SYSTEMATIC REVIEW

Can Artificial Intelligence Reliably and Accurately Measure Lower Limb Alignment: A Systematic Review and Meta-Analysis

Yashar Khani, MD; Amir Bisadi, MD; Ali Salmani, MD; Negarsadat Namazi, MD; Iman Elahi Vahed, MD; Joben Kianparsa, MD; Mohammad Nouroozi, MD; Fateme Mansouri Rad, MD; Mohammad Poursalehian, MD

Research performed at Akhtar Hospital, Elahie, Tehran, Iran

Received: 25 December 2024 *Accepted: 12 February 2025*

Abstract

Objectives: Lower limb alignment (LLA) measurements are vital for pre-operative assessments and surgical planning in orthopedics. Artificial intelligence (AI) can enhance the precision and consistency of these measurements. This systematic review and meta-analysis evaluates the accuracy and reliability of AI-based approaches in detecting anatomical landmarks and measuring LLA angles, highlighting both their strengths and limitations.

Methods: Adhering to PRISMA guidelines, we searched PubMed, Scopus, Embase, and Web of Science on July 2024 and included observational studies validating AI-driven LLA measurements. Pooled intraclass correlation coefficients (ICCs) were computed to assess inter-rater reliability between AI and manual measurements. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was used to assess study quality.

Results: We reviewed 28 studies with 47,200 patients and 61,253 images; AI demonstrated high reliability in measuring 15 lower limb angles, with pooled ICCs ranging from 0.9811 to 1.0597. Angles like the hip-knee-ankle (HKA; ICC = 0.9987, 95% CI: 0.9975–0.9998) and the mechanical tibiofemoral angle (mTFA; ICC = 1.0001, 95% CI: 1.0001–1.0001) showed near-perfect agreement. In contrast, the joint line convergence angle (JLCA) and femoral anatomical-mechanical angle (FAMA) exhibited lower reliability and significant publication bias. Heterogeneity was substantial across most angles (I^2 = 63%–100%). These findings highlight the potential of AI for clinical applications while also identifying areas that require refinement and standardization.

Conclusion: AI exhibits high reliability and accuracy in measuring key LLA angles, often outperforming manual techniques in both speed and consistency. It holds significant promise as a clinical tool, though challenges with less reliable angles warrant further refinement. Future studies should focus on standardizing landmark definitions and addressing implementation barriers to maximize AI's potential in orthopedic practice.

Level of evidence: IV

Keywords: Artificial intelligence, Hip-knee-ankle angle, Joint line congruency angle, Lower limb alignment, Mechanical axis deviation, Neural network

Introduction

Using a full-length (weight-bearing) leg radiograph, orthopedic surgeons measure the anatomical and mechanical axes in lower limb alignment (LLA).[1] LLA is essential for various investigations, including long-leg discrepancy, pre- and post-operative assessments, deformity assessments, and surgical planning.[1] Malalignment is a considerable risk factor for cartilage damage, pain, gait disturbances, and the development of osteoarthritis.[2,3] Furthermore, patient-reported outcomes, prosthesis survival, and postoperative complications are

Corresponding Author: Amir Bisadi, Department of Orthopedic Surgery, Akhtar Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Email: bisadi.md@sbmu.ac.ir

THE ONLINE VERSION OF THIS ARTICLE
ABJS.MUMS.AC.IR

associated with LLA following total knee arthroplasty.[4,5] Despite the importance of LLA measurement, it's time-consuming and challenged by variability among surgeons and inconsistencies.[6,7] Artificial intelligence (AI) systems represent a promising area of research for improving workflow efficiency and enhancing the accuracy of LLA measurement.

AI has significantly advanced medical imaging by offering tools that enhance diagnostic accuracy, reduce variability, and improve workflow efficiency.[8] Numerous studies have explored the applications of AI in orthopedic imaging, including fracture detection, osteoarthritis classification, implant identification, and the measurement of sub-specialty angles.[8-10]

Because these systems can fully automate the processing of enormous amounts of data, they have gained popularity in medical imaging.[11] Convolutional neural networks (CNNs) are the most frequently used architecture to develop these machine learning software.[12] The capabilitiesof AI in measuring LLAs remain underexplored, with notable gaps in model variations, landmark detection techniques, and imaging protocols. Key questions include which angles are most reliably measured by AI, its accuracy in complex cases, and the practical challenges of integrating AI into clinical workflows. Addressing these issues is crucial for advancing AI from research to widespread clinical application.

This study systematically reviews and analyzes existing research on AI-based LLA measurement, comparing its accuracy and reliability to traditional manual methods. By identifying the strengths and limitations of AI in this context, we aim to evaluate its readiness for clinical adoption and highlight areas that require further development. Ultimately, our work addresses a crucial question: Can AI replace manual LLA measurement while maintaining the accuracy and consistency demanded in orthopedic practice?

Materials and Methods

This systematic review was conducted in complete accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and was prospectively registered with PROSPERO (CRD42023437952).[13]

Literature Search

Scopus, Embase, PubMed, and Web of Science databases were searched comprehensively in July 2024, using a combination of keywords: ("artificial intelligence" OR "machine learning" OR "neural network*") AND ("alignment" OR "malalignment" OR "bone" OR "knee" OR "limb" OR "valgus cut angle" OR "Q angle"). No restriction on time or language were applied during the search. Addittionally, Google Scholar was searched, and the citations of included studies were screened for any relevant studies that may have been overlooked.

Study Eligibility Criteria

Observational studies that validate AI methods for measuring LLA from X-ray radiographs were eligible for inclusion in this study. We also included studies that reported on models that categorize lower limb posture as neutral, valgus, or varus. We excluded papers that presented surgical decision support systems for balancing LLA but did not disclose alignment measurements or status. Additionally, reviews, conference abstracts, non-human studies, and papers not published in English were excluded. This review adhered to the Population Intervention Comparison Outcome (PICO) framework: (P) patients undergoing lower limb alignment assessments; (I) AI-based measurement of lower limb alignment angles from X-ray radiographs; (C) manual or conventional methods of alignment measurement and (O) the accuracy and reliability of AI-based methods, as evaluated by intraclass correlation coefficient (ICC) and error metrics.

Study Selection

After removing duplicate entries, two reviewers (AS and NN) systematically screened the titles and abstracts. They assessed the full texts based on stringent inclusion and exclusion criteria, and any disagreements were effectively resolved through discussion and consultation with the first author (YK).

Data Extraction

A data extraction sheet was developed (YK) that included demographic data, study date, imaging modality, AI model details, number of landmarks, measured angles, and overall conclusions. Two reviewers (MN and JK) independently extracted the data. If multiple machine learning models were utilized in a study, information regarding all of them was recorded. The measured angles included the lateral proximal femoral angle (LPFA), medial proximal femoral angle (MPFA), mechanical lateral distal femoral angle (mLDFA), medial proximal tibial angle (MPTA), lateral distal tibial angle (LDTA), mechanical axis deviation (MAD), hip-knee-ankle (HKA) angle, anatomic tibiofemoral angle (aTFA), mechanical tibiofemoral angle (mTFA), joint line convergence angle (JLCA), weight-bearing line (WBL) ratio, joint line orientation angle (JLOA), femoral anatomical-mechanical angle (FAMA), femoral component alignment (FCA), and tibial component alignment (TCA).[14]

Risk of Bias Assessment

Quality Assessment of Studies of Diagnostic Accuracy-Revised tool was utilized to assess the risk of bias in the included studies.[15] This tool was tailored by considering specific items from the Checklist for Artificial Intelligence in Medical Imaging (CLAIM).[16] Two independent authors (FM and MP) conducted the assessment to ensure clarity and consensus. Any conflicts were resolved through group discussion.

Statistical Analysis

A meta-analysis was conducted to estimate the pooled ICC and the corresponding 95% confidence intervals (CIs) for inter-rater reliability across the studies included in the analysis. We used the Fisher Z-transformation method to stabilize the variance of the ICCs. Following the meta-analysis, the pooled Z-transformed ICC and its 95% confidence interval were back-transformed to the ICC scale. The data were categorized based on the different AI models used in the studies, and a separate meta-analysis was performed for each group. Only AI models with more than three studies were included in the analysis.

Restricted Maximum Likelihood (REML) random effects model was applied in the analyses. Heterogeneity and inconsistency were assessed using Cochran's Q statistics and I^2 tests. P-values less than 0.05 were considered statistically significant. All analyses were performed using R software (version 4.3.2, released on October 31, 2023), using the "meta" and "metafor" packages.

Results

Search Results

We searched all available databases for this systematic review, including PubMed (468), Embase (483), Scopus (634), and Web of Science (415). After eliminating duplicate articles, 996 articles remained. After two individuals reviewed the titles and abstracts, the full texts of 64 articles were obtained. Two team members conducted the full-text screening, resulting in 28 articles that met our criteria [Figure 1].[11,17-43]

Study Characteristics

After extensive research and data extraction, 28 studies conducted between 2020 and 2024 were included in this revies. South Korea,[23,24,26,30,32,33] Germany,[20,34,36,39,43] USA,[18,22,25,40,41] and Austria[11,28,37,38] were the most prominent countries. Only one article utilized knee X-ray (KXR) to measure alignment angles [32]; while three articles used both KXR and full-length limb X-ray (FLXR).[21,30,42] The remaining studies exclusively used FLXR. The included studies processed data from 47,200 patients, including a total of 61,253 images. Ten of these studies implemented data augmentation techniques, and 20 utilized regions of interest recognition before assessing the anatomical landmarks. Four articles focused on measuring alignment in patients either pre- or post-arthroplasty.[20,34,37,40] Two studies examined the accuracy of AI in pediatrics application,[31,41] while three articles evaluated angle measurements in corrective knee osteotomy.[28,29,38] The number of anatomical landmarks varied between three to 31 in different models [Table 1]. The only commercially available product investigated was the Leg Angle Measurement Assistant (LAMA™) by Image Biopsy Lab.[11,18,28,34,37,38]

The reviewed studies used various AI models and architectures, primarily CNNs such as U-NET and ResNet, for image segmentation and landmark detection. Advanced object detection models, including YOLOv3 and YOLOv4 were used.[39,40] DenseNet and Inception ResNet were used for feature extraction and alignment measurements.[31,39,42,43] Segmentation-focused models, such as SegNet and HRNet, were also used to improve accuracy in identifying regions of interest.[17,26]

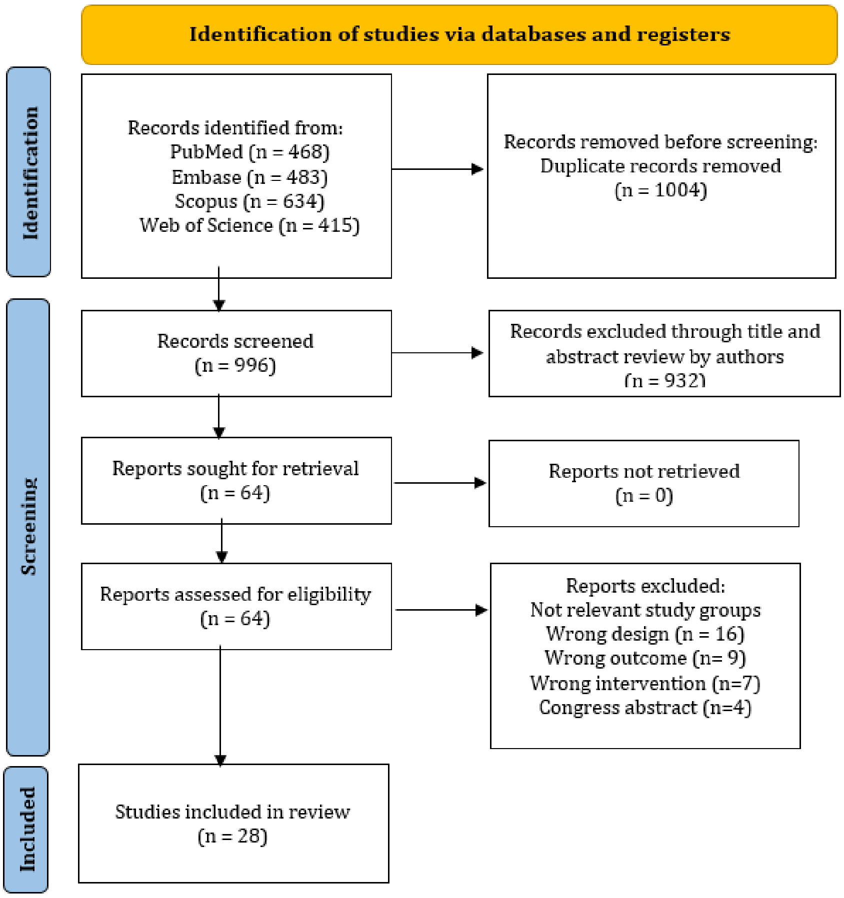

Figure 1. PRISMA flow diagram

Table 1. Study characteristics

Author, year	Country	Images (n)				Data augmentation	Patients (n)	Age	Female (%)	Imaging modality	AI model		ROI	No. Landmarks (hip/knee/ankle)	Angle measured
		Training	*Validation*	*Test*	*Total*						*architecture*	*brand*			
Archer[18] 2023	USA	NR	NR	164	NR	NO	85	NR	58%	FL XR	NR	LAMA™	-	NR	LPFA mLDFA MPTA LDTA MAD HKA JLCA ATFA
Bernard [19] 2023	France	6510	30	561	7101	YES	561	NR	NR	FL XR	CNN (U-NET)	NR	+	1/12/1	mLDFA MPTA HKA JLCA
Erne[20] 2022	Germany	202	200	400	802	YES	119	66 ±9.3	61%	FL XR	CNN (U-NET)	NR	+	9/20/2	mLDFA MPTA LDTA ATFA HKA
Gielis[21] 2020	Netherlands	293	110	NR	NR	NO	100	54±7.4	50%	FL XR + K XR	NR	NR	+	111	HKA
Jang[22] 2024	USA	NR	NR	NR	1078	NO	1078	61.4±9.1	53%	FLXR	CNN (U-NET)	NR	-	NR	FAMA
Jo[23] 2022	South Korea	10407	500	305	11212	NO	11212	62.7±13	73.8%	FL XR	CNN	NR	+	1/11/3	mLDFA MPTA HKA JLCA
Kim[24] 2024	South Korea	11212	300	NR	NR	NO	11212	62.7± 13.0	73.8%	FL XR	CNN	NR	+	5/10/4	mLDFA MPTA LDTA HKA JLCA WBL ratio JLOA
Kunze[25] 2023	USA	200	50	1011	1261	NO	1011	61.2 ±9.0	52.3%	FL XR	CNN (U-NET)	NR	+	1/5/1	HKA
Lee[26] 2024	South Korea	299	37	68	404	NO	434	A: 44.3/ B: 30.2	42.8%	FLXR	seg-Net	NR	+	1/6/1	mLDFA MPTA JLCA
Meng[27] 2022	China	700	100	200	1000	YES	1000	NR	NR	FL XR	CNN (U-NET)	NR	+	2/6/2	mLDFA MPTA JLCA

Table 1. Continued

Mitterer [28] 2023	Austria	NR	110	104	NR	NO	102	40.8 ± 11.8	39.2%	FL XR	NR	LAMA™	+	NR	LPFA mLDFA MPTA LDTA MAD HKA JLCA
Miyama [29] 2024	Japan	20	NR	87	107	NO	107	60.6 ± 9.4	62%	FLXR	NR	NR	+	8/10/2	mLDFA MPTA LDTA HKA WBL ratio
Moon [30] 2021	South Korea	3249	362	386	3997	NO	2001	64.8 ± 12.8	82%	FL XR + K XR	CNN	NR	-	NR	WBL ratio
Murad [31] 2024	Canada	1943	470	470	2883	YES	46	NR	NR	FL XR	ResNet-50	NR	+	2/3/3	mLDFA
Nam [32] 2023	South Korea	2319	287	290	2896	YES	2410	65.1 ± 9.31	83.2%	K XR	DL	NR	-	4	WBL ratio
Nguyen [33] 2020	South Korea	2100	80	80	2260	NO	NR	NR	NR	FL XR	CNN	NR	+	2/2/1	mLDFA MPTA HKA FAMA
Pagano [34] 2023	Germany	NR	NR	200	NR	NO	100	66.8±8.2	50%	FL XR	NR	LAMA™	-	NR	LPFA mLDFA MPTA LDTA MAD HKA JLCA
Pei [35] 2020	China	541	135	120	796	NO	398	NR	72%	FL XR	CNN (U-NET)	NR	+	1/1/1	HKA
Schock [36] 2021	Germany	109	40	106	255	YES	225	26	NR	FL XR	CNN (U-NET)	NR	+	1/2/1	HKA FAMA
Schwarz [37] 2022	Austria	8660	174	1312	10146	NO	146	68.8 ± 10.0	NR	FL XR	NR	LAMA™	-	NR	HKA TCA FCA
Simon [11] 2022	Austria	NR	NR	289	NR	NO	284	65	34.2%	FL XR	CNN (U-NET)	LAMA™	+	NR	mLDFA MPTA HKA
Stotter [38] 2023	Austria	NR	NR	190	NR	YES	95	46.9 ± 7.6	43.2%	FL XR	CNN (U-NET)	LAMA™	+	NR	mLDFA MPTA MAD HKA JLCA

Table 1. Continued

Tack[39] 2021	Germany	NR	NR	NR	900	NO	3843	63.16±9.1	57%	FL XR	ResNet/ YOLOv4	NR	+	6/2/2	HKA
Tanner[40] 2024	USA	854	245	122	1221	YES	1379	65	56%	FLXR	YOLOv3	NR	+	NR	HKA
Tsai[41] 2021	USA	NA	NA	NA	528	YES	517	10.8±4.2	45%	FL XR	CNN/ Pyro4	NR	-	1/1/1	HKA
Wang[42] 2023	UK	5950	2550	2550	8500	NO	7518	NR	NR	FL XR + K XR	DenseNet/ Inception ResNet	NR	-	NR	HKA ATFA
Wilhelm [43] 2024	Germany	356	60	178	594	NO	594	41.1 ± 13.2	30.6%	FLXR	RCNN / ResNet	NR	+	2/12/5	mLDFA MPTA MAD JLCA FAMA
Yang[17] 2024	China	837	101	204	1142	YES	623	62.72±8.16 / 65.33±5.11	NR	FLXR	HRNet	NR	-	10	mLDFA MPTA HKA JLCA FAMA

ROI: region of interest, FLXR: full-length x-ray, KXR: knee x-ray, mLDFA: mechanical lateral distal femoral angle, MPTA: medial proximal tibial angle, LDTA: lateral distal tibial angle, ATFA: anatomical tibiofemoral angle, HKA: hip-knee-ankle angle, WBL: weight-bearing line, LPFA: lateral proximal femoral angle, JLCA: joint line convergence angle, MAD: mechanical axis deviation, JLOA: joint line orientation angle, TCA: tibial component angle, FCA: femoral component angle, FAMA: femoral anatomical-mechanical angle, CNN: convolutional neural network, LAMA: Leg Angle Measurement Assistant

Fifteen angles were analyzed to assess LLA using a machine learning (ML) technique. The HKA angle, referenced in 21 studies, was the angle that was evaluated the most frequently among these angles. The ICC values for the measured angles ranged from 0.71 to 1.0, indicating that the automated AI measurement method exhibited a high level of accuracy and comparability with manual measurements. The ML algorithm significantly correlated with manual measurements across all cases. Among the angles measured, LDTA, mLDFA, and JLCA exhibited the lowest correlation with manual measurements. In contrast, MAD and HKA angles exhibited the highest correlation and accuracy. Yang et al. measured five angles in both native and prosthetic knee joints, reporting a lower discrepancy in angle measurements, particularly for JLCA, mLDFA, and MPTA.[17] A summary of the measured angles is presented in the supplementary material.

Meta-analysis

The pooling of the ICCs for individual angles revealed that AI models consistently attained excellent inter-rater reliability, with ICCs ranging from 0.9811 to 1.0597. Despite this robust performance, there was considerable heterogeneity for most angles, with I^2 values ranging from 63% to 100%, indicating significant variability. Publication bias was identified in several measurements, especially for angles such as the JLCA and the FAMA. Critical angles, including the HKA and the mTFA, demonstrated near-perfect reliability, highlighting their clinical significance. These findings underscore the reliability of AI in standard measurement and the necessity for further refinement in areas characterized by variability and bias [Figure 2 and Table 2].

Quality Assessment

The risk of bias assessment revealed that seven studies exhibited a high risk of bias in patient selection due to inappropriate exclusions. In contrast, several other studies presented an unclear risk owing to insufficient data or unclear patient selection processes. Within the reference standard domain, two studies were classified as having high risk of bias due to a lack of information regarding human annotators and replication details, while two additional studies demonstrated an unclear risk. In the flow and timing domain, nine studies exhibited a high risk because not all patients were analyzed or did not receive the same reference standard. Additionally, six studies were categorized as having an unclear risk due to insufficient information [Figure 3 and Table 3].

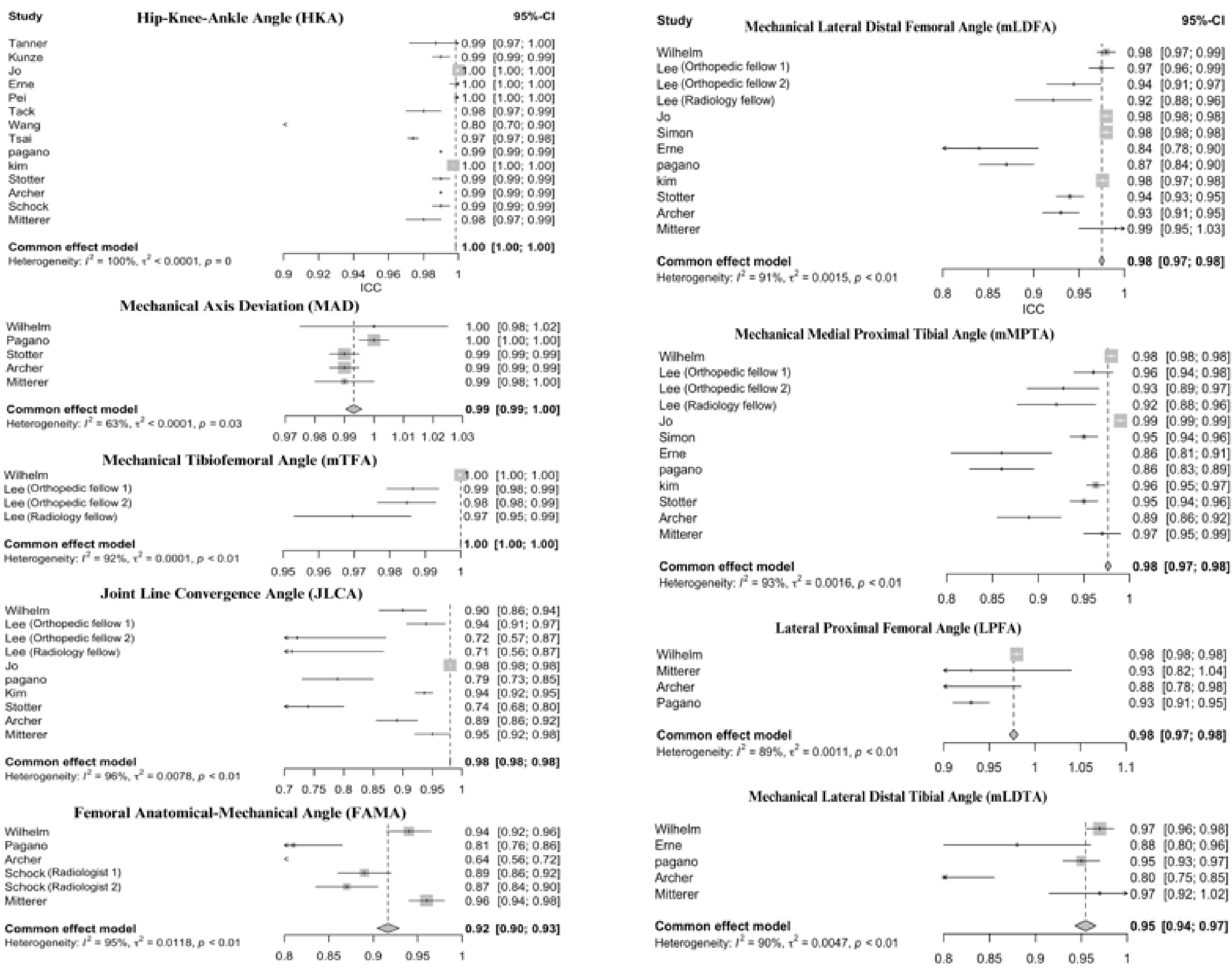

Figure 2. Meta-analysis forest plots

Table 2. Summary of meta-analyses for each measured angle

Angle	No. studies	Pooled ICC (95% CI)	Heterogeneity%	Publication bias (p)
HKA	14	0.9987 (0.9975-0.9998)	100	0.3094
MAD	5	0.9928 (0.9719-1.0136)	63	0.9499
mTFA	2	1.0001 (1.0001-1.0001)	92	< 0.0001
JLCA	8	0.9811 (0.9799-0.9823)	96	0.0001
mLDFA	10	0.9888 (0.9781-0.9995)	91	0.0040
mMPTA	10	0.9965 (0.9867-1.0063)	93	0.0001
LPFA	4	0.9850 (0.9519-1.0182)	89	0.2400
mLDTA	5	0.9930 (0.9055-1.0805)	90	0.1945
FAMA	5	1.0597 (1.0181-1.1012)	95	0.0005

Hip-Knee-Ankle Angle (HKA); Mechanical Axis Deviation (MAD); Mechanical Tibiofemoral Angle (mTFA); Joint Line Convergence Angle (JLCA); Mechanical Lateral Distal Femoral Angle (mLDFA); Mechanical Medial Proximal Tibial Angle (mMPTA); Lateral Proximal Femoral Angle (LPFA); Mechanical Lateral Distal Tibial Angle (mLDTA); Femoral Anatomical-Mechanical Angle (FAMA); Confidence interval (CI)

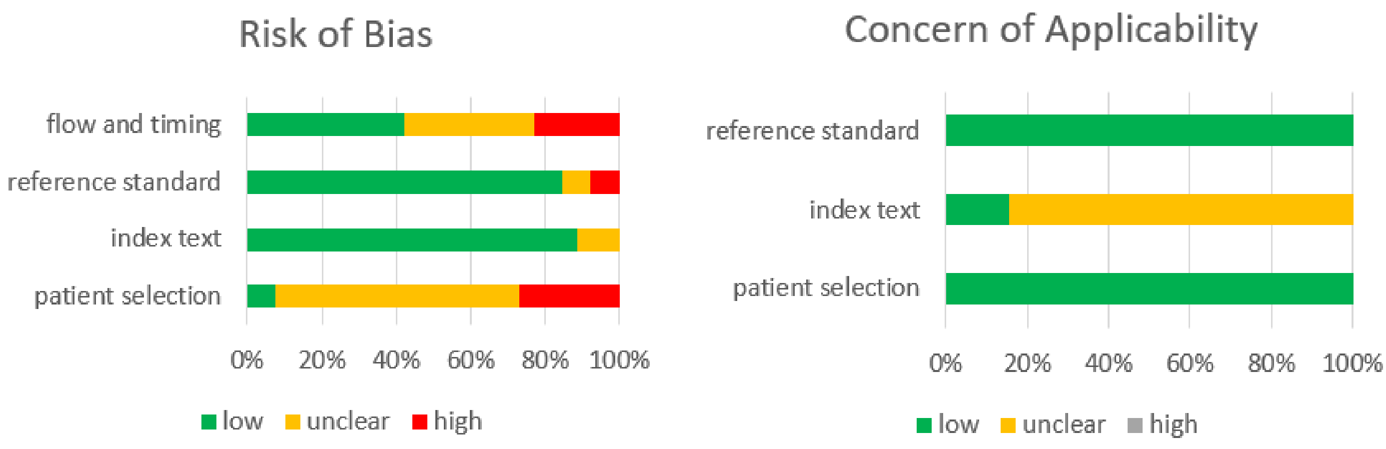

Figure 3. Quality assessment graphs

Table 3. Details of Quality assessment

	Risk of bias				**Concern of Applicability**		
	patient selection	**Index Test (Machine Learning)**	**Reference Standard**	**Flow and Timing**	**Patient Selection**	**Index Test**	**Reference Standard**
Jang et al.	unclear	unclear	low risk	low risk	low risk	unclear	low risk
Miyama et al.	unclear	unclear	low risk	low risk	low risk	unclear	low risk
Lee et al.	unclear	unclear	low risk	low risk	low risk	low risk	low risk
Tanner et al.	unclear	unclear	low risk	low risk	low risk	unclear	low risk
Wilhelm et al.	unclear	low risk	low risk	low risk	low risk	low risk	low risk
yang et al.	unclear	low risk	low risk	low risk	low risk	unclear	low risk
Simon et al.	unclear	unclear	low risk	high risk	low risk	unclear	low risk
Stotter et al.	high risk	low risk	low risk	high risk	low risk	unclear	low risk
Schwarz et al.	high risk	unclear	low risk	high risk	low risk	unclear	low risk
Mitterer et al.	high risk	unclear	low risk	high risk	low risk	low risk	low risk
Archer et al.	high risk	unclear	low risk	high risk	low risk	low risk	low risk
Wang et al.	unclear	unclear	unclear	low risk	low risk	unclear	low risk
Tack et al.	unclear	unclear	low risk	high risk	low risk	unclear	low risk
tsai	low risk	unclear	unclear	high risk	low risk	unclear	low risk
Pei et al.	high risk	unclear	low risk	low risk	low risk	unclear	low risk
Bernard et al.	high risk	unclear	low risk	unclear	low risk	unclear	low risk
Jo et al.	unclear	unclear	low risk	high risk	low risk	unclear	low risk
Meng et al	unclear	unclear	low risk	high risk	low risk	unclear	low risk
nguyen et. al	unclear	unclear	high risk	unclear	low risk	unclear	low risk
jang et al.	unclear	unclear	low risk	unclear	low risk	unclear	low risk
erne et al.	unclear	unclear	low risk	unclear	low risk	unclear	low risk
kunze et al.	low risk	unclear	high risk	unclear	low risk	unclear	low risk
Schock et al.	high risk	unclear	low risk	low risk	low risk	unclear	low risk

Table 3. Continued

Moon et al.	unclear	unclear	low risk	low risk	low risk	unclear	low risk
Nam et al.	unclear	unclear	low risk	unclear	low risk	unclear	low risk
Zheng et al.	unclear	unclear	low risk	low risk	low risk	unclear	low risk

Discussion

We conducted a systematic review of 28 studies published between 2020 and 2024. These studies included data from 47,200 patients and 61,253 images used to measure alignment angles. The HKA angle was the most commonly assessed and demonstrated high accuracy, comparable to manual measurements. Our meta-analyses showed that AI technologies are reliable for clinical applications, with the MAD, HKA, and mTFA angles exhibiting the highest reliability.

This systematic review assessed the predictive value of AI in measuring LLA and compared its accuracy to that of manual measurement. A significant advantage of AI is its high processing speed; for instance, the measurement times for different models per image range from 0.3 to 20.63 seconds, whereas manual measurement takes between 135.4 and 616.8 seconds.[22,26,43] Additionally, AI helps prevent human error. The accuracy of AI detection was influenced by the angle of measurement, with higher correlations reported in the MAD and HKA compared to the LDTA, mLDFA, and JLCA.[17,34] Another factor affecting detection accuracy was the extent of joint deformity; for instance, landmark detection was easier in a prosthetic knee joint than in a knee with severe osteoarthritis and multiple osteophytes.[37] Therefore, the timing of the measurement—whether pre- or post-operation— should be carefully considered when evaluating the advantages and disadvantages of using AI for alignment measurement.

AI has demonstrated the potential to revolutionize routine tasks performed in an orthopedic surgeon's practice. Accurate and consistent LLA measurement is essential for achieving optimal outcomes. Kurmis and Ianunzio[44] emphasize AI's capacity to analyze extensive datasets and improve surgical accuracy, which supports our results that AI models attained near-perfect correlation with seasoned professionals who possess decades of experience and training. The aggregated results highlight AI's potential to transform orthopedic practices through enhanced precision, efficiency, and improved patient outcomes.

AI has transformed the field of orthopedics by enhancing surgical techniques and improving patient care. It greatly improves the diagnosis of musculoskeletal issues, intraoperative navigation, and preoperative planning.[45-47] AI assists in identifying types of implants from radiographs, classifying osteoarthritis in the knee, and determining the stages of osteoarthritis with accuracy comparable to that of orthopedic surgeons.[10,48] According to recent studies, AI performs better than humans due to its speed and capacity for parallel analysis of multiple types of data. For instance, Chong et al. found that AI detects periprosthetic joint infections (PJI) more sensitively than surgeons.[49] Xu et al highlight how AI, instead of conventional techniques, can expedite orthopedic procedures.[50] Furthermore, Zhang et al. demonstrate that AI excels in identifying fractures, surpassing expert assessments, especially in less experienced settings.[51]

While there is currently no specific tool for assessing the risk of bias in AI diagnostic studies, recent protocols for developing such tools have been published.[52,53] Therefore, in accordance with current recommendations,[54] we used the QUADAS-2 tool modified by selected items from the CLAIM checklist, which was specifically designed for manuscripts on AI in medical imaging. Additionally, the AI extension for the Transparent Reporting of a multivariable prediction model of Individual Prognosis or Diagnosis (TRIPOD-AI) has recently been developed.[55] To improve the quality of future studies in AI-aided measurement of LLA, researchers should adhere to these reporting guidelines.

In our reviewed studies, the reliability and consistency of AI models were assessed using metrics such as the ICC (most common),[11,28] mean absolute error (MAE),[40] and root mean squared error (RMSE).[11,27,34,39] Using Bland-Altman graphs and calibration metrics provides a more insightful report on biases and practical applications.[34] Standardizing these parameters across research could improve comparability and facilitate the integration of AI into clinical practice.[55]

This study has several limitations, notably population and measurement heterogeneity. Additionally, the ground truth varied across different studies, leading to inter-rater variability. Limited access to raw data for reanalysis in most studies posed a significant challenge. The absence of standardization in the number and selection of anatomical landmarks complicates the comparison of study findings. Furthermore, the review may not fully address the practical challenges and limitations of implementing AI-based measurements in clinical settings, such as the need for specialized equipment, training, and integration with existing workflows. Future studies should examine the long-term clinical outcomes associated with AI-based lower limb angle measurements and address ethical and privacy concerns related to the use of patient data in AI research. There is a pressing need for standardized protocols regarding anatomical landmarks, imaging techniques, and data processing methods.

Conclusion

The accuracy of AI's prognostic role in measuring LLA suggests that it is highly reliable and accurate compared to manual measurements, while also being considerably faster. AI presents a promising alternative to manual measurement in clinical settings.

Acknowledgement

N/A

Authors Contribution: Authors who conceived and designed the analysis: Amir Bisadi, Yashar Khani/ Authors who collected the data: Ali Salmani. Negarsadat Namazi, Joben Kianparsa, Fateme Mansouri Rad/ Authors who contributed data or analysis tools: Mohammad Nouroozi, Iman Elahi Vahed, Mohammad Poursalehian/ Authors who performed the analysis: Iman Elahi Vahed, Mohammad Porsalehian/ Authors who wrote the paper: Yashar Khani, Ali Salmani. Negarsadat Namazi, Joben Kianparsa, Mohammad Nouroozi, Fateme Mansouri Rad, Mohammad Poursalehian

Declaration of Conflict of Interest: The authors do NOT have any potential conflicts of interest for this manuscript.

Declaration of Funding: The author(s) received NO financial support for the preparation, research, authorship, and publication of this manuscript.

Declaration of Ethical Approval for Study: N/A

Declaration of Informed Consent: N/A

Yashar Khani MD [1]
Amir Bisadi MD [2]
Ali Salmani MD [3]
Negarsadat Namazi MD [1]
Iman Elahi Vahed MD [1]
Joben Kianparsa MD [4]
Mohammad Nouroozi MD [5]
Fateme Mansouri Rad MD [6]
Mohammad Poursalehian MD [7]

1 Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2 Department of Orthopedic Surgery, Akhtar Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3 Student Research Committee, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
4 Student Research Committee, School of Medicine, Shahed University, Tehran, Iran
5 Clinical Research Development Unit (CRDU), Shohada-e-Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
6 Birjand University of Medical Sciences and Health Services, Birjand, Iran
7 Joint Reconstruction Research Center, Tehran University of Medical Sciences, Tehran, Iran

References

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer. 2018; 18(8):500-510. doi:10.1038/s41568-018-0016-5.
2. Felson DT, Niu J, Gross KD, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: findings from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. Arthritis Rheum. 2013; 65(2):355-62. doi: 10.1002/art.37726.
3. Sharma L, Chmiel JS, Almagor O, et al. The role of varus and valgus alignment in the initial development of knee cartilage damage by MRI: the MOST study. Ann Rheum Dis. 2013; 72(2):235-40. doi: 10.1136/annrheumdis-2011-201070.
4. Choong PF, Dowsey MM, Stoney JD. Does accurate anatomical alignment result in better function and quality of life? Comparing conventional and computer-assisted total knee arthroplasty. J Arthroplasty. 2009; 24(4):560-9. doi: 10.1016/j.arth.2008.02.018.
5. Tew M, Waugh W. Tibiofemoral alignment and the results of knee replacement. J Bone Joint Surg Br. 1985; 67(4):551-6. doi: 10.1302/0301-620X.67B4.4030849.
6. Laskin RS. Alignment of total knee components. Orthopedics. 1984; 7(1):62-72. doi: 10.3928/0147-7447-19840101-09.
7. Wright JG, Treble N, Feinstein AR. Measurement of lower limb alignment using long radiographs. J Bone Joint Surg Br. 1991; 73(5):721-3. doi: 10.1302/0301-620X.73B5.1894657.
8. Gitto S, Serpi F, Albano D, et al. AI applications in musculoskeletal imaging: a narrative review. Eur Radiol Exp. 2024; 8(1):22. doi:10.1186/s41747-024-00422-8.
9. Jung J, Dai J, Liu B, Wu Q. Artificial intelligence in fracture detection with different image modalities and data types: A systematic review and meta-analysis. PLOS Digit Health. 2024; 3(1):e0000438. doi:10.1371/journal.pdig.0000438.
10. Ren M, Yi PH. Artificial intelligence in orthopedic implant model classification: a systematic review. Skeletal Radiol. 2022; 51(2):407-416. doi:10.1007/s00256-021-03884-8.
11. Simon S, Schwarz GM, Aichmair A, et al. Fully automated deep learning for knee alignment assessment in lower extremity radiographs: a cross-sectional diagnostic study. Skeletal Radiol. 2022; 51(6):1249-1259. doi: 10.1007/s00256-021-03948-9.
12. Li Z, Liu F, Yang W, Peng S, Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Trans Neural Netw Learn Syst. 2022; 33(12):6999-7019. doi: 10.1109/TNNLS.2021.3084827.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021:372:n71. doi: 10.1136/bmj.n71.
14. Marques Luís N, Varatojo R. Radiological assessment of lower limb alignment. EFORT Open Rev. 2021; 6(6):487-494. doi: 10.1302/2058-5241.6.210015.
15. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009.
16. Mongan J, Moy L, Kahn Jr CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020; 2(2):e200029. doi: 10.1148/ryai.2020200029.

17. Yang J, Ren P, Xin P, et al. Automatic measurement of lower limb alignment in portable devices based on deep learning for knee osteoarthritis. J Orthop Surg Res. 2024; 19(1):232. doi: 10.1186/s13018-024-04658-3.
18. Archer H, Reine S, Xia S, et al. Deep learning generated lower extremity radiographic measurements are adequate for quick assessment of knee angular alignment and leg length determination. Skeletal Radiol. 2024; 53(5):923-933. doi:10.1007/s00256-023-04502-5.
19. Bernard de Villeneuve F, Jacquet C, El Kadim B, et al. An artificial intelligence based on a convolutional neural network allows a precise analysis of the alignment of the lower limb. Int Orthop. 2023; 47(2):511-518. doi:10.1007/s00264-022-05634-4.
20. Erne F, Grover P, Dreischarf M, et al. Automated Artificial Intelligence-Based Assessment of Lower Limb Alignment Validated on Weight-Bearing Pre- and Postoperative Full-Leg Radiographs. Diagnostics (Basel). 2022; 12(11):2679. doi: 10.3390/diagnostics12112679.
21. Gielis WP, Rayegan H, Arbabi V, et al. Predicting the mechanical hip-knee-ankle angle accurately from standard knee radiographs: a cross-validation experiment in 100 patients. Acta Orthop. 2020; 91(6):732-737. doi:10.1080/17453674.2020.1779516.
22. Jang SJ, Kunze KN, Casey JC, et al. Variability of the femoral mechanical-anatomical axis angle and its implications in primary and revision total knee arthroplasty. Bone Jt Open. 2024; 5(2):101-108. doi:10.1302/2633-1462.52.Bjo-2023-0056.R1.
23. Jo C, Hwang D, Ko S, et al. Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. Knee Surg Sports Traumatol Arthrosc. 2023; 31(4):1388-1397. doi:10.1007/s00167-022-07124-x.
24. Kim SE, Nam JW, Kim JI, Kim JK, Ro DH. Enhanced deep learning model enables accurate alignment measurement across diverse institutional imaging protocols. Knee Surg Relat Res. 2024; 36(1):4. doi:10.1186/s43019-023-00209-y.
25. Kunze KN, Jang SJ, Li T, et al. Radiographic findings involved in knee osteoarthritis progression are associated with pain symptom frequency and baseline disease severity: a population-level analysis using deep learning. Knee Surg Sports Traumatol Arthrosc. 2023; 31(2):586-595. doi:10.1007/s00167-022-07213-x.
26. Lee HS, Hwang S, Kim SH, et al. Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning. Sci Rep. 2024; 14(1):7226. doi:10.1038/s41598-024-57887-1.
27. Meng X, Wang Z, Ma X, et al. Fully automated measurement on coronal alignment of lower limbs using deep convolutional neural networks on radiographic images. BMC Musculoskelet Disord. 2022; 23(1):869. doi:10.1186/s12891-022-05818-4.
28. Mitterer JA, Huber S, Schwarz GM, et al. Fully automated assessment of the knee alignment on long leg radiographs following corrective knee osteotomies in patients with valgus or varus deformities. Arch Orthop Trauma Surg. 2024; 144(3):1029-1038. doi:10.1007/s00402-023-05151-y.
29. Miyama K, Akiyama T, Bise R, Nakamura S, Nakashima Y, Uchida S. Development of an automatic surgical planning system for high tibial osteotomy using artificial intelligence. Knee. 2024; 48:128-137. doi:10.1016/j.knee.2024.03.008.
30. Moon HD, Choi HG, Lee KJ, Choi DJ, Yoo HJ, Lee YS. Can Deep Learning Using Weight Bearing Knee Anterio-Posterior Radiograph Alone Replace a Whole-Leg Radiograph in the Interpretation of Weight Bearing Line Ratio? J Clin Med. 2021; 10(8). doi: 10.3390/jcm10081772.
31. Murad Y, Chhina H, Cooper A. Fully Automated Analysis of the Anatomic and Mechanical Axes From Pediatric Standing Lower Limb Radiographs Using Convolutional Neural Networks. J Pediatr Orthop. 2024; 44(4):244-253. doi:10.1097/bpo.0000000000002611.
32. Nam HS, Park SH, Ho JPY, Park SY, Cho JH, Lee YS. Key-Point Detection Algorithm of Deep Learning Can Predict Lower Limb Alignment with Simple Knee Radiographs. J Clin Med. 2023; 12(4).doi:10.3390/jcm12041455.
33. Nguyen TP, Chae DS, Park SJ, Kang KY, Lee WS, Yoon J. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. Comput Biol Med. 2020; 120:103732. doi:10.1016/j.compbiomed.2020.103732.
34. Pagano S, Müller K, Götz J, et al. The Role and Efficiency of an AI-Powered Software in the Evaluation of Lower Limb Radiographs before and after Total Knee Arthroplasty. J Clin Med. 2023; 12(17). doi:10.3390/jcm12175498.
35. Pei Y, Yang W, Wei S, et al. Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning. Phys Eng Sci Med. 2021; 44(1):53-62. doi: 10.1007/s13246-020-00951-7.
36. Schock J, Truhn D, Abrar DB, et al. Automated Analysis of Alignment in Long-Leg Radiographs by Using a Fully Automated Support System Based on Artificial Intelligence. Radiol Artif Intell. 2021; 3(2):e200198. doi:10.1148/ryai.2020200198.
37. Schwarz GM, Simon S, Mitterer JA, et al. Artificial intelligence enables reliable and standardized measurements of implant alignment in long leg radiographs with total knee arthroplasties. Knee Surg Sports Traumatol Arthrosc. 2022; 30(8):2538-2547. doi:10.1007/s00167-022-07037-9.
38. Stotter C, Klestil T, Chen K, et al. Artificial intelligence-based analyses of varus leg alignment and after high tibial osteotomy show high accuracy and reproducibility. Knee Surg Sports Traumatol Arthrosc. 2023; 31(12):5885-5895. doi: 10.1007/s00167-023-07644-0.
39. Tack A, Preim B, Zachow S. Fully automated Assessment of Knee Alignment from Full-Leg X-Rays employing a "YOLOv4 And Resnet Landmark regression Algorithm" (YARLA): Data from the Osteoarthritis Initiative. Comput Methods Programs Biomed. 2021:205:106080. doi: 10.1016/j.cmpb.2021.106080.
40. Tanner IL, Ye K, Moore MS, et al. Developing a Computer Vision Model to Automate Quantitative Measurement of Hip-Knee-Ankle Angle in Total Hip and Knee Arthroplasty Patients. J Arthroplasty. 2024; 39(9):2225-2233. doi:10.1016/j.arth.2024.04.062.
41. Tsai A. A deep learning approach to automatically quantify lower extremity alignment in children. Skeletal Radiol. 2022; 51(2):381-390. doi:10.1007/s00256-021-03844-2.
42. Wang J, Hall TAG, Musbahi O, Jones GG, van Arkel RJ. Predicting

hip-knee-ankle and femorotibial angles from knee radiographs with deep learning. Knee. 2023; 42:281-288. doi:10.1016/j.knee.2023.03.010.

43. Wilhelm NJ, von Schacky CE, Lindner FJ, et al. Multicentric development and validation of a multi-scale and multi-task deep learning model for comprehensive lower extremity alignment analysis. Artif Intell Med. 2024; 150:102843. doi:10.1016/j.artmed.2024.102843.
44. Kurmis AP, Ianunzio JR. Artificial intelligence in orthopedic surgery: evolution, current state and future directions. Arthroplasty. 2022;4(1):9. doi:10.1186/s42836-022-00112-z.
45. Droppelmann G, Rodríguez C, Jorquera C, Feijoo F. Artificial intelligence in diagnosing upper limb musculoskeletal disorders: a systematic review and meta-analysis of diagnostic tests. EFORT Open Rev. 2024; 9(4):241-251. doi:10.1530/eor-23-0174.
46. Gupta P, Haeberle HS, Zimmer ZR, Levine WN, Williams RJ, Ramkumar PN. Artificial intelligence-based applications in shoulder surgery leaves much to be desired: a systematic review. JSES Rev Rep Tech. 2023; 3(2):189-200. doi:10.1016/j.xrrt.2022.12.006.
47. Morya VK, Lee HW, Shahid H, et al. Application of ChatGPT for Orthopedic Surgeries and Patient Care. Clin Orthop Surg. 2024; 16(3):347-356. doi:10.4055/cios23181.
48. Joseph GB, McCulloch CE, Sohn JH, Pedoia V, Majumdar S, Link TM. AI MSK clinical applications: cartilage and osteoarthritis. Skeletal Radiol. 2022; 51(2):331-343. doi:10.1007/s00256-021-03909-2.
49. Chong YY, Chan PK, Chan VWK, et al. Application of machine learning in the prevention of periprosthetic joint infection following total knee arthroplasty: a systematic review. Arthroplasty. 2023; 5(1):38. doi:10.1186/s42836-023-00195-2.
50. Xu D, Lou W, Li M, Xiao J, Wu H, Chen J. Current status of robot-assisted surgery in the clinical application of trauma orthopedics in China: A systematic review. Health Sci Rep. 2022; 5(6):e930. doi:10.1002/hsr2.930.
51. Zhang X, Yang Y, Shen YW, et al. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis. Eur Radiol. 2022; 32(10):7196-7216. doi:10.1007/s00330-022-08956-4.
52. Collins G, Dhiman P, Andaur Navarro C, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021; 11(7):e048008. doi: 10.1136/bmjopen-2020-048008.
53. Sounderajah V, Ashrafian H, Aggarwal R, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med. 2020; 26(6):807-808. doi: 10.1038/s41591-020-0941-1.
54. Jayakumar S, Sounderajah V, Normahani P, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. NPJ Digit Med. 2022; 5(1):11. doi: 10.1038/s41746-021-00544-y.
55. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024:385:e078378. doi: 10.1136/bmj-2023-078378.