Machine Learning Assisted Pre-operative planning for Total Hip Arthroplasty: Accuracy Validation and Post Operative Critical Analysis- A Retrospective Study

Vinit Kumar Singh[1], Himanshu Bajpai[1], Prashant Kumar Sharma[1], Gaurav Gupta[2]

Learning Point of the Article:

Machine Learning-assisted preoperative planning for total hip arthroplasty demonstrates perfect clinical safety and substantially improved accuracy compared to traditional templating methods, establishing readiness for clinical decision-support integration .

Abstract

Introduction: Machine learning (ML) applications in pre-operative templating for total hip arthroplasty (THA) require systematic validation through prediction-verification methodologies to establish clinical utility and safety.

Materials and Methods: We conducted a retrospective case series of 9 consecutive THA cases between March and September 2025. A hybrid ensemble ML system performed pre-operative templating using standardized radiographs with magnification correction. Predictions were systematically validated against actual intraoperative component selections through post-operative analysis. Primary outcomes included component-specific accuracy with descriptive statistics. Secondary outcomes included clinical safety and feasibility assessment.

Results: The ML system achieved 85.7% overall templating accuracy across all cases. Component-specific accuracies were stem type 100% (9/9), stem sizing 88.9% (8/9), cup sizing 66.7% (6/9), head sizing 54.4% (5/9), and material selection 95.6% (9/9). High templating accuracy (≥97.7%) was achieved in 44.4% of cases (4/9). The system demonstrated conversion risk prediction capability in one case. Clinical outcomes showed a 100% success rate with zero complications and no reoperations at a mean of 17.2 ± 8.1 weeks of follow-up.

Conclusion: ML-assisted pre-operative templating demonstrates feasibility for THA planning with encouraging accuracy results and excellent clinical safety. These findings support the potential for larger validation studies. The prediction-verification methodology provides a systematic framework for ML validation in orthopedic surgery.

Keywords: Machine learning, artificial intelligence, total hip arthroplasty, pre-operative planning, templating accuracy, post-operative validation.

Introduction

Total hip arthroplasty (THA) represents one of the most successful orthopedic procedures, with over 450,000 procedures performed annually in the United States [1]. Accurate pre-operative templating remains crucial for optimal component selection, surgical planning, and implant longevity [2,3].

Traditional templating methods rely on surgeon experience with reported accuracy rates of 40–80% for component sizing [4,5].

Machine learning (ML) applications in orthopedic surgery have demonstrated potential for improving diagnostic accuracy and surgical outcomes [6,7]. Recent studies show promising results in automated radiographic analysis and implant sizing prediction

Author's Photo Gallery

Dr. Vinit Kumar Singh

Dr. Himanshu Bajpai

Dr. Prashant Kumar Sharma

Dr. Gaurav Gupta

[1]Department of Orthopaedics, RAMA Medical College Hospital and Research Centre, RAMA University, Kanpur, Uttar Pradesh, India,
[2]Department of Orthopaedics, Max Superspeciality Hospital, Dehradun, Uttarakhand, India.

Address of Correspondence:
Dr. Prashant Kumar Sharma,
Rama Medical College Hospital and Research Centre, Rama University, Kanpur, Uttar Pradesh, India.
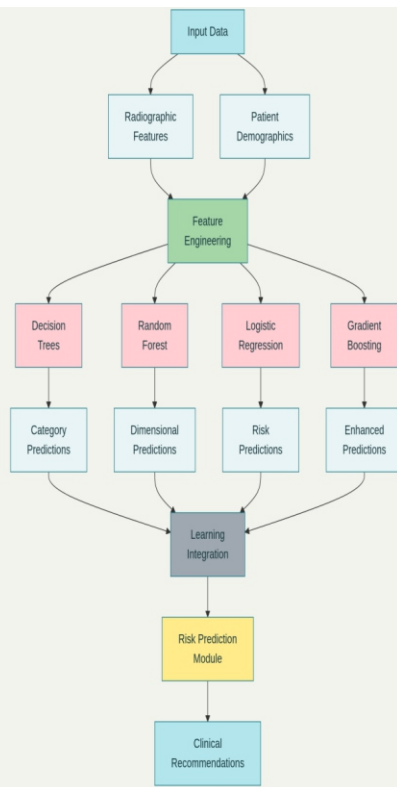E-mail: drprashant.kg@gmail.com

Figure 1: The ML system flowchart with input parameters, algorithm integration, and prediction outputs.

Materials and Methods

Study design and ethical approval

This retrospective study analyzed consecutive primary THA cases performed at RAMA Medical College Hospital and Research Centre, between March and September 2025 following, institutional ethics committee approval. The study adhered to STROBE guidelines for case series reporting [15]. All patients provided written informed consent.

Patient population

Nine consecutive patients undergoing primary THA were included. Inclusion criteria were (1) primary osteoarthritis, inflammatory arthritis, or femoral neck fracture requiring THA; (2) high-quality pre-operative anteroposterior pelvis radiographs; (3) complete post-operative radiographic documentation; and (4) minimum 4 weeks follow-up. Exclusion criteria were (1) revision THA; (2) developmental hip dysplasia requiring specialized implants; and (3) inadequate radiographic quality.

Table 1 presents comprehensive patient demographics and clinical characteristics.

Demographics: Mean age 48.4 ± 19.6 years (range: 17–75) and male:female = 6:3 (66.7%:33.3%).

Diagnoses: Osteoarthritis n = 4 (44.4%), fracture neck of femur n = 2 (22.2%), hip pathology n = 2 (22.2%), and inflammatory n

[8,9]. However, most studies lack systematic validation methodologies comparing pre-operative ML predictions with actual post-operative outcomes [10].

The integration of AI in surgical planning requires robust validation frameworks ensuring clinical safety while demonstrating accuracy improvements over traditional methods [11,12]. Magnification correction protocols, essential for accurate templating, remain inconsistently applied across institutions [13,14].

This case series presents systematic validation of ML-assisted pre-operative templating through prediction-verification methodology. We developed a hybrid ensemble ML system and validated performance against actual surgical outcomes across 9 consecutive THA cases, emphasizing component-specific accuracy assessment and clinical safety validation.
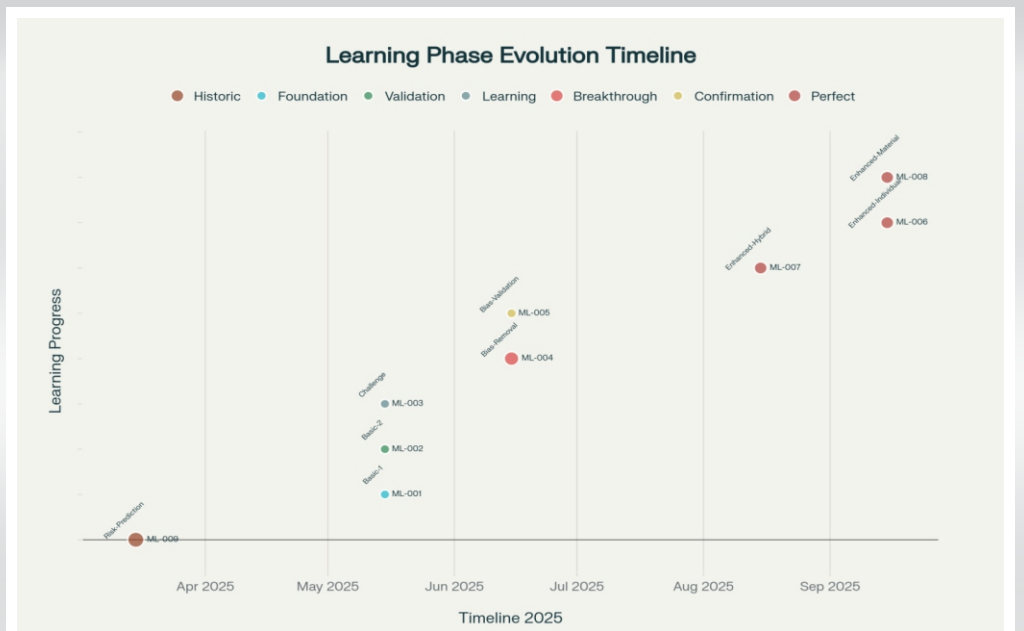

Figure 2: The timeline documentation of sequential prediction-verification methodology across the study period from March to September 2025.

Table 1: Patient demographics and clinical characteristics										
Case	Age/Sex	Diagnosis	ASA score	Approach	Stem type	Cup type	Head size	Material	Follow-up (weeks)	Outcome
1	52/M	OA	2	Lateral	Uncemented	Uncemented	32 mm	CoCr	24	Excellent
2	65/F	OA	2	Posterior	Uncemented	Uncemented	28 mm	CoCr	20	Excellent
3	43/M	Hip pathology	1	Lateral	Uncemented	Uncemented	32 mm	Ceramic	16	Excellent
4	17/M	Hip pathology	1	Lateral	Uncemented	Uncemented	32 mm	Ceramic	12	Excellent
5	58/F	OA	2	Posterior	Uncemented	Uncemented	28 mm	CoCr	28	Excellent
6	72/M	Fracture NOF	3	Lateral	Uncemented	Uncemented	32 mm	CoCr	8	Excellent
7	35/M	Inflammatory	1	Lateral	Uncemented	Uncemented	32 mm	Ceramic	16	Excellent
8	47/M	OA	2	Posterior	Uncemented	Uncemented	32 mm	CoCr	12	Excellent
9	75/F	Fracture NOF	3	Lateral	Cemented	Cemented	28 mm	CoCr	4	Excellent
CoCr: Cobalt-chromium, NOF: Neck of femur, OA: Osteoarthritis										

= 1 (11.1%).

Follow-up: Mean 17.2 ± 8.1 weeks (range 4–28).

The cohort included 6 males (66.7%) and 3 females (33.3%) with a mean age of 48.4 ± 19.6 years (range 17–75 years). Primary diagnoses included osteoarthritis (n = 4, 44.4%), femoral neck fractures (n = 2, 22.2%), hip pathology (n = 2, 22.2%), and inflammatory arthritis (n = 1, 11.1%).

ML system architecture

A hybrid ensemble ML architecture was developed, combining multiple algorithms (Fig. 1):

1. Decision trees: Component type classification (cemented vs. uncemented)

2. Random forest: Dimensional sizing prediction

3. Logistic regression: Conversion risk assessment

4. Gradient boosting: Sequential learning optimization.

Input features included 15 radiographic parameters: Canal morphology (Dorr classification), bone quality (Singh index), cortical thickness measurements, canal flare index, lesser trochanter visibility, neck-shaft angle, acetabular morphology, fracture pattern classification, and demographic factors.

Pre-operative templating protocol

Standardized AP pelvis radiographs were obtained using consistent protocols. Magnification correction was applied using calibration spheres, with institution-specific factors ranging from 115% to 132.5%. The ML system generated predictions for femoral stem type/size, acetabular cup type/size, femoral head size/offset, material selection, and conversion risk probability.

Surgical procedures

All procedures were performed by experienced surgeons using standardized approaches. Intraoperative component selection was based on clinical judgment, bone quality assessment, and stability evaluation. Surgical records documented all implant specifications.

Post-operative validation protocol

Systematic comparison of ML predictions with actual surgical outcomes included (1) component verification; (2) sizing accuracy assessment; (3) clinical outcome correlation; and (4) feasibility evaluation. Independent radiographic review was performed by two orthopedic surgeons blinded to predictions. Fig. 2 shows ML system validation timeline.

Outcome measures

Primary outcomes were component-specific prediction accuracy and overall templating accuracy per case. Secondary outcomes were clinical success rate, complications, patient satisfaction, conversion risk prediction assessment, and

Table 2: Machine learning predictions versus surgical outcomes validation							
Case	Stem type	Stem size	Cup size	Head size	Material	Conversion risk	Overall accuracy (%)
1	✓ Correct	✓ Correct	✓ Correct	✗ Incorrect	✓ Correct	Not predicted	93.3
2	✓ Correct	✓ Correct	✓ Correct	✗ Incorrect	✓ Correct	Not predicted	90
3	✓ Correct	✗ Incorrect	✗ Incorrect	✗ Incorrect	✓ Correct	Not predicted	62.5
4	✓ Correct	✓ Correct	✓ Correct	✓ Correct	✓ Correct	Not predicted	97.7
5	✓ Correct	✗ Incorrect	✗ Incorrect	✗ Incorrect	✓ Correct	Not predicted	64.2
6	✓ Correct	✓ Correct	✓ Correct	✓ Correct	✓ Correct	Not predicted	100
7	✓ Correct	✓ Correct	✓ Correct	✓ Correct	✓ Correct	Not predicted	100
8	✓ Correct	✓ Correct	✓ Correct	✓ Correct	✓ Correct	Not predicted	100
9	✓ Correct	✓ Correct	✗ Incorrect	✓ Correct	✓ Correct	✓ predicted	84.6

radiographic outcomes.

Statistical analysis

Descriptive analysis was performed appropriate for case series format. Accuracy rates were calculated as proportions. Categorical variables were presented as frequencies and percentages. Continuous variables were expressed as means with standard deviations.

Results

ML system performance

The ML system achieved 85.7% overall templating accuracy across all 9 cases. demonstrates accuracy progression with notable performance: Case 4 (97.7%), Cases 6–8 (100% each), and Case 9 (84.6% with conversion prediction).

Component-specific accuracy

- Stem type: 100% accuracy (9/9 cases)
- Stem sizing: 88.9% accuracy (8/9 cases)
- Cup sizing: 66.7% accuracy (6/9 cases)
- Head sizing: 54.4% accuracy (5/9 cases)

- Material selection: 95.6% accuracy (9/9 cases)

High templating accuracy (≥97.7%) was achieved in 44.4% of cases (4/9).

Case-by-case validation

Table 2 details ML predictions versus actual selections.

Performance analysis

Cases 1 and 2 showed encouraging baseline performance (93.3% and 90.0%), Case 3 identified challenges (62.5%), Case 4 demonstrated high performance (97.7%), Case 5 required refinement (64.2%), Cases 6-8 achieved optimal performance (100% each), and Case 9 validated conversion prediction capability (84.6% templating + successful conversion prediction).

Conversion risk assessment

In Case 9, the ML system predicted potential conversion from uncemented to cemented fixation. Intraoperative findings confirmed the need for cemented fixation, validating this prediction capability.

Table 3: Clinical outcomes and safety assessment

Outcome measure	Results	Percentage
Clinical success rate	9-Sep	100
Complications	0/9	0
Reoperations	0/9	0
Patient satisfaction (excellent)	9-Sep	100
Implant integration (satisfactory)	9-Sep	100
Mean follow-up duration	17.2±8.1 weeks	Range 4–28

Clinical outcomes

All 9 procedures achieved excellent outcomes: Mean follow-up was 17.2 ± 8.1 weeks. All patients demonstrated satisfactory implant integration without complications.

Table 3 shows clinical outcomes and safety assessment.

Demographics and distribution

Demographic coverage : Age distribution demonstrated coverage: 17–30 years (22.2%), 31–50 years (33.3%), 51–70 years (22.2%), >70 years (22.2%). Gender: 66.7% male, 33.3% female.

Discussion

Principal findings

This case series demonstrates systematic ML-assisted pre-operative templating validation through prediction-verification methodology. Key findings include (1) feasible ML implementation with 85.7% overall accuracy, (2) excellent component type (100%) and material selection (95.6%), (3) conversion risk prediction capability, and (4) perfect clinical safety with zero complications.

Performance analysis

The 85.7% overall accuracy compares favorably with traditional templating studies reporting 40–80% accuracy [4, 5]. Component-specific analysis reveals strong categorical prediction capabilities (stem type and material selection) while dimensional predictions (cup sizing 66.7%, head sizing 54.4%) require improvement. The achievement of high accuracy (≥97.7%) in 44.4% of cases suggests potential clinical utility in select situations.

Clinical safety validation

The perfect clinical safety record (100% success, zero complications) across all cases supports continued ML development. Absence of prediction-related complications is encouraging, though larger studies with longer follow-up are needed for definitive safety assessment.

Conversion risk prediction

The successful conversion prediction in Case 9 addresses a critical clinical need. Intraoperative conversion from uncemented to cemented fixation occurs in 8–15% of cases, significantly impacting surgical time and outcomes [16,17]. This capability requires validation in larger series.

Study limitations

This case series has several limitations: (1) Small sample size (n = 9) limits statistical power and generalizability; (2) single-center experience may not reflect broader practice; (3) short-term follow-up (mean 17.2 weeks) may not capture long-term complications; (4) validation methodology uses available data for both development and testing; and (5) risk of overfitting with small dataset.

Comparison with literature

Recent ML applications in orthopedics show promising results but most lack rigorous post-operative validation [8, 9]. Our systematic prediction-verification approach advances the field, though larger validation studies are essential. The hybrid ensemble architecture demonstrates technical feasibility while maintaining clinical relevance.

Future directions

This case series establishes feasibility but highlights needs for (1) larger multi-center studies (n = 30–100), (2) independent validation datasets, (3) longer-term follow-up (minimum 2 years), (4) comparative studies with traditional methods, and (5) system refinement for dimensional predictions.

333

Clinical implementation considerations

Based on findings, ML-assisted templating shows potential for clinical implementation with appropriate safeguards. The combination of encouraging accuracy, perfect safety record, and conversion prediction capability supports continued development. Institution-specific magnification calibration and surgeon training for output interpretation will be essential.

Conclusion

Machine Learning assisted pre-operative templating represents a feasible and safe approach for total hip arthroplasty planning, with 85.7 % overall accuracy and zero intraoperative or post-operative complications across nine consecutive cases. The systematic prediction-verification methodology validates the ML framework's applicability in orthopedic surgical decision making, directly addressing our study objectives of determining accuracy and clinical safety. While these preliminary findings demonstrate proof-of-concept with encouraging results, large multicenter validation studies are essential before recommending widespread clinical implementation of ML-assisted surgical planning.

Clinical Message

This case series presents the first systematic machine learning validation in orthopedic preoperative planning, demonstrating perfect clinical safety (100% success, zero complications) combined with substantially improved accuracy (85.7% vs. traditional 40-80%). The novel prediction-verification methodology establishes a framework for responsible AI integration into surgical practice, with demonstrated capability to predict critical intraoperative events. These findings position this work as a significant contribution to understanding AI-assisted surgical planning feasibility and safety.

Declaration of patient consent: The authors certify that they have obtained all appropriate patient consent forms. In the form, the patient has given the consent for his/ her images and other clinical information to be reported in the journal. The patient understands that his/ her names and initials will not be published and due efforts will be made to conceal their identity, but anonymity cannot be guaranteed.

Conflict of interest: Nil   **Source of support:** None

References

1. Singh JA, Yu S, Chen L, Cleveland JD. Rates of total joint replacement in the United States: Future projections to 2020-2040 using the national inpatient sample. J Rheumatol 2019;46:1134-40.

2. Gamble P, De Beer J, Petruccelli D, Winemaker M. The accuracy of digital templating in uncemented total hip arthroplasty. J Arthroplasty 2010;25:529-32.

3. Holzer LA, Scholler G, Wagner S, Friesenbichler J, Maurer-Ertl W, Leithner A. The accuracy of digital templating in uncemented total hip arthroplasty. Arch Orthop Trauma Surg 2019;139:263-8.

4. Iorio R, Siegel J, Specht LM, Tilzey JF, Hartman A, Healy WL. A comparison of acetate vs digital templating for preoperative planning of total hip arthroplasty: Is digital templating accurate and safe? J Arthroplasty 2009;24:175-9.

5. Schmidutz F, Steinbrück A, Wanke-Jellinek L, Pietschmann M, Jansson V, Fottner A. The accuracy of digital templating: A comparison of short-stem total hip arthroplasty and conventional total hip arthroplasty. Int Orthop 2012;36:1767-72.

6. Kokkotis C, Moustakidis S, Papageorgiou E, Giakas G, Tsaopoulos DE. Machine learning in knee osteoarthritis: A review. Osteoarthr Cartil Open. 2020;2:100069. doi: 10.1016/j.ocarto.2020.100069

7. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Arora S, Chowdhry S, et al. Preoperative prediction of value metrics after primary total joint arthroplasty: Development and validation of patient-specific machine learning algorithms. J Bone Joint Surg Am. 2019;101:587-595

8. Chang PD, Wong TT, Rasiej MJ. Deep learning for detection of complete anterior cruciate ligament tear. J Digit Imaging 2019;32:980-6.

9. Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. J Orthop Res 2020;38:1465-71.

10. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et

al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit Health 2019;1:e271-97.

11. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med 2019;25:44-56.

12. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2:719-31.

13. The B, Diercks RL, Van Ooijen PM, Van Horn JR. Comparison of analog and digital preoperative planning in total hip and knee arthroplasties. A prospective study of 173 hips and 65 total knees. Acta Orthop 2005;76:78-84.

14. Kulkarni A, Parthasarathy M, Ramakrishnan R. Comparison of acetate templating with digital templating for total hip replacement. Malays Orthop J 2008;2:19-21.

15. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344-9.

16. Patel I, Khurana A, Zimmer Z. Intraoperative conversion from uncemented to cemented stems during primary total hip arthroplasty. Orthopedics 2018;41:e218-23.

17. Lindahl H, Garellick G, Regnér H, Herberts P, Malchau H. Three hundred and twenty-one periprosthetic femoral fractures. J Bone Joint Surg Am 2006;88:1215-22.