Machine-Learning-Generated Predictive Model for Acetabular Implant Sizing in Total Hip Arthroplasty

Eric Wang, Maor Shir, Catherine Di Gangi, Ran Schwarzkopf, Morteza Meftah

INTRODUCTION:

Accurate preoperative prediction of acetabular implant size can streamline inventory management and reduce intraoperative inefficiencies in total hip arthroplasty (THA). This study aimed to develop and validate a machine learning model using easily accessible patient demographic data to predict acetabular implant size.

METHODS:

Patients who underwent THA at a large, urban hospital from 2016-2022 were retrospectively reviewed. Sex, age, race, American Society of Anesthesiologists (ASA) classification, body mass index (BMI), height, weight, and acetabular shell implant size were collected. Shells of sizes 44-64 were included in this study; sizes 40, 42, 66 and 72 were excluded due to small sample sizes that impeded model training. The dataset was split into training (80%) and testing (20%) subsets using stratified random sampling.

An initial extreme gradient boosting (XGBoost) model was developed to evaluate feature importance and identify predictive variables. Gain-based feature importance from XGBoost and clinical interpretability guided the selection of four features: age, sex, height, and BMI. These features were used to train a decision tree classifier. Decision trees were chosen for their transparency and ease of clinical integration. Hyperparameter tuning (e.g., maximum depth, minimum samples per split, and splitting criterion) was conducted using a 10-fold cross-validated grid search on the training dataset to optimize model performance.

Model performance was assessed on the testing dataset using accuracy tolerance thresholds of ±1 and ±2 cup sizes. Additionally, confusion matrices and class-specific prediction accuracies were analyzed to understand prediction trends across cup sizes and evaluate error margins.

RESULTS:

In total, 9779 patients with a mean age of 64.6 years (interquartile range [IQR]: 58-72), mean BMI of 29.3 (IQR: 25.1-32.9), and mean height 65.7 inches (IQR: 63-68.9) were included. Model accuracy was 78.4% and 96.8% with tolerance thresholds of +/- 1 and 2 shell sizes, respectively. Prediction performance was reduced for the largest shell sizes (62 and 64 mm), likely due to limited sample sizes.

DISCUSSION AND CONCLUSION:

We developed an interpretable, accurate decision tree model capable of predicting acetabular implant size based on demographic data readily available preoperatively. This tool may enhance surgical planning and inventory optimization in THA.

Table 1: Overall Patient Cohort - Descriptive Statistics

Variables	N = 9,779
Sex, n (%)	
Female	5,670 (58)
Male	4,109 (42)
Age, Median (IQR)	65 (58 – 72)
Race, n (%)	
Asian	164 (1.7)
Black	1,263 (13)
Other	1,080 (10)
White	7,352 (75)
ASA, n (%)	
1	657 (6.5)
2	5,843 (60)
3	3,140 (32)
4	139 (1.4)
BMI (kg/m²), Median (IQR)	28.6 (25.1 – 32.9)
Height (inches), Median (IQR)	65.7 (63.0 – 68.9)
Weight (pounds), Median (IQR)	177 (149 – 208)

IQR, Interquartile Range; ASA, American Society of Anesthesiologists; BMI, Body Mass Index; kg/m², kilograms per square-meter

Table 2: Training and Testing Data Subsets - Descriptive Statistics

	Subset		
Variables	Test N = 1,956	Train N = 7,823	P-value
Sex, n (%)			0.33
Female	1,153 (59)	4,517 (58)	
Male	803 (41)	3,306 (42)	
Age, Median (IQR)	65 (58 – 72)	65 (58 – 72)	0.62
Race, n (%)			0.5
Asian	25 (1.3)	139 (1.8)	
Black	254 (13)	1,009 (13)	
Other	199 (10)	801 (10)	
White	1,478 (76)	5,874 (75)	
ASA Classification, n (%)			0.17
1	122 (6.2)	515 (6.6)	
2	1,134 (58)	4,709 (60)	
3	663 (34)	2,477 (32)	
4	37 (1.9)	122 (1.6)	
BMI, Median (IQR)	28.7 (25.2 – 33.2)	28.6 (25.1 – 32.9)	0.59
Height (inches), Median (IQR)	65.4 (62.8 – 68.9)	65.7 (63.0 – 68.9)	0.38
Weight (pounds), Median (IQR)	177 (149 – 208)	177 (149 – 208)	0.84

[1]Pearson's Chi-squared test; Wilcoxon rank sum test
IQR, Interquartile Range; ASA, American Society of Anesthesiologists; BMI, Body Mass Index; kg/m², kilograms per square-meter

Table 3: Distribution of Acetabular Implant Sizes

	Subset		
Characteristic	Test N = 1,956	Train N = 7,823	P-value
Acetabular Implant Size, n (%)			>0.99
44	5 (0.3)	21 (0.3)	
46	32 (1.6)	130 (1.7)	
48	268 (14)	1,071 (14)	
50	309 (16)	1,234 (16)	
52	490 (25)	1,958 (25)	
54	327 (17)	1,306 (17)	
56	292 (15)	1,170 (15)	
58	156 (8.0)	626 (8.0)	
60	54 (2.8)	218 (2.8)	
62	18 (0.9)	71 (0.9)	
64	5 (0.3)	18 (0.2)	

[1]Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)

Table 4: Accuracy by Acetabular Implant Size (error +/- 1 size)

Acetabular Implant Size	N	Prediction, n (%)	
		Correct	Incorrect
44	5	1 (20)	4 (80)
46	32	25 (78)	7 (22)
48	268	161 (60)	107 (40)
50	309	296 (96)	13 (4.2)
52	490	374 (76)	116 (24)
54	327	289 (88)	38 (12)
56	292	253 (87)	39 (13)
58	156	122 (78)	34 (22)
60	54	12 (22)	42 (78)
62	18	0 (0)	18 (100)
64	5	0 (0)	5 (100)

Table 5: Accuracy by Acetabular Implant Size (error +/- 2 sizes)

Acetabular Implant Size	N	Prediction, n (%)	
		Correct	Incorrect
44	5	4 (80)	1 (20)
46	32	30 (94)	2 (6.3)
48	268	265 (99)	3 (1.1)
50	309	301 (97)	8 (2.6)
52	490	490 (100)	0 (0)
54	327	316 (97)	11 (3.4)
56	292	285 (98)	7 (2.4)
58	156	149 (96)	7 (4.5)
60	54	46 (85)	8 (15)
62	18	7 (39)	11 (61)
64	5	0 (0)	5 (100)