Content available at: https://www.ipinnovative.com/open-access-journals

IP International Journal of Orthopaedic Rheumatology

Journal homepage: https://www.ijor.org/

Review Article

Development and validation of AI-driven radiographic biomarkers for predicting orthopedic outcomes

Shivam Dubey[1]*

[1]Rani Durgavati Vishwavidyalaya, Jabalpur, Madhya Pradesh, India

Abstract

The increasing prevalence of degenerative and post-traumatic musculoskeletal disorders has amplified the demand for accurate prognostic tools to guide orthopedic decision-making. Radiographs remain the most accessible imaging modality worldwide, yet traditional radiographic grading systems (e.g., Kellgren–Lawrence) are limited by subjectivity and relatively poor predictive power for long-term outcomes. Recent advances in artificial intelligence (AI), radiomics, and deep learning now enable extraction of quantitative, reproducible radiographic biomarkers that may predict important orthopedic outcomes — such as progression to joint replacement, implant sizing accuracy, risk of revision, or functional recovery. This review critically examines the current state of development and validation of AI-driven radiographic biomarkers in orthopedics, summarizing candidate biomarker classes, methodological pipelines, validation strategies, reporting standards, and challenges for clinical translation. We emphasize the need for rigorous external validation, transparent reporting (e.g., via TRIPOD+AI and CLAIM), and demonstration of clinical utility to ensure safe and equitable adoption.

Keywords: AI, Deep learning, Radiomics, Radiographic biomarker, Joint replacement, Knee osteoarthritis, Tripod-AI.

Received: 10-10-2025-2025; **Accepted:** 17-11-2025; **Available Online**: 10-12-2025

1. Introduction

Musculoskeletal disorders remain a major contributor to global disability, with degenerative joint diseases such as osteoarthritis (OA) accounting for a substantial proportion of the orthopedic disease burden worldwide.[1] Radiography continues to serve as the first-line imaging modality for evaluating structural joint changes because of its affordability, accessibility, and established clinical utility. However, traditional radiographic grading frameworks—most notably the Kellgren–Lawrence (KL) system—provide only semi-quantitative assessments and exhibit substantial inter-observer variability, limiting their prognostic utility for long-term orthopedic outcomes.[2] Recent advancements in medical imaging analytics have catalyzed a paradigm shift. Artificial intelligence (AI), encompassing radiomics and deep learning, enables automated extraction of complex, high-dimensional image features that far exceed human visual capacity. These AI-derived radiographic biomarkers have shown potential to characterize subtle bone texture, trabecular morphology, joint space morphology, and shape-based features linked to disease severity and progression.[3,4] For example, deep learning systems have demonstrated robust performance in grading knee OA severity and identifying structural abnormalities predictive of disease progression toward total knee replacement (TKR).[5,6]

Radiographs offer a unique advantage for AI-driven biomarker development: vast repositories of routinely acquired images, often coupled with long-term clinical outcomes. This facilitates training of prediction models for clinically meaningful endpoints such as risk of conversion to arthroplasty, likelihood of implant failure or revision, and postoperative functional recovery. Studies leveraging textural radiomics and convolutional neural networks have shown that radiographic features can predict TKR several years before clinical decision-making, outperforming traditional radiographic measures.[6] Despite these promising developments, significant methodological and translational challenges remain. Many published models rely on limited datasets, insufficient external validation, or inconsistent reporting of technical details. To ensure real-world applicability, prediction models must adhere to recognized

Corresponding author: Shivam Dubey
Email: shivamdubey20@gmail.com

https://doi.org/10.18231/j.ijor.17450.1765359123

reporting frameworks such as TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis)[7] and the Checklist for Artificial Intelligence in Medical Imaging (CLAIM).[8] These frameworks emphasize reproducibility, transparency, generalizability, and avoidance of bias—key prerequisites for successful clinical adoption. This review synthesizes current evidence on the development and validation of AI-driven radiographic biomarkers for predicting orthopedic outcomes. Key aims include:

1. Classify current radiographic biomarkers made possible by AI;
2. Detailing model development pipelines;
3. Validation practices evaluation;
4. To assess reproducibility and reporting standards; and
5. Barriers and future directions necessary for robust clinical translation.

2. Classification of AI-Driven Radiographic Biomarkers

AI-driven radiographic biomarkers can be broadly classified into four major categories based on the type of imaging information extracted and the analytic methodology used. These include: (1) handcrafted radiomic biomarkers, (2) deep-learning–derived biomarkers, (3) hybrid biomarkers that integrate radiomics and deep learning, and (4) shape, geometry, and biomechanical surrogate biomarkers obtained through automated extraction. These categories represent complementary approaches for quantifying radiographic structure and predicting clinically significant orthopedic outcomes.

2.1. Handcrafted radiomic biomarkers

Radiomics refers to extraction of predefined, mathematically engineered quantitative features from radiographs. These features commonly capture texture, intensity, shape, and spatial heterogeneity of bone and joint structures. Textural features such as gray-level co-occurrence matrix (GLCM) homogeneity, entropy, fractal dimension, and run-length metrics have shown significant correlation with OA severity and progression toward arthroplasty.[9,10] In orthopedic prognostication, radiomic analysis of the subchondral bone plate and trabecular texture has been particularly informative. Studies evaluating radiomic signatures in knee radiographs have demonstrated that bone-texture patterns can predict future joint-space narrowing and the likelihood of total knee replacement (TKR), often outperforming Kellgren–Lawrence grades and simple anatomical measurements.[11] These biomarkers are explainable, relatively stable, and computationally efficient, making them attractive for large-cohort analyses and multi-site validation.

2.2. Deep learning–derived biomarkers

Deep learning, especially convolutional neural networks (CNNs), enables automated extraction of high-dimensional features directly from raw radiographs without predefined mathematical descriptors. These deep features encode complex radiographic attributes—such as global joint morphology, microstructural intensity patterns, and hierarchical shape characteristics—representing latent biomarkers predictive of orthopedic outcomes. Deep-learning models trained on longitudinal radiograph datasets have shown strong ability to predict disease progression and surgical conversion. For example, CNN-based models have achieved high accuracy in forecasting TKR within 5–7 years by learning multi-scale image representations beyond human-interpretable features.[12] In addition, attention-based architectures such as Vision Transformers (ViT) and hybrid CNN-transformers have recently demonstrated improved sensitivity for detecting early pathological cues in bone and cartilage regions.[13] These biomarkers remain less interpretable than radiomic descriptors, necessitating post-hoc explainability methods (e.g., Grad-CAM, integrated gradients). However, their predictive strength and capacity to capture complex radiographic patterns make them essential for state-of-the-art modeling.

2.3. Hybrid biomarkers-radiomics + Deep learning

Hybrid approaches include the interpretability of handcrafted radiomics with the representational power of deep learning. This could be achieved through:

1. Feature fusion, where radiomic and deep features are concatenated before either classification or regression, or
2. Two-stream models, where both pipelines run in parallel and outputs are integrated.

Such hybrid biomarker systems have shown improved prognostic performance in musculoskeletal imaging, particularly when datasets are moderately sized—benefiting from radiomics' stability and deep learning's abstraction capacity.[14] This approach is especially suited for predicting multifactorial outcomes such as revision risk or postoperative functional recovery, where multiple radiographic pathways contribute to disease trajectory.

2.4. Shape, geometry, and biomechanical surrogate biomarkers

AI methods can automatically extract geometric features such as joint-space width (JSW), limb alignment angles, osteophyte morphology, femoral/tibial contour metrics, and bone curvature parameters. These computationally derived geometric biomarkers have shown strong association with surgical indications and outcomes. For instance, automated alignment measurements from knee radiographs have been linked to TKR prediction and implant selection accuracy.[15] Similarly, AI-based quantification of hip morphology (e.g., center–edge angle, alpha angle) can forecast progression of femoroacetabular impingement and subsequent orthopedic intervention. Additionally, biomechanical surrogate biomarkers—such as estimated cartilage load distribution or

bone mineral surrogates derived from radiograph texture—are emerging. These biomarkers approximate physiologic stress or degeneration patterns that are otherwise not directly measurable from radiographs, offering a new dimension for risk modelling.[16]

2.5. Biomarker categories aligned to orthopedic outcomes

The integration of these biomarker types into robust predictive frameworks provides a basis for the next generation of precision orthopedics. (Table 1)

Various types of AI-driven radiographic biomarkers have shown utility for predicting different orthopedic outcomes:

Orthopaedic Outcome	Effective Biomarker Types
Progression to TKR	Radiomics; Deep-learning features; Shape biomarkers
Implant failure or revision	Geometric biomarkers; Hybrid models
Postoperative functional recovery	Deep learning; Hybrid fusion markers
Early detection of degenerative changes	Radiomic texture; CNN latent features
Fracture risk prediction	Texture-based radiomics; Bone-structure radiomics

3. Development Pipeline for AI-Driven Radiographic Biomarkers

The development of AI-driven radiographic biomarkers for predicting orthopedic outcomes follows a structured multi-stage pipeline. Although specific workflows vary across studies, most adhere to core methodological phases:

1. Data acquisition and cohort design,
2. Image preprocessing and annotation,
3. Biomarker extraction,
4. Model training and optimization,
5. Validation and performance evaluation, and
6. Clinical utility assessment.

Adherence to rigorous pipelines is essential for generating reliable, reproducible, and clinically meaningful biomarkers.

3.1. Data acquisition and cohort design

The foundation of biomarker development is a sufficiently large, diverse, and well-characterized dataset. Retrospective radiograph archives, multi-centre registries, and prospective observational cohorts serve as primary data sources. Large datasets such as the Osteoarthritis Initiative (OAI) have been instrumental for training predictive models due to their inclusion of longitudinal radiographs and robust outcome labels.[17] The cohort design should prioritize:

1. Sufficient sample size to avoid model underfitting
2. Outcome clarity: Progression to TKR, revision surgery, alignment correction, and functional score change among others
3. Longitudinal follow-up for the development of prognostic biomarkers
4. Representation across demographics and disease severity reduces bias and limits overfitting.
5. High-quality ground truth outcomes, such as surgeon-confirmed arthroplasty, imaging-based progression indices, or validated functional scores are fundamental for the reliable development of models.

3.2. Image pre-processing and annotation

Standardized preprocessing improves the generalizability of models by reducing image variability from acquisition differences. Common steps include:

1. Intensity normalization and histogram equalization
2. Rescaling or standardizing pixel spacing
3. ROI extraction including automated segmentation of the knee joint, hip joint, or vertebrae
4. Artefact removal: markers, noise, and exposure variation

Accurate annotation is critical. Labels for tasks such as grade of joint-space narrowing, osteophyte severity, implant loosening, and fracture classification are often obtained from radiologists or expert clinicians. Semi-automated annotation tools and AI-assisted segmentation systems are increasingly employed to automate this process.[19]

3.3. Biomarker extraction

This may involve the following, depending on the methodological approach:

1. Radiomic Feature Extraction: Quantitative features such as texture, intensity distribution, and shape descriptors are extracted from ROIs. Features must be checked for stability through test-retest analysis and inter-reader variability to ensure the features are reproducible.[20]
2. Deep Learning Feature Extraction: CNNs or Vision Transformers learn hierarchical features from entire radiographs or targeted ROIs in a self-supervised manner, resulting in latent biomarkers for subtle structural changes that cannot be captured by manual interpretation alone.[21]
3. Hybrid Extraction: Integrating deep-learning embedding into radiomic signatures provides a higher predictive granularity, especially for complex orthopedic endpoints.

3.4. Model training, optimization and calibration

Computer algorithms utilised for orthopedic outcome prediction include:

1. Random forests and gradient boosting machines - Radiomics

2. Convolutional neural networks
3. Vision Transformers (ViT)
4. Hybrid deep-radiomic ensembles
5. Multi-task learning models that jointly predict severity and progression
6. Class imbalance: this often occurs in progression or revision datasets, and must be balanced through oversampling, focal loss, or cost-sensitive learning. Hyperparameter tuning is done through cross-validation or Bayesian optimization.

Model calibration ensures that predicted probabilities-e.g., risk of conversion to TKR-reflect real-world outcome frequencies, an important requirement for clinical usability.

3.5. Internal, temporal, and external validation

Robust validation delineates clinically reliable biomarkers from overfitted models. Stages of validation recommended include:

1. Internal validation: using k-fold cross-validation or bootstrapping
2. Temporal validation: application of the model to later-acquired radiographs, which enables evaluation for robustness against time-related drift
3. External validation using datasets from various hospitals, different geographic regions, or different imaging protocols. Indeed, external validation is considered the most powerful determinant of real-world performance and thus essential before clinical translation.[23]

3.6. Clinical utility, interpretability, and decision curve analysis

Biomarkers have to show practical clinical value for adoption in orthopedics beyond just statistical accuracy. Key components of evaluation include:

1. Interpretability through heat maps, feature importance scores, or attention visualization
2. Decision curve analysis to evaluate the clinical benefit over risk thresholds
3. Subgroup analysis ensuring fair performance across age groups, sex, body habitus, and radiographic severity
4. Integration feasibility with PACS, EMRs, and surgical planning pipelines
5. Randomized diagnostic impact studies represent the next step in the evaluation of real-world utility for AI-driven radiographic biomarkers.

4. Clinical Applications of AI-Driven Radiographic Biomarkers in Orthopedics

AI-enabled radiographic biomarkers have now started to change clinical decision-making in many orthopedic subspecialties. Their value comes from the extraction of quantitative imaging features, often invisible to the human eye, correlating strongly with disease progression, structural integrity, implant performance, and functional outcomes.

4.1. Osteoarthritis progression and treatment response prediction

AI-derived imaging biomarkers from plain radiographs, MRI, and CT have shown robust capability in forecasting structural deterioration in knee, hip, and hand osteoarthritis. Deep learning models that quantify joint-space narrowing, trabecular texture, subchondral bone shape, and osteophyte morphology offer significantly higher predictive accuracy compared to traditional Kellgren–Lawrence grading alone.[18,19] Additionally, machine learning–derived bone shape vectors and cartilage surface biomarkers have demonstrated value in identifying "fast progressors," enabling earlier intervention and optimized treatment selection.[20] Radiomics models integrated with clinical variables can further enhance prediction of response to intra-articular corticosteroids, hyaluronic acid injections, and conservative management strategies.[21]

4.2. Trauma and fracture risk stratification

AI-driven radiographic features help predict fracture risk, treatment failure, and postoperative complications. Texture-based and shape-based radiomic markers extracted from DXA scans, radiographs, or CT can quantify bone fragility beyond conventional bone mineral density (BMD) measurements.[22] In acute trauma care, AI biomarkers have shown potential in estimating fracture displacement risk, monitoring healing trajectories, and predicting delayed union or non-union based on early postoperative radiographs.[23] These models provide objective early indicators for clinical decision-making, especially in high-risk fractures such as tibial shaft, scaphoid, and proximal femoral fractures.

4.3. Spine degeneration and surgical outcome prediction

AI biomarkers extracted from lumbar and cervical spine imaging—such as disc height metrics, Modic change quantification, vertebral endplate morphology, and radiomic signatures of neural foraminal narrowing—are emerging as reliable predictors of postoperative outcomes in spine surgery.24 Machine learning models can predict the likelihood of persistent radiculopathy, hardware complications, or re-operation following lumbar fusion or decompression procedures, supporting personalized surgical planning.[25] Predictive radiographic biomarkers are also being integrated into spine triage algorithms to differentiate patients requiring surgery versus conservative therapy.[26]

4.4. Joint replacement prognosis and implant performance assessment

AI-enabled imaging biomarkers are increasingly used to assess implant alignment, periprosthetic bone adaptation, risk of aseptic loosening, and long-term survivorship after knee, hip, and shoulder arthroplasty. The deep learning models can automatically quantify component positioning, radiolucent

lines, and migration patterns with accuracy comparable to RSA but using conventional radiographs.[27] Predictive models leveraging pre-operative imaging biomarkers have demonstrated abilities in predicting post-operative functional outcomes, patient-reported satisfaction scores, and risk of complications such as stiffness, instability, or infection.[28]

4.5. Inflammatory and metabolic bone disorders

AI-driven radiographic biomarkers are being explored for early detection and monitoring of ankylosing spondylitis, rheumatoid arthritis, and metabolic bone diseases. Deep learning–based erosion scoring systems using radiographs have shown excellent reproducibility and sensitivity in tracking disease progression in rheumatoid arthritis, significantly reducing inter-observer variation.[29] Similarly, automated quantification of syndesmophytes, vertebral squaring, and sacroiliitis markers supports precision disease monitoring in axial spondyloarthritis and enables earlier therapeutic decision-making.

5. Validation Frameworks for AI-Driven Radiographic Biomarkers

Robust validation is necessary for AI-derived radiographic biomarkers to be clinically reliable, reproducible, and generalizable across populations, imaging systems, and clinical contexts. Structured, multilayered validation frameworks help make the transition from algorithmic discovery to the real world in orthopedic applications.

5.1. Technical validation

Technical validation focuses on whether the AI system performs consistently and accurately under different conditions of imaging. Key elements here are:

1. Internal validation (Train–Test Splits, Cross-Validation): This internal validation involves k-fold cross-validation or bootstrapping analysis on the same dataset to determine algorithmic stability. This will mitigate overfitting and ensure that the biomarkers extracted capture true radiographic patterns and not dataset-specific noise.[30]
2. External validation across institutions and imaging vendors: External validation on independent data coming from various scanners, protocols, and patients' demographics is important in establishing generalizability. AI biomarkers indeed degrade in performance across institutions as a result of variation in contrast, positioning, acquisition parameters, and calibration standards.[31] Models validated across ≥3 centres show significantly higher reproducibility and lower bias.
3. Robustness and stress testing: Stress testing evaluates model resilience under noisy, incomplete, rotated, or low-resolution images that reflect real-world radiographic imperfections. This includes various techniques like adversarial perturbation analysis, synthetic noise injection, and domain-shift simulation to quantify technical robustness of imaging biomarkers.[32]

5.1.1. Clinical validation

Clinical validation is necessary to ensure that AI biomarkers correlate with real clinical outcomes and offer added value beyond existing scoring systems.

1. Association with established clinical endpoints: Biomarkers should be strongly and statistically significantly associated with established clinical or surgical outcomes, such as the KOOS, PROMIS, ODI, WOMAC scores, implant survivorship, or reoperation rates of subjects.[33]
2. Comparison with Expert Assessments and Standard Methods: Benchmarking against radiologist or orthopedic surgeon grading, classical radiographic indices, and conventional biomarkers will help establish clinical relevance. AI biomarkers often outperform human reader variability by capturing high-dimensional texture and shape features.[34]
3. Incremental Value Analysis: Techniques such as NRI, decision curve analysis, and improvements in AUC quantify whether AI biomarkers add meaningful predictive value beyond age, sex, disease severity, or traditional imaging metrics.

5.1.2. Analytical validation and radiomic reliability

Here, consistency analysis of biomarker extraction is performed.

1. Repeatability and Reproducibility Testing: Repeatability is assessed with same scan, while reproducibility deals with different scanners or settings. ICCs, Bland–Altman analyses, and coefficients of variation are performed to study repeatability and reproducibility. Reliable biomarkers should maintain ICC >0.80 across acquisition settings.[36]
2. Harmonization Techniques: ComBat harmonization, intensity standardization, and feature normalization are methods that reduce variability dependent on the scanners. These are increasingly recognized as necessary steps in radiomics pipelines to ensure stability of AI-extracted features.[37]

5.1.3. Regulatory and reporting standards

The AI radiographic biomarkers have to adhere to evolving regulatory and reporting frameworks set globally.

1. TRIPOD, CONSORT-AI, SPIRIT-AI, and DECIDE-AI: These provide standardized checklists for reporting on prediction models, clinical trial protocols, and preliminary AI decision-support evaluations. Their application ensures that biomarkers are developed in ways that guarantee transparency, reproducibility, and interpretability.

2. FDA, EMA, and PMDA Considerations for Software as a Medical Device: Regulatory agencies emphasize algorithm explainability, the need for diversity in data sets, performance monitoring, and predefined protocol for change control of adaptive AI algorithms alone.[39]

5.1.4. Real-world validation and post-deployment surveillance

Real-world validation assesses performance in uncontrolled, high-variability clinical environments.

1. Prospective observational validation: Prospective multi-center trials allow determination of the predictive value of biomarkers in routine workflows, which ensure performance is reproduced outside curated research data sets.[40]
2. Drift detection and continuous performance monitoring: This monitoring after deployment helps the identification of model drift due to changes in patient population, imaging protocols, or surgical techniques. Such automated monitoring systems and updating pipelines maintain the long-term reliability of biomarkers.[41]

6. Discussion

AI-driven radiographic biomarkers represent a major advancement in orthopedic imaging, offering objective, reproducible, and high-dimensional quantitative features that can meaningfully predict clinical and surgical outcomes. Across osteoarthritis, trauma, spine disorders, inflammatory arthropathies, and arthroplasty, AI-enabled biomarkers have demonstrated superior performance compared to traditional radiographic grading systems, primarily due to their ability to capture subtle textural, shape-based, and biomechanical cues often missed by human observers.[8,14] A key theme emerging from the literature is the need to balance innovation with rigorous validation. While numerous studies report promising internal performance, external and multi-institutional validation still remains inadequate.

As shown repeatedly, model performance declines when applied to imaging from different scanners, vendors, or demographically distinct cohorts, underscoring the importance of diversification and harmonization.[9,12] Similarly, radiomic features—particularly those derived from plain radiographs—may suffer from instability due to variations in acquisition parameters and noise, highlighting the necessity for standardized pipelines and robust feature-selection strategies.[10,11] From a clinical perspective, AI biomarkers hold the potential to address long-standing challenges in orthopedics, such as early identification of fast-progressing osteoarthritis, objective quantification of fracture healing, risk stratification for postoperative complications, and timely detection of implant loosening or failure.[18] Integrating biomarker predictions into PACS-driven workflows, surgical planning systems, and triage pathways may improve efficiency and support evidence-based, personalized decision-making. However, challenges in explainability and clinician trust remain significant.

Orthopedic surgeons often require transparent reasoning to justify treatment decisions; opaque “black-box” models limit clinical confidence and hinder implementation. Emerging neuro-symbolic AI and interpretable deep learning architectures may provide a solution by combining high accuracy with meaningful visual and rule-based explanations.[5,7] Ethical and regulatory dilemmas also persist, especially regarding bias, data governance, and continuous model updates. Clearer international SaMD guidelines and hospital-level AI governance frameworks will be essential to ensure safety and maintain long-term performance stability.[6,13]

Looking ahead, multimodal integration—combining imaging with clinical, biomechanical, and biochemical datasets—may yield far more powerful biomarkers capable of predicting individualized trajectories and treatment responses.[19,20] Federated learning and large-scale collaborative networks can overcome data barriers while preserving privacy, enabling more diverse, generalizable model development.[12,21] Yet, real-world effectiveness can only be established through prospective studies demonstrating tangible improvements in patient outcomes, cost-efficiency, and workflow optimization.[16,17] Overall, AI-driven radiographic biomarkers are transitioning from experimental tools to clinically meaningful instruments. While challenges remain, the field is progressing rapidly toward scalable, trustworthy, and integrative solutions that can reshape the future of orthopedic care.

6.1. Challenges and Limitations

Along with the rapid evolution of AI-driven radiographic biomarkers in orthopedics, several methodological, clinical, technical, and regulatory limitations impede large-scale adoption and real-world translation. Most of these issues emphasize the need for detailed standardization, robust strategies regarding validation, and multidisciplinary collaboration.

6.1.1. Data quality, heterogeneity, and annotation bottlenecks

Radiographic datasets often suffer from heterogeneity in acquisition protocols, exposure parameters, and patient positioning, which significantly affects feature extraction and model stability. Variability across imaging vendors and institutions leads to domain shift, reducing generalizability of trained models.[8,12] High-quality annotation—especially for musculoskeletal imaging—requires expert radiologists or orthopedic surgeons, making the process time-consuming and costly. Inter-observer variability further complicates the creation of reliable ground truth labels.[14]

6.1.2. Risk of model over fitting and limited external generalizability

Many AI models exhibit strong performance within internal datasets but deteriorate when tested externally. Small, single-center training cohorts remain common in orthopedic AI studies, increasing risk of overfitting and limiting reproducibility.[9,15] Without large-scale, demographically diverse datasets, biomarker robustness across populations (age groups, ethnicities, comorbidity profiles) remains insufficiently validated.

6.1.3. Radiomic instability and sensitivity to imaging parameters

Radiomic features—especially texture-based descriptors—can be sensitive to noise, reconstruction kernels, beam angles, and exposure levels. This instability affects the reliability of biomarkers derived from plain radiographs, CT, and MRI.[10,16] Although harmonization and feature standardization techniques improve reproducibility, they are not universally applied, and their performance varies across modalities and anatomical regions.[11]

6.1.4. Explainability, interpretability, and clinical acceptance

Orthopedic surgeons require transparent decision-support tools that clearly link imaging biomarkers to clinically meaningful outcomes. Deep learning models, especially convolutional neural networks, often function as "black boxes," complicating interpretation of predictions.[5] The absence of standardized explainability frameworks—such as consistent saliency maps, feature attribution methods or decision logic visualization—limits clinician trust and hinders integration into surgical planning or patient discussions.[7]

6.1.5. Regulatory, ethical, and workflow integration barriers

The regulatory pathways of AI-enabled biomarkers are still hard to navigate and vary among different jurisdictions. International SaMD frameworks have strict requirements in terms of continuous monitoring, dataset representativeness, and software updates.[6] There are also ethical issues surrounding algorithmic bias, the privacy of imaging data, and performance drift over time. Again, integration with the AI system into PACS/RIS and orthopedic workflows remains difficult because of interoperability constraints and a lack of standardized deployment pathways.[13]

6.1.6. Limited prospective and real-world evidence

Most studies remain retrospective and lack prospective multi-centre validation. Real-world data are required to assess performance in routine clinical variability, particularly for emergency trauma imaging and postoperative follow-up.[17] A significant gap still exists between technical model performance and improvements demonstrated in patient-reported outcomes, surgical decision-making, or long-term implant survivorship.[18]

7. Future Directions and Opportunities

The future generation of AI-driven radiographic biomarkers is likely to transform diagnostics, prognostics, and personalized decision-making for orthopedics. Innovation, coupled with strong validation frameworks, will enable further improvements in accuracy, interpretability, and clinical integration. Several promising directions stand out.

7.1. Multimodal biomarkers integrating imaging, clinical, and genomic data

Future systems will go beyond single-modality analyses toward integrated models combining radiographs, CT, MRI, wearable sensor metrics, EHR data, and—when relevant—genomic or proteomic information. The advantages of multimodal fusion are multidimensional in capturing structural, biological, and functional aspects of musculoskeletal disease, which greatly improve predictive power.[19,20] Such models can enable precision orthopedics by forecasting individualized risk trajectories, treatment responsiveness, and postoperative recovery profiles.

7.2. Federated learning and privacy-preserving model development

Federated learning and secure computation techniques will allow researchers to train models collaboratively across institutions without sharing raw imaging data—a major barrier in orthopedics. These strategies reduce bias, increase dataset diversity, and maintain compliance with data privacy regulations while preserving high performance across scanners and demographics.[12,21]

7.3. Automated, real-time decision support in clinical workflows

AI biomarkers will increasingly be embedded directly within PACS/RIS systems to provide real-time risk assessments, annotated overlays, and prognostic alerts. Examples include automated predictions of osteoarthritis progression, fracture healing trajectories, implant loosening risk, or postoperative complications, enabling earlier and more targeted interventions.[18] Real-time integration will also support triage systems in emergency trauma care, streamlining care pathways and reducing diagnostic delays.

7.4. euro-symbolic AI and improved explainability

Advanced hybrid models, which incorporate deep learning with rule-based reasoning or neuro-symbolic AI, are improved for enhanced interpretability while sustaining accuracy. Such models could articulate the underlying radiographic patterns driving such predictions as quantifiable changes in trabecular microarchitecture, bone shape, or implant migration pattern, for a much-enhanced clinician trust and adoption.[5,7]

7.5. Standardization of radiomic pipelines and consensus biomarker panels

There is an increasing consensus among international expert consortia for standardized acquisition parameters, radiomic feature sets, harmonization protocols, and reporting guidelines. The consensus-driven biomarker panels for specific orthopedic conditions, such as early knee osteoarthritis, degenerative spine disease, or post-arthroplasty evaluation, will decrease the variability and enhance reproducibility of studies.[10,11]

7.6. Perspective trials and evidence-based clinical translation

1. Large-scale, prospective, multi-center studies are needed to confirm clinical utility and cost-effectiveness.
2. Future studies will determine whether AI biomarkers meaningfully enhance decision-making, precision of treatment, patients' satisfaction, and long-term outcomes beyond standard approaches.[16,17]
3. Integration of AI into pragmatic clinical workflows will be pivotal upon their safe deployment with automated quality checks, drift detection, and clinician feedback loops.

7.7. Adaptive and continuously learning biomarker systems

Adaptive AI models that can continuously learn from the inflow of new data, advances in imaging techniques, surgical innovation, and shifting populations will enable long-term model relevance. Such systems will have controlled updates monitored under regulated frameworks for retaining performance, managing drift, and ensuring stability in evolving clinical environments.[6,13]

8. Conclusion

AI-driven radiographic biomarkers are transforming orthopedic imaging by enabling precise, reproducible, and high-dimensional assessment of musculoskeletal structures. These biomarkers—derived from radiomics, deep learning, or hybrid approaches—have demonstrated substantial potential in predicting disease progression, fracture risk, implant performance, and postoperative outcomes across a wide range of orthopedic conditions. Robust development pipelines, standardized preprocessing, rigorous technical and clinical validation, and adherence to regulatory guidelines are critical to translating AI biomarkers into routine clinical practice. Future directions, including multimodal integration, federated learning, interpretable AI, and adaptive continuous learning, promise to enhance predictive accuracy, clinical trust, and personalized orthopedic care. Although challenges persist with regard to data heterogeneity, interpretability, and workflow integration, AI-driven radiographic biomarkers will likely become indispensable in evidence-based decision-making for orthopedic treatments by offering early intervention, optimized treatment planning, and improved patient outcomes.

9. Source of Funding

None.

10. Conflict of Interest

None.

References

1. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci Rep.* 2018;8(1):1727. https://doi.org/10.1038/s41598-018-20132-7.
2. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. *Radiology.* 2020;296(3):584-593. https://doi.org/10.1148/radiol.2020192091.
3. Tiulpin A, Saarakkala S. Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs Using Deep Convolutional Neural Networks. Diagnostics (Basel). 2020;10(11):932. https://doi.org/10.3390/diagnostics10110932.
4. Kijowski R, Fritz J, Deniz CM. Deep learning applications in osteoarthritis imaging. *Skeletal Radiol.* 2023;52(11):2225-38. https://doi.org/10.1007/s00256-023-04296-6.
5. Gurung B, Liu P, Harris PDR, Sagi A, Field RE, Sochart DH, et al. Artificial intelligence for image analysis in total hip and total knee arthroplasty : a scoping review. *Bone Joint J.* 2022;104-B(8):929-37. https://doi.org/10.1302/0301-620X.104B8.BJJ-2022-0120.R2.
6. Zhao H, Ou L, Zhang Z, Zhang L, Liu K, Kuang J. The value of deep learning-based X-ray techniques in detecting and classifying K-L grades of knee osteoarthritis: a systematic review and meta-analysis. *Eur Radiol.* 2025;35(1):327-40. https://doi.org/10.1007/s00330-024-10928-9.
7. Kim SE, Nam JW, Kim JI, et al. Enhanced deep learning model enables accurate alignment measurement across diverse institutional imaging protocols. *Knee Surg Relat Res.* 2024 ;36(1):4. https://doi.org/10.1186/s43019-023-00209-y.
8. Liu Z, Wu J, Gao X, Qin Z, Tian R, Wang C. Deep learning-based automatic measurement system for patellar height: a multicenter retrospective study. *J Orthop Surg Res.* 2024;19(1):324. https://doi.org/10.1186/s13018-024-04809-6.
9. Hwang D, Ahn S, Park YB, Kim SH, Han HS, Lee MC, et al. Deep Learning-Based Muscle Segmentation and Quantification of Full-Leg Plain Radiograph for Sarcopenia Screening in Patients Undergoing Total Knee Arthroplasty. *J Clin Med*. 2022;11(13):3612. https://doi.org/10.3390/jcm11133612.
10. Tolpadi AA, Lee JJ, Pedoia V, Majumdar S. Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images. *Sci Rep*. 2020;10(1):6371. https://doi.org/10.1038/s41598-020-63395-9.
11. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage.* 2007;15 Suppl A:A1-56. https://doi.org/10.1016/j.joca.2006.11.009.
12. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* 1957;16(4):494-502. https://doi.org/10.1136/ard.16.4.494.
13. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. *ArXiv.* .2014. https://arxiv.org/abs/1409.4842
14. Nguyen HH, Saarakkala S, Blaschko MB, Tiulpin A. Semixup: In- and Out-of-Manifold Regularization for Deep Semi-Supervised Knee Osteoarthritis Severity Grading From Plain Radiographs. *IEEE Trans Med Imaging.* 2020;39(12):4346-56. https://doi.org/10.1109/TMI.2020.3017007.
15. Babazadeh S, Dowsey MM, Bingham RJ, Ek ET, Stoney JD, Choong PFM. The long leg radiograph is a reliable method of assessing alignment when compared to computer-assisted

navigation and computer tomography. *Knee.* 2013;20(4):242-9. https://doi.org/10.1016/j.knee.2012.07.009.
16. Skyttä ET, Lohman M, Tallroth K, Remes V. Comparison of standard anteroposterior knee and hip-to-ankle radiographs in determining the lower limb and implant alignment after total knee arthroplasty. *Scand J Surg.* 2009;98(4):250-3. https://doi.org/10.1177/145749690909800411.
17. Federer SJ, Jones GG. Artificial intelligence in orthopaedics: A scoping review. *PLoS One.* 2021;16(11):e0260471. https://doi.org/10.1371/journal.pone.0260471.
18. Tiulpin A, Thevenot J, Rahtu E, Saarakkala S. A novel method for automatic localization of joint area on knee plain radiographs. *ArXiv.* 2017. https://arxiv.org/abs/1701.08991
19. Bugday B, Bingol H, Yildirim M, Alatas B. Enhancing knee osteoarthritis detection with AI, image denoising, and optimized classification methods and the importance of physical therapy methods. *Peer J Comput Sci.* 2025;11:e2766. https://doi.org/10.7717/peerj-cs.2766.
20. Favreau H, Chennen K, Feruglio S, Perennes E, Anton N, Vandamme T, et al. Knee Osteoarthritis Diagnosis: Future and Perspectives. *Biomedicines.* 2025;13(7):1644. https://doi.org/10.3390/biomedicines13071644.
21. Rajamohan HR, Wang T, Leung K. Prediction of total knee replacement using deep learning analysis of knee MRI. *Sci Rep*. 2023;13(1):6922. https://doi.org/10.1038/s41598-023-33934-1.
22. Jiang T, Lau SH, Zhang J, Chan LC, Wang W, Chan PK, et al. Radiomics signature of osteoarthritis: Current status and perspective. *J Orthop Translat*. 2024;45:100-6. https://doi.org/10.1016/j.jot.2023.10.003.
23. Kwolek K, Grzelecki D, Kwolek K, Marczak D, Kowalczewski J, Tyrakowski M. Automated patellar height assessment on high-resolution radiographs with a novel deep learning-based approach. *World J Orthop.* 2023;14(6):387-98. https://doi.org/10.5312/wjo.v14.i6.387.
24. Jamshidi A, Pelletier JP, Labbe A, Abram F, Martel-Pelletier J, Droit A. Machine Learning-Based Individualized Survival Prediction Model for Total Knee Replacement in Osteoarthritis: Data From the Osteoarthritis Initiative. *Arthritis Care Res (Hoboken).* 2021;73(10):1518-27. https://doi.org/10.1002/acr.24601.
25. Demehri S, Kasaeian A, Roemer FW, Guermazi A. Osteoarthritis year in review 2022: imaging. *Osteoarthritis Cartilage.* 2023;31(8):1003-1011. https://doi.org/10.1016/j.joca.2023.03.005.
26. Liawrungrueang W, Kim P, Kotheeranurak V, Jitpakdee K, Sarasombath P. Automatic Detection, Classification, and Grading of Lumbar Intervertebral Disc Degeneration Using an Artificial Neural Network Model. *Diagnostics (Basel).* 2023;13(4):663. https://doi.org/10.3390/diagnostics13040663.
27. Üreten K, Arslan T, Gültekin KE, Demir AND, Özer HF, Bilgili Y. Detection of hip osteoarthritis by using plain pelvic radiographs with deep learning methods. *Skeletal Radiol.* 2020;49(9):1369-74. https://doi.org/10.1007/s00256-020-03433-9.
28. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis.* 2014;73(7):1323-30. https://doi.org/10.1136/annrheumdis-2013-204763.
29. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594. https://doi.org/10.1136/bmj.g7594.
30. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiol Artif Intell.* 2020;2(2):e200029. https://doi.org/10.1148/ryai.2020200029.
31. Roemer FW, Demehri S, Omoumi P, et al. State of the Art: Imaging of Osteoarthritis-Revisited 2020. *Radiology.* 2020;296(1):5-21. https://doi.org/10.1148/radiol.2020192498.
32. Jia H, Zhang J, Ma K, Qiao X, Ren L, Shi X. Application of convolutional neural networks in medical images: a bibliometric analysis. *Quant Imaging Med Surg*. 2024;14(5):3501-3518. https://doi.org/10.21037/qims-23-1600.
33. Soliman MAS, Kelahan LC, Magnetta M, Savas H, Agrawal R, Avery RJ, et al. A Framework for Harmonization of Radiomics Data for Multicenter Studies and Clinical Trials. *JCO Clin Cancer Inform.* 2022;6:e2200023. https://doi.org/10.1200/CCI.22.00023.
34. Pati S, Kumar S, Varma A, Edwards B, Lu C, Qu L, et al. Privacy preservation for federated learning in health care. *Patterns (N Y).* 2024;5(7):100974. https://doi.org/10.1016/j.patter.2024.100974.
35. Kong SH. Incorporating Artificial Intelligence into Fracture Risk Assessment: Using Clinical Imaging to Predict the Unpredictable. *Endocrinol Metab (Seoul).* 2025;40(4):499-507. https://doi.org/10.3803/EnM.2025.2518.
36. Rouzrokh P, Ramazanian T, Wyles CC, Philbrick KA, Cai JC, Taunton MJ, et al. Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty From Postoperative Radiographs. *J Arthroplasty.* 2021;36(6):2197-203.e3. https://doi.org/10.1016/j.arth.2021.02.028.
37. Conrozier T, Tron AM, Balblanc JC, Mathieu P, Piperno M, Fitoussi G, et al. Mesure de l'interligne coxofémoral par un analyseur automatique d'images numérisées [Measurement of the hip joint space using automatic digital image analysis]. *Rev Rhum Ed Fr.* 1993;60(2):137-43.
38. Wu L, Wang B, Lin B, Li M, Wu Y, Si H, et al. A Deep Learning-Based Clinical Classification System for the Differential Diagnosis of Hip Prosthesis Failures Using Radiographs: A Multicenter Study. *J Bone Joint Surg Am.* 2025;107(16):1798-809. https://doi.org/10.2106/JBJS.24.01601.
39. Yu H, Mu Q, Wang Z, Guo Y, Zhao J, Wang G, et al. A study on early diagnosis for fracture non-union prediction using deep learning and bone morphometric parameters. *Front Med (Lausanne).* 2025;12:1547588. https://doi.org/10.3389/fmed.2025.1547588.
40. Ong W, Liu RW, Makmur A, Low XZ, Sng WJ, Tan JH, et al. Artificial Intelligence Applications for Osteoporosis Classification Using Computed Tomography. *Bioengineering (Basel).* 2023;10(12):1364. https://doi.org/10.3390/bioengineering10121364.
41. Wang T, Liu H, Zhao W, et al. Predicting knee osteoarthritis progression using neural network with longitudinal MRI radiomics, and biochemical biomarkers: A modeling study. *PLoS Med*. 2025;22(8):e1004665. https://doi.org/10.1371/journal.pmed.1004665.

Cite this article: Dubey S. Development and validation of AI-driven radiographic biomarkers for predicting orthopedic outcomes. *IP Int J Orthop Rheumatol*. 2025;11(2):62-70.