

Advancements in Artificial Intelligence for Foot and Ankle Surgery: A Systematic Review

Puneet Gupta, BS¹ , Kiera A. Kingston, MD², Martin O'Malley, MD^{2,3}, Riley J. Williams, MD^{2,3}, and Prem N. Ramkumar, MD, MBA^{2,3}

Abstract

Background: There has been a rapid increase in research applying artificial intelligence (AI) to various subspecialties of orthopaedic surgery, including foot and ankle surgery. The purpose of this systematic review is to (1) characterize the topics and objectives of studies using AI in foot and ankle surgery, (2) evaluate the performance of their models, and (3) evaluate their validity (internal or external validation).

Methods: A systematic literature review was conducted using PubMed/MEDLINE and Embase databases in December 2022. All studies that used AI or its subsets machine learning (ML) and deep learning (DL) in the setting of foot and ankle surgery relevant to orthopaedic surgeons were included. Studies were evaluated for their demographics, subject area, outcomes of interest, model(s) tested, model(s)' performance, and validity (internal or external).

Results: A total of 31 studies met inclusion criteria: 14 studies investigated AI for image interpretation, 13 studies investigated AI for clinical predictions, and 4 studies were grouped as "other." Studies commonly explored AI for ankle fractures, calcaneus fractures, hallux valgus, Achilles tendon pathologies, plantar fasciitis, and sports injuries. For studies reporting the area under the receiver operating characteristic curve (AUC), AUCs ranged from 0.64 (poor) to 0.99 (excellent). Two studies (6.45%) reported external validation.

Conclusion: Applications of AI in the field of foot and ankle surgery are expanding, particularly for image interpretation and clinical predictions. Current model performances range from poor to excellent, and most studies lack external validation, demonstrating a need for further research prior to deploying AI-based clinical applications.

Level of Evidence: Level III, retrospective cohort study.

Keywords: artificial intelligence, machine learning, foot, ankle, technology, orthopaedics

Introduction

Artificial intelligence (AI) and its subsets machine learning (ML) and deep learning (DL) are being increasingly explored for applications in medicine and orthopaedic surgery.^{3,7,8,10,12,15,19,25,29,31,49} The essentials of AI, ML, and DL for orthopaedic surgeons, clinicians, and researchers have been thoroughly described in previous literature.^{6,12,34,37,38,45} Briefly, AI and its subsets involves the use of technology to simulate human intelligence. Algorithms or models can be developed that learn and understand complex relationships from data sets. These models can then be applied for many different purposes, such as automating

analysis of radiographic images, predicting surgical outcomes, or predicting injuries in sports players.

¹Department of Orthopaedic Surgery, George Washington University School of Medicine and Health Sciences, Washington, DC, USA

²Hospital for Special Surgery, New York, NY, USA

³Brooklyn Nets, National Basketball Association (NBA), Brooklyn, NY, USA

*Prem N. Ramkumar is also affiliated to Long Beach Orthopaedic Institute, Long Beach, CA

Corresponding Author:

Prem N. Ramkumar, MD, MBA, Hospital for Special Surgery, 535 E 70th St, New York, NY 10021-4898, USA.

Email: premramkumar@gmail.com



AI models are being developed in nearly all orthopaedic subspecialties, including hip, knee, spine, and pediatric surgery.^{21,27,28,35,55} Klemm et al²⁷ developed and validated ML models for predicting the risk of early revision surgery after primary total hip arthroplasty (THA). Jo et al²¹ developed and validated an ML model for predicting the risk of transfusion following primary TKA. Merali et al³⁵ developed and validated a DL model for detecting cervical spinal cord compression in magnetic resonance imaging (MRI) scans. Kunze et al²⁸ trained and tested several ML models for predicting patients that would achieve the minimal clinically important difference (MCID) in Hip Outcome Score-Sports Subscale (HOS-SS) following hip arthroscopy for femoroacetabular impingement syndrome. Xu et al⁵⁵ developed a DL-assisted system for automated measurements and classifications pertinent to developmental dysplasia of the hip directly from plain pelvic radiographs.

Potential applications for AI in foot and ankle surgery are vast and are at least partly similar to other orthopaedic subspecialties. Given the impact that AI and its subsets may have on clinical and operative practice, it is important for surgeons to understand the current advancements that have been made thus far in applying AI in foot and ankle surgery. Therefore, the purpose of this systematic review is to (1) characterize the topics and objectives of studies using AI in foot and ankle surgery, (2) evaluate the performance of their models, and (3) evaluate their validity (internal or external validation). We hypothesized that most studies would investigate AI for imaging analysis, have models that are not performing excellently, and have models that are not externally validated.

Methods

Search Strategy

We performed a systematic literature review in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Two reviewers independently completed structured searches using the PubMed/MEDLINE and Embase databases on December 11, 2022, to search for all available articles on the databases before December 11, 2022. The search query used the terms as follows: (artificial intelligence OR machine learning OR deep learning) AND (foot OR ankle OR hallux valgus OR tibial tendon insufficiency OR hallux rigidus OR Lisfranc OR Achilles OR peroneal OR metatarsal OR plantar fasciitis OR midfoot OR talus OR cuboid OR ankle arthroscopy OR ankle arthroplasty). Two experienced orthopaedic researchers independently screened all titles, abstracts, and full-text articles. The reference lists of the final articles were also reviewed and cross-referenced to identify any other additional pertinent studies that were not found from the keyword search. The search strategy used in this study is displayed in Figure 1.

Eligibility Criteria

Standardized inclusion and exclusion criteria were used to determine study eligibility. Any disagreements or discrepancies were resolved by consensus. Inclusion criteria were as follows: (1) involve foot and ankle surgery; (2) involve AI; (3) clinically or operatively relevant to orthopaedic surgeons; (4) published in English; (5) available between January 1, 2005, and December 11, 2022; (6) original studies with level I to IV evidence; (7) published studies providing extractable outcome data. Exclusion criteria were as follows: (1) not involving foot and ankle surgery; (2) not involving AI; (3) not clinically or operatively relevant to orthopaedic surgeons; (4) not published in English; (5) no original, extractable clinical data (ie, review articles, commentaries, letters to the editor); (6) no full-text available; and (7) systematic review, meta-analysis, abstracts, conference proceedings.

Data Items

The primary outcomes of interest were (1) subject area in which AI was being applied, (2) best model performance metrics, and (3) whether the model(s) were internally or externally validated. Other variables for which data were sought included outcomes of interest, number of participants, median or average age of patients, percentage of males in the study, and the models evaluated.

Studies were grouped into 3 categories based on their subject area: clinical predictions, image interpretation, or other. Image interpretation studies were any that used AI for detection, classification, or diagnosis using plain radiographs, magnetic resonance imaging (MRI) images, computed topography (CT) images, or ultrasonographic images.

The best performance metrics were only recorded for studies applying AI for clinical predictions or image interpretation. The primary metrics used for evaluating the performance of models with dependent variables that were categorical/classes were area under the curve of the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). The ROC is a plot of a test's sensitivity and specificity, with sensitivity on the y axis and $1 -$ specificity on the x axis. AUC values range from 0 to 1.0. A value of 1.0 indicates a test has perfect discriminative ability. AUC values were interpreted as follows for the models: >0.90 was considered excellent performance, 0.80 to 0.89 was considered good, 0.70 to 0.79 was considered fair, and 0.51 to 0.69 was considered poor.³² The primary metrics used for evaluating the performance of models with dependent variables that were continuous were root mean squared error (RMSE) and coefficient of determination (R^2). If none of the aforementioned metrics were available, any other pertinent metrics reported by the study were recorded.

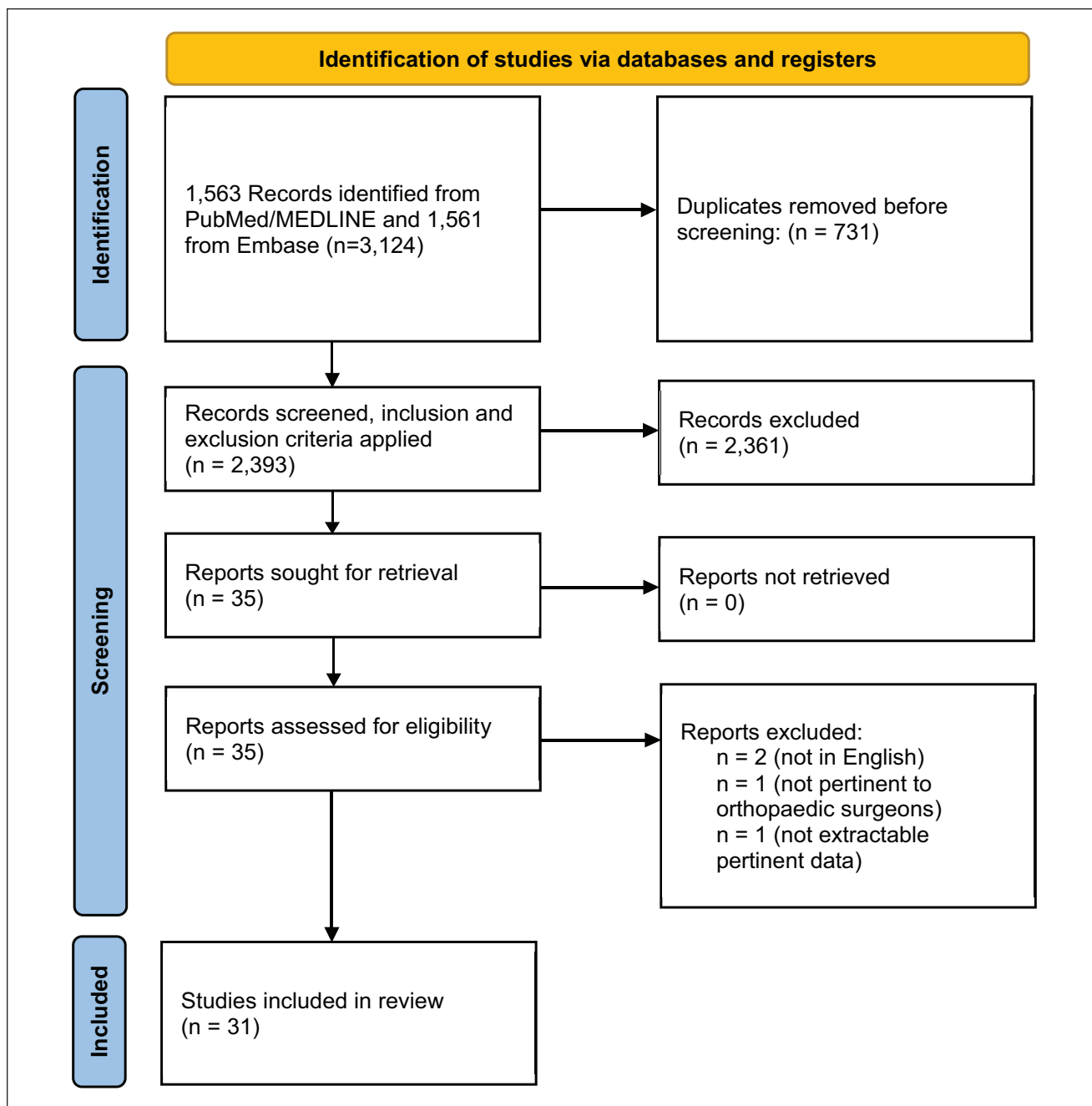


Figure 1. PRISMA diagram.

Validation method was recorded for studies applying AI for clinical predictions or image interpretation. Internal validation was defined as when a model is tested on a population that is similar to that on which it was trained on. External validation was defined as evaluating the performance of an algorithm when applied to an external cohort, such as that from a different institution or national database. Studies in which data from a single population was split into training, validation, and independent test sets were not

considered to have externally validated their models. Determining whether a model has been externally validated is useful for assessing its generalizability.

Data Analysis

No pooled analysis for AUC, accuracy, or other performance metrics was able to be performed because of significant methodological heterogeneity including the models

tested, the types of outcomes, and patient characteristics that increases the risk for bias and inaccurate conclusions.

Results

A total of 31 studies met criteria for inclusion in the final analysis. No additional articles were identified after cross-referencing and reviewing the reference lists. Fourteen studies investigated AI for image interpretation, 13 studies investigated AI for clinical predictions, and 4 studies were grouped as “other.”

Image Interpretation

Of the 14 image interpretation studies, topics included general foot and ankle fractures, Lisfranc malalignment, hallux valgus parameters, calcaneus fractures, and the Achilles tendon (Table 1). Two of the 14 studies externally validated their models (14.3%). DL models were used in all of the studies except for 1 (92.9%) (Table 2). Of the 14 studies, 8 studies reported AUCs, for which the best values ranged from 0.85 (good) to 0.99 (excellent). Eight studies reported accuracies, for which the best values ranged from 72% to 99%.

Ashkani-Esfahani et al⁵ internally validated 2 deep convolutional neural networks (DCNNs) for identifying ankle fractures from radiographs and achieved a near-perfect AUC of 0.99. Kitamura et al²⁶ internally validated 5 separate CNNs for detecting ankle fractures from plain radiographs and achieved a fair fracture detection accuracy of 81%. Prijs et al⁴⁴ internally and externally validated a DL model for detecting, classifying, and localizing ankle fractures from plain radiographs and achieved an excellent AUC of 0.92 and accuracy of 99% (classifying “no fracture”) on external validation. Guermazi et al¹⁴ internally validated a DL model for detecting fractures from foot and ankle plain radiographs, which performed excellently with an AUC of 0.97, sensitivity per patient of 93%, and specificity per patient of 93%. Olczak et al³⁹ internally validated neural network models for classifying ankle fractures from radiographs according to the AO Foundation/Orthopaedic Trauma Association (AO/OTA) 2018 classification, which performed fair to excellent with AUCs ranging from 0.79 to 0.99 in classifying AO types. Pinto Dos Santos et al⁴² internally validated a CNN for detecting fractures in anteroposterior ankle radiographs, which performed good with an AUC of 0.85.

Li et al³⁰ aimed internally validated a DL model for automated detection of 18 anatomical landmarks and measurement of the first-second intermetatarsal angle (IMA), hallux interphalangeal angle (HIA), hallux valgus angle (HVA), and distal metatarsal articular angle (DMAA) from weight-bearing, dorsoplantar radiographs. The observed (manual

by radiologist) and predicted (model) values of the 4 angles correlated well (ICC 0.89-0.96, r 0.81-0.97).³⁰

Wang et al⁵³ internally validated several radiomics-based ML models for diagnosing Achilles tendinopathy from ultrasonographic images in skiers and achieved an excellent AUC of 0.99, 90% sensitivity, and 100%. Kapiński et al²² internally validated several DL models for classifying Achilles tendons as injured or healthy from MRI images and achieved a maximum accuracy of 97.6%, sensitivity of 98.3%, and specificity of 99.45%.

Wang et al⁵⁴ internally and externally tested a DL system for detecting and grading fatigue fractures (a type of stress fracture) from plain radiographs, which performed excellent (AUC 0.911, sensitivity 90.8%) in detection of fatigue fractures for the foot images and good (AUC 0.877, sensitivity 85.5%) for the tibiofibula images. External validity for grading of fatigue fractures was not demonstrated, as the DL system performed poorly with an overall accuracy of 62.9% for the tibiofibula images and an accuracy of 61.1% for the foot images.

Ashkani-Esfahani et al⁴ internally validated 2 DCNN models for detecting Lisfranc instability from single-view (anteroposterior) and 3-view radiographs (anteroposterior, lateral, oblique), which performed excellently with AUCs ranging from 0.925 to 0.994.

Day et al⁹ aimed to assess the performance of an AI-based software that automatically measures the M1-M2 IMA from weightbearing cone beam computed tomography (WBCT) scans in hallux valgus patients. The AI-based software was faster than manual measurements, correlated well with manual measurements, and had higher and nearly perfect test-retest reliability (0.99 intrasoftware intraclass correlation coefficient for both 3D and 2D IMA).⁹

Aghnia Farda et al¹ internally validated a CNN model for classifying calcaneal fractures on CT images into the Sanders system, which performed well with a classification accuracy of nearly 72% after augmenting the data. Pranata et al⁴³ internally validated 2 separate DCNN models for detecting the presence or absence of calcaneal fractures on CT images and achieved an excellent accuracy of 98%.

Clinical Predictions

Of the 13 clinical prediction studies, topics were wide ranging and included predicting outcomes following surgery for ankle fractures, predicting lower extremity sports injuries, predicting recovery of peroneal nerve palsy, and more (Table 3). Zero of the 13 studies externally validated their models (0%). The number of ML and DL models tested ranged from 1 model to 11 models (Table 4). Of the 13 studies, 9 studies reported AUCs, for which the best values ranged from 0.64 (poor) to 0.97 (excellent). Six studies reported accuracies, for which the best values ranged from 70.4% to 93.18%.

Table 1. Artificial Intelligence for Image Interpretation in Foot and Ankle Surgery.

Author	Title	Subject Area	Specific Topic	Number of Participants	Mean or Median Age (y)	% Male Participants	External Validation?
Ashkani-Esfahani et al ⁵	Detection of Ankle Fractures Using Deep Learning Algorithms	Image interpretation	Ankle fractures	2100 patients	41.5 (mean)	58.80	No
Ashkani-Esfahani et al ⁴	Deep Learning Algorithms Improve the Detection of Subtle Lisfranc Malalignments on Weightbearing Radiographs	Image interpretation	Lisfranc malalignment	1280 patients	38.4 (mean)	58.50	No
Guermazi et al ¹⁴	Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence	Image interpretation	Foot and ankle fractures	60 patients (foot and ankle only)	54 (mean)	30.00	No
Li et al ³⁰	Feasibility Study of Hallux Valgus Measurement With a Deep Convolutional Neural Network Based on Landmark Detection	Image interpretation	Hallux valgus imaging parameters	1023 radiographs	44.92 (mean)	37.50	No
Prijs et al ⁴⁴	Development and External Validation of Automated Detection, Classification, and Localization of Ankle Fractures: Inside the Black Box of a Convolutional Neural Network (CNN)	Image interpretation	Ankle fractures	12 000 ankle studies	Not reported	Not reported	Yes
Wang et al ⁵³	Musculoskeletal Ultrasound Image-Based Radiomics for the Diagnosis of Achilles Tendinopathy in Skiers	Image interpretation	Achilles tendonopathy diagnosis (ultrasonography)	139 patients	28.64 (mean)	79.10	No
Wang et al ⁵⁴	Lower-Extremity Fatigue Fracture Detection and Grading Based on Deep Learning Models of Radiographs	Image interpretation	Lower extremity fatigue fractures	3993 images	21.7 (mean)	86.80	Yes
Day et al ⁹	Evaluation of a Weightbearing CT Artificial Intelligence-Based Automatic Measurement for the M1-M2 Intermetatarsal Angle in Hallux Valgus	Image interpretation	Hallux valgus imaging parameters	93 patients (128 feet)	49.6 (mean)	29.00	No
Aghnia Farda et al ¹	Sanders Classification of Calcaneal Fractures in CT Images With Deep Learning and Differential Data Augmentation Techniques	Image interpretation	Calcaneus fractures	760 images	Not reported	Not reported	No

(continued)

Table 1. (continued)

Author	Title	Subject Area	Specific Topic	Number of Participants	Mean or Median Age (y)	% Male Participants	External Validation?
Olczak et al ³⁹	Ankle Fracture Classification Using Deep Learning: Automating Detailed AO Foundation/Orthopedic Trauma Association (AO/OTA) 2018 Malleolar Fracture Identification Reaches a High Degree of Correct Classification	Image interpretation	Ankle fractures	5086 images	Not reported	Not reported	No
Kapinski et al ²²	Monitoring of the Achilles Tendon Healing Process: Can Artificial Intelligence Be Helpful?	Image interpretation	Achilles tendon pathology	90 patients	Not reported	Not reported	No
Kitamura et al ²⁶	Ankle Fracture Detection Utilizing a Convolutional Neural Network Ensemble Implemented With a Small Sample, De Novo Training, and Multiview Incorporation	Image interpretation	Ankle fractures	596 images	Not reported	Not reported	No
Pranata et al ⁴³	Deep Learning and SURF for Automated classification and Detection of Calcaneus Fractures in CT Images	Image interpretation	Calcaneus fractures	1931 images	Not reported	Not reported	No
Pinto Dos Santos et al ⁴²	Structured Report Data Can Be Used to Develop Deep Learning Algorithms: A Proof of Concept in Ankle Radiographs	Image interpretation	Ankle fractures	157 patients/images	43.0 (mean)	51.59	No

Table 2. Summary of Artificial Intelligence Models for Image Interpretation in Foot & Ankle Surgery.

Author	Purpose/Outcome(s)	Models Tested	Best Metrics Achieved
Ashkani-Esfahani et al ⁵	Detect ankle fractures on conventional radiographs	DCNN(s)	3-views Inception V3: AUROC: 0.99 F score: 0.99 Accuracy: 99% NPV: 99% PPV: 99% Specificity: 99%
Ashkani-Esfahani et al ⁴	Detect Lisfranc malalignment from weightbearing radiographs	DCNN(s)	3-views Inception V3: AUROC: 0.99 F score: 0.96 Accuracy: 98.6% 3-views Resnet-50: NPV: 96.9 PPV: 97.8 Specificity: 97.7%
Guermazi et al ¹⁴	Detect foot and ankle fractures	DL model	AUROC 0.97 Sensitivity: 93% Specificity: 93%
Li et al ³⁰	Hallux valgus imaging parameters	DCNN(s)	ICC: 0.96 (HVA) r: 0.97 (HVA)
Prijs et al ⁴⁴	Detect, classify, localize ankle fractures	DCNN(s)	External validation metrics Accuracy: 99% (no fracture) Sensitivity: 99% (no fracture) Specificity: 100% (Weber A and C) AUROC: 0.98 (Weber B)
Wang et al ⁵³	Ultrasonographic diagnosis of Achilles tendinopathy	ML models	AUROC: 0.99 Sensitivity: 90.0% Specificity: 100%
Wang et al ⁵⁴	Detect and grade tibiofibular and foot fatigue fractures	DCNN(s)	AUROC: 0.965 Sensitivity: 96.4% Specificity: 80.1% PPV: 77.6% NPV: 98.4%
Day et al ⁹	Automatic measurement of M1-M2 intermetatarsal angle (IMA) from weightbearing cone beam CT images	Neural network	Correlation coefficients between manual and automatic measurements ranging from 0.52 to 0.63
Aghnia Farda et al ¹	Classify calcaneal fractures (Sanders system) on CT images	DCNN(s)	Accuracy: 72% ^a
Olczak et al ³⁹	Classify ankle fractures according to AO/OTA 2018 classification	Neural network	Weighted mean AUROC: 0.90
Kapiński et al ²²	Detect Achilles tendon injuries from MRI images	DCNN(s) ensembles	Accuracy: 97.6% Specificity: 99.45% Sensitivity: 98.3%
Kitamura et al ²⁶	Detect ankle fracture on plain radiographs	DCNN(s) ensembles	Ensemble A: Accuracy: 81.0% Sensitivity: 80.0% Ensemble B: Specificity: 0.88 PPV: 85.0% NPV: 76.0%
Pranata et al ⁴³	Detect and classify calcaneus fractures in CT images	DCNN(s)	Accuracy: 98%
Pinto Dos Santos et al ⁴²	Detect ankle fracture on plain radiographs	DCNN(s)	Accuracy: 76.9% AUROC: 0.850 Sensitivity: 0.625 Specificity: 1.0 PPV: 1.0 NPV: 0.625

Abbreviations: AUROC, area under the receiver operating characteristic curve; CT, computed tomography; DCNN, deep convolutional neural network; DL, deep learning; HVA, hallux valgus angle; ICC, intraclass correlation coefficient; ML, machine learning; NPV, negative predictive value; PPV, positive predictive value.

^aHighest accuracy when ensuring that patient samples from training set are not in test set.

Table 3. Artificial Intelligence for Clinical Predictions in Foot and Ankle Surgery.

Author	Title	Subject Area	Specific Topic	Number of Participants	Mean or Median Age (y)	% Male Participants	External Validation?
Diniz et al ¹¹	Pre-injury Performance Is Most Important for Predicting the level of Match Participation After Achilles Tendon Ruptures in Elite Soccer Players: A Study Using a Machine Learning Classifier	Clinical predictions	Level of soccer match participation	209 participants	28.3 (mean)	Not reported	No
Lu et al ³³	Machine Learning for Predicting Lower Extremity Muscle Strain in National Basketball Association Athletes	Clinical predictions	Lower extremity muscle strain in professional basketball players	2103 participants	26 (median)	100	No
Oosterhoff et al ⁴⁰	Feasibility of Machine Learning and Logistic Regression Algorithms to Predict Outcome in Orthopaedic Trauma Surgery	Clinical predictions	Posterior malleolar fracture in tibial shaft fracture patients	263 patients	41.00 (mean)	75	No
Vasavada et al ⁵¹	Predictors Using Machine Learning of Complete Peroneal Nerve Palsy Recovery After Multiligamentous Knee Injury: A Multicenter Retrospective Cohort Study	Clinical predictions	Recovery of peroneal nerve palsy	16 participants	27.7 (mean)	87.50	No
Wang et al ⁵²	Identification of Radiographic Characteristics Associated With Pain in Hallux Valgus Patients: A Preliminary Machine Learning Study	Clinical predictions	Hallux valgus pain	72 feet	58.72 (mean)	50.00	No
Jauhiainen et al ²⁰	New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes	Clinical predictions	Predict knee and ankle sports injuries	314 participants	15.72 (mean)	48.40	No
Ruiz-Pérez et al ⁴⁶	A Field-Based Approach to Determine Soft Tissue Injury Risk in Elite Futsal Using Novel Machine Learning Techniques	Clinical predictions	Predict lower extremity, noncontact, soft tissue injury	139 participants	22.45 (mean)	51.80	No
Hendrickx et al ¹⁷	A Machine Learning Algorithm to Predict the Probability of (Occult) Posterior Malleolar Fractures Associated With Tibial Shaft Fractures to Guide "Malleolus First" Fixation	Clinical predictions	Posterior malleolar fracture in tibial shaft fracture patients	263 patients	41.00 (mean)	75	No
Suda et al ³⁰	Recognition of Foot-Ankle Movement Patterns in Long-Distance Runners With Different Experience Levels Using Support Vector Machines	Clinical predictions	Classifying running experience level	78 patients	40.7 (mean)	50.00	No
Sharif Bidabadi et al ⁴⁷	Classification of Foot Drop Gait Characteristic due to Lumbar Radiculopathy Using Machine Learning Algorithms	Clinical predictions	Foot drop classification	86 patients	Not reported	Not reported	No
Merrill et al ³⁶	Machine Learning Accurately Predicts Short-term Outcomes Following Open Reduction and Internal Fixation of Ankle Fractures	Clinical predictions	Ankle fracture outcomes	50005 patients	59.00 (mean)	30.80	No
Yin et al ⁵⁶	Use of Artificial Neural Networks to Identify the Predictive Factors of Extracorporeal Shock Wave Therapy Treating Patients With Chronic Plantar Fasciitis	Clinical predictions	Chronic plantar fasciitis	210 patients	54.1 (mean)	53.33	No
Keijsers et al ²⁴	Classification of Forefoot Pain Based on Plantar Pressure Measurements	Clinical predictions	Forefoot pain	297 patients / 594 feet	51.43 (mean)	34.34	No

Table 4. Summary of Artificial Intelligence Models for Clinical Predictions in Foot & Ankle Surgery.

Author	Purpose/Outcomes(s)	Models Tested	Best Metrics Achieved
Diniz et al ¹¹	Level of match participation following Achilles tendon rupture	Extreme gradient boosting	AUROC: 0.81 Brier loss score: 0.12
Lu et al ³³	Sustaining a lower extremity muscle strain (calf, hamstring, quadriceps, groin)	Random forest Extreme gradient boosting ^a Neural network Support vector machine Elastic net penalized logistic regression Generalized logistic regression	AUROC: 0.840 (XGBoost) Brier score: 0.029 (RF)
Oosterhoff et al ⁴⁰	Presence of a posterior malleolar fracture in patients with tibial shaft fracture	BPM ^a Support vector machine Neural network BDT LR	AUROC: 0.89 (all except BDT) Brier score: 0.11 (all except LR)
Vasavada et al ⁵¹	Recovery of peroneal nerve palsy from MLKI	Random forest	AUROC: 0.64 Accuracy: 75% F1 score: 0.86
Wang et al ⁵²	Painful or pain-free hallux valgus	Support vector machine	Accuracy: 76.4%
Jauhiainen et al ²⁰	Predict moderate and severe knee and ankle injuries	Random forest L1-regularized logistic regression	AUROC: 0.65 (LR)
Ruiz-Pérez et al ⁴⁶	Predict lower extremity noncontact soft tissue injury in futsal players	C4.5 ADTree Support vector machine with SMO ^a K-nearest neighbor Random forest	AUROC: 0.767 Sensitivity: 85.1% Specificity: 62.1%
Hendrickx et al ¹⁷	Presence of a posterior malleolar fracture in patients with tibial shaft fracture	BPM ^a Support vector machine Neural network BDT	AUROC: 0.89 (all except BDT) Brier score: 0.106 (BPM)
Suda et al ⁵⁰	Classify running experience level based on foot-ankle kinematic and kinetic patterns	Support vector machines	Accuracy: 88.5% (less experienced) Precision: 100% (moderately experienced) F1 score: 0.80 (less experienced) Recall: 76.7% (most experienced)
Sharif Bidabadi et al ⁴⁷	Classify foot drop due to L5 radiculopathy	Multilayer perceptron k-nearest neighbor Logistic regression Bayes Net Naive Bayes C4.5 decision tree Random forest Random tree Support vector machine OneR (1R) Deep learning model	Accuracy: 93.18% (RF) AUROC: 0.97 (RF)
Merrill et al ³⁶	1. Morbidity and mortality 2. LOS > 3 d 3. 30-d readmission	Gradient boosting Logistic regression	Accuracy: 85.0% (GB) Sensitivity: 0.57 (LR) Specificity: 0.88 (GB) AUROC: 0.75 (LR)
Yin et al ⁵⁶	Decrease in VAS by at least 60% after extracorporeal shockwave therapy	Artificial neural network	Sensitivity: 95.0% Specificity: 90.0%
Keijsers et al ²⁴	Classify forefoot pain using plantar pressure data	Artificial neural network	Accuracy: 70.4%

Abbreviations: AUROC, area under the receiver operating characteristic curve; BDT, boosting decision tree; BPM, Bayes point machine; GB, gradient boosting; LOS, length of stay; LR, logistic regression; MLKI, multiligamentous knee injury; RF, random forest; SMO, sequential minimal optimization; VAS, visual analog scale.

^aDeemed best model in the study.

Diniz et al¹¹ internally validated one ML model for predicting whether soccer players would return to a similar level of match participation following an Achilles tendon rupture, which achieved a good AUC of 0.81 and Brier score loss of 0.12.

Lu et al³³ internally validated many ML models for predicting the occurrence of a lower extremity muscle strain (calf, groin, quadriceps, hamstring) in professional basketball players, among which the XGBoost model achieved the highest AUC of 0.840 and was deemed the best-performing model when also considering Brier score and calibration. Jauhiainen et al²⁰ internally validated 2 ML models for predicting moderate and severe knee and ankle injuries in young basketball and floorball players (age ≤ 21 years), which performed poorly with an AUC of 0.63 for the random forest model and 0.65 for the logistic regression model. Ruiz-Pérez et al⁴⁶ internally validated many ML models to predict lower extremity noncontact soft tissue injury in elite futsal players, which generally performed fairly, with the best model achieving an AUC of 0.767, sensitivity of 85.1%, and specificity of 62.1%.

Vasavada et al⁵¹ internally validated one random forest model for predicting complete recovery of a peroneal nerve palsy following a multiligamentous knee injury, which performed poorly with an AUC of 0.64, accuracy of 75%, and F1 score of 0.86.

Wang et al⁵² internally validated a support vector machine model for classifying hallux valgus patients as having painful feet or pain-free feet using radiographic metrics such as hallux valgus angle (HVA), intermetatarsal angle (IMA), and distal metatarsal articular angle (DMAA), which performed fair with an accuracy of 76.4%.

Hendrickx et al¹⁷ internally validated 4 ML and DL models for predicting which patients with tibial shaft fractures have an occult posterior malleolar fracture. The models performed good with AUCs ranging from 0.81 to 0.89.

Oosterhoff et al⁴⁰ internally validated 5 models for predicting posterior malleolar involvement in distal tibial shaft fractures using the same data set as that in the previously described study by Hendrickx et al.¹⁶ Oosterhoff et al⁴⁰ found that all the models performed good with AUCs >0.80 (highest 0.89) and 4 of 5 having a Brier score of 0.11.

Suda et al⁵⁰ internally validated several support vector machine models for classifying running experience level based on foot-ankle kinematic and kinetic patterns to potentially assist with running rehabilitation and training. The models performed well with classification accuracies of 88.5% for less experienced runners, 87.2% for moderately experienced runners, and 84.6% for experienced runners.⁵⁰

Merrill et al³⁶ internally validated a logistic regression and gradient boosting model for predicting short-term complications, including readmissions and mortality, following open reduction and internal fixation for ankle fractures. Both models performed similarly, with AUCs for gradient

boosting ranging from 0.6979 to 0.7580 and AUCs for logistic regression ranging from 0.7101 to 0.7583.³⁶

Yin et al⁵⁶ internally validated a neural network model for predicting patients that would achieve the minimum clinically successful therapy (decrease in visual analog score [VAS] by 60% or more from baseline) at 6 months after extracorporeal shock wave therapy for chronic plantar fasciitis. The model performed well, with an overall accuracy of 92.5%, sensitivity of 95.0%, and specificity of 90.0%.⁵⁶

Sharif Bidabadi et al⁴⁷ internally validated many models for classifying gait patterns as normal or due to L5 radiculopathy using data from sensors called inertial measurement units (IMUs). Their best model performed excellently as evidenced by an AUC of 0.97 and accuracy of 93.18%.⁴⁷

Keijsers et al²⁴ internally validated a neural network model for differentiating patients who have forefoot pain and those that do not using plantar pressure data, which performed satisfactorily with an accuracy of 70.4%.

Other

Ardhianto et al² applied DL to help with automated measurement of the foot progression angle (FPA) from plantar pressure images to help clinicians assess gait abnormalities. Pakhomov et al⁴¹ applied ML to automate identification and classification of foot examination findings from clinical notes as normal, abnormal, or not assessed, and their models performed well with overall accuracies ranging from 81% to 87%. Hernigou et al¹⁸ applied AI and ML to assist in conducting their study for developing a method of defining the ideal and patient-specific motion axes of the tibiotalar joint, with the goal of improving how total ankle arthroplasty is performed with robotics. Zhu et al⁵⁷ aimed to assess whether ultrasonography-guided needle knife therapy with AI assistance can improve patient outcomes for plantar fasciitis better than the same therapy without AI. The AI technology used in this study assisted with processing of the ultrasonographic images. Those receiving the intervention with AI had significantly lower plantar fascia thickness, lower plantar fascia elasticity scores, and higher American Orthopaedic Foot & Ankle Society (AOFAS) ankle-hindfoot scores at 2, 4, and 8 weeks posttreatment compared to those without AI assistance.⁵⁷

Discussion

There is early optimism of the transformative impact that AI may have on the health care system and change how we practice medicine. As such, it is necessary for orthopaedic surgeons to be aware of the advancements in AI in their respective areas. This systematic review is the first of its kind in orthopaedic foot and ankle surgery to explore the subject areas in which AI is being applied, the performance

of AI models, and the validity for the AI models. This study found that most studies are using AI for image interpretation, especially for ankle fractures, calcaneus fractures, and hallux valgus. The performance of current AI models is wide ranging, from poor to excellent, but there is significant heterogeneity in study methodologies that prevents any pooled analysis. Additionally, very few studies have externally validated their models.

This systematic review found that most current studies involve AI applications for imaging analysis, particularly fracture identification and classification. This is a common trend seen in other orthopaedic subspecialties as well. For example, in TJA, Karnuta et al²³ externally validated a DL system for classifying hip arthroplasty femoral implants from radiographs that performed excellently with a near perfect AUC of 0.999. Many investigators are likely driven to explore AI's utility in image analysis because of their optimism that AI will outperform or enhance humans in speed and accuracy, translating to potential time-savings, cost-savings, and better patient outcomes.¹³

This systematic review found that models for image interpretation are mostly performing excellent, with 75% (9 of 12) of those reporting accuracy or AUCs achieving a value ≥ 0.90 . In contrary, almost no clinical prediction model (8.33%, 1 of 12 studies) performed excellent (AUC or accuracy ≥ 0.90). There is a need for more research on improving the performance of the clinical prediction models. Many factors, including the quality and size of the data sets, types of models, and how models are optimized, influence model performance and need to be further investigated for foot and ankle surgery. Clinicians can play a vital role in ensuring high-quality data are available to train and test models by helping with accurate data collection, data annotation, and data auditing.

Internal validation may lead to false optimism as it does not allow assessment of how generalizable a model is to other populations, such as those of a different region, age, or insurance status. It has been shown that predictive models often perform significantly worse during external validation.⁴⁸ Thus, external validation is necessary prior to clinical translation of any ML or DL model. None of the clinical prediction studies in this systematic review performed external validation of their models and only 2 of the imaging interpretation studies did. Therefore, it is important that clinicians are aware that most models developed in current foot and ankle surgery studies, although promising, are not yet ready for clinical translation.

Conclusion

AI applications are being increasingly explored in foot and ankle surgery, but most models lack external validation. Most models are being used for image interpretation and are performing excellently in doing so, but model performance

is not robust for clinical predictions. More subject areas need to be explored in foot and ankle surgery, and models with better performance and external validation are needed.

Author Contributions

All authors were involved in conceiving the study, researching background literature, data analysis, and manuscript writing. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Ethical approval

Ethical approval for this study was waived by the Institutional Review Board.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. ICMJE forms for all authors are available online.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Puneet Gupta, BS,  <https://orcid.org/0000-0001-8274-6970>

References

1. Aghnia Farda N, Lai JY, Wang JC, Lee PY, Liu JW, Hsieh IH. Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. *Injury*. 2021;52(3):616-624. doi:10.1016/j.injury.2020.09.010
2. Ardhiyanto P, Subiakto RBR, Lin CY, et al. A deep learning method for foot progression angle detection in plantar pressure images. *Sensors (Basel)*. 2022;22(7):2786. doi:10.3390/s22072786
3. Arvind V, London DA, Cirino C, Keswani A, Cagle PJ. Comparison of machine learning techniques to predict unplanned readmission following total shoulder arthroplasty. *J Shoulder Elbow Surg*. 2021;30(2):e50-e59. doi:10.1016/j.jse.2020.05.013
4. Ashkani-Esfahani S, Mojahed-Yazdi R, Bhimani R, et al. Deep learning algorithms improve the detection of subtle lisfranc malalignments on weightbearing radiographs. *Foot Ankle Int*. 2022;43(8):1118-1126. doi:10.1177/10711007221093574
5. Ashkani-Esfahani S, Mojahed Yazdi R, Bhimani R, et al. Detection of ankle fractures using deep learning algorithms. *Foot Ankle Surg*. 2022;28(8):1259-1265. doi:10.1016/j.fas.2022.05.005
6. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J Arthroplasty*. 2018;33(8):2358-2361. doi:10.1016/j.arth.2018.02.067
7. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K,

- eds. *Artificial Intelligence in Healthcare*. Academic Press; 2020:25-60. doi:10.1016/B978-0-12-818438-7.00002-2
8. Chen Y, Jiang Y. Construction of prediction model of deep vein thrombosis risk after total knee arthroplasty based on XGBoost algorithm. *Comput Math Methods Med*. 2022; 2022:3452348. doi:10.1155/2022/3452348
 9. Day J, de Cesar Netto C, Richter M, et al. Evaluation of a weightbearing CT artificial intelligence-based automatic measurement for the M1-M2 intermetatarsal angle in hallux valgus. *Foot Ankle Int*. 2021;42(11):1502-1509. doi:10.1177/10711007211015177
 10. DelSole EM, Keck WL, Patel AA. The state of machine learning in spine surgery: a systematic review. *Clin Spine Surg*. 2022;35(2):80-89. doi:10.1097/BSD.0000000000001208
 11. Diniz P, Abreu M, Lacerda D, et al. Pre-injury performance is most important for predicting the level of match participation after Achilles tendon ruptures in elite soccer players: a study using a machine learning classifier. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(12):4225-4237. doi:10.1007/s00167-022-07082-4
 12. Federer SJ, Jones GG. Artificial intelligence in orthopaedics: a scoping review. *PLoS One*. 2021;16(11):e0260471. doi:10.1371/journal.pone.0260471
 13. Groot OQ, Bongers MER, Ogink PT, et al. Does artificial intelligence outperform natural intelligence in interpreting musculoskeletal radiological studies? A systematic review. *Clin Orthop Relat Res*. 2020;478(12):2751-2764. doi:10.1097/CORR.0000000000001360
 14. Guermazi A, Tannoury C, Kompel AJ, et al. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology*. 2022;302(3):627-636. doi:10.1148/radiol.210937
 15. Harris AHS, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? *Clin Orthop Relat Res*. 2019;477(2):452-460. doi:10.1097/CORR.0000000000000601
 16. Hendrickx LAM, Cain ME, Sierevelt IN, et al. Incidence, predictors, and fracture mapping of (occult) posterior malleolar fractures associated with tibial shaft fractures. *J Orthop Trauma*. 2019;33(12):e452-e458. doi:10.1097/BOT.0000000000001605
 17. Hendrickx LAM, Sobol GL, Langerhuizen DWG, et al. A machine learning algorithm to predict the probability of (occult) posterior malleolar fractures associated with tibial shaft fractures to guide "malleolus first" fixation. *J Orthop Trauma*. 2020;34(3):131-138. doi:10.1097/BOT.0000000000001663
 18. Hernigou P, Olejnik R, Safar A, Martinov S, Hernigou J, Ferre B. Digital twins, artificial intelligence, and machine learning technology to identify a real personalized motion axis of the tibiotalar joint for robotics in total ankle arthroplasty. *Int Orthop*. 2021;45(9):2209-2217. doi:10.1007/s00264-021-05175-2
 19. Hinterwimmer F, Lazic I, Suren C, et al. Machine learning in knee arthroplasty: specific data are key—a systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(2):376-388. doi:10.1007/s00167-021-06848-6
 20. Jauhiainen S, Kauppi JP, Leppänen M, et al. New machine learning approach for detection of injury risk factors in young team sport athletes. *Int J Sports Med*. 2021;42(2):175-182. doi:10.1055/a-1231-5304
 21. Jo C, Ko S, Shin WC, et al. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. *Knee Surg Sports Traumatol Arthrosc*. 2020;28(6):1757-1764. doi:10.1007/s00167-019-05602-3
 22. Kapiński N, Zieliński J, Borucki BA, et al. Monitoring of the Achilles tendon healing process: can artificial intelligence be helpful? *Acta Bioeng Biomech*. 2019;21(1):103-111.
 23. Karnuta JM, Luu BC, Roth AL, et al. Artificial intelligence to identify arthroplasty implants from radiographs of the knee. *J Arthroplasty*. 2021;36(3):935-940. doi:10.1016/j.arth.2020.10.021
 24. Keijsers NLW, Stolwijk NM, Louwerens JWK, Duysens J. Classification of forefoot pain based on plantar pressure measurements. *Clin Biomech (Bristol, Avon)*. 2013;28(3):350-356. doi:10.1016/j.clinbiomech.2013.01.012
 25. Kim JS, Merrill RK, Arvind V, et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976)*. 2018;43(12):853-860. doi:10.1097/BRS.0000000000002442
 26. Kitamura G, Chung CY, Moore BE 2nd. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. *J Digit Imaging*. 2019;32(4):672-677. doi:10.1007/s10278-018-0167-7
 27. Klemm C, Laurencin S, Alpaugh K, et al. The utility of machine learning algorithms for the prediction of early revision surgery after primary total hip arthroplasty. *J Am Acad Orthop Surg*. 2022;30(11):513-522.
 28. Kunze KN, Polce EM, Clapp I, Nwachukwu BU, Chahla J, Nho SJ. Machine learning algorithms predict functional improvement after hip arthroscopy for femoroacetabular impingement syndrome in athletes. *J Bone Joint Surg Am*. 2021;103(12):1055-1062. doi:10.2106/JBJS.20.01640
 29. Langenberger B, Thoma A, Vogt V. Can minimal clinically important differences in patient reported outcome measures be predicted by machine learning in patients with total knee or hip arthroplasty? A systematic review. *BMC Med Inform Decis Mak*. 2022;22(1):18. doi:10.1186/s12911-022-01751-7
 30. Li T, Wang Y, Qu Y, Dong R, Kang M, Zhao J. Feasibility study of hallux valgus measurement with a deep convolutional neural network based on landmark detection. *Skeletal Radiol*. 2022;51(6):1235-1247. doi:10.1007/s00256-021-03939-w
 31. Li YC, Chen HH, Horng-Shing Lu H, Hondar Wu HT, Chang MC, Chou PH. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? *Clin Orthop Relat Res*. 2021;479(7):1598-1612. doi:10.1097/CORR.0000000000001685
 32. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA. Artificial learning and machine learning decision guidance applications in total hip and knee arthroplasty: a systematic review. *Arthroplast Today*. 2021;11:103-112. doi:10.1016/j.artd.2021.07.012

33. Lu Y, Pareek A, Lavoie-Gagne OZ, et al. Machine learning for predicting lower extremity muscle strain in National Basketball Association athletes. *Orthop J Sports Med.* 2022;10(7):23259671221111742. doi:10.1177/23259671221111742
34. Makhni EC, Makhni S, Ramkumar PN. Artificial intelligence for the orthopaedic surgeon: an overview of potential benefits, limitations, and clinical applications. *J Am Acad Orthop Surg.* 2021;29(6):235-243. doi:10.5435/JAAOS-D-20-00846
35. Merali Z, Wang JZ, Badhiwala JH, Witiw CD, Wilson JR, Fehlings MG. A deep learning model for detection of cervical spinal cord compression in MRI scans. *Sci Rep.* 2021;11(1):10473. doi:10.1038/s41598-021-89848-3
36. Merrill RK, Ferrandino RM, Hoffman R, Shaffer GW, Ndu A. Machine learning accurately predicts short-term outcomes following open reduction and internal fixation of ankle fractures. *J Foot Ankle Surg.* 2019;58(3):410-416. doi:10.1053/j.jfas.2018.09.004
37. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med.* 2020;3:126. doi:10.1038/s41746-020-00333-z
38. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: an introduction for clinicians. *J Bone Joint Surg Am.* 2020;102(9):830-840. doi:10.2106/JBJS.19.01128
39. Olczak J, Emilson F, Razavian A, Antonsson T, Stark A, Gordon M. Ankle fracture classification using deep learning: automating detailed AO Foundation/Orthopedic Trauma Association (AO/OTA) 2018 malleolar fracture identification reaches a high degree of correct classification. *Acta Orthop.* 2021;92(1):102-108. doi:10.1080/17453674.2020.1837420
40. Oosterhoff JHF, Gravesteyn BY, Karhade AV, et al. Feasibility of machine learning and logistic regression algorithms to predict outcome in orthopaedic trauma surgery. *J Bone Joint Surg Am.* 2022;104(6):544-551. doi:10.2106/JBJS.21.00341
41. Pakhomov SVS, Hanson PL, Bjornsen SS, Smith SA. Automatic classification of foot examination findings using clinical notes and machine learning. *J Am Med Inform Assoc.* 2008;15(2):198-202. doi:10.1197/jamia.M2585
42. Pinto Dos Santos D, Brodehl S, Baeßler B, et al. Structured report data can be used to develop deep learning algorithms: a proof of concept in ankle radiographs. *Insights Imaging.* 2019;10(1):93. doi:10.1186/s13244-019-0777-8
43. Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW, Hsieh IH. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. *Comput Methods Programs Biomed.* 2019;171:27-37. doi:10.1016/j.cmpb.2019.02.006.
44. Puijs J, Liao Z, To MS, et al. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN). *Eur J Trauma Emerg Surg.* Published online November 14, 2022. doi:10.1007/s00068-022-02136-1
45. Rowe M. An introduction to machine learning for clinicians. *Acad Med.* 2019;94(10):1433-1436. doi:10.1097/ACM.0000000000002792
46. Ruiz-Pérez I, López-Valenciano A, Hernández-Sánchez S, et al. A field-based approach to determine soft tissue injury risk in elite futsal using novel machine learning techniques. *Front Psychol.* 2021;12:610210. doi:10.3389/fpsyg.2021.610210
47. Sharif Bidabadi S, Murray I, Lee GYF, Morris S, Tan T. Classification of foot drop gait characteristic due to lumbar radiculopathy using machine learning algorithms. *Gait Posture.* 2019;71:234-240. doi:10.1016/j.gaitpost.2019.05.010
48. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol.* 2015;68(1):25-34. doi:10.1016/j.jclinepi.2014.09.007
49. Song BM, Lu Y, Wilbur RR, et al. Machine learning model identifies increased operative time and greater BMI as predictors for overnight admission after outpatient hip arthroscopy. *Arthrosc Sports Med Rehabil.* 2021;3(6):e1981-e1990. doi:10.1016/j.asmr.2021.10.001
50. Suda EY, Watari R, Matias AB, Sacco ICN. Recognition of foot-ankle movement patterns in long-distance runners with different experience levels using support vector machines. *Front Bioeng Biotechnol.* 2020;8:576. doi:10.3389/fbioe.2020.00576
51. Vasavada K, Shankar DS, Bi AS, et al. Predictors using machine learning of complete peroneal nerve palsy recovery after multiligamentous knee injury: a multicenter retrospective cohort study. *Orthop J Sports Med.* 2022;10(9):23259671221121410. doi:10.1177/23259671221121410
52. Wang C, Li C, Zhang R, et al. Identification of radiographic characteristics associated with pain in hallux valgus patients: a preliminary machine learning study. *Front Public Health.* 2022;10:943026. doi:10.3389/fpubh.2022.943026
53. Wang L, Wen D, Yin Y, et al. Musculoskeletal ultrasound image-based radiomics for the diagnosis of achilles tendinopathy in skiers. *J Ultrasound Med.* Published online July 16, 2022. doi:10.1002/jum.16059
54. Wang Y, Li Y, Lin G, et al. Lower-extremity fatigue fracture detection and grading based on deep learning models of radiographs. *Eur Radiol.* 2023;33(1):555-565. doi:10.1007/s00330-022-08950-w
55. Xu W, Shu L, Gong P, et al. A deep-learning aided diagnostic system in assessing developmental dysplasia of the hip on pediatric pelvic radiographs. *Front Pediatr.* 2022;9:785480. <https://www.frontiersin.org/article/10.3389/fped.2021.785480>
56. Yin M, Ma J, Xu J, et al. Use of artificial neural networks to identify the predictive factors of extracorporeal shock wave therapy treating patients with chronic plantar fasciitis. *Sci Rep.* 2019;9(1):4207. doi:10.1038/s41598-019-39026-3
57. Zhu S, Niu Y, Wang J, Xu D, Li Y. Artificial intelligence technology combined with ultrasound-guided needle knife interventional treatment of PF: improvement of pain, fascia thickness, and ankle-foot function in patients. *Comput Math Methods Med.* 2022;2022:3021320. doi:10.1155/2022/3021320