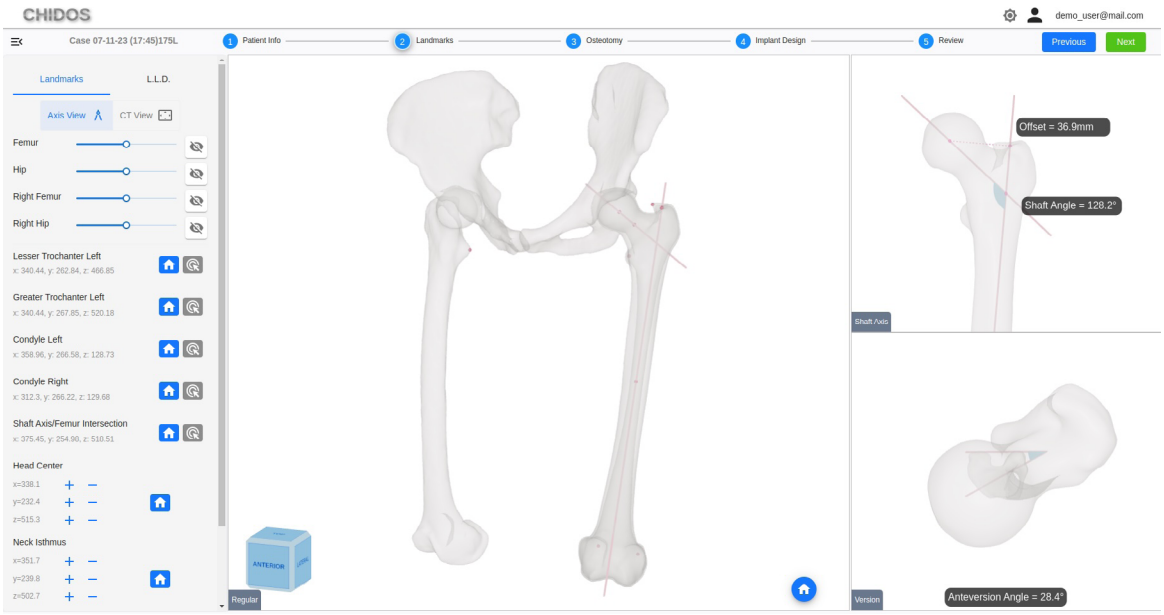

Figure 3. Three-dimensional geometry of the proximal femur with the automatically detected anatomical landmarks, shown from three different views

Table 1. Anatomical definitions of the measured landmarks

Landmark	Definition	Unit
Femoral head center	The center of the largest sphere that fits in the innate femoral head	mm
Femoral neck isthmus	The center of the smallest cross-sectional slice perpendicular to the neck axis	mm
Lesser trochanter	Point located at the midpoint of the prominence height on the lesser trochanter surface	mm
Femoral anteversion angle	The angle between the femoral neck axis and a line connecting the centers of the medial and lateral condyles, projected onto the XY plane	°
Neck-shaft angle	The angle formed between the femoral neck axis and the anatomical femoral shaft axis, projected onto the XZ plane	°

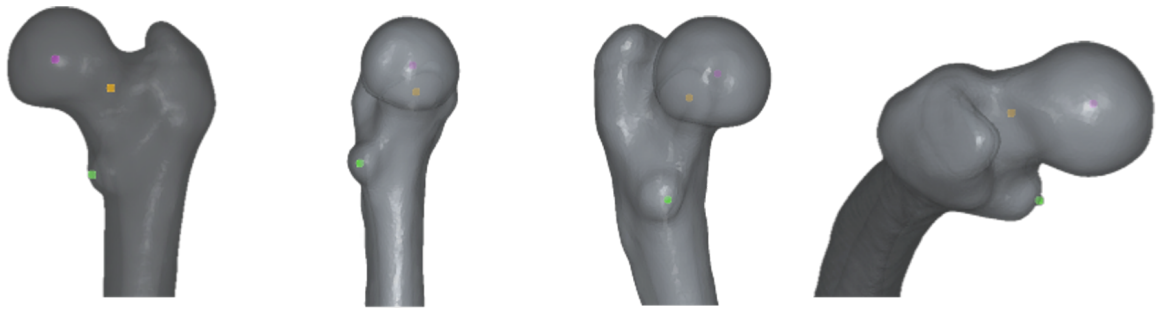

Figure 4. Manual landmark localization within the FreeCAD environment. Green represents the lesser trochanter, yellow represents the femoral neck isthmus, and pink represents the femoral head center.

condyles, projected onto the XY plane.

(v) The neck-shaft angle was computed as the angle between the neck axis and the anatomical femoral shaft axis, projected onto the XZ plane.

2.5. Statistical analysis

Agreement between software and manual measurements was assessed using mean absolute error, Bland–Altman analysis, and intraclass correlation coefficients (ICC) with corresponding 95% confidence intervals (CIs). ICC values were calculated using a two-way mixed-effects model for absolute agreement for angular measurements. In terms of 3D landmarks, it is methodologically appropriate to quantify landmark variability using Euclidean geometric distances derived from x, y, and z Cartesian coordinates. Three pairwise comparisons were performed for each endpoint: CHIDOS versus Clinician 1, CHIDOS versus Clinician 2, and Clinician 1 versus Clinician 2 (interobserver variability). The mean of the two CHIDOS–clinician deviations was additionally computed as a composite secondary metric. Continuous variables are presented as mean ± standard deviation. Normality of the deviation series and of the two one-sided *t*-tests (TOST) difference series was evaluated using the Shapiro–Wilk test. Paired Wilcoxon signed-rank tests (paired, repeated-measures design at the specimen level) were used for the pairwise comparisons of the per-specimen deviation series. As an exploratory post-hoc sensitivity analysis, the composite CHIDOS-clinician deviation was compared with the interobserver deviation using TOST at the 5% significance level, against literature-published margins of ± 2° for the angular endpoints, based on the interobserver reproducibility of 1.8° and 2.4° reported for CT-based femoral anteversion measurement,[28] and ±3 mm for the linear endpoints, based on the 3–5 mm landmark errors reported for previously published automated planning systems.[19,24] The observed mean interobserver variability of each endpoint was additionally used as a sensitivity margin. The sensitivity analysis was not prespecified. The margins were selected post-hoc; this analysis is exploratory and is not a definitive test of equivalence, non-inferiority, or clinical interchangeability. Bland–Altman analysis was performed for angular measurements, including femoral anteversion and neck-shaft angle, variables appropriate for agreement analysis. 3D anatomical landmarks were not evaluated using Bland–Altman methodology because landmark localization is represented by Cartesian spatial coordinates rather than scalar measurements. A p-value < 0.05 was considered statistically significant. All angular measurements were expressed in degrees, and linear measurements in millimeters. The analysis was conducted using RStudio (version 2023.03.0+386).

The choice of a two-way mixed-effects ICC model was based on the fact that both the measurement system and the individual observers were factors of interest, and that absolute agreement rather than mere consistency was the relevant criterion for a clinical decision-support tool. For 3D landmarks, the Euclidean distance was preferred over a per-axis comparison because the clinically relevant error is a vector quantity and the three coordinate deviations are strongly correlated. The number of specimens was determined by the availability of de-identified preoperative planning CT datasets meeting the inclusion criteria; the sample size was therefore not derived from an a priori power calculation. Instead, the minimum detectable mean difference between CHIDOS and interobserver variability at 80% power was computed for each endpoint (0.65° for femoral anteversion, 0.67° for the neck-shaft angle, 0.60 mm for the femoral head center, 0.66 mm for the lesser trochanter, and 0.86 mm for the femoral neck isthmus), all below the published margins used in the exploratory sensitivity analysis. This post-hoc observation does not establish adequate power for equivalence. The decision to restrict the Bland–Altman analysis to angular measurements reflects the fact that the method presumes a single scalar measurement for each specimen, which does not correspond to the three spatial coordinates of a landmark.

3. Results

Forty-three CT scans from patients undergoing preoperative planning for THA were included in the analysis (Table 2). Automated landmark extraction using the CHIDOS platform was successfully performed in all cases. Agreement between software and clinician measurements was acceptable across all evaluated proximal femoral landmarks. Mean absolute deviations ranged from 2.79 ± 1.47 mm for lesser trochanter localization to 5.01 ± 1.73 mm for femoral neck isthmus, while angular deviations ranged from 1.33 ± 0.97° for femoral anteversion to 2.38 ± 1.75° for neck-shaft angle measurements (Table 3). The lowest variability was observed for femoral anteversion measurements, whereas femoral neck isthmus determination demonstrated the greatest variability.

The cohort comprised 43 patients (43 hips) with a mean age of 66.7 ± 4.6 years (range: 58–76 years). Twenty-two patients were female, and 21 were male. The operated side was left in 20 cases and right in 23 cases (Table 2). All automated extractions were completed successfully, and no case required exclusion from the analysis. For the 3D landmarks, the mean deviations of the automated measurements from each clinician and the corresponding interobserver differences are reported in Table 3. For the femoral head center, the deviations were 4.01 ± 1.38 mm

(Clinician 1) and 2.96 ± 1.08 mm (Clinician 2), compared with an interobserver difference of 2.85 ± 1.25 mm. For the femoral neck isthmus, the corresponding values were 3.57 ± 1.80 mm, 5.01 ± 1.73 mm, and 4.73 ± 2.04 mm, and for the lesser trochanter, 3.78 ± 2.57 mm, 2.79 ± 1.47 mm, and 2.98 ± 1.75 mm. For the angular measurements, the deviations were 2.35 ± 1.40° and 1.33 ± 0.97° for femoral anteversion (interobserver: 2.72 ± 1.87°) and 2.38 ± 1.75° and 1.80 ± 1.14° for the neck-shaft angle (interobserver: 2.77 ± 1.97°). Among the reported means, the greatest deviation was 5.01 mm for the femoral neck isthmus relative to Clinician 2, and the smallest was 1.33° for femoral anteversion relative to Clinician 2. For the lesser trochanter, the standard deviation of the deviation relative to Clinician 1 (2.57 mm) exceeded that relative to Clinician 2 (1.47 mm), indicating a case-dependent pattern of agreement for this landmark. The pairwise comparisons of the deviation series (paired Wilcoxon signed-rank tests) showed that CHIDOS deviations differed between the two clinicians for the femoral head center ($p < 0.001$), the femoral anteversion ($p < 0.001$), the femoral neck isthmus ($p = 0.0005$), and the lesser trochanter ($p = 0.006$), but not for the neck-shaft angle ($p = 0.122$). Compared with the interobserver deviation, deviations were lower for femoral anteversion ($p < 0.001$) and for the neck-shaft angle relative to Clinician 2 ($p = 0.004$). Relative to Clinician 1, they were higher for the femoral head center ($p < 0.001$) and for the lesser trochanter ($p = 0.008$), whereas they were lower for the femoral neck isthmus ($p = 0.006$). The asymmetric agreement with the two clinicians reflects the absence of an adjudicated consensus reference: each clinician's manual localization contains its own observer-specific error, and a third measurement system is expected to deviate differently from each observer (Figure 5).

Table 2. Patient demographics

Variable	Value
Number of hips	43
Number of patients	43
Age, years (mean ± standard deviation)	66.7 ± 4.6
Female	22
Male	21
Side	20 left; 23 right

The mean CHIDOS–clinician deviation relative to interobserver deviations were 1.84° ± 0.96° and 2.72° ± 1.87° for femoral anteversion, 2.09° ± 0.95° and 2.77° ± 1.97° for the neck–shaft angle, 3.48 ± 1.01 mm and 2.85 ± 1.25 mm for the femoral head center, 4.29 ± 1.31 mm and 4.73 ± 2.04 mm for the femoral neck isthmus, and 3.28 ± 1.76 mm and 2.98 ± 1.75 mm for the lesser trochanter, respectively (Table 4). The 95% confidence intervals for the mean paired differences between the two deviations were within the equivalence margins of ±2° for angular measurements and ±3 mm for linear measurements. In an exploratory post hoc sensitivity analysis, the TOST yielded $p < 0.001$ for each measurement.

These results are exploratory and do not establish definitive non-inferiority or clinical interchangeability. Normality of the difference series was accepted for four of five endpoints by the Shapiro–Wilk test (all $p > 0.05$); for the lesser trochanter ($W = 0.858$, $p < 0.001$), a non-parametric bootstrap (95% CI: −0.10 to +0.77 mm; B = 2,000) was additionally reported. Agreement analysis demonstrated high ICCs for angular measurements (0.89–0.98), while Bland–Altman analysis showed small mean biases but wide limits of agreement in several angular comparisons (Table 5); the mean deviations for femoral neck isthmus and lesser trochanter localization were among the largest observed. High ICCs should therefore not be equated with clinical interchangeability. Bland–Altman plots are provided in Figures S1 and S2. The minimum detectable mean difference between CHIDOS and interobserver variability at 80% power was 0.65° for femoral anteversion, 0.67° for the neck-shaft angle, 0.60 mm for the femoral head center, 0.66 mm for the lesser trochanter, and 0.86 mm for the femoral neck isthmus; all values were descriptively below the literature-anchored margins used in the exploratory sensitivity analysis. These post-hoc calculations do not establish adequate power for equivalence; the findings should be interpreted through the observed confidence intervals. CHIDOS–clinician deviations compared with clinician interobserver variability: exploratory post-hoc equivalence analysis (TOST) and minimum detectable mean differences at 80% power are provided in Table S1.

The landmark-specific agreement statistics are summarized in Table 5. For femoral anteversion, the ICCs were 0.94 (95% CI: 0.76–0.98) for the software versus Clinician 1, 0.98 (95% CI: 0.96–0.99) for the software versus Clinician 2, and 0.90 (95% CI: 0.69–0.96) between the two clinicians. For the neck-shaft angle, the corresponding values were 0.92 (95% CI: 0.86–0.96), 0.96 (95% CI: 0.92–0.98), and 0.89 (95% CI: 0.80–0.94). Bland–Altman analysis demonstrated no large systematic bias: the mean bias ranged from −1.92° (between the two clinicians for femoral anteversion) to +1.68° (software versus Clinician 1 for femoral anteversion), and the width of the limits of agreement ranged from 6.43° to 13.41° across the six comparisons.

Table 3. Deviations between CHIDOS and manual measurements

Landmark	CHIDOS vs. Clinician 1 (mean deviation ± SD)	CHIDOS vs. Clinician 2 (mean deviation ± SD)	Interobserver error (Clinician 1 vs. Clinician 2)
Femoral head center (mm)	4.01 ± 1.38	2.96 ± 1.08	2.85 ± 1.25
Femoral neck isthmus (mm)	3.57 ± 1.80	5.01 ± 1.73	4.73 ± 2.04
Lesser trochanter (mm)	3.78 ± 2.57	2.79 ± 1.47	2.98 ± 1.75
Femoral anteversion (°)	2.35 ± 1.40	1.33 ± 0.97	2.72 ± 1.87
Neck-shaft angle (°)	2.38 ± 1.75	1.80 ± 1.14	2.77 ± 1.97

Note: For the femoral head center, $n = 42$ (one case lacked complete manual head-center localization); for all other endpoints, $n = 43$.
Abbreviation: SD: Standard deviation.

Table 4. Average artificial intelligence–clinician variability *vs.* clinician interobserver variability

Measurement	Artificial intelligence–clinician variability	Clinician interobserver variability	*p*-value of one-sided *t*-tests
Femoral anteversion (°)	1.84 ± 0.96	2.72 ± 1.87	<0.001
Neck-shaft angle (°)	2.09 ± 0.95	2.77 ± 1.97	<0.001
Femoral head center (mm)	3.48 ± 1.01	2.85 ± 1.25	<0.001
Lesser trochanter (mm)	3.28 ± 1.76	2.98 ± 1.75	<0.001
Femoral neck isthmus (mm)	4.29 ± 1.31	4.73 ± 2.04	<0.001

Note: Values are presented as mean absolute differences ± standard deviation. TOST *p*-values from the exploratory post-hoc sensitivity analysis described in Methods 2.5 (margins ±2° angular, ±3 mm linear).

Table 5. Agreement analysis of angular measurements between automated and manual assessment

Angle	Comparison	ICC 95% CI	Bland–Altman bias (95% limits of agreement)
Femoral anteversion (°)	AI vs. Clinician 1	0.94 (0.76–0.98)	1.68 (−2.59 to 5.95)
Femoral anteversion (°)	AI vs. Clinician 2	0.98 (0.96–0.99)	−0.24 (−3.46 to 2.97)
Femoral anteversion (°)	Clinician 1 vs. Clinician 2	0.90 (0.69–0.96)	−1.92 (−7.23 to 3.38)
Neck-shaft angle (°)	AI vs. Clinician 1	0.92 (0.86–0.96)	−0.2 (−6.03 to 5.62)
Neck-shaft angle (°)	AI vs. Clinician 2	0.96 (0.92–0.98)	−0.36 (−4.5 to 3.79)
Neck-shaft angle (°)	Clinician 1 vs. Clinician 2	0.89 (0.8–0.94)	−0.15 (−6.86 to 6.55)

Abbreviations: AI: Artificial intelligence; CI: Confidence interval; ICC: Intraclass correlation coefficient.

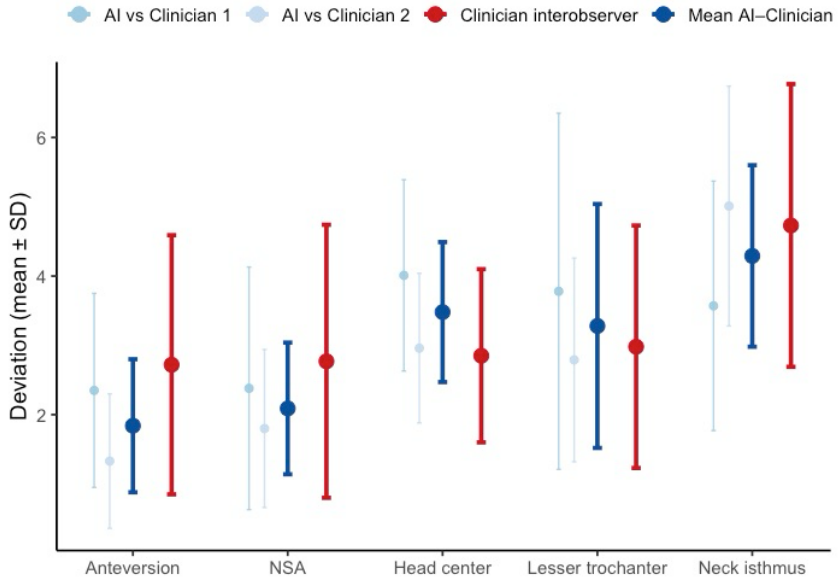

Figure 5. Comparison of AI–clinician to inter-clinician variability across angular and three-dimensional anatomical landmarks. The mean AI–clinician variability (dark blue) was calculated as the average absolute deviation between automated and clinician-derived measurements for each specimen. Individual AI versus Clinician 1 and AI versus Clinician 2 comparisons are shown in light blue, while clinician interobserver variability is shown in red. Error bars represent one standard deviation.
Abbreviations: AI: Artificial intelligence; NSA: Neck-shaft angle; SD: Standard deviation.

4. Discussion

Recent advances in AI and 3D imaging have accelerated the development of automated landmark detection systems for orthopedic applications, particularly in preoperative planning for THA. Rostamian *et al.*[22] presented a deep learning multi-view 3D landmark detection system for lower limbs with malalignment, yielding an average positional error of 2.05 mm and angular error under 1°, while in this study, the mean absolute angular deviations were 1.84° for anteversion and 2.09° for the neck-shaft angle. Xiao *et al.*[16] proposed an automated method for femoral neck anteversion assessment from CT scans, which achieved excellent reproducibility (ICC up to 0.96) and strong agreement with manual techniques. In a comparative evaluation of automated methods, Berger *et al.*[23] found that artificial neural networks and statistical shape models achieved accuracy comparable to expert observers, while geometric approaches demonstrated higher error rates. In 2025, Chan *et al.*[19] developed a deep learning model for automatic implant positioning in THA. They calculated the mean root-square error in measurements between the model and four trained raters. For the identification of the greater and lesser trochanter, they found values ranging from 3.19 mm to 5.37 mm, results in accordance with this study. Accordingly, Jang *et al.*[24] documented deviations of 3 to 5 mm in identifying the femoral head center by a deep learning algorithm. However, there is no universal threshold defining an "acceptable" error for all THA planning parameters. Interobserver reproducibility of 1.8° to 2.4° has been reported for CT-based femoral anteversion measurement,[28] and errors at the level of expert interobserver reproducibility are generally considered acceptable for version estimation, whereas the acceptability of a 3–5 mm error depends on the clinical task: it may be acceptable for preoperative planning support and landmark screening, but may be insufficient for applications requiring millimeter precision.

Most prior studies have focused primarily on implant prediction accuracy, segmentation performance, or clinical outcomes rather than direct variability compared to natural human observer variability. The present study evaluated the agreement of AI-assisted automated proximal femoral landmark measurements with independent manual 3D measurements. The principal finding was that CHIDOS–clinician agreement operated within the range of human interobserver variability for all five endpoints, with endpoint- and observer-dependent asymmetries. This finding is consistent with the more uniform application of geometric definitions and the elimination of observer-dependent landmark selection; it does not, however,

establish that automated angular measurements are more reproducible than manual assessment. The ICC values ranging from 0.89 to 0.98 indicate good to excellent agreement for the angular measurements, while landmark detection demonstrated measurement-dependent patterns. For the femoral head center, the CHIDOS deviation relative to Clinician 1 was modestly but significantly greater than the interobserver deviation (+1.16 mm; 95% CI: 0.72 to 1.60 mm), whereas the deviation relative to Clinician 2 was not significantly different ($p = 0.54$). Although statistically significant, the absolute difference between AI-related and clinician variability remained modest. The femoral head center is derived from 3D geometric fitting procedures and may be influenced by variations in femoral morphology, osteophytes, segmentation boundaries, and image quality. Therefore, small differences between automated and manual approaches may be expected even when overall agreement remains high. For the lesser trochanter and the femoral neck isthmus, the 95% CIs of the mean differences relative to interobserver variability (+0.31 mm, −0.16 to +0.77 mm; and −0.44 mm, −1.05 to +0.17 mm, respectively) were well within the ±3 mm margin; the exploratory TOST sensitivity analysis yielded concordant results for both endpoints. Greater variability in landmark localization was observed for the femoral neck isthmus and lesser trochanter, likely reflecting the greater anatomical complexity and less clearly defined geometric boundaries of these structures compared with the angular measurements. The authors interpret these findings as indicating that the variability of an AI-assisted software relative to human measurements is anatomy-dependent in terms of 3D landmark location. The asymmetric agreement of CHIDOS with the two clinicians (**Table 3**) is expected when two independent observers serve as the reference without an adjudicated consensus target: each clinician's manual localization contains its own observer-specific error, and a third measurement system will generally deviate differently from each.

From a clinical perspective, the magnitude of the observed deviations is relevant beyond statistical significance. The mean AI–clinician deviation for femoral anteversion (1.84 ± 0.96°) was below the approximately 2° interobserver reproducibility reported for CT-based anteversion measurement,[28] and it was also smaller than the variability observed between two experienced clinicians (2.72 ± 1.87°). In other words, for the angular parameters, the mean software–clinician deviations were within the range observed between two experienced clinicians working independently. For the femoral head center, the mean deviation of 3.48 mm falls within the 3–5 mm range previously reported for deep learning-based identification of the joint center on radiographs[24] and is comparable to the 3.19–5.37 mm mean root-square errors reported for automated implant positioning.[19] Because the manual reference measurements in the present study were performed on 3D reconstructions, the comparison reflects the variability of expert 3D assessment rather than that of conventional 2D templating, which is associated with larger errors and greater interobserver variability.[2,3] In this context, deviations that remain within expert 3D interobserver ranges are consistent with the software serving as a preliminary first-pass measurement in planning workflows; any advantage over 2D techniques, and clinical utility, remain to be established.

Two further practical considerations follow. First, the automated pipeline provides a consistent first-pass measurement for every case, which standardizes the starting point of the planning process. Second, whether this translates into reduced verification time or workflow benefit is plausible but unproven and requires prospective assessment. For applications in which millimeter precision is required, such as the design of patient-specific instrumentation,[4] the landmark-specific differences reported here should be taken into account, particularly for the femoral neck isthmus and the femoral head center.

The interpretation of the agreement statistics also benefits from the established intraclass correlation conventions. Under commonly used interpretation criteria, intraclass correlation values of 0.90 or higher indicate excellent agreement and values between 0.75 and 0.89 indicate good agreement. In the present study, five of the six angular comparisons reached or exceeded 0.90, and the single comparison below that threshold (0.89 between the two clinicians for the neck-shaft angle) remains in the good-to-excellent range. This pattern is consistent with the observation that angular parameters derived from well-defined axes are less susceptible to observer-dependent landmark selection than the 3D points that define the lesser trochanter and the femoral neck isthmus, whose anatomical boundaries are less distinct.

There are four main limitations in this study. First, the study cohort was limited to patients undergoing primary THA planning and excluded developmental dysplasia, post-traumatic deformity, and revision arthroplasty cases. Therefore, algorithmic performance in severely altered femoral anatomies remains uncertain and should be investigated in future studies. The cohort comprised 43 consecutive hips, a number determined by the availability of de-identified preoperative planning CT datasets meeting the inclusion criteria rather than by an a priori power calculation. The minimum detectable mean differences between CHIDOS and interobserver variability at 80% power were 0.65° for femoral anteversion, 0.67° for the neck–shaft angle, and 0.60–0.86 mm for the linear

endpoints. A larger multicenter cohort would further narrow the confidence intervals. Furthermore, the present study focused specifically on landmark extraction agreement and did not evaluate downstream effects on implant selection, surgical execution, or postoperative clinical outcomes. Finally, the human comparator comprised two independent clinicians without repeated measurements or an adjudicated consensus reference. The study therefore estimates agreement with expert two-observer practice rather than accuracy against a ground-truth reference, and the asymmetric agreement of CHIDOS with the two clinicians should be interpreted in that light.

Additional limitations should be considered. The study was conducted at a single center, and all CT acquisitions were performed on a single scanner model using a standardized protocol with a slice thickness of 0.625 mm and a bone reconstruction algorithm. The generalizability of the observed performance to other scanner vendors, slice thicknesses, and reconstruction protocols has not been established, and prior validation work on CT-based segmentation has suggested that image acquisition parameters can influence segmentation accuracy.[9] Second, the manual reference measurements were performed on 3D models in a dedicated visualization environment, which represents a more favorable reference condition than the 2D radiographs used in routine practice; the software's performance relative to a conventional 2D workflow therefore remains to be quantified in future studies.

Future studies should validate the approach in larger, multicenter cohorts that include the anatomically complex populations excluded from the present study, such as developmental dysplasia of the hip, post-traumatic deformity, and revision cases, where both the planning difficulty and the potential for clinical benefit of automated measurement may be greater. External validation on CT data acquired on different scanner platforms and reconstruction protocols is a prerequisite before the method can be considered robust across institutions. Equally important is the assessment of downstream effects: whether consistency in landmark identification translates into improved implant sizing accuracy, shorter operative planning times, and ultimately better clinical outcomes remains to be demonstrated in prospective studies. Finally, the integration of automated landmark detection into the wider ecosystem of surgical navigation, robotic-assisted arthroplasty, and patient-specific instrumentation warrants investigation, as their performance depends on the accuracy of the underlying anatomical landmarks.[4] Because the CHIDOS platform is deployed as a cloud-based service, prospective use of the pipeline in routine clinical planning would also allow the quantification of workflow-level effects, including the time required from CT upload to completed landmark verification, in real-world conditions.

5. Conclusion

In this selected cohort, AI-assisted proximal femoral landmark extraction showed preliminary technical agreement with clinician measurements, with mean deviations within the range of interobserver variability and high ICCs for angular measurements. However, Bland–Altman limits of agreement were wide for the angular endpoints, and this study does not establish clinical equivalence, superiority, accuracy against an independent reference standard, generalizability, or clinical utility. These findings support further evaluation of AI-assisted landmark extraction as a reproducible first-pass step in CT-based THA planning workflows; effects on implant selection, operative performance, and clinical outcomes remain unproven.

Acknowledgments

None.

Funding

None.

Conflict of interest

Athanasios Batagiannis is the owner and CEO of Arthroniq SMPC, the company that develops CHIDOS, the commercial software used in this study. All other authors declare no competing interests.

Author contributions

Conceptualization: Styliani Stergiadou, Antonios Koutalos, Konstantinos Malizos, Aristeidis Zibis, Athanasios Batagiannis
Data curation: Styliani Stergiadou
Formal analysis: Styliani Stergiadou, Michael Klontzas
Investigation: Styliani Stergiadou, Antonios Koutalos, Konstantinos Malizos
Methodology: Styliani Stergiadou, Athanasios Batagiannis, Michael Klontzas
Project administration: Aristeidis Zibis
Resources: Aristeidis Zibis, Michael Klontzas, Grégoire Micicoi, Athanasios Batagiannis
Software: Athanasios Batagiannis
Supervision: Aristeidis Zibis, Grégoire Micicoi, Konstantinos Malizos
Validation: Styliani Stergiadou, Antonios Koutalos, Konstantinos Malizos
Visualization: Styliani Stergiadou, Athanasios Batagiannis

Writing–original draft: Styliani Stergiadou
Writing–review & editing: All authors

Ethics approval and consent to participate

This retrospective study was conducted in accordance with the Declaration of Helsinki and approved by the Ethical Review Board of the University of Thessaly, Larissa (Acceptance Protocol Number: 407; [28.03.2023]). All patients provided written informed consent for the use of anonymized imaging data for research purposes.

Consent for publication

Written informed consent was obtained from all patients for the publication of their anonymized data and images. All identifying information was removed prior to analysis and publication.

Availability of data

The datasets generated and analyzed during the current study are not publicly available due to patient privacy regulations but are available from the corresponding author on reasonable request.

AI tools statement

ChatGPT (OpenAI Codex, based on GPT-6) was used solely for English-language editing and grammatical improvement. All scientific content, analyses, interpretations, and conclusions were developed and verified by the authors, who take full responsibility for the final manuscript.

References

1. Guo L, Naghavi SA, Wang Z, *et al*. On the design evolution of hip implants: a review. *Mater Des*. 2022;216:110552.

   doi: 10.1016/j.matdes.2022.110552

2. Huppertz A, Radmer S, Wagner M, Roessler T, Hamm B, Sparmann M. Computed tomography for preoperative planning in total hip arthroplasty: what radiologists need to know. *Skeletal Radiol*. 2014;43(8):1041-1051.

   doi: 10.1007/s00256-014-1853-2

3. Parisi FR, Zampogna B, Zampoli A, *et al*. Planning accuracy and stem offset assessment in digital two-dimensional versus three-dimensional planning in cementless hip arthroplasty: a systematic review and meta-analysis. *J Clin Med*. 2024;13(21):6566.

   doi: 10.3390/jcm13216566

4. Moralidou M, Di Laura A, Henckel J, Hothi H, Hart AJ. Accuracy of a three-dimensional (3D)-printed patient-specific (PS) femoral osteotomy guide: a computed tomography (CT) study. *Bioengineering*. 2022;9(11):667.

   doi: 10.3390/bioengineering9110667

5. Schmid J, Chênes C, Chagué S, *et al*. MyHip: supporting planning and surgical guidance for a better total hip arthroplasty. *Int J Comput Assist Radiol Surg*. 2015;10(10):1547-1556.

   doi: 10.1007/s11548-015-1177-2

6. Lattanzi R, Viceconti M, Zannoni C, Quadrani P, Toni A. Hip-Op: an innovative software to plan total hip replacement surgery. *Med Inform Internet Med*. 2002;27(2):71-83.

   doi: 10.1080/14639230210150346

7. Eskandar K. The role of artificial intelligence in orthopedic surgery: current applications and future perspectives—a systematic review of the literature. *Rev Esp Cir Ortop Traumatol*. 2026;70(3):248-257.

   doi: 10.1016/j.recot.2025.09.005

8. Khaje Mozafari J, Moshtaghioon SA, Mahdavi SM, Ghaznavi A, Behjat M, Yeganeh A. The role of artificial intelligence in preoperative planning for total hip arthroplasty: a systematic review. *Front Artif Intell*. 2024;7:1417729.

   doi: 10.3389/frai.2024.1417729

9. Wang D. Clinical validation of the segmentation algorithms underlying an automatic surgery planning system for total hip arthroplasty. Conference abstract. European Congress of Radiology; 2025.

   doi: 10.26044/ECR2025/C-16661

10. Wang H, Zhang X, Li S, *et al*. A 3D multi-task network for the automatic segmentation of CT images featuring hip osteoarthritis. *Biomed Phys Eng Express*. 2025;11(5):055042.

    doi: 10.1088/2057-1976/ae0593

11. Spinciu IT, Duşa S, Petean CE, Sultana AE, Cristea DD, Liciu E. Automatic multi-class segmentation and landmark-based osteotomy planning for hip arthroplasty. In: *Proceedings of the 2025 IEEE 21st International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE; 2025:1-9.

    doi: 10.1109/ICCP68926.2025.11427145

12. Wan J, Liu L, Wang H, *et al*. UNSCT-HRNet: modeling anatomical uncertainty for landmark detection in total hip arthroplasty. Preprint. arXiv; 2024.

    doi: 10.48550/arXiv.2411.08488

13. Zhai H, Chen Z, Li L, *et al*. Two-stage multi-task deep learning framework for simultaneous pelvic bone segmentation and landmark detection from CT images. *Int J Comput Assist Radiol Surg*. 2024;19(1):97-108.

    doi: 10.1007/s11548-023-02976-1

14. Li L, Ye Q, Hu T, Fang D, Wan R, Shi Y. Fully automated measurement of femoral neck axis line in anatomical coordinate system using CT images. *Med Biol Eng Comput*.

2024;62(6):1837-1849.

doi: 10.1007/s11517-024-03037-8

15. Schmaranzer F, Movahhedi M, Singh M, *et al*. Computed tomography-based automated 3D measurement of femoral version: validation against standard 2D measurements in symptomatic patients. *J Orthop Res*. 2024;42(10):2237-2248.

doi: 10.1002/jor.25865

16. Xiao H, Yibulayimu S, Zhao C, *et al*. Rethinking femoral neck anteversion assessment: a novel automated 3D CT method compared to traditional manual techniques. *BMC Musculoskelet Disord*. 2025;26(1):466.

doi: 10.1186/s12891-025-08697-7

17. Chen X, Liu X, Wang Y, *et al*. Development and validation of an artificial intelligence preoperative planning system for total hip arthroplasty. *Front Med*. 2022;9:841202.

doi: 10.3389/fmed.2022.841202

18. Wu L, Zhao X, Lu Z, Yang Y, Ma L, Li P. Accuracy analysis of artificial intelligence-assisted three-dimensional preoperative planning in total hip replacement. *Jt Dis Relat Surg*. 2023;34(3):537-547.

doi: 10.52312/jdrs.2023.1059

19. Chan PYW, Baker CE, Suh Y, Moyer D, Martin JR. Development of a deep learning model for automating implant position in total hip arthroplasty. *J Arthroplasty*. 2025;40(8):2092-2100.

doi: 10.1016/j.arth.2025.01.032

20. Xie H, Yi J, Huang Y, *et al*. Application and evaluation of artificial intelligence 3D preoperative planning software in developmental dysplasia of the hip. *J Orthop Surg Res*. 2024;19(1):176.

doi: 10.1186/s13018-024-04588-0

21. Clement ND, Clement R, Clement A. Predicting functional outcomes of total hip arthroplasty using machine learning: a systematic review. *J Clin Med*. 2024;13(2):603.

doi: 10.3390/jcm13020603

22. Rostamian R, Shariat Panahi M, Karimpour M, Kashani HG, Abi A. A deep learning-based multi-view approach to automatic 3D landmarking and deformity assessment of lower limb. *Sci Rep*. 2025;15(1):534.

doi: 10.1038/s41598-024-84387-z

23. Berger L, Brößner P, Ehreiser S, Tokunaga K, Okamoto M, Radermacher K. Validation and comparison of three different methods for automated identification of distal femoral landmarks in 3D. *Biomed Tech*. 2025;70(5):425-431.

doi: 10.1515/bmt-2025-0026

24. Jang SJ, Kunze KN, Vigdorchik JM, Jerabek SA, Mayman DJ, Sculco PK. John Charnley Award: deep learning prediction of hip joint center on standard pelvis radiographs. *J Arthroplasty*. 2022;37(7):S400-S407.e1.

doi: 10.1016/j.arth.2022.03.033

25. Vidhani FR, Woo JJ, Zhang YB, Olsen RJ, Ramkumar PN. Automating linear and angular measurements for the hip and knee after computed tomography: validation of a three-stage deep learning and computer vision-based pipeline for pathoanatomic assessment. *Arthroplasty Today*. 2024;27:101394.

doi: 10.1016/j.artd.2024.101394

26. Edgar HJH, Daneshvari Berry S, Moes E, Adolphi NL, Bridges P, Nolte KB. New Mexico Decedent Image Database. Dataset. Published online 2020.

doi: 10.25827/5s8c-n515

27. Wolf I, Vetter M, Wegner I, *et al*. The Medical Imaging Interaction Toolkit. *Med Image Anal*. 2005;9(6):594-604.

doi: 10.1016/j.media.2005.04.005

28. Lee YS, Oh SH, Seon JK, Song EK, Yoon TR. 3D femoral neck anteversion measurements based on the posterior femoral plane in ORTHODOC® system. *Med Biol Eng Comput*. 2006;44(10):895-906.

doi: 10.1007/s11517-006-0104-7