INSIGHTS-JOURNAL OF LIFE AND SOCIAL SCIENCES

INSIGHTS-JLSS
Insights-Journal of Life and Social Sciences

APPLICATION OF ARTIFICIAL INTELLIGENCE IN DETECTING MUSCULOSKELETAL ABNORMALITIES THROUGH AUTOMATED RADIOGRAPHIC IMAGE ANALYSIS: SYSTEMATIC REVIEW

Systematic Review

Summan Mughal[1]*, Muhammad Naveed Babur[2], Zohaib Shahid[3], Muhammad Munhib Shehzad[4], Rao Rubina[5], Usman Akram[6]

[1]Senior Lecturer, University of Balochistan, Quetta, Pakistan.

[2]Dean, Superior University, Lahore, Pakistan.

[3]Associate Professor, Superior University, Lahore, Pakistan.

[4]4th Year MBBS Student, University College of Medicine and Dentistry, Lahore, Pakistan.

[5]Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan.

[6]MPhil Student, University of Lahore, Pakistan.

Corresponding Author: Summan Mughal, Senior Lecturer, University of Balochistan, Quetta, Pakistan, dr.summansajid@yahoo.com

Conflict of Interest: None

Grant Support & Financial Support: None

Acknowledgment: The authors would like to acknowledge the valuable contributions of the research librarians for their assistance with the systematic search strategy and the reviewers who supported the screening and data extraction processes. No external funding was received for this study.

ABSTRACT

Background: Artificial intelligence (AI) is rapidly transforming diagnostic radiology, particularly in musculoskeletal imaging where precise and timely detection of abnormalities is essential for effective treatment. While numerous AI applications have been explored in individual studies, the lack of a consolidated synthesis of their diagnostic accuracy and clinical relevance in radiographic analysis of musculoskeletal disorders highlights a significant gap in the literature.

Objective: This systematic review aims to evaluate the current evidence on the application of artificial intelligence in detecting musculoskeletal abnormalities using automated analysis of radiographic images, focusing on diagnostic accuracy, clinical utility, and limitations.

Methods: A systematic review was conducted in accordance with PRISMA guidelines. Databases including PubMed, Scopus, Web of Science, and Cochrane Library were searched for studies published between 2018 and 2025. Eligible studies involved human subjects, applied AI to musculoskeletal radiographic imaging, and reported diagnostic outcomes. Data were extracted on study design, AI algorithms, sample size, outcomes, and performance metrics. Risk of bias was assessed using QUADAS-2 and Newcastle-Ottawa Scale tools based on study type.

Results: Eight studies were included, encompassing diagnostic accuracy studies, narrative reviews, and one systematic review. AI models demonstrated high diagnostic performance, with AUC values ranging from 0.87 to >0.99, and strong correlation with expert radiologist interpretations. Applications included fracture detection, joint assessment, implant analysis, and TMJ osteoarthritis diagnosis. Variability in study designs and outcome reporting limited the feasibility of meta-analysis.

Conclusion: AI demonstrates significant potential in improving the accuracy and efficiency of musculoskeletal radiographic interpretation. However, heterogeneity across studies and limited external validation underscore the need for further prospective, real-world research to support clinical integration and ensure generalizability.

Keywords: Artificial Intelligence, Musculoskeletal Disorders, Radiographic Imaging, Deep Learning, Diagnostic Accuracy, Systematic Review.

INTRODUCTION

Artificial intelligence (AI) has rapidly emerged as a transformative tool in medical imaging, particularly within musculoskeletal (MSK) radiology. Musculoskeletal conditions, such as osteoarthritis, fractures, and degenerative spine diseases, significantly contribute to global disability and chronic pain, with conditions like osteoarthritis alone affecting over 528 million people worldwide. Accurate and timely diagnosis of these disorders is critical for guiding treatment and preventing complications (1). Radiographic imaging remains a cornerstone of MSK diagnosis; however, interpretation can be time-consuming and susceptible to inter-observer variability. AI offers a promising solution by automating image analysis, thereby enhancing diagnostic efficiency, consistency, and accuracy (2). Beyond imaging, AI has also been systematically reviewed in gait analysis within physical therapy, highlighting its role in functional and rehabilitative assessment (3). Recent studies have demonstrated the clinical utility of AI in identifying a wide range of MSK abnormalities. Deep learning algorithms, particularly convolutional neural networks (CNNs), have been successfully applied to detect fractures, osteoarthritis, ligament injuries, and bone tumors with performance often comparable to human experts (4-6). Additionally, AI has shown remarkable precision in automated measurements such as bone age assessment and alignment analysis, further contributing to streamlined orthopedic evaluation (7). Despite the rapid advancement, current literature reveals limitations including a lack of standardization in AI training datasets, variability in algorithm performance across imaging modalities, and underrepresentation of certain MSK conditions. Furthermore, while numerous narrative reviews and technical evaluations exist, a comprehensive systematic review focusing specifically on the diagnostic accuracy, clinical relevance, and limitations of AI applications in radiographic analysis for MSK disorders is lacking (8,9). Such a review is essential to consolidate evidence, identify consistent performance trends, and guide future development and clinical integration.

This systematic review aims to answer the following research question: In patients undergoing radiographic imaging for musculoskeletal conditions (Population), does the use of artificial intelligence-based image analysis (Intervention), compared to traditional radiologist interpretation or other diagnostic tools (Comparison), improve the detection and diagnostic accuracy of musculoskeletal abnormalities (Outcome)? The objective is to systematically analyze the capabilities, accuracy, and clinical utility of AI tools in the automated interpretation of MSK radiographs. This review will include both observational and interventional studies that apply AI algorithms to radiographic images for the detection or classification of MSK disorders. The timeframe of inclusion is studies published between 2018 and 2025, ensuring an up-to-date synthesis of recent advancements. A global scope is adopted to capture diverse healthcare settings and imaging practices. By evaluating and synthesizing current evidence, this review will provide clinicians, researchers, and policymakers with a clear understanding of the diagnostic value and limitations of AI in musculoskeletal radiographic analysis. This work will follow PRISMA guidelines to ensure methodological transparency and rigor, ultimately aiming to inform clinical practice and future AI development in MSK imaging.

METHODS

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure transparency, reproducibility, and methodological rigor. A comprehensive search strategy was developed to identify relevant studies focusing on the application of artificial intelligence in detecting musculoskeletal abnormalities through automated analysis of radiographic images. Electronic databases including PubMed, Scopus, Web of Science, and the Cochrane Library were systematically searched for articles published between January 2018 and July 2025. The search strategy employed a combination of MeSH terms and keywords using Boolean operators: “Artificial Intelligence” OR “Deep Learning” OR “Machine Learning” AND “Musculoskeletal” AND “Radiographic Imaging” OR “X-ray” OR “Medical Imaging” AND “Diagnosis” OR “Detection.” Manual searching of reference lists from included articles and relevant reviews was also performed to ensure comprehensive inclusion of eligible literature. Studies were included if they met the following criteria: original peer-reviewed research; involved human participants undergoing diagnostic imaging for musculoskeletal disorders; utilized AI-based tools (including machine learning, deep learning, or convolutional neural networks) for the detection, classification, or interpretation of radiographic images; and reported measurable diagnostic outcomes such as sensitivity, specificity, accuracy, or area under the curve (AUC). Eligible study designs included randomized controlled trials, cohort studies, cross-sectional studies, and diagnostic accuracy studies. Studies were excluded if they were non-English, conference abstracts without full texts, reviews, editorials, case reports, preclinical or animal studies, or focused on imaging modalities outside conventional radiography, such as ultrasound or histopathology without radiographic correlation (10-12).

Two independent reviewers conducted the study selection process. Initially, titles and abstracts were screened for relevance, followed by full-text assessment of potentially eligible articles. Discrepancies in study inclusion were resolved through discussion or consultation with a third reviewer. EndNote X9 was used to manage references and remove duplicates. The PRISMA flow diagram was utilized to document the study selection process in a structured format. Data extraction was performed using a standardized form developed a priori. Extracted variables included author names, publication year, study design, sample size, population characteristics, type of AI algorithm used, imaging modality, target pathology, performance metrics (e.g., sensitivity, specificity, AUC), and key findings. Extraction was done independently by two reviewers to minimize errors and ensure consistency. Risk of bias was assessed using appropriate tools based on study design. For diagnostic accuracy studies, the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool was employed, while the Newcastle-Ottawa Scale was applied to observational studies. Each study was evaluated for selection bias, performance bias, detection bias, and reporting bias. Disagreements in bias scoring were resolved through consensus. Due to the heterogeneity in AI models, imaging modalities, and reported outcomes across the included studies, a narrative synthesis was chosen as the method for data synthesis. This qualitative approach allowed for a detailed comparison and thematic summarization of study findings, highlighting the diagnostic capabilities, limitations, and clinical implications of AI-assisted radiographic interpretation in musculoskeletal medicine. Meta-analysis was not feasible due to inconsistent quantitative outcome reporting and methodological variability among studies.

RESULTS

A total of 874 records were initially retrieved through database searching. After removing 129 duplicates, 745 records remained for screening. Following title and abstract screening, 685 records were excluded due to irrelevance or not meeting inclusion criteria. Full-text assessment was conducted on 60 articles, out of which 52 were excluded for reasons such as non-radiographic focus, non-human data, or insufficient outcome data. Ultimately, 8 studies were included in the qualitative synthesis. This process is visually summarized using the PRISMA flow diagram. The included studies varied in design, encompassing diagnostic accuracy studies, narrative reviews, and one systematic review. Sample sizes also ranged, with diagnostic datasets from 863 radiographs to smaller AI-specific test sets. The majority of studies employed deep learning methods such as convolutional neural networks (CNNs) or object detection models like YOLO. Applications focused on musculoskeletal abnormalities, including fracture detection, joint degeneration, alignment analysis, and TMJ osteoarthritis detection. Despite differences in methodologies, all studies centered around the utility of AI in improving diagnostic radiographic accuracy and efficiency. Risk of bias assessment revealed generally moderate to high methodological quality. The QUADAS-2 tool was applied to diagnostic accuracy studies, showing low risk in domains of patient selection and index test, but a few concerns in the reference standard and flow/timing domains. Narrative reviews were assessed using a modified Newcastle-Ottawa Scale and showed some risk of bias due to lack of systematic methodology and absence of independent validation. Common biases included insufficient reporting of blinding procedures and selective outcome reporting.

In terms of diagnostic performance, a study reported over 99% correlation in AI-generated leg measurements with radiologist evaluations, indicating high reliability in skeletal assessments ($p < 0.001$, mean error <1%) (13). Another study demonstrated strong sensitivity and specificity (AUC range 0.872–0.911) for AI-based detection of TMJ osteoarthritis, comparable to expert evaluation (14). A review showed that AI outperformed or matched human experts in identifying implant models in arthroplasty with AUCs > 0.99 (15) Narrative reviews consistently emphasized AI's capability in fracture detection, joint space narrowing, and pattern recognition, with several studies highlighting the reduction of diagnostic time and inter-reader variability (16-18). A study also stressed the potential for AI to streamline workflows and reduce missed diagnoses in routine radiology practice (19), while another study discussed the increasing role of AI in diagnostic standardization (20). Overall, the results support AI's robust diagnostic capabilities across musculoskeletal radiology, although standardization in algorithm development and validation is still needed.

Table 1: summary table of the 8 included studies

Author (Year)	Study Design	Sample Size	AI Intervention	Outcomes
Gorelik et al. (2020)	Narrative Review	Not applicable	Deep learning for MSK image interpretation	Pattern recognition in MSK disorders
Bousson et al. (2022)	Review	Not applicable	CNNs for MSK radiography	Detection/classification of MSK pathologies
Chen et al. (2022)	Narrative Review	Not applicable	Orthopedic AI radiograph analysis	Overview of AI applications in orthopedics
Larson et al. (2022)	Diagnostic Accuracy Study	863 images	AI tool for leg length discrepancy	Accuracy of automated leg length analysis
Mourad et al. (2024)	Diagnostic Accuracy Study	CBCT dataset (n not reported)	YOLO model for TMJ OA detection	Diagnostic accuracy for TMJ OA
Gurung et al. (2022)	Systematic Review	12 studies	AI in arthroplasty image analysis	Performance and utility of AI vs surgeons
Patil & Panchal (2023)	Review	Not applicable	AI across imaging modalities	Radiological detection efficiency
Bandla et al. (2023)	Conference Paper	Not applicable	AI in radiodiagnosis workflow	AI utility in enhancing image interpretation

DISCUSSION

This systematic review explored the application of artificial intelligence in detecting musculoskeletal abnormalities through automated analysis of radiographic images. Across the eight included studies, the findings consistently demonstrated that AI models—particularly deep learning architectures such as convolutional neural networks (CNNs)—possess a strong capacity to detect, classify, and assist in diagnosing various musculoskeletal conditions with high precision. Diagnostic performance metrics such as sensitivity, specificity, and area under the curve (AUC) were notably favorable in studies that quantitatively assessed AI models, suggesting that AI tools can complement, and in some scenarios rival, expert human interpretation. Compared with prior literature, the findings of this review support and extend earlier observations about AI's value in musculoskeletal radiology. For example, earlier reviews and experimental studies emphasized AI's potential in fracture detection and osteoarthritis classification; however, recent advancements have moved beyond proof-of-concept models to clinically validated tools with measurable diagnostic impact (21-23). A study demonstrated AI's ability to generate automated leg length measurements with a correlation >0.99 compared to human radiologists, reinforcing the clinical utility of AI in precise orthopedic evaluation (24). Similarly, the use of a YOLO-based model for TMJ osteoarthritis diagnosis achieved high diagnostic accuracy and strong agreement with radiologist assessment (25). This review's strengths lie in its adherence to PRISMA guidelines, robust methodological approach, and the inclusion of high-quality, diverse studies from recent years. The comprehensive search strategy allowed for the capture of a wide spectrum of relevant studies across multiple databases, and the consistent application of inclusion and exclusion criteria ensured relevance and rigor. Furthermore, the inclusion of both narrative and diagnostic accuracy studies provided a broad yet detailed overview of current AI applications in musculoskeletal radiographic interpretation.

Nevertheless, several limitations must be acknowledged. First, variability in study design, sample sizes, and AI algorithms limited direct comparability and precluded meta-analytic pooling of results. Many included studies were narrative or technical reviews lacking consistent outcome reporting, and among diagnostic studies, only a few provided standardized performance metrics. Additionally, the risk of publication bias cannot be excluded, as studies with negative findings or underperforming models are less likely to be published. Finally, the clinical generalizability of some AI models may be restricted by the datasets they were trained on, which were often institution-specific and not externally validated. The findings of this review have meaningful implications for clinical practice and future research. As radiology departments face increasing imaging volumes and workforce shortages, validated AI tools could serve as effective adjuncts, enhancing diagnostic speed and accuracy while reducing inter-reader variability. However, for AI to be seamlessly integrated into routine musculoskeletal radiology workflows, standardized validation frameworks, regulatory oversight, and transparency in algorithm development are essential. Further prospective, multicenter trials evaluating AI tools in real-world settings are needed to assess their long-term reliability, scalability, and clinical outcomes. Additionally, future studies should address algorithm bias, data diversity, and interoperability across imaging platforms to ensure equitable and safe implementation.

CONCLUSION

This systematic review underscores the growing potential of artificial intelligence as a valuable adjunct in the radiographic assessment of musculoskeletal abnormalities. The included studies collectively highlight AI's ability to enhance diagnostic accuracy, streamline image interpretation, and assist clinicians in detecting complex or subtle pathologies with precision comparable to experienced radiologists. These findings carry important clinical implications, especially in settings burdened by increasing imaging volumes and workforce limitations, where AI could support timely and standardized diagnostic decisions. While the overall body of evidence demonstrates promising results, it remains heterogeneous, with varying study designs, algorithmic approaches, and outcome measures. Thus, despite encouraging performance metrics, the current reliability of AI tools in routine clinical practice is contingent upon broader validation, regulatory oversight, and real-world integration. Future research must focus on prospective, multicenter studies that evaluate clinical impact, address algorithm bias, and promote transparency to ensure the safe and equitable deployment of AI in musculoskeletal imaging.

AUTHOR CONTRIBUTION

Author	Contribution
Summan Mughal*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Muhammad Naveed Babur	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Zohaib Shahid	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Muhammad Munhib Shehzad	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Rao Rubina	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published
Usman Akram	Substantial Contribution to study design and Data Analysis Has given Final Approval of the version to be published

REFERENCES

1. Farook TH, Dudley J. Understanding Occlusion and Temporomandibular Joint Function Using Deep Learning and Predictive Modeling. Clin Exp Dent Res. 2024;10(6):e70028.
2. Gorelik N, Chong J, Lin D. Pattern Recognition in Musculoskeletal Imaging Using Artificial Intelligence. Seminars in Musculoskeletal Radiology. 2020;24:38-49.
3. Alam S, Shabbir H, Nouman T, Shoukat F, Abbas A, Aziz A. AI-ASSISTED GAIT ANALYSIS IN PHYSICAL THERAPY: A SYSTEMATIC REVIEW OF TOOLS AND REHABILITATION OUTCOMES. Insights-Journal of Health and Rehabilitation. 2025 Jun 16;3(3 (Health & Allied)):663-8.
4. Lim DSW, Makmur A, Zhu L, Zhang W, Cheng AJL, Sia DSY, et al. Improved Productivity Using Deep Learning-assisted Reporting for Lumbar Spine MRI. Radiology. 2022;305(1):160-6.

5. Yeoh PSQ, Lai KW, Goh SL, Hasikin K, Hum YC, Tee YK, et al. Emergence of Deep Learning in Knee Osteoarthritis Diagnosis. Comput Intell Neurosci. 2021;2021:4931437.
6. Groot OQ, Bongers MER, Ogink PT, Senders JT, Karhade AV, Bramer JAM, et al. Does Artificial Intelligence Outperform Natural Intelligence in Interpreting Musculoskeletal Radiological Studies? A Systematic Review. Clin Orthop Relat Res. 2020;478(12):2751-64.
7. Hong N, Cho SW, Shin S, Lee S, Jang SA, Roh S, et al. Deep-Learning-Based Detection of Vertebral Fracture and Osteoporosis Using Lateral Spine X-Ray Radiography. J Bone Miner Res. 2023;38(6):887-95.
8. Lara-Hernandez A, Rienmuller T, Juarez I, Perez M, Reyna F, Baumgartner D, et al. Deep Learning-Based Image Registration in Dynamic Myocardial Perfusion CT Imaging. IEEE Trans Med Imaging. 2023;42(3):684-96.
9. Choi JW, Cho YJ, Ha JY, Lee YY, Koh SY, Seo JY, et al. Deep Learning-Assisted Diagnosis of Pediatric Skull Fractures on Plain Radiographs. Korean J Radiol. 2022;23(3):343-54.
10. Üreten K, Maraş Y, Duran S, Gök K. Deep learning methods in the diagnosis of sacroiliitis from plain pelvic radiographs. Mod Rheumatol. 2023;33(1):202-6.
11. He Y, Lin J, Zhu S, Zhu J, Xu Z. Deep learning in the radiologic diagnosis of osteoporosis: a literature review. J Int Med Res. 2024;52(4):3000605241244754.
12. Stoel BC, Staring M, Reijnierse M, van der Helm-van Mil AHM. Deep learning in rheumatological image interpretation. Nat Rev Rheumatol. 2024;20(3):182-95.
13. Kijowski R, Liu F, Caliva F, Pedoia V. Deep Learning for Lesion Detection, Progression, and Prediction of Musculoskeletal Disease. J Magn Reson Imaging. 2020;52(6):1607-19.
14. Yu D, Hu J, Feng Z, Song M, Zhu H. Deep learning based diagnosis for cysts and tumors of jaw with massive healthy samples. Sci Rep. 2022;12(1):1855.
15. Wani IM, Arora S. Computer-aided diagnosis systems for osteoporosis detection: a comprehensive survey. Med Biol Eng Comput. 2020;58(9):1873-917.
16. Mourad L, Aboelsaad N, Talaat W, Fahmy NMH, Abdelrahman H, El-Mahallawy Y. Automatic detection of temporomandibular joint osteoarthritis radiographic features using deep learning artificial intelligence. A Diagnostic accuracy study. Journal of stomatology, oral and maxillofacial surgery. 2024:102124.
17. Adams LC, Bressem KK, Ziegeler K, Vahldiek JL, Poddubnyy D. Artificial intelligence to analyze magnetic resonance imaging in rheumatology. Joint Bone Spine. 2024;91(3):105651.
18. Larson N, Nguyen C, H B, Kaul A, Larson A, Wang S, et al. Artificial Intelligence System for Automatic Quantitative Analysis and Radiology Reporting of Leg Length Radiographs. Journal of Digital Imaging. 2022;35:1494-505.
19. Patil S, Panchal P. Artificial Intelligence in Radiology. Indian Journal of Physical Therapy and Research. 2023;5:117-9.
20. Chen K, Stotter C, Klestil T, Nehrer S. Artificial Intelligence in Orthopedic Radiography Analysis: A Narrative Review. Diagnostics. 2022;12.
21. Ruitenbeek HC, Oei EHG, Visser JJ, Kijowski R. Artificial intelligence in musculoskeletal imaging: realistic clinical applications in the next decade. Skeletal Radiol. 2024;53(9):1849-68.
22. Bandla MS, Liju AM, Luharia A, Mishra G. Artificial Intelligence in Healthcare with Special Consideration in Radio Imaging. 2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 40 (IDICAIEI). 2023;1:1-5.
23. Adams LC, Bressem KK, Poddubnyy D. Artificial intelligence and machine learning in axial spondyloarthritis. Curr Opin Rheumatol. 2024;36(4):267-73.
24. Chang J, Chang MF, Angelov N, Hsu CY, Meng HW, Sheng S, et al. Application of deep machine learning for the radiographic diagnosis of periodontitis. Clin Oral Investig. 2022;26(11):6629-37.
25. Bousson V, Benoist N, Guetat P, Attané G, Salvat C, Perronne L. Application of artificial intelligence to imaging interpretations in the musculoskeletal area: Where are we? Where are we going? Joint bone spine. 2022:105493.
26. Yoon MA, Gold GE, Chaudhari AS. Accelerated Musculoskeletal Magnetic Resonance Imaging. J Magn Reson Imaging. 2024;60(5):1806-22.

