

Automated measurements of long leg radiographs in pediatric patients: a pilot study to evaluate an artificial intelligence-based algorithm

Lelij, T.J.N. van der; Grootjans, W.; Braamhaar, K.J.; Witte, P.B. de

Citation

Lelij, T. J. N. van der, Grootjans, W., Braamhaar, K. J., & Witte, P. B. de. (2024). Automated measurements of long leg radiographs in pediatric patients: a pilot study to evaluate an artificial intelligence-based algorithm. *Children*, *11*(10). doi:10.3390/children11101182

Version:Publisher's VersionLicense:Creative Commons CC BY 4.0 licenseDownloaded from:https://hdl.handle.net/1887/4179370

Note: To cite this publication please use the final published version (if applicable).





Article Automated Measurements of Long Leg Radiographs in Pediatric Patients: A Pilot Study to Evaluate an Artificial Intelligence-Based Algorithm

Thies J. N. van der Lelij ¹, Willem Grootjans ², Kevin J. Braamhaar ¹ and Pieter Bas de Witte ^{1,*}

- ¹ Department of Orthopaedics, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; t.j.n.van_der_lelij@lumc.nl (T.J.N.v.d.L.)
- ² Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; w.grootjans@lumc.nl
- * Correspondence: p.b.de_witte@lumc.nl

Abstract: Background: Assessment of long leg radiographs (LLRs) in pediatric orthopedic patients is an important but time-consuming routine task for clinicians. The goal of this study was to evaluate the performance of artificial intelligence (AI)-based leg angle measurement assistant software (LAMA) in measuring LLRs in pediatric patients, compared to traditional manual measurements. Methods: Eligible patients, aged 11 to 18 years old, referred for LLR between January and March 2022 were included. The study comprised 29 patients (58 legs, 377 measurements). The femur length, tibia length, full leg length (FLL), leg length discrepancy (LLD), hip-knee-ankle angle (HKA), mechanical lateral distal femoral angle (mLDFA), and mechanical medial proximal tibial angle (mMPTA) were measured automatically using LAMA and compared to manual measurements of a senior pediatric orthopedic surgeon and an advanced practitioner in radiography. Results: Correct landmark placement with AI was achieved in 76% of the cases for LLD measurements, 88% for FLL and femur length, 91% for mLDFA, 97% for HKA, 98% for mMPTA, and 100% for tibia length. Intraclass correlation coefficients (ICCs) indicated moderate to excellent agreement between AI and manual measurements, ranging from 0.73 (95% confidence interval (CI): 0.54 to 0.84) to 1.00 (95%CI: 1.00 to 1.00). Conclusion: In cases of correct landmark placement, AI-based algorithm measurements on LLRs of pediatric patients showed high agreement with manual measurements.

Keywords: artificial intelligence; leg angle measurement assistant; LAMA; long leg radiographs; pediatric; orthopedics

1. Introduction

Long leg radiographs (LLRs) serve as a crucial diagnostic tool for assessing bone length, lower limb alignment, and joint line orientation. Specifically in children, LLRs play an important role in the diagnosis and quantification of various limb malalignments and deformities, including genu varum (bow legs), genu valgum (knock knees), and leg length discrepancy (LLD) [1]. However, performing and interpreting length and angle measurements manually on LLRs of patients is very time-consuming for clinicians and prone to intra-and interobserver bias [2–4]. In this study, we evaluated the performance of an artificial intelligence (AI)-based software application for automatic assessment of LLRs in pediatric patients.

LLD and lower limb malalignment are common pediatric orthopedic issues that are associated with various musculoskeletal disorders including gait deviation, scoliosis, low back pain, osteoarthritis, and compromised postural control [5,6]. Although an LLD < 1 cm is often asymptomatic and present in up to 90% of the population, LLD in children can be progressive, and LLD > 2 cm may become symptomatic later in life [6–9]. Similarly, malalignment (e.g., valgus and varus leg angles) developed during childhood may increase



Citation: van der Lelij, T.J.N.; Grootjans, W.; Braamhaar, K.J.; de Witte, P.B. Automated Measurements of Long Leg Radiographs in Pediatric Patients: A Pilot Study to Evaluate an Artificial Intelligence-Based Algorithm. *Children* **2024**, *11*, 1182. https://doi.org/10.3390/ children11101182

Academic Editor: Niels Wedderkopp

Received: 30 August 2024 Revised: 13 September 2024 Accepted: 17 September 2024 Published: 27 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the risk of early osteoarthritis in adulthood [10]. Minimally invasive procedures (i.e., guided growth procedures) are available to manage LLD and lower limb malalignment to prevent future symptoms [8]. LLR measurements play a critical role in the clinical decision-making and follow-up for children treated with guided growth procedures [11–14]. Automation of such measurements with the use of artificial intelligence (AI) techniques, particularly deep learning (DL) algorithms, has the potential to improve the speed, accuracy, and efficiency of these evaluations. This could save time for clinicians and subsequently for patients while at the same time improving the consistency and accuracy of LLR measurements [15–17].

Recent studies have investigated the use, performance, and added value of AI-based algorithms in orthopedic radiology [15]. However, little is known about AI-based measurement programs for LLR in children. We set out to explore an AI-based leg angle measurement assistant (LAMA) that can automate length and angle measurements on LLRs. In previous studies, the performance of this AI-based algorithm has been studied in adults [16–20]. To our knowledge, there are no studies that have used the LAMA software to evaluate a comprehensive set of LLR measurements, including bone length and joint angle measurements, specifically in the pediatric population. The latter is actually one of the largest groups of patients undergoing these radiologic investigations. If the measurements of the AI-based algorithm are consistent with manual measurements, the software may be used in clinical practice as an adjunct or even substitute for the traditional manual measurements, thereby saving valuable time for both clinicians and patients.

Therefore, the aim of this study was to assess the accuracy and reliability of measurements performed by the LAMA software compared to manual measurements on LLRs in patients under the age of 18 years.

2. Materials and Methods

2.1. Patient Inclusion and Image Acquisition

For this observational cohort study, the study population consisted of pediatric patients referred for LLR between January and March 2022. Patients were included if they were aged 11 to 18 years at the time of LLR. The following exclusion criteria were used: visual artifacts or poor visibility on radiographs, incorrect positioning, non-weight-bearing or abnormal cropping. Abnormal cropping refers to the issue where an LLR does not capture the entire area of the leg(s) or where part of the leg(s) is unintentionally excluded from the radiograph. This can occur due to incorrect patient positioning, improper image capture settings, or technical errors during the imaging process. Artifacts were defined by abnormal or misleading image features that were not caused by the patient's anatomy. LLRs were acquired in a standardized, weight-bearing manner using a digital Aseco+ X-ray system with CXDI detectors (Canon Medical Systems Corporation, Otawara, Japan). The imaging parameters included a tube voltage of 85 kVp and a tube current of 450 mA. Three separate X-ray images were taken: (1) pelvis to mid-femur, (2) mid-femur to mid-tibia, and (3) mid-tibia to foot. These images were subsequently stitched together to create a single LLR image.

2.2. Measurements

All measurements on the LLRs were performed by two observers and LAMA. Firstly, a senior pediatric orthopedic surgeon (>5 years of experience) and an advanced practitioner in radiography (>15 years of experience) independently performed all measurements manually. Observers were blinded to each other's measurements. Secondly, automatic assessment of LLRs was performed using commercially available software based on DL technology (LAMA, version 1.03, ImageBiopsy Lab, Vienna, Austria). The LAMA application was trained on a dataset comprising over 15,000 radiographs sourced from various studies, including the Osteoarthritis Initiative, the Multicenter Osteoarthritis Study, the Cohort Hip and Cohort Knee study, and five sites in Austria [21–23]. The training cohort contained LLRs from adult patients of different ages and ethnic backgrounds and data acquired using different radiography systems. In order to perform the measurements, the

LAMA application identifies anatomical bony landmarks and provides measurements of angles and lengths. For additional details on the model training, readers are directed to the supplement provided by Simon et al. [17].

2.3. Image Analysis

Manual assessment of LLRs was performed using IDS7 software (Sectra AB, IDS7 version 25.2, Linköping, Sweden). With regard to automated measurements, results of the LAMA analysis were visually evaluated in the IDS7 software to assess the correct placement of landmarks, including the top of the femoral head, the medial femoral condyle, the mid-tibial roof, the mechanical axis of the tibia, the proximal tibial knee joint orientation line, the mechanical axis of the femur, and the distal femoral knee joint orientation line (Table 1). With these landmarks, the following measurements were obtained: femur length, tibia length, full leg length (FLL), leg length discrepancy (LLD), hip–knee–ankle angle (HKA), mechanical lateral distal femoral angle (mLDFA), and mechanical medial proximal tibial angle (mMPTA) (Figure 1). Cases with incorrect measurements due to the inability of LAMA to identify the correct landmarks were excluded.

Table 1. Overview of landmarks and measurement variables.

Landmarks		Description		
1. Top of the femoral head		Most superior point of the femoral head		
2. Medial femoral condyle		Most distal point of the medial femoral condyle		
3. Mid-tibial roof		Middle of tibial plafond in the tibiotalar joint		
4. Mechanical axis of the tibia		Axis passing through the center of the ankle joint and the midpoint of the knee joint		
5. Proximal tibial knee joint orientation line		Line crossing the two lowest points of the tibia plateau		
6. Mechanical axis of the femur		Axis passing through the center of the femoral head and the midpoint of the knee joint		
7. Distal femoral ki	nee joint orientation line	Line passing through the most distal points of the femoral condyles		
Measurement Variables	I	Landmarks Used		
Femur length	Distance between the most the most distal poir	1 and 2		
Tibia length	Distance between the mo condyle	2 and 3		
Full leg length	Distance between the most m	1 and 3		
Leg length discrepancy	Difference between the fu s	1 and 3		
mMPTA	Angle between the mechan joint	4 and 5		
mLDFA	Angle between the mechani joint	6 and 7		
НКА	Angle between the m	6 and 4		
mMPTA, mechanical medial proximal tibial angle; mLDFA, mechanical lateral distal femoral angle; HKA				

mMPTA, mechanical medial proximal tibial angle; mLDFA, mechanical lateral distal femoral angle; HKA, hip-knee-ankle angle.



Figure 1. Schematic overview of measurements performed on long leg radiographs (LRRs), including (**a**) femur length (F) and tibia length (T), (**b**) full leg length (FLL), (**c**) hip–knee–ankle angle (HKA), (**d**) mechanical lateral distal femoral angle (mLDFA), and mechanical medial proximal tibial angle (mMPTA).

2.4. Statistical Analyses

Patient demographics and distributions of all measurements (2 observers) were evaluated using descriptive statistics: means with standard deviations (in case of normally distributed data) and medians with ranges (in case data were not normally distributed). The primary outcome was correct landmark placement by LAMA. The secondary outcome of this study was the comparison of quantitative analyses (i.e., agreement) between LAMA and manually performed measurements. The manually performed measurements of 2 observers were compared to each other and to the LAMA results.

Paired *t*-tests were used to compare the mean of manual measurements (2 observers) to the measurements obtained by using LAMA for each of the 29 patients (total of 58 legs). For analysis of LLD, left and right legs were compared. Analysis of interobserver agreement was performed by comparing manual measurements between the two observers. Agreement was determined by calculating the intraclass correlation coefficients (ICCs) [24]. We assessed the agreement between the observers using an absolute agreement ICC in a two-way random effects model [25]. Furthermore, agreement was determined between the mean values of the manual measurements (of the two observers) and measurements obtained by LAMA. The agreement between the manual and AI measurements was assessed using an absolute agreement ICC in a two-way mixed effects model [25]. The categorization used for interpreting the ICC values were as follows: values less than 0.50—poor reliability, between 0.50 and 0.75—moderate reliability, between 0.75 and 0.90—good reliability, and greater than 0.90—excellent reliability [25]. Furthermore, manual and LAMA measurements were

visually presented using Bland–Altman plots, including the lower and upper limits of the 95% confidence interval (CI) of agreement [26]. Analyses were performed using SPSS (version 29.0, IBM SPSS Statistics, Armonk, NY, USA).

3. Results

A total of 29 patients (58 legs) were included in this study. The median age of the patients was 13.7 years (range 12–16). Furthermore, a total of 12 patients were male (41%) and 17 were female (59%). LLRs were taken for (suspected) LLD (12 patients), tall stature (11 patients), screening for fibrous dysplasia (1 patient), genu valgum (3 patients), or genu varum (2 patients). In a total of eight LLRs, there was erroneous placement of the anatomical landmarks by LAMA. Figure 2 demonstrates a number of these erroneously placed landmarks.



Figure 2. Examples of erroneously placed landmarks by the leg angle measurement assistant (LAMA) showing incorrect identification of (**a**) the top of the femoral head and femoral head center, (**b**) the proximal tibial knee joint orientation line, (**c**) the distal femoral knee joint orientation line.

The erroneous placements were due to failure in the identification of the top of the femoral head (for length measurements) in seven legs, the femoral head center (for angle measurements) in five legs, the placement of the distal femoral knee joint orientation line in one leg, and the proximal tibial knee joint orientation line in one leg. This led to the exclusion of the following LAMA angle measurements: two HKA, five mLDFA, and one

mMPTA measurement in six LLRs. With regard to length measurements by LAMA, seven femur lengths, seven FLLs, and seven LLDs (22 of 29 pairs of legs had correct landmark placements in both legs) were omitted. Landmark placement was correct in 91% to 98% of the cases with regard to angle measurements and in 76 to 100% with regard to length measurements (Table 2).

Table 2. Overview of correct landmark placements with AI software per measurement variable.

Measurement Variables	cLMP (%)	Legs Analyzed (n)
Femur length	88%	51
Tibia length	100%	58
FLL	88%	51
LLD	76%	44
mMPTA	98%	57
mLDFA	91%	53
НКА	97%	56

cLMP, correct landmark placement; FLL, full leg length; LLD, leg length discrepancy; mMPTA, mechanical medial proximal tibial angle; mLDFA, mechanical lateral distal femoral angle; HKA, hip-knee-ankle angle.

3.1. Comparison of Manual Observations

One observer (an advanced practitioner in radiography) reported slightly higher mean values for HKA and mLDFA, while the second observer (a pediatric orthopedic surgeon) measured higher means for LLD. However, the differences between both observers for all measurement variables were small (and considered clinically not to be relevant) and statistically non-significant. Moderate to excellent agreement between the observers for all measurement variables was observed (Table 3).

 Table 3. Mean measurement and interobserver agreement of manual measurements on long leg radiographs (LLRs).

Measurements	Legs Analyzed (n)	Observer 1 (AP)	Observer 2 (OS)	Mean Difference	ICC
Femur length [mm]	51	512.2 (47.6)	512.2 (47.8)	0.0 (-0.4 to 0.5)	1.00 (1.00 to 1.00)
Tibia length [mm]	58	410.6 (43.3)	410.5 (43.3)	0.1 (-0.1 to 0.4)	1.00 (1.00 to 1.00)
FLL [mm]	51	923.5 (89.7)	923.5 (89.6)	0.0 (-0.5 to 0.4)	1.00 (1.00 to 1.00)
LLD [mm]	44	8.9 (9.2)	9.1 (9.2)	-0.2 (-0.8 to 0.3)	0.99 (0.98 to 1.00)
mMPTA [°]	57	88.6 (2.0)	88.5 (1.9)	0.1 (-0.2 to 0.4)	0.88 (0.79 to 0.93)
mLDFA [°]	53	87.0 (1.9)	86.4 (2.0)	0.8 (0.5 to 1.0)	0.91 (0.62 to 0.97)
HKA [°]	56	-0.2 (2.5)	-0.3 (2.5)	0.2 (0.0 to 0.3)	0.98 (0.97 to 0.99)

Measurements of both observers are presented as mean with standard deviation. Mean differences and intraclass correlation coefficients (ICCs) are presented with a 95% confidence interval. Lengths are presented in millimeters and angles in degrees. AP, advanced practitioner in radiography; OS, orthopedic surgeon; FLL, full leg length; LLD, leg length discrepancy; mMPTA, mechanical medial proximal tibial angle; mLDFA, mechanical lateral distal femoral angle; HKA, hip–knee–ankle angle.

3.2. Comparison of LAMA with Manual LLR Measurements

After the exclusion of measurements with erroneously placed landmarks, the LAMA software showed comparable mean values for all lengths and angles compared to the mean manual measurements. For length measurements, the agreement was excellent (ICC ≥ 0.99). For angle measurements, the ICC ranged from moderate to excellent agreement (ICC 0.73 (95%CI 0.54 to 0.84) to 0.97 (95%CI 0.70 to 0.99)) (Table 4). Based on inspection of the Bland–Altman plots, the difference in femur length measurements seemed to increase with larger femoral length (Bland–Altman Figure S1, see Supplementary Material). For the mMPTA, the difference between LAMA and manual measurements seemed to increase when the angle became smaller (Bland–Altman Figure S5, see Supplementary Material).

Measurements	Legs Analyzed (n)	LAMA	Manual	Mean Difference	ICC
Femur length [mm]	51	511.3 (47.3)	512.2 (47.7)	0.9 (0.6 to 1.2)	1.00 (1.00 to 1.00)
Tibia length [mm]	58	411.4 (43.6)	410.6 (43.3)	-0.9 (-1.1 to -0.6)	1.00 (1.00 to 1.00)
FLL [mm]	51	923.8 (89.8)	923.5 (89.7)	-0.3 (-0.7 to 0.1)	1.00 (1.00 to 1.00)
LLD [mm]	44	9.0 (9.4)	9.0 (9.1)	0.0 (-0.5 to 0.4)	0.99 (0.99 to 1.00)
mMPTA [°]	57	87.4 (2.8)	88.5 (1.8)	1.2 (0.7 to 1.6)	0.73 (0.54 to 0.84)
mLDFA [°]	53	87.1 (2.2)	86.8 (1.9)	-0.3 (-0.6 to 0.0)	0.91 (0.84 to 0.95)
HKA [°]	56	0.3 (2.5)	-0.3 (2.5)	-0.6 (-0.8 to -0.4)	0.97 (0.82 to 0.99)

Table 4. Comparison of mean long leg radiograph (LLR) measurements performed by two observers with the measurements obtained by the leg angle measurement assistant (LAMA).

Measurements obtained with LAMA or manually are presented as mean with standard deviation. Mean differences and ICCs are presented with a 95% confidence interval. Lengths are presented in millimeters and angles in degrees. The mean values of the measurements of the two observers are compared to the measurements obtained with the leg angle measurement assistant (LAMA). Lengths are presented in millimeters and angles in degrees. FLL, full leg length; LLD, leg length discrepancy; mMPTA, mechanical medial proximal tibial angle; mLDFA, mechanical lateral distal femoral angle; HKA, hip–knee–ankle angle.

4. Discussion

LLRs are frequently performed on children to assess and follow up on leg alignment and length differences, as well as for surgical planning. However, performing length and angle measurements on these radiographs is labor-intensive and time-consuming. To our knowledge, this was the first study in pediatric patients to assess the performance of automatically analyzed LLRs using a DL-based LAMA software application compared to manually annotated LLRs by two observers. We found that LAMA was able to accurately identify the anatomical landmarks that are needed for length and angle measurements in the vast majority of cases. However, because correct landmark placement ranged between 76% and 100% for different LLR measurements, LAMA should not be used in clinical practice to analyze the LLR of pediatric patients without oversight of landmark placement by a clinician. In cases where landmark placement was correct, the agreement between LLR measurements obtained with LAMA and manual measurements was high, as depicted in the ICCs for both length and angle measurements.

Considering the existing literature on LAMA, Schwarz et al. reported correct landmark placement in 92% and produced an output rate (angle measurements) of 96% [16]. Simon et al. found an overall accurate landmark placement in 89% of cases and a higher output rate (length and angle measurements) of 98.0% [17]. Although we obtained an output rate in all of the cases with LAMA, the percentage of correctly placed anatomical landmarks in our study was slightly lower compared with these studies in adult patients. One explanation could be that ossification is still ongoing in (younger) children or that children with the indication for LLR have anatomical abnormalities (i.e., LLD or varus/valgus due to an underlying disease), making it difficult for LAMA to identify the right landmarks and draw the correct lines. Based on some of the observed erroneous landmark placements in our study by LAMA on the proximal tibial knee joint orientation line, it may seem that it can be difficult to find the most distal point in the tibia plateau groove in children (as depicted in Figure 2b). Also, for the human eye, finding the exact most distal central point of a reasonably flat surface on a 2D image can be rather difficult. Thus, resulting in more variability between human observers and LAMA. The rate of correct landmark placement may be further improved by providing data for the retraining of skeletally immature patients.

Length measurements with LAMA resulted in high output rates and landmark placements, except for LLD. As shown, correct landmark placement was considerably lower in cases of LLD measurements. To obtain the LLD, two correctly measured FLLs of a patient are needed (i.e., from both legs), requiring four correct landmarks for each measurement. When a single landmark required for FLL in one leg is incorrectly placed, LLD cannot be determined, thus explaining the somewhat higher exclusion rate for LLD measurements in our study. Also, the top of the femur head is one of the landmarks that is needed to calculate FLL and LLD. The presence of a deformed femoral head in children who have been identified as needing an LLR may hamper the correct identification of the top of the femur head landmark by LAMA.

With regard to angle measurements, specifically, mMPTA measurements appeared difficult in our dataset, as reflected by lower ICC compared to other measurements. This finding does not correspond with Archer et al., who evaluated the agreement in LLD and knee alignment measurements between LAMA AI software and two manual observers in adult patients [20]. The latter study reported an ICC of 0.89 (95%CI 0.85 to 0.92) for mMPTA when comparing the output of an AI model with manual observers. Another study by Erne et al. on adult patients, using an algorithm based on AI for automated leg measurements on LLR, also showed a higher ICC for mMPTA (ICC > 0.83) between the AI model and manual measurements compared with our study [27].

A recent study conducted by Zheng et al. investigated a different DL-based model on an LLR dataset of children and found a high consistency of LLD measurements between automated DL-based and manual measurements (Pearson correlation coefficient (r) 0.94) [28]. Although r and ICC evaluate validity and reliability, they highlight different aspects, hampering a direct comparison of their results to the findings of our study. Whereas r assesses a linear relationship, ICC provides an absolute and more robust agreement between the two methods. Lastly, de Villeneuve et al. compared an algorithm based on a machine learning process with 11 orthopedic surgeons and found mean differences for mLDFA of 2.1°, MPTA 1.6°, and HKA 1.3° [29]. In our study, similar differences between manual and LAMA measurements were found.

Our results suggest that AI applications like LAMA have the potential to enhance the time efficiency of LLR assessment in the pediatric orthopedic setting. This could be a significant gain in high-volume clinics, as it reduces the time and effort that is required for leg length and joint angle measurements. Furthermore, improved efficiency and reduced variability could potentially lead to better patient outcomes by enabling early and accurate detection of deviating growth in pediatric patients. As for imaging, the precision of measurements may be further improved by using Cone–Beam Computed Tomography (CBCT) or low-dose biplanar digital X-ray systems [30], hereby eliminating factors such as rotation and fan effect distortion [31].

There are some limitations of our study to take into account. Firstly, given that the measurements performed by the senior pediatric orthopedic surgeon and advanced practitioner in radiography are not deemed flawless, the question arises as to whether measurement discrepancies can be attributed to the flawed measurements of LAMA or variability and inaccuracy of the observers (i.e., there is no gold standard). To assess the reliability of LAMA fully, it is important to underline that both inter- and intraobserver variability exist in manual observations. Whereas automated software applications, such as LAMA, will always provide the exact same measurements (i.e., there is no intraobserver variability), intraobserver variability will be present for manual observers. We found moderate to excellent interobserver agreement for all length and angle measurements on LLRs in our cohort of children, but we did not assess the intraobserver variability within the observers. However, the possible intraobserver variability within the manual observers is expected to be very small [4,32]. Therefore, we do not expect that the latter would have significantly influenced our findings with regard to the agreement between LAMA and manually performed measurements. Another limitation is that the reason for landmark misplacement by LAMA is not always clear due to the opaque nature of the software application. In our study, a case was observed where the hip joint anatomy was abnormal, and LAMA completely missed this landmark. In the latter example, it may be clear why the placement of a specific landmark went wrong, but in other cases without obvious osseous or other structural deformations or image artifacts, it may not always be clear why the placement of specific landmarks was performed erroneously. The use of explainable AI methods could help to understand the reason for incorrect landmark placement and indicate what could be done to improve landmark placement by the addition of specific

data for retraining the model. Lastly, our sample size was relatively small, and we collected patients from a single center, which may limit the generalizability of our findings.

5. Conclusions

In conclusion, our findings suggest that the LAMA software is a reliable tool for LLR measurements in a pediatric setting that can potentially save valuable time for the treating physician. The LAMA software application demonstrated correct landmark placement in 91% to 98% of cases with regard to angle measurements and in 76 to 100% with regard to length measurements. The latter underlines that manual oversight of landmark placement by the LAMA software in LLRs of pediatric patients is important. If the landmark placement was correct, high agreement of LAMA with manually performed measurements on LLRs was observed.

Supplementary Materials: The following supporting information can be downloaded at: https:// www.mdpi.com/article/10.3390/children11101182/s1, Figure S1: Bland-Altman plot of the artificial intelligence (AI) and manual femur length measurements. Figure S2: Bland-Altman plot of the artificial intelligence (AI) and manual tibia length measurements. Figure S3: Bland-Altman plot of the artificial intelligence (AI) and manual FLL measurements. Figure S4: Bland-Altman plot of the artificial intelligence (AI) and manual LLD measurements. Figure S5: Bland-Altman plot of the artificial intelligence (AI) and manual mMPTA measurements. Figure S6: Bland-Altman plot of the artificial intelligence (AI) and manual mLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual MLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual MLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual MLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual MLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual MLDFA measurements. Figure S7: Bland-Altman plot of the artificial intelligence (AI) and manual HKA measurements.

Author Contributions: Conceptualization, P.B.d.W. and W.G.; methodology, P.B.d.W., W.G. and K.J.B.; software, W.G.; formal analysis, W.G.; investigation, P.B.d.W., W.G., K.J.B. and T.J.N.v.d.L.; resources, W.G.; data curation, K.J.B.; writing—original draft preparation, P.B.d.W. and K.J.B.; writing—review and editing, P.B.d.W., W.G., K.J.B. and T.J.N.v.d.L.; visualization, K.J.B. and T.J.N.v.d.L.; supervision, P.B.d.W. and W.G.; project administration, P.B.d.W. and W.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study by the local medical ethics committee (entry no. nWMODIV2_20240034) because the burden associated with requesting permission was disproportionate to the expected return and approval date was 25 June 2023. The researchers had no access to information that could be used to identify the patients. Patients (and/or caregivers) could indicate if they did not want to participate in scientific research.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank Samir Challiui for performing the manual measurements and Jelle Goeman for supervising the statistical analyses.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Babazadeh, S.; Dowsey, M.M.; Bingham, R.J.; Ek, E.T.; Stoney, J.D.; Choong, P.F. The long leg radiograph is a reliable method of assessing alignment when compared to computer-assisted navigation and computer tomography. *Knee* 2013, 20, 242–249. [CrossRef] [PubMed]
- Vaishya, R.; Vijay, V.; Birla, V.P.; Agarwal, A.K. Inter-observer variability and its correlation to experience in measurement of lower limb mechanical axis on long leg radiographs. *J. Clin. Orthop. Trauma* 2016, 7, 260–264. [CrossRef] [PubMed]
- Bowman, A.; Shunmugam, M.; Watts, A.R.; Bramwell, D.C.; Wilson, C.; Krishnan, J. Inter-observer and intra-observer reliability of mechanical axis alignment before and after total knee arthroplasty using long leg radiographs. *Knee* 2016, 23, 203–208. [CrossRef] [PubMed]
- 4. Braun, S.; Brenneis, M.; Holder, J.; Meurer, A.; Stief, F. Intra- and interobserver reliability analysis of pediatric lower limb parameters on digital long leg radiographs. *J. Orthop. Surg. Res.* **2023**, *18*, 69. [CrossRef]

- 5. Khamis, S.; Carmeli, E. A new concept for measuring leg length discrepancy. J. Orthop. 2017, 14, 276–280. [CrossRef]
- Murray, K.J.; Molyneux, T.; Le Grande, M.R.; Castro Mendez, A.; Fuss, F.K.; Azari, M.F. Association of Mild Leg Length Discrepancy and Degenerative Changes in the Hip Joint and Lumbar Spine. J. Manip. Physiol. Ther. 2017, 40, 320–329. [CrossRef]
- Tallroth, K.; Ristolainen, L.; Manninen, M. Is a long leg a risk for hip or knee osteoarthritis? *Acta Orthop.* 2017, 88, 512–515. [CrossRef]
- 8. Gordon, J.E.; Davis, L.E. Leg Length Discrepancy: The Natural History (And What Do We Really Know). *J. Pediatr. Orthop.* 2019, 39, S10–S13. [CrossRef]
- 9. Knutson, G.A. Anatomic and functional leg-length inequality: A review and recommendation for clinical decision-making. Part I, anatomic leg-length inequality: Prevalence, magnitude, effects and clinical significance. *Chiropr. Osteopat.* 2005, *13*, 11. [CrossRef]
- Sharma, L.; Song, J.; Dunlop, D.; Felson, D.; Lewis, C.E.; Segal, N.; Torner, J.; Cooke, T.D.; Hietpas, J.; Lynch, J.; et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. *Ann. Rheum. Dis.* 2010, 69, 1940–1945. [CrossRef]
- Carli, D.; De Pellegrin, M.; Franceschi, L.; Zinali, F.; Paonessa, M.; Spolaore, S.; Cardaropoli, S.; Cravino, M.; Marcucci, L.; Andreacchio, A.; et al. Evolution over Time of Leg Length Discrepancy in Patients with Syndromic and Isolated Lateralized Overgrowth. J. Pediatr. 2021, 234, 123–127. [CrossRef] [PubMed]
- Stief, F.; Holder, J.; Braun, S.; Brenneis, M.; van Drongelen, S.; Byrnes, S.K.; Layher, F.; Dussa, C.U.; Meurer, A.; Böhm, H. Relevance of instrumented gait analysis in the prediction of the rebound phenomenon after guided growth intervention. *Sci. Rep.* 2024, 14, 16060. [CrossRef] [PubMed]
- Vogt, B.; Gosheger, G.; Wirth, T.; Horn, J.; Rödl, R. Leg Length Discrepancy- Treatment Indications and Strategies. *Dtsch. Arztebl. Int.* 2020, 117, 405–411. [CrossRef] [PubMed]
- 14. Whitaker, A.T.; Vuillermin, C. Lower extremity growth and deformity. Curr. Rev. Musculoskelet. Med. 2016, 9, 454–461. [CrossRef]
- 15. Lalehzarian, S.P.; Gowd, A.K.; Liu, J.N. Machine learning in orthopaedic surgery. World J. Orthop. 2021, 12, 685–699. [CrossRef]
- Schwarz, G.M.; Simon, S.; Mitterer, J.A.; Frank, B.J.H.; Aichmair, A.; Dominkus, M.; Hofstaetter, J.G. 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, 2538–2547. [CrossRef]
- 17. Simon, S.; Schwarz, G.M.; Aichmair, A.; Frank, B.J.H.; Hummer, A.; DiFranco, M.D.; Dominkus, M.; Hofstaetter, J.G. Fully automated deep learning for knee alignment assessment in lower extremity radiographs: A cross-sectional diagnostic study. *Skelet. Radiol.* **2022**, *51*, 1249–1259. [CrossRef]
- Stotter, C.; Klestil, T.; Chen, K.; Hummer, A.; Salzlechner, C.; Angele, P.; Nehrer, S. 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, 5885–5895. [CrossRef]
- Mitterer, J.A.; Huber, S.; Schwarz, G.M.; Simon, S.; Pallamar, M.; Kissler, F.; Frank, B.J.H.; Hofstaetter, J.G. 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, 1029–1038. [CrossRef] [PubMed]
- Archer, H.; Reine, S.; Xia, S.; Vazquez, L.C.; Ashikyan, O.; Pezeshk, P.; Kohli, A.; Xi, Y.; Wells, J.E.; Hummer, A.; et al. Deep learning generated lower extremity radiographic measurements are adequate for quick assessment of knee angular alignment and leg length determination. *Skelet. Radiol.* 2023, *53*, 923–933. [CrossRef]
- 21. The Osteoarthritis Innitiative. Available online: https://nda.nih.gov/oai (accessed on 25 April 2024).
- 22. Segal, N.A.; Nevitt, M.C.; Gross, K.D.; Hietpas, J.; Glass, N.A.; Lewis, C.E.; Torner, J.C. The Multicenter Osteoarthritis Study: Opportunities for rehabilitation research. *PM R* **2013**, *5*, 647–654. [CrossRef]
- 23. Wesseling, J.; Boers, M.; Viergever, M.A.; Hilberdink, W.K.; Lafeber, F.P.; Dekker, J.; Bijlsma, J.W. Cohort Profile: Cohort Hip and Cohort Knee (CHECK) study. *Int. J. Epidemiol.* **2016**, *45*, 36–44. [CrossRef] [PubMed]
- 24. Liljequist, D.; Elfving, B.; Skavberg Roaldsen, K. Intraclass correlation—A discussion and demonstration of basic features. *PLoS ONE* **2019**, *14*, e0219854. [CrossRef] [PubMed]
- Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J. Chiropr. Med. 2016, 15, 155–163. [CrossRef] [PubMed]
- 26. Ludbrook, J. Comparing methods of measurement. Clin. Exp. Pharmacol. Physiol. 1997, 24, 198–203. [CrossRef]
- Erne, F.; Grover, P.; Dreischarf, M.; Reumann, M.; Saul, D.; Histing, T.; Nussler, A.; Springer, F.; Scholl, C. Automated Artificial Intelligence-Based Assessment of Lower Limb Alignment Validated on Weight-Bearing Pre- and Postoperative Full-Leg Radiographs. *Diagnostics* 2022, *12*, 2679. [CrossRef]
- Zheng, Q.; Liu, B.; Tong, X.; Liu, J.; Wang, J.; Zhang, L. Automated measurement of leg length discrepancy from infancy to adolescence based on cascaded LLDNet and comprehensive assessment. *Quant. Imaging Med. Surg.* 2023, 13, 852–864. [CrossRef]
- 29. Bernard de Villeneuve, F.; Jacquet, C.; El Kadim, B.; Donnez, M.; Coue, O.; Poujade, T.; Khakha, R.; Argenson, J.-N.; Ollivier, M. An artificial intelligence based on a convolutional neural network allows a precise analysis of the alignment of the lower limb. *Int. Orthop.* **2023**, *47*, 511–518. [CrossRef]
- 30. Lintz, F.; Beaudet, P.; Richardi, G.; Brilhault, J. Weight-bearing CT in foot and ankle pathology. *Orthop. Traumatol. Surg. Res.* 2021, 107, 102772. [CrossRef]

- 31. Lazennec, J.Y.; Chometon, Q.; Folinais, D.; Robbins, C.B.; Pour, A.E. Are advanced three-dimensional imaging studies always needed to measure the coronal knee alignment of the lower extremity? *Int. Orthop.* **2017**, *41*, 917–924. [CrossRef]
- 32. Feldman, D.S.; Henderson, E.R.; Levine, H.B.; Schrank, P.L.; Koval, K.J.; Patel, R.J.; Spencer, D.B.; Sala, D.A.; Egol, K.A. Interobserver and Intraobserver Reliability in Lower-Limb Deformity Correction Measurements. *J. Pediatr. Orthop.* **2007**, *27*, 204–208. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.